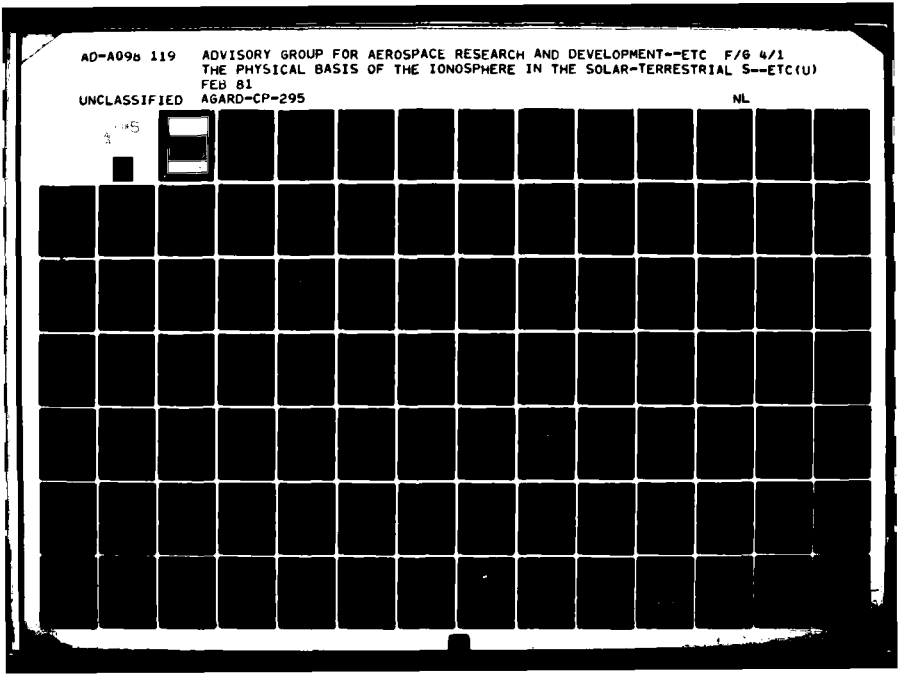
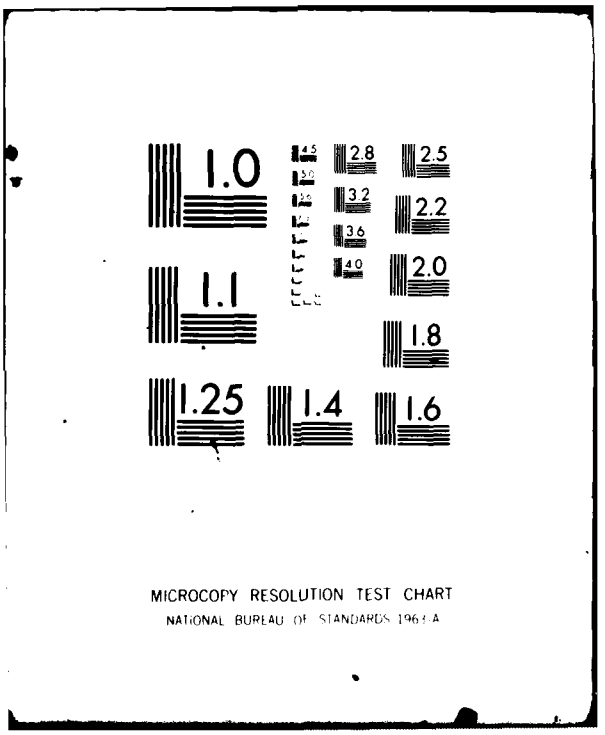
AD-A098 119 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 4/1
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL S--ETC(U)
FEB 81
UNCLASSIFIED AGARD-CP-295 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AGARD-CP-295

AGARD-CP-295

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No. 295

The Physical Basis of the Ionosphere in the Solar-Terrestrial System

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

CP 295
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM
EPP

AGARD-CP-295

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.295

THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM.

Conference proceedings.

Feb 81

400

Papers and Discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel held in Pozzuoli, Italy 28–31 October, 1980

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

Published February 1981

ISBN 92-835-0284-1

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

PREFACE

Our understanding of the solar-terrestrial environment has advanced enormously with the development of satellite technology. Satellites not only provide in-situ measurements of the atmosphere, ionosphere and magnetosphere, they are also able to monitor the incoming radiations and particles, and even provide, at the same time, much broader and denser geographical coverage of parameters that can be measured from the ground. Now that satellites have provided data of various kinds over a complete 22-year solar cycle, the Electromagnetic Wave Propagation Panel deemed it appropriate to hold a symposium on "The Physical Basis of the Ionosphere in the Solar-Terrestrial System". The meeting took place in Pozzuoli, Italy, 28–31 October 1980, and was basically dedicated to reviewing our current understanding of the processes that produce and control the distribution of ionization in the ionosphere as a component of the solar-terrestrial system, with some examples of modelling and prediction techniques that can be applied by terrestrial users of the electromagnetic spectrum.

The symposium consisted of six sessions and a round-table discussion. The first three sessions covered the solar source, primarily the ultraviolet emissions and solar wind, the production of the magnetosphere and ionosphere, coupling between the magnetosphere, ionosphere and atmosphere, dynamical processes, special features of the auroral regions and man-made perturbations. The last three sessions covered morphology, modelling, irregularities and predictions.

During the round-table discussion, the session chairmen summarized their views of the current state of knowledge, and work still to be done. The sources of ionization are now well understood. The seasonal variation is understood if the neutral atmosphere is taken as a lower boundary condition, so that to some extent, our ignorance is referred back to the atmosphere. There are still significant gaps in our understanding of coupling, dynamics and the magnetospheric current systems that feed significant amounts of energy to the ionosphere at auroral latitudes. Coupling effects are more important than is generally realized, since atmospheric ions have been detected throughout the magnetosphere, and solar wind ions are regularly seen in the ionosphere. The whole concept of a quasi-static system has given place to a magnetosphere in continual motion, even during "quiet" conditions, charging up with energy from the solar wind and discharging this energy downward in a geomagnetic substorm that typically occurs several times a day and is seen most spectacularly in the auroral regions, but has global consequences on the ionosphere.

These proceedings contain papers presented at the symposium and reproduced directly from copies furnished by the authors. I am pleased to acknowledge the help provided by session chairmen in submitting the summaries and editing the discussions. Thanks are due to the members of the program committee for their efforts in obtaining the speakers who provided such a broad range of topics.

Special thanks are due to the EPP Executive and his secretary for making all necessary arrangements before, during and after the meeting. We are also grateful to the Italian Delegate to AGARD and the Commander of the Accademia Aeronautica for providing such a delightful setting for the symposium.

ERWIN SCHMERLING

THEME

In connection with aerospace systems, detailed knowledge of the ionosphere is important for a number of purposes. When using ionospheric paths for communications and navigation, it is clear that frequencies, look-angles and fading rates must be estimated, and their short and long term variations understood. At very low and very high frequencies the ionosphere limits the accuracy of radars and navigation systems, and ionospheric irregularities limit bandwidths and communication rates. Further progress in modeling and prediction techniques requires new approaches which take into account other parts of the solar-terrestrial system that influence the ionosphere.

In recent years there has been increasing interest and significant progress in a broad class of problems linked together under the general title of "Solar terrestrial system studies". These start with the characteristics of the sun, and the temporal variations of its emissions. These emissions control the terrestrial environment in a very complex way. To first order, the solar wind particles control the magnetosphere, and the solar photons the ionosphere and atmosphere. These studies have shown, however, the importance of coupling between different parts of this system; especially between the hot magnetospheric plasma, the cool ionosphere plasma, and the atmosphere. Motions of both the hot and cool plasma across the geomagnetic field set up current systems and these, in turn, control dynamical processes in other parts of the "circuit". The closure of the current systems presents a non-trivial problem, and seems to be related to the storage and periodic precipitation of particles in substorms – an effect of particular importance in the high latitude ionosphere.

Ionospheric data are now available for several solar cycles, and a wealth of information is being developed concerning the sun, the inter-planetary medium, the magnetosphere and the atmosphere. The next advance in understanding the ionosphere will come by considering it in the context of the solar terrestrial system.

ELECTROMAGNETIC WAVE PROPAGATION PANEL

CHAIRMAN: Dr J.Aarons
AF Geophysics Lab.
Hanscom Field
Bedford, Mass.01731
USA

DEPUTY CHAIRMAN: Dr J.Belrose
Communications Research Center
Department of Communications
P.O.Box 11490, Station H
Ottawa K2H 8S2,
Canada

PROGRAM COMMITTEE

CHAIRMAN: Dr Schmerling

MEMBERS:
Prof. Bossy, BE
Prof. Cutolo, IT
Dr Gendrin, FR
Dr King, UK
Dr Nisbet, US
Prof. Rawer, FRG
Dr Risbeth, UK
Dr Thrane, NO

PANEL EXECUTIVE

Lt. Col. J.B.Catiller
AGARD/NATO
7, rue Ancelle
92200 Neuilly-sur-Seine
France

CONTENTS

* Not available at time of printing

SUMMARY

by

S.H.Gross
AGARD Meeting in Naples, Italy
Session I – The Sun, Solar Wind and Magnetosphere

This session encompassed four papers, two on solar radiation, one on the solar wind and one on magnetospheric and ionospheric flow in relation to the interplanetary magnetic field. Unfortunately, the third paper on the solar wind could not be presented. This summary is therefore on the papers presented.

The first paper entitled "The Solar Ultraviolet Source for the Ionosphere and Its Variation" was by Dr H.E.Hinteregger. It was directed at the formation of the ionosphere by solar EUV and its variation. He presented information on the considerable variability of the spectra as measured by the Atmospheric Explorer satellites during the present and previous solar cycles. With respect to the ionosphere he pointed out the importance of the effects of the variation of composition and the feedback effect of the solar radiation on the composition and distribution, as well. The relationship or non-relationship of solar EUV to popular non-EUV indices, such as $F_{10.7}$ and sunspot number, and the feasibility of EUV indices were discussed. He described the complexity of the problem of modeling and the need for more work.

The second paper was by Dr G.Schmidtke and was entitled "Solar Radiations and Their Interaction with the Terrestrial Environment". This paper was also concerned with the variability of solar radiation. Dr Schmidtke discussed the problem of determining changes in solar radiation from ionospheric measurements and noted that some large local fluctuations cannot be directly explained in this fashion, requiring consideration of dynamics as well. He stated that sources of energy originating from the solar wind, such as Joule heating, also compete with EUV radiation under some conditions. A further difficulty is the changes of parameters with time, such as heating efficiency.

The need for simpler indices more directly related to EUV was brought out in the questions following these two papers.

The last paper of the session was given by Dr S.W.H.Crowley and was entitled "Magnetospheric and Ionospheric Flow and the Interplanetary Magnetic Field". As the title suggests, Dr Crowley reviewed the relationship between the IMF and flow patterns in the ionosphere. Major differences in cell structures were described when the z-component of the IMF takes on negative and positive values. Asymmetries in the field and flow structures arising from the dawn-dusk component were also described, as well as the effects of a component in the sunward direction. Questions after the paper noted that not all ionospheric effects are directly correlated with the IMF; some associated with substorms appear to be much more complex.

THE SOLAR ULTRAVIOLET SOURCE FOR THE IONOSPHERE AND ITS VARIATION

Hans E. Hinteregger
Air Force Geophysics Laboratory, Hanscom Air Force Base
Bedford MA 01731, U.S.A.

ABSTRACT

Ion production in the ionospheric E and F regions is primarily due to solar EUV fluxes at wavelengths below 102.7 nm, where the flux ratios of maximum/minimum for the present solar cycle were found to range from values around 3 for the dominant chromospheric emissions to much higher values for coronal emissions, resulting in a factor of about 4 in the integrated flux of the range 14-102.7 nm. The vertical distribution of ion production depends not only on the incident solar EUV fluxes but also on the structure of the absorbing neutral atmosphere (thermosphere). As the latter is strongly affected by solar EUV fluxes above 130 nm (production of atomic oxygen and heating), the observed increase, with ratios ranging from about 1.2 around 175-185 nm to about 2.3 around 140 nm, is at least indirectly important in the solar-cycle variation of the ionosphere. EUV variability models of the ionosphere using non-EUV indices such as $F_{10.7}$, R_Z, or CaII plage indices are valuable as a crude guide only, but not for quantitative representations.

1. INTRODUCTION

Ionospheric physics, like all other branches of the physics of the earth's upper atmosphere, experienced an enormous enrichment by the advent of rockets and satellites used as powerful tools of experimental space research. The latter led to a fairly rapid progress in many areas of ionospheric research. Unfortunately, the progress toward establishing observational data on full-disk solar EUV fluxes, including all essential spectral detail as well as the various types of temporal variabilities, appears rather painfully slow from the viewpoint of the quantitative requirements of advanced theories of ionospheric physics and aeronomy in general. The aeronomical significance of this situation has been reviewed by various investigators (Hinteregger, 1970, 1976; Timothy, 1977; Schmidtke, 1978; Hinteregger, 1979).

Observations of solar EUV irradiance at wavelengths from 14 to 185 nm made by the so-called EUVS experiment (Hinteregger et al., 1973) on the satellites AE-C (1974-1978) and AE-E (1976-present) have revealed unexpectedly great differences between the present solar cycle 21 and the previous cycle 20. This is the first time that measurements by the same set of instruments monitored long-term variation of solar EUV irradiance over a period as long as that from the minimum through the maximum of a sunspot cycle.

1.1. Solar EUV Fluxes Incident on Top of the Earth's Atmosphere

The solar EUV fluxes incident on top of the earth's atmosphere, $I_{o\lambda}$ [photons m^{-2} sec^{-1}], responsible for the day-time ionization in the E, F_1, and F_2 regions, consist of wavelengths (λ) shortward of 102.7 nm [1 nm=10 Å], including many different types of emissions stemming from grossly different source regions in the solar atmosphere, mainly the chromosphere, transition region, and corona. Hence it is not surprising that both short-term and long-term variabilities of $I_{o\lambda}$ are generally strongly λ-dependent, often even within small wavelength intervals. The brightness distribution of virtually all solar EUV emissions across the visible part of Sun (the solar disk) is generally far from uniform, showing more or less drastic contrasts between quiet-disk areas and active-region areas. The latter are unimportant only under nearly ideal conditions of solar minimum such as those met by AE-E satellite spectrometer observations in the period of 13-28 July 1976. Both the neutral atmosphere and the ionosphere obviously respond only to the total so-called "full-disk fluxes".

Therefore, a division of $I_{o\lambda}$ into a "quiet-disk component" and an "active-region component", $I_{o\lambda}=I^{Q}_{o\lambda}+I^{A}_{o\lambda}$, might appear superfluous for aeronomical applications. However, some distinction is important for the understanding of the variability of the full-disk fluxes, e.g. the often strikingly different patterns of 27-day variations for different periods of solar activity. A similar distinction appears even in the formulations of the conventional thermospheric and ionospheric models which do not include any actual EUV data input, but use the most readily available quantitative index of solar activity, i.e. the 2800 MHz solar radio flux. Using the latter one distinguishes a daily index, $F_{10.7}$ (same or previous day), from a slowly variable index, $\langle F_{10.7} \rangle$ (e.g. defined as 81-day running mean value of $F_{10.7}$).

1.2. The Intertwinement of Neutral Atmospheric and Ionospheric Effects

The earth's ionosphere is commonly divided into the D, E, F_1, and F_2 regions identified by layers or ledges in the vertical profile of electron density. This ionospheric designation refers to the charged particles consisting of various ion species (i) and electrons. Their distribution is described in terms of densities, $n^{(i)}[m^{-3}]$, $n^{+}=\sum_i n^{(i)}$, and $n_e[m^{-3}]$, and effective values of temperatures, $T^{(i)}$ and T_e, where the latter are generally not the same as the kinetic temperature of the neutral atmospheric gas, T_n. The atmospheric medium in this height range called ionosphere is indeed only weakly ionized, i.e. most of the particles are neutral. The distribution of the neutral particle densities, $n_j[m^{-3}]$, in the so-called thermosphere (neutral atmosphere above the level of the temperature minimum called mesopause, which occurs around 90 km) reflects thermal equilibrium, $T_j=T_n$, and approaches gravitational diffusive equilibrium above about 115-125 km ("turbopause") at least for the three major constituents, j=1(O), 2(N_2) and 3(O_2). This means that the corresponding partial pressures have scale heights inversely proportional to the particle mass, $H_j=kT/m_jg$. This explains the well known dominance of atomic oxygen at greater heights in the upper thermosphere, i.e. in the topside ionosphere.

The intertwinement of neutral atmospheric and ionospheric effects caused by variations of solar EUV is clearly apparent in the formulation of the production rate of ions of any given species (i) resulting from EUV absorption by a given neutral particle species (j).

$$R_j^{(i)}[m^{-3}sec^{-1}] = \int \sigma_{\lambda j}^{(i)}\, n_j(P) I_\lambda(P)\, d\lambda \qquad (1)$$

where $\sigma_{\lambda j}^{(i)}$ is the photon interaction cross section of j-particles absorbing radiation of wavelength λ resulting in the creation of a certain ion species (i). This production is in general not simply a function of the incident solar EUV, $I_{o\lambda}$, but involves the optical depth of the atmospheric column extending from the given reference point (P) in the solar direction, $\tau_\lambda(P)$. This means that the local photon flux density $I_\lambda(P)$, appearing in equ. (1) above,

$$I_\lambda(P) = I_{o\lambda}\exp[-\tau_\lambda(P)] \qquad (2)$$

in general involves attenuation ratios, $I_\lambda(P)/I_{o\lambda}$, which are different for different wavelengths, since the optical depth, $\tau_\lambda(P)$, is more or less strongly wavelength-dependent. This is due to the dependence of τ_λ on the absorption cross sections of the absorbing neutral species in the atmospheric column extending from P in the solar direction (s), given as

$$\tau_\lambda(P) = \int \sum_j \sigma_{\lambda j} n_j(s)\, ds \qquad (3)$$

where $\sigma_{\lambda j}$ is the total absorption cross section of j-particles for wavelength λ, i.e. $\sigma_{\lambda j} = \sum_i \sigma_{\lambda j}^{(i)} + \sigma_{\lambda j}^{(N)}$ (sum of all ionizing as well as non-ionizing photon absorption cross sections).

This implicit involvement of the neutral atmospheric effects in the total ionospheric response to EUV variability increases in complexity as one progresses from the highest ionospheric regions (topside F_2) downward. The simplest situation exists in the upper F_2 region, at altitudes sufficiently high to neglect the attenuation for all relevant EUV wavelengths, so that one can set $I_\lambda(P)$ practically equal to $I_{o\lambda}$. The probability of the solar EUV induced creation of one ion of a given species (i) per unit time from one neutral particle of a given species j, called production rate coefficient or ionization frequency of that ion species (i) $q_j^{(i)}=R_j^{(i)}/n_j$, is then given as

$$q_j^{(i)}\Big|_{\text{all } \tau_\lambda \to o} = q_{oj}^{(i)} \equiv \int \sigma_{\lambda j}\, I_{o\lambda}\, d\lambda \qquad (4)$$

For this low-attenuation region, $\tau_\lambda \ll 1$, these so-called "unattenuated ionization frequencies" then become pure functions of the incident EUV fluxes, $I_{o\lambda}$, as shown in equ. (4) above.

Proceeding to lower altitudes in the F_2, F_1, and E region, one meets a much more complicated situation. This complication is caused mainly by two effects, i.e. (a) the spectra of the local EUV fluxes, $I_\lambda(P)$, now become functions of altitude which are generally different for different wavelengths (λ-dependent τ_λ) and (b) atomic oxygen starts losing its role as the only important absorber. Therefore, the local ionization frequencies, $q_j^{(i)}$, are then no longer given by the simple expression of equ. (4) but decrease below the value of $q_j^{(i)}(\tau_\lambda \to o)$ for decreasing altitudes below the F_2 peak. The functions of their altitude dependence are then also generally different for different wavelengths. A detailed discussion of these complications is obviously not within the scope of the present paper.

1.3. Significance of EUV Wavelength Range of 120-195 nm

The ionospheric D region is distinctly different from the ionosphere above. Ranging from about 40 to 90 km, this ionospheric region is mainly due to (a) the photoionization of a minority constituent of the neutral atmosphere (NO ionized by solar H Ly-α 121.6 nm) as well as (b) the photoionization of N_2 and O_2 by solar x-rays of wavelengths below about 1 nm. As the present review is limited to effects of solar EUV of wavelengths ranging from about 14 to 185 nm, it will touch D-region problems but briefly, i.e. by referring to the variability of the solar H Ly-α flux.

The great importance of variations in the neutral atmospheric structure induced by the absorption of solar EUV of all wavelengths including those not capable of creating ionization, is perhaps best illustrated by the photodissociation of O_2 in the Schumann continuum ($\lambda < 175$ nm) and the Schumann-Runge bands (175-195 nm) showing a major impact on the composition and on the temperature profile of the lower thermosphere and upper mesosphere, respectively.

2. REPRESENTATIONS OF THE SOLAR EUV SPECTRUM AND ITS VARIATIONS

The present section first deals with the requirements of spectral detail (Section 2.1.) and then briefly reviews "EUV-Variability Models" of various categories (Sections 2.2.-2.6.). The entire Section 2 is based on a paper on the same topic presented at the most recent COSPAR meeting (Hinteregger, 1980a).

2.1. Requirements for Spectral Details in EUV Flux Data

The aeronomical (ionospheric) requirements for details of the solar EUV spectrum can be divided into two categories, in accordance with typically different objectives summarized as follows:

(a) The practically desirable use of a reasonably small number of EUV-variability indices for an at least approximately correct representation of the temporal variability of the incident fluxes, I_o, for all of the known solar emission wavelengths throughout the total range of interest, say for $\lambda \leq 200$ nm (or at least for the range of 14-185 nm which covers the ionospherically most important part and also represents the range of actually existing long-term observations for the present solar cycle 21).

(b) The scientifically compelling need to account for the known spectral details of EUV photon-interaction cross sections, at least for the major constituents of the upper atmosphere.

Attempts at approaching the latter objective (b) by the crude substitution of "effective values" of cross sections for a very small number of very wide wavelength intervals, pursued throughout the earlier stages of ionospheric and aeronomical theories, have been essentially rejected in the more recent developments of computer-aided, quantitative theoretical models. For the latter, the actual use of known details in cross-section data is obviously much more sensible than any simplistic substitution of poorly defined mean values. Therefore, the statement of objective (b) now certainly does not imply that the available data on $\sigma_{\lambda j}$ and $\sigma_{\lambda j}^{(i)}$ should be converted into some compressed data set listing only the mean values for some small number of wavelength intervals. This means that a compression, if deemed necessary at all, is now considered suitable for quantitative ionospheric-thermospheric applications only if a sufficiently large number of wavelength intervals is used. This situation is illustrated by the selection of 37 wavelength intervals shown in Table 1 which has been extracted from a recent paper on ionization frequencies (Torr, et al., 1979). These wavelength intervals are essentially the same as those used by Heroux and Hinteregger (1978) in their summary table for the aeronomical reference spectrum F74113 which actually includes records for 1957 different wavelengths. Compared with the latter, the number of 37 wavelength intervals used in Table 1 is obviously small. On the other hand, this same number would appear rather unattractively large from the viewpoint of the first objective, (a) above.

Regarding the first objective, (a) above, from the viewpoint of thermospheric model developments, one can hardly ignore the very remarkable success of models such as MSIS (Hedin et al., 1977a,b) or the Jacchia Models (e.g. Jacchia, 1977) which have used no more than two variability indices for the representation of the "EUV effect". This success appears even more remarkable, if one appreciates that the two most commonly used variability indices, $F_{10.7}$ (daily value for same or previous date) and $\langle F_{10.7} \rangle$ (e.g. defined as 81-day mean value centered on given date), express the solar radio flux at 2800 MHz, which indeed penetrates the upper atmosphere without any physical interactions.

Shortcomings inherent in this simple procedure of using $F_{10.7}$ and $\langle F_{10.7} \rangle$ as "EUV" indices have been discussed by many investigators in the past. However, the level of verified inconsistencies between actual aeronomical observations and corresponding model predictions is apparently not serious enough to advocate

the use of a very large number of solar variability indices for the practical approach to objective (a). Hence, a subdivision of the EUV spectrum into the 37 intervals of Table 1(a,b), with the same large number of interval-peculiar variability indices, would hardly be desirable. On the other hand, an even finer subdivision into significantly more than 37 intervals would be needed to avoid the inclusion of solar emissions of clearly different variability character within any interval. For instance, interval 3 of Table 1(a,b) would require a number of different variability indices for different parts of that interval, 15.0 - 20.0 nm, which is known to be dominated by emissions of FeIX, FeX, FeXI, FeXII and FeXIII, with relative variability amplitudes spread over a range in excess of a factor of two.

Table 1.† Irradiance and Cross Sections (O,N_2) for 37 Wavelength Intervals (5-105 nm)

(a) INCIDENT SOLAR EUV FLUXES 10^{13} photons. m^{-2} sec^{-1}

(b) IONIZATION AND TOTAL ABSORPTION CROSS SECTIONS OF O AND N_2 (effective values for given wavelength interval) in units of megabarn [1 megabarn = 10^{-22} m^2]

Interval	λλ/nm	Ident.	AVG.JUL 1976	19 FEB 1979	$O^+(4S)$	$O^+(2D)$	$O^+(2P)$	$O^+(4P)$	$O^+(2P^*)$	O(ABS)	N_2(ABS)	N_2(ION)
1	5-10		.4382	1.3710	.32	.34	.22	.10	.03	1.06	.60	.60
2	10-15		.1687	.4675	1.03	1.14	.75	.34	.27	3.53	2.32	2.32
3	15-20		1.8692	5.7024	1.62	2.00	1.30	.58	.46	5.96	5.40	5.40
4	20-25		1.3951	7.1448	1.95	2.62	1.70	.73	.54	7.55	8.15	8.15
5	25.6	HeII, SiX	.5064	1.0832	2.15	3.02	1.95	.82	.56	8.43	9.65	9.65
6	28.415	FeXV	.0773	5.7229	2.33	3.39	2.17	.89	.49	9.26	10.60	10.60
7	25-30		1.3556	12.1600	2.23	3.18	2.04	.85	.52	8.78	10.08	10.08
8	30.331	SiXI	.6000	4.6908	2.45	3.62	2.32	.91	.41	9.70	11.58	11.58
9	30.378	HeII	7.7625	14.3956	2.45	3.63	2.32	.91	.41	9.72	11.60	11.60
10	30-35		.8671	6.8315	2.61	3.98	2.52	.93	.00	10.03	14.60	14.60
11	36.807	MgIX	.7394	1.5355	2.81	4.37	2.74	.92	.00	10.84	18.00	18.00
12	35-40		.2121	2.5423	2.77	4.31	2.70	.92	.00	10.70	17.51	17.51
13	40-45		.4073	1.5310	2.99	4.75	2.93	.55	.00	11.21	21.07	21.07
14	46.522	NeVII	.3299	.7358	3.15	5.04	3.06	.00	.00	11.25	21.80	21.80
15	45-50		.3081	1.8228	3.28	5.23	3.13	.00	.00	11.64	21.85	21.85
16	50-55		.5085	1.6486	3.39	5.36	3.15	.00	.00	11.91	24.53	24.53
17	55.437	OIV	.7992	1.5163	3.50	5.47	3.16	.00	.00	12.13	24.69	24.69
18	58.433	HeI	1.5875	4.3005	3.58	5.49	3.10	.00	.00	12.17	23.20	23.20
19	55-60		.4843	1.0477	3.46	5.30	3.02	.00	.00	11.90	22.38	22.38
20	60.976	MgX	.6333	2.4838	3.67	5.51	3.05	.00	.00	12.23	23.10	23.10
21	62.973	OV	1.8484	3.8701	3.74	5.50	2.98	.00	.00	12.22	23.20	23.20
22	60-65		.4002	1.3672	3.73	5.50	2.97	.00	.00	12.21	23.22	23.22
23	65-70		.2623	.5388	4.04	5.52	.47	.00	.00	10.04	29.75	25.06
24	70.331	OIII	.3915	.7461	4.91	6.44	.00	.00	.00	11.35	26.30	23.00
25	70-75		.1667	.4287	4.20	3.80	.00	.00	.00	8.00	30.94	23.20
26	76.515	NIV	.1997	.4386	4.18	.00	.00	.00	.00	4.18	35.46	23.77
27	77.041	NeVIII	.2425	1.1873	4.18	.00	.00	.00	.00	4.18	26.88	18.39
28	78.936	OIV	.7831	1.5140	4.28	.00	.00	.00	.00	4.28	19.26	10.18
29	75-80		.8728	2.4541	4.23	.00	.00	.00	.00	4.23	30.71	16.75
30	80-85		1.9311	4.8538	4.38	.00	.00	.00	.00	4.38	15.05	.00
31	85-90		4.4325	12.2187	4.18	.00	.00	.00	.00	4.18	46.63	.00
32	90-95		3.6994	9.8513	2.12	.00	.00	.00	.00	2.12	16.99	.00
33	97.702	CIII	4.8400	10.2165	.00	.00	.00	.00	.00	.00	.70	.00
34	95-100		1.7155	4.0779	.00	.00	.00	.00	.00	.00	36.16	.00
35	102.572	HI	4.3750	11.8519	.00	.00	.00	.00	.00	.00	.00	.00
36	103.191	OVI	1.9425	6.1049	.00	.00	.00	.00	.00	.00	.00	.00
37	100-105		2.4775	6.0928	.00	.00	.00	.00	.00	.00	.00	.00

$F_{10.7}$	($10^{-22}Wm^{-2}Hz^{-1}$)	68.0	243.0
$<F_{10.7}>$	($10^{-22}Wm^{-2}Hz^{-1}$)	71	202

†all data of Table 1 (a,b,c) have been extracted from a recent paper by Torr, et al. (1979).

(c) UNATTENUATED IONIZATION FREQUENCIES [sec^{-1}] FOR MAJOR IONOSPHERIC SPECIES

	O_2^+	N_2^+	He^+	O^+-TOT.	$O^+(4S)$	$O^+(2D)$	$O^+(2P)$	$O^+(4P)$	$O^+(2P^*)$
13-28 JUL 1976	5.43 x10^{-7}	3.89 x10^{-7}	4.89 x10^{-8}	2.80 x10^{-7}	1.09 x10^{-7}	0.96 x10^{-7}	0.56 x10^{-7}	1.34 x10^{-8}	6.17 x10^{-9}
19 FEB 1979	1.73 x10^{-6}	1.28 x10^{-6}	1.93 x10^{-7}	9.35 x10^{-7}	3.39 x10^{-7}	3.25 x10^{-7}	1.94 x10^{-7}	5.34 x10^{-8}	2.45 x10^{-8}

These considerations lead to the conclusion that the method of subdividing the EUV spectrum into wavelength intervals is a fundamentally poor approach to objective (a) above.

The parts (a) and (b) of Table 1 illustrate only the range of ionizing solar EUV radiation of wavelengths shortward of 102.7 nm (i.e. omitting the D region sources, H Ly-α at 121.6 nm, as well as x-rays of $\lambda\lambda \lesssim 1.0$ nm. Results of recent calculations of ionization frequencies of the major thermospheric constituents as a function of solar cycle 21 (Torr et al., 1979) are abstracted here in Table 1 (c), restricted

to a comparison of the solar minimum period, represented by data averages for 13-28 July 1976, with the near-maximum conditions of solar activity of 19 February 1979.

2.2. Model Representations of EUV Variability

The various possible representations of EUV variations by models using a certain number of daily indices can be divided into categories depending on the physical nature of the quantities used as indices and on the number of such indices actually used. For instance, one may regard the conventional indices, $F_{10.7}$ and $\langle F_{10.7} \rangle$ either as two different indices or as two different quantities based on the same basic index, i.e. $F_{10.7}$. However, the more important distinction is that of identifying both $F_{10.7}$ and $\langle F_{10.7} \rangle$ as "Non-EUV Indices". Other Non-EUV Indices of solar activity, actually used but rarely in aeronomical studies, are the daily values of Zürich sunspot numbers or the CaII plage-area index.

The other major category of variability models, introduced only recently, is that of using actual "EUV Indices", defines as a set of daily values of the incident solar full-disk flux for some given EUV wavelength(s) or wavelength interval(s). The development of variability models based on such indices derived from actual EUV flux observations originated only very recently (Schmidtke, 1978; Hinteregger, 1980a). Much more detailed investigations remain to be completed before one may formulate a sufficiently matured model that should eventually replace essentially all of those presently used variability models which use Non-EUV Indices.

The basic characterization of any EUV-variability model requires not only a statement of the type of daily indices used as variables but also a detailed statement of the specific formulations converting the input of these daily indices into a sufficiently detailed list of EUV flux values for the given date. The most important part of the second statement is represented by a list of adjustment constants. The size of this array of constants will of course depend on the degree of spectral detail. Since the latter indeed should involve more than the 37 wavelength groups of Table 1 (for reasons already discussed in the preceeding section of this paper), these types of variability models are obviously recommended only for computer-aided investigations. Accepting this fact, one is left with no serious objection against the actual use of a fully detailed reference spectrum.

2.3. "Non-EUV Indices" Used "Without EUV-Observation Fits"

The traditional type of EUV-variability representation in terms of the daily indices $F_{10.7}$ and $\langle F_{10.7} \rangle$ used in the well-established thermospheric models (e.g. MSIS or Jacchia Models) actually by-passes the entire problem of spectral detail in both solar EUV emission and terrestrial atmospheric absorption. They define specific formulations for the conversion of the daily indices into the desired model values of thermospheric constituent densities and temperatures, using a list of adjustment constants determined empirically by fitting to a very large data base which include satellite-drag observations and results of in-situ measurements of thermospheric quantities, none of which are formally related to any characteristics of the EUV spectrum.

The recently published "Two-Component Model" called CBV below (Cook et al., 1980) expresses variations of solar EUV irradiance in the range of 117.5-210 nm simply as a function of the fractional coverage of the solar disk by active regions. The authors of the CBV model actually do not recommend it for any quantitative aeronomical applications, particularly in the range of wavelengths below about 120 nm, where the CBV model would indeed belong to the same category as the traditional models.

2.4. "Non-EUV Indices" Used "With EUV-Observation Fits"

The most dissatisfactory aspect of the treatment of the "EUV-effect" in the traditional atmospheric models (see Section 2.4) has been the use of mean values of the solar radio flux at 2800 MHz taken over a long period such as that of 81 days centered on the given date of the desired calculation of atmospheric model quantities. The authors of these models have often emphasized that they never suggested that the atmospheric response to solar variations could really extend 40 days into the past, let alone the nonsensical assumption of any response to 40 days of the future. Hence one had to accept the use of $\langle F_{10.7} \rangle$ as a purely pragmatically justified method of accounting for the long-term variability in solar activity, as distinguished from day-to-day variability or the well known variations with the 27-day period of solar rotation.

Our preliminary study of AE-E observations during 1976-1979 has included regression analysis for several important solar EUV emissions (Hinteregger, 1980a). These analyses were carried out for two types of associations with 2800 MHz fluxes, i.e. (A) fitting F_λ (EUV) data to a linear expression of the type $\phi_A = A_o + A_1 F_{10.7}$ and (B) fitting to the two-component expression, $\phi_B = B_o + B_1 \langle F_{10.7} \rangle + B_2 (F_{10.7} - \langle F_{10.7} \rangle)$. The comparison of the correlation coefficients for the two fit types showed that ϕ_B is clearly superior, thus proving that EUV fluxes indeed correlate with the parameter $\langle F_{10.7} \rangle$ more strongly than with the daily $F_{10.7}$. This result provides the first physically meaningful explanation for the indisputably great success of thermospheric models such as MSIS or the Jacchia Models.

The numerical values of the coefficients B_o, B_1 and B_2 of the fit function ϕ_B are of course different for different wavelengths. This is illustrated here by giving a numerical example for the Ly-α lines of hydrogen and HeII. Using a preliminary AE-E data base limited to the EUV observing period from 3 June 1977 to 14 August 1979, we obtained $B_o = 22627$, $B_1 = 3885$ and $B_2 = 740$ for H Ly-α; $B_o = 3782$, $B_1 = 57.5$ and $B_2 = 21.7$ for "30.4 nm" (blend of HeII and SiXI lines shown separately in Table 1). With these coefficients one can then calculate "predicted" values of EUV fluxes (ϕ_λ) for any date by inserting the corresponding data of $F_{10.7}$ and $\langle F_{10.7} \rangle$ into the expression of ϕ_B noted above [using the conventional units of 10^{-22} W m^{-2} sec^{-1} for $F_{10.7}$ we obtain ϕ_λ in units of 10^{10} photons m^{-2} sec^{-1}]. This predicted value of ϕ_λ is obviously not recommended for quantitative application to any date for which the value of the actually observed flux, F_λ, is available. The actual agreement between ϕ_λ and F_λ is considered surprisingly good in a statistical sense (Hinteregger, 1980a) but may be quite poor for some specific dates or entire subsections of dates even if one stays within the period of the observations of F_λ to which the coefficients of ϕ_B have been fitted. On the other hand, the agreement is sometimes surprisingly good even for some prediction date outside the original fitting period. For instance, using the aforementioned coefficients for Ly-α and 30.4 nm to calculate ϕ_B for 10 November 1979 ($F_{10.7} = 381$, $\langle F_{10.7} \rangle = 224$, i.e. $F_{10.7} - \langle F_{10.7} \rangle = 157$) we found ϕ (Ly-α) $\doteq 1.01 \times 10^{16}$ and ϕ (30.4 nm) $\doteq 2.01 \times 10^{14}$ [ph m^{-2} sec^{-1}] in surprisingly good agreement with actual measurements for that date giving F(Ly-α) = 9.15×10^{15} and F(30.4 nm) = 2.07×10^{14}. This numerical example is given here only for the purpose of illustrating the importance of the $\langle F_{10.7} \rangle$-associated term which is seen to exceed the other variable term quite substantially, in spite of the rather large magnitude of $F_{10.7} - \langle F_{10.7} \rangle = 157$. For that limited purpose, the quality of the preliminary incomplete data base was certainly adequate.

2.5. EUV Class-Variability Model

The EUV class-variability model (CVM), discussed most recently in some detail elsewhere (Hinteregger, 1980a), is being developed mainly for the aeronomically desirable purpose of transforming a fully detailed EUV reference spectrum, $F_{o\lambda}$, into equally detailed flux data for some other date, $F_{D\lambda}$, without actually requiring new data for each of the thousand or more individual wavelengths listed. Since the best available set of long-term satellite observations of solar EUV fluxes is related to the present solar cycle 21, it seemed undesirable to use F74113 (Heroux and Hinteregger, 1978) as base-line reference spectrum for AE-E data analysis (F74113 applies to 23 April 1974, i.e. still within cycle 20). Therefore the development of our CVM is based on a detailed reference spectrum names F76REF which applies to the period of 13-28 July 1976 which was selected for two reasons. First, the day-to-day variations within that period of spotless sun with no visible active regions were indeed found to be practically negligible, allowing a great enhancement of statistical significance of even weak line signals by averaging all good instrumental scan records obtained during that period. Secondly, it seemed most profitable theoretically to work with a reference spectrum reflecting the simplest possible solar conditions, where the full-disk fluxes include essentially no contributions of active regions.

Letting the subscript K = 0, 1, 2,...designate a class identification, defining $F_{o\lambda}$ as the F76REF value of flux for any given wavelength λ [also given class K(λ)], and denoting the corresponding updated flux as $F_{D\lambda}$, we define our variability model by the relation

$$R = F_{D\lambda}/F_{o\lambda} = 1 + (R_K - 1)C_\lambda \qquad (5)$$

where $K = K(\lambda)$ and C_λ are listed for each λ in F76REF and R_K is the ratio of $F_{D\lambda}/F_{o\lambda}$ for one or more wavelengths listed with $C_\lambda = 1$ and same class K. Hence, the quantity

$$R_K = \left(\frac{F_{D\lambda}}{F_{o\lambda}} \right)_{C_\lambda = 1} \qquad (6)$$

will be briefly called "class ratio" below. The physical implication of this relation is obvious, i.e. the

fluxes in all emissions (λ) of the same class (K) and same adjustment constant (C_λ) are assumed to vary identically.

The primary goal of our CVM development is to express EUV flux variabilities in terms of a reasonably small number of variable class parameters, R_K, *in connection with sets of adjustment parameters*, C_λ, treated as constants (over a reasonably limited date range). The first phase of this development (Hinteregger, 1980a) was based on the definition of only three variability classes (K = 0, 1, 2) with the following "key wavelengths" (defined by setting $C_\lambda = 1$): λ_K(K=0) = $\lambda\lambda$ 177.5-185 nm (quasi-continuum); λ_K(K=1) = 58.433 nm (HeI); λ_K(K=2) = 33.541 nm (FeXV$^+$); assigning class K=1 to the various solar emissions with ionization potentials E_i <200 eV and K = 2 for those with $E_i \geq 200$ eV. The class K = 0 consists of those emissions in the range 110-185 nm identified as "quasi-continuum", denoted as QUASI(C) in F76REF (as well as F74113). The continuing development of our EUV variability model will aim at an improved trade-off between simplicity and accuracy. This trade-off condition is inherent in the method of determining best-fitting values of C_λ for some selected period of existing observations. If this period is extended over a truly long time span, say 2-3 years, the best-fitting C_λ values indeed differ more or less significantly from those found by fitting a shorter period of observations [see Table 3 of paper by Hinteregger (1980a)]. It is hardly surprising that the model representations for the shorter periods were found to be somewhat more accurate.

2.6. Critique of Simplified Model Concepts

The specific form of equ.(5) might suggest the adoption of an attractively simple model, which regards the increase of full-disk solar EUV fluxes after 1976 as a straightforward consequence of the increasing number of active regions, with C_λ playing the role of products of active-region areas and specific relative contrast ratios of active-region/quiet-disk brightness, i.e. assuming that there is no solar-cycle variation of the quiet-disk component itself and also implying that there is no long-term variability in these contrast ratios. However, this model can be justified only as a greatly simplified concept. For instance, the assumption of an invariant quiet-disk radiance appears to be in conflict with a number of completely independent observations such as those reported by Hinteregger (1977, 1979), by Mount et al., (1980) [$\lambda\lambda$ >120 nm], and by Vidal-Madjar and Phissamay (1980) [H Ly-α].

3. IRRADIANCE VARIATIONS RELATED TO THE PHASE IN SOLAR CYCLE AND SOLAR ROTATION

The compilation of relevant solar irradiance data by Delaboudiniere et al., (1978) covers the period of the solar cycle 20, showing intercomparison graphs of results obtained by the various investigators. The emphasis of this intercomparison was clearly on the long-term variation rather than day-to-day variations or the so-called 27-day variation due to solar rotation. Considerable amplitudes of the latter are still apparent in the AE-C satellite observations throughout the year 1974, i.e. relatively close to the end of the solar cycle 20 [see Fig. 1 of paper by Hinteregger et al. (1977)]. The irradiance minimum of solar chromospheric emissions between the activity cycles 20 and 21 appears to have occurred already around 20 April 1975, i.e. remarkably sooner than the minimum in the monthly mean value of the Zürich sunspot number around June-July 1976. This early minimum of solar chromospheric EUV, first identified from AE-C satellite observations (Hinteregger, 1977) has been confirmed by AEROS-B satellite data reported by Schmidtke [included in Fig. 1b of paper by Hinteregger (1980a)] as well as by OSO-5 satellite observations of the H Ly-α flux reported most recently by Vidal-Madjar and Phissamay (1980).

A graphic reproduction of the aforementioned features of the declining phase of cycle 20 and the transition toward cycle 21 is shown only on slides in oral presentation of this paper, but has been omitted here in the interest of conserving enough space to illustrate EUV variations for the present solar cycle 21 for which much better observations have been accumulated.

3.1. Overview of the Development of Solar Cycle 21

The appearance of a new solar cycle is certainly not a sudden phenomenon. There is a relatively long period of several years during which ground observations of specific sunspots and H_α-plage regions are known to be classified either as "old-cycle activity" or as "new-cycle activity". Therefore, the declaration of a specific month, let alone a specific date, as "the solar minimum" or as "the start of the new solar cycle" is meaningful only in reference to some specific value of one of the conventional activity indices (R_Z,$F_{10.7}$), or, where available, the observed variation in the solar full-disk flux of EUV for some specified wavelength or for a specified category of solar emissions. For instance, the aforementioned finding of an EUV minimum in April 1975 applies only to chromospheric emissions, whereas the minimum of

EUV irradiance for coronal emissions appears to have occurred around June-July 1976, i.e. essentially coinciding with the minimum seen in the smoothed values of the conventional indices of solar activity, $<R_Z>$ and $<F_{10.7}>$. Our reasons for selecting the period of 13-28 July 1976 as base-line reference for studies of the solar cycle 21 (Hinteregger, 1980a,b) have been explained in Section 2.5. above.

The present solar cycle appears to have gone through its "main maximum" for both the conventional indices ($<F_{10.7}>$, $<R_Z>$) and EUV irradiance during roughly the same period, say November 1979 - February 1980. The solar EUV "pre-maximum" of January - February 1979 already came fairly close to the main-maximum in the EUV fluxes of essentially all wavelengths. AE-E observations of the development of cycle 21 are summarized in Fig. 1 (reproduced from a recent paper (Hinteregger, 1980b)). The full range of solar EUV wavelengths observable by the EUVS experiment on AE-E is illustrated in simplified graphical form by Fig. 2 below (Hinteregger, 1980a).

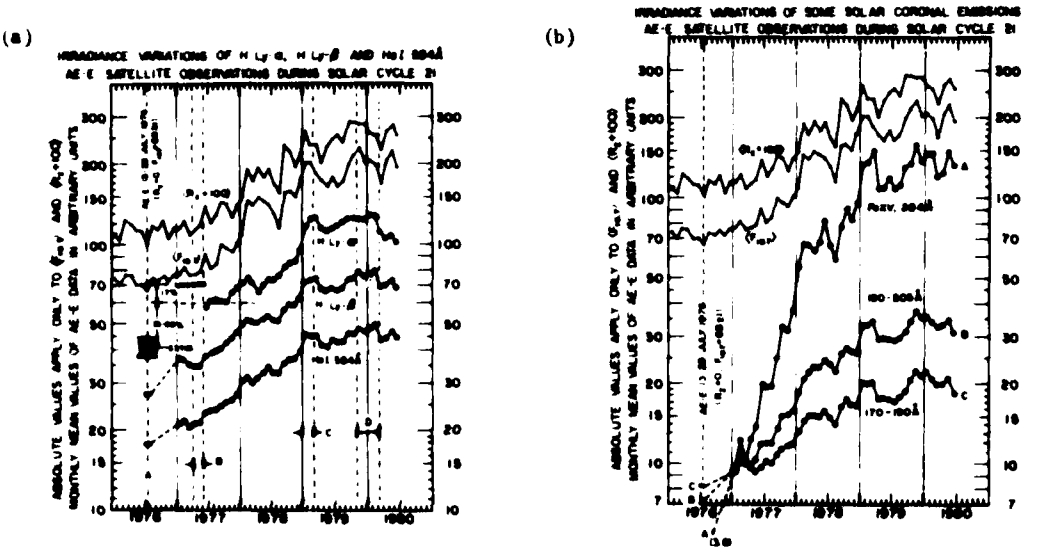


Fig. 1 Monthly mean values of solar irradiance for six different EUV fluxes are represented by the solid circles shown for the period of January 1977 through June 1980. The points were calculated as 31-day mean values (centered on the middle of the month) of all existing AE-E observations (i.e. covering about 60-70% of the dates of a given 31-day interval). The bars on the time line marked A show the corresponding data for the observations of spotless solar conditions during the period of 13-28 July 1976. The connections to data circles of January 1977 are drawn as straight dashed lines only, reflecting a regrettably long period of instrument-diagnostical turnons, new command-sequence developments, and most unfavorable conditions of orbital-solar geometry during the scheduled exposures. Both diagrams (a) and (b) include plots of the associated quantities of $<F_{10.7}>$ and $<R_Z>$ on the same absolute ordinate scale. This logarithmic ordinate scale has no absolute meaning for the EUV plots, which have been shifted to provide a good view of the temporal variations, completely ignoring the actual intensity relationships among these six EUV fluxes.

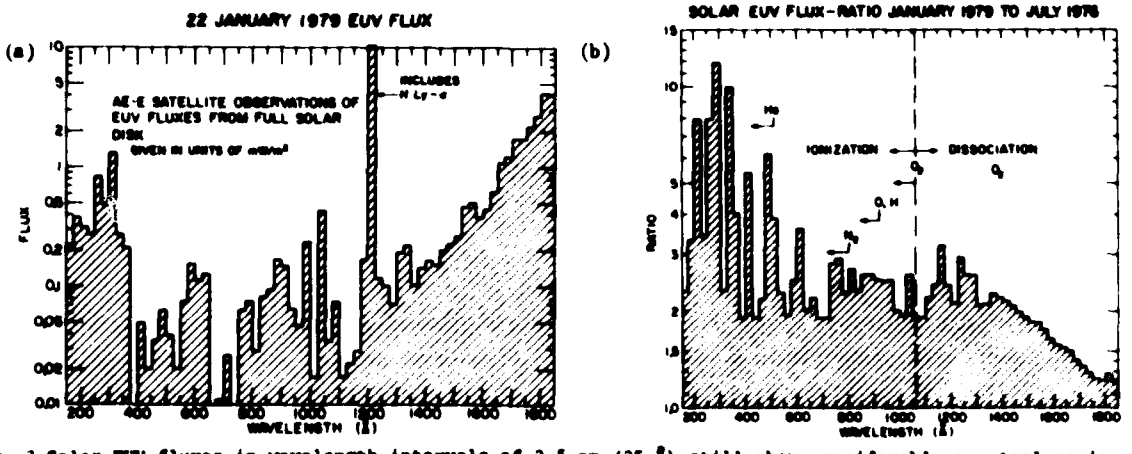


Fig. 2 Solar EUV fluxes in wavelength intervals of 2.5 nm (25 Å) still show considerable spectral variations illustrated by diagram (a) which represents incident energy flux densities [mW m^{-2}/2.5nm] for each interval, obtained for a specific date of relatively high fluxes within the "pre-maximum" period of cycle 21. Diagram (b) illustrates the ratios of these fluxes relative to our reference period of 13-28 July 1976. Appraisals of the importance of the spectral region of 120-185 nm have been given in various recent reports (Torr et al., 1980; Cook et al., 1980; Mount et al., 1980; Hinteregger, 1980b).

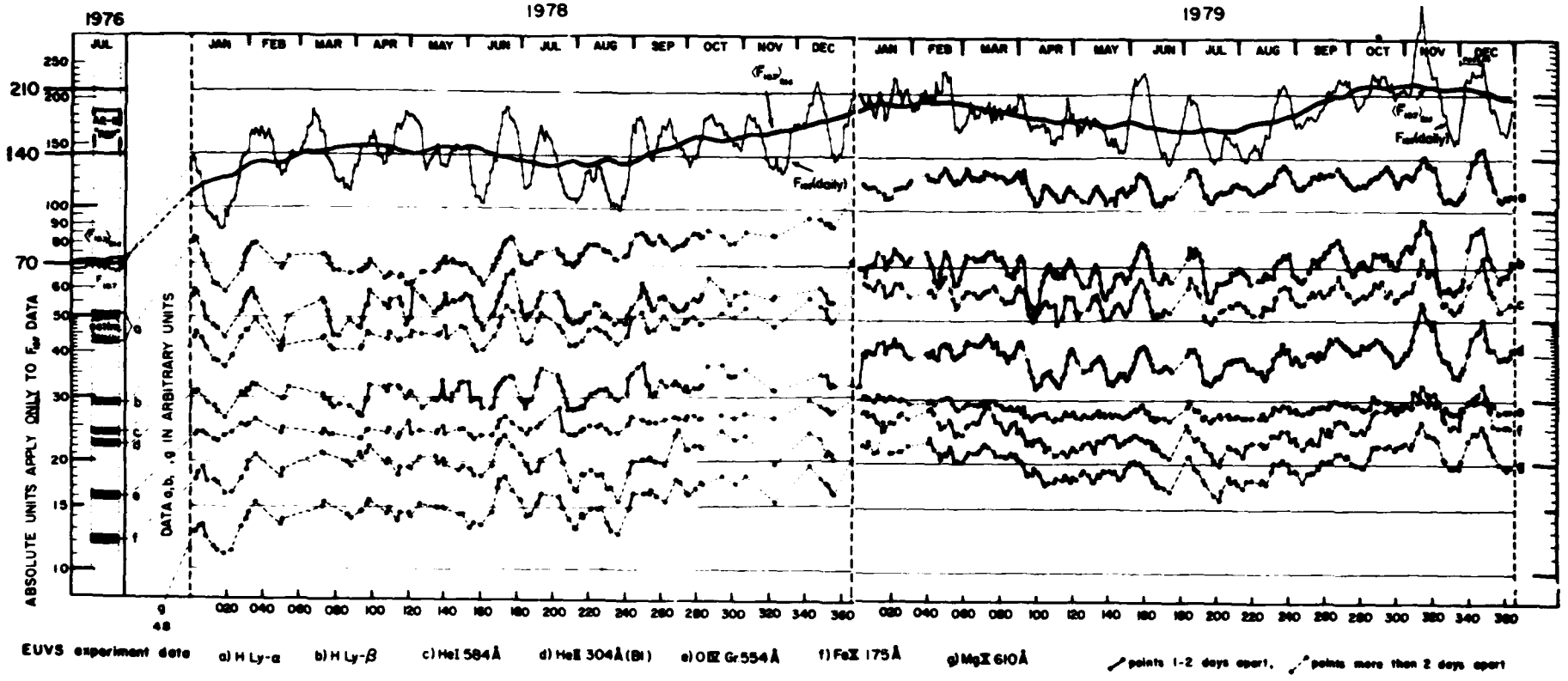


Fig. 3 AE-E satellite observations of solar EUV irradiance at wavelengths ranging from 14-185 nm (140-1850 Å) had been scheduled for one orbit per day of actual turnon, with a duty cycle adjusted to spacecraft power conditions (depending on solar-orbital configuration) resulting in an average of approximately 30% of dates for which no measurements exist. The plots for the seven selected EUV wavelengths a, b,...,g. show full circles for all existing daily data, illustrating the rather non-uniform distribution of actual turnon duty cycles. The top of the diagram includes the full-disk solar radio flux at 2800 MHz in units of 10^{-22} W m^{-2} Hz^{-1} showing both the daily values and the 81-day sliding mean values. Computer-listed data in absolute units for these 7 wavelengths, covering all useful AE-E observations in the period from 3 June 1977 to the end of 1979 as well as the reference values for July 1976 (indicated graphically at the left side of the diagram) was included (file named $EUVFLX) as part of a data package on magnetic tape released to the National Space Science Data Center for public availability. A preliminary supplement ($EUVFLXA) for the period 1 January - 12 September 1980 is presently available only by access to the AE-dedicated computer (printout copies may be obtained from the author). The present diagram has been assembled from illustrations prepared for recent paper (Hinteregger, 1980a).

3.2. Variation of Daily Values of EUV Irradiance

The observed changes in the daily average values of solar EUV irradiance such as that depicted for the period of 1978-1979 in Fig. 3 are well known to be primarily due to the non-uniform solar-longitude distribution of relatively bright localized source regions (active regions, plage areas). If both the quiet and the active regions remained invariant, all full-disk EUV fluxes would exhibit variations with exactly the same periodicity of about 27 days per solar rotation. The real situation is obviously much more complicated, for a number of well-known reasons related to (a) the long-term variations in the average fractional plage-area coverage of the solar surface, showing rather drastic differences between the maximum and the minimum of a solar cycle, (b) the variations in the brightness of any given active region during its lifetime which is generally different for different active regions, and (c) the poorly understood but apparently significant variations in the so-called quiet-disk brightness which may remain unimportant for several years within the same solar cycle, but should not be neglected over an entire solar cycle or the transition between two solar cycles, e.g. the period of 1974 - 1976.

4. CONCLUSIONS

Variations in the neutral, thermospheric structure caused by variations in the incident solar EUV fluxes, $I_{o\lambda}$, lead to important modifications in the purely ionospheric response to these EUV fluxes. The complexity of this intertwinement increases from the topside F_2 region downward. The accounting for the important details of wavelength-dependent ionization cross sections, at least in the form illustrated by Table 1(b), has now become characteristic of any advanced ionospheric studies. A generally less-satisfactory situation appears to exist in the treatment of the aforementioned intertwinement with the neutral atmospheric response. Even otherwise advanced recent ionospheric studies had to resort to grossly simplifying assumptions such as adoption of an invariant neutral atmosphere model for the study of the ionospheric response to solar EUV variations. This obvious shortcoming has been clearly due to the very slow progress toward more satisfactory representations of the variabilities of the spectrum of the incident solar EUV fluxes for quantitative aeronomical use. The presently accomplished progress in this direction (Hinteregger, 1980a) still reflects only an exploratory stage of development.

Acknowledgements. The EUVS experiment was sponsored by the National Aeronautics and Space Administration (NASA DPR S50030-AG). The author wishes to thank his colleagues at AFGL, K. Fukui, for invaluable help in many parts of this study. The work as a whole would have been impossible without the excellent support of scientific data analysis by analysts of Computer Sciences Corporation under the expert leadership of B.R. Gilson.

REFERENCES

Cook, J.W., Brueckner, C.E., and VanHousier, M.R., 1980, "Variability of the solar flux in the far ultraviolet 1175-2100 Å", J.Geophys. Res. 85, 2257

Delaboudinière, R.P., Donnelly, R.F., Hinteregger, H.E., Schmidtke, G., and Simon, P.C., 1978, "Intercomparison/compilation of relevant solar flux data related to aeronomy (Solar cycle 20)", COSPAR Technique manual series, No. 7

Hedin, A.E., Salah, J.E., Evans, J.V., Reber, C.A., Newton, G.P., Spencer, N.W., Kayser, D.C., Alcayde, C., Bauer, P., Cogger, L., and McClure, J.P., 1977a, "A global thermospheric model based on mass spectrometer and incoherent scatter data, MSIS 1,Composition", J. Geophys. Res., 82, 2139

Hedin, A.E., Reber, C.A., Newton, G.P., Spencer, N.W., Brinton, H.C., Mayr, H.G.,and Potter, W.E., 1977b, "A global thermospheric model ....,MSIS 2,Composition", J. Geophys. Res. 82, 2148

Heroux, L., and Hinteregger, H.E., 1978, "Aeronomical reference spectrum for solar UV below 2000 Å", J. Geophys. Res., 83, 5305

Hinteregger, H.E., 1970, "The extreme ultraviolet solar spectrum and its variation during a solar cycle", Ann. Geophys., 26, 547

Hinteregger, H.E., Bedo, D.E., and Manson, J.E., 1973, "The EUV spectrophotometer on Atmosphere Explorer", Radio Science, 8, No.4, 349

Hinteregger, H.E., 1976, "EUV fluxes in the solar spectrum below 2000 Å", J. Atm. & Terr. Phys., 38, 791

Hinteregger, H.E., Bedo, D.E., Manson, J.E., and Skillman, D.R., 1977, "EUV flux variations with solar rotation observed during 1974-1976 from the AE-C satellite", COSPAR Space Research XVII, 533

Hinteregger, H.E., 1977, "EUV flux variation during end of solar cycle 20 and beginning cycle 21, observed from AE-C satellite", Geophys. Res. Letters, 4, 231

Hinteregger, H.E., 1979, "Development of solar cycle 21 observed in EUV spectrum and atmospheric absorption", J. Geophys. Res., 84, 1933

Hinteregger, H.E., 1980a, "Representations of solar EUV fluxes for aeronomical applications", COSPAR Space Research XXI (in press)

Hinteregger, H.E., 1980b, "AE-E experiences of solar irradiance monitoring for 1200-1850 Å", in Proceedings of Workshop on Solar UV Monitoring held 31 July -1 August 1980 in Boulder, Colorado (in press)

Jacchia, L.G., 1977, "Thermospheric temperature, density, and composition; new models" Smithsonian Astrophysical Observatory, Special Report No. 375

Mount, G.H., Rottman, G.J., and Timothy, J.G., 1980, "The solar spectral irradiance 1200-2550 Å at solar maximum", J. Geophys. Res., 85, 4271

Schmidtke, G., 1978, "Todays knowledge of the solar EUV output and the future needs for more accurate measurements for aeronomy", Planet. Space Sci.,26, 347

Timothy, J.G., 1977, "The solar spectrum between 300 and 1200 Å", in The Solar Output and Its Variation (ed. O. White), University of Colorado Press, pp 237-259

Torr, M.R., Torr, D.G., Ong, R.A., and Hinteregger, H.E., 1979, "Ionization frequencies for major thermospheric constituents as a function of solar cycle 21", Geophys. Res. Letters, 6, No.10, 771

Torr, M.R., Torr, D.G., and Hinteregger, H.E., 1980, "Photodissociation frequency for O_2 by Schumann-Runge continuum absorption as a function of solar cycle 21", J. Geophys. Res. (in press)

Vidal-Madjar, A., and Phissamay, B., "The solar Lα flux solar minimum", 1980, Solar Physics 66, 259

SOLAR RADIATIONS AND THEIR INTERACTIONS WITH THE TERRESTRIAL ENVIRONMENT

G. Schmidtke

Fraunhofer-Institut für Physikalische Messtechnik
Heidenhofstrasse 8
D-7800 Freiburg, FRG

SUMMARY

Most of the energy of the earth atmosphere is originated from solar electromagnetic radiation and from solar wind. The measurement of these parameters and the tracing of the energy conversions are rather difficult especially in view of the solar variability. Therefore, quantitative results in the energy ranges of interest for different atmospheric layers could be achieved in the last few years. Not only the measurement of the solar wind parameters but also of the solar electromagnetic radiation in the different spectral ranges must be considered of being at an early stage, yet. The variability of the radiations and some aspects of these competing energy sources for the terrestrial atmosphere will be discussed.

1. INTRODUCTION

Variations of the solar electromagnetic radiation are predicted since about a century due to evidence of climatological changes with solar cycle. Long series of mountain top measurements were performed to define the range of possible variation. Out of those, the measurements of Charles Abbot are of special interest. By reevaluating the data collected from about 1908 through 1952, Aldrich and Hoover (1954) limited the range of possible variations up to about ± 1 %. A closer limit could not be defined in view of the uncertainties of the variability of radiation absorbing atmospheric parameters.

The search for the variability of the solar electromagnetic radiation has been strongly promoted by the modern technological developments of experimental platforms

aboard balloons, aircrafts and rockets. In addition the importance of particle radiation from the sun and its variation could be measured by satellites. Some aspects of these competing energy sources for the terrestrial atmosphere will be discussed here.

2. VARIABILITY OF THE SOLAR IRRADIANCE

One of the best known sun-related terrestrial effects is the ionization enhancement in the ionospheric D-region during solar flares and during high levels of solar activity caused by increased X-ray emission from active regions of the sun. This radiation can be measured by relatively simple devices such as Geiger counters or ionization chambers as developed in the laboratory since long. For these reasons X-rays were measured aboard many rockets and satellites exhibiting a very high degree of variability ranging from two orders of magnitude during flares up to five orders of magnitude and more on a solar cycle time scale (Fig. 1). Of special interest are satellite measurements with the longest series of measurements performed by the Naval Research Laboratories. Aboard the SOLRAD 11 satellites broadband sensors were operated during solar minimum through solar maximum conditions of the solar cycle 21 (Horan and Kreplin, 1980). Currently the geostationary satellites GOES-2, 3 and 4 are monitoring the solar X-rays within the spectral band of 0.05 - 0.4 nm and 0.1 - 0.8 nm with samples every three seconds. The data evaluation is still underway.

In the soft X-ray spectral region between 2.5 - 15 nm very few measurements are known (Fig. 2). Fig. 3 delineates the variability derived from a careful study of these measurements.

A survey of the EUV spectral region is presented in the preceding talk by Hinteregger describing the most exciting results derived from the experiments aboard Atmospheric Explorers-C and E. Combining these and the data from the satellites AEROS-A and B (Schmidtke, 1976; Schmidtke et al., 1977) solar EUV flux data on a daily base are available now almost continuously from end of 1972 through today (Hinteregger, 1980). Between 15 - 125 nm the integral variability ranges up to about 25 % during very strong solar flares, up to 50 % during very pronounced solar rotations and up to about a factor of three during the solar cycle 21.

Very few experiments are performed in the spectral region from 125 - 180 nm (Delaboudiniere et al., 1978). From Atmospheric Explorers-C and E data are collected from solar minimum conditions from 1974 through solar maximum in 1979 exhibi-

ting solar cycle variations up to about a factor of two at 140 nm down to a change of about 25 % at 180 nm (Hinteregger, 1980).

For the solar cycle 21 a maximum variation of about 20 % at 200 nm is estimated by Simon (1980). Changes up to 3 % are measured with solar rotation for the same wavelength. These values strongly decrease with increasing wavelengths. However, for wavelengths longer than 200 nm the experimental accuracy still exceeds the expected long-term solar variability resulting in a lack on data representing solar cycle variations. On a medium-term scale e.g. for solar rotations, data on a relative scale show about 1 % variation at 250 nm important for the ozone layer.

For longer wavelengths similar measurements with high accuracy are not known, since data including the visible spectral range primarily represent the solar "constant" containing infra-red and ultra-violet spectral regions as well. Though the total flux from the sun was determined for many years, so far, there is no indication for a solar cycle variation (Fröhlich, 1980). However, during the last solar minimum no measurements were performed anymore after the solar "constant" was considered to be known. On the other hand, a long-term steady decrease of the total solar flux by 0,03 % per year from 1966 through 1980 seems to be possible to derive from balloon and satellite observations being correlated with a decrease in the global mean temperature (Fröhlich, 1980).

On a medium-term base total flux decreases up to 0.2 % over seven to ten day periods (Fig. 4) are reported (Willson et al., 1980) being highly correlated with the development of specific sunspot groups. These data showing significant fluctuations even on an orbit-to-orbit base (96 min per orbit) represent the first high precission spacecraft measurements. They are carried out aboard the Solar Maximum Mission by the Active Cavity Radiometer Irradiance Monitor experiment implementing statistical uncertainties as small as $\pm$0.001 %. Measurements and data analysis are still going on.

3. SOLAR RADIATIONS AS ATMOSPHERIC ENERGY SOURCES

After solar EUV radiation was predicted by astronomers, ion-electron production and ionization equilibrium profiles were calculated on a relative base before the ionosphere was discovered (Rawer, 1981). Later the interpretation of ground-based ionospheric observations required a quantitative determination of the ionizing radiation requesting "ultraviolet excess factors" up to 10^6 with respect to black body radiation. Though the measured electron densities could be explained by a

total ionizing energy flux of about 0.1 mWM^{-2} because of a compensating error in the recombination coefficient, the EUV flux measured directly by rocket experiments (Hinteregger et al., 1965) was almost two orders of magnitude larger.

Ionospheric observations gave strong indications for the variability of the solar EUV radiation on a short-term (flares) and long-term base. However, a quantitative determination of solar radiation changes derived from ionospheric measurements is still not possible due to the complexity of the series of processes involved. On the other hand, since direct radiation measurements revealed changes of this important energy source, a model study could be performed (Roble and Schmidtke, 1979) to compare variations of ionospheric/atmospheric parameters observed and of those calculated from a model of the ionospheric E- and F-regions with a constant neutral atmospheric model (Figs. 5 and 6) interacting with a variable solar EUV radiation. The latter data are based on the AEROS measurements (Schmidtke et al., 1977). Similar conclusions were drawn when the MSIS atmospheric model (Hedin et al., 1977) was used applied to the time period of interest (Schmidtke, 1979): Changes of the solar EUV flux cannot directly explain the large local fluctuations e.g. of electron density and electron temperature as measured by numerous experiments. Ionospheric and atmospheric dynamics seem to have stronger impacts on these parameters, though on a global scale similar studies are yet to be done. Probably this will be a topic for the next future.

Another problem is related to the energy budget of the upper atmosphere. If the energy loss by airglow is neglected and if all heat sources other than EUV radiation are ignored, the exospheric temperature T_{exos} and the solar EUV input "at the top" of the upper atmosphere, I_o, could be related by (Bauer, 1973)

$$T_{exos} \propto I_o^{1/s}$$

with the parameter s depending on the constituents. This relation could not yet be proofed by experimental data.

The semi-empirical models of the upper atmosphere based on a wealth of satellite data are trying to take into account solar EUV radiation and geomagnetic energy sources originated mostly from solar particle radiation. Both solar energy sources cannot be monitored from ground observations. Also direct measurements in space on a routine base are not yet planned. Therefore, two indices, $F_{10.7}$ and A_p, both derived from ground measurements are used as substitutes to represent the solar heat sources in the atmospheric models. Though the physical background of these

two indices and the solar radiations they are supposed to represent is quite different, the atmospheric models (e.g. Jacchia, 1971; Hedin et al., 1977) describe the atmospheric parameters quite well for most of the applications. However, at certain times especially during solar minimum conditions, the agreement between exospheric temperatures derived from EUV absorption measurements and those computed from the MSIS atmospheric model are not satisfactory. For example, for the period of AEROS-A (December 1972 through August 1973) the total solar EUV flux decreased by about 30 % with the declining solar cycle 20 (Schmidtke, 1976). The exospheric temperatures derived from EUV extinction data of the same experiment neither reflect the long-term decrease of the solar EUV irradiance nor the medium-term changes with solar rotations. Geomagnetic activity is controlling the exospheric temperatures stronger. Also, for higher values of the index Ap the MSIS model represents too low temperatures for the same periods. Similar observations are made by Hinteregger (1979). At the beginning solar cycle 21 when the solar EUV radiation increased significantly exospheric temperatures were rather constant on a long-term scale (Hinteregger, 1978).

The decrease of energy of the EUV radiation during the mission of AEROS-A was for the whole earth of the order 10^{11} Watts, which is about the order of magnitude for the energy involved in a geomagnetic storm. Thus Joule heating and particle dissipation of solar wind origin are competing with solar EUV radiation as upper atmospheric heat sources at least during solar minimum conditions.

It is rather difficult to investigate the influence of the different heat sources in the upper atmosphere separately in view of the dynamics of the earth-atmospheric system. In addition, most of the parameters involved change with time. For example the heating efficiency of the solar EUV radiation is not constant (Torr et al., 1981), neither with atmospheric height nor with time (Fig. 7). Since the energy conversion seems to be more efficient for higher exospheric temperatures and vice versa, this parameter must be taken into account for a quantitative analysis of the energy budget of the upper atmosphere. This latter is not yet possible to be performed, because too little is known concerning the atmospheric heating efficiency of the solar wind and its variability and the role of the atmospheric airglow. These fields are yet to be investigated in more detail.

REFERENCES

ALDRICH, L.B., and HOOVER, W.H., 1954, Ann. Ap. Obs. Smithonian Inst. 7, 165.

BAUER, S.J. Physics and Chemistry in Space, 1973, Springer-Verlag Berlin, 3.

DELABOUDINIERE, J.P., DONNELLY, R.F., HINTEREGGER, H.E., SCHMIDTKE, G., and SIMON, P.C., 1978, Intercomparison/Compilation of relevant Solar Flux Data related to Aeronomy, COSPAR Technique Manual Series No. 7, 4.

FRÖHLICH, C., 1980, Solar Radiation and its Variation with Time, 14th ESLAB Symposium, Scheveningen.

HEDIN, A.E., SALAH, J.E., EVANS, J.V., REVER, C.A., NEWTON, G.P., SPENCER, N.W., KAYSER, D.C., ALCAYDE, D., BAUER, P., LOGGER, L., and McCLURE, J.P., 1977, A global thermospheric Model based on Mass Spectrometer and Incoherent Scatter Data, J. Geophys. Res. 82, 2139.

HINTEREGGER, H.E., HALL, L.A., and SCHMIDTKE, G., Solar XUV Radiation and Neutral Particle Distribution in July 1963 Thermosphere, Space Research V, 1175.

HINTEREGGER, H.E., 1978, private communication.

HINTEREGGER, H.E., 1979, Private communication.

HINTEREGGER, H.E., 1980, The Solar Ultraviolet Source for the Ionsphere and its Variations, AGARD Symposium, Naples.

HORAN, D.M., and KREPLIN, R.W., 1980, Simultaneous Measurements of EUV and Soft X-ray solar flare emission, 14th ESLAB-Symposium, Scheveningen.

JACCHIA, L.G., 1971, Smithonian Astrophysical Observatory Report 332, 21.

RAWER, K., 1981, Energy Sources for the Ionosphere - a Survey, Space Research XXI (in press).

ROBLE, R.G., and G. SCHMIDTKE, 1979, Calculated Ionospheric Variations due to Changes in the Solar EUV Flux measured by the AEROS Spacecraft, J. Atmosph. Terr. Phys. 41, 153.

SCHMIDTKE, G., 1976, EUV Indices for Solar-Terrestrial Relations, Geophys. Res. Let., 3, 573.

SCHMIDTKE, G., RAWER, K., BOTZEK, H., NORBERT, D., and HOLZER, K., 1977, Solar EUV Photon Fluxes Measured Aboard AEROS-A, J. Geophys. Res., 82, 2423.

SCHMIDTKE, G., 1979, Change of Primary Ion-Electron Production Rates with Solar EUV Flux, Ann. Géophys. 35, 141.

SIMON, P.C., 1980, Solar Irradiance between 120 and 400 nm and its Variations, 14th ESLAB Symposium, Scheveningen.

TORR, M.R., RICHARDS, P.G., and TORR, D.G., 1981, Solar EUV Energy Budget of the Thermosphere, Space Research XXI (in press).

WHITE, O.R., 1977, The Solar Output and its Variation, Colorado Associated University Press, Boulder, 306.

WILLSON, R.C., GULKIS, S., JANSSEN, M., HUDSON, H.S., and CHAPMAN, G.A., 1980, Observations of solar Irradiance Variability, 14th ESLAB-Symposium, Scheveningen.

IV SOLAR SPECTRUM

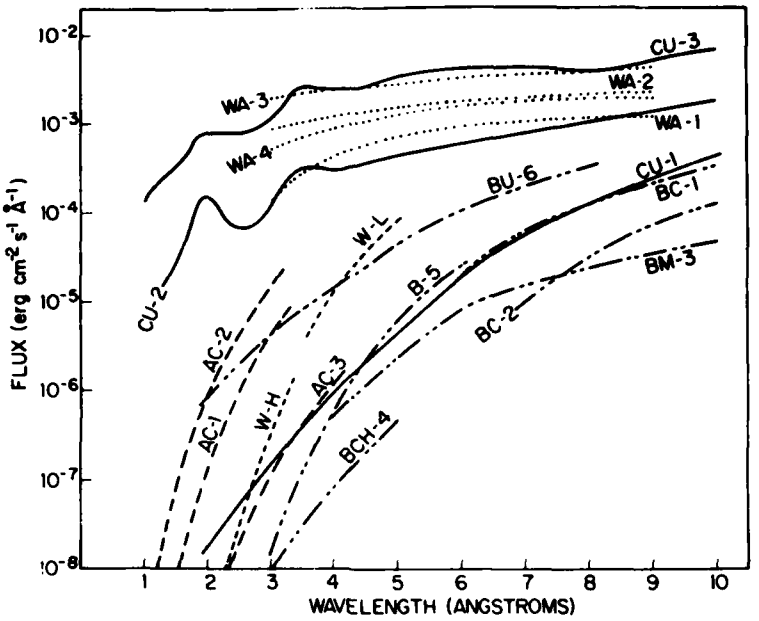


Fig. 1: Variability of the solar X-ray spectral region (White, 1977)

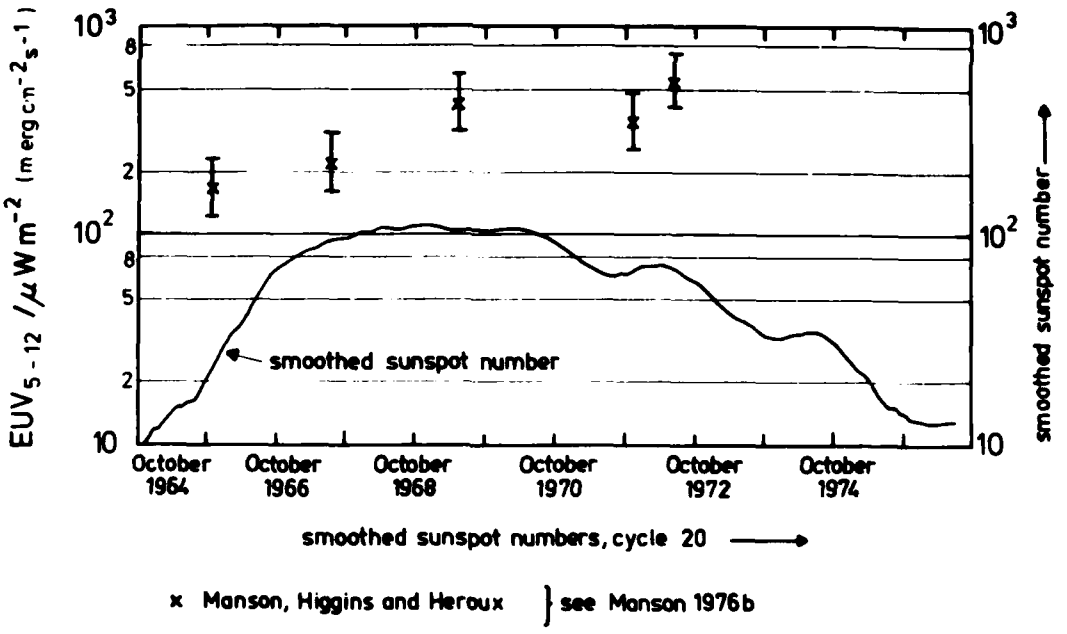


Fig. 2: Measurements in the wavelength range 5 - 12 nm (Delaboudiniere et al., 1980)

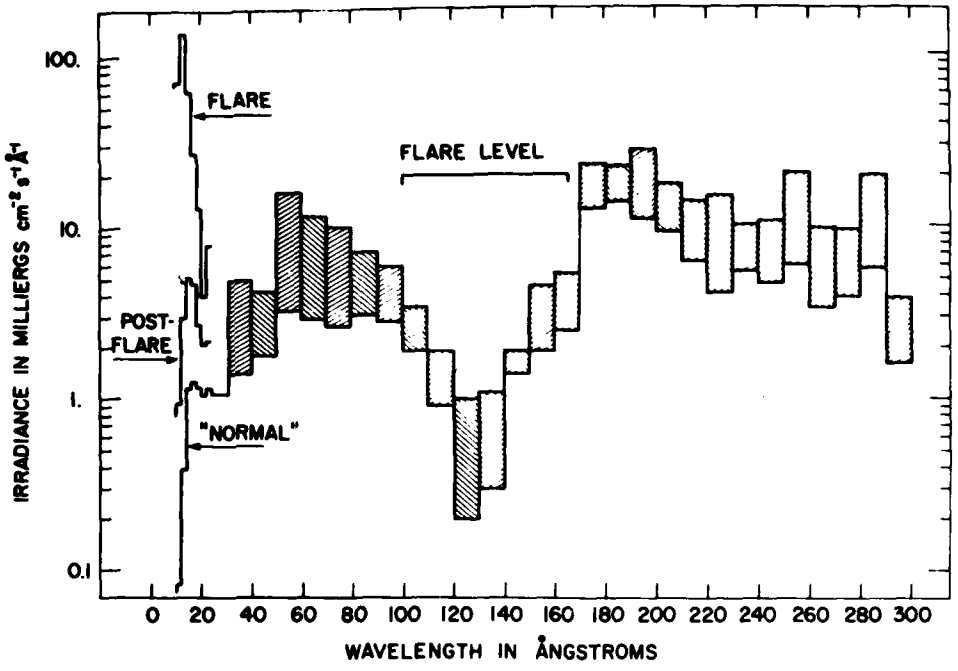


Fig. 3: Solar spectral irradiance between 1 - 30 nm (White, 1977).

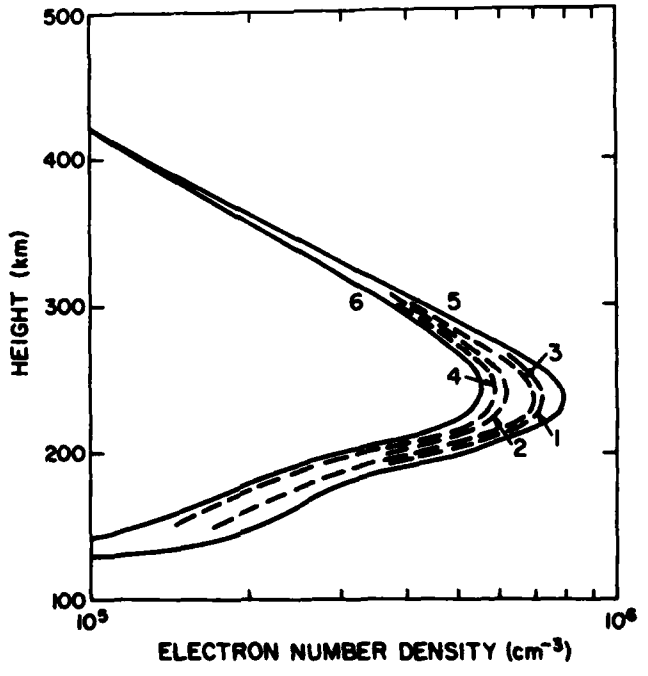


Fig. 5: Calculated electron density profile in 1973 (Roble and Schmidtke, 1979)

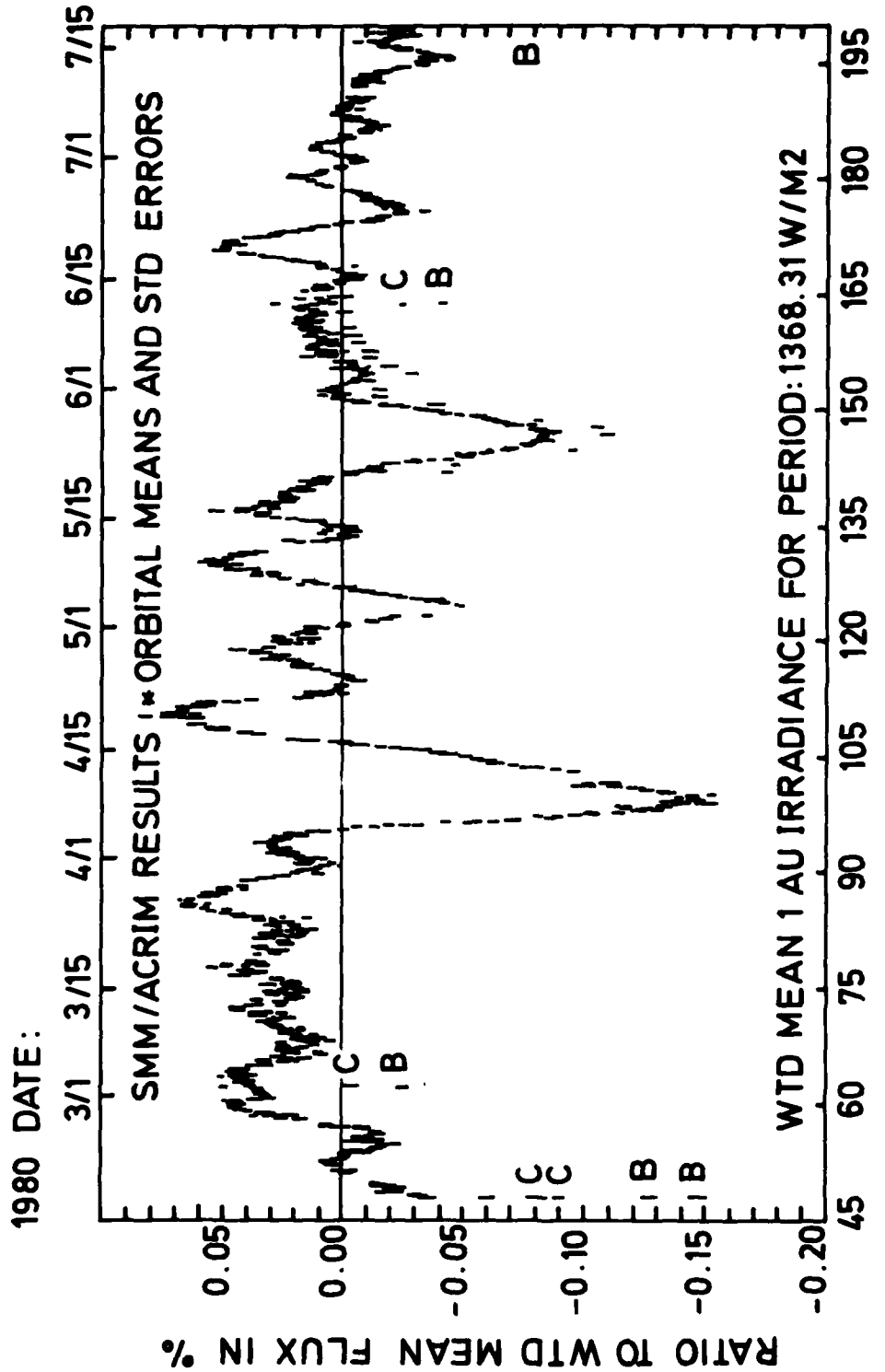


Fig. 4: Solar total irradiance in 1980 (Willson, 1980)

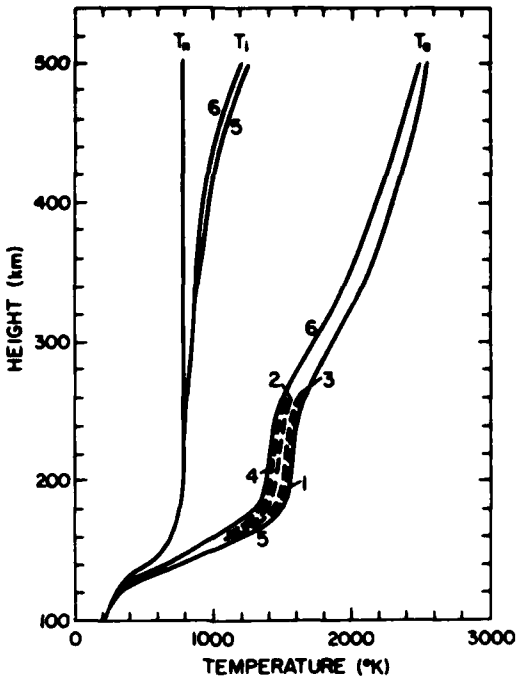


Fig. 6: Calculated electron and ion temperature profiles in 1973 (Roble and Schmidtke, 1979)

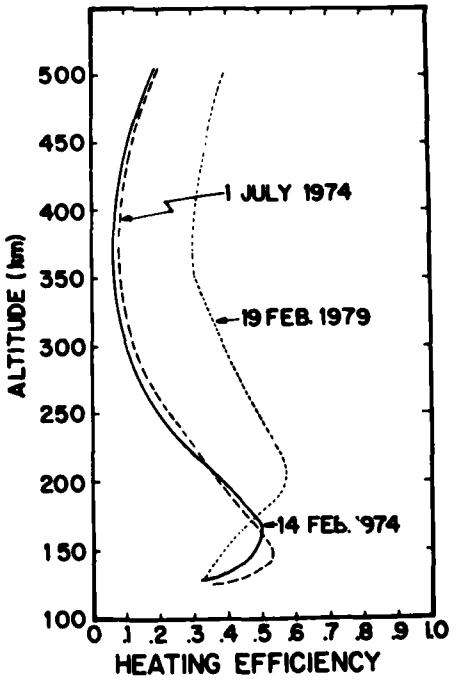


Fig. 7: Heating efficiency profiles for solar minimum and maximum conditons (Torr et al. 1981).

MAGNETOSPHERIC AND IONOSPHERIC FLOW AND THE INTERPLANETARY MAGNETIC FIELD

S.W.H. Cowley
Blackett Laboratory
Imperial College
London, SW7 2BZ, U.K

SUMMARY

The flux tube and plasma flow within the Earth's magnetosphere and the resulting flow in the ionosphere are discussed, with particular reference to effects associated with the interplanetary magnetic field (IMF). When IMF $B_z \lesssim 0$, or is at least not strongly positive, two-cell convection occurs at high latitudes with antisunward flow over the polar cap. The size of the cells depends on IMF B_z, as expected from the Dungey open magnetosphere model, but part of the convection appears to be driven by 'viscous' magnetopause boundary layers. Asymmetries occur in the two-cell flow about the noon-midnight meridian in response to IMF B_y which form a consistent set with a ready qualitative interpretation in terms of the open model. Sunward-tailward shifts in the cells also occur in response to IMF B_x, which may be similarly interpreted. When IMF B_z is strongly positive a qualitatively different ionospheric flow pattern occurs at high latitudes with sunward flow in at least part of the polar cap. Observations pertaining to this condition are reviewed, together with theoretical interpretations.

1. INTRODUCTION

The single most important factor governing the morphology and properties of the Earth's magnetosphere and its internal plasma populations is the convection of flux tubes and plasma within it driven by the flow of the solar wind. The magnetospheric electric fields associated with the flow map along equipotential (at least on a gross scale) magnetic field lines into the ionosphere and drive ionospheric flows and current systems which dominate at high latitudes, above the plasmapause. The flow is important in determining the distribution of high latitude ionization, and the resulting Joule heating can be an important energy source. Particle precipitation from the magnetosphere also contributes to both the energy and ionization budgets.

In this paper the ionospheric flows driven at high latitudes by magnetospheric convection will be discussed. For recent reviews of the effects of convection in determining magnetospheric plasma populations and their ionospheric images the reader is referred to the recent works by, eg Vasyliunas (1979), Wolf and Harel (1980) and Cowley (1980).

The flow at high latitudes is observed to be strongly influenced by the strength and direction of the interplanetary magnetic field. When IMF B_z (the south to north component) is negative, or at least not strongly positive, a twin-cell convection pattern occurs, with antisunward convection over the polar cap. The magnitude of this convection depends upon the magnitude of B_z. Asymmetries appear in the convection pattern about the noon-midnight meridian in connection with the IMF B_y (dawn to dusk) component, and these are discussed in detail in the next section. Effects associated with the IMF B_x component (Sunward directed) will also be commented upon. When the IMF B_z component is strongly positive, however, a qualitatively different convection pattern is observed, with sunward flow in at least part of the polar cap. This behaviour forms the basis of the discussion in section (3).

2. CONVECTION WITH IMF $B_z < 0$ AND THE EFFECTS ASSOCIATED WITH IMF B_x AND B_y

2.1. The Basic Two-Cell Convection Pattern

When the IMF B_z component is negative, or at least not strongly positive, the plasma convection pattern observed over the Earth's polar regions consists of two flow cells, with anti-sunward convection over the polar cap and return sunward flow in the auroral zones, as sketched in Figure (1a). The known correlation between the size of the polar cap, the potential drop across it and the IMF B_z component (eg Holzworth and Meng, 1975; Akasofu, 1977, chapter 4 and references therein; Meng, 1980) strongly indicates the importance of the role played by magnetic merging in driving this convection, as originally suggested by Dungey (1961). The convection cycle for a southward-directed IMF is illustrated by the numbered field line sequence in Figure (1b). An X-type neutral line ring encircles the Earth in the equatorial plane, across the dayside section of which closed field lines become opened into the solar wind (field line 1). The solar wind flow carries the open'ends' downstream, stretching them out into a long geomagnetic tail, and the solar wind electric field E_y, mapping down the open field lines into the ionosphere, results in anti-sunward convection over the polar cap. The field lines become closed again across the nightside section of the X-line, possibly mainly impulsively during substorms (lines 5) and then return to the dayside via the dusk and dawn auroral zones where the cycle repeats. Typical polar cap convection speeds of $\sim$500 m s^{-1}, result in typical transpolar convection times of $\sim$2 to 4 hours, so that the length of the tail connected to the polar regions is $\sim$1000 R_E (Dungey, 1965). The tail neutral line, however, may typically lie much closer to the Earth. Similarly, the time taken for closed tubes to convect back from the tail to the dayside may typically be $\sim$8 hours, so that the total convection cycle time is around half a day, generally encompassing several substorms. While the correlation of this flow with IMF B_z indicates the importance of the above process, at least part of the flow appears to be driven by a viscous-like interaction at the magnetopause on closed field lines (Eastman et al., 1976; Eastman and Hones, 1979, McDiarmid et al., 1979), as originally proposed by Axford and Hines (1961). This flow is indicated schematically in Figure (1a) by the regions of antisunward flow external to the open field line boundary, where the field lines remain closed throughout the convection cycle. The magnetospheric magnetic structure envisaged is that proposed by Crooker (1977). At the present time the relative importance of the two convection-driving processes (measured by the relative potential drops imposed across the magnetosphere) and its dependence on IMF conditions is not well determined, but magnetic merging may be dominant under typical, and certainly under disturbed conditions.

2.2 Observed IMF B_y-dependent asymmetries

The convection pattern in Figure (1a) is shown symmetrical about the noon-midnight meridian, but under the influence of the IMF B_y component which is generally present a number of asymmetries have been observed to occur about the meridian. These asymmetries are oppositely directed in opposite hemispheres and reverse in sense when the IMF B_y reverses. The effects observed in the northern hemisphere for IMF $B_y > 0$ are sketched in Figure (2a) and itemized below:

(1) Fairfield (1977), Potemra and Saflekos (1979) and McDiarmid et al. (1979) have suggested on somewhat slender evidence that the narrow 'throat' of flows into the dayside polar cap (originally described by Heelis et al., 1976) may be displaced from noon in a direction dependent on the IMF B_y component ie eastward in the northern hemisphere and westward in the southern hemisphere when IMF $B_y > 0$, and vice-versa for IMF $B_y < 0$. The existence of this effect is somewhat questionable, however, and most recently Primdahl et al. (1980) have present evidence which they interpret as indicating that this local time shift may not in fact occur. This question will be briefly discussed further below.

(2) Newly opened flux tubes in the dayside cusp have a strong azimuthal flow whose direction depends on the IMF B_y direction (Galperin et al., 1978; Heelis, 1979). The azimuthal flow drives an oppositely-directed Hall current along the cusp whose magnetic signature on the ground is the Svalgaard-Mansurov effect, or DPY disturbance (eg Mansurov, 1969; Friis-Christensen et al., 1972; Svalgaard, 1973; Berthelier et al., 1974; Friis-Christensen and Wilhjelm, 1975; Akasofu et al., 1980). The electric field associated with the flow will also drive a north-south Pedersen current across the cusp which will close in the solar wind via roughly azimuthally aligned field-aligned current (FAC) sheets, as predicted by Leontyev and Lyatsky (1974). This current system results in azimuthal 'transverse field disturbances' above the ionosphere between the FAC sheets as observed by McDiarmid et al. (1978b, 1979), Wilhjelm et al. (1978), Iijima et al. (1978), Saflekos et al. (1979), and Saflekos and Potemra (1980). These transverse fields are generally expected to be directed opposite to the flow in the northern hemisphere and roughly parallel to the flow in the southern hemisphere, as is necessary to realize the required downward flux of electromagnetic energy and momentum into the ionosphere. In the cusp these 'transverse' fields thus have the same direction as the IMF B_y component in both hemispheres, such that the field above the ionosphere is slightly tilted in a manner consistent with that expected for a field mapping into a dissipative medium and being azimuthally 'pulled' at large distances.

(3) The magnitude of the antisunward flow in the polar cap, eg in the vicinity of the dawn-dusk meridian usually exhibits a strong IMF B_y dependent dawn-dusk asymmetry which is consistent and continuous with the preferred sense of azimuthal flow in the dayside cusp (Heppner, 1972, 1973; Mozer et al., 1974). The effect is also seen in ground magnetic perturbations, (Maezawa, 1976). The polar cap 'transverse field perturbations' show a corresponding asymmetry (McDiarmid et al., 1978a, 1979), suggestive of a distributed and fairly uniform FAC flow into or out of the polar cap resulting from the flow electric field gradient. For IMF $B_y > 0$ the suggested current flow is out of the northern polar cap and into the southern polar cap, and vice-versa for IMF $B_y < 0$, ie the same directions as the dominant poleward cusp FAC.

(4) The two-cell pattern appears to become shifted towards dawn or dusk by a few degrees of latitude depending upon the IMF B_y component, the shift being opposite in opposite hemispheres. The northern hemisphere pattern is shifted towards dawn for IMF $B_y > 0$ and towards dusk for IMF $B_y < 0$. This shift has been observed in the electric field (flow) pattern (Heppner, 1972, 1973; Mozer et al., 1974), the resulting ground magnetic disturbance pattern (Friis-Christensen and Wilhjelm, 1975), and most recently in the location of the aurorae (Meng, 1980), although the latter result is preliminary at the present time.

(5) In the auroral zone at a given invariant latitude the flows in and near the dawn-dusk meridian are of unequal magnitude on either side of the pole, the sunward flow being strongest on the side in which the polar cap flow is strongest (Mozer and Lucht, 1974). This effect may be considered to be a corollary of the shift in location of the flow pattern, since auroral zone flows often become stronger with increasing latitude before reversing across the flow boundary.

It is important to note that the cusp and polar cap asymmetries described in (2) and (3) above need involve only open flux tubes which map into the solar wind via the geomagnetic tail lobes, but that asymmetries (4), (5) and (1) (should it in fact occur) involve closed auroral zone flux tubes as well. The dawn-dusk shift in the pattern, for example, may mean that at a given invariant latitude and given longitude a field line may belong to the polar cap region in one hemisphere but to the auroral zone in the other hemisphere. Mozer and Lucht (1974) previously drew attention to the appearance of asymmetries on closed flux tubes, noting that their results implied either a "lack of conjugacy" on auroral zone field lines (to be suggested here), or non-equipotential closed field lines on a gross scale (which seems to have been their preferred explanation).

With regard to the proposed shift in the dayside 'throat' (effect (1)) we note that effect (2), the preferred sense of azimuthal flow in the cusp, can occur independently of the latter, although the local time extent of such flows may be expected to depend on the 'throat' location. If the full local time width of the 'throat' is only about two hours, as suggested by Heelis et al. (1976) on the basis of the AE-C flow data, and the centre of the 'throat' remains fixed at noon independent of IMF B_y, then the B_y-dependent azimuthal cusp flow poleward of the throat and its associated magnetic effects (DPY on the ground and azimuthal transverse disturbances at satellite altitudes) would be expected to be observed essentially only on one side of noon or the other, ie prenoon in the northern hemisphere and postnoon in the southern hemisphere for IMF $B_y > 0$ and vice-versa for IMF $B_y < 0$. On the other side of noon a much weaker and oppositely directed azimuthal flow and field perturbation pattern would be expected, the cusp FAC pattern being essentially that originally suggested by Iijima and Potemra (1976) (see also the review by Potemra et al. (1980)) but with asymmetrical magnitudes on either side of noon. On the contrary, however, observations have been presented which indicate that the 'preferred' azimuthal cusp flow and associated magnetic disturbances extend continuously across noon in a zone which is at least ~5 hours LT wide and centred roughly on noon (eg Friis-Christensen and Wilhjelm, 1975; McDiarmid et al., 1978b; Burch and Heelis, 1980). These results are not consistent with a narrow throat which remains undisplaced near noon, and they thus indicate that either the throat does become displaced as described in item (1) above, or that the throat

is much wider than reported by Heelis et al. (1976) (ie typically ∿5 hours LT wide and therefore essentially non-existent as a 'throat-like' feature). The evidence presented by Primdahl et al. (1980) which is interpreted as indicating that the 'throat' does not move concerns the occurrence of the 'Slant E Condition' (SEC) in high-latitude ionograms, a condition which appears to be related to an instability in the E-region driven by rapid flows. It was previously shown by Olesen et al. (1975) that SEC-occurrence in the central polar cap (at Thule, $\Lambda \simeq 86^{o}$) had an IMF B_y-associated dusk-dawn asymmetry in line with the observed flow asymmetry (item (3) above), a result which Primdahl et al. confirm. The latter authors then go on to show that at lower latitudes typical of the poleward part of the cusp region (Godhavn, $\Lambda = 77.5^{o}$) no significant asymmetry is observed, a result which they interpret in terms of the non-displacement of the throat. However, as pointed out above, if the throat is narrow and does not become displaced then the strong azimuthal flows in the poleward part of the cusp, to which the Godhavn ionosonde will presumably respond, would be expected to show a marked asymmetry about noon. In fact the symmetry of SEC occurrance about noon at Godhavn seems to be quite in agreement with the continuity across noon of the other B_y-associated cusp phenomena discussed above, which we interpreted as indicating that either the 'throat' does become displaced, or is much wider in LT than has been reported. We therefore conclude that Primdahl et al.'s (1980) results may, in fact, be consistent with a displaced 'throat', but would equally emphasize the need for further experimental study of the dayside cusp region in order to elucidate this matter.

IMF-B_y associated effects have also been observed at large distances in the geomagnetic tail as sketched in Figure (2b):

(6) The occurrence frequency of the plasma mantle at the lunar distance shows a strong IMF-B_y dependent dawn-dusk asymmetry which is oppositely directed in the two tail lobes (Hardy et al., 1976, 1979). These authors thus deduced the existence of a dusk-dawn asymmetry in the tail lobe flow speed and hence electric field (which convects the mantle plasma inward toward the tail centre), an asymmetry which is entirely consistent with the ionospheric polar cap flow asymmetry in item (3) above. The observed mantle occurrence frequency asymmetry has been interpreted in Figure (2b) as indicating that on average the mantle is thicker on one side of the tail than the other depending on the direction of IMF B_y (thicker on the dawn side of the northern lobe and on the dusk side of the southern lobe for IMF $B_y > 0$ and vice-versa for IMF $B_y < 0$).

(7) Fairfield (1979) found that a fairly uniform B_y field appears across the tail lobes and plasma sheet in the distance range $X \sim -20$ to $-30\ R_E$, having the same sign as the IMF B_y. The field is roughly proportional to the IMF B_y but about a factor ten less in magnitude. It is important to emphasize that the B_y field appears to occur not only in the open tail lobe region but also on closed field lines in the plasma sheet.

2.3. Interpretation of the IMF B_y-associated effects in terms of the reconnection model, and comments on IMF B_x effects

The above described suite of effects associated with the y component of the IMF form a mutually consistent set which find a ready qualitative explanation in terms of the reconnection model of the magnetosphere, outlined in its simplest form in section (2.1.). The inclusion of IMF B_y, however, requires modification of the description of the details of reconnection process and the magnetic field configuration, as discussed by Cowley (1973) (see also Dungey, 1963, Stern 1973, Mozer et al. 1974 and Gonzales and Mozer, 1974). The gross picture remains as shown in Figure (1b), except that the X-type reconnection line becomes tilted out of the equatorial plane by the presence of B_y. Crooker (1979) has presented a somewhat different qualitative picture, but, irrespective of these details essentially any *open model will be qualitatively consistent with the observations.*

Beginning at the day side, the tension on newly opened flux tubes has a net east-west component in the presence of the B_y field, which is oppositely directed in northern and southern hemispheres. These oppositely directed forces constitute a torque exerted by the IMF on the magnetosphere, and, in response, oppositely directed azimuthal flows occur in the dayside cusp in the directions observed, as sketched in Figure (3a) (Atkinson, 1972, Jørgensen et al., 1972; Russell and Atkinson, 1973; Stern, 1973; Gonzales and Mozer, 1974; Crooker, 1979). The azimuthal flows imply asymmetric addition of open flux to the tail lobes, flux being added preferentially to the dawn side of the northern lobe and to the dusk side of the southern lobe for IMF $B_y > 0$, and vice versa for IMF $B_y < 0$. The flux tube motion is such as to reduce the torque which the IMF would continue to exert on the down-stream tail.

The asymmetric addition of flux tubes to the tail lobes may be expressed in terms of an asymmetry in the distribution of the field normal to the tail lobe magnetopause about the noon-midnight meridian over the boundary as sketched in Figure (3b). The figure shows a cross-section through the tail, the external field lines attached to the boundary representing projections of magnetosheath field which connect across the boundary at the location of the cut. To a lowest approximation sheath field lines are draped around the magnetopause (ie the magnetosphere is closed), but a narrow bundle connects across it due to dayside reconnection, whose transverse width well away from the magnetosphere is typically around one fifth to one tenth of the tail diameter, as inferred from the relative internal and external electric field strengths. The cross-tail potential is typically ∿50 to 100 kV, while the potential drop in the solar wind along a line of length equal to the tail diameter and parallel to the solar wind electric field is ∿500 kV(eg Stern, 1973). This factor, having a value around a few tenths is sometimes referred to as the "efficienty of reconnection". The connected interplanetary field lines have been drawn equally spaced in the uniform IMF region at large distance from the magnetopause in Figure (3b), so that equal magnetic flux is contained between them. The tail flux asymmetry implied by the cusp flows is then indicated by the uneven distribution of their points of intersection with the magnetopause. Equal potential differences exist between each of the sheath field lines depicted, so that an electric field asymmetry is also directly implied. Quantitatively, the tangential electric field imposed by the sheath flow v_s at the tail magnetopause boundary is $E = v_s B_n$ where B_n is the field strength normal to the boundary. Consequently, assuming v_s to be roughly constant over the boundary, the internal electric fields will be strongest and the inward flow speeds fastest in those quadrants of the tail in which most magnetic flux crosses the magnetopause. This effect is clearly related to the observed polar cap electric field asymmetry and plasma mantle

occurrence asymmetry noted in items (3) and (6) above, and most recently Cowley (1980) has constructed simple open tail lobe and plasma flow models which explicitly demonstrate these effects. In Figure (3b) the tail electric field asymmetry is illustrated by the short-dashed lines within the tail which indicate equipotentials in the plane of the cut with equal potential drops between them. These lines are also the projections of equipotential tail lobe field lines, and plasma streamlines as indicated by the arrowheads.

The asymmetry of the tail magnetopause normal flux distribution about the midnight meridian also necessarily implies the cross-tail magnetic field component observed by Fairfield (1979). Flux enters the tail preferentially on one side of the meridian and exits preferentially on the other, so that by flux conservation net flux must cross the midnight meridian, associated with an internal B_y field of the same sign as the IMF B_y field. The magnitude of this internal field is readily estimated from Figure (3b), on assuming that the figure represents a cut through a roughly two-dimensional system. It is clear that when substantial asymmetry exists in the boundary normal flux distribution (such that most of the flux enters on one side of the tail central meridian and leaves on the other) the internal B_y field will be comparable to the mean normal field component $\bar{B}_n$, and will be less than the interplanetary field strength B_I by the ratio of internal to external electric field strengths, since both the potential drop and the magnetic flux in the plane of the diagram are conserved between the connected bundle of IMF lines and the tail lobes. Thus in general the internal B_y field and the IMF are connected in magnitude by the "efficiency of reconnection" factor of around one or two tenths in agreement with Fairfield's (1979) results.

The above qualitative argument can be quantitatively elaborated as follows. Suppose the total open flux in each tail lobe is Φ_T, and that of this total a fraction $(1 + K)/2$ enters the northern lobe on the dawn side (and $(1 - K)/2$ on the dusk side) while, assuming symmetry, a fraction $(1 + K)/2$ leaves the southern lobe on the dusk side (and $(1 - K)/2$ on the dawn side). The net flux entering the tail on the dawn side is then $K\Phi_T$, equal to the net flux leaving on the dusk side. This same quantity of flux must then cross the midnight meridian inside the tail. The sign of K is the same as the sign of IMF B_y, and $|K| < 1$. Assuming this 'transverse' flux is uniformly distributed over the meridian in the tail, of area $A \simeq 2 R_T L_T$ where R_T is the tail radius and L_T its length, then the mean internal B_y field $\bar{B}_y$ is given by

$$\bar{B}_y = \frac{K\Phi_T}{2R_T L_T} = (\frac{\pi K}{2})\bar{B}_n = (\frac{\pi K}{4}) B_T (\frac{R_T}{L_T}) \qquad (1)$$

where $\bar{B}_n$ is the mean magnetopause normal field and B_T is the near-Earth tail lobe field strength. Thus for $|K| \sim 1$ we have $\bar{B}_y \sim \bar{B}_n$ and $\bar{B}_y << B_T$ (the latter by factors $\sim$50 since $R_T \sim 20\ R_E$ and $L_T \sim 1000\ R_E$). Referring to Figure (3b) flux and potential conservation then give

$$\frac{\bar{B}_n}{B_{\perp_I}} \simeq \frac{2}{\pi}(\frac{E_T}{E_I}) \qquad (2)$$

where $B_{\perp_I}$ is the interplanetary field transverse to the solar wind flux vector and E_T and E_I are the tail and interplanetary electric fields respectively. Substituting (2) into (1) then gives

$$\bar{B}_y \simeq K(\frac{E_T}{E_I})B_{\perp_I} \qquad (3)$$

so that for $|K| \sim 1$, $\bar{B}_y$ is related to $B_{\perp_I}$ by the ratio of the electric fields. This result may be elaborated somewhat further by introducing simple functional forms for K and E_T/E_I. If ϕ is the angle between $\underline{B}_{\perp_I}$ and the solar magnetospheric $-\hat{z}$ axis, measured positive toward the $\hat{y}$ axis, then the simplest functional form for K is $\sin(\phi/2)$, such that, eg K is zero for a purely southward field and equal to $\pm 1/\sqrt{2}$ for $\underline{B}_{\perp_I}/|B_{\perp_I}| = \pm\hat{y}$ respectively. Similarly the simplest form for E_T/E_I is $\beta\cos(\phi/2)$ where β is the "efficiency of reconnection" for a purely southward interplanetary field. Then (3) becomes

$$\bar{B}_y \simeq \beta\frac{\sin\phi}{2} B_{\perp_I} = \frac{\beta}{2} B_{y_I} \qquad (4)$$

ie with these simple and plausible, but ad hoc assumptions we obtain a linear dependence of the mean internal field $\bar{B}_y$ on the IMF B_y, with the proportionality constant being half the "efficiency of reconnection" factor for a purely southward IMF. Fairfield's (1979) observations indeed indicate a linear dependence between $\bar{B}_y$ and B_{y_I}, the proportionality constant being $\sim$0.13. This result is in good agreement with (4) and observed values of the magnetospheric cross-tail potential which indicate $\beta \sim 0.1$ to 0.3.

Having thus discussed the IMF B_y effects observed on open flux tubes (items (2), (3), (6) and (7) in section (2.2.)) we now turn to the effects ((1), (4) and (5)) which also involve closed auroral zone flux tubes, and which seem to indicate a lack of conjugacy between northern and southern hemispheres. In this connection we would point out that if an IMF-associated B_y field exists in the open flux tube regime in the geomagnetic tail, then in general we may expect this component to be retained on closed flux tubes following reconnection in the tail. Fairfield's (1979) observations certainly seem to support this idea as previously noted. Then, to a lowest approximation the tail lobe B_y field, of order the IMF B_y diminished by the "efficiency of reconnection" factor, may pervade the entire magnetosphere in a steady state, such that the simple illustrative 'dipole plus uniform IMF field' open magnetosphere models described eg by Dungey (1963), Stern (1973) and Cowley (1973) may have more detailed applicability than their authors might originally have supposed.

A B_y field present in the region of closed flux tubes will introduce shifts in the conjugate points of the field lines which are opposite in opposite hemispheres and which will reverse in sense when the IMF B_y reverses. The direction of the shift depends upon local time. In the dawn-dusk meridian the shifts are latitudinal and the resulting asymmetries are those associated with the polar cap boundary and the auroral zone flow speed (items (4) and (5) in section (2.2.)). In the noon-midnight meridian the shift is longitudinal, relating to the possible shift of the 'throat' location (item (1)).

The effect of these conjugate point displacements on the mapping of a uniform sunward flow in the equatorial plane of the Earth's dipole is shown in Figure (4), taken from Cowley (1980). Here we have projected equally-spaced equipotentials (corresponding to the lines Y = 0, $\pm$3 R_E, ... $\pm$12 R_E) from the equatorial plane along numerically integrated field lines, assumed equipotential, into the northern hemisphere. Figure (4a) shows the resulting ionospheric flow for the undisturbed dipole field, while in Figure (4b) we have added a 5nT B_y field, a much larger value than generally expected (by a factor $\sim$10) in order to clearly demonstrate the resulting effects. The flow lines are terminated at the dashed line which is the projection of an equatorial circle of 14.93 R_E radius (ie $\Lambda_o = 75^o$). The diagram represents the closed field line counterpart of that previously presented for the open field line regime on a similar basis by Stern (1973). The shift of the pattern towards dawn (for $B_y > 0$), the auroral zone flow speed asymmetry and the local time shift of the 'centre' of flows into the dayside polar cap to the post-noon hours are all evident. We emphasize that the simultaneous asymmetries are all reversed in sense in the southern hemisphere, and that they are all produced solely by the change in the field mapping.

Figure (4b) indicates that longitudinally-directed flow asymmetries should also occur in the nightside auroral zone. The flow shown in reminiscent of the Harang discontinuity (eg Heppner, 1977), but, of course, the sense of the IMF-related asymmetry reverses with IMF B_y. However, Heelis (1979) and Heelis and Hanson (1980) find both senses of asymmetry in AE-C flow data, although it has not been shown that these relate to IMF B_y. The sense of the IMF-associated asymmetry is such that in the northern hemisphere nightside auroral zone eastward flows and westward Hall currents should be enhanced when IMF $B_y > 0$, and westward flows and eastward Hall currents should be enhanced when IMF $B_y < 0$ (and vice-versa in the southern hemisphere). A hint that a weak effect of this kind may be occurring is contained in the northern hemisphere ground magnetic perturbations analysed by Fraus-Christensen and Wilhjelm (1975) (see the 'residual plots' in their even-numbered figures). The most distinct effect of this nature which they found was a significant enhancement of the westward electrojet for IMF $B_y > 0$ during winter months, a result which is at least consistent in sign with the above suggestions.

The conjugate point displacements are also relevant to studies of the conjugacy of aurorae. It is well known that equivalent discrete auroral forms in northern and southern hemispheres can be displaced a few 100 km at times relative to their expected locations (eg Bond, 1969; Stenbaek-Nielsen et al., 1972). The possible role of IMF B_y has never been demonstrated for these detailed auroral observations, but the recent preliminary results presented by Meng (1980) from the analysis of DMSP images do indicate that the location of the aurorae on a gross scale does respond to IMF conditions. In the northern hemisphere he found dawnward shifts of the centre of the auroral oval for IMF $B_y > 0$ and duskward shifts for IMF $B_y < 0$ as expected. The magnitude of the shift is generally $\leq 1^o$ latitude, but can be much larger. In the southern hemisphere no clear trend was observed with IMF B_y from the limited data sample available. He also investigated the possibility of sunward-tailward shifts of the oval with IMF B_x, and indeed found that the northern oval is shifted tailward for $B_x > 0$ and sunward for $B_x < 0$ and vice versa in the southern hemisphere. The shifts are again $\sim 1^o$. This result appears to indicate that like the B_y component, a B_x component may also occur through the magnetosphere whose sign depends on the IMF B_y. In a simple 'dipole plus uniform field' model the shift of the oval is oppositely directed to the uniform field direction in the northern hemisphere, and in the same direction as the uniform field in the southern hemisphere. Consequently tailward shifts in the northern hemisphere and sunward shifts in the southern hemisphere observed for IMF $B_x > 0$ are, for example, compatible with $B_x > 0$ appearing within the magnetosphere. Our conclusion that the results are compatible with simple models differs from the conslusions reached by Meng (1980), due, it seems, to some typographical errors concerning the sign of B_x in Stern's (1973) paper.

The possible appearance of B_x within the magnetosphere may, like B_y, be considered as being due to inter-hemispheric differences in the flux tube tension on newly opened dayside field lines. As shown in Figure (5), when IMF B_x is positive the flux tube tension resulting in poleward contraction of the newly opened tubes will be greater in the northern than in the southern hemispheres (and vice-versa for IMF $B_x < 0$). Consequently the poleward contraction rate of the tubes in the northern hemisphere will be greater than in the southern hemisphere, and the normal flux through the magnetopause correspondingly less. In other words, if we follow the 'ends' of the interplanetary field line following reconnection on the dayside and consider where these 'ends' intersect the magnetopause at subsequent times, then the northern end intersection should always, following reconnection, be displaced tailward of the southern end (for IMF $B_x > 0$, and vice versa for IMF $B_x < 0$), as shown in Figure (5). The effect is not dependent on the sign of the IMF y component. It is clear that this displacement is equivalent to the imposition of a net x-directed flux along the magnetosphere in the same direction as IMF B_x. In the tail at a given X, therefore, the open lobe flux will be slightly greater in one tail lobe than the other (greater in the northern lobe than the southern for IMF $B_x > 0$ and vice versa for IMF $B_x < 0$), and consequent upon the fact that the two lobe field strengths must be very nearly equal from pressure balance considerations across the central current sheet, we then conclude that the central current sheet must itself be displaced in the z-direction (southward, ie to -ve z for IMF $B_x > 0$ and northward for IMF $B_x < 0$). The displacements are likely to be a few tenths of an R_E in magnitude. Within the near-Earth magnetosphere, however, the imposed cross-magnetosphere x-directed flux may appear more straight forwardly as a perturbation field vector which shifts the conjugate points of field lines in a manner consistent with Meng's (1980) observations, and tilting the nightside field lines in a manner consistent with the north-south shift of the central current sheet in the tail.

Finally, before concluding this section we should point out that the x-component also appears to exert significant control on particle access and precipitation in the polar caps. The effects on high-energy solar particles have been extensively reviewed by Paulikas (1974), while Mizera and Fennell (1978) have reviewed low-energy magnetosheath electron precipitation. With regard to the latter, the presence of the x-component results in the northern polar cap being magnetically connected to a hotter magnetosheath electron population than the southern polar cap when the IMF $B_x < 0$ (away sector), and vice versa for IMF $B_x > 0$. This presumably accounts for Yeager and Frank's (1976) result that few 100 eV electron fluxes over the northern polar cap are much more intense during away sectors than during toward sectors.

3. CONVECTION WITH IMF B_z POSITIVE

3.1. Appearance of a new convection pattern for IMF $B_z \gtrsim 1$ nT

Consideration of the simple dipole plus uniform field model indicates that a magnetically open magnetosphere having the same topological characteristics as shown in Figure (1b) can occur for all orientations of the IMF except due north, when a closed magnetosphere is expected for strictly steady conditions (eg Dungey, 1963; Stern, 1973; Cowley, 1973; Yeh, 1976). Similarly, intensive theoretical investigation of the reconnection process has failed to reveal any finite minimum angle between merging fields below which reconnection cannot take place, independent of the relative magnitudes of the fields (Cowley, 1976, and references therein). Consequently, we might then expect that as the IMF B_z turns from south to north the high-latitude flows would decrease both in magnitude and in spatial extent (consistent with decreasing open flux in the tail lobes) but would remain qualitatively as shown in Figure (1a). That is, we might expect a weakening two cell pattern to remain, presumably dominated at sufficiently positive IMF B_z by a closed field line two-cell flow driven by the magnetopause boundary layers.

This seems not to be the case. When the IMF B_z is sufficiently positive (a consensus of opinion favours $B_z \gtrsim 1$ nT) a steadily accumulating body of evidence appears to indicate that a qualitatively different flow pattern occurs at very high latitudes, with a sunward-directed flow appearing in the noon sector of the polar cap whose magnitude increases as IMF B_z increases. When IMF $|B_z| < 1$ nT only the two-cell pattern seems to occur, the flow magnitude of which depends upon the magnitude of IMF B_y (Maezawa, 1976).

3.2. Observational evidence

Observation of sunward flow in the polar cap and related magnetic disturbances is relatively infrequent ($\lesssim$ 10%, say, of typical polar orbiting satellite passes), no doubt reflecting in part the unusual nature of the IMF conditions required to produce such flows, and in part the spatially restricted and very high-latitude region in which they occur. Consequently, the experimental literature is sparse. In the monoaxial OGO-6 electric field data Heppner (1972) observed that a region of dusk-to-dawn electric field occasionally appears in a localized region of the central polar cap when magnetic activity is low, as sketched in Figure (6a) (see also Langel, 1975). However, in the absence of vector information it was not possible to infer whether this was due to the occurrence of a qualitatively different multi-cell flow pattern, or to a distortion of the usual two-cell flow, as was considered more likely by Heppner (1977). Flows having a sunward component but primarily directed east-west can certainly occur in the dayside polar cap, for example, as a result of the distortions of the 2-cell pattern associated with the IMF B_y component discussed in the previous section (Heelis et al., 1976; Burch and Heelis, 1980). That this is not the cause of sunward polar cap flows when IMF $B_z \gtrsim 1$ nT is indicated by the S3-2 electric field measurements presented by Burke et al. (1979), who found that the east-west flows were not large compared with the sunward flows in such regions, and who argued on the basis of the extent of the region where such flows were observed that they could not be due to a distortion of the normal two-cell system. Rather, they suggested that a four-cell system was being observed, as sketched in Figure (6b). They also found that such flow patterns occur only when the IMF B_z is positive, and that they always occurred in their data set when $B_z > 0.7$ nT. From electric field measurements on balloons Mozer and Gonzales (1973) and Mozer et al. (1974) had also inferred the existence of dusk-to-dawn electric fields in the high latitude polar cap in a number of cases when the IMF B_z component was positive.

The analysis of ground magnetic perturbations has produced some conflicting results. Friis-Christensen and Wilhjelm (1975) found that when IMF $B_z > 1$ nT the effects of the two-cell convection pattern disappear entirely. There remains, however, a one-cell IMF B_y-dependent DPY disturbance pattern whose form and current magnitude is essentially the same as that which occurs for $B_z \lesssim 0$ (then superimposed upon the two-cell pattern), but displaced towards higher latitudes. Maezawa (1976) found a similar B_y dependence, but, on statistically removing this effect then found a two-cell flow pattern at latitudes above 78° whose direction is the reverse of the 'normal' pattern for IMF $B_z < 0$. Similar results and their IMF B_z dependence were first reported by Iwasaki (1971). These results are broadly in agreement with the electric field observations of Burke et al. (1979) (which eg also show IMF B_y-dependent asymmetries), except that the lower-latitude 'normal' flow cells as shown in Figure (6b) are not represented. Maezawa's (1976) analysis may have been confined to too high a latitude ($\Lambda \gtrsim 78°$) to have detected these, particularly on the dayside, while Friis-Christensen's (1975) conclusions may have been affected by their choice of baseline. More recently Horwitz and Akasofu (1979) have re-examined high-latitude magnetograms for times of IMF $B_z > 0$ and have suggested that the perturbation pattern is quite consistent with the flow pattern sketched in Figuer (6b). In particular they find that the region of sunward flow in the polar cap at $\Lambda \simeq 85°$ occurs only in the noon sector and does not extend into the nightside hours as suggested by Maezawa's (1976) results. On the nightside the ground magnetic perturbations are consistent with anti-sunward polar cap flow but at a considerably reduced level to that which occurs for IMF $B_z < 0$. It may be noted that Burke et al's. (1979) observations of sunward flow also in general appear to be confined to the dayside of the polar cap.

Most recently McDiarmid et al. (1980) and Saflekos and Potemra (1980) have presented observations of 'transverse field perturbations' during periods of strongly positive B_z which show essentially identical behaviour to Burke et al's. (1979) and Heppner's (1972) electric field observations shown in Figure (6a). The region of sunward flow appearsas an antisunward field perturbation in the northern hemisphere (McDiarmid et al., 1980) and as a sunward-directed perturbation in the southern hemisphere (Saflekos and Potemra, 1980). Again, we note that all such observations published to date occurred on the dayside of the polar cap. It is to be emphasized that the perturbations associated with sunward flow are by no means small, and appear to increase in magnitude as B_z increases, reaching ~4000 nT in one case published by McDiarmid et al. (1980) when IMF B_z probably had the unusually large value of ~30 nT. The dependence on B_z is illustrated in Figure (7) where the maximun perturbation associated with the region of sunward flow is plotted versus IMF B_z for the six satellite passes published to date for which IMF B_z is known. The solid circles are from McDiarmid et al. (1980), the open circles from Saflekos and Potemra (1980). By comparison, it is known that the peak field perturbations associated with the sunward 'auroral zone' flows (of the lower latitude cells in Figure (6b)) do not increase with IMF B_z, but instead appear to continuously

fall from values of typically several 100 nT when IMF B_z is substantially negative to values 100 - 200 nT when IMF B_z is substantially positive (eg McDiarmid et al., 1978a).

It is of obvious significance to know whether the region of sunward flow occurs on open or closed flux tubes. McDiarmid et al. (1980) investigated this question by examining the pitch angle distributions of energetic solar electrons which were present during the three passes with IMF $B_z > 0$ which they presented. In one case, with IMF $B_z \sim 9$ nT, they found the region of sunward flow to be on open field lines, as was the surrounding region of antisunward flow. All of the sunward 'auroral zone' flow is on closed lines, together with part of the antisunward flow adjacent to it, as is usually the case (see Figure 1a). These results are in apparent agreement with the observations of Burke et al. (1979) who indicated a rough position of the polar cap boundary obtained from energetic electron data. In their diagrams the 'polar cap' so defined appears to encompass both the region of sunward flow and the two adjacent regions of antisunward flow as well. This may be the situation under normal circumstances. However in two passes observed during very disturbed conditions ($B_z \sim 30$ nT) McDiarmid et al. (1980) found that at least the major part of the central region of sunward flow was unequivocally on closed field lines, while the surrounding region of antisunward flow remained on open field lines.

3.3. Theoretical Interpretations

Interpretation of the sunward-directed flow in the polar cap has traditionally centred on the possibility of magnetic merging between a northward-directed IMF and tail lobe field lines poleward of the dayside cusp. A variety of descriptions of the process are then possible, depending upon whether the tail lobe fields are taken to be open or closed, and whether a given interplanetary field line is taken to connect with only one or with both lobes. Several possibilities are sketched in schematic form in Figure (8). In Figures (8a - c) the interplanetary line connects to both lobes, while in Figures (8d - e) it connects only to one. In Figures (8a, b, d) the lobes are closed, while in Figures (8c, d) they are open. The numbers indicate the sequence of the merging process in each case.

The case shown in Figure (8a) corresponds to that described by Dungey (1963) for a strictly northward interplanetary magnetic field. The field topology is that of a strictly closed magnetosphere (eg a dipole field plus uniform northward field), except that the north and south neutral points must presumably be imagined as being elongated in the dawn-dusk direction into lines so that finite flux transfer can take place. The only field lines which connect both to the Earth and to the IMF pass through these neutral lines, ie there is zero open magnetic flux. In the process depicted in Figure (8a) the northward IMF on the dayside reconnects simultaneously in the northern and southern hemispheres, with the net result that closed nightside flux tubes become closed dayside flux tubes. These dayside flux tubes can then be transported back to the tail via the magnetopause boundary layers so that a steady state can be envisaged. In the equatorial plane closed flux tubes 'appear' over a segment of the dayside magnetopause mapping to the neutral lines, along which a dusk-to-dawn electric field exists. These flux tubes then flow tailward in a layer adjacent to the magnetopause before 'disappearing' again over a similar nightside segment of the boundary of closed field lines. The convection cycle of these tubes corresponds to the high-latitude two-cell system shown in Figure (6b). The ionospheric image of the neutral lines is a line joining the centres of the two cells, the nightside to dayside flux transfer taking place as the streamlines cross the line. The two lower latitude flow cells in Figure (6b) then correspond to an inner part of the boundary layer flow where the flux tubes return to the dayside by flowing back through the 'middle' of the magnetosphere connected to the lower latitude 'auroral zones' in the usual way. In this system, therefore, closed flux tubes are transferred from the dayside to the tail via the magnetopause boundary layers and return to the dayside either via the auroral zones or via the flux transfer process depicted in Figure (8a).

Figure (8b) shows a simple variation on the above theme which may occur with either an IMF B_x field or with a tilted dipole. In this case the north and south reconnection processes do not occur simultaneously on one interplanetary field line such that the two neutral lines are not magnetically connected. In the case depicted in Figure (8b) the line is first connected to the northern lobe and field lines in both hemispheres are temporarily open before connection occurs to the southern lobe. The ionospheric image of the two neutral lines are therefore not now coincident, and between them there exists a "hole" of open magnetic flux. The flow in this hole will be sunward in direction as is readily seen on mapping the dusk-to-dawn directed interplanetary electric field along the open magnetic field lines into the ionosphere. The overall flow pattern remains as shown in Figure (6b).

In Figure (8c) the equivalent situation is shown for open tail lobe flux tubes, the latter being connected to downstream IMF lines having a differing orientation (southward as shown in the figure). Here the net effect is to remove open flux tubes from the lobes and to increase the closed flux on the dayside. The newly closed tubes are then swept back into the tail via the boundary layer flow. The magnetic topology in this case may be imagined to be that obtained by taking a dipole field plus infinitely conducting sheet in the y - z plane on the dayside (as in the Chapman-Ferraro problem) and adding to it a weak southward-directed uniform field, such that the equatorial neutral line ring which occurs in the absence of the sheet no longer encircles the Earth but is confined to the nightside, and terminates at two equatorial neutral points on the sheet. The dayside magnetopause then resembles that of a closed magnetosphere, with a bounded surface of field lines mapping from the southern to northern 'cusp' neutral points (which we imagine as being extended into lines so that finite flux transfer can take place via the process shown in Figure (8c)), while the nightside field remains that of a classical open magnetosphere. The bounding field lines of the closed dayside magnetopause surface map from the northern and southern 'cusp' neutral lines to the termination of the equatorial nightside neutral line on the sheet, so that the three neutral lines (northern and southern cusp and equatorial nightside) are magnetically connected. They map in the ionosphere to a closed ring which encircles and bounds the region of open flux. Over a dayside segment of this ring the electric field is dusk-to-dawn directed, corresponding to the flux transfer process in Figure (8c), while over the remaining (nightside) segment the electric field is from dawn to dusk, corresponding to continued connection to the downstream southward interplanetary field and continued reconnection of the open flux tubes in the tail. The region of open flux thus continuously shrinks, consistent with Faraday's law applied to this ring, and the end state is the closed magnetosphere of Figure (8a). This occurs with the closing around the nightside of the field lines mapping to the northern

and southern cusp neutral lines, such that the equatorial neutral line and the region of open flux simultaneously disappear. While the open flux tubes are in the process of disappearing the ionospheric flow will retain antisunward flowsin the region of open flux tubes while the reversed high-latitude two-cell system will emerge on the dayside. The flow will reverse in direction across the dayside open field line boundary (mapping to the cusp neutral lines), which may appear peculiar at first, but it should be remembered that this boundary will be moving polewards as the open flux decreases, at a rate somewhat in excess of the antisunward flow speed on the high-latitude side of the boundary, consistent with net flux transfer from open lobe tubes to closed dayside tubes. In this picture the sunward flow on the dayside is again on closed tubes, but if the two cusp reconnection processes are not simultaneous, as taken in the above discussion for simplicity (ie the picture is modified to that shown in Figure (8b)), part of the sunward flow adjacent to the original boundary of open flux tubes will also be on open tubes, as discussed above in relation to Figure (8b).

The discussion so far has assumed that each incoming interplanetary field line connects, at least eventually, with both northern and southern lobes. This process may occur when the IMF is very nearly northward pointing, and the dayside sunward flow then occurs mainly on closed flux tubes as discussed above, possibly relating to the observations of McDiarmid et al. (1980) when $B_z \simeq 30$ nT. However, when a significant IMF B_y field is also present, a given interplanetary line is likely to connect with only one lobe, as first pointed out by Russell (1972) and discussed further by Maezawa (1976). In Crooker's (1979) model only this latter process occurs; northward field lines never become connected to both lobes. The process is illustrated for closed and open tails respectively in Figures (8d, e). In Figure (8d) closed tail field lines become open dayside field lines so that the process is inherently non-steady and involves increasing open flux, while in Figure (8e) open flux tubes are transferred from the lobes to the dayside, the net amount of open flux remaining unchanged. The former case is that discussed by McDiarmid et al. (1980) in relation to their observations of sunward flow on closed tubes. The latter case corresponds to that discussed by Russell (1972) and Maezawa (1976), (see also Crooker, 1979). In both cases the ionospheric flow associated with the open dayside flux tubes will have a sunward component, but the field line tension associated with the IMF B_y component will also introduce a preferred sense of azimuthal flow (ie the Salgaard-Mansurov effect) and asymmetric addition of the open flux tubes to the tail lobes exactly as previously described in section (2). Although it is not easy to depict in two-dimensional sketches such as Figures (8d, e) connection of differing interplanetary lines may (but need not) proceed simultaneously in both northern and southern hemispheres since the dayside field lines shown do not lie in the plane of the figure due to the presence of B_y. For the case shown in Figure (8e) Maezawa (1976) and Crooker (1979) suggest the possibility of continuous circulation of open flux tubes within the polar cap. Maezawa suggests two cells, while Crooker suggests one, whose direction depends on B_y. In the latter work this cell coexists with the normal two-cell pattern when B_z is small but negative, and these in general also coexist for positive B_z. This type of continuous open flux tube circulation appears to agree with the observations of Burke et al. (1979) and McDiarmid et al. (1980) of sunward flow on apparently open flux tubes, although a full account of the process remains as yet to be given. The remaining possible explanation is in terms of a highly distorted two-cell system, which has not yet quite been decisively rejected. The observational picture, at least, should become clearer when more extensive observations of the ionospheric flow under positive IMF B_z conditions are analysed. In this regard integration of electric field data to provide the electrostatic potential pattern would be especially useful.

ACKNOWLEDGEMENTS

This work was performed during the tenure of a U.K. SRC Advanced Fellowship. The author would like to thank Mrs. H.M. Southwood for her help in preparing the figures and Mrs. F. Josselyn for her careful typing.

REFERENCES

Akasofu, S.-I., 1977, "Physics of Magnetospheric Substorms", D. Reidel, Dordrecht-Holland.

Akasofu, S.-I., Romick, G.J. and H.W. Kroehl, 1980, The IMF B_y effects observed by the IMS Alaska meridian chain, J. Geophysic. Res., 85, 2079.

Atkinson, G., 1972, Magnetospheric flows and substorms, In "Magnetosphere-Ionosphere Interactions", (ed. K. Folkestad), p.203, Universitetsforlaget, Oslo, Norway.

Axford, W.I. and Hines, C.O., 1961, A unifying theory of high-latitude geophysical phenomena and geomagnetic storms, Can. J. Phys., 39, 1433.

Berthelier, A., Berthelier, J.J., and Guerin, C., 1974, The effect of the east-west component of the interplanetary magnetic field on magnetospheric convection as deduced from magnetic perturbations at high latitudes, J. Geophys. Res., 79, 3187.

Bond, F.R., 1969, Auroral morphological similarities at two magnetically conjugate stations: Buckles Bay and Kotzebue, Aust.J. Phys., 22, 421.

Burch, J.L., and Heelis, R.A., 1980, IMF changes and polar-cap electric fields and currents, In "Dynamics of the Magnetosphere", (ed. S.-I.Akasofu), D. Reidel, Dordrecht, Holland, p.47.

Burke, W.J., Kelley, M.C., Sagalyn, R.C., Smiddy, M., and Lai, S.T., 1979, Polar cap electric field structures with a northward interplanetary mangetic field, Geophys. Res. Lett., 6, 21.

Cowley, S.W.H., 1973, A qualitative study of the reconnection between the Earth's magnetic field and an interplanetary field of arbitrary orientation, Radio Sci., 8, 903.

Cowley, S.W.H., 1976, Comments on the merging of non antiparallel magnetic fields,J. Geophys. Res., 81, 3455.

Cowley, S.W.H., 1980, Plasma populations in a simple open model magnetosphere, Space Sci. Rev., 26, 217.

Cowley, S.W.H., 1980, Magnetospheric asymmetries associated with the Y-component of the IMF, Planet. Space Sci., to be published.

Crooker, N.U., 1977, The magnetospheric boundary layers: A geometrically explicit model, J. Geophys. Res., 82, 3629.

Crooker, N.U., 1979, Dayside merging and cusp geometry, J. Geophys. Res., 84, 951.

Dungey, J.W., 1961, Interplanetary magnetic field and the auroral zones, Phys. Rev. Lett., 6, 47.

Dungey, J.W., 1963, The structure of the exosphere or adventures in velocity space. In "Geophysics, The Earth's Environment", (eds. C. De Witt, J. Hieblot, and L. Le Beau), Gordon and Breach, New York, p.503.

Dungey, J.W., 1965, The length of the magnetospheric tail, J. Geophys. Res., 70, 1753.

Eastman, T.E., Hones, E.W.,Jr., Bame, S.J., and Asbridge, J.R., 1978, The magnetospheric boundary layer: Site of plasma, momentum and energy transfer from the magnetosheath into the magnetosphere, Geophys. Res. Lett., 3, 685.

Eastman, T.E., and Hones, E.W.,Jr., 1979, Characteristics of the magnetospheric boundary layer as observed by Imp 6, J. Geophys. Res., 84, 2019.

Fairfield, D.H., Electric and magnetic fields in the high-latitude magnetosphere, Revs. Geophys. Space Phys., 15, 285, 1977.

Fairfield, D.H., 1979, On the average configuration of the geomagnetic tail, J. Geophys. Res., 84, 1950.

Friis-Christensen, E., Lassen, K., Wilhjelm, J., Wilcox, J.M., Gonzales, W., and Colburn, D.S., 1972, Critical component of the interplanetary magnetic field resonsible for large geomagnetic effects in the polar cap, J. Geophys. Res., 77, 3371.

Friis-Christensen, E., and Wilhjelm, J., Polar cap currents for different directions of the interplanetary magnetic field in the y-z plane, J.Geophys. Res., 80, 1248,1975.

Galperin, Yu., I., Ponomarev, V.N., and Zosimova, A.G., 1978, Equatorial ionospheric anomaly and interplanetary mangetic field, J. Geophys. Res., 83, 4265.

Gonzales, W.D., and Mozer, F.S., 1974, A quantitative model for the potential resulting from reconnection with an arbitrary interplanetary magnetic field, J. Geophys. Res., 79, 4186.

Hardy, D.A., Freeman, J.W.,Jr., and Hills, H.K., 1976, Plasma observations in the magnetotail, In "Magnetospheric Particles and Fields", (ed. B.M. McCormac), p.89, D. Reidel, Hingham, Mass., U.S.A.

Hardy, D.A., Hills, H.K. and Freeman, J.W., 1979, Occurrance of the lobe plasma at lunar distance, J. Geophys. Res., 84, 72.

Heelis, R.A., 1979, Ionospheric convection at high latitudes, Proc. Magnetospheric Boundary Layers Conference, ESA SP-148, ESTEC, Noordwijk, The Netherlands, p.175.

Heelis, R.A., Hanson, W.B., and Burch, J.L., 1976, Ion convection velocity reversals in the dayside cleft, J. Geophys. Res., 81, 3803.

Heelis, R.A., and Hanson, W.B., 1980. High latitude ion convection in the nighttime F region, J. Geophys. Res., 85, 1995.

Heppner, J.P., 1972, Polar cap electric field distributions related to the interplanetary magnetic field direction, J.Geophys. Res., 77, 4877.

Heppner, J.P., High latitude electric fields and the modulcations related to interplanetary magnetic field parameters, Radio Sci., 8, 933 , 1973.

Holzworth, R.H., and Meng, C.-I., 1975, Mathematical representation of the auroral oval, Geophys. Res. Lett., 2, 377.

Horwitz, J.L., and Akasofu, S.-I., 1979, On the relationship of the polar cap current system to the north-south component of the interplanetary magnetic field, J. Geophys. Res., 84, 2567.

Iijima, T., and Potemra, T.A., 1976, Field-aligned currents in the dayside cusp observed by Triad, J. Geophys. Res., 81, 5971.

Iijima, T., Fujii, R., Potemra, T.A., and Saflekos, N.A., 1978, Field-aligned currents in the south polar cusp and their relationship to the interplanetary magnetic field, J. Geophys. Res., 83, 5595.

Iwasaki, N., 1971, Localized abnormal geomagnetic disturbance near the geomagnetic pole and simultaneous ionospheric variation, Rep. Ionos. Res. Space Res. Jap., 25, 163.

Jørgensen, T.S., Friis-Christensen, E., and Wilhjelm, J., 1972, Interplanetary magnetic field direction and high-latitude ionospheric current, J. Geophys. Res., 77, 1976.

Langel, R.A., 1975, A comparison of electric field and magnetic field data from the Ogo 6 spacecraft, J. Geophys. Res., 80, 4661.

Leontyev, S.V., and Lyatsky, W. B., 1974, Electric field and currents connected with the Y-component of the interplanetary magnetic field, Planet. Space Sci., 22, 811.

Maezawa, K., 1976, Magnetospheric convection induced by the positive and negative Z components of the interplanetary magnetic field: Quantitative analysis using polar cap magnetic records, J. Geophys. Res., 81, 2289.

Mansurov, S.M., 1969, New evidence of a relationship between magnetic fields in space and on earth, Geomagn. Aeron. USSR, 9, 622.
McDiarmid, I.B., Burrows, J.R., and Wilson, M.D., 1978a, Comparison of magnetic field perturbations at high latitudes with charged particle and IMF measurements, J. Geophys. Res., 83, 681.
McDiarmid, I.B., Burrows, J.R., and Wilson, M.D., 1978b, Magnetic field perturbations in the dayside cleft and their relationship to the IMF, J. Geophys. Res., 83, 5753.
McDiarmid, I.B., Burrows, J.R., and Wilson, M.D., 1979, Large-scale magnetic field perturbations and particle measurements at 1400 km on the dayside, J. Geophys. Res., 84, 1431.
McDiarmid, I.B., Burrows, J.R., and Wilson, M.D., 1980, Comparison of magnetic field perturbations and solar electron profiles in the polar cap, J. Geophys. Res., 85, 1163.
Meng, C.-I., 1980, Polar cap variations and the interplanetary magnetic field, In "Dynamics of the Magnetosphere", (ed. S.-I. Akasofu), D. Reidel, Dordrecht, Holland, p.23.
Mizera, P.F., and Fennell, J.F., 1978, Satellite observations of polar, magnetotail lobe, and interplanetary electrons at low energies, Rev. Geophys. Space Phys., 16, 147.
Mozer, F.S., and Gonzales, W.D., 1973, Response of polar cap convection to the interplanetary magnetic field, J. Geophys. Res., 78, 6784.
Mozer, F.S., Gonzales, W.D., Bogott, F., Kelley, M.C., and Schutz, S., 1974, High-latitude electric fields and the three-dimensional interaction between the interplanetary and terrestrial magnetic fields, J. Geophys. Res., 79, 56.
Mozer, F.S., and Lucht, P., 1974, The average auroral zone electric field, J. Geophys. Res., 79, 1001.
Olesen, J.K., Primdahl, F., Spangslev, F., and D'angelo, N., 1975, On the Farley instability in the polar cap E region, J. Geophys. Res., 80, 696.
Paulikas, G.A., 1974, Tracing of high latitude magnetic field lines by solar particles, Rev. Geophys. Space & Phys., 12, 117.
Potemra, T.A., and Saflekos, N.A., Birkeland currents and the interplanetary magnetic field, Proc. Magnetospheric Boundary Layers Conference, ESA SP-148, ESTEC, Noordwijk, The Netherlands, p.193,1979.
Potemra, T.A., Iijima, T., and Saflekos, N.A., 1980, Large-scale characteristics of Birkeland currents, In "Dynamics of the Magnetosphere", (ed. S.-I.Akasofu), D. Reidel, Dordrecht, Holland, p.165.
Primdahl, F., Primdahl, I.S. and Spangslev, F., 1980, Longitudinal position of polar cap plasma flow entry region, J. Geophys. Res., 85, 33.
Russell, C.T., 1972, The configuration of the magnetosphere, In "Critical Problems of Magnetospheric Physics", (ed. E.R. Dyer), Inter-Union Committee on Solar Terrestrial Physics, National Academy of Sciences, Washington, D.C., p.1.
Russell, C.T., and Atkinson, G., 1973, Comments on a paper by J.P. Heppner, 'Polar cap electric field distributions related to interplanetary magnetic field direction', J. Geophys. Res., 78, 4001.
Saflekos, N.A., Potemra, T.A., Kintner, P.M.,Jr., and Green, J.L., 1979, Field-aligned currents, convection electric fields, and ULF-ELF waves in the cusp, J. Geophys. Res., 84, 1391, 1979.
Saflekos, N.A., and Potemra, T.A., 1980, The orientation of Birkeland current sheets and the dayside polar region and its relationship to the IMF, J. Geophys. Res., 85, 1987.
Stenbaek-Nielsen, H.C., Davis, T.N., and Glass, N.W., 1972, Relative motion of auroral conjugate points during substorms, J. Geophys. Res., 77, 1844.
Stern, D.P., 1973, A study of the electric field in an open magnetospheric model, J. Geophys. Res., 78, 7292.
Svalgaard, L., 1973, Polar cap magnetic variations and their relationship with the interplanetary magnetic sector structure, J. Geophys. Res., 78, 2064.
Vasyliunas, V.M., 1979, Interaction between the magnetospheric boundary layers and the ionosphere, Proc. Magnetospheric Boundary Layers Conference, ESA SP-148, ESTEC, Noordwijk, The Netherlands, p.387.
Wilhjelm, J., Friis-Christensen, E., and Potemra, T.A., 1978, The relationship between ionospheric and field-aligned currents in the dayside cusp, J. Geophys. Res., 83, 5586.
Wolf, R.A., and Harel, M., 1980, Dynamics of the magnetospheric plasma, In "Dynamics of the Magnetosphere", (ed. S.-I- Akasofu), D. Reidel, Dordrecht, Holland, p. 143.
Yeager, D.M., and Frank, L.A. 1976, Low energy electron intensities at large distances over the Earth's polar cap, J. Geophys. Res., 81, 3966.
Yeh, T., Dayside reconnection between a dipolar geomagnetic field and a uniform interplanetary field, J. Geophys. Res., 81, 2140, 1976.

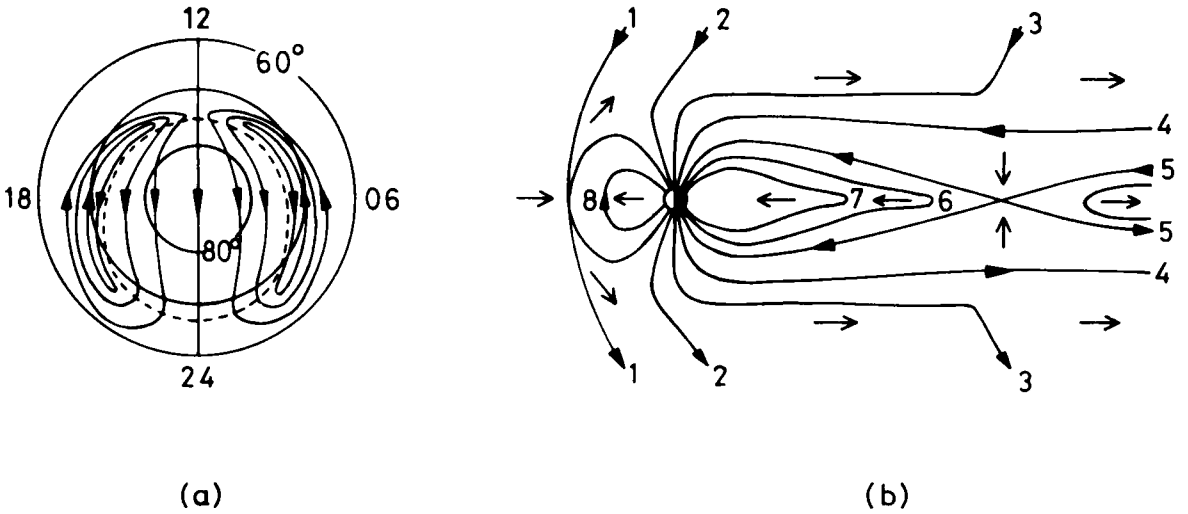


Figure 1. (a) Sketch of the basic two-cell convection pattern observed under usual conditions, when the IMF B_z is not strongly positive. The dashed line is the boundary of open field lines, and the antisunward convection just equatorward of it maps to the magnetopause boundary layers on closed field lines. The flow in the vicinity of noon is represented as a narrow 'throat' structure in conformity with the observations of Heelis et al. (1976), but no attempt has been made to represent the 'Harang discontinuity' on the nightside (see eg. Heppner, 1977). The coordinates are magnetic latitude and local time.

(b) Sketch of the convection cycle proposed by Dungey (1961) initiated by dayside reconnection between the Earth's magnetic field and a southward-directed interplanetary magnetic field. The numbers indicate the successive positions of a flux tube in the cyclic flow.

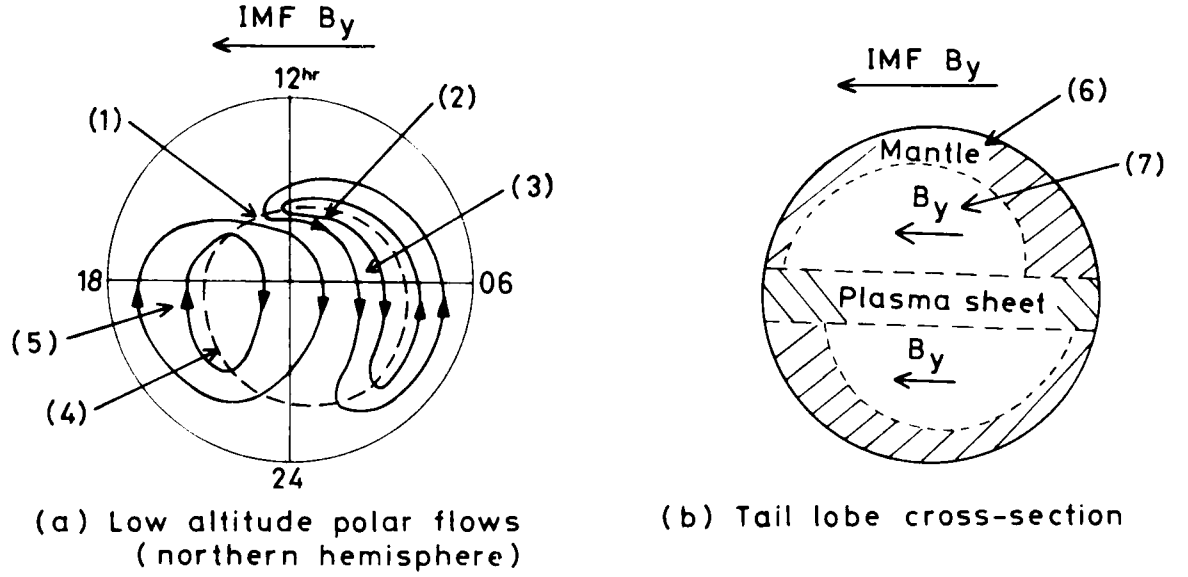


Figure 2. Sketch of observed magnetospheric asymmetries associated with the IMF B_y component (a) in low altitude polar flows (norther hemisphere shown), (b) in the geomagnetic tail, for $B_y > 0$. When $B_y < 0$ the asymmetries reverse in sense. These asymmetries are: (1) Local time displacements of the 'throat' of flows into the dayside polar cap (?); (2) preferred sense of azimuthal flow in the dayside cusp; (3) asymmetric polar cap flow speed; (4) dawn-dusk shift of the polar cap boundary; (5) auroral zone flow speed asymmetry; (6) plasma mantle occurrence frequency asymmetry; (7) B_y field in the tail.

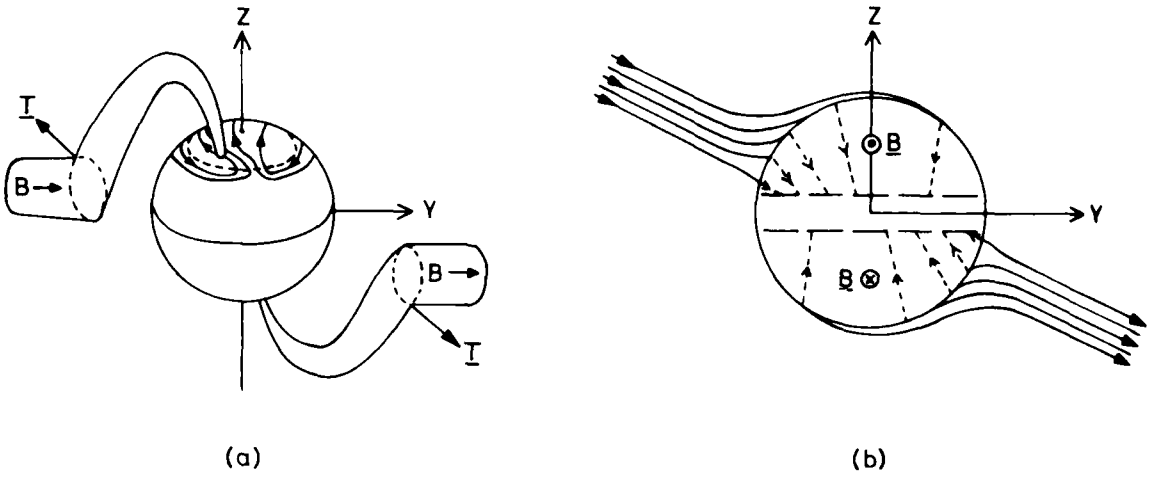


Figure 3. Sketches of the effects of the IMF B_y component on magnetospheric convection on the dayside and in the tail.

(a) Schematic view of newly opened flux tubes on the dayside, illustrating the field line tension T which results in oppositely directed azimuthal flows in northern and southern cusp regions. The case shown is for IMF $B_y > 0$ and the flows reverse for IMF $B_y < 0$. The dashed line represents the open field line boundary.

(b) Sketch of a cross-section through the geomagnetic tail, showing the IMF flux bundle (for IMF $B_y > 0$ as in (a)) which connects across the magnetopause at the location of the cut. Equal magnetic flux and potential drops occur between each of the IMF lines drawn, and the magnetopause normal flux asymmetry and applied electric field asymmetry are then indicated by the uneven distribution of the points of intersection of the lines with the magnetopause. The asymmetry results from, and is continuous with the azimuthal flow depicted in (a). The short dashed lines show the same equipotentials within the tail lobes; these lines also represent projections of equipotential tail lobe field lines and also plasma streamlines, as indicated by the arrows. The long dashed lines indicate the boundary between the open tail lobes and the central layer of reconnected field lines.

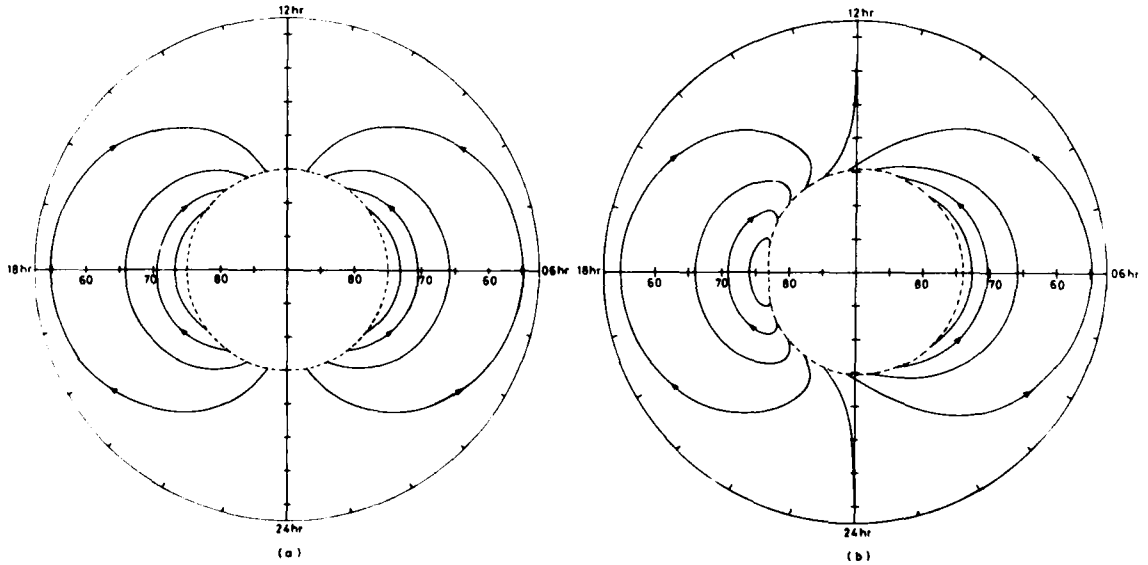


Figure 4. Shown here are the streamlines in the northern auroral zone ionosphere obtained by mapping a uniform sunward flow in the equatorial plane along equipotential closed field lines. In (a) a pure dipole field has been used, while in (b) a 5 nT uniform B_y field has been added, and the field lines numerically integrated. The dashed line shows the mapping of a circle at R = 14.93 R_E ($\Lambda = 75^o$) in the equatorial plane. In (b) note the dusk-dawn asymmetries in and near the dusk-dawn meridian and the local time shifts in and near the noon-midnight meridian. The simultaneous asymmetries are reversed in sense in the southern hemisphere, and are produced solely by the change in the mapping of the field resulting from the presence of the B_y field.

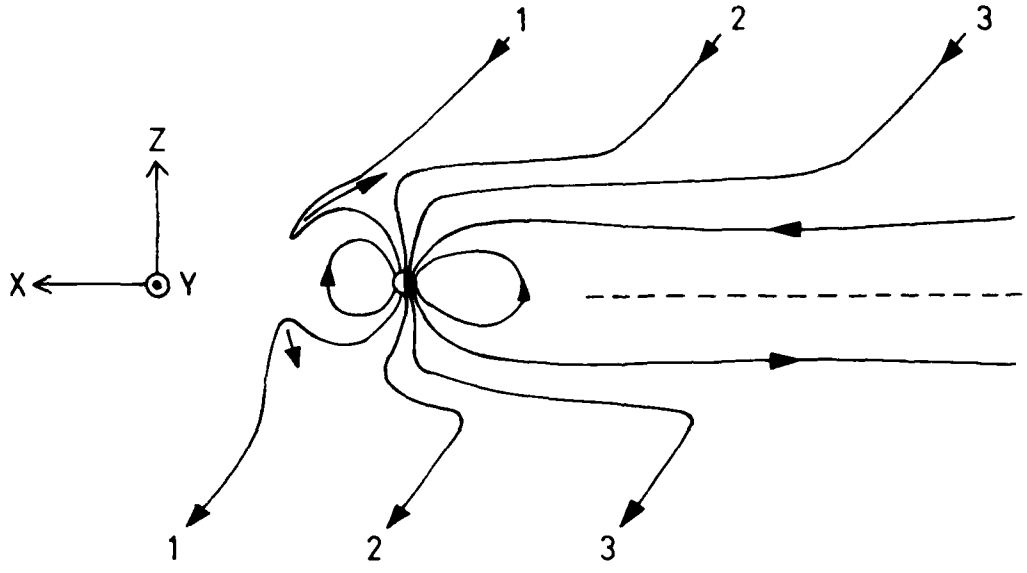


Figure 5. Sketch of the reconnection model in the presence of an IMF B_y field (and $B_z \lesssim 0$) showing the unequal flux tube tensions on newly opened dayside flux tubes, and the resulting displacement along the x-axis between the magnetopause intersection points of the two 'ends' of the IMF field line. The displacement requires the existence of a net x-directed flux internal to the magnetosphere, which will result in sunward-tailward shifts in the polar cap boundary and low altitude flow pattern, and a north-south displacement in the central current sheet in the geomagnetic tail.

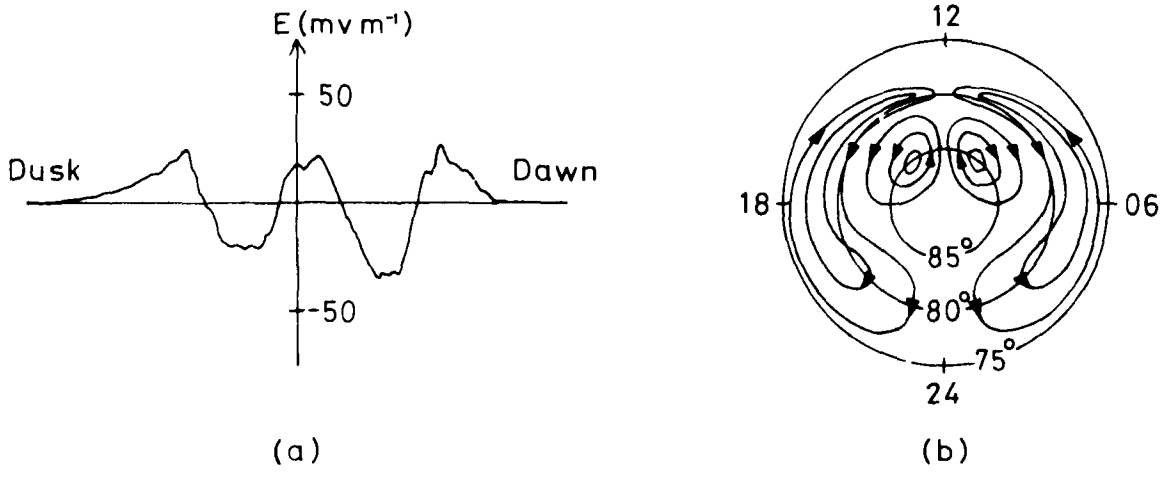


Figure 6. Sketch of the electric fields and inferred flow pattern in northernhemisphere high latitudes when IMF $B_z \gtrsim 1$ nT. (a) Electric field pattern observed in a near dawn-dusk pass in the dayside part of the flow. Positive values correspond to a dusk-to-dawn electric field direction and sunward flow. Larger values of antisunward flow are shown on the dawn side of the polar cap, corresonding in the northern hemisphere to IMF $B_y > 0$.

(b) Inferred flow pattern from polar electric field measurements, transverse magnetic field perturbation measurements, and observations of ground magnetic field disturbances for IMF $B_z \gtrsim 1$ nT. The pattern shown is symmetric about the noon meridian, but asymmetries occur dependent on IMF B_y as indicated above.

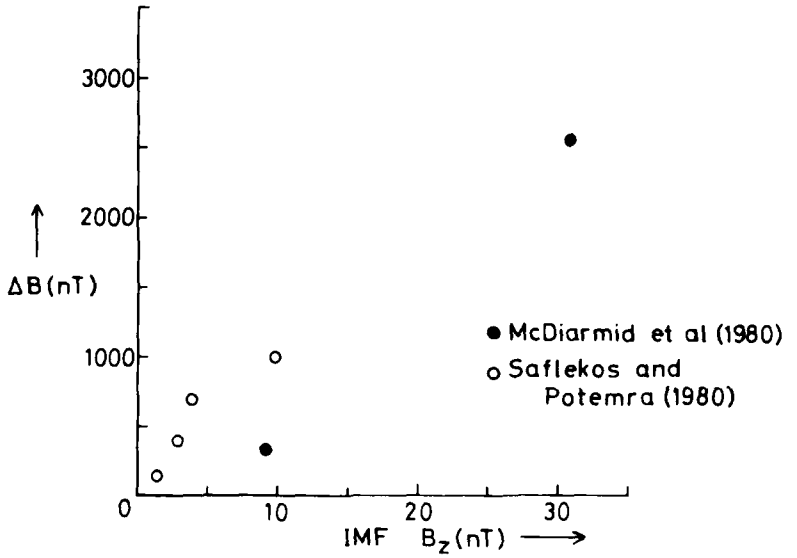


Figure 7. Graph of the maximum transverse field perturbation observed in the region of sunward polar cap flow versus IMF B_z for the six published passes where the IMF is known. Solid circles are from McDiarmid et al. (1980), open circles from Saflekos and Potemra (1980). A positive correlation is evident.

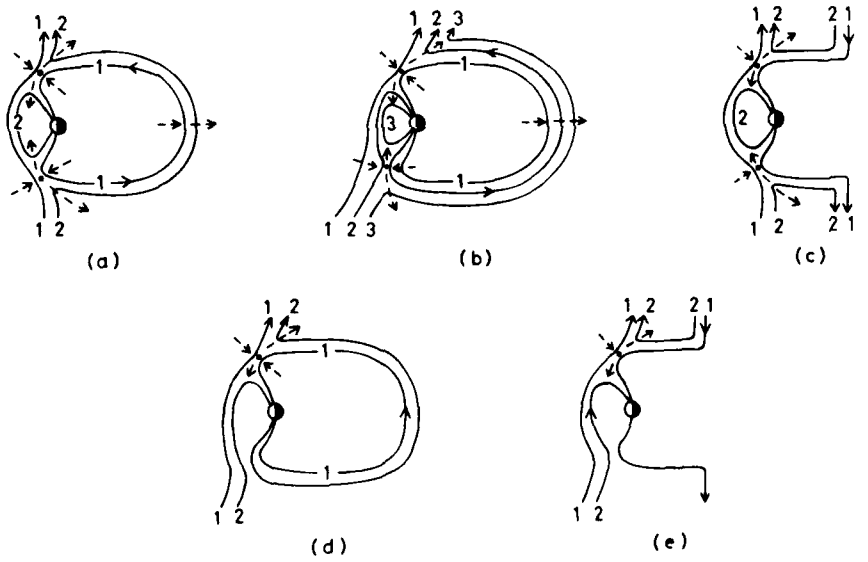


Figure 8. Schematic sketches of reconnection between northward directed IMF lines and open and closed tail lobes. In (a) to (c) connection occurs to both lobes, not simultaneously in (b), while in (d) and (e) connection of a given IMF line occurs to only one lobe due to IMF B_y although other IMF lines may similarly be connecting to the other lobe. The numbers indicate the field line sequence and the heavy dot the location of the neutral line. The dashed lines indicate the flow in the vicinity of the latter. Case (a) is that discussed by Dungey (1963), case (d) by McDiarmid et al., (1980) and case (e) by Russell (1972), Maezawa (1976) and Crooker (1979).

SESSION DISCUSSION

SESSION I – THE SUN, SOLAR WIND AND MAGNETOSPHERE

Chairman and Editor – Prof. S.H.Gross
Polytechnic Institute of New York
Route 110
Farmingdale, N.Y. 11735, USA

THE SOLAR ULTRAVIOLET SOURCE FOR THE IONOSPHERE AND ITS VARIATION
by H.E.Hinteregger

J.S.Nisbet, US

It would be very useful for many purposes associated with ionospheric modeling to have simple solar indices based on EUV measurements. It would be desirable, for example, to have an index of the intensities of those lines ionizing the F region, others for the D and E regions, and one for thermospheric heating. These would be much more satisfactory than current indices such as R_Z or $F_{10.7}$. Do you think these indices are possible?

Author's Reply

My answer is either yes or no, depending on how we approach the problem. If the preparatory steps of the computational procedure include (a) all of the known spectral details of all relevant types of absorption and ionization cross sections in connection with a fully detailed EUV reference spectrum and (b) detailed instructions for the conversion of the reference spectrum into an equally detailed one for the required modeling date, based on the input of date-peculiar values of several EUV indices such as those for at least three "classes" described in my COSPAR 1980 paper, *then* one can provide any desirable "simple solar indices" such as those you mentioned or any desirable others as well. My conviction is simply to emphasize that this desirable goal of simplicity should be accomplished as a *result of detailed computations*, i.e., not via an irretrievable loss of detailed known aspects within the computations.

J.Forbes, US

Wouldn't the product of absoprtion coefficient times solar flux, integrated over the appropriate wavelength ranges for the appropriate constituents, form a more physical data base for construction of indices to predict perturbations in ionospheric and neutral atmosphere densities?

Author's Reply

My answer to this question is to emphasize that both "simple indices" and any desirable set of "composite indices" could be produced most reliably as an appendix to detailed standard computations, indeed giving special results from suitable programs accessing the physically most significant data base available. I have no comment on predictive capabilities other than expecting progress from first studying variabilities in past observations.

R.Sharp, US

A comment on a possibly related phenomena that may be of interest to this group is that certain satellites are showing anomalous long term warming trends which have been attributed to be at least partially due to thermal control surface degradation associated with the unusually high EUV fluxes reported by Dr Hinteregger during this current solar cycle.

Author's Reply

I cannot offer any specific comments on either long-term warming or surface-degradation. In general terms, let me just repeat that both the conventional solar activity indices and the EUV irradiance measurement have indeed shown that the present solar cycle has a maximum much higher than the previous cycle 20.

SOLAR RADIATIONS AND THEIR INTERACTIONS WITH THE TERRESTRIAL ENVIRONMENT
by G.Schmidtke

H.E.Hinteregger, US

Dr Schmidtke (Paper No. 2) mentioned calculations of ionospheric response to solar EUV variations published by Roble and Schmidtke; I believe their use of an invariant model of the neutral atmosphere in connection with

otherwise very advanced ionospheric modeling details reflects the great persisting difficulties of a proper treatment of the very complex intertwinement of neutral atmospheric and ionospheric response to solar EUV variations. This complexity indeed increases drastically from the upper F_2 region downward, in the F_1 and E regions in particular.

Author's Reply

The study by Roble and Schmidtke has been conducted further applying the MSIS atmospheric model [Schmidtke, Ann. Geophys., t. 35, 141 (1979)]. The conclusions are still the same as derived from the static atmospheric model. However, it is true the problems involved are of such complex nature, that these studies must be considered as being at an early stage, yet.

E.R.Schmerling, US

Nature seems to have partly solved the difficult problem discussed by the two last speakers in providing two surprisingly good indices: the sunspot number and the 10.7 cm flux. We now need to make further progress in two directions:

1. To develop our understanding of the sun; to understand the generation of sunspots and the production of the 10.7 cm emissions.
2. To provide a theoretical framework for the formation of the ionosphere that accounts for:
   (a) The solar radiation that controls the neutral atmosphere.
   (b) The solar radiation that ionizes the neutrals.
   (c) The effects of coupling-in the solar wind.

From such a theoretical framework it should be possible to derive a model with indices better than the sunspot number or the 10.7 cm flux, based on an understanding of the physical interactions of the sun with the earth's environment and the variability of the solar emissions.

Author's Reply

It must be recalled that the successful application of those indices to stratospheric modeling was performed by semi-empirical fitting to a large amount of experimental data from mass spectrometers and incoherent scatter stations. The application of the models to other periods are not always satisfactory [Schmidtke, Space Research XIX, 193 (1979)]. Because of the complexity involved and of the different physical backgrounds of the indices mentioned compared to the EUV, I do not see a realistic way to construct the theoretical framework requested.

Supplementary answer by H.E.Hinteregger

Let me just emphasize what I believe to be the most relevant requirement for future "theoretical" model developments, i.e., to retain all known aspects of all known spectral detail of solar emissions and cross sections, etc., within the computational structure of any modern theoretical program and to try to obtain the various desirable goals, such as specific "simplicity" from a given viewpoint, or better theoretical transparency as results of continuing research in the future.

MAGNETOSPHERIC AND IONOSPHERIC FLOW AND THE INTERPLANETARY MAGNETIC FIELD
by S.W.H.Cowley

A.Egeland, No

During substorm activity in the dayside cusp aurora both optical and magnetic activity correlate both in time and magnitude with night-time activity, and probably less closely with the Interplanetary Magnetic Field (IMF). Would you comment on that? Do you believe that IMF is more important for the dayside cusp during quiet than disturbed conditions?

Author's Reply

There seems little doubt that some magnetospheric processes, notably substorms, are not related on a one-to-one basis with IMF conditions, in the sense of an IMF "trigger" for substorms, although negative IMF B_z is necessary to predispose the magnetosphere to substorms. In the dayside cusp we should expect to see responses both to substorms and more directly to IMF conditions via magnetic connection of cusp field lines into the solar wind. Only the latter effects were discussed in the paper due to the emphasis on direct solar-wind-ionosphere coupling, but it is certain that direct response to the IMF and to substorms both occur. I think we still have a lot to learn about the high-latitude dayside region and the relative importance, the relationship and the interaction between the two sources of influence which you mention. I don't think we can be definite about these points at the present time.

J.Forbes, US

What is the response time-constant associated with full shifts of the asymmetric pattern of the polar ionospheric current system following a change in direction of B_y, and how small is this compared to the average time between changes in the sign of B_y?

Author's Reply

In general the solar wind B_x and B_y field components are governed by the Archimedes spiral direction and the section structure, so that B_y can have a preferred sign for many days on end. However, within each sector B_y can fluctuate in sign on shorter time scales, such that, for example, within a given sector of positive B_y one will often find hourly averaged B_y fields which are negative. With regard to the magnetospheric-ionospheric response time to changes in B_y, I do not think that this is as yet well determined experimentally, but if one considers for example the response time for changes to occur on newly opened flux tubes in the dayside cusp then the time scale should be very rapid, corresponding to the time scale for Alfven wave travel between the dayside magnetopause to the ionosphere, which is minutes. On the other hand, for the asymmetry patterns to develop fully throughout the magnetosphere one may have to wait for time scales of the order of the convection time scale, which is hours. B_y is often constant in sign on such time scales, however.

M.Lockwood, US

Recently published observations from Atmospheric Explorer C indicate that convection flow reversals do not always occur at the boundary of the open field lines. Can you explain how this arises?

Author's Reply

Antisunward convection occurs on open field lines over the polar cap and on adjacent closed field lines mapping into the boundary layer. The flow reversals, therefore, actually occur on closed field lines, at least in and near the dawn-dusk meridian. There seems no reason to suppose that reconnection at the dayside magnetopause and viscous interaction are mutually exclusive processes, and observations show pretty clearly that both occur. The real question is the relative importance of these processes, measured in terms of their contribution to the cross-magnetosphere potential. Information on this is rather hard to come by, but under normal conditions it seems that they may be roughly comparable in importance, each process contributing a few tens of kV. During disturbed times the cross-magnetosphere potential may go as high as 200 kV, reconnection then probably being completely dominant.

J.S.Nisbet, US

Satellite measurements in the cusp show that cusp currents are present at all times although they increase with increasing B_z. Does this not mean that the basic ionospheric current systems associated with the four cell convection pattern are always present even though they may only be seen in electric field and magnetometer measurements 10% of the time when they become large?

Author's Reply

No, I don't think it does. I don't see any evidence to suggest that small high-latitude cells occur in the polar cap under usual conditions, and that these simply become bigger as B_z gets more positive. It is, I suppose, just possible, but I think it much more likely that the 3 or 5 cell patterns are due to a new process occurring when B_z gets rather positive as I discuss in the paper. I should also mention that the cusp currents are very different in direction for B_z positive ($\gtrsim$ 1nT) and negative. In the latter case the resulting transverse field perturbations are largely east-west directed depending upon IMF B_y, consistent with a distorted two-cell pattern. In the former case, they are predominantly noon-midnight directed, consistent with a 3 or 4 cell pattern.

MAGNETOSPHERIC INFLUENCES ON THE IONOSPHERE

H. Kohl
Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, FRG

SUMMARY

The magnetosphere influences the ionosphere mainly at high latitudes where magnetospheric particles, electric fields and waves have direct access to the ionosphere along the linking magnetic field lines. The resulting effects like production of ionisation by penetrating particles, fast convection due to electric fields, heating by dissipation etc. make the high latitude ionosphere basically different from that at lower latitudes and certainly much more variable. A number of the problems like the polar cap ionosation, the formation of the trough, the energy input from Joule heat and others will be discussed.

1. INTRODUCTION

The subject of ionosphere-magnetosphere coupling is a very wide one and here only a few problems can be discussed, but more aspects of the subject will be presented by several authors in this symposium, e.g. by Axford, Cowley, Brekke and others.

In general the magnetosphere influences the ionosphere by particles, electric fields and plasma waves originating in the magnetosphere and propagating downwards. The penetrating particles produce ionisation and optical emissions, the electric fields make the ionosphere above about 150 km to take part in magnetospheric convection, and they also induce electric current flows in the lower ionosphere. The waves associated with Pc5 effects are an example of plasma waves of magnetospheric origin that are effective in the ionosphere. All of the three phenomena also deposit energy in the ionosphere or upper atmosphere.

2. PARTICLES

Mainly particles in the energy range below some tens of keV are responsible for producing ionisation. Sometimes harder particles penetrate deeper and cause anomalous D-layer ionisation.

Fig. 1 shows calculated ion/electron production rates for different energies (Banks, Chappell and Nagy, 1974). It can be seen that soft particles produce ionisation higher up than hard particles. This is not only because particles of lower energy get stuck in the atmosphere after a few ionizing collisions, but also because the cross section for ionisation is larger for low energy electrons. It can be clearly seen in Fig. 1 that at 200 km height, for instance, 10 keV electrons are less ionizing than 1 keV electrons. In general F-layer ionisation is produced by particles around 1 keV and below, while E layer ionisation needs electrons of about 10 keV.

It may be noted that the model calculations shown in Fig. 1 have been performed in a very sophisticated way including the pitch angle distribution of the incident particles, pitch angle scattering, the contribution of secondary electrons and so on.

Fig. 2 (Knudsen, Banks, Winningham and Klumpar, 1977) shows an ionisation rate vs. height profile for different regions of the polar ionosphere. Measurements of energetic particle fluxes, which were considered typical, were used together with calculations like in Fig. 1 to derive such curves. The results show that particles can produce ionisation rates of the same order as EUV-radiation in the auroral oval and in the cleft, which is the region where the magnetic field lines are connected to the dayside boundary of the magnetosphere and where magnetosheath plasma has more or less direct access to the magnetosphere and even to the ionosphere.

It is also possible to do the reverse operation, that is to deduce from observed electron density vs. height profiles the flux and energy spectrum of incident electrons. One has first to derive the ionisation production rate from the measured electron density using assumptions about the loss rate, and then, so to say, expand the production rate into a series of functions of the kind shown in Fig. 1. The coefficients of the expansion form the energy spectrum. Of course, this procedure is only possible when sunlight is absent, because otherwise it can not be distinguished between particles and EUV as causes. A further restriction is that at greater altitudes diffusion increases and complicates the relation between electron density and production rate. An example is given in Fig. 3, where the electron density was measured by the Chatanika Radar in several beam directions. Direction "D" goes right through an auroral arc and the corresponding energy spectrum has the hardest particles (Vondrak and Baron, 1976).

Sometimes, during proton events for instance, particles in the MeV range penetrate to very low altitudes and produce anomalous D-layer ionisation. Fig. 4 is an example measured by the Chatanika Radar (Watt, 1975). The production rate Q was calculated from proton flux measurements by the 1971-089 satellite; the effective loss coefficient α was determined as the ratio of production and density.

The energy that is carried by the penetrating particles is finally converted into heat, except a small fraction that goes into optical emissions. The amount of such heat releases is typically of the order of tens of mW/m^2 for the region above 100 km during particle events. This is about the same order as the heat input from solar radiation.

3. CONVECTION DUE TO ELECTRIC FIELDS

Electric fields map down from the solar wind along open field lines and produce a potential drop across the polar cap that leads to an antisunward convection of the ionospheric plasma above about 150 km. However, the electric field can not be kept restricted to the polar cap, but it will spread out to lower latitudes. The formal reason is that curl $\underline{E} = 0$; but the mechanism that determines the electric potential distribution towards lower latitudes is the anisotropic conductivity, which together with the condition of divergence free current flow leads to the well-known two cell convection pattern. For more detailed explanation see Vasyliunas (1975) and for details about the shape of the pattern Cowley (this volume).

It may be noted here, that the extension of the electric field to lower latitudes is limited by the influence of the ring current. At the inner edge of the ring current east-west gradients in particle density occur that lead to field-aligned currents. Vasyliunas (1972) has shown that this process is equivalent to a strong increase in integrated Hall conductivity shielding the electric field from low latitudes. Fig. 5 shows results of a model calculation for the case that the equivalent Hall conductivity is ten times the real one. It can be seen that at the inner edge of the ring current, which was assumed at several L-values in the calculation, the north-south electric field drops by more than an order of magnitude.

The velocity of the convection at F-layer heights is about 1 km/s depending on the degree of disturbance. This high velocity leads to several consequences that make the behaviour of the high latitude ionosphere much different from that at medium latitudes. They will be discussed now.

Firstly, an anomalously high loss rate can result from the large difference of speed between the charged particles and the neutral gas. This is equivalent to an increase of temperature, because for a charge exchange reaction between an ion and a neutral it is unimportant whether the relative velocity is due to random thermal motion or bulk motion. Banks, Schunk and Raitt (1974) found that the rate of the reaction $O^+ + N_2 \rightarrow NO^+ + N$ is increased by a factor of 3 for a convection velocity of 1 km/s and by a factor of 18 for 2 km/s compared to zero velocity and T = 1000 K. It is evident that such strong changes in rate coefficients can appreciably affect the F-layer electron distribution.

A question that is related to F-layer convection is what causes the ionisation over the polar cap in winter. Fig. 6 shows measurements from ISIS 2 over one orbit from midnight to noon in winter (Whitteker, Shepherd, Anger, Burrows, Wallis, Klumpar and Walker, 1978). In the lower part the electron density variation is drawn for several altitude levels and the upper part gives energetic electron fluxes from three energy channels. At $\Lambda \approx 78°$ on the noon side there is a pronounced maximum of electron density at all heights. It corresponds very well with a peak in the electron precipitation in the lowest energy channel. This is in agreement with the above mentioned result that the F-layer ionisation should be mainly produced by electrons at about 1 keV and below. The region of this maximum is the so-called cleft which has been mentioned in relation to Fig. 2. Above the polar cap the measured electron density is about an order of magnitude lower than at the cleft, while the soft electron flux is smaller by about two orders of magnitude. Thus, it seems that the polar cap F layer can not completely be explained as a consequence of precipitation. Alternatively, one has to discuss what the antisunward convection that moves ionisation from the dayside cleft across the polar cap contributes to the electron density above the cap. Possibly the ionisation observed there is originally produced in the cleft and then spread over the cap. The difficulty with this idea is that if the convection across the cleft is fast, say 1 km/s, then all ionisation produced there will be swept over the a few hundred kilometers wide cleft region within a few hundred seconds into the polar cap, and one would not expect such a pronounced cleft maximum to develop. On the other hand, a slow convection means that a maximum can build up, but it would take a long time to move the ionisation across the polar cap and it would decay in between. I do not see an obvious solution of this problem. It must, however, be remembered that Fig. 6 shows one particular orbit, which allows us to discuss typical problems. But by no means it represents a standard high latitude F-layer, which is something that hardly ever exists.

A further problem related to convection is the existence of a region of low F-layer electron density equatorwards of the auroral oval, the so-called trough, which is most pronounced on the nightside. Several explanations have been suggested including vibrational excitation of N_2 (Schunk and Banks, 1975) and the above mentioned anomalously high loss rate due to the relative bulk velocity between ionisation and neutral air (Banks, Schunk and Raitt, 1974). Here another point shall be discussed. Fig. 7 shows a schematic convection pattern, on the right hand side seen from an observer rotating with the earth and on the left hand side seen from an observer from outside. That means, on the left side of the diagram the rotation velocity of the earth is added. On the evening side of the left hand pattern a stagnation point appears, which means that the westward convection is cancelled by the eastward rotation of the earth. In the left hand side of Fig. 8 the same situation is presented close to the stagnation point in a schematic way. The dots on the flow lines denote the distance over which a plasma volume moves in hourly intervals. It can be seen that lower latitude flow lines just corotate with the earth, while the high latitude flow joins magnetospheric convection. In between there is a region where plasma moves slowly from the dayside into the nightside, then swings around and slowly flows back. On the innermost flow line on that graph it takes 11 hours for the plasma to move from the 18h meridian to the stagnation point. The idea is that it is this long time that makes the electron density decay by recombination and, thus, causes the trough. The right hand side of Fig. 8 shows the result of a model calculation by Spiro, Heelis and Hanson (1978). The development of a sharp trough is clearly seen. Of course, this can only explain the trough formation at the evening side. Other processes, as mentioned above, may be responsible elsewhere.

4. ELECTRIC CURRENTS CAUSED BY ELECTRIC FIELDS.

The same electric fields that cause the convection of the ionosphere above 150 km produce electric currents at lower altitudes. This happens because the electrons still have the same convection velocity $(\underline{E} \times \underline{B})/B^2$ while the ions as a consequence of increasing collision frequency now move slower and in a different direction, which means that an electric currents is set up. Such a current will release Joule heat in the lower ionosphere, but it is only that component of the current parallel to the driving field that contributes to Joule heat, because according to Pointing's theorem the only way to draw energy from an electric

field is to have a field parallel current.

The amount of Joule heat under disturbed conditions can be easily larger than the solar EUV-input. Fig. 9 shows results from measurements with the Chatanika Radar (Brekke and Rino, 1978). The Joule heat was derived from measurements of the electric field and the ion velocity. The peak energy dissipation is 100 mW/m² in the altitude region above 90 km. This is clearly more than the solar EUV input, although the latter is difficult to compare, because it is also variable and one has to know the spectral range that is absorbed above 90 km. A reasonable estimate may be 10 or 20 mW/m². Such high Joule dissipation, however, occurs only sometimes and in a limited spatial area. Nevertheless, Joule heat is probably of importance for the total energy balance of the upper atmosphere.

This is not only true for the high latitude atmosphere, but is also relevant for lower latitudes, as Joule heat produced in the auroral atmosphere can be transported equatorwards by gravity waves and wind circulation. These latter phenomena are set up when the energy input leads to a heating and expansion of the upper atmosphere. A drastic heat release on a short time scale will excite gravity waves, while a more steady heating will generate a neutral air circulation system. Fig. 10 shows neutral air compositions for several orbits of ESRO 4 (Jacchia, Slowey and v. Zahn, 1977) during disturbed conditions. It is obvious that an atmospheric distortion spreads from high latitudes over the whole globe and the wave-like structure is apparent on the graph.

An interesting approach has been made by Schlegel and St. Maurice (in press), who also used the Chatanika Radar and measured electron and ion temperatures. Fig. 11 shows their result for an active period. It turns out that both temperatures are very well correlated with the electric field. However, at an altitude of 137 km the temperature increases are strong in ion temperature but not in electron temperature, while at 112 km the opposite happens. Here T_e is dramatically peaking at 1400 K.

The behaviour of the ions is understandable, although temperatures of up to 1700 K at 137 km were not expected. The maximum of the ion velocity parallel to the E-field should happen to be at about the upper gate i.e. at 137 km. This is because at altitudes well above this level the ions more or less perform an $\underline{E} \times \underline{B}/B^2$-drift perpendicular to $\underline{E}$, while at much lower altitudes, e.g. at 112 km, collisions are so frequent that the ion velocity is small. Thus, it is not surprising that the ion draw much energy from the E-field at 137 km, but not at 112 km.

The situation is different for the electrons. They should join the $\underline{E} \times \underline{B}/B^2$ - drift at both altitudes with a negligible velocity component antiparallel to the electric field. In order to explain the discrepancy to the observations Schlegel and St. Maurice have worked out a theory that involves an additional energy input from the dissipation of plasma instabilities. In my interpretation that would mean that such instabilities finally lead to a velocity component of the electrons antiparallel to the E-field or, in other words, the electrons can contribute appreciably to Pedersen conductivity. If this is true, then, Joule heating in the lower E region can be roughly twice as high as has been thought so far.

5. WAVES

Only one example for waves of magnetospheric origin may be mentioned here. STARE (Scandinavien Twin Auroral Radar Experiment) does Doppler-measurements of instabilities in the auroral E-layer, which are interpreted in terms of an $\underline{E} \times \underline{B}/B^2$ - drift of the electrons. During Pc5 events a regular behaviour of the spatial velocity pattern was observed. Fourier analysis of both north-south and east-west components of the velocity was performed in the time domain for each spatial point of the observation grid and the result presented in graphs like Fig. 12 (Villain, to be published). An interesting polarization pattern is achieved. The upper left of the figure shows a counterclockwise elliptical polarization, i.e. the velocity vector or the corresponding E-vector rotates with the period of the pulsation. Then there is a region of linear polarisation and finally of clockwise elliptical polarization. This is in agreement with theoretical considerations by Southwood (1974), who derived this kind of polarization behaviour for the time before noon from a field line resonance model.

6. References

Banks, P.M., Chappell, C.R., and Nagy, A.F., 1974a, A new model for the interaction of auroral electrons with the atmosphere: spectral degradation, backscatter, optical emission, and ionization, J. Geophys. Res. 79, 1459-1470.

Banks, P.M., Schunk, R.W. and Raitt, W.J., 1974b, NO^+ and O^+ in the hign latitude F-region, J. Geophys. Res. Let., 1, 239-241.

Brekke, A. and Rino, C.L., 1978, High-resolution altitude profiles of the auroral zone energy dissipation due to ionospheric currents, J. Geophys. Res. 83, 2517-2524.

Jacchia, L.G., Slowey, J.W. and v. Zahn, U., 1977, Temperature, density, and composition in the disturbed thermosphere from ESRO 4 gas analyzer measurements, J. Geophys. Res. 82, 684-688.

Knudsen, W.C., Banks, P.M., Winningham, J.D., and Klumpar, D.M., 1977, Numerical model of the convecting F2 ionosphere at high latitudes, J. Geophys. Res. 82, 4784-4792,

Schlegel, K., St. Maurice, J.P., Anomalous heating of the polar E region by unstable plasma waves.I. Observations, J.Geophys. Res., in press.

Schunk, R.W. and Banks, P.M., 1975, Auroral N_2 vibrational excitation and the electron density trough, Geophys. Res. Let., 2, 239-242.

Spiro, R.W., Heelis, R.A. and Hannon, W.B., 1978, Ion convection and the formation of the mid.-latitude F-region, J. Geophys. Res., 83, 4255-4264.

Vasyliunas, V.M., 1972, The interrelationship of magnetospheric processes, in McCormac (ed.): Earth's magnetospheric processes, D. Reidel Publ., 29-38.

Vasyliunas, V.M., 1975, Concepts of magnetospheric convection, in Formisano (ed.): The magnetosphere of the Earth and Jupiter, D. Reidel Publ., 179-188.

Villain, J.P., Characteristics of Pc5 micropulsations as determined with the STARE Experiment, to be published.

Vondrak, R.R. and Baron, M.J., 1976, Radar measurements of the latitudinal variation of auroral ionization, Radio Sci. 11, 939-946.

Watt, T.M., 1975, Effective recombination coefficient of the polar D-region under conditions of intense ionizing radiation, Stanford Research Institute, Rep. DNA 3663T.

Whitteker, J.H., Shepherd, G.G., Anger, C.D., Burrows, J.R., Wallis, D.D., Klumpar, D.M. and Walker, J.D., 1978, The winter polar ionosphere, J. Geophys. Res., 83, 1503.

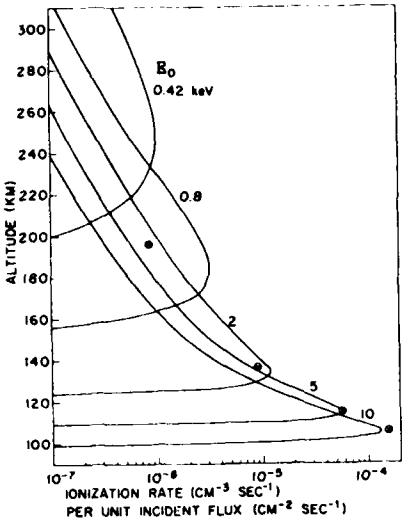


Fig. 1 Ionisation rate as a function of altitude for penetrating electrons of various energies (after Banks et al., 1974a).

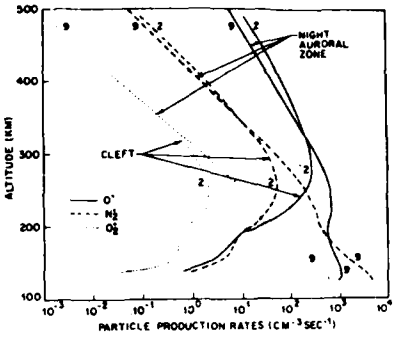


Fig. 2 Production rate of ionisation at the cleft and midnight auroral oval for typical particle penetration fluxes (after Knudsen et al., 1977).

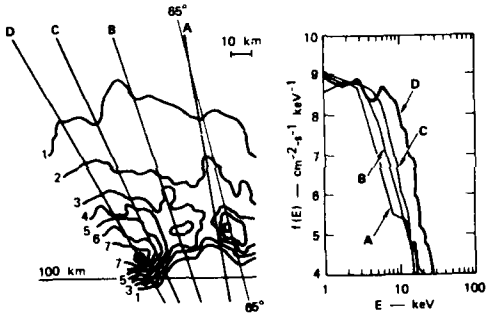


Fig. 3 Left: Electron density distribution around an auroral arc measured by the Chatanika Radar. Right: Derived electron energy spectra for four beam directions (A-D) of the radar (after Vondrak and Baron, 1976).

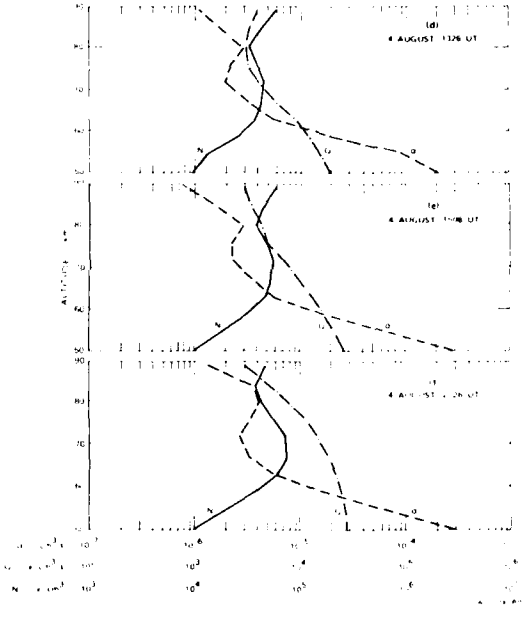


Fig. 4 Electron density N measured by the Chatanika Radar during a proton event. Production rate Q derived from particle measurements of satellite 1971-089 (after Watt, 1975).

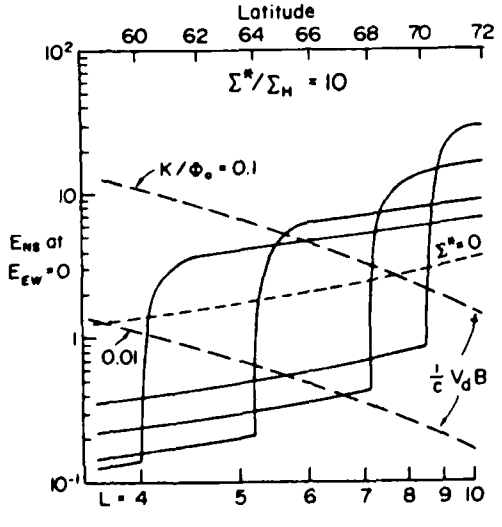


Fig. 5 Shielding of the north-south electric field by the ring current for an assumed ratio of effective to real Hall conductivity $\Sigma^*/\Sigma_H = 10$ for different positions of the inner edge of the ring current (after Vasyliunas, 1972).

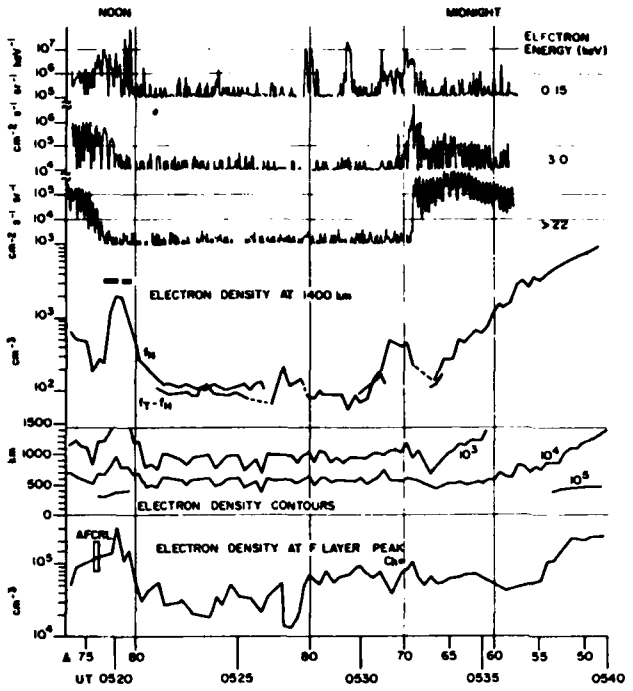


Fig. 6 Electron density variations and energetic electron fluxes for an ISIS 2 pass over the polar cap from the noonside to the midnightside (after Whitteker et al., 1978).

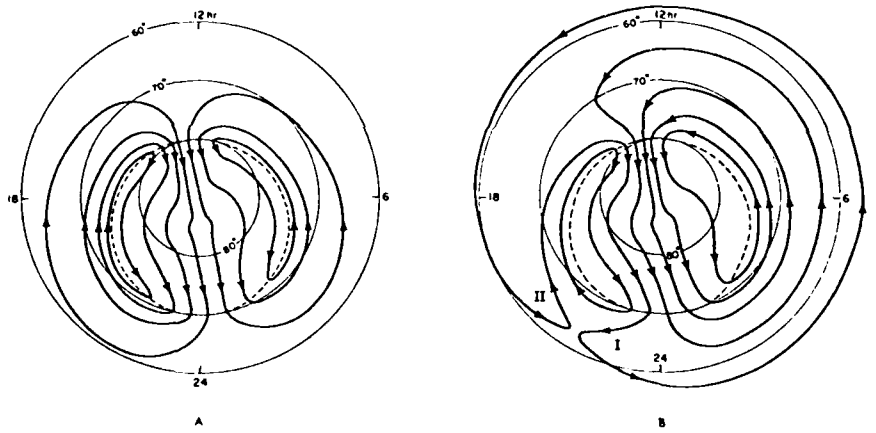


Fig. 7 Schematic diagram of two cell magnetospheric convection. A: seen from an observer rotating with the Earth; B: seen from an observer from outside.

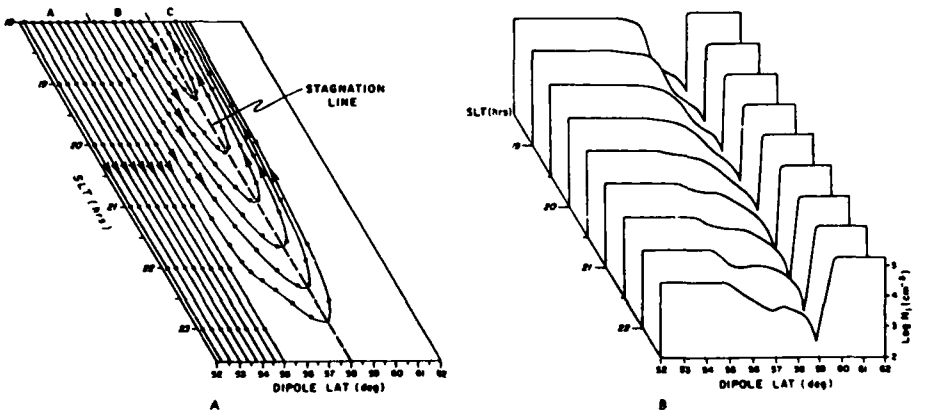


Fig. 8 Model calculation for the formation of a trough on the eveningside. Left: Schematic convection model for the eveningside seen by an observer from outside. Dots denote intervals of one hour. Right: Result of model calculation (after Spiro et al., 1978).

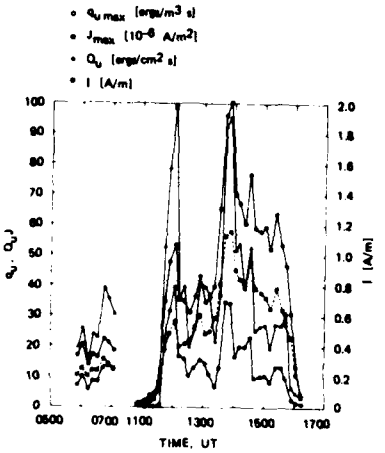


Fig. 9 Joule heat and electric current in the E layer, maximum value and height integrated, from measurements of the Chatanika Radar (after Brekke and Rino, 1978).

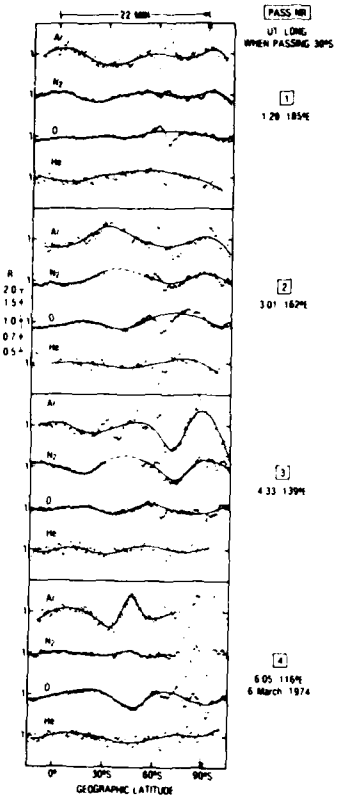


Fig. 10 Composition of the atmosphere along several passes of ESRO 4 (after Jacchia et al., 1977).

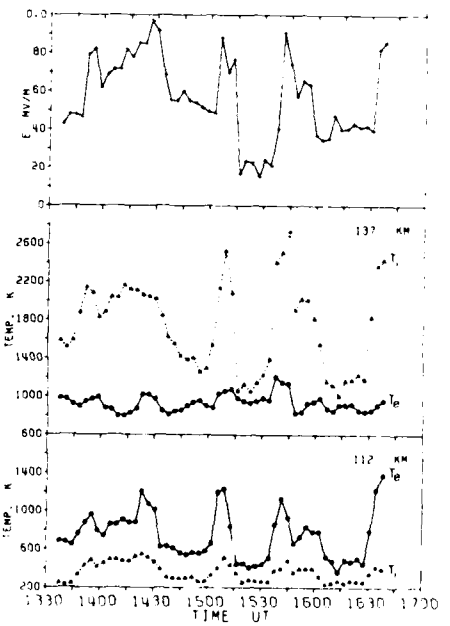


Fig. 11 Electron and ion temperatures at 112 km and 137 km altitude together with the electric field measured by the Chatanika Radar during an active period (after Schlegel and St. Maurice, in press).

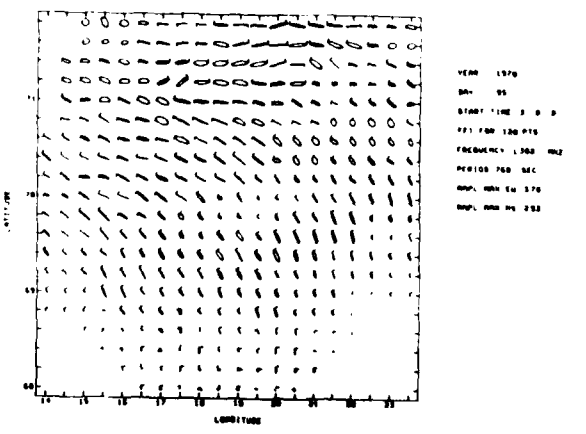


Fig. 12 STARE measurement of electron drift during a Pc5 event. In the upper left the vector rotates counterclockwise, in the lower right clockwise, in between it oscillates linearly (after Villain, to be published).

IONOSPHERIC DISTURBANCES OF MAGNETOSPHERIC ORIGIN

H. Poeverlein
Technische Hochschule Darmstadt
D 6100 Darmstadt, Fed.Rep.Germany

E. Neske
Fraunhofer-Institut für Physikalische Meßtechnik
Heidenhofstr. 8
D 7800 Freiburg, Fed.Rep.Germany

ABSTRACT

Various types of coupling between magnetosphere and ionosphere are distinguished. Some phenomena occurring in connection with geomagnetic disturbances give a clear evidence of magnetosphere-ionosphere coupling. The displacement of the midlatitude trough of the ionosphere to lower L values (lower latitudes) with increased geomagnetic activity,which is shown on the basis of AEROS-B observations, corresponds to a similar displacement of the plasmapause. Some of the related observational facts may give a hint at underlying processes. The large-scale magnetospheric convection is demonstrated by theoretically derived pictures of streamlines. Though there is general agreement on the explanation of the plasmapause in terms of these motions, certain characteristic details depend on various factors. From some crude approximative relationships an expression for the location of the plasmapause in dependence on K_p is obtained. Ionospheric storms require various processes, involving electrodynamic coupling in the positive phase of the storms and varying composition of the atmosphere in the negative phase.

1. INTRODUCTION

The regular behavior of the ionosphere is determinend by ionization processes, by mechanisms of decay of ionization, and by some state of worldwide motion. Ionization is due to solar UV and corpuscular radiation that is preferably incident at high latitudes. Disturbances of the ionosphere may be caused by solar events that affect the incident radiations or they may originate from magnetospheric phenomena, which ultimately are also the result of variable solar influences.

The following section 2 gives a short list of coupling processes linking the ionosphere with the magnetosphere. Section 3 deals with the midlatitude trough of the ionosphere and the magnetospheric plasmapause. The trough and the plasmapause offer a clear case of evidence of a relation between magnetosphere and upper ionosphere (F2 layer). The trough, on the other hand, is also interesting as separation line between midlatitide and high latitude ionosphere, both showing quite different characteristics. In section 4 large-scale magnetospheric motion is discussed with a view at the plasmapause as the boundary between regions of closed and open magnetic field lines. The boundary is known to shift with varying K_p, in other words, with varying solar wind intensity. Ionospheric storms, i.e. F2-layer storms accompanying geomagnetic disturbances, are the subject of section 5. Positive and negative phases of these storms seem to be expressions of quite different processes.

2. BASIC COUPLING PHENOMENA

The magnetosphere is characterized by the presence of atmospheric plasma and the geomagnetic field and by the validity of the laws of magnetohydrodynamics. Since these laws are still valid in the upper ionosphere, down to roughly 150 km, there is a strong electromagnetic coupling between ionosphere and magnetosphere. (Some authors even consider the ionosphere from 150 km up as part of the magnetosphere, despite the typical ionospheric processes observed there.) In consequence of the coupling the large-scale motions of the magnetosphere lead in the ionosphere to an electric field, to drift motions, currents and modifications of the ionosphere (electron density and vertical structure).

Precipitation of particles from the magnetosphere is another source of ionospheric phenomena. Protons are precipitated at higher latitudes on field lines passing through the radiation belt of ring current region. Disturbances, e.g. the penetration of protons in the energy range 1 to 100 keV from the magnetotail into the ring current region with presence of thermal plasma, may cause instability, perhaps cyclotron instability with a fall-out of the fast particles.

A variety of magnetospheric phenomena can lead to a deposit of energy in the ionosphere. Currents and motions in the ionosphere, resulting from electromagnetic coupling, may,for example,be accompanied by a dissipation of energy in the lossy medium. Particles incident from the ring current deliver their kinetic energy (at geomagnetic latitudes around and

beyond 60°). Waves originating in magnetospheric instabilities, perhaps in connection with corpuscular events, are another carrier of energy. Deposit of energy means heating and possible modification of ionospheric chemistry. An elevation of the turbopause with an increase of the concentration of diatomic molecules at a particular height might result from the temperature increase. Processes of this type apparently play a role at middle and high latitudes.

Diffusion of charged particles at a large rate takes place at latitudes beyond the ionospheric trough, where field lines extend into the magnetospheric region of low plasma density, the 'plasma trough' of the magnetosphere. Plasma departing from the polar region with a fairly high velocity (~ 10 km/s) and consisting of light ions with electrons has been observed and is called the 'polar wind'. At night the ionization decay in the ionosphere may entail a diffusion of plasma from higher altitudes into the lower F region.

3. IONOSPHERIC TROUGH AND PLASMAPAUSE

The midlatitude trough (MLT) is an electron density minimum quite striking in observations of the topside ionosphere over a range of geomagnetic latitudes around roughly 60° (NESKE, 1978). At night it is almost always observed and even seen at F2 maximum or below. In daytime it is present frequently, but not at all local times (see table 1). The appearance during the day is more common at higher altitudes (> 2500 km); a light-ion trough (LIT) may be seen at lower altitudes.

Remarkable is the displacement of the MLT to lower geomagnetic latitude with increasing geomagnetic activity. MLT and plasmapause are located nearly on the same field line shell, i.e. at the same value of the McILWAIN parameter L, varying in the same way with K_p. Fig. 1 shows the L values in dependence on K_p. The dots indicate locations of the MLT at night according to measurements on the satellite AEROS-B in 1974 and 1975(NESKE and KIST, 1974; SPENNER and DUMBS, 1974).Empirical approximations for the position of MLT and plasmapause based on results of various observation techniques are also represented. The steeper solid line in Fig. 1 refers to a theoretical approach discussed in the next section. The trough is in the average found at slightly lower L than the plasmapause.The good correspondence seen on a statistical basis is however not evident in the course of disturbances, which involve some time delay.

Table 1

PLASMAPAUSE / MIDLATITUDE TROUGH OBSERVATIONS

		LOCATION	OCCURRENCE	PROPERTIES / MOTIONS	DEPENDENCES
DAY	PLASMAPAUSE	L ≥ 5.5 ✳ L = 4 - 5 ▽	PERMANENT	T_E - PEAK ▽ NO AFTERNOON BULGE (MIDLAT.) ▽ AFTERNOON BULGE (EQUATORIAL) ✕	LITTLE DIURNAL VARIATION ▽ $L_{pp} = (4.47-0.18K_p) \pm 0.27$ L ▽
	TROUGH	<1500 KM L= 12.5±1 ✳ >1500 KM L= 12(T_1) ✳ L= 2-6 (T_2) ✳	DISAPPEARS NEAR 06 LT ┼ REOBSERVABLE NEAR 15 LT ┼ 12 LT : 48 % ✳ HIGHER WITH HIGHER K_p ○	REGULAR TROUGH DIFFICULT TO DISTINGUISH ⧺ LOCATION NORTH OF SCINTILLATION BOUNDARY ◇ DIFFERENCES TO LIT ⧺	DIURNAL ✳ NO SEASONAL ✳ UNIVERSAL TIME ○
NIGHT	PLASMAPAUSE	L = 4 (MIDNIGHT) ✳ L = 4 (K_p = 3) ▭ L = 5.6 (P_{K_p} = 0) ▭ L = 2 ($K_p \geq 7$) ▭ L = 3.5 - 5.5 (MIDNIGHT) ▽	PERMANENT	T_E - PEAK INDICATES LOCATION DURING MAGNETIC STORMS □ RECOVERY TIMES OF DAYS AFTER STORMS △	$L_{pp} = 5.64-(0.78 \pm 0.12)\sqrt{K_p}$ ▭ $L_{pp} = 5.7 - 0.47\ K_p$ ✕
	TROUGH	>1500KM L=4.4 (MIDNIGHT) ✳ <1500KM L=3.8 ✳ L = 3 - 6 ○	00 - 03 LT : >90 % ✳ HIGHER DURING WINTER ○	T_E - PEAK △] STEEPER EDGE POLEWARD ○ EQUATORWARD MOVEMENT 0.1 L/H ▭ LOCATION SOUTH OF SCINTILLATION BOUNDARY ◇ POLEWARD WALL CORRESPONDS TO THE EQUATORWARD ELECTRON PRECIPITATION REGION △	NO APPERENT ON LONGITUDE ○ DIURNAL/NO SEASONAL ✳ UT ○ WIDTH DECREASES WITH INCREASING K_p ○ $L_T = 5.64-(1.09 \pm 0.22)\sqrt{K_p}$ ▭ $L_T = 5.3 - 1.04\sqrt{K_p}$ ▯
DAWN/ DUSK	PLASMAPAUSE	DUSK L ≥ 6 (EQUATORIAL) ▽ L ~ 4 (MIDLAT.) ▽ DAWN L = 3 -5 ▽	PERMANENT	DUSK : NO BULGE (MIDLAT.) ▽ BULGE (EQUATORIAL) ✕	
	TROUGH	$L_{DAWN} < L_{DUSK}$ ○	60 % ✳	DUSK : NO BULGE ✳ DAWN : SEASONAL (SUMMER POLEWARD) ✕	DEPTH : DUSK > DAWN (SOUTHERN HEMISPHERE) △

REFERENCES :

✳ M.AHMED ET AL.,JGR,VOL82,NO A2,1979
▭ M.J.RYCROFT AND J.O.THOMAS,PLANET.SPACE SCI,VOL18,1970
— D.B.MULDREW,JGR,VOL70,NO11,1965
○ Y.TULUNAY AND J.SAYERS,JATP,VOL33,1971
◺ H.BRINTON ET AL.,JGR,VOL83,NO A10,1978
□ L.H.BRACE ET AL.,JGR,VOL79,NO 34,1974
◇ Y.TULUNAY ET AL.,JATP,VOL38,1976
▯ E.NESKE,SPACE RESEARCH,VOLXVIII,1978
⧺ H.TAYLOR AMDW.J.WALSH,JGR,VOL77,NO34,1972
▽ L.H.BRACE AND R.F.THEIS,JGR,VOL79,NO13,1974
✕ D.L.CARPENTER AND C.G.PARKS,REV.GEOPHYS.SPACE PHYS.,VOL11,1973

The variation of the positions of plasmapause and trough during geomagnetic storms is shown in Fig.2. A rapid onset of the displacement and a much slower recovery (in comparison with the recovery of K_p) are apparent. Emptying the plasmapause seems to be much faster than refilling it. The L value of the plasmapause (L_{pp}) in Fig. 2 was determined indirectly from a maximum of the electron temperature in the upper ionosphere.

In Table 1 observations referring to the plasmapause and the MLT are listed. Particularly interesting is the persistent appearance of a maximum of the electron temperature at the nighttime MLT. In daytime there is no such coincidence.

The plasmapause is the boundary of thermal plasma in the magnetosphere. Outside, only a very low density of thermal plasma is observed, but energetic (suprathermal) particles can accumulate.Presence of both thermal and suprathermal particles at the same location, perhaps near the plasmapause, is supposed to lead to cyclotron instability with particle precipitation and supply of wave energy to the ionosphere. The same process has been known as a possible cause of stable auroral red (SAR) arcs near 40° latitude.

The MLT is by some authors thought of as the transition between the region of ionization by UV and corpuscular ionization. This seems to be compatible with the idea that the plasmapause, located at nearly the same L, confines the region that contains suprathermal particles. The electron temperature maximum may be connected with energy deposit from the instability near the boundary. A consequent change of the composition of the ionosphere may offer an alternative explanation of the trough (cf. Section 5). For discussions and references concerning the present topics see the review articles by CHAPPELL (1975), BANKS (1975 and 1979) and THORNE (1975).

4. MAGNETOSPHERIC CONVECTION

The plasmapause ist generally explained as a field line shell separating closed field lines from field lines that are at least some time during the day open (or widely streched out). Field lines are supposed to 'move with the plasma'. The motion consists of two parts: a circulatory convection caused by the solar wind and rotation with the earth. Streamlines in the equatorial plane (Fig. 3) were developed in accordance with this idea. Fig. 3 (OBAYASHI and NISHIDA,1968) shows the typical picture of the equatorial plane. The streamlines are at the same time equipotential lines. The dashed line is supposed to be the plasmapause. Voltages are noted in kV. The figure is not based on strict analytical derivations.

An analytical or numerical derivation seems to be possible only with severe simplifications. The laws of stationary magnetohydrodynamics to be used are

$$\vec{E} + \vec{v} \times \vec{B} = 0$$

$$\nabla \times \vec{E} = 0$$

or

$$\nabla \times (\vec{v} \times \vec{B}) = 0.$$

The ionosphere is considered as the bottom surface of the magnetosphere and is characterized by its two-dimensional conductivity tensor, which may also be given in a simple fashion[1]. A reasonable assumption for the surface current density in the ionosphere is

$$\nabla \cdot \vec{I} = 0.$$

In place of a boundary condition at the magnetopause it is common to assume a homogeneous electric field in the polar cap of the ionosphere corresponding to a 'polar cap voltage'.

Equipotential lines calculated in this way by various authors are represented in Figs.4 - 7 (see the references in the captions). The equipotential lines are again streamlines. It has to be noticed that Fig. 4 shows convective motion only, as seen from the rotating earth. In Figs. 5 - 7 the rotation is included. Fig. 4 shows motions in the ionosphere from 50° geomagnetic latitude up. Fig. 5 represents equipotential lines in the equatorial plane for a motion model similar to that of Fig. 4a, though in some respect refined. In these figures a dipole magnetic field was for simplification assumed. Figs. 6 and 7, referring to the equatorial plane were obtained with a more realistic magnetic field, in particular paying regard to the magnetotail. Empty spaces in these two figures are due to a failure of the computational method in certain regions.

The ionospheric conductivities (HALL and PEDERSEN conductivity) used in Fig. 4a varied smoothly with location, whereas in Fig. 4b enhanced conductivities in the auroral zone were assumed. This led to an increase of the electric field in the auroral zone and a decrease at lower latitudes. The electric field at lower latitudes (L < 4)can be even more reduced by various effects. The ring current can cause a reduction (VASYLIUNAS, 1972) and certain boundary conditions may act in the same sense (cf. next paragraph). As a consequence there should be nearly no influence of the magnetospheric convection at lower latitudes (L < 4). The effect of field reduction, however, depends on the time constant of the state of motion. Dynamic (or other) processes seem to prevent the reduction for time constants $\lesssim$ 1 hr. This means that disturbances of such a duration (e.g. substorm

[1] usually without regard to the diurnal variation.

phenomena, ionospheric storms, and magnetic DP2 disturbances) may well penetrate to the lower latitude ionosphere.

Table 1 presents two examples of convective velocities as function of L, again calculated with extreme simplifications, but not based on a homogeneous field in the polar cap (POEVERLEIN, unpublished). Velocities at the boundary were prescribed instead. The relative velocities in the ionosphere (L ≈ 1) are found to depend decisively on boundary conditions. Short time constants seem also in the present model to support penetration to low L [case a) of table 2]. The limiting time constant depends on coupling between plasma and neutral gas in the ionosphere, thus increasing at night.

Table 2

Calculated convection velocities
in the equatorial plane

Two examples

	L	Maximum radial velocity (outward)		Maximum azimuthal velocity (eastward)	
		relative value	position	relative value	position
a)	10	1	0°	0	-
	4	$1.79 \cdot 10^{-1}$	1.8°	$4.38 \cdot 10^{-2}$	110.1°
	~1	$1.80 \cdot 10^{-2}$	21.3°	$1.54 \cdot 10^{-2}$	190.1°
b)	10	1	106.6°	$7.34 \cdot 10^{-1}$	21.8°
	4	$7.14 \cdot 10^{-2}$	102.1°	$7.89 \cdot 10^{-2}$	17.7°
	~1	0	-	$1.39 \cdot 10^{-4}$	0°

The motion models calculated by LANDSCHULZ (Figs. 6 and 7)indicate a significant change with increasing polar cap voltage, i.e. with increasing K_p. For higher K_p a separate flow loop appears. This may with additional processes (LEMAIRE, 1974; LANDSCHULZ, 1976) be the reason for detached plasma islands outside the plasmasphere - though an alternative explanation of the islands by plasma removal is found in the literature.

Pictures of streamlines in the equatorial plane for various polar cap voltages should indicate how the position of the plasmapause varies with the polar cap voltage. Identification of the last closed streamline with the plasmapause in pictures for four values of the polar cap voltage ($2V_o$) that LANDSCHULZ represents lead to Fig. 8. The L values were chosen for midday and midnight and are certainly very inaccurate since the last closed streamline is not well defined and no attention was paid to the pecularities of the streamline pictures. At the right side of Fig. 8 an appropriate K_p scale is introduced. A voltage of approximately 28 kV is seen to correspond to the quiet state ($K_p = 0$).

In some theoretical pictures of equatorial streamlines there appears a stagnation point of the flow on the last closed streamline. The L value of this stagnation point may be used to denote the position of the plasmapause and to derive an approximative expression for its dependence on K_p. Perfect symmetry may be assumed in a crude approximation.

At the stagnation point the convective velocity v equals the rotary velocity:

$$v = R_E L \omega$$

with the earth's angular velocity ω. The solar wind velocity has been found to be roughly proportional to K_p+5 (VASYLIUNAS, 1968). If a certain fraction of the corresponding electric field enters the magnetosphere it generates a velocity near the magnetopause that is

$$v_{MP} = \frac{E}{B_{MP}}$$

or, leaving the numerical factor open,

$$v_{MP} \sim (K_p+5)\ L_{MP}^3$$

with the magnetopause situated at L_{MP}. Numerical examples such as case b) in table 1 suggest

$$\frac{v}{v_{MP}} \approx \left(\frac{L}{L_{MP}}\right)^3$$

as a crude approximation. The given relationships with elimination of v_{MP} and v and adjustment of the numerical factor yield

$$L^2 = \frac{135}{K_p + 5}.$$

This approximation of the plasmapause position is represented in Fig. 1 (cf. the related treatment by VASYLIUNAS, 1968).

5. IONOSPHERIC STORMS

Ionospheric storms (F2-region storms) occur in connection with geomagnetic storms. In general two phases of the ionospheric storm are observed: an initial increase of N_mF2 for a few hours, called the positive phase, and a decrease for two or more days, the negative phase. Individual storms, however, can be quite varied. The appearance depends to some extent on geomagnetic latitude, season, and local time. At fairly low latitudes there is only a positive phase. The negative phase is stronger in summer than in winter. The total electron content also shows some variation, increase and decrease. The two phases are not coincident with the initial and main phases of the geomagnetic storm. The transition in the two types of disturbances does not seem to be of the same origin. For material and references on ionospheric storms refer to a recent review paper by PRÖLSS (1980).

There is no doubt that an ionospheric storm is a complex phenomenon, built up of several processes. A geomagnetic storm is initiated by a compression of the magnetosphere. Descending motion of the ionospheric plasma is to be expected from the extension of the disturbance E-field into the ionosphere, i.e. electrodynamic or electromagnetic coupling. An ensuing compression of the ionosphere would primarily mean an increace of the electron concentration.A descent of plasma all around the earth is however unlikely. It would be accompanied by an unrealistically strong increase of the magnetic field strength - if field lines 'move with the plasma'. There may be a descent (or compression) on the day side in conjunction with an ascent on the night side. An ascent of F2-layer plasma at night means a reduction of the ion loss rate, again with the consequence of an increase of the electron concentration. Height displacements of 100 km or more may well be possible. The calculated relative velocities of case a) in table 2, the case referring to short-time events, indicate a displacement of 100 km for a shift of the magnetopause at 10 R_E by nearly 1 R_E.

The long-lasting negative phase of ionospheric storms can be ascribed to an increased loss rate, which is expected from a deposit of energy in the ionosphere. The energy supply presumably is at least partly due to particle events in the magnetosphere. Particles incident from the magnetotail or from the cleft may enter the ring current region or penetrate through the plasmapause. The question to which extent the energy deposited in the ionosphere comes from Joule heating and from deceleration of particles still requires further investigation.

For the process by which the deposited energy affects the loss rate there are again various possibilities. Increased temperature enhances the rate of the reaction between O^+ and N_2 with subsequent recombination. An excitation of N_2 molecules acts in the same sense. The decisive process, however, seems to be an increase of the N_2 concentration due to the energy deposit that may contribute to turbulent mixing. An increase of the N_2 concentration and a decrease of the O concentration support the loss of ions. Very good examples of parallel variations of the maximum electron concentration and the O/N_2 density ratio in negative storm phases are shown by PRÖLSS (1980). It might be emphasized that the second phase of the storm has a complex nature. Transport phenomena and large scale circulation are essential for it.

The variations of the O/N_2 density ratio are limited to middle and high latitudes. In stronger storms the low latitude limit, though, is shifted equatorward. No O/N_2 variations are observed in connection with positive storm effects. This may characterize them as the result of a displacement of ionospheric plasma.

REFERENCES

BANKS, P.M., 1975, Ionosphere-magnetosphere coupling, 2. Electric fields. Rev. Geophys. Space Phys. 13, No. 3, 874-878 and 905-908.

BANKS, P.M., 1979, Magnetosphere, ionosphere and atmosphere interactions, in "Solar System Plasma Physics", Vol. 2 (ed. by C.F. Kennell, L.J. Lanzerotti, and E.N. Parker), pp. 57-103, North-Holland Publ. Co.

BRACE, L.H., E.J. MAIER, J.H. HOFFMAN, J. WHITTEKER, and G.G. SHEPHERD, 1974, Deformation of the night side plasmasphere and ionosphere during the August 1972 geomagnetic storm. J. Geophys. Res. 79, No. 34, 5211-5218.

CHAPPELL, C.R., 1975, Ionosphere-magnetosphere coupling, 1. Thermal plasma. Rev. Geophys. Space Phys. 13, No. 3, 872-873 and 904-905.

DUMBS, A., G. EMMENEGGER, R. KIST, D. KLUMPAR, E. NESKE, J. SLAVIK, K. SPENNER, and H. WOLF, 1979, Results from the plasma experiments on AEROS. J. Geomag. Geoelectr. 31, Suppl., S 125 - S 139.

LANDSCHULZ, H.-D., 1976, Quantitatives Modell der magnetosphärischen Konvektion und Abbildung des magnetosphärischen elektrischen Feldes. Dissertation, Technische Hochschule Darmstadt.

LEMAIRE, J. 1974, The Roche-limit of ionospheric plasma and the formation of the plasmapause. Planet. Space Sci. 22, No. 5, 757-766.

NESKE, E., 1978, High latitude structures in ionspheric plasma. COSPAR: Space Res., Vol.18 (ed.by M.J. Rycroft and A.C. Stickland), pp. 237-240, Pergamon Press.

NESKE, E.,and R. KIST, 1974, Impedance probe, The AEROS-B electron density experiment. J. Geophys. 40, No. 5, 593-600.

OBAYASHI, T., and A. NISHIDA, 1968, Large-scale electric field in the magnetosphere. Space Sci. Rev. 8, No. 1, 3-31.

PRÖLSS, G.W., 1980, Magnetic storm associated perturbations of the upper atmosphere: recent results obtained by satellite-borne gas analyzers. Rev. Geophys. Space Phys. 18, No. 1, 183-202.

SPENNER, K. and A. DUMBS, 1974, The retarding potential analyzer on AEROS-B. J. Geophys. 40, 585-592.

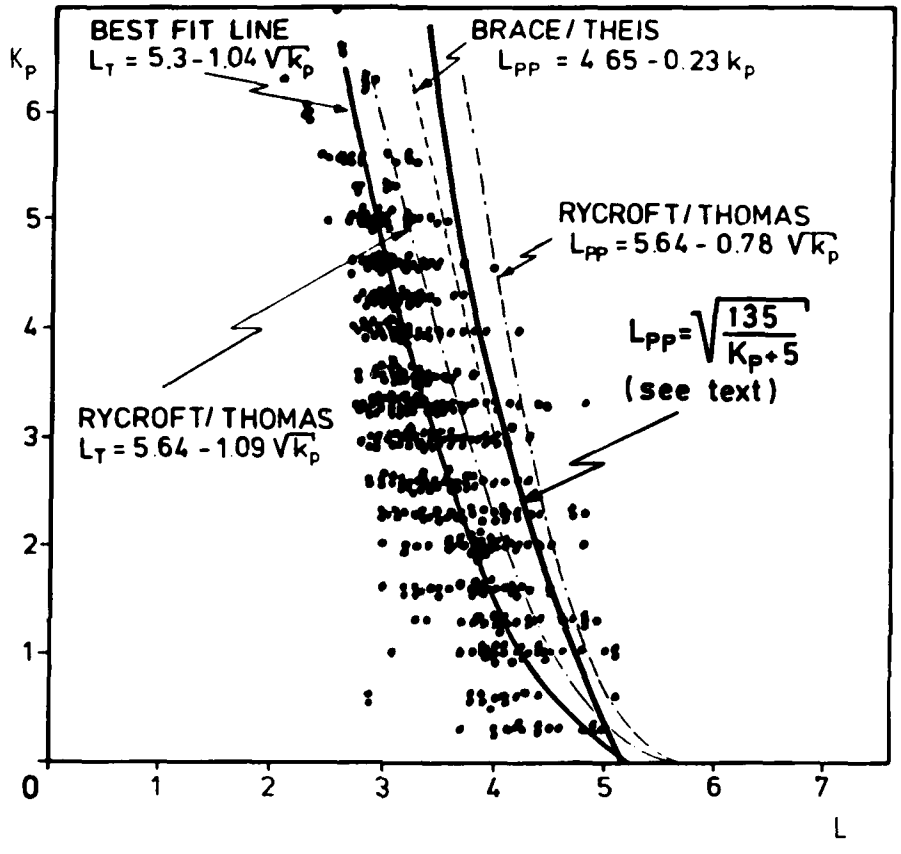
THORNE, R.M., 1975, Ionosphere-magnetosphere coupling, 3. A review of the role of wave-particle interactions. Rev. Geophys. Space Phys. 13, No. 3, 878-881 and 908-909.

VASYLIUNAS, V.M., 1968, A crude estimate of the relation between the solar wind speed and the magnetospheric electric field. J. Geophys. Res. 73, No. 7, 2529-2530.

VASYLIUNAS, V.M., 1970, Mathematical models of magnetospheric convection and its coupling to the ionosphere, in "Particles and Fields in the Magnetosphere" (ed. by B.M. McCormac), D. Reidel Publ. Co., pp. 60-71.

VASYLIUNAS, V.M., 1972, The interrelationship of magnetospheric processes, in "Earth's Magnetospheric Processes" (ed. by B.M. McCormac), D. Reidel Publ. Co., pp. 29-38.

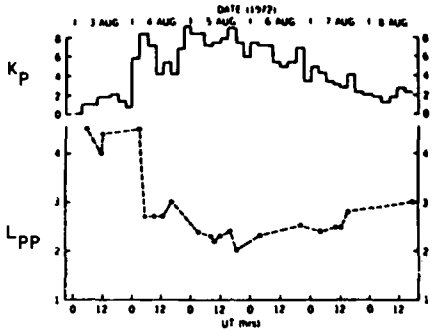
VOLLAND, H., 1975, Models of global electric fields within the magnetosphere. Ann. Géophys. 31, No. 1, 159-173.



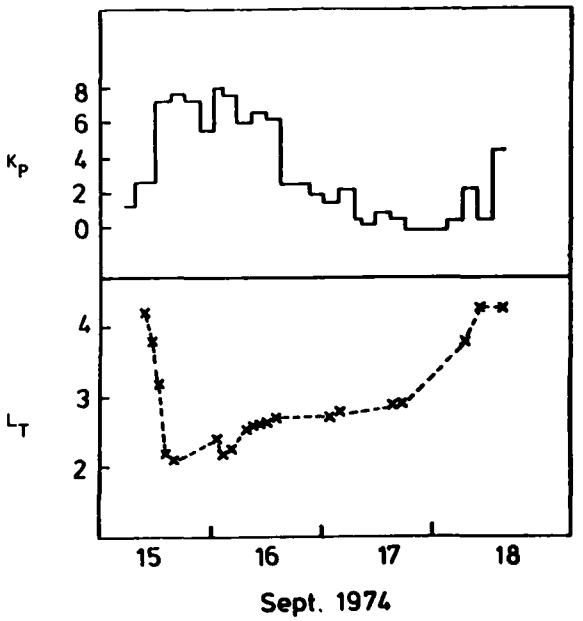
Location of the nighttime midlatitude electron density trough in dependence on K_p (Neske 1978).

The dots represent values observed in the altitude range 250 - 750 km around 04.30 local time.
The lines correspond to empirical and theoretical expressions for the trough (L_T) and the plasmapause (L_{PP}).

Fig. 1



MOVEMENTS OF THE PLASMAPAUSE L_{PP} BASED ON THE POSITION OF THE CORRESPONDING τ_E PEAKS (AFTER BRACE ET AL .1974)



MOVEMENTS OF THE NIGHTTIME MIDLATITUDE TROUGH L_T BASED ON THE OBSERVED ELECTRON DENSITY MINIMA (AFTER DUMBS ET AL.1979)

Fig. 2

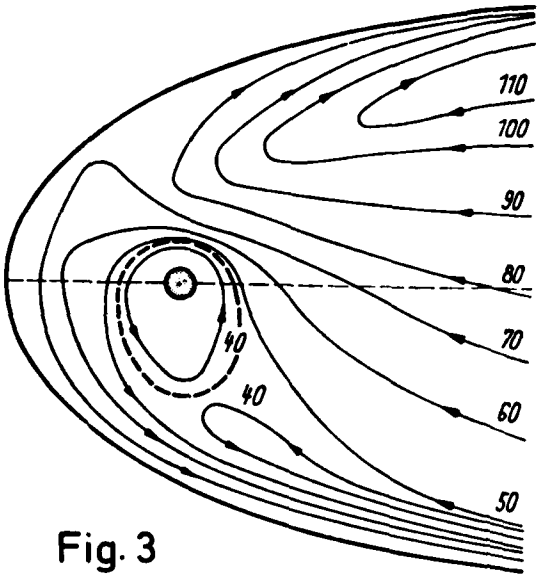
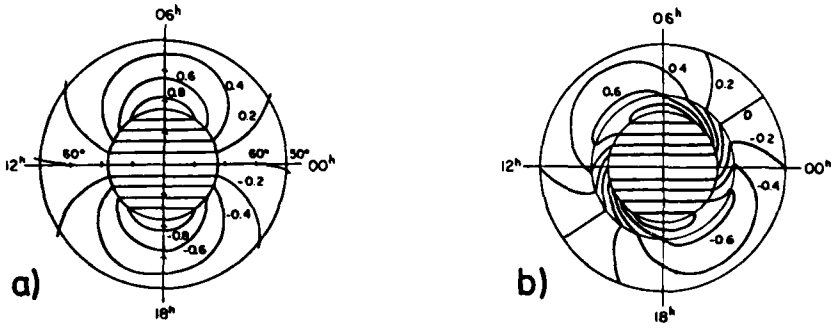


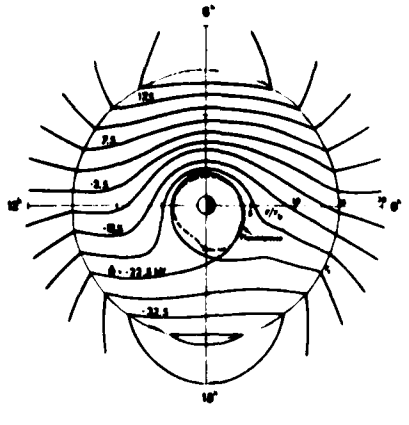
Fig. 3

Model of equipotential lines in the equatorial plane (Obayashi and Nishida 1968). Potentials are given in kV.



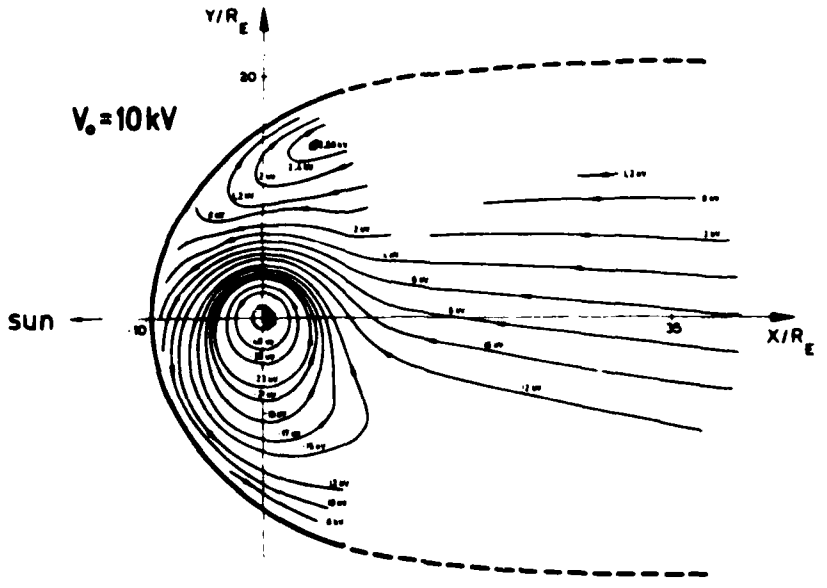
Equipotential contours at high latitudes in the ionosphere (Vasyliunas 1970)
a) without enhancement of the conductivity,
b) with enhanced conductivity from 65° to 72° latitude.
Relative values of the potential are given.

Fig. 4



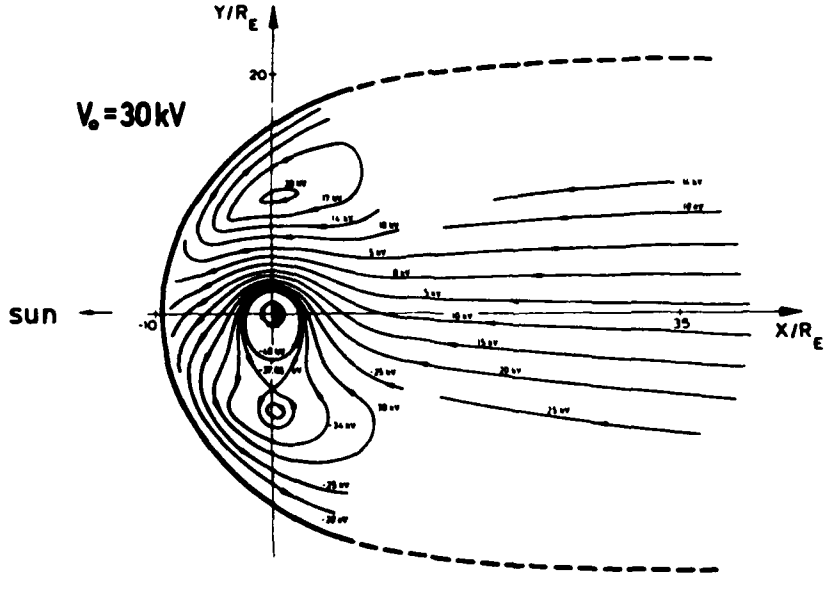
Theoretical model of equipotential lines in the equatorial plane (Volland 1975)

Fig. 5



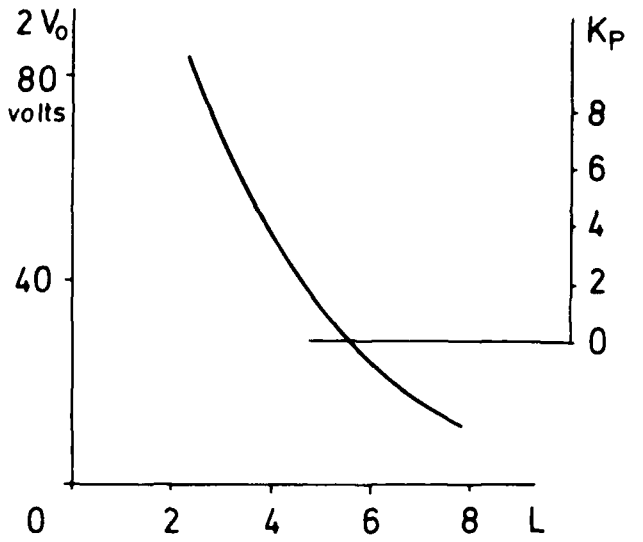
Equipotential lines in the equatorial plane (Landschulz 1976)

Fig. 6



Equipotential lines in the equatorial plane (Landschulz 1976)

Fig. 7



Polar cap voltage ($2V_0$) vs. position of plasmapause (L)

Fig. 8

MAGNETOSPHERE-IONOSPHERE COUPLING THROUGH THE AURORAL ACCELERATION REGION

R. D. Sharp and E. G. Shelley
Lockheed Palo Alto Research Laboratory
3251 Hanover Street
Palo Alto, California 94304 USA

ABSTRACT

It has recently been discovered that an important form of coupling between the magnetosphere and the ionosphere occurs through acceleration mechanisms operative in the high altitude ionosphere on magnetic field lines connecting to the auroral zone. Energetic ion mass spectrometer data from within these auroral acceleration regions are presented to illustrate the characteristics of the mechanisms. Observations of ionospheric plasmas in the ring current, the distant plasma sheet, and the magnetotail lobes are shown illustrating the extent of their circulation and the importance of their contribution to the plasma in each regime. Finally the precipitating plasmas in the auroral region and the extent and peculiar effects of the O^+ component of that precipitation on the ionosphere are illustrated.

1. INTRODUCTION

Until recently research in ionospheric physics and magnetospheric physics have proceeded relatively independently, although it has of course long been recognized that the magnetosphere provides major energy inputs to the ionosphere, particularly in the high latitude regions of auroral particle precipitation and the joule heating driven by the magnetospheric electric fields. Newly discovered coupling mechanisms between the two regimes have emphasized that they in fact form a mutually interactive system and that neither regime can be understood without an understanding of the other. The most dramatic new evidence of this coupling is found in the high altitude auroral ionosphere and will be described in this report. Other processes act in other regions. A historical review and detailed discussion of the state of our current understanding in this area is contained in the NASA report entitled "Dynamics Explorer" (Hoffman, 1978) which describes the next major thrust in research in this field. This will be a dedicated mission to specifically investigate the interactive nature of the magnetosphere-ionosphere system with two simultaneous satellites making coordinated observations.

This paper will briefly describe some of the characteristics of the recently discovered ionospheric acceleration mechanisms. We will then show how the circulation of the energized ionospheric ions throughout the magnetosphere can be traced through the use of a newly developed class of plasma diagnostic instruments known as energetic ion mass spectrometers (Balsiger et al., 1976; Shelley et al., 1978). We will then show how the energized ionospheric ions, particularly the O^+ ions, precipitate back into the ionosphere and affect it in ways that are not expected on the basis of pure proton precipitation.

2. AURORAL ACCELERATION REGION

The S3-3 satellite was launched in the summer of 1976 into a polar orbit with an apogee of [illegible] km. This was the first scientifically instrumented satellite to explore this region and it discovered a number of phenomena that affect our basic understanding of plasma physics as well as locating the source of the much-sought-after auroral electron acceleration mechanism (Shelley et al., 1976; Mozer et al., 1977; Torbert and Mozer, 1978; Mizera and Fennel, 1977).

Quasistatic electric potential drops of the order of 10 kV were found to be a commonly occurring phenomenon in this plasma regime (Croley et al., 1978; Sharp et al., 1979; Cladis and Sharp, 1979). These potentials act to violate the frozen-in field condition which has been extensively employed to analyze the dynamics of the hot magnetospheric plasma and other plasma physics processes ranging from fusion reactors to astrophysical systems (Falthammar et al., 1978). The potentials preferentially occur in such a direction that they energize precipitating auroral electrons and accelerate ambient ionospheric ions and inject them in large quantities into the magnetosphere (Shelley et al., 1976; Ghielmetti et al., 1978a, 1979). This injection represents a direct form of ionospheric- magnetospheric coupling that profoundly affects the plasma dynamics of the entire magnetosphere, so we shall examine the process here in some detail.

Figure 1 from Shelley et al. (1976) shows data from the Lockheed experiment on the S3-3 satellite on 13 July 1976 in one of the acceleration regions. The top panel shows the pitch angle of the instrument view direction. The two center panels show the response of an energetic ion mass spectrometer to O^+ and H^+ ions respectively. The energy per charge of the measured ions is indicated. One sees sharply peaked pitch angle distributions corresponding to ions streaming up the field lines from the ionosphere. The flux intensely corresponding to the peak response is 0.3×10^8 $(cm^2 \text{ sec ster keV})^{-1}$ for the 1.28 keV O^+ ions and 0.7×10^8 $(cm^2 \text{ sec ster keV})^{-1}$ for the .94 keV H^+ ions. The lowest panel shows electron fluxes in the energy range from 0.37 to 1.28 keV. One sees that the electron loss cones are substantially widened and deepened in the region of the upflowing ions relative to the adjacent regions. This is the signature of a quasistatic electric field parallel to the magnetic field along the electron's trajectory between the satellite and the atmosphere (Sharp et al., 1979). Such signatures are a common feature in the S3-3 satellite data. They are generally found in association with broad regions of upflowing ions, low frequency wave turbulence and field aligned currents (Mozer et al., 1980). A mass-per-unit-charge spectrum measured at 1.28 keV/e for the upflowing ions at about 0715:47 UT is shown in Figure 2. One sees that the entire flow consists of O^+ and H^+ ions with no significant He^+ or He^{++} present. This is typical of most of the upflowing ion events although occasionally a measureable He^+ component is encountered.

An example of a second class of ionospheric acceleration mechanism not specifically associated with quasistatic parallel electric fields is shown in Figure 3 (Sharp et al., 1977a). The upflowing ions in this event result from a process acting primarily transversely to the magnetic field. The pancake-like distributions resulting from such a transverse acceleration fold inwards as the particles move upward adiabatically in the diverging geomagnetic field resulting in the observed maxima at ~130° pitch angle. This is often referred to as a "conical" pitch angle distribution. This class of phenomena has also been observed at lower altitudes from the ISIS satellites (Klumpar, 1979) and from a rocket experiment (Whalen et al., 1978) in association with intense electron precipitation events.

For ease in discussion we will refer to the upflowing ion events without a detectable minimum along B in their pitch angle distribution as "beams" to distinguish them from the "conics". However, as discussed by Ghielmetti et al. (1978a) and Gorney et al. (1980) the ability to detect such a minimum depends strongly on the specific experimental parameters. Both types of distributions are consistent with combined parallel and perpendicular acceleration processes, although they exhibit considerable differences in their morphologies (Gorney et al., 1980).

During quiet times (Kp $\lesssim$ 3) the conics are observed with constant frequency in altitude above ~1000 km consistent with their being energized at approximately that altitude. Their local time distribution maximizes on the dayside and may be associated with the low altitude extension of the polar cusp (Gorney et al., 1980). Beams on the other hand are observed primarily above 5000 km in the pre-midnight sector (Ghielmetti et al., 1978a; Gorney et al., 1980). In more active times (Kp $\gtrsim$ 3) conics also appear to be generated in this high altitude evening sector (Gorney et al., 1980).

The ion beams are often found in association with broad inverted-V shaped potential structures such as the one illustrated in Figure 4. In this event the magnitude of the quasistatic parallel electric field both above and below the satellite has been determined from an analysis of the electron distributions (Cladis and Sharp, 1979). The circles and triangles denote the potential difference below the satellite determined independently from the loss cone widths and depths respectively. The squares denote the total potential difference which includes the potential difference above the satellite.

It is of interest to compare the magnitude of the potential difference below the satellite with the energy spectrum of the upflowing ions in order to determine the importance of the electrostatic contribution to their energization. In this example, the upflowing ions were almost entirely protons. The ion energy spectra measured on satellite spins 2, 3, and 4 (which were identified on Figure 4) are shown in Figure 5. Data from both the Lockheed ion mass spectrometer and the Aerospace Corporation electrostatic anlayzer on S3-3 are shown. The mass spectrometer provides only a three point energy spectrum on each spin (~18 seconds). The electrostatic analyzer provides more spectral points and, since in this example the flux is dominantly of one species, there is no ambiguity introduced by the lack of mass discrimination. The Aerospace analyzer is mounted on the spinning spacecraft such that it samples the upflowing ions about 5 seconds before the Lockheed ion measurements and the times at which the potentials were derived from the electron distributions. The more complete spectrum determined by the electrostatic analyzer exhibit peak fluxes at energies roughly comparable to the potential below the satellite determined from Figure 4 at the appropriate times, i.e. at points on the abscissa about one-fourth of a spin period prior to the numbered spins. The fewer Lockheed points are consistent with the peaks being at somewhat higher energies on spins 2 and 3 and at lower energy on spin 4, as would be expected from Figure 4. The data suggest that electrostatic acceleration was the dominant mechanism in this instance. However, it should be noted that the energy spectrums are broader than would be expected for a pure electrostatic acceleration and imply that some additional energy transfer process is taking place.

Collin et al. (1980) have made intercomprisons of the O^+ and H^+ constituents of the upflowing ion events. This study was conducted in the evening local time sector, during primarily quiet times, for events with ion energies greater than 500 eV. Under these conditions Gorney et al. (1980) report that almost all the events are beams rather than conics. For the Collin et al. study a few wide conics (pitch angle distributions $\gtrsim$ 50° FWHM) were deleted from the sample base in order to focus on a single phenomenon.

A comparison of the average energies of the upflowing O^+ and H^+ ions observed during the same events is shown in Figure 6. The oxygen ions are seen to be systematically more energetic than the protons. The ratio of average energies for the data set shown was 1.7. An examination of the pitch angle distributions of the two species shows that the O^+ ions also had significantly wider beam widths than the H^+. A histogram of the occurrence probability versus the half width (at half maximum) of the beams is shown in Figure 7.

These data (Figures 6 and 7) allow us to make an estimate of the relative importance of the transverse and parallel contributions to the energization of the observed ions. From the results of Ghielmetti et al. (1978a) and Gorney et al. (1980) we know that essentially all the energy in the upflowing beams is acquired above ~5000 km. If the ions derived as much as 50% of their energy from a purely transverse acceleration, then their pitch angles in the altitude range of this study (6000 to 8000 km) would be greater than 39° to 30° respectively. From Figure 7, therefore, we can infer that the bulk of the proton energy was derived electrostatically in agreement with the inference from Figure 5. Since the electrostatic process is mass independent, the added O^+ energy shown in Figure 6 could have resulted from their having entered the parallel field region at a lower altitude or from some additional mass dependent mechanism. In the former case the O^+ ions would be expected to have narrower cone widths and to be accelerated to detectable energies ($\geq$ 500 eV) at a lower altitude than the H^+. As we see in Figure 7 the O^+ beams are in fact wider than the H^+ beams, and Ghielmetti (1978b) has reported that the relative frequency of occurrence of O^+ and H^+ was not significantly dependent on altitude. We can infer therefore that the additional energy of the O^+ beams relative to the H^+ beams derives from some mass dependent transverse acceleration mechanism. Candidate mechanisms have been proposed by Lysak et al. (1980) involving resonant wave-particle interactions and by Lennartsson (1980) involving fluctuating or small scale transverse electric fields.

To summarize this section we conclude that the magnetosphere-ionosphere coupling in the auroral acceleration region proceeds through at least two mechanisms; one energizing the ions transverse to the magnetic field and one parallel. The parallel mechanism can be attributed to electrostatic acceleration. For the ion beams we conclude that the parallel mechanism is primarily responsible for the proton energization and the transverse mechanism preferentially acts on the O^+ ions and provides, on the average, about half of their energy.

3. ION CIRCULATION

The circulation of the energized ionospheric ions resulting from the above described acceleration mechanisms is best studied with the data from the International Sun-Earth Explorer (ISEE). This satellite has a near equatorial orbit with a high apogee (23 R_E) and penetrates all of the diverse magnetospheric plasma regimes. Its energetic ion mass spectrometer (Shelley et al., 1978) allows us to utilize the O^+ component of the upflowing ion streams as a tracer of the ionospheric ions since this species does not originate to a significant extent from any other source.

An illustration of the extent of the energetic O^+ contribution to the hot plasmas of the inner magnetosphere during active times is shown in Figure 8. This is one result of a synoptic study including all of the principal magnetic storms that occurred during approximately the first year of the ISEE experiment (Lennartsson et al., 1980). The measurements covered the energy per charge range from 0.1 to 17 keV/e and pitch angles in the interval $90^o \pm 45^o$. Data near the peak of DST or in the early recovery phase of 10 magnetic storms were utilized covering the full range of local time. Each sample represents the average of the data over a radial distance of ~1 R_E.

The histograms represent the fraction of the measured number density in the form of O^+. The remaining density was primarily H^+. He^+ and He^{++} were typically less than a few percent. The effects of the background on the density have been treated in two extreme manners represented by the solid and dashed histograms. The most probable frequency of occurrence is expected to lie in between the two. We see that there is a wide range of O^+ compositions extending to well over 50% of the total density. When we consider that some fraction of the H^+ also represents plasma of ionospheric origin we see that the coupling to the ionospere is a major determinant of the plasma dynamics of the inner magnetosphere during magnetic storms.

Moving from the inner magnetosphere to the distant magnetotail we show as an example in Figure 9 the ISEE data during a 90 minute interval on 2 March 1978 (Peterson et al., 1980). The satellite was in the plasma sheet at a geocentric radial distance of 19 R_E at 03 hours local time. Kp was 3^+. The plasma sheet is generally characterized by a high temperature plasma (kT typically of the order of a few keV) undergoing strong flow (bulk velocities typically of the order of 10^2 km/sec). Number densities are typically in the range of a few tenths (cm^{-3}) (Akasofu et al., 1973). The data in Figure 9 have been averaged over four angular ranges centered on the directions indicated in the insert. (The spacecraft spins with the instrument's view direction approximately in the equatorial plane). "Earth" represents earthward flowing plasma, etc. The He^{++} data and the O^+ data above 2 keV/e (represented by dots) did not have sufficient statistical precision to subdivide into quadrants and have been averaged over 360^o.

One sees in Figure 9 that H^+ is the dominant constituent in terms of number flux; the O^+ however forms 46% of the number density in the energy range below 17 keV. He^{++} is a useful tracer of plasma of solar wind origin as O^+ is for plasma of ionospheric origin (Balsiger et al., 1980). The presence of both the He^{++} and O^+ suggests that some fraction of the H^+ in this example is of solar, and some of ionospheric origin.

The large O^+ density indicates that independent of the origin of the H^+, the distant plasma sheet is at times significantly populated by ions of ionospheric origin. No statistical studies of the plasma sheet composition have yet been completed so we do not have a quantitative measure of how often the O^+ contribution is as high as it was during this period. Preliminary indications are that it is not a frequent occurrence. The H^+ component typically dominates both the flux and density in this energy range in the distant plasma sheet.

This generalization is not true however for the magnetotail lobes. In this regime the plasma is typically in the form of streams with low temperatures (kT of the order of 10 eV), and low densities (of the order of 10^{-3} cm^{-3}) (Sharp et al., 1980). This is a dramatically different population than that of the plasma sheet.

Figure 10 shows the result of a statistical study of approximately 70 hours of data when ISEE was in the central magnetotail lobes during the Spring of 1978. Each "case" in the histogram represents the observation of an ion stream of the indicated species with properties similar to those noted above. The measurements pertain only to the energy range above 100 eV. The peak flux intensity in the bulk flow direction (approximately tailward) is indicated on the abscissa. The frequency of occurrence of such streams during the period of the study was 10.7 ± 1.4% for the H^+ ions and 16.6 ± 1.7% for the O^+ ions. Thus the O^+ streams were more frequently occurring than the H^+ streams and, as we see in Figure 10, they had on the average a somewhat higher flux intensity. A detailed analysis of the properties of these central tail lobe streams (Sharp et al., 1980) suggests that they are a different population than the magnetotail boundary layer plasmas which are primarily of solar origin, and that both the O^+ and H^+ constituents are probably ionospheric, originating from acceleration mechanisms similar to those discussed above, which operate to produce the polar cap auroral arcs (Burch et al., 1979).

4. PRECIPITATION

The ISEE results have demonstrated the ubiquitous nature of the ionospheric plasma throughout the magnetosphere. As a consequence of the large O^+ content in the trapped plasma population, particularly during magnetic storms, there is also a large O^+ content to the precipitating plasma and this leads to some interesting consequences for the high latitude ionosphere where this energy source is significant. Figure 11 shows the result of a study of the precipitating ion fluxes in the energy-per-charge range from

0.7 to 12 keV/e during the magnetic storms of 16-17 December 1971 at ~0300 hours local time (Johnson et al., 1975). We see that the O^+ ions provide the dominant portion of the precipitating energy flux in this energy range over a substantial latitudinal interval. In contrast to precipitating energetic protons the O^+ ions lose their energy primarily through momentum transfer collisions with the ambient oxygen atoms (Torr et al., 1974). They deposit this energy at F region altitudes in the form of heating. The measured O^+ precipitation rates during major storms are such that the energy deposited by them in the F region is comparable to the solar energy input (Torr and Torr, 1979). Because this energy is deposited at a higher altitude than would be expected for precipitating protons, some "anomalous" storm time phenomena in the F region may now be understood.

Figure 12 taken from Maier et al. (1975) shows plasma and airglow parameters as measured from ISIS-2 on 18 December 1971 during the same magnetic storm as discussed above. The measurements were made in the region of a SAR arc which is evident as the 700 R peak in the 6300 Å intensity at L = 3.1. Note the high ion temperatures (over 6000°K) associated with the arc. These are "anomalous" in the sense that they are not expected from the thermal electron conductance model of the SAR arc without some independent mechanism for heating the ions (Rees and Roble, 1975).

Observations over this same SAR arc were made with an energetic ion mass spectrometer on the 1971-89A satellite approximately 3 hours later in local time and 4 hours later in universal time (Sharp et al., 1977b). The precipitating particle fluxes were very low, as expected from spectral purity considerations. That is, the lack of 5577 Å emission in SAR arcs typically establishes an upper limit on the energetic electron and proton precipitation. Precipitating ion fluxes did exist however as predicted by the currently accepted model for the formation of the arc (Cornwall et al., 1971) which involves pitch angle diffusion of equatorially trapped energetic ions into the loss cone. The 1971-89A measurements indicated that O^+ was the dominant constituent of this precipitation. The peak response was observed in the 8.6 keV energy channel and the estimated precipitating O+ energy flux was about 10^{-2} ergs/cm^2 sec. The detailed modeling necessary to establish whether this flux is quantitatively consistent with the observed enhanced ion temperatures has not been carried out. Rees and Roble (1975) have however modeled a SAR arc assuming an ad hoc topside ion heat flow into the region of the arc and find that ion temperatures comparable to those found here can be obtained with an energy input of about an order of magnitude less than was provided by the O^+ in this case. We therefore suggest that the ion temperature peak in Figure 12 represents the first direct observation of the effects of precipitating O^+ ions on the high altitude ionosphere.

5. SUMMARY AND CONCLUSIONS

We have shown that a direct and important coupling between the ionosphere and the magnetosphre occurs through at least two acceleration mechanisms operative on magnetic field lines connecting to the auroral zone. One mechanism electrostatically accelerates and precipitates both thermal and plasma sheet electrons, causing auroral arcs and inverted-V precipitation regions. It simultaneously accelerates thermal ions from the ionosphere and injects them upward into the magnetosphere.

Another class of phenomena primarily acts to energize the transverse component of the ion energy and also acts preferentially on the O^+ constituent of the ionospheric plasma. The subsequent adiabatic motion of the transversely accelerated ions along the diverging magnetic field lines transports them upward into the magnetosphere. In some, perhaps all, of the ionospheric acceleration regions both mechanisms appear to be acting simultaneously.

The principal ion species accelerated from the ionosphere are H^+ and O^+ and the O^+ ions form a convenient tracer of these processes since they do not originate to any significant extent from the other principal source for the magnetosphere plasma; the solar wind. The energized O^+ ions are found throughout the magnetosphere including the distant magnetotail where they form the principal component of the central tail lobe plasma streams and a lesser but still significant component of the distant plasma sheet. These plasmas convect inward and mix with the directly injected ionospheric plasma forming the ring current, which is often dominated by plasmas of ionospheric origin in the energy range observable with the present generation of instruments ($E/Q \leq 32$ keV/e).

The precipitation of the ring current oxygen ions causes an anomalously high rate of heating in the upper ionosphere since they stop at higher altitudes than do protons of comparable energies, and deposit their energy in a region of low heat capacity. Thus the O^+ circulation pattern both begins and ends in newly recognized ionospheric disturbance phenomena, and in its course involves all of the diverse magnetospheric plasma regimes.

6. ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research, the Division of Atmospheric Sciences of the National Science Foundation under Grant ATM-7911174, and NASA under contracts NAS 5-25773 and NASW-3395.

7. REFERENCES

Balsiger, H., P. Eberhardt, J. Geiss, A. Ghielmetti, H. P. Walker, D.T. Young, H. Loidl, and H. Rosenbauer, 1976, "A Satellite-Borne Ion Mass Spectrometer for the Energy Range 0 to 16 keV," Space Science Instrum., 2, 499.

Balsiger, H., P. Eberhardt, J. Geiss, and D. T. Young, 1980, "Magnetic Storm Injection of 0.9 - 16 keV/e Solar and Terretrial Ions into the High Altitude Magnetosphere," J. Geophys. Res., 85, 1645.

Birch, J. L., S. A. Fields, and R. A. Heelis, 1979, "Polar Cap Electron Acceleration Regions," J. Geophys. Res., 84, 5863.

Cladis, J. B., and R. D.Sharp, 1979, "Scale of Electric Field Along Magnetic Field in an Inverted- V Event," J. Geophys. Res., 84, 6564.

Collin, H. L., R. D. Sharp, E. G. Shelley, and R. G. Johnson, 1980, "Some General Characteristics of Upflowing Ion Beams over the Auroral Zone and Their Relationship to Auroral Electrons," Lockheed Palo Alto Research Laboratory, preprint.

Cornwall, J. M., F. V. Coroniti, and R. M. Thorne, 1971, "Unified Theory of SAR Arc Formation at the Plasmapause," J. Geophys. Res., 76, 4428.

Croley, Jr., D. R., P. F. Mizera, and J. F. Fennell, 1978, "Signature of a Parallel Electric Field in Ion and Electron Distributions in Velocity Space," J. Geophys. Res., 83, 2701.

Falthammar, C. G., S. I. Akasofu,and H. Alfven, 1978, "The Significance of Magnetospheric Research for Progress in Astrophysics," Nature, 275, 5677.

Ghielmetti, A. G., R. G. Johnson, R. D. Sharp, and E. G.Shelley, 1978a, "The Latitudinal, Diurnal, and Altitudinal Distributions of Upflowing Energetic Ions of Ionospheic Origin," Geophys. Res. Lett., 5, 59.

Ghielmetti, A. G., 1978b, "Upward Flowing Ion Characteristics in the High Latitude Ionospheric Acceleration Region," EOS Trans. Am. Geophys. U., 59, 1155.

Ghielmetti, A. G., R. D. Sharp, E. G. Shelley, and R. G. Johnson, 1979, "Downward Flowing Ions and Evidence for Injection of Ionospheric Ions into the Plasma Sheet," J. Geophys. Res., 84, 5781.

Gorney, D. J., A. Clarke, D. R. Croley, J. F. Fennell, J. M. Luhmann, and P. F. Mizera, 1980, "The Distribution of Ion Beams and Conics Below 8000 km," Aerospace Corporation Report ATR-80 (7774)-1.

Hoffman, R. A., 1978, "The Dynamics Explorer Mission - report from the Dynamics Explorer Science Team," NASA-GSFC Report.

Klumpar, D. M., 1979, "Transversely Accelerated Ions: An Ionospheric Source of Hot Magnetospheric Ions," J. Geophys. Res., 84, 4229.

Lennartsson, W., R. D. Sharp, E. G. Shelley, R. G. Johnson, and H. Balsiger, 1980, "Ion Composition and Energy Distribution During 10 Magnetic Storms," Lockheed Palo Alto Research Laboratory, preprint.

Lennartsson, W., 1980, "On the Consequences of the Interaction Between the Auroral Plasma and the Geomagnetic Field," Planet. Space Sci., 58, 135.

Lysak, R. L., M. K. Hudson, and M. Temerin, 1980, "Ion Heating by Strong Electrostatic Ion Cyclotron Turbulence," J. Geophys. Res., 85, 678.

Maier, E. J., S. Chandra, L. Brace, J. H. Hoffman, G. G. Shepherd, and J. H. Whitteker, 1975, "The SAR Arc Event Observed During December 1971 Magnetic Storm," J. Geophys Res., 80, 4591.

Mozer, F. S., C. W. Carlson, M. K. Hudson, R. B. Torbert, B. Parady, J. Yatteau, and M. C. Kelley, 1977, "Observations of Paired Electrostatic Shocks in the Polar Magnetosphere," Phys. Rev. Lett., 38, 292.

Mozer, F. S., C. A. Cattell, M. K. Hudson, R. L. Lysak, M. Temerin, and R. B. Torbert, 1980, "Satellite Measurements and Theories of Low Altitude Auroral Particle Acceleration," University of California Berkeley, preprint.

Peterson, W. K., R. D. Sharp, E. G. Shelley, R. G. Johnson, and H. Balsiger, 1980, "Energetic Ion Composition of the Plasma Sheet," J. Geophys. Res., (in press).

Rees, M. H., and R. G. Roble, 1975, "Observations and Theory of Formation of the Stable Auroral Red Arcs," Rev. Geophys. Space Phys., 13, 201.

Sharp. R. D., R. G. Johnson, and E. G. Shelley, 1977a, "Observation of an Ionospheric Acceleration Mechanism Producing Energetic (keV) Ions Primarily Normal to the Geomagnetic Field Direction," J. Geophys. Res., 82, 3324.

Sharp, R. D., E. G. Shelley, and R. G. Johnson, 1977b, "A Search for Helium Ions in the Recovery Phase of a Magnetic Storm," J. Geophys. Res., 82, 2361.

Sharp, R. D., E. G. Shelley, R. G. Johnson, W. K. Peterson, A. Ghielmetti, and H. Rosenbauer, 1980, "Composition of Ion Streams on the Magnetotail," Lockheed Palo Alto Research Laboratory, preprint.

Shelley, E.G., R. D. Sharp, and R. G. Johnson, 1976, "Satellite Observations of an Ionospheric Acceleration Mechanism," Geophys. Res., Lett., 3, 654.

Shelley, E. G., R. D. Sharp, R. G. Johnson, J. Geiss, P. Eberhardt, H. Balsiger, G. Haerendel, and H. Rosenbauer, 1978, "Plasma Composition Experiment on ISEE-A," ISEE Trans. on Geoscience Electronics, Vol. GE-16, 266.

Torbert, R. B., and F. S. Mozer, 1978, "Electrostatic Shocks as the Source of Discrete Auroral Arcs," Geophys. Res. Lett., 5, 135.

Torr, M. R., J. C. G. Walker, and D. G. Torr, 1974, "Escape of Fast Oxygen Atoms from the Atmosphere During Geomagnetic Storms," J. Geophys. Res., 79, 5267.

Torr, M. R., and D. G. Torr, 1980, "Energetic Oxygen: A Direct Coupling Mechanism Between the Magnetosphere and the Thermosphere," 1979, Geophys. Res. Lett., 6, 700.

Whalen, B. A., W. Bernstein, and P. W. Daly, 1978, "Low Altitude Acceleration of Ionospheric Ions," Geophys. Res. Lett., 5, 55.

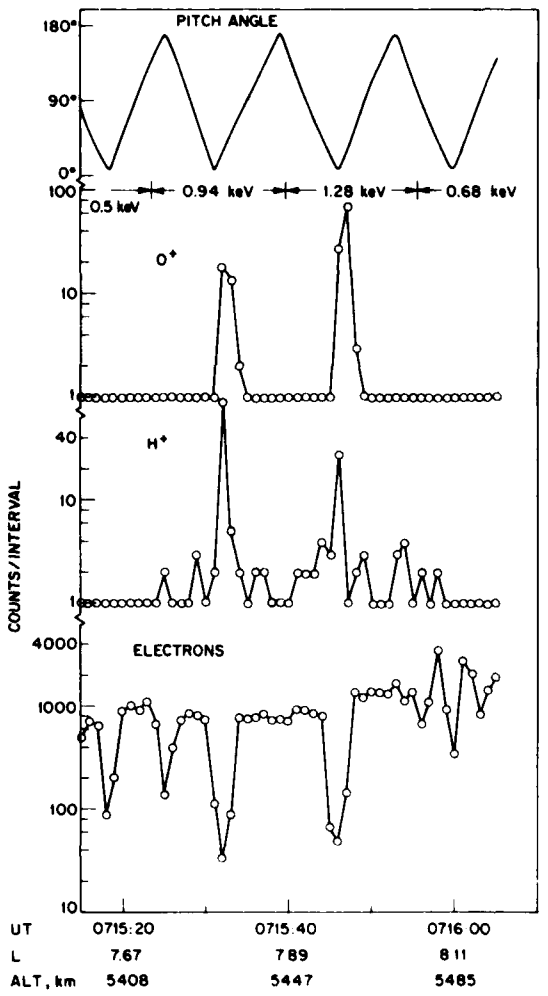


Figure 1. Upflowing ions beams observed from the S3-3 satellite in association with signatures of a parallel electric field in the electron data (Shelley et al., 1977).

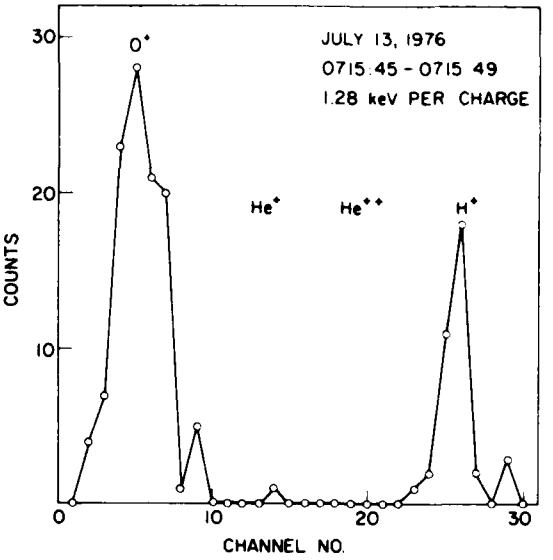


Figure 2. Ion mass-per-unit-charge spectrum of one of the beams illustrated in Figure 1 (Shelley et al., 1976).

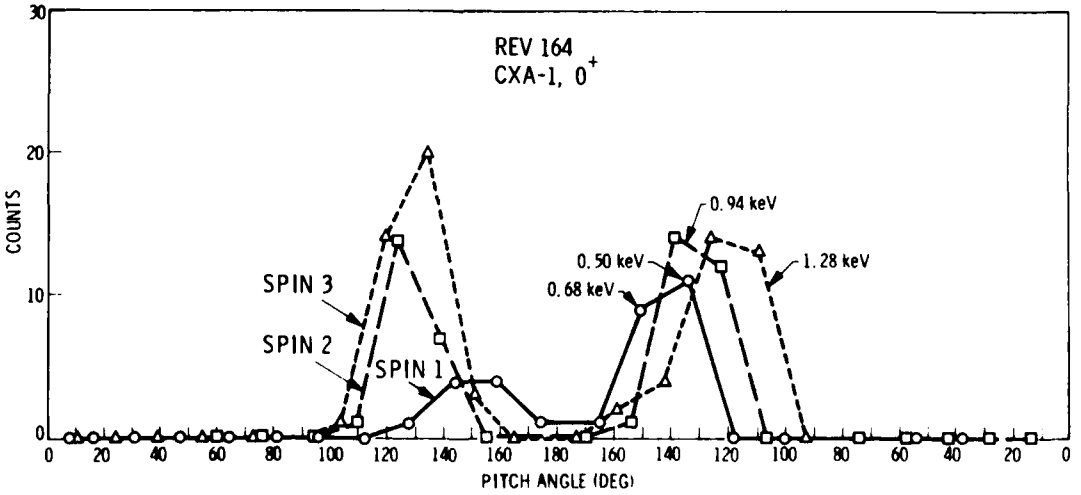


Figure 3. Oxygen conic observed at an altitude of 7600 km in the dayside cusp (Sharp et al., 1977a).

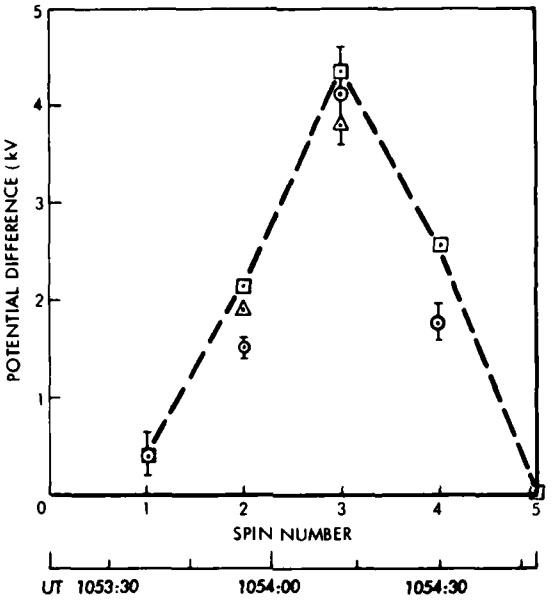


Figure 4. Potential difference along the magnetic field through the structure of an inverted-V event (Cladis and Sharp, 1979).

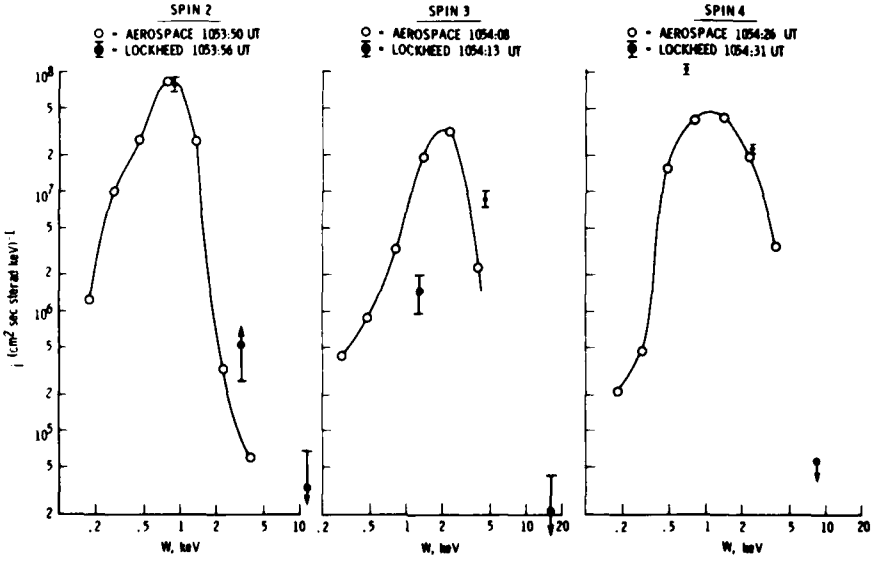


Figure 5. Energy spectra of upflowing ions within the inverted-V event shown in Figure 4 (Cladis and Sharp, 1979).

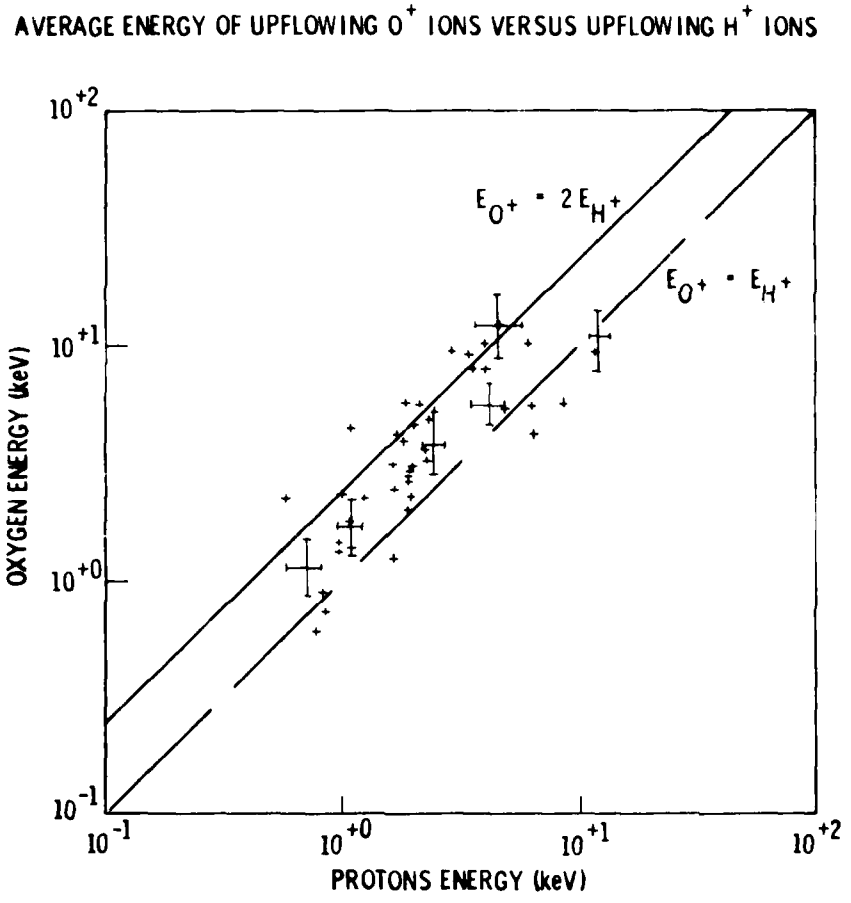


Figure 6. Scatter plot of average O^+ and H^+ energies in upflowing ion beams (Collin et al, 1980).

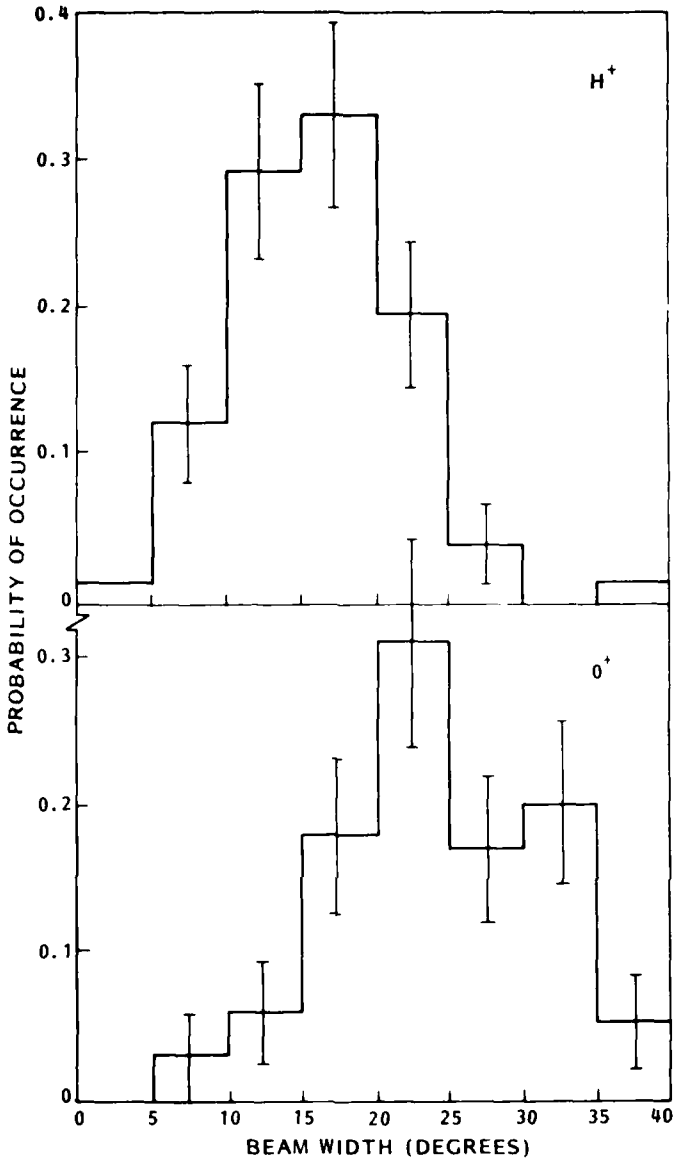


Figure 7. A comparison of the O+ and H^+ beams widths from the upflowing ion events shown in Figure 6 (Collin et al., 1980).

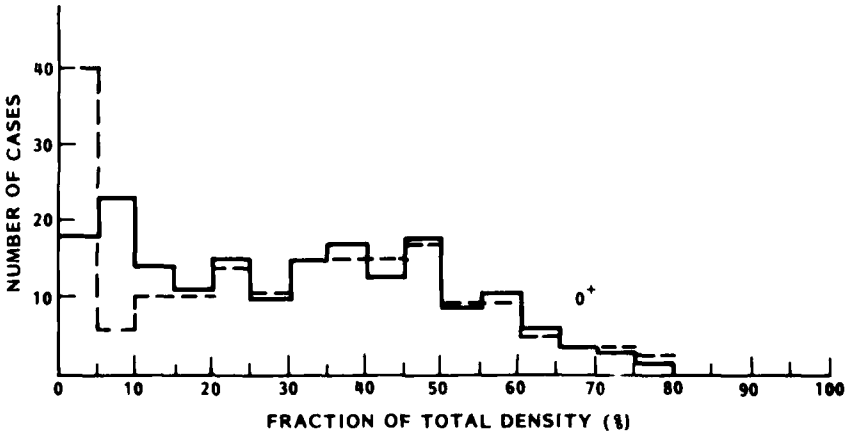


Figure 8. Results of a statistical study of the composition of the storm time ring current during 10 magnetic storms. The fractional number density in the energy per charge range $1.0 \leq E/q \leq 17$ keV in the form of O^+ is indicated. Almost all of the remaining density was in the form of H^+ (Lennartsson et al., 1980).

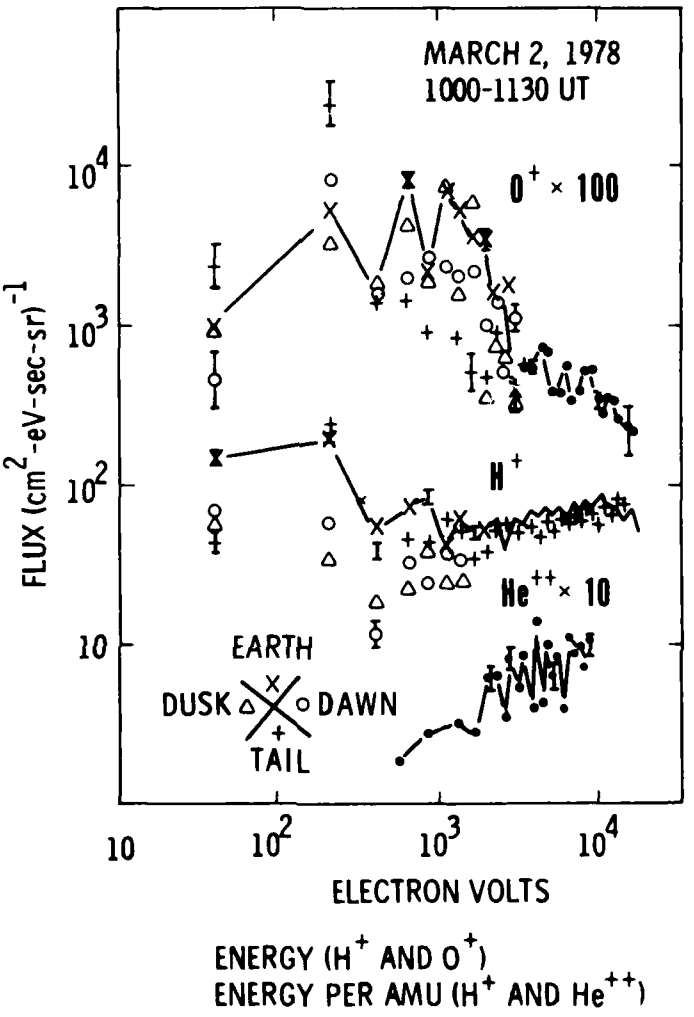


Figure 9. ISEE composition data from the plasma sheet at 19 R_E geocentric radial distance. The data have been averaged over 90° angular sectors as indicated. The O^+ data are plotted versus energy per nucleon (Peterson et al., 1980).

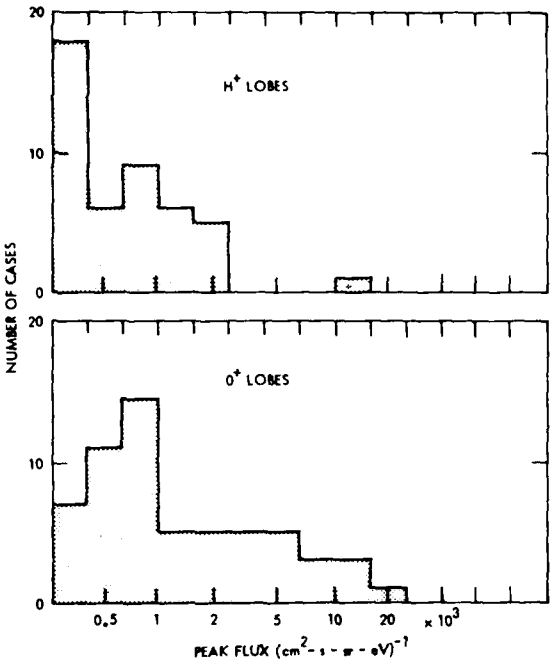


Figure 10. Results of a statistical study of the composition of ion streams observed by ISEE in the lobes of the magnetotail (Sharp et al., 1980).

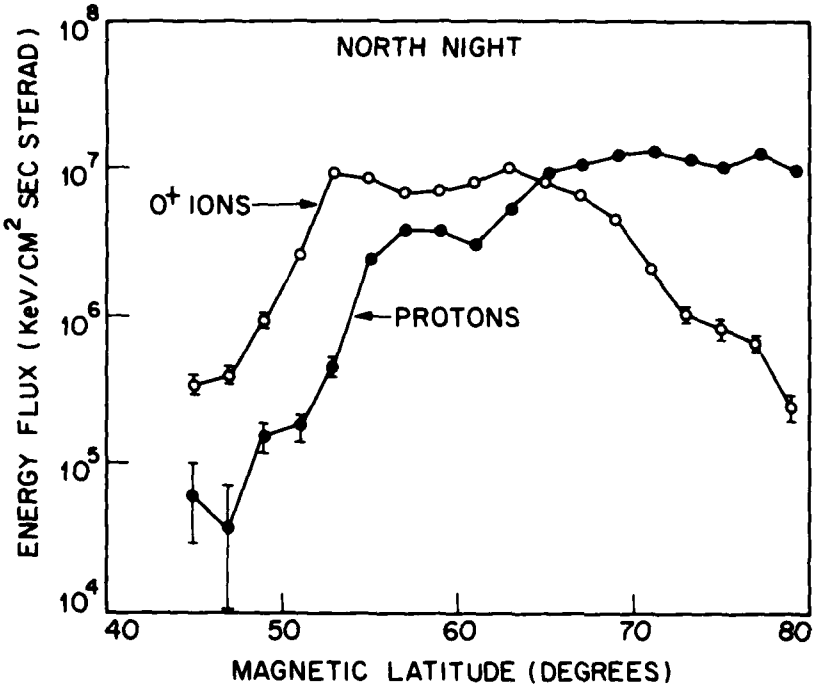


Figure 11. Latitudinal variation of the average precipitating energy flux of O^+ and H^+ ions during the time period from 0500 UT on December 17 to 1200 UT on December 18, 1971 (Johnson et al., 1975).

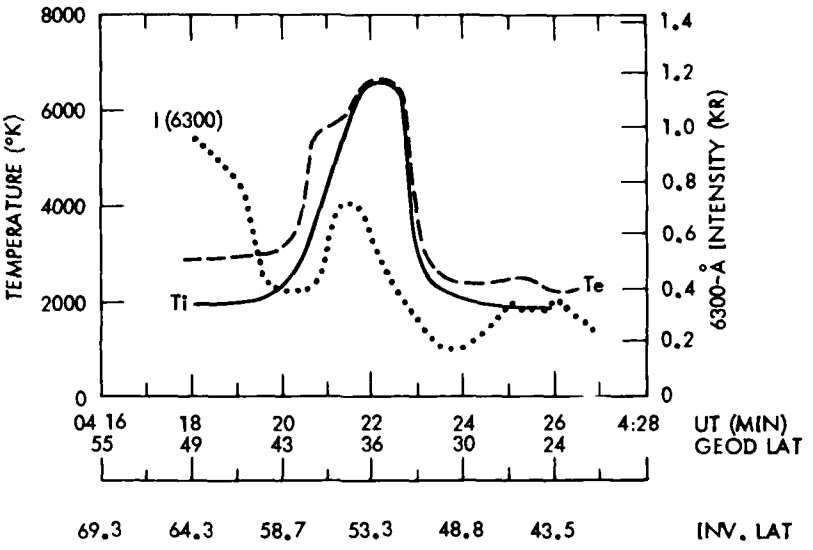


Figure 12. Plasma and airglow parameters as measured from ISIS-2 at 1400 km altitude on December 18, 1971. The I(6300) data are projected to the appropriate L value at 350 km altitude (Maier et al., 1975).

MODELLING THE HIGH-LATITUDE IONOSPHERE

W.J. Raitt, R.W. Schunk and J.J. Sojka

Center for Atmospheric and Space Sciences
Utah State University
Logan, Utah 84321
U.S.A.

ABSTRACT

Results of an ionospheric model program are presented which demonstrate the extreme variability of the steady state, daytime, ionospheric F-region electron density and ion composition due to both neutral atmospheric changes with solar cycle, season and magnetic activity, and to the effects of ionospheric drifts caused by perpendicular electric fields. Consideration is given to the time history of the ionospheric plasma as it undergoes convective motion due to the combined effects of co-rotation forces and electromagnetic forces, which results from the mapping of the magnetospheric cross-tail electric field to the rotating ionosphere. A simple model of the convection pattern is described. The model calculates the net effect of the tendency for the plasma to co-rotate about the geographic pole and the $\underline{E} \times \underline{B}$ velocity induced by a perpendicular electric field mapped to a circle centered about a point 5^{o} anti-sunward of the geomagnetic pole and oriented such that the equipotentials are parallel to the noon-midnight meridian. This convection pattern shows the generally accepted features of high latitude convection, but because of the offset between the geographic and geomagnetic poles a marked universal time dependence in these features is predicted. The results of a comparison of the convection model with ground-based incoherent scatter radar measurements from Chatanika and Millstone Hill are shown to be in good general agreement. In addition, differences in the convection pattern between the northern and southern polar ionospheres are shown to be important, particularly in the case of asymmetric magnetospheric electric fields. Finally, maps are shown of the modelled ionospheric composition and density for a convecting polar ionosphere demonstrating the formation of commonly observed features such as the mid-latitude trough and the polar hole, and their dependence on universal time.

1. INTRODUCTION

The characteristics of the ionized region extending outwards from the Earth starting at an altitude of about 60 km differ considerably at high latitudes compared to low and mid latitudes. All three regions are generally called the ionosphere, but some workers regard the high latitude ionosphere as the earthward boundary of the magnetosphere. The difference between the low, mid, and high latitude regions is related to the dynamical processes acting within these regions, which are determined by the plasma, electric field, and magnetic field environments that surround these regions. These environments comprise the Earth's magnetosphere, which is shown schematically in Figure 1. The ionosphere at low and mid latitudes is contained in the region labeled the plasmasphere, a torus-shaped volume that surrounds the Earth and contains a relatively cool, high density plasma. The dominant processes affecting the plasma in the plasmasphere are photochemical reactions and ambipolar diffusion. The main source of ionization and heating in this region is related to solar EUV radiation. The plasma in the plasmasphere corotates with the Earth, but it can also flow along geomagnetic field lines from one hemisphere to the other. In the equatorial plane, the plasmasphere has a radial extent of about 4 R_E, and its boundary, called the plasmapause, is marked by a large decrease in electron density as you leave the plasmasphere. The plasmapause is essentially the boundary between plasma that corotates with the Earth and plasma that doesn't.

At high latitudes photochemical processes and ambipolar diffusion operate just as they do at low and mid latitudes. However, additional processes are also important in this region, such as production and heating of ionospheric plasma due to energetic particle precipitation from the plasma sheet and polar cusps, the escape of plasma from the topside ionosphere along 'open' geomagnetic field lines (the polar wind), and the disruption of corotation due to electric fields of magnetospheric origin.

Owing to the interaction of the shocked solar wind with the geomagnetic field, an electric potential difference is generated across the tail of the magnetosphere, with the resulting electric field pointing from dawn to dusk (cf. Stern, 1977). This cross-tail potential difference is typically 76 kilovolts, but can vary from 20 to 120 kilovolts depending on the level of geomagnetic activity. Except for isolated regions and brief periods, the geomagnetic field lines are equipotentials due to the high electrical conductivity along field lines. Consequently, this cross-tail potential difference is mapped into the high latitude ionosphere as an electric field that is directed perpendicular to the geomagnetic field. At ionospheric heights, this perpendicular (or convection) electric field is typically 25-50 mVm^{-1} in the polar cap (Banks and Doupnik, 1975), but can be much greater than 100 mVm^{-1} in restricted latitudinal bands (Heelis et al, 1976; Smiddy et al, 1977; Spiro et al, 1978). Only the high latitude ionosphere is influenced by the magnetospheric electric field, since most of the time the ring current provides an effective shield for the plasma within the plasmasphere.

Over the last several years we developed a theoretical model of the convecting high-latitude ionosphere in order to determine the extent to which various chemical and transport processes affect the ion composition and electron density in both the sunlit and dark hemispheres (Banks et al, 1974; Schunk and

Banks, 1975; Schunk et al, 1975, 1976; Schunk and Raitt, 1980). From these studies, as well as those by Knudsen (1974), Knudsen et al (1977), Spiro et al (1978), Watkins (1978) and Brinton et al (1978), it is apparent that the high-latitude ionosphere can, by varying degrees, be influenced by solar EUV radiation, energetic particle precipitation, diffusion, thermospheric winds, electrodynamic drifts, polar wind escape, energy-dependent chemical reactions and magnetic storm induced neutral composition changes.

In our most recent study (Schunk and Raitt, 1980), we improved our high-latitude ionospheric-atmospheric model so that we could study the solar cycle, seasonal, and geomagnetic activity variations of the daytime high-latitude F layer. The improvements included updating the various chemical reaction rates, the adoption of the latest solar EUV fluxes measured by the Atmosphere Explorer satellites, the adoption of the most recent MSIS neutral atmosphere model (N_2, O_2, O, He), and the adoption of the Engebretson et al (1977) atomic nitrogen model. Also, our theoretical model was improved by including N^+ and He^+ in addition to the ions NO^+, O_2^+, N_2^+ and O^+.

In parallel with the improvement of our high-latitude ionospheric-atmospheric model, we developed a simple model of high-latitude plasma convection (Sojka et al, 1979a, b; 1980a, b). Our convection model includes the offset between the geographic and geomagnetic poles, the tendency of plasma to corotate about the geographic pole, and a dawn/dusk magnetospheric electric field mapped to a circular region in the ionosphere about a center which is offset in the antisunward direction from the magnetic pole. The convection model allows for asymmetries in the mapped magnetospheric electric field.

We have combined our convection and ionospheric-atmospheric models in order to study the high-latitude F-region. Our model produces time-dependent, 3-dimensional, ion density distributions for the ions NO^+, O_2^+, N_2^+, O^+, N^+ and He^+. We cover the high-latitude ionosphere above 54° N magnetic latitude and at altitudes between about 160-800 km for one complete day. The input parameters used by our model are shown schematically in Figure 2. The neutral composition and thermospheric wind are inputs to our model as are the neutral, ion, and electron temperatures; these parameters vary over the polar cap. The EUV solar radiation spectrum and ion production due to both auroral particle precipitation and resonantly scattered solar radiation are also inputs. Finally, our model requires the cross-tail magnetospheric potential, which is mapped down along dipole geomagnetic field lines. These parameters are used in the calculation of plasma convection, plasma diffusion, and photochemical processes, which in turn yield the ion density distributions.

2. PLASMA CONVECTION

Our basic convection model includes the offset between the geographic and geomagnetic poles, the tendency of plasma to corotate about the geographic pole, and a dawn/dusk magnetospheric electric field mapped to a circular region in the ionosphere about a center offset by 5° in the antisunward direction from the magnetic pole. The radius of the circle corresponds to 17° of latitude and the electric potentials are aligned parallel to the noon/midnight meridian within the circular region. Equatorward of the circle the potential diminishes radially and varies inversely as the fourth power of sine magnetic co-latitude (Volland, 1975). Extensions to our basic convection model include allowance for the mapping of both asymmetric magnetospheric electric potentials and electric potentials that are not aligned parallel to the noon/midnight meridian.

Magnetospheric Electric Potentials

Representative magnetospheric electric potentials mapped into the polar ionosphere are shown in Figure 3. Magnetic local time in hours and magnetic latitude at 10° intervals are shown in the figure. Contours of potential have been computed at 5 kilovolt intervals; five of these contours have been labelled. A total potential drop of 90 kilovolts was used for this calculation. This potential is representative of moderately active magnetospheric conditions. Model (a) contains a uniform magnetospheric electric potential distribution. By enhancing the dawn or dusk polar cap electric field it is possible to simulate the effects of different orientations of the interplanetary magnetic field on the magnetospheric convection electric field (cf. Fairfield, 1977; Heppner, 1977; Stern, 1977). Model (b) shows the effect of an enhanced dusk electric field on the magnetospheric potential distribution. The symmetry present in model (a) about the zero potential contour is no longer present; the pattern now has one small cell (dusk) and one large cell (dawn). The magnetospheric potential distribution for an enhanced dawn electric field is the mirror image of model (b) and is not shown. Model (c) contains a rotation of the whole magnetospheric electric field by two hours around the point about which the field is mapped into the ionosphere. Empirically it was found that this adjustment gave very good agreement between the model and the finer details of magnetospheric convection as observed at Chatanika and Millstone Hill. Model (c) shows the result of rotating the magnetospheric electric field of model (a) by two hours such that the zero potential contour lies approximately along the 10 to 22 hour magnetic meridian.

Plasma Convection Velocities

The plasma drift velocities in the high latitude ionosphere are determined by the combined action of the magnetospheric electric field and corotation. However, the magnetospheric electric field mapping into the ionosphere is determined by the geomagnetic axis, while corotation is relative to the geographic axis. Since these two axes are not aligned but subtend an angle of about 11.5° in the northern hemisphere, the eventual plasma drift velocity depends on the relative location of the axes.

In the geomagnetic quasi-inertial frame the combined action of plasma corotation and magnetospheric convection results in a flow pattern similar to that shown in Figure 4. In this geomagnetic frame the plasma streamlines coincide with the electric equipotentials. A total cross-tail potential drop of 64 kV has been used in the model for the magnetospheric electric field. Lines of constant magnetic latitude are represented by concentric dashed circles, while magnetic longitude is represented by tick marks at hourly MLT intervals. An important aspect of the convection pattern shown in Figure 4 concerns the time it takes a field tube of plasma to complete a full circulation. This circulation period will exhibit significant variation from trajectory to trajectory because the various trajectories have different lengths and the flow speed varies from one to the other and also from point to point along a given trajectory. To

illustrate this point, we show in Figure 4 the circulation times for eight representative plasma drift trajectories. The circulation times vary from a few hours for the smaller trajectories 3, 7, and 8 to much more than a day for trajectory 4, which passes through the band of low-speed flow. The lower-latitude trajectories 1 and 2 have a circulation period of 1 day, which indicates corotation about the geomagnetic pole in this geomagnetic quasi-inertial frame.

Universal Time Dependence

If the displacement between the geographic and geomagnetic poles is ignored, the plasma flow pattern predicted for the geographic inertial frame would be UT independent and identical to that shown in Figure 4. However, the displacement between the poles cannot be neglected, and the motion of the geomagnetic pole about the geographic pole acts to introduce a UT dependence when the plasma drift velocities are transformed from the geomagnetic to the geographic frame.

This UT dependence is clearly illustrated in Figure 5, where we present flow directions and speed distributions in the geographic inertial frame for the four universal times corresponding to the geomagnetic pole being on the (a) midnight, (b) dawn, (c) noon, and (d) dusk meridians. In these plots the arrows are unit vectors that represent flow direction. The speed distribution is shown by the gray scale display. Also, our definition of inertial frames is such that the sun is fixed at 12 hours in the geomagnetic frame but moves by approximately 1 degree per day in the geographic frame. This motion is sufficiently small that for our purposes the sun can be taken to be at 12 hours local time in both frames.

In addition to the obvious UT dependence shown in Figure 5, several other features are noteworthy. First, there is a band of low-speed plasma flow that is located symmetrically about 1800 MLT in the geomagnetic frame, which is in general not symmetrical about 1800 LT when viewed from the geographic frame. Furthermore, the region of very low speed flow (<50 m/s) moves from noon to midnight local time during the course of a day. Certain flow features are also evident in the geographic frame at certain universal times, such as shear reversals, rotational reversals, and throats. The throatlike feature appears when the magnetic pole is on the midnight meridian (Figure 5a) and is located on the sunward side of the magnetic pole. Rotational reversals of the plasma flow direction can be seen in the low-speed flow regions, and a high-speed shear reversal is evident at 80° latitude on the noon meridian in Figure 5b. Also, the anti-sunward flow over the polar cap is not always parallel to the noon-midnight meridian in the geographic inertial frame.

It should be emphasized that the movement of the low-speed region with UT and the appearance of a throatlike feature at a certain UT are a consequence of the relative motion of the geomagnetic and geographic poles. These features are not evident in the geomagnetic inertial frame and hence are not connected with our model of the magnetospheric electric field.

Although the plasma trajectories have simple shapes in the magnetic quasi-inertial frame, they are more complex in the geographic inertial frame because of the motion of the geomagnetic pole about the geographic pole. This is illustrated in Figure 6, where three representative plasma trajectories have been followed for a 24-hour period in the geographic inertial frame. Also shown in this figure are the positions of the terminator at winter solstice (W), equinox (E), and summer solstice (S).

The trajectory in panel A corresponds to trajectory 3 in Figure 4. This trajectory has a circulation period of about 2 hours, and hence the plasma executes many cycles per day. Depending on the location of the terminator, the plasma may drift entirely in darkness, entirely in sunlight, or move in and out of sunlight many times during the course of the day. The trajectory in panel B corresponds to trajectory 4 in Figure 4, and its circulation period is longer than 24 hours. For equinox conditions, plasma following this trajectory would cross the terminator three times in a 24-hour period. Panel C corresponds to trajectory 7 in Figure 4. The behavior of this trajectory in the geographic inertial frame is similar to that of trajectory 3, but its circulation period is longer, and hence fewer cycles are executed in 24 hours.

Comparison of Northern and Southern Hemisphere Plasma Convection

We studied the differences expected to be found in comparing plasma convection patterns in the northern and southern hemispheres. The differences due to different offsets between the geographic and geomagnetic poles (Bond, 1968) were studied by considering a quiet-time uniform magnetospheric electric field model. The anti-symmetric merging of the interplanetary magnetic field with the geomagnetic field in the northern and southern hemispheres (Heppner, 1973) was also considered by using a model of the magnetospheric electric field which contains enhanced flow in the dawnside northern hemisphere in conjunction with enhanced flow in the duskside southern hemisphere.

From our study we found the following: (1) The UT dependence of the gross plasma flow distributions in the two hemispheres is similar, but there is a phase shift of about half a day between them. Even taking acount of this half-day phase shift these plasma flow distributions reveal differences of up to 5° in latitude and an hour in local time between similar features in the two hemispheres. These differences are due to differences in the latitudes and longitudes of the magnetic poles in the geographic inertial frame. The 180° longitude difference results in the half-day phase shift, while the difference in latitudes is mainly responsible for the detailed differences; (2) A complete transformation of both the plasma location and velocity from the geographic inertial frame to the geomagnetic quasi-inertial frame results in the same non-UT dependent flow distribution for both hemispheres; and (3) The relative motion between the geographic and geomagnetic frames is a velocity of magnitude 96 m/s and 130 m/s for the northern and southern hemispheres, respectively. The direction of this velocity in the geographic inertial frame is UT dependent and is approximately anti-parallel in the two hemispheres. A consequence of these two anti-parallel velocities is that the northern and southern hemisphere plasma flows in the geographic inertial frame can differ in magnitude by up to 200 m/s and can display significantly different directions even if the same uniform magnetospheric electric field is mapped into the two polar ionospheres.

Comparison of Model Predictions with Incoherent Scatter Measurements

In order to determine the extent to which our plasma convection model could account for specific observations, plasma convection patterns predicted by our model were compared to those observed simultaneously at Chatanika, Alaska and Millstone Hill, Massachusetts (Sojka et al, 1979b). These two incoherent scatter facilities operated over the same period of four days in June, 1978 and provided data sets which were averaged to 24 hours in order to minimize the effects of individual substorms. In general, good qualitative agreement was obtained between our plasma convection model and the different diurnal patterns observed simultaneously at Chatanika and Millstone Hill.

Figure 7 shows an example of the comparison between predicted and observed diurnal patterns of plasma convection for Chatanika. For the comparison shown we used the magnetospheric potential distribution given by model (b) in Figure 3. At Chatanika (Figure 7), features of the predicted and observed regions of zonal flow match quite closely. Near local midnight the zonal flow is closest to Chatanika, an invariant latitude of 65^{o}. Poleward of this zonal flow is a region which contains both a discernable antisunward flow component before midnight and a reversed zonal flow in the early morning. The latitudinal extent of the zonal flow pattern and its magnitude as well as the flow characteristics poleward of this zonal flow are well represented by the model predictions. The effects of substorms are not included in the model, thus resulting in its smooth overall appearance when compared to the observed data.

The predictions for the Millstone Hill site also show good general agreement with the data. However, the model predictions and observations for Millstone Hill primarily show the main region of zonal flow; only around local midnight is there any evidence of antisunward flow. The magnitude and direction of the predicted and observed flows compare favorably, which is noteworthy since the convection pattern predicted for Millstone Hill is significantly different from that predicted for Chatanika even though the two sites cover approximately the same magnetic latitude range.

Even better agreement between model predictions and observations for both Chatanika and Millstrone Hill were obtained when we used the magnetospheric electric potential given by model (c) in Figure 3. In particular, the times of the zonal flow reversal match more closely. However, we are reluctant to recommend the use of the skewed magnetospheric potential at this time because of the limited data base used in the comparison and the lack of a physical basis for introducing the skewed feature in our model.

3. STEADY STATE DAYTIME IONOSPHERIC DENSITIES

We studied the behavior of O^+ and N^+ ions in a convecting daytime F layer for a range of geophysical conditions covering solar cycle, seasonal, and geomagnetic variations. Both zonal and meridional convection electric fields were considered. Without allowance for electric fields, the peak O^+ and N^+ densities varied by an order of magnitude and the altitudes of the peaks varied by about 100 km over the range of geophysical conditions studied. Convection electric fields can also produce about an order of magnitude change in the O^+ and N^+ densities. These electric field induced changes could either assist or oppose the solar cycle, seasonal, and geomagnetic activity variations depending on the ionospheric conditions. In general, N^+ was the second most abundant ion in the upper F region, but there were cases when He^+ was more abundant than N^+ even though He^+ was in a state of outflow.

Seasonal Variations

Although our primary interest was to determine the solar cycle, seasonal, and geomagnetic activity variations of N^+ and O^+ in the daytime high-latitude F region, six different ion species were considered, including NO^+, O_2^+, N_2^+, O^+, N^+, and He^+. Altitude profiles of all of the ion densities are shown in Figure 8 for two geophysical situations. The ion densities in both panels were calculated for a model atmosphere appropriate to solar maximum and high geomagnetic activity conditions, but the convection electric field was set to zero, and consequently, the results are representative of a region of small ion-neutral relative velocities. The two panels show a seasonal effect, with the top panel calculated for summer solstice and the bottom one for winter solstice.

There are several aspects concerning the ion densities shown in Figure 8 that should be noted. First, the summer-winter comparison does not display a 'winter anomaly'; that is, the peak electron density in winter is not greater than that in summer. The lack of a winter anomaly is related to the fact that the ionosphere is in darkness in winter below about 300 km at the latitude of our calculations. This lack of sunlight at low altitudes in winter is also evident in the N_2^+ density profile. In developing our ionospheric model, we assumed that N_2^+ was in chemical equilibrium at all altitudes, and therefore the absence of sunlight implies the absence of N_2^+ under steady state conditions.

Another noteworthy feature shown in Figure 8 is that the He^+ density in winter is much greater than that in summer, with the result that He^+ is the second most abundant ion above 600 km in winter. This is a consequence of the so-called 'winter helium bulge.' However, in summer, N^+ is the second most abundant ion at all altitudes above about 350 km.

Solar Cycle Variations

The effect of changing solar activity on the N^+ and O^+ densities is shown in Figure 9, where we compare profiles calculated at solar maximum and solar minimum for summer solstice and low geomagnetic activity. At solar maximum the peak O^+ density and the height of the peak are greater than at solar minimum. Also, except for altitudes below about 240 km, the O^+ density is greater at solar maximum than at solar minimum. This general behavior is consistent with that expected for a neutral atmosphere temperature increase from solar minimum to solar maximum (cf. Banks and Kockarts, 1973).

The N^+ density, on the other hand, exhibits a variation with solar activity that is, in general, opposite to that of O^+, i.e., the N^+ density is lower at solar maximum than at solar minimum. At solar maximum all of the neutral species have enhanced densities, and although the N^+ production rate is greater above 200 km due to the enhanced N_2 and N densities, the enhancement in the $N^+ + O_2$ loss rate due to the

increased O_2 density is more than sufficient to offset the increased production. Therefore, the N^+ density is lower at solar maximum than at solar minimum.

As was found for O^+, the altitude of the peak N^+ density is higher at solar maximum than at solar minimum. The peak N^+ density occurs at the altitude where diffusion becomes comparable to chemical loss. At solar maximum the neutral densities and the O^+ density are greater, which acts to reduce the N^+ diffusion coefficient and raise the altitude at which diffusion and chemical loss are comparable.

The solar activity comparison shown in Figure 9 was for summer solstice and low geomagnetic activity. The other three solar activity comparisons obtained from the different combinations of summer and winter solstice and high and low geomagnetic activity produced results that were very similar to that shown in Figure 9, but the difference between corresponding solar maximum and solar minimum densities was less.

Geomagnetic Activity Variations

The variations of the N^+ and O^+ densities with geomagnetic activity are more complicated than those found for seasonal or solar cycle changes. Typical results are shown in Figure 10, which were calculated for summer solstice and solar maximum. For these calculations, the O^+ density was lower at all altitudes for high geomagnetic activity than for low geomagnetic activity, while the N^+ density exhibited very little change with geomagnetic activity. The reduction in the O^+ density resulted from an increased $O^+ + N_2$ reaction due to the significantly enhanced N_2 density (factor of 10 at 370 km) that is characteristic of high geomagnetic activity. The small difference between N^+ density profiles resulted primarily from compensating changes in the N^+ production and loss rates, since diffusion was larger than chemical loss only above 500 km for the high geomagnetic activity case and above 440 km for the low geomagnetic activity case. On the one hand, the N^+ production rate was increased for high geomagnetic activity due to the significantly increased N_2 density, but opposing this was a reduction in the N^+ production rate due to a decreased N density (factor of 2 at 370 km) and an enhancement in the N^+ loss rate due to an increased O_2 density (factor of 5 at 370 km). These compensating changes in the N^+ production and loss rates led to the small difference in the N^+ density profiles calculated for high and low geomagnetic activity.

Meridional Electric Fields

The magnitude of the meridional or north-south convection electric field generally is about 50 mVm^{-1}, but can be as high as 200-300 mVm^{-1} on occasion. The primary effect of a meridional electric field is to heat the plasma through frictional interaction with the neutral atmosphere, and the resulting elevated ion temperature acts to alter the ion chemical reaction rates and high-altitude scale heights. In previous studies [Schunk et al, 1975, 1976] we have shown that large meridional electric fields (>100 mVm^{-1}) significantly affect the ion composition and the molecular ion to atomic ion transition altitude. However, the occurrence of large electric fields is generally restricted spatially and temporally, while electric fields of the order of 50 mVm^{-1} are more common over most of the polar cap.

We repeated our calculations covering solar cycle, seasonal, and geomagnetic activity variations including the 50 mVm^{-1} meridional electric field, and we found that the electric field acted to either increase or decrease the N^+ and O^+ densities depending on the ionospheric conditions. The changes in the N^+ and O^+ densities were either in or out of phase, again depending on the ionospheric conditions. Also, in some cases the changes were negligibly small.

A typical case is shown in Figure 11, which corresponds to solar maximum, summer, and high geomagnetic activity. For these conditions the inclusion of the meridional electric field acts to decrease the O^+ density at all altitudes. The N^+ density was also decreased slightly above 350 km. These density changes resulted from changes in the ion chemical reaction rates and diffusion coefficients due to electric field heating. For O^+ the dominant change involved the $O^+ + N_2$ reaction. At 300 km the electric field raised the ion temperature from 1414°K to 2380°K, which acted to increase the $O^+ + N_2$ reaction rate and hence reduce the O^+ density.

Zonal Electric Fields

The magnitude of the zonal or east-west electric field is generally much smaller than that of the meridional electric field, with the maximum value being about 30 mVm^{-1}. Whereas a meridional electric field acts primarily to heat the plasma, the main effect of a zonal electric field is to induce an upward or downward electrodynamic drift. Even though the magnitude of the zonal electric field is typically much smaller than the meridional electric field, a zonal electric field can, at times, have a more dramatic effect on the ion composition [cf. Schunk et al., 1976]. In our model studies, we adopted a value of 15 mVm^{-1} for the zonal electric field, since this value is more typical of general polar cap conditions.

We repeated our calculations covering solar cycle, seasonal, and geomagnetic activity variations including both westward and eastward electric fields of 15 mVm^{-1}, and we found several interesting effects. Figure 12 shows N^+ and O^+ density profiles for solar maximum, summer, and low geomagnetic activity. The curves are numbered from (1) to (3), which corresponds to a westward field (downward drift), no field, and an eastward field (upward drift), respectively. The progression from a downward drift to no drift, to an upward drift leads to an increase in both N_mF2 and h_mF2. This simply results from the fact that an upward drift raises the F layer to an altitude where the loss rate is lower, and hence the peak O^+ density is increased. The reverse is true for a downward electrodynamic drift. The minor ion N^+ also exhibits a behavior similar to that of O^+, but major ion drag is important in addition to the direct effect of the electrodynamic drift.

4. HIGH-LATITUDE F-REGION MORPHOLOGY

Our combined plasma convection and ionospheric-atmospheric model was used to study the high-latitude winter F-region at solar minimum for low magnetic activity. Our numerical study produced time-dependent, 3-dimensional, ion density distributions for the ions NO^+, O_2^+, N_2^+, O^+, N^+, and He^+. We covered the high-latitude ionosphere above 54°N magnetic latitude and at altitudes between 160-800 km for a time period

of one complete day. The main result we obtained was that high-latitude ionospheric features, such as the 'main trough', the 'ionization hole', the 'tongue of ionization', the 'aurorally produced ionization peaks', and the 'universal time effects', are a natural consequence of the competition between the various chemical and transport processes known to be operating in the high-latitude ionosphere.

O^+ Density Morphology

A convenient way to display the UT variation of the high-latitude ionosphere is by the use of contour plots of ion density at fixed altitudes. Figure 13 shows the resulting contours of the O^+ density at 300 km for four universal times. In this figure the gray scale range was chosen to emphasize low density regions. The contour plots in Figure 13 clearly show a marked UT variation of the O^+ density distribution over the region poleward of 55^o latitude. Both contour plots show the gross features of a mid-latitude (or main) ionization trough, a ring of enhanced ionization in the vicinity of the auroral oval, and a high-latitude ionization hole around local dawn. However, the detailed characteristics of these features differ for the four UT times shown. For example, the depth and extent of the mid-latitude ionization trough differ considerably.

With regard to the depth of the polar hole, our model calculations produced O^+ densities as low as 2×10^2 cm^{-3}, but there was also a significant UT variation. This UT variation can be clearly seen by comparing the four UT plots shown in Figure 13. Not only is there a significant variation in the size and location of the polar hole, but at 08h UT a region of low O^+ density extends into the noon sector. This feature is only just discernable on the 20 h UT plot. It should also be noted that the location of the polar hole corresponds very well with that measured by Brinton et al (1978).

Simulated Satellite Measurements

An important result to emerge from our study is that certain ionospheric processes can be better analyzed in different reference frames. For example, since the terminator is essentially fixed on the time scale of several days in the GEI frame, effects related to solar radiation should be studied in this frame. Likewise, since the auroral oval and the convection electric field are related to magnetospheric effects, they should be studied in the MLT frame. However, because of the competition between the various high-latitude processes, neither the GEI nor the MLT frame is completely satisfying. Also, the UT dependence of the high-latitude ionosphere introduces a further complication with regard to the interpretation of either ground-based or satellite measurements.

To determine exactly what is observed in a given reference frame, we simulated a 300 km circular satellite orbit that traverses the polar region in the dawn-dusk plane, and we plotted the O^+ densities that would be observed along the satellite track in both the GEI and MLT frames. The results are shown in Figure 14, with the left panels corresponding to the GEI frame and the right panels to the MLT frame. The top panel of each column shows two simulated satellite trajectories a half day apart (08h and 20h UT), as viewed in the GEI and MLT frames. The O^+ densities that would be observed along these satellite tracks are shown in the middle panels, while the bottom panels show the O^+ densities that would be observed along a series of orbital tracks. In these panels negative co-latitude is defined to be in the dusk sector, while positive values correspond to the dawn sector.

As the satellite moves from dusk toward dawn, the 'observed' O^+ density first decreases in the region of the main trough and then abruptly increases as the satellite enters the auroral oval. For the 08h UT orbit the O^+ density remains enhanced until the satellite leaves the oval on the dawn side. For the 20h UT orbit, on the other hand, the satellite leaves the oval and enters the polar cap, which at this time contains low O^+ densities, with the minimum in the O^+ density occurring near the location of the polar hole. For this 20h UT orbit, the satellite then enters the dawn side of the auroral oval and the O^+ density increases again.

In the GEI frame the two orbits show very different locations for the equatorward edge of the oval, while in the MLT frame the equatorward edge of the oval appears at approximately the same location for the two orbits. Also, in the GEI frame the variation of the O^+ density with co-latitude in the afternoon trough is similar for both orbits, while in the MLT frame this variation is different for the two orbits. These results demonstrate the importance of displaying and analyzing data in the reference frame of the dominant or controlling process. In the main trough the densities vary systematically in the GEI frame, while in the auroral oval they vary systematically in the MLT frame. A region like the polar cap is more complex, since even if particle precipitation is not important, the coupling of magnetospheric convection and solar radiation processes makes it difficult to find a unique reference frame for data presentation.

The bottom panels in Figure 14 show the confusion that could arise from accumulating and displaying data from a whole day's worth of satellite crossings of the polar region. Each symbol in these panels represents data from a particular satellite crossing. As with measured satellite data, the interpretation of the results is not always straightforward. However, with the knowledge of what our model predicts it is possible to identify some of the key ionospheric features in these two panels.

Region A in the GEI frame shows how the trough densities vary smoothly with co-latitude, while in the MLT frame this region is rather variable. In the GEI frame, region B shows a higher density band parallel to region A, which corresponds to plasma field tubes which have entered the oval, enhanced their densities, left the oval in the polar region, and then proceeded to convect toward dusk. This band is not clearly seen in the MLT reference frame. Feature E in the MLT frame represents the equatorward edge of the oval; there is no single corresponding feature in the GEI frame. The low densities in the polar hole are feature D, while the morning sector trough characteristics are feature C. The morning sector auroral oval is indicated in the MLT frame as feature F.

Ion Composition

As the ionospheric plasma convects through the different regions of the high latitude ionosphere, it

is subjected to a myriad of dynamical and chemical processes. The competition between these different processes produces a wide range of ion compositions in the polar F-region at different locations and times. Four examples are shown in Figure 15, where we present altitude profiles of the various ionospheric ions for selected locations and times. Panels A and B are for the same MLT location, near the polar hole, but at different UT's (24h and 05h, respectively). In both examples the altitude of the peak electron density is considerably below 300 km. For panel A the peak electron density is located at about 200 km and is composed of almost equal amounts of NO^+, O_2^+ and O^+, while for panel B the peak is located at 170 km and is essentially all NO^+. The densities in Panel C are from the polar cap at local midnight near the poleward edge of the auroral oval at a UT of 21h. In this case the F-peak is an O^+ peak and is located at about 240 km; however, at higher altitudes the N^+ density becomes comparable to the O^+ density. This plasma has convected across the polar cap from the dayside oval in about 1/3 of a day. In contrast, panel D is from inside the oval at local midnight and 06h UT. Here the effect of auroral particle production is highlighted. The dominant F-region ion is O^+ and the peak is at 240 km; however at low altitudes the auroral production of the molecular ions is evident in the form of enhanced molecular ion densities.

The morphology of the high-latitude ion composition is shown in Figure 16, where we present contours of the mean ion mass at two universal times for each of three altitudes. It is apparent that there is a marked UT variation of the distribution of the mean ion mass at the three altitudes. At a given UT, the mean ion mass decreases with increasing altitude, as expected. However, at 300 km there is still a significant abundance of molecular ions at some locations indicated by the region where the mean mass is greater than 16 amu. Also, at 200 km certain regions have a large fraction of atomic ions (mean ion mass= 16 amu). This large variation of mean ion mass with UT and altitude has important implications for the interpretation of incoherent scatter radar spectra.

5. SUMMARY

Recently, we developed a dynamic model of the high-latitude ionosphere and we employed this model to study several aspects of high-latitude plasma convection, ion composition, and ionospheric features, such as the main trough, the polar ionization hole, the aurorally produced ionization peaks, and the universal time effects.

From our convection study we found the following:

1. In the magnetic quasi-inertial frame the combination of UT-independent corotation and a time-independent convection electric field produces a flow pattern that does not vary with universal time.
2. This UT-independent flow pattern becomes UT dependent when viewed in a geographic inertial frame because of the motion of the magnetic pole about the geographic pole.
3. The UT variation of the plasma flow pattern in the geographic inertial frame occurs on a time scale that is comparable to satellite orbital periods and that is much less than typical plasma flow times over the polar cap.
4. In the geographic inertial frame the main region of very low speed flow is not centered at 1800 LT but moves from about 1300 to 2300 LT during the course of a day. The size and shape of this region also vary with universal time.
5. In the geographic inertial frame, additional low-speed regions occur in the morning sector at certain universal times.
6. Low cross-tail electric potentials, which correspond to quiet geomagnetic activity, produce a plasma drift pattern which has a pronounced low-speed region in the polar cap in addition to the one that normally occurs between 1300 and 2300 LT.

From our study of the daytime high-latitude ionosphere, we found the following:

1. The changes in atmospheric composition due to solar cycle, seasonal, and geomagnetic activity variations had a pronounced effect on the high-latitude ion densities and composition. The solar zenith angle also played a very important role at high-latitudes.
2. The peak O^+ and N^+ densities varied by about a factor of 10 over the range of conditions studied.
3. The altitude of the peak O^+ density varied from about 220 km to about 350 km over the range of geophysical conditions. In general, the N^+ peak was broader and occurred at a higher altitude than the O^+ peak. The altitude of the N^+ peak varied from about 350 to 430 km for the same range of geophysical conditions.
4. The peak O^+ density was greater and the height of the peak was higher at solar maximum than at solar minimum. Even though the height of the peak N^+ density was higher at solar maximum than at solar minimum, the N^+ density was smaller at all altitudes.
5. Seasonal variations of the N^+ and O^+ densities were greater than either solar cycle or geomagnetic activity variations, and this was due primarily to the fact that in winter the ionosphere was in darkness below 300 km at the geographic latitude of our calculations (80^o). In winter the O^+ density was a factor of 10 lower and the N^+ density a factor of 4 lower than in summer.
6. The variations of the N^+ and O^+ densities with geomagnetic activity were more complicated than those found for either solar cycle or seasonal changes. The O^+ density displayed a regular pattern at low altitudes in that it was reduced in going from low to high geomagnetic activity due to the increased N_2 density. However, differing scale heights acted to complicate the situation at altitudes above the F region peak density. The N^+ density, on the other hand, exhibited a negligibly small change or was reduced in going from low to high geomagnetic activity depending on seasonal and solar cycle conditions. This N^+ variation was due primarily to changes in the N^+ photochemical reaction rates.
7. In general, electric fields had a large effect on the N^+ and O^+ density profiles, and their effect could either assist or oppose the solar cycle, seasonal, and geomagnetic activity variations discussed above depending on the magnitude and direction of the electric field.
8. The meridional or north-sourth electric field acted to either increase or decrease the N^+ and O^+ densities depending on the ionospheric conditions. The changes in the N^+ and O^+ densities were either in or out of phase, again depending on the ionospheric conditions. These density changes resulted from changes in the chemical reaction rates and diffusion coefficients due to the electric field enhanced ion temperatures.
9. For summer solstice the progression from a westward field (downward drift) to no field, to an eastward field (upward drift) led to a significant increase in the peak density and the altitude of the

peak for both N^+ and O^+. However, at low altitudes (below 250 km for O^+ and 350 km for N^+) the electrodynamic drift had a negligibly small effect on the ion densities.

10. For winter solstice the effect of the electrodynamic drift was more complicated. Whereas the summer calculations produced a smaller $N_m F_2$ for high geomagnetic activity than for low geomagnetic activity, the reverse was true for the winter calculations due to the greater importance of diffusion in winter.

From our study of the morphology of the F-region, we found the following:

1. High-latitude ionospheric features, such as the main trough, the ionization hole, the tongue of ionization, the aurorally produced ionization peaks, and the universal time effects, are a natural consequence of the competition between the various chemical and dynamical processes known to be operating in the high-latitude ionosphere.

2. The high-latitude F-layer ionization exhibits a significant UT variation owing to the displacement between the geomagnetic and geographic poles. The peak electron density at a given location and local time can vary by an order of magnitude due to this UT effect.

3. Some ionospheric processes can be better analyzed in certain reference frames. For example, processes related to solar radiation should be studied in the geographic inertial frame, while the effects of magnetospheric processes should be studied in a magnetic frame. However, because of competition between the various high-latitude processes, neither the geographic inertial frame nor a magnetic frame is completely satisfying.

4. Because of the competition between the various dynamical and chemical processes, a wide range of ion compositions can occur in the polar F-region at different locations and times. Also, the height of the F-region peak electron density displays a significant spatial and temporal variation.

5. The minimum value for the electron density in the main trough is sensitive to nocturnal maintenance processes, such as ion production due to resonantly scattered radiation, ion production due to a small flux of precipitation particles, and a neutral wind induced upward ionization drift.

6. The depth and longitudinal extent of the main trough exhibit a significant UT dependence.

7. The spatial extent, depth, and location of the polar ionization hole are UT dependent.

8. Ion production due to resonantly scattered radiation can maintain the electron density at 300 km in the polar ionization hole at about 2×10^2 cm^{-3}, in agreement with the measurements of Brinton et al (1978).

9. The mean ion mass exhibits a large variation with both universal time and altitude, and this variation has important implications for the interpretation of incoherent scatter radar spectra.

Acknowledgements This research was supported by Air Force contract USAF/ESD F19628-79-C-0025, NSF grant ATM78-10501 and NASA grant NAGW-77 to Utah State University.

References

Bahnsen, A., 1978, Recent techniques of observations and results from the magnetopause regions, J. Atmos. Terr. Phys., 40, 235-256.

Banks, P.M., and J.R. Doupnik, 1975, A review of auroral zone electrodynamics deduced from incoherent scatter radar observations, J. Atmos. Terr. Phys., 37, 951.

Banks, P.M. and G. Kockarts, 1973, Aeronomy, Academic, New York.

Banks, P.M., Schunk, R.W. and Raitt, W.J., 1974, NO^+ and O^+ in the high latitude F-region, Geophys. Res. Letters 1, 239-242.

Bond, F.R., 1968, Magnetic and auroral conjugacy, Ann. Geophys., 24.

Brinton, H.C., J.M. Grebowsky, and L.H. Brace, 1978, The high-latitude winter F region at 300 km: Thermal plasma observations from AE-C, J. Geophys. Res., 83, 4767-4776.

Engebretson, M.J., K. Mauersberger, D.C. Kayser, W.E. Potter, and A.O. Nier, 1977, Empirical model of atomic nitrogen in the upper thermosphere, J. Geophys. Res., 82, 461-471.

Fairfield, D.H., 1977, Electric and magnetic fields in the high-latitude magnetosphere, Rev. Geophys. Space Phys., 15, 285-298.

Heelis, R.A., W.B. Hanson, and J.L. Burch, 1976, Ion convection velocity reversals in the dayside cleft, J. Geophys. Res., 81, 3803.

Heppner, J.P., 1973, High latitude electric fields and the modulation related to interplanetary magnetic field parameters, Radio Sci., 8, 933-948.

Heppner, J.P., 1977, Empirical models of high-latitude electric fields, J. Geophys. Res., 82, 1115-1125, 1977.

Knudsen, W.C., 1974, Magnetospheric convection and the high-latitude F2 ionosphere, J. Geophys. Res., 79, 1046-1055.

Knudsen, W.C., P.M. Banks, J.D. Winningham, and D.M. Klumpar, 1977, Numerical model of the convecting F_2 ionosphere at high latitudes, J. Geophys. Res., 82, 4784.

Schunk, R.W. and Banks, P.M., 1975, Auroral N_2 vibrational excitation and the electron density trough, Geophys. Res. Lett., 2, 239-242.

Schunk, R.W. and W.J. Raitt, 1980, Atomic nitrogen and oxygen ions in the daytime high-latitude F-region, J. Geophys. Res., 85, 1255-1272.

Schunk, R.W., Raitt, W.J. and Banks, P.M., 1975, Effect of electric fields on the daytime high-latitude E-and F-regions, J. Geophys. Res., 80, 3121-3130.

Schunk, R.W., P.M. Banks and W.J. Raitt, 1976, Effects of electric fields and other processes upon the nighttime high latitude F-layer, J. Geophys. Res., 81, 3271-3282.

Smiddy, M., M. Kelley, W. Burke, F. Rich, B. Sagalyn, B. Schuman, R. Hays, and S. Lai, 1977, Intense poleward directed electric fields near the ionospheric projection of the plasmapause, Geophys. Res. Lett., 4, 543-546.

Sojka, J.J., W.J. Raitt and R.W. Schunk, 1979a, Effect of displaced geomagnetic and geographic poles on high-latitude plasma convection and ionospheric depletions, J. Geophys. Res., 84, 5943-5951.

Sojka, J.J., W.J. Raitt and R.W. Schunk, 1979b, High latitude plasma convection: Predictions for EISCAT and Sondre Stromfjord, Geophys. Res. Lett., 6, 877-880.

Sojka, J.J., J.C. Foster, W.J. Raitt, R.W. Schunk and J.R. Doupnik, 1980a, High-latitude convection: Comparison of a simple model with incoherent scatter observations, J. Geophys. Res., 85, 703-709.

Sojka, J.J., W.J. Raitt and R.W. Schunk, 1980b, A comparison of model predictions for plasma convection in the northern and southern polar regions, J. Geophys. Res., 85, 1762-1768.

Sojka, J.J., W.J. Raitt and R.W. Schunk, 1980c, A theoretical study of the high-latitude winter F-region at solar minimum for low magnetic activity, J. Geophys. Res., submitted.

Sojka, J.J., W.J. Raitt and R.W. Schunk, 1980d, Theoretical predictions for ion composition in the high-latitude winter F-region for solar minimum and low magnetic activity, J. Geophys. Res., submitted.

Spiro, R.W., R.A. Heelis, and W.B. Hanson, 1978, Ion convection and the formation of the mid-latitude F region ionization trough, J. Geophys. Res. 83, 4255-4264.

Stern, D.P., 1977, Large-scale electric fields in the Earth's magnetosphere, Rev. Geophys. Space Phys., 15, 156-194.

Volland, H., 1975, Models of global electric fields within the magnetosphere, Annl. Geophys., 31, 159.

Watkins, B.J., 1978, A numerical computer investigation of the polar F-region ionosphere, Planet. Space Sci., 26, 559-569.

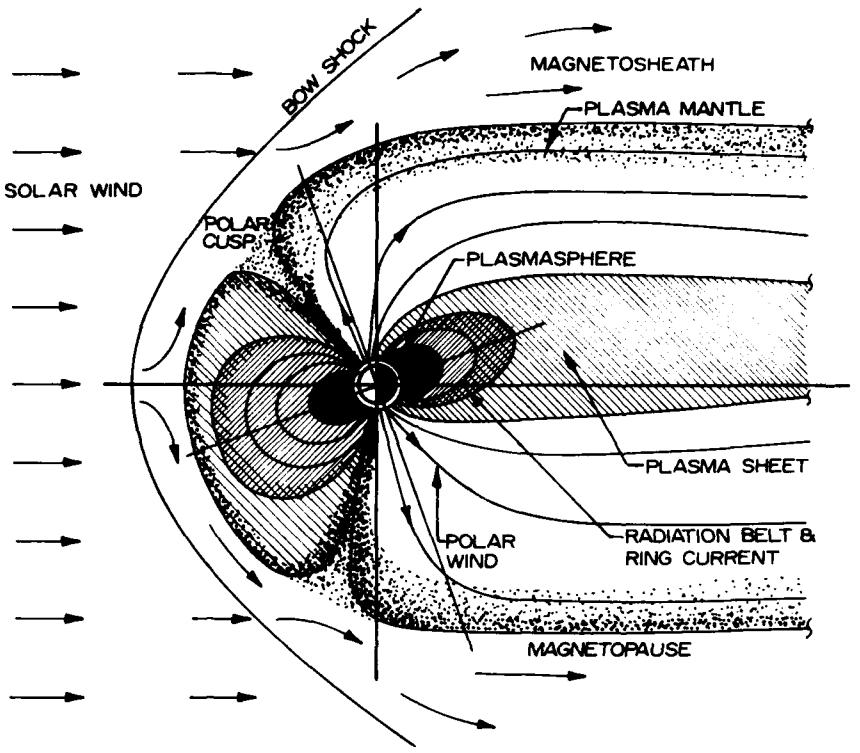


Fig. 1. Schematic illustration of the earth's magnetosphere. From Bahnsen (1978).

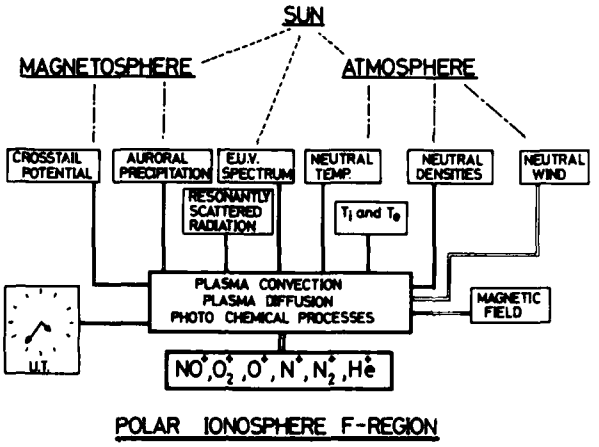


Fig. 2. Diagram showing the various physical processes affecting the F-region polar ionosphere.

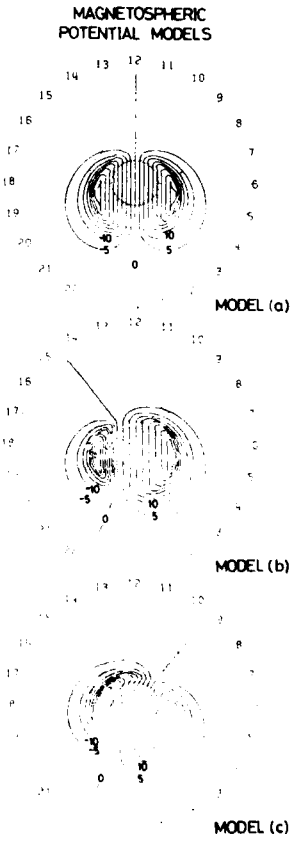


Fig. 3. Magnetospheric electric potentials mapped into the polar ionosphere shown in magnetic coordinates. Model (a) is for a uniform cross-tail electric field mapped about a point offset by 5° in the anti-sunward direction from the magnetic pole. Model (b) has the same characteristics as model (a), except the magnetospheric electric field is asymmetric with the enhancement on the dusk side. Model (c) has the same characteristic as model (a), except the magnetospheric potential distribution is rotated to be parallel to the 10-22 MLT meridian. All models are for a total cross-tail potential of 90 kV and the contour interval is 5 kV. From Sojka et al. (1979b).

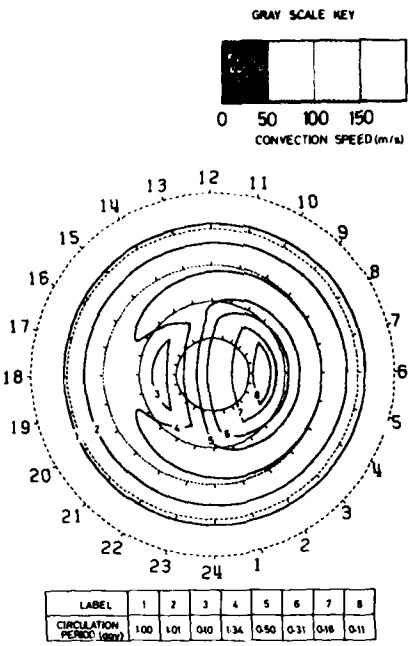


Fig. 4. Plasma drift trajectories over the polar cap viewed in a magnetic inertial frame. Each of the eight trajectories shown has a different circulation period, tabulated in the figure. For these calculations the total cross-tail potential was maintained at 64 kV. From Sojka et al. (1979a).

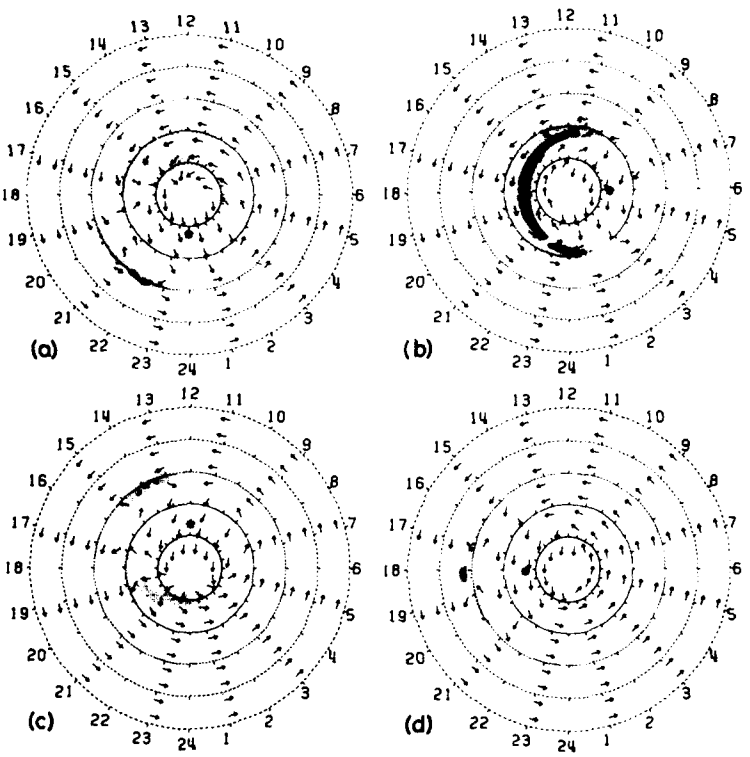


Fig. 5. Plasma flow directions and speed distributions viewed in the geographic inertial frame for nonaligned geomagnetic and geographic axes are shown for four universal times. The four universal times correspond to the location of the magnetic pole (shown by an asterisk) on the (a) midnight, (b) dawn, (c) noon, and (d) dusk meridians. The total cross-tail potential is 64 kV. From Sojka et al. (1979a).

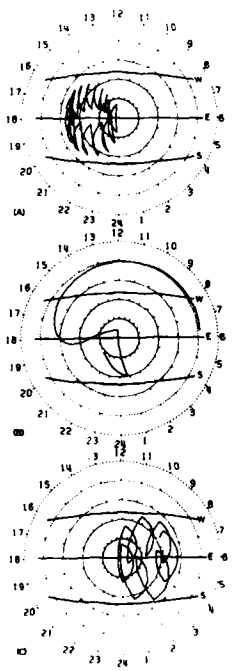


Fig. 6. Plasma drift trajectories over the polar cap viewed in a geographic inertial frame for nonaligned geomagnetic and geographic axes. The trajectories in panels A, B, and C correspond to those shown in Figure 4 and labelled 3, 4, and 7, respectively. These trajectories have been followed for a 24-hour period. The positions of the terminator for winter solstice (W), equinox (E), and summer solstice (S) are shown. From Sojka et al. (1979a).

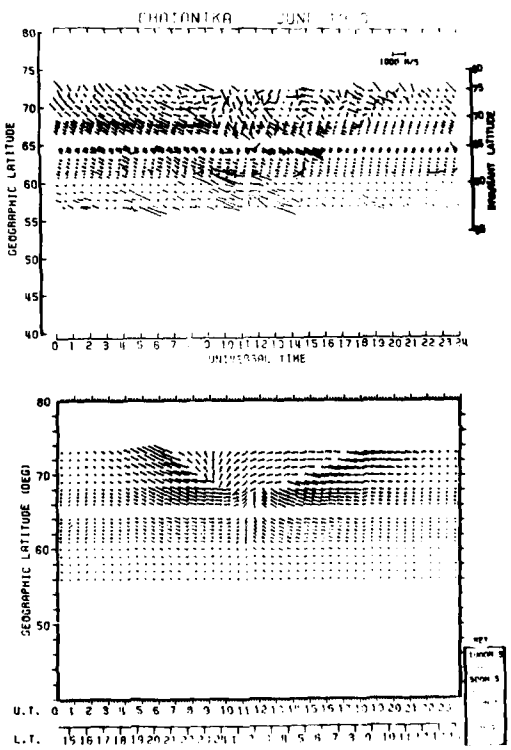


Fig. 7. Distribution of horizontal plasma drift velocities observed from Chatanika (upper panel) and the predictions for this location of our convection model (lower panel). From Sojka et al. (1980a).

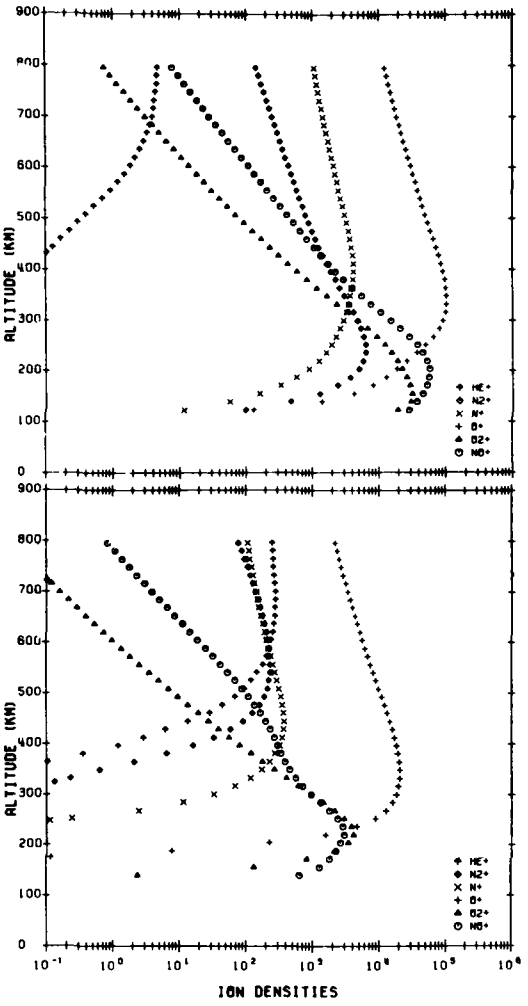


Fig. 8. Ion densities versus altitude for summer (top panel) and winter (bottom panel) solstice and for solar maximum and high geomagnetic activity. From Schunk and Raitt (1980).

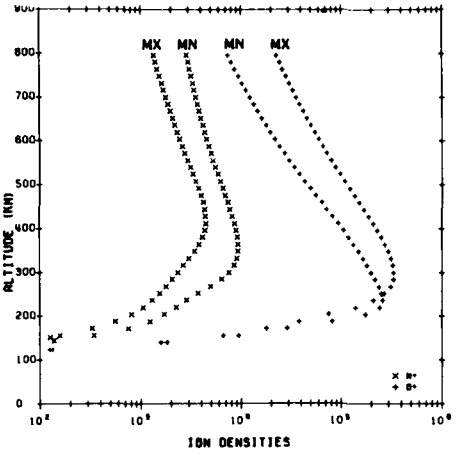


Fig. 9. N^+ and O^+ densities versus altitude at solar maximum (MX) and solar minimum (MN) for summer solstice and low geomagnetic activity. From Schunk and Raitt (1980).

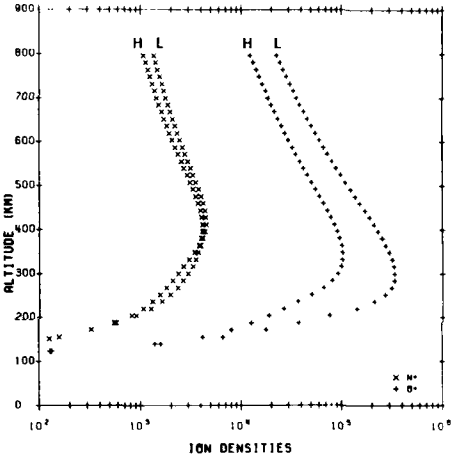


Fig. 10. N^+ and O^+ densities versus altitude at solar maximum for high (H) and low (L) geomagnetic activity and for summer solstice. From Schunk and Raitt (1980).

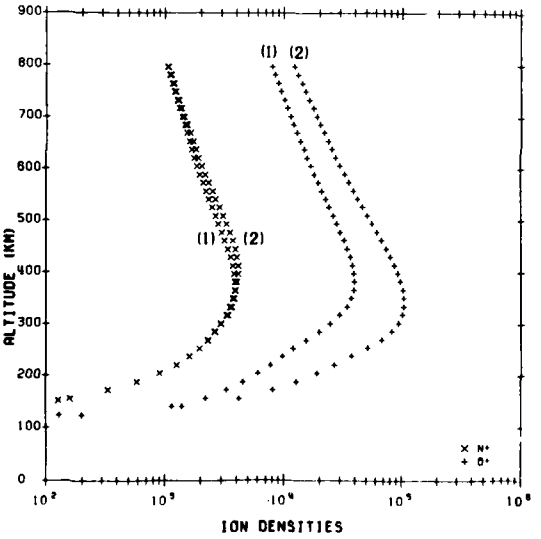


Fig. 11. N^+ and O^+ density profiles calculated with (1) and without (2) allowance for a meridional electric field of 50 mV m^{-1} for solar maximum, summer, and high geomagnetic activity. From Schunk and Raitt (1980).

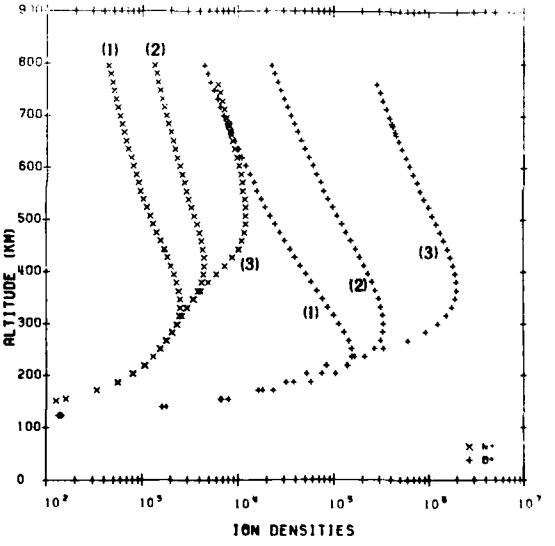


Fig. 12. N^+ and O^+ density profiles for several zonal electric fields. The electric fields are: (1) 15 mV m^{-1} westward; (2) 0 mV m^{-1}; and (3) 15 mV m^{-1} eastward. These calculations are for solar maximum, summer, and low geomagnetic activity. From Schunk and Raitt (1980).

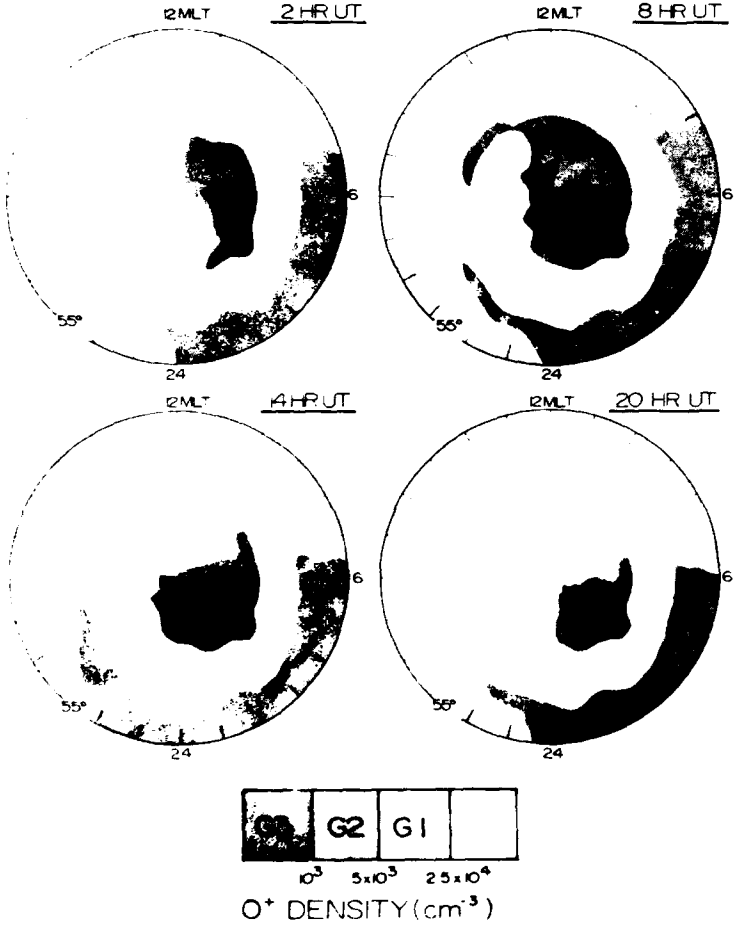


Fig. 13. O^+ Density contours at 300 km in the magnetic quasi-inertial frame for four universal times. From Sojka et al. (1980d).

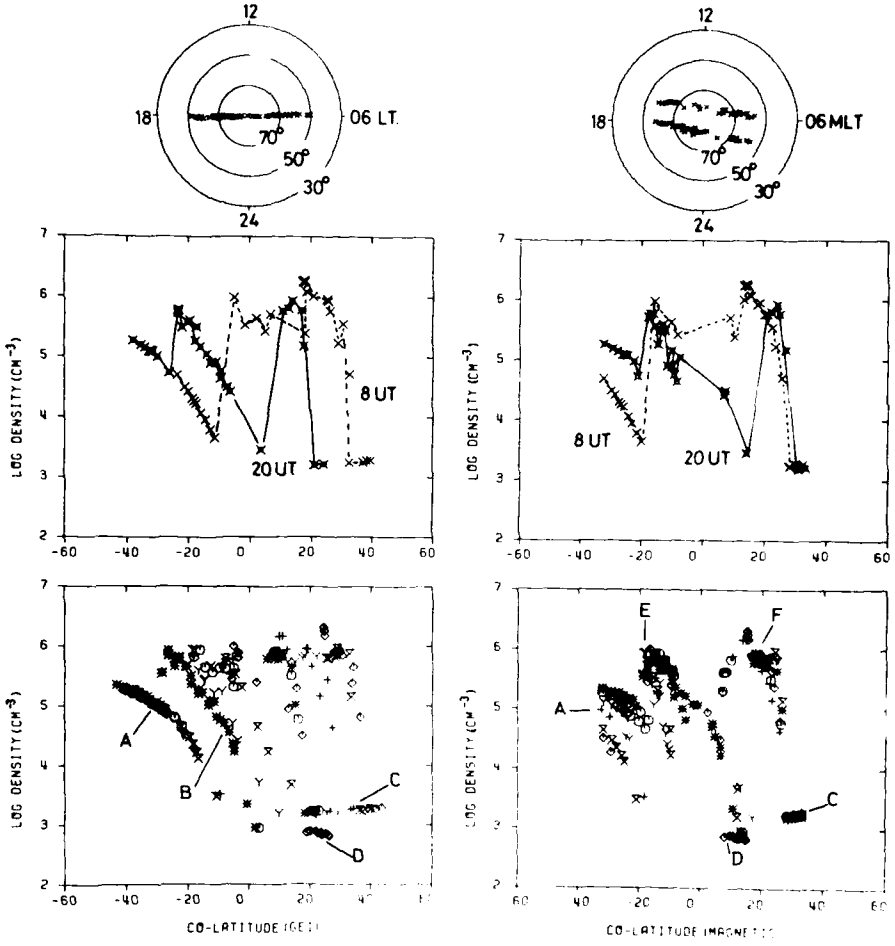


Fig. 14. Simulated satellite crossings of the northern polar cap in the dawn-dusk plane at 300 km. The top panels show two representative crossings 12 hours apart in both the geographic inertial frame (left panel) and the magnetic quasi-inertial frame (right panel). The center panels display the O^+ density along the satellite track for these crossings in both the geographic inertial frame (left panel) and the magnetic quasi-inertial frame (right panel). The bottom panels display the O^+ density along the satellite track for one full day. From Sojka et al. (1980c).

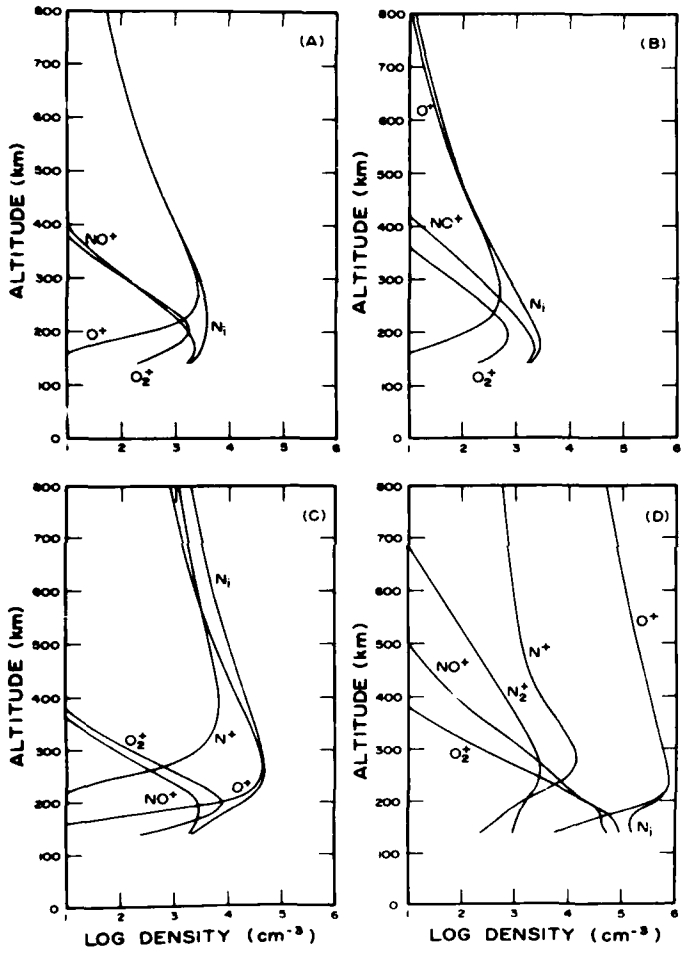


Fig. 15. Representative ion density profiles as a function of altitude. These profiles were selected to show the wide variation in composition that is possible depending on location and universal time. The locations and universal times are: (A) The polar hole at 24 UT; (B) The polar hole at 5 UT; (C) In the polar cap adjacent to the oval at local midnight and 12 UT; and (D) In the oval at local midnight and 6 UT. From Sojka et al. (1980c).

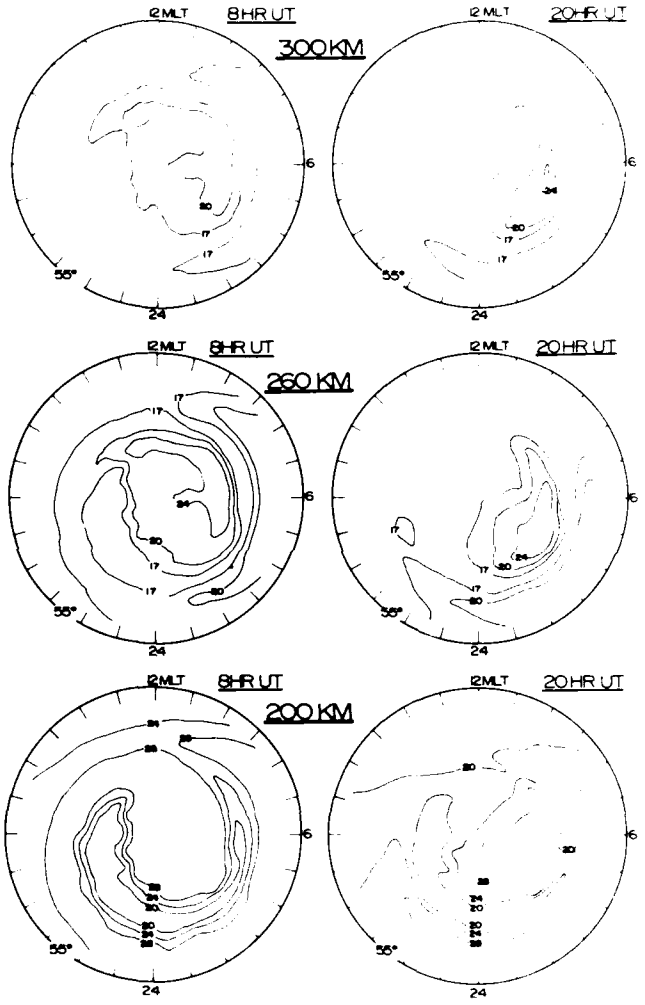


Fig. 16. Contours of mean ion mass at two universal times for each of three altitudes. The mean ion mass is in atomic mass units. The contours are plotted in a MLT, magnetic latitude reference frame. From Sojka et al. (1980d).

THE PHYSICS OF RADIO WAVE ABSORPTION

E V Thrane
Norwegian Defence Research Establishment
P O Box 25 - N-2007 Kjeller, Norway

ABSTRACT

Present knowledge of the physical causes of radio wave absorption in the ionosphere is reviewed. The theory for propagation of radio waves through a partially ionized atmosphere is well developed, but our understanding of the processes that govern the behaviour of the ionospheric plasma is still incomplete. The most important absorbing region comprise the D-region and lower E-layer. The paper discusses physical models of these layers that may explain the observed variation of absorption.

1. INTRODUCTION

The attenuation of radio waves reflected from and penetrating the ionosphere is of great practical importance for radio wave communication and navigation systems. The first attempts to measure the absolute magnitude of this attenuation were made by Appleton and Ratcliffe (1930), using the frequency-change method for medium frequency waves reflected from the E-region. They concluded from their experiment that the main part of the attenuation occured below the reflection level and named the absorbing region the D-region of the ionosphere. The later application of the pulse sounding (A1) method for absorption measurements confirmed a distinction between the D-region absorption and the attenuation suffered by a wave during the process of reflection in the E- or F-regions (Appleton 1937, Piggott 1953). Extensive experimental and theoretical investigations of both types of absorption have been made since that time. We now have a large amount of information on where and when absorption occurs. The basic magneto-ionic theory for propagation of radio waves in the ionospheric plasma has been further developed, so that, given a detailed description of the propagation medium, we can with good accuracy trace the path of the wave through the medium and compute the attenuation.

The present paper will not empahsize wave propagation theory, but rather discuss current understanding of the various physical processes that cause radio wave absorption in the ionosphere. Section 2 will give a brief introduction to the theory of absorption. Section 2 briefly deals with methods for absorption measurements and Section 3 describes some aspects of the morphology of absorption. In the subsequent sections the division of the subject is based on the global morphology, since the dominant physical causes of absorption change significantly with latitude.

2. SOME THEORETICAL CONSIDERATIONS

In the context of this paper absorption means dissipative attenuation of a radio wave through collisions of electrons, vibrating in the wavefield, with other particles. We shall not deal with other loss processes, such as spatial attenuation, scattering from irregularities in the medium etc. The absorption may be expressed as

$$L = \int_s \kappa ds \quad \text{(Nepers)} \tag{1}$$

when the index κ is absorption per unit length along the raypath s. The Appleton-Hartree theory for quasilongitudinal propagation gives the well known formula:

$$\kappa = \frac{e^2}{2\varepsilon_o mc}\,\frac{1}{\mu}\,\frac{N_e \nu}{(\omega \pm \omega_L)^2 + \nu^2} \tag{2}$$

where the + and - sign refer to ordinary and extraordinary waves, and

- e is the electronic charge
- m the electronic mass
- ε_o the dielectric constant for free space
- c the velocity of light
- μ the real part of the refractive index
- N_e the electron number density
- ν the electron-neutral collision frequency
- ω the angular wave frequency and
- ω_L the longitudinal component of the angular gyrofrequency ω_H for electrons in the earth's magnetic field.

It may be useful to remind the reader of a few of the properties of Equation (2).

2.1 Non-deviative absorption

When the refractive index $\mu \approx 1$ the rays suffer little bending and Equation (2) simplifies. This condition normally holds in the D-region. If in addition $\omega >> \nu$, the absorption in dB may be written (N_e, m^{-3}; ω Hz)

$$L = 4.6 \cdot 10^{-5} \frac{1}{(\omega \pm \omega_L)^2} \int_s N_e \nu ds \quad \text{(dB)} \tag{3}$$

For HF-waves this formula is a useful approximation, and illustrates the well-known inverse square frequency law. An absorption parameter which is independent of frequency may be defined as $A = \frac{L}{\pi^2}(\omega \pm \omega_L)^2$. For low frequency waves or absorption in the lower D-region, i e $\omega << \nu$, we find

$$L = 4.6 \cdot 10^{-5} \int_s \frac{N_e}{\nu} ds \quad \text{(dB)} \tag{4}$$

Figure 1 shows typical normal daytime values of N_e, ν and $N_e\nu$ in lower ionosphere, and illustrates that the main part of the absorption for HF-waves occurs below the E-region where $N_e\nu$ has a maximum.

2.2 Deviative absorption

Deviative absorption occurs where the refractive index μ is large, that is in the region where the wave is reflected. Using certain simplifying assumptions the deviative absorption may be expressed as

$$\int_s \kappa ds \propto \bar{\nu} \int_s (\frac{1}{\mu} - \mu) ds \tag{5}$$

where $\bar{\nu}$ is an average collision frequency. This absorption will depend upon the gradient in electron density near the point of reflection, and may vary strongly with frequency. Figure 2 shows an example of absorption measurements for waves reflected at vertical incidence. The smooth curve illustrates the expected frequency variation for the non-deviative absorption. The large values of absorption near the critical frequencies are due to deviative absorption. At oblique incidence the reflection occurs at a non-zero value of μ ($\mu = \sin I$, where I is the angle of incidence) and the deviative absorption is less important, and may for many purposes be neglected. Current methods for prediction of HF circuit reliability include estimates of deviative absorption based upon Equation (5) (Lloyd et al 1978).

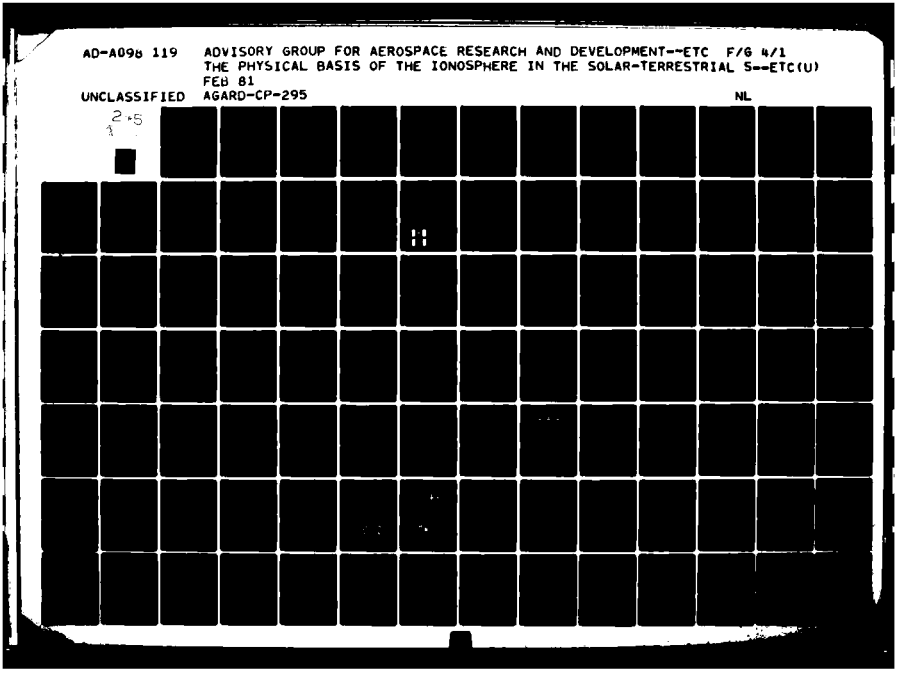
2.3 The generalized magneto-ionic theory

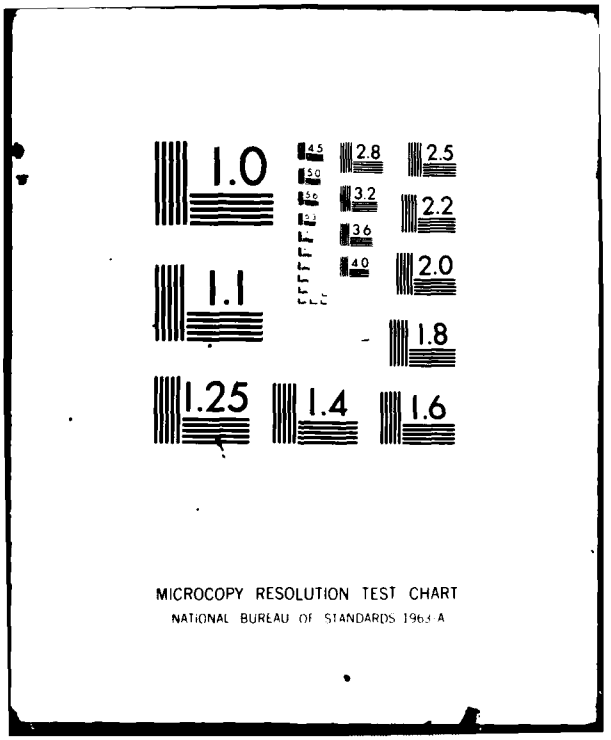
Sen and Wyller (1960) developed a generalized magnetoionic theory to take into account the velocity dependent momentum transfer cross section Q(v) for electrons with a maxwellian velocity distribution colliding with neutral molecules. Laboratory measurements of cross sections in nitrogen were available (Phelps and Pack 1959) at thermal energies, showing $Q(v) \propto v$. This implies that the collision frequency ν is proportional to electron energy u, $\nu(u) \propto u$. This relation was also assumed to hold for air, and the now familiar Sen and Wyller formulae for the complex refractive index were developed. These are employed in most modern radio propagation work where accurate computation of the refractive index is important. Their use is essential in regions where the angular radio frequency ω is comparable to the electron-neutral collision frequency ν. In the limiting cases $\omega \pm \omega_H >> \nu$ and $\omega \pm \omega_H << \nu$, the Appleton-Hartree equations may be used with appropriate effective collision frequencies.

Although the generalized formulae for the refractive index in the ionosphere derived by Sen and Wyller are accepted and in general use, a word of caution is in order. In the lower ionosphere the temperature is in the range 130 K to 300 K, and the values of the momentum transfer collision cross section in molecular oxygen are very uncertain in this temperature range. The only available laboratory data were derived from measurements of microwave conductivity (Mentzoni 1965, Veach et al 1966), and these differ from theoretical estimates made for example by Hake and Phelps (1967). A recent review of collision frequencies in the ionosphere by Aggarwal et al (1979) seems to have overlooked this difficulty. It is however correct, as stated by these authors, that the best available estimates indicate that molecular oxygen contributes a small amount (about 10%) of the total electron-neutral collision frequency in the entire region from 50 to 100 km. Both theoretical estimates and radio wave propagation experiments (Thrane and Piggott 1966) indicate that the collision frequency is proportional to atmospheric pressure in this height range. A commonly used relation is

$$\nu_M = Kp \tag{6}$$

where ν_M is the collision frequency of electrons with energy kT and p is the total atmospheric pressure. The parameter K should be constant if the collision cross-section Q is proportional to electron velocity as laboratory measurements in nitrogen indicate. K may be determined from measurements of pressure combined with values of ν_M derived from propagation experiments. One should not forget, however that this derivation of ν_M is based on the assumption that $Q \propto v$, (i e $\nu \propto u$), so that there is a danger of arguing in a circle. A discrepancy could show up as a variation in the value of K, with height or with atmospheric conditions (for example temperature changes). A


AD-A098 119
ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 4/1
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL S--ETC(U)
FEB 81
UNCLASSIFIED
AGARD-CP-295
NL
2-5


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

weak height dependence of K has indeed been found (Friedrich and Torkar 1980, Mechtly 1974). It seems unlikely that the difficulties discussed here could seriously affect computations of absorption except possibly under extreme conditions. The problem is unresolved, however, and if the collision cross section in molecular oxygen, contrary to expectations, is large and/or strongly varying at thermal energies, the magnetoionic theory may need revision.

3. MEASUREMENTS OF ABSORPTION

Most of the available data on ionospheric absorption have been obtained by three ground based experimental techniques:

a) The A1 method, using amplitude measurements of pulsed HF radio signals reflected from the E- or F-layer. (Appleton and Piggott 1954).

b) The A2 method, (riometer) using observations of intensity of galactic cosmic radio noise penetrating the ionosphere at VHF and UHF. (Hargreaves 1969).

c) The A3 method using amplitude measurements of MF or LF signals propagating via the ionosphere over medium length paths. (Lauter and Triska 1965).

4. THE MORPHOLOGY OF ABSORPTION

We have noted that the radio wave absorption in the ionosphere depends upon the electron density and collision frequency. For HF waves in the D-region it is simply proportional to $\int N_e \nu_M \, ds$. The collision frequency decreases exponentially with height, but the function $\nu_M(h) = K p(h)$ varies slowly with time. The diurnal variation is small, but at high latitudes there is a significant seasonal variation which agrees well with the seasonal pressure changes. (CIRA 1972), (Thrane 1968). It follows that all rapid variations of absorption are due to changes in electron density, including day-to-day and diurnal changes. In the following we will briefly describe the observed morphology of absorption.

4.1 The normal diurnal seasonal and sunspot cycle variation of absorption

The measured radio-wave absorption show certain irregular variations which will be discussed later, but also a regular, systematic dependence upon the solar zenith angle χ, and upon solar activity represented by the 13-month smoothed number R_z. For a radio wave passing through a simple theoretical Chapman-layer the absorption may be expressed as

$$L \propto \cos^n \chi \qquad (7)$$

where n = 3/2. If the relation (7) is fitted to the measured diurnal variation of absorption it is found (George 1971) that the exponent n varies with latitude from $n = 0.85 \pm 0.15$ at low latitudes to $n = 0.75 \pm 0.15$ at middle latitudes. At high latitudes the diurnal exponent is small, $n \approx 0.3$. Obviously the lower ionosphere does not behave like a Chapman-layer.

George (1971) has used the absorption measurements made during the IGY and IQSY to derive the global morphology of the quantity $A_o(T)$ which is a frequency independent absorption index for noon conditions, normalized to $\cos\chi = 1$ (an overhead sun) using Equation (6) with n = 0.8. $A_o(T)$ represents the total absorption suffered by a wave passing through the D- and E-regions and is proportional to $\int_s N_e \nu_M ds$ for a path through the regions.

Figure 3 shows $A_o(T)$ plotted against season and magnetic latitude, represented by a modified magnetic dip angle. According to George (1971), the use of this coordinate system gives significantly less scatter in the experimental values than the use of geographic latitude. This fact indicates a geomagnetic control of the absorption. A surprising feature of Figure 3 is the semiannual variation at low latitudes with a trough near the equator, and maxima near the equnioxes at dip angles between 25 and 30^o. The large absorption in winter at middle latitudes is the well known winter anomaly. This absorption is not only large but extremely variable. Figure 4 shows a typical variation of noon absorption in winter at middle latitudes.

The graph in Figure 3 was derived for a sunspot number $R_z = 200$. The data indicate that $A_o(T) \propto (1+0.0067R_z)$ and there is no indication of a significant change in the global pattern with sunspot number.

The average pattern of absorption in polar regions is shown in Figure 5. Here the zone of maximum occurence of absorption conicides in broad terms with the auroral zone. The high latitude regions are characterized by extreme variability in the observed absorption at all seasons.

5. THE PHYSICAL CAUSES OF ABSORPTION

We have noted that the short term variations in absorption are due to changes in electron density. The continuity equation for the electron density in the lower ionosphere may be written

$$(1+\lambda)\frac{dN_e}{dt} = q-\psi N_e^2-\nabla\cdot N\vec{v} \qquad (8)$$

where q is the ion pair production rate, ψ an effective recombination coefficient, $\lambda = \frac{N-}{N_e}$, the ratio of negative ion density to electron density, and $\vec{v}$ a velocity of the gas. Equation (8) must be the starting point for our discussion of the physics of absorption. It is convenient first to discuss the diurnal and sunspot cycle variation in general terms, and then deal with three different latitude regions, since some of the important physcial mechanisms vary significantly with latitude.

5.1 The diurnal variation of absorption

The regular diurnal variation of electron density in the D- and E-regions is caused mainly by the changes in the penetration depth of solar ultraviolet and X-rays with solar zenith angle. We now know that the total ionization density is a result of a complex interaction between ion production and loss processes, and it is hardly surprising that the diurnal variation of absorption deviates from that expected from a simple Chapman-layer, which is formed by mono-chromatic radiation ionizing a single atmospheric constituent. Can present lower ionosphere models explain the observed diurnal variations of absorption as described in section 4.1? Figure 6 illustrates the problem further by presenting the diurnal variation of the absorption index $A = L(f+f_L)^2$ during sunspot minimum conditions at Kjeller (60°N 11 E) for a winter and a summer month. Thrane (1972). Note the striking difference between the measurements and the Chapman-law. Using such graphs most experimenters have fitted a straight line through the measured values to derive an exponent n for the diurnal variation. For HF waves, however, the measurements often indicate a decrease in the slope (or n) with increasing zenith angle (see for example Singer et al 1979). Near noon in summer $n \approx 0.8$ at Kjeller, decreasing to $n \approx 0.2$ near sunrise and sunset. Thrane (1972) used a model for the ion production and loss processes in the lower ionosphere to compute the HF absorption as a function of solar zenith angle, and was able to simulate the observed diurnal variation as shown by the solid line in Figure 6. The model for the ion production is illustrated in Figure 7a and b for two zenith angles. The model for the loss rate assumed a seasonal change, but no diurnal changes. The loss rate models used in this work were crude and need some revision in view of recent results. The excellent agreement between model and observation may thus be fortuitous, but nevertheless two mechanisms were identified that can explain the observed slow diurnal variation of absorption. Firstly, the ionization of nitric oxide, NO, by solar H-Lyman-α is very important in the lower E-region near sunrise and sunset. This source of free electrons is turned on in the morning as soon as Lyman-α radiation illuminates this region. Since Lyman-α suffers little atmospheric attenuation above 100 km, this ionization source will not vary appreciably with solar zenith angle throughout the day in the upper D-region and adds a constant term to the absorption of radio waves penetrating the region. As the solar zenith angle decreases towards noon, X-rays and Lyman-β radiation will soon dominate the ionization in the 90-100 km range and the electron density will increase rapidly. There is, however, a second factor that will compensate for this rapid increase. Waves reflected from the E-region will see a lowering of their reflection level as the electron density increases, and the integral $\int N_e \nu ds$ will not increase as rapidly as predicted by simple theory for waves penetrating the E-layer. Most stations using the A1 or A3 techniques depend upon observations of waves reflected from the E-region during daytime.

The modelling referred to above used an effective loss rate which was constant in time and neglected the divergence term in the continuity equation. Some observations indicate that one or both of these assumptions may not hold. Lastovicka (1978) have shown that the absorption of radio waves in middle latitudes show a diurnal asymmetry, with a maximum after local noon. This asymmetry has a marked seasonal variation. Figure 8 illustrates this result. The possibility that transport of minor constituents, such as NO, could cause such variations should be considered, but so far no adequate model including dynamics has been suggested.

It is interesting to note that the tentative explanation, discussed above, of the observed diurnal change in absorption, may also explain the change with latitude of the diurnal exponent n. At high latitudes the solar zenith angle always remains large, and the rapid development of the E-layer at small zenith angles never occurs. The exponent n therefore remains small, as in the morning and sunrise situations as low and middle latitudes. At high latitudes there is also a contribution to the ion production from precipitation of energetic particles. This contribution is of course not linked to the solar elevation, and adds a variable term to the total absorption. In the monthly averages of absorption at a given solar zenith angle, this term will represent a constant addition to the mean.

5.2 The sunspot cycle variation of absorption

The dependence of radio wave absorption on solar activity is firmly establish from long time series of measurements at several locations (Appleton and Piggott 1954, Lauter et al 1979, Gnanalingam 1974). It is also clear that there is a significant increase in Solar UV and X-ray intensity with sunspot activity, at wavelengths that will ionize in the lower ionosphere. No really adequate modelling from first principles, of the solar cycle dependence of absorption has been done.

The International Reference Ionosphere (IRI, Rawer 1977) provides a set of electron density distributions for the lower ionosphere based upon rocket and ground based data. Recently Singer et al (1979) have calculated absorption for radio waves reflected from these model N_e distributions, and compared the calculations with observations. They find significant discrepancies, in that the IRI 77 gives larger solar cycle variations in absorption than observed.

5.3 The seasonal variation of absorption at low latitudes

From a physical point of view the equatorial lower ionosphere seems relatively simple. Solar X-rays and UV radiation are the dominant ionizing agents, and Oyinloye (1978) has found a good correlation between absorption and solar X-ray fluxes. The seasonal changes in absorption follows the law $L \alpha \cos^n x$, where $n \approx 1.7$, in contrast to value $n \approx 1.0$ for the diurnal variation. This seasonal "anomaly" can apparantly be explained by changes in solar flux and zenith angle. Oyinloye (1978) concludes that there is no evidence to suggest that the absorption variations are associated with dynamical phenomena, such as mesospheric circulation or ionosphere-stratosphere coupling.

In his analysis Oyinloye used data from Colombo and Ibadan, both near 7°N with magnetic dip angles near 5°. Further away from the equator, at dip angles of between $20^\circ$and 30°, Figure 3 shows a maximum in the latitudinal variation of absorption, and a semiannual variation of the intensity of this maximum. The form of this latitudinal variation is similar to the equatorial anomaly in the F-region. Furthermore, as mentioned earlier, the absorption seems to be geomagnetically controlled. The physcial mechanisms causing the observed variations are not established. Pradhan & Shrike (1978) suggest that downward transport of NO, produced in the F-region, is responsible for the enhanced electron densities in the lower ionosphere "equatorial anomaly". Dyson and Bennett (1980) have pointed out that George (1971) may have overestimated the $\int N_e \nu ds$ at low latitudes due to neglect of a term in the magneto-ionic formula for the refractive index. Thus the "anomaly" may be smaller than shown by George, or nonexistent. However, rocket measurements by Mechtly et al (1969) seem to confirm the existence of an anomaly in electron density. No definite conclusion can be drawn at present on this point.

5.4 The absorption at middle latitudes

In the latitude region from 35° to 60° the lower ionosphere, and the radio wave absorption, exhibit much greater variability than at lower latitudes. Our present picture of the physics of this region is very complex and probably far from complete, but some important mechanisms causing enhanced ionization and absorption will be discussed below.

5.4.1 The "Storm after effect"

The "Storm after effect" is a period of enhanced absorption occuring a few days after the main phase of a geomagnetic storm (Belrose & Thomas (1968)). Dickinson and Bennett (1978) have used rocket borne experiments to show that the absorption is due to enhanced D-region electron densities. Larsen et al (1976) have shown by means of satellite observations that energetic electrons (10-200 keV) precipitate at subauroral latitudes during such conditions. These particles will penetrate into the D-region. Spjeldvik and Thorne (1975) have produced a simplified theoretical model for the D-region under past storm conditions, and their estimates agree well with Dickinson and Bennett's measurements.

5.4.2 The winter anomaly in absorption

The winter anomaly in absorption is illustrated in Figure 4 which shows that large values of absorption are observed on days or groups of days in winter at middle latitudes. Although the phenomenon has been known since the early days of ionospheric physics, real progress in our understanding has only been made in recent years. Beynon and Jones (1965) and Shapley and Beynon (1965) were the first to show that ionospheric absorption could be associated with dynamical phenomena in the neutral atmosphere. Figure 9 shows the similarity between the seasonal variation in absorption, mesospheric densities and stratospheric temperature. Shapley and Beynon (1965) demonstrated a connection between absorption and stratospheric warnings. Sechrist (1967) suggested that enhanced densities of nitric oxide could occur in winter, in connection with "D-region warmings", and could cause large absorption through increased ion production.

A major effort was made to study the phenomenon in the Western Europe D-region Winter Anomaly Campaign 1975/76. The results are presented in a separate issue of J Atm Terr Phys 41 Oct/Nov 1979. The campaign has also been described by Offermann (1977) and (1979). In summary the winter anomaly in radio wave absorption is caused by an increase in electron density in the 70-95 km height region (Friedrich et al 1979). This increased electron density is due to several processes: (Thrane et al 1979)

i) An enhanced ion production due to increased NO densities

ii) A decrease in effective electron loss rate associated with a temperature increase and a lowering of the region where complex cluster ions with large recombination rates occur

iii) A possible increase in ion production due to increased density of the excited species $O_2(^1\Delta g)$.

The processes are listed in probable order of importance (Thomas 1979). If increased nitric oxide densities in the D-region is indeed the major cause for the winter anomaly, a plausible mechanism must be found to bring NO into and out of the D-region with time scales of the order of one day, since this is a typical time scale for the absorption changes. We know that NO is produced in the E-region, and that the species has a photochemical lifetime of many days in this height region. Offermann et al (1980) have recently suggested that, in a winter situation, atmospheric waves may propagate from below into the mesosphere and create enhanced turbulence and temperature. The turbulence can rapidly transport NO from above into the D-region, where changing winds can blow the NO away from or towards the observer. This suggestion needs of course more precise experimental verification, but is a plausible hypothesis.

5.5 Sudden Ionospheric Disturbances (SID)

X-rays emitted from solar flares cause increased D-region ionization over the entire sunlit hemisphere. The excess absorption caused by these events is observed as short wave fade-outs (SWF) and can lead to break-down of HF-communications for previous from ½-2 hours. Detailed knowledge of the incident X-ray spectrum makes it possible to estimate ion production rates, but estimates of the resulting electron densities are complicated by a decrease in loss rate, associated with the intense radiation. (Desphande and Mitra 1972).

5.6 The absorption in high latitudes

The absorption in latitudes beyond 60° is most likely influenced by the meteorological processes described in the preceding section. The absorption measurements from Kjeller (60°N) (see Figure 6) certainly show a well developed winter anomaly (Armstrong, Lied & Thrane 1972). However, with increasing latitude beyond 60° such effects are masked by the strong influence of precipitating energetic particles. These penetrate into the lower ionosphere, creating enhanced ionization densities which sometimes completely disrupt HF communications and seriously affect VLF and LF communication systems. Detailed discussions of the different types of disturbances in the high latitude ionosphere have been given in Thrane (1979) and Thrane et al (1980) and references therein. In the present paper emphasis will be on a few points of interest for understanding the physics of high latitude absorption.

We normally distinguish between three important types of high latitude absorption events, Polar Cap Absorption (PCA), Auroral Absorption (AA) and Realistic Electron Precipitation events (REP).

A PCA is caused by energetic solar protons impinging on the polar atmosphere, causing long-lasting and large changes of plasma density in the lower ionosphere over the entire polar caps, sometimes down to latitudes of 60°. Protons with energies of many tens of MeV may penetrate down to stratospheric heights in extreme cases. The ion production rate in the lower D-region have been observed to increase by four orders of magnitude (the August 1972 event, Reagan 1977). These large production rates also profoundly influence the ion chemistry, preventing the formation of large cluster ions, and thereby decreasing the loss rate. The resulting electron densities are large, causing absorption of 10-20 dB of cosmic radio noise observed by riometers operating near 30 MHz. A strong PCA event may last for many days, and although the particle flux may remain relatively constant, the absorption will be smaller during nighttime. The reason for this diurnal variation is that the free electrons in the lower D-region will attach to molecular oxygen forming negative ions. A sudden increase in absorption may be observed at sunrise when visible light causes photo-detachment of the electrons.

The A2, riometer method has proved a very simple and useful technique for studies of high latitude absorption events. Figure 5 shows a map of occurence probability of auroral absorption in the northern hemisphere, and illustrates a characteristic diurnal variation with a maximum near 70°N in the early morning hours. AA events are caused by precipiation of energetic electrons (energies > 10 keV) which ionize in the D- and lower E-regions. These hard particles are associated with, but not necessarily coincident with the softer particles causing the visible aurora. The ionization intensity and height distribution are very variable. There has been some discussion in the literature about the heigths at which the auroral absorption occurs, and to what extent the observed riometer absorption can be used to indicate the state of the lower ionosphere. Torkar (1977) has collected a set of more than 20 electron density profiles derived from rocket experiments during disturbed ionospheric conditions in Northern Scandinavia. Figure 10 shows these profiles and illustrates the great variability in the observed plasma densities. The electron density at different heights may be plotted against the riometer absorption measured during each rocket flight, and a regression line fitted to the points. Figure 11 shows the results for a height at 85 km. Clearly there is a considerable scatter of the values, but a

correlation between riometer absorption and electron density is evident as expected. The slope of the regression line is steepest and the scatter smallest near 85 km. This indicates that most of the absorption occurs between 80 and 90 km, for AA events of the magnitude considered here. Figure 12 shows height distributions of absorption for five rocket flights made during different degrees of disturbance. In all cases the absorption is largest between 70 and 100 km. The code names for the five rockets will be found in Figure 11, which shows the associated riometer absorption. The largest event, during the S-18-2 flight has maximum absorption at the lowest height. The measured particle energy spectrum and the derived ion production is shown in Figure 13 for the rocket S-18-2 (Bjørn et al 1979). It should be noted that the rockets code named F-33 and F-34 were launched during an REP event and a very hard component in the energy spectrum of the electrons created subsidiary maxima in the 60-70 km height range. We conclude that during AA and REP events with riometer absorption less than 3 dB the main absorbing region is above 75 km. This means that auroral absorption should be frequency dependent in the HF range (see Section 2.1).

Attempts have been made to include high latitude absorption into HF propagation prediction methods. Three such methods, the HFM loss (CCIR Report 252-2), the "Bludeck" (Haydon et al 1976), and the "Ioncap" (Lloyd et al 1978) introduce a term, the "excess system loss". This term has a latitudinal, seasonal and diurnal variation, but no frequency dependence or dependence upon magnetic activity. It is claimed that high latitude absorption occurs in the lowest part of the D-region (50-70 km, Lloyd et al 1978), contrary to the results quoted above. Foppiano (1975) proposed a new method for predicting auroral absorption in which the absorption zone is centered at the auroral oval and its intensity varies with frequency and the Kp index. This method has been included in CCIR recommendations and in the Applab prediction method (Bradley 1975). A further development of the Foppiano model has been discussed by Vondrak et al (1977). Thrane and Bradley (1980) have compared the different prediction methods with measurements of circuit reliability on high latitude paths. The find that none of the current methods are adequate for transmission paths within the disturbed auroral zone.

The physical processes causing aurora and auroral absorption are complex and a detailed treatment beyond the scope of this paper. A recent, interesting result should, however, be mentioned. Oksman and Ranta (1980 private communication) have studied the relation between the IMF (Interplanetary Magnetic Field) sector polarity and the diurnal and annual variation of riometer absorption in the auroral zone. They find significant differences in the absorption measured during A (Away) and T (Toward) polarities. The A periods have the smallest absorption in the spring and the greatest absorption in the autumn. This pattern follows the pattern in geomagnetic activity, and is a consequence of high transfer rate of energy from the solar wind into the magnetosphere during conditions favourable for magnetic coupling between the solar and terrestrial fields (Russel and McPherron, 1975).

6. CONCLUSIONS

Our understanding of the physics of radio wave absorption in the ionosphere is linked to our general understanding of upper atmosphere physics, and the radio wave absorption is a sensitive indicator of the state of the lower ionosphere. A wide range of physical problems have been mentioned, such as electron collision processes in atmosphere gases, magneto-ionic theory, the meteorology of the mesosphere and stratosphere, and the link from the solar wind and interplanetary field to the magnetosphere and lower ionosphere. Even though our knowledge has increased rapidly in later years, physical models are in most cases not sufficiently well developed to be useful for prediction purposes (Thrane et al 1979). Radio wave frequency predictions still depend upon empirical formula for the attenuation of the waves. Clearly further research is needed.

REFERENCES

Aggarwal, K.M., Narinder Nath and Setty, CSGK (1979). Collision frequencies and transport properties of electrons in the ionosphere. Planet Space Sci 27, 753-768.

Appleton, E.V. (1937). Regularities and irregularities in the ionosphere. Proc Roy Soc 162, 451.

Appleton, E., W R Piggott (1954). Ionospheric absorption measurements during a sunspot cycle. J Atm Terr Phys 5, 141-172.

Appleton, E.V., J A Ratcliffe (1930). Some simultaneous observations of downcoming wireless waves. Proc Roy Soc (A) 128, 133.

Armstrong, R.J., F Lied and E V Thrane (1970). A possible explanation of the winter anomaly in ionospheric absorption at Kjeller. Physica Norvegica 4, 157-160.

Belrose, J.S. and L Thomas (1968). Ionization changes in the middle latitude D-region associated with geomagnetic storms. J Atmos T P 30, pp 1397-1413.

Beynon, W.J.G. and E S O Jones (1965). Meteorological Influences in Ionospheric Absorption Measurements. Proc Roy Soc 288, 558.

Bjørn, L.G., F Arnold, D Krankowsky, B Grandal, O Hagen and E V Thrane (1979). Lower ionosphere ionproduction, density and composition in an auroral absorption event. J Atmos Terr Phys 41, 1184-1194.

CIRA (1972). Cospar Reference Atmosphere Akademic-Verlag, Berlin.

Desphande, S.P. and A P Mitra (1972). Ionospheric effects of solar flares III. The quantitative relationship of flare X-rays to SID's. J Atmos Terr Phys 34, 243-253.

Dickinson, P.H.G. and F D G Bennett (1978). Diurnal variations in the D-region during a storm after-effect. J Atmos Terr Phys 40, 549-558.

Dyson, P.L. and J A Bennett (in press). General formulae for absorption of radio waves in the ionosphere. J Atmos Terr Phys

Foppiano, A.J. (1975). A new method for predicting the auroral absorption of HF sky waves. CCIR Interim Working Party Document 3.

Friedrich, M., K M Torkar, K Spenner, G Rose and H U Widdel (1979). Electron densities during winter anomalous conditions of different intensity. J Atmos Terr Phys 41, 1121.

Friedrich, M. and K M Torkar (1980). Derived D-region collision frequencies. INW 8011 Technische Universität Graz.

George, P.L. (1971). The global morphology of the quantity $\int N_e \nu . dh$ in the D- and E-regions of the ionosphere. J Atmos Terr Phys 33, pp 1893-1906.

Gnanalingam, S. (1974). Equatorial ionospheric absorption during half a solar cycle. J Atm Terr Phys 36, 1335-1354.

Hake, R.D. and A V Phelps (1967). Momentum-Transfer and Inelastic - Collision Cross Sections for Electrons in O_2, CO and CO_2. Phys Rev 158, 71-84.

Hargreaves, J.K. (1969). Auroral absorption of HF radio waves in the ionosphere: A review of results from the first decade of riometry. Proc IEEE 57, 1348.

Haydon, G.W., M Leftin and R K Rosich (1976). Predicting the performance of high frequency skywave telecommunication systems. OT Report 76-102 Boulder Colo, USA.

Larsen, T.R., J B Reagan, W L Imhof, L E Mountbriand and J S Belrose. A coordinated study of energetic electron precipitation and D-region densities over Ottawa during disturbed conditions. J Geophys Res 81, 2200-2212.

Lastovicka, J. (1977). Seasonal variation in the asymmetry of diurnal variation. J Atm Terr Phys 39, 891-894.

Lauter, E.A. and P Triska (1965). Annual pattern of ionospheric absorption in the medium and long-wave ranges. Geomagn Aeron 5, 822-825.

Lloyd, J.L., G W Haydon, D L Lucas and L R Teters (1978). Estimating the performance of telecommunication systems using the ionospheric transmission channel. Institute for Telecommunication Sciences Boulder, Colo USA.

Mechtly, E.A., Rao, Skaperdas and L G Smith (1969). Latitudinal variation of the lower ionosphere. Radio Sci 4, 517-520.

Mechtly, E.A. (1974). Accuracy of rocket measurements of lower ionosphere electron concentration. Radio Sci 19, 373-378.

Mentzoni, M.H. (1965). J Res Natl Bur Std 69D, 213.

Offermann, D. (1977). Some results from the European Winter Anomaly Campaign 1975-76. In: Dynamical and chemical coupling of neutral and ionized atmosphere. Ed B Grandal J A Holtet. Reidel Publ Co Dordrecht/Holland, pp 235-252.

Offermann, D. (1979). An integrated GBR campaign for the study of the D-region winter anomaly in Western Europe 1975/76. J Atm Terr Phys 41, 1047-1050.

Offermann, D., P Curtis, J M Cisneros, J Satrustegui, H Lauche, G Rose and K Petzholdt. (1979). Atmospheric temperature structure during the Western Europe Winter Anomaly Campaign 1975/76. J Atm Terr Phys 41, 1051-1062.

Offermann, D., H G Bückelman, K M Torkar and H U Widdel. (1980). Mesosphere, stratosphere and D-region Winter Anomaly. Space Research. (In press).

Oyinloye, J.O. (1978). On the seasonal variation of absorption of radio waves in the equatorial ionosphere. J Atmos Terr Phys 40, 793-798.

Phelps, A.V. and J L Pack. (1959). Phys Rev Letters 3, 340.

Piggott, W.R. (1953). The reflection and absorption of radio waves in the ionosphere. Proc IEE 100 Part III 61-72.

Pradhan, S.N. and J S Shirke. (1978). Geomagnetic control of mesospheric nitric oxide concentrations from simultaneous D- and F-region ionization measurements. Ann Geophys 34, 125-131.

Rawer, K. (1977). Mid- and high latitude reference ionosphere. In: Dynamical and Chemical Coupling of Neutral and Ionized Atmosphere. Eds B Grandal and J Holtet. Reidel Publ Co Holland pp 145-160.

Reagan, J.B. (1977). Ionization processes. In: Dynamical and Chemical Coupling of Neutral and Ionized Atmosphere. Eds: B Grandal and J A Holtet. Reidel Publ Co Holland pp 145-160.

Russel, C.T. and R L McPherron. (1975). Semiannual variation of geomagnetic activity. J Geophys Res 78, 92.

Sechrist Jr, C.F. (1967). A theory of the winter absorption anomaly at middle latitudes. Journ Atm Terr Phys 29, 113-136.

Sen, H.K. and A A Wyller. (1960). On the Generalization of the Appleton-Hartree Magneto-ionic Formulas. J Geophys Res 65, 3931.

Shapley, A.H. and W J G Beynon. (1965). Winter anomaly in Ionospheric Absorption and Stratospheric Warmings. Nature 206, 1242.

Singer, W., J Taubenheim and J Bremer. (1980). A test of IRI lower ionosphere models by comparison with radio propagation data. J Atm Terr Phys 42, 241-248.

Spjeldvik, W.N. and R M Thorne. (1975). J Atmos Terr Phys 37, 1313.

Thomas, L. (1979). A study of the enhanced electron concentrations in the mid-latitude D-region on winter days in terms of the positive ion chemistry. J Geomag Geoelectr 31, 567-583.

Thrane, E.V. (1972). On the diurnal and seasonal variation of the D- and E-regions above Kjeller. In: Ed K Folkestad, Magnetosphere - Ionosphere Interactions. Proceedings of NATO Adv Study Inst Oslo University Press pp 29-38.

Thrane, E.V. (1979). Geophysical disturbance effects on the state of the propagation medium and their predictability. AGARD (NATO) Lecture Series No 99 Aerospace propagation media... Lecture No 8.

Thrane, E.V. et al. (1979). D-region predictions, a working group report. In: Solar-Terrestrial Prediction Proceedings. Ed R F Donnelly, Volume II pp 573-598. US Gvt Printing Office Stock no 003-017-00471-6.

Thrane, E.V. (1968). Collision frequency in the high latitude D-region. In: Ed K Folkestad, Ionospheric Radio Communications, pp 63-72 Plenum Press.

Thrane, E.V. and P A Bradley. (1981). High frequency skywave prediction methods and observational data for high latitude communication circuits. Submitted to IEE Symposium on Antennas and propagation, London.

Thrane, E.V., B Grandal, O Hagen, F Ugletveit, W Bangert, D Beran, M Friedrich, A Loidl, H Schwentek and K M Torkar. (1979). Ion production and effective loss rate in the mesosphere during the Western European Winter Anomaly Campaign 1975/76. J Atmos Terr Phys 41, 1097-1103.

Thrane, E.V. and W R Piggott. (1966). The collision frequency in the E- and D-region of the ionosphere. J Atm Terr Phys 28, 721.

Torkar, K.M. (1977). Die untere Ionosphäre und ihre Bedeutung im Energie haushalt der hohen Atmosphäre. INW 7709 Institut für Nachrichten-technik und Wellenausbreitung, Technische Universität, Graz.

Veatch, G.E., J T Verdeyen and J H Cahn. (1966). Bull Atm Phys Soc I 11, 496.

Vondrak, R.R., G Smith, V E Hatfield, R T Tsunoda, V R Frank and P D Perrault. (1977). Chatanika model of the high latitude ionosphere for application to HF propagation predictions. Stanford Research Institut Report on Contract F19628-77-C-0102.

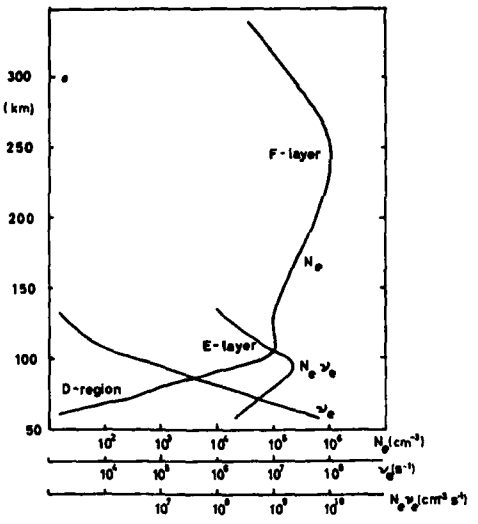


Figure 1 Normal daytime height variations of electron density N_e, collision frequency ν and $N_e\nu$

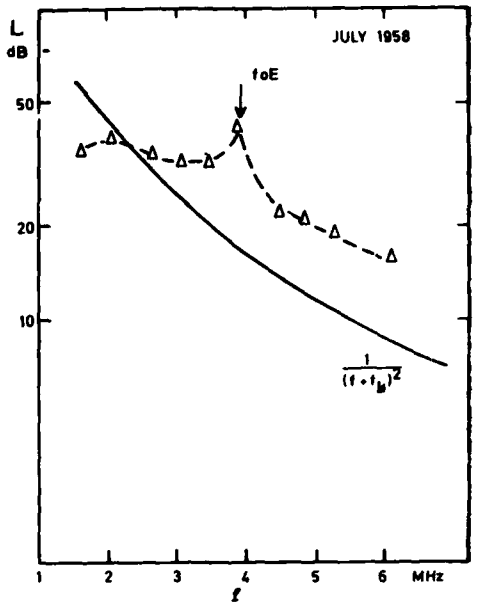


Figure 2 Absorption versus frequency measured by the A1 method at vertical incidence. The triangles indicate the average noon values for July 1958 at Kjeller (Thrane 1972). The solid line show the theoretical frequency dependence for the non-deviative absorption

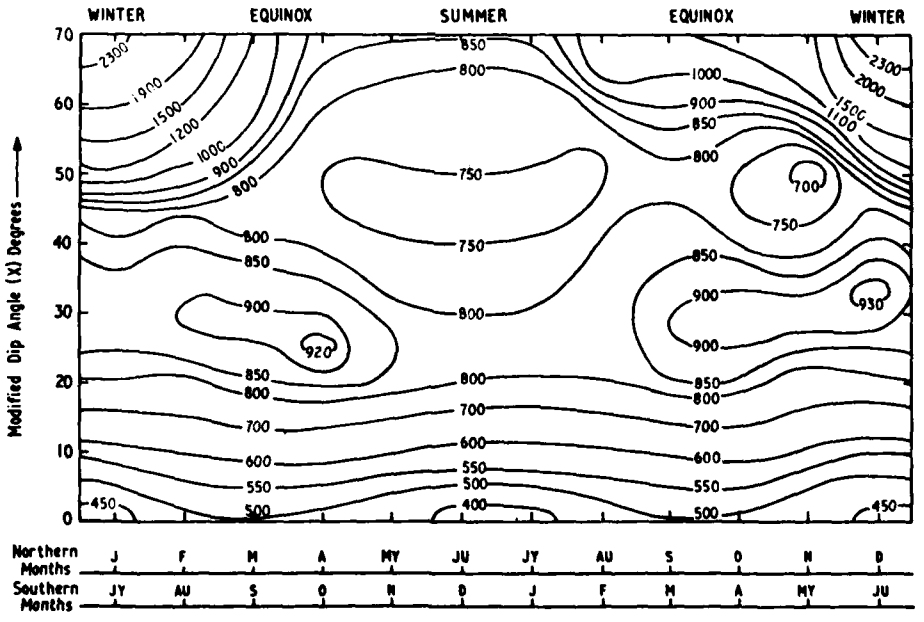


Figure 3 Seasonal and dip-angle variation of the index A_e(T) (George 1971)

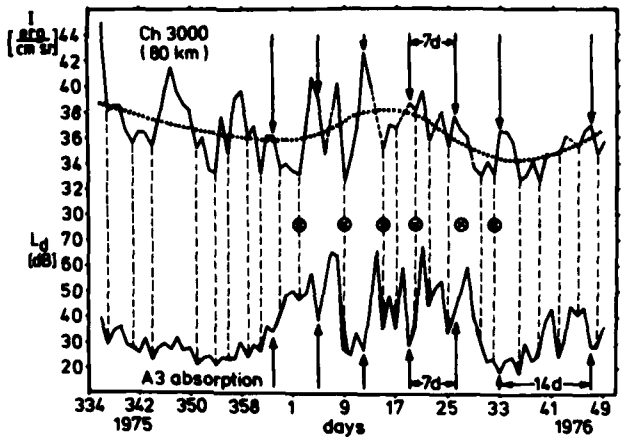


Figure 4 The bottom curve shows typical winter variations of diurnal averages of A3 absorption measured at a European midlatitude station (offermann et al 1979). The top curve shows infrared radiation fluxes measured from a satellite. These fluxes can be taken as measures of atmospheric temperature near 80 km

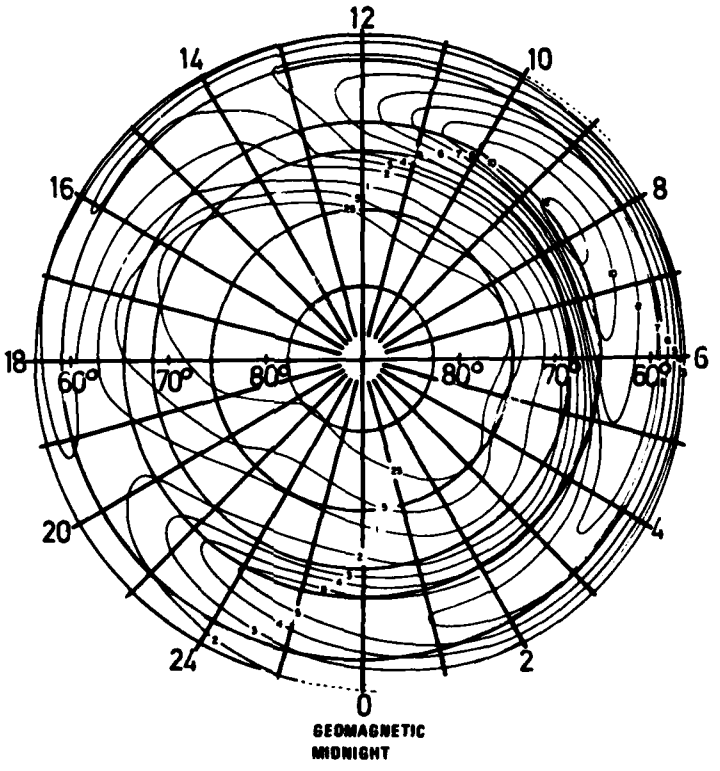


Figure 5 A map in geomagnetic coordinates of the northern polar region, showing percentage probability that the riometer absorption at 30 MHz exceeded 1.0dB. (Harz T R, L E Montbriand, E L Vogan Canad J Phys 41 581 (1963))

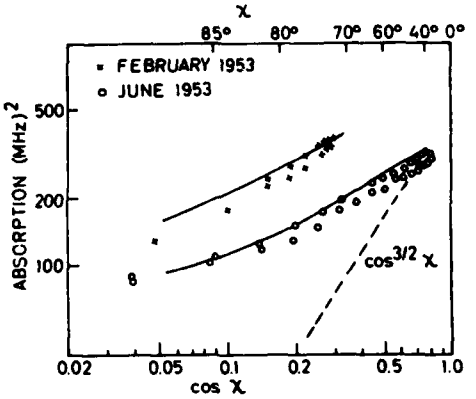


Figure 6 Diurnal variation of the absorption index $A=L(f+f_L)^2$ at Kjeller for a summer and a winter month. The circles are monthly averages of the measurements, the solid lines represent model computations. The dashed line shows the variation expected from a Chapman layer. (Thrane 1972)

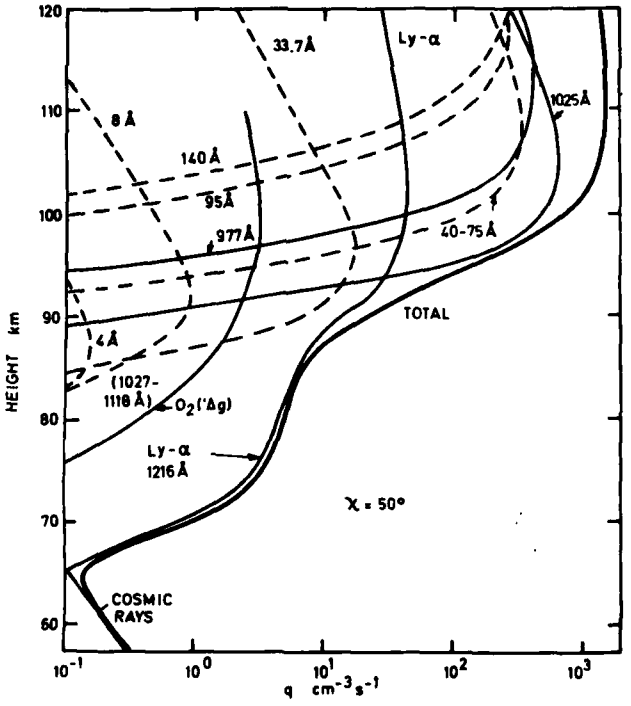


Figure 7 Models of ion production rate versus height for sunspot minimum conditions at two solar zenith angles a) $\chi = 50°$. (Thrane 1972)

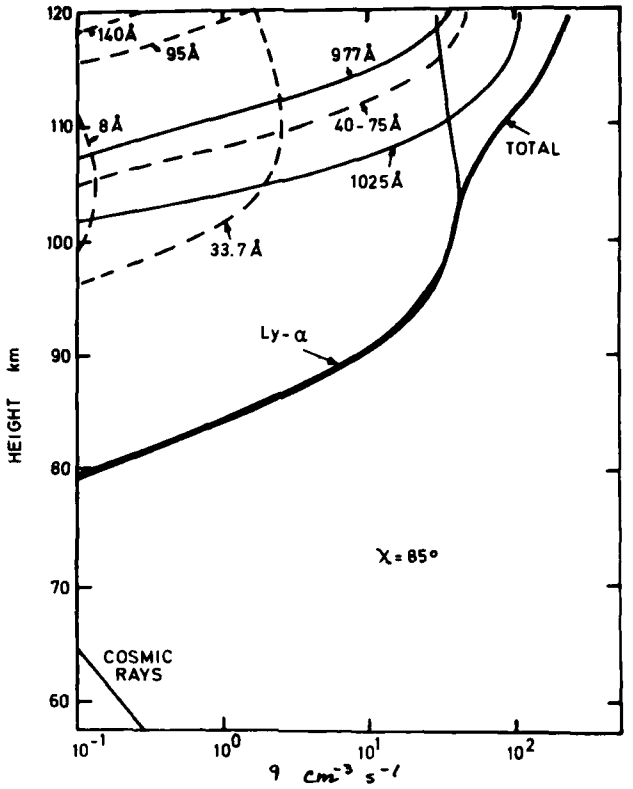


Figure 7 Models of ion production rate versus height for sunspot minimum conditions at two solar zenith angles b) χ = 85°. (Thrane 1972)

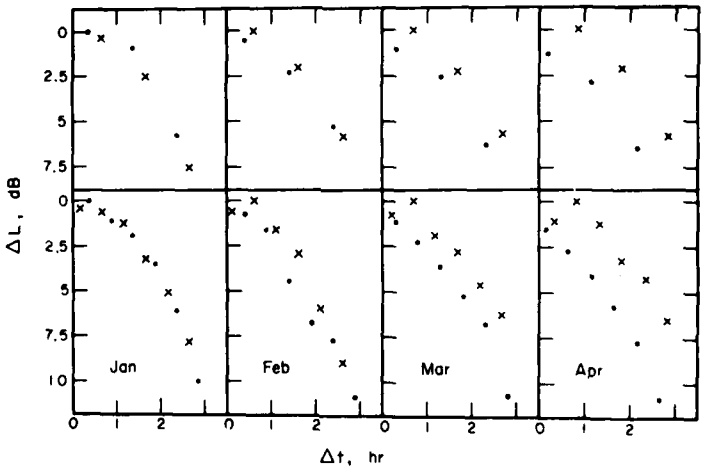


Figure 8 Mean diurnal variation of absorption at mid-latitude stations (Laštovička 1978). Dots represent morning values, crosses afternoon values. Δt represents time deviation from local noon. Upper part hourly mean values 1961-65, lower part half hour mean values 1966-73

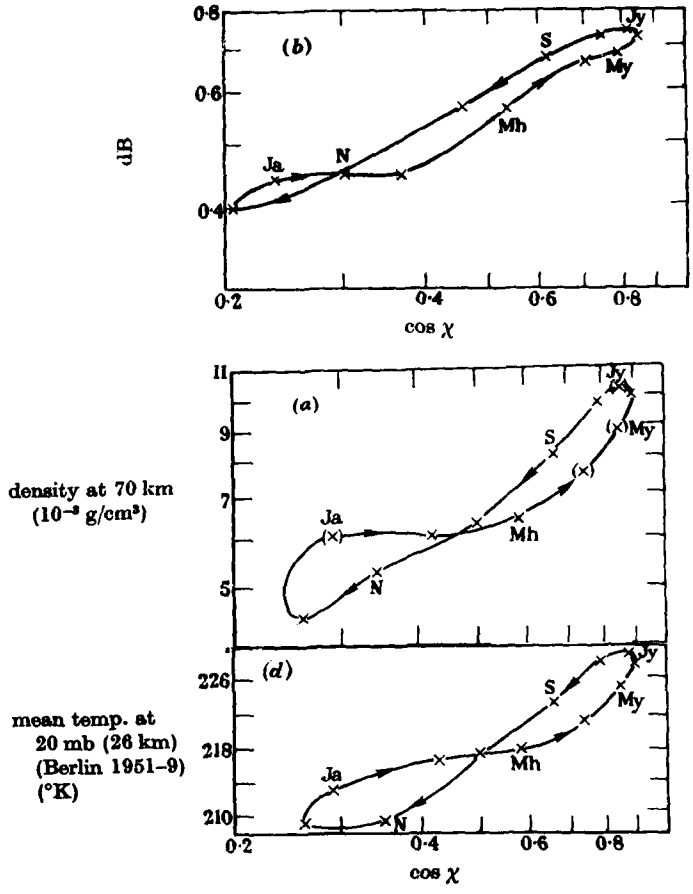


Figure 9 Seasonal variation of
a) Monthly mean absorption at Kuhlungsborn 1959, 245 kHz, b) Density at 70 km, c) Temperature at 26 km (Berlin 1951-59) (Beynon and Jones 1965)

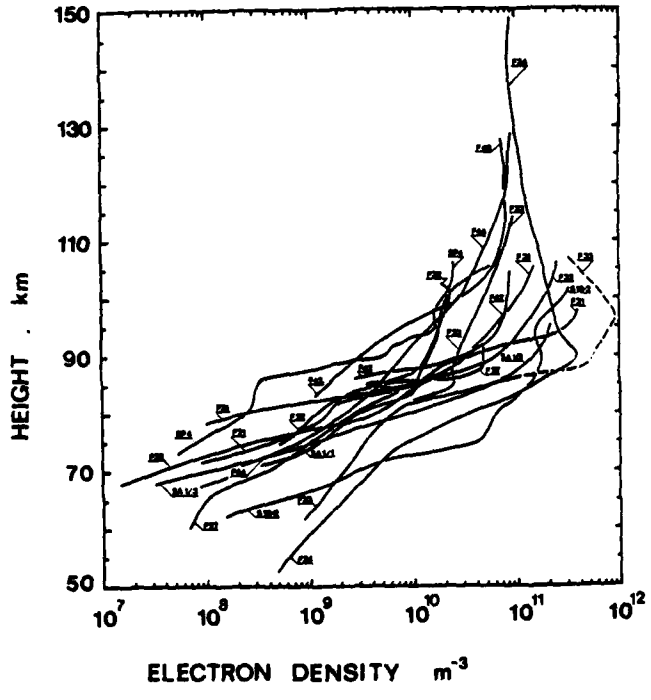


Figure 10 Electron density distributions measured by rocket techniques in the auroral zone (Torkar 1977). The code names for the rockets are given for each curve

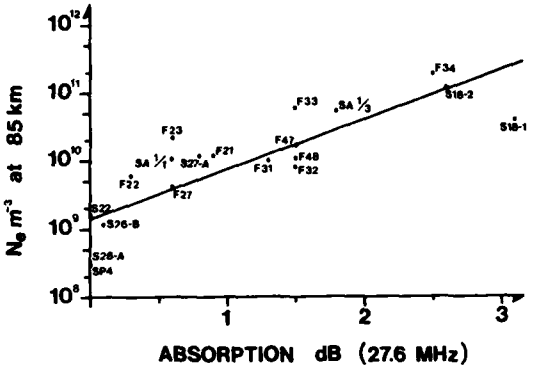


Figure 11 The electron density measured by the rocket techniques at 85 km versus riometer absorption at 27.6 MHz, measured during launch at the rocket range. The regression line is log N_e = 0.73A+ 9.14 where A is the absorption in dB

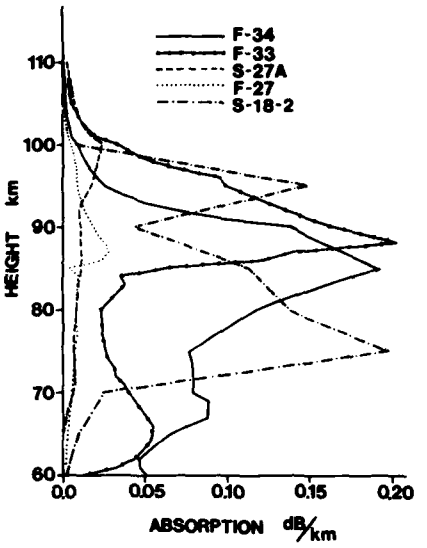


Figure 12 Absorption index in dB/km for five rocket flight during different degrees of disturbances (see Figure 11) (Friedrich, private communication)

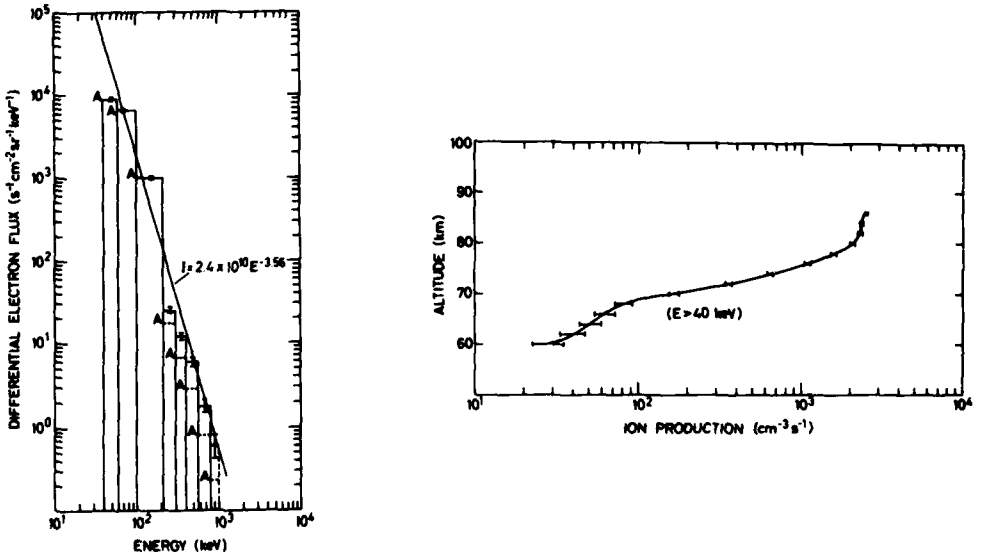


Figure 13 Measured differential electron flux and derived ion production rates during the S-18-2 flight. (Bjørn et al 1979). The riometer absorption was 2.5 dB

SESSION DISCUSSION

SESSION II – SOLAR WIND, MAGNETOSPHERE IONOSPHERE COUPLING

Chairman and Editor – Dr L.G.Smith
Dept. of Electrical Engineering
University of Illinois
Urbana, Illinois 61801, USA

A REVIEW OF SOLAR WIND-MAGNETOSPHERE-IONOSPHERE COUPLING
by W.I.Axford

H.Poeverlein, Ge

To which plasma did you refer when you said it consists, on the average, of 90 percent solar-wind plasma and 10 percent ionospheric plasma?

Author's Reply

I referred only to the hot plasma, well above the energies of a few eV which could result from purely ionospheric effects. Naturally, the total magnetospheric population, including the F-region and plasmaspheric plasma, is dominated by cold particles produced by ionization of the upper atmosphere.

IONOSPHERIC DISTURBANCES OF MAGNETOSPHERIC ORIGIN
by H.Poeverlein and E.Neske

D.G.Torr, US

Can you explain why the light ion trough occurs at lower L shells than does the low altitude midlatitude trough?

Author's Reply

The light ion trough (LIT) and the midlatitude trough (MLT) seem to be different phenomena. In particular, differences exist in the occurrence frequency with respect to local time and the LIT is preferably observed at higher altitudes (see reference in table 1).

P.Spalla, It

Please comment on the behaviour of the total electron content during magnetic storms.

Author's Reply

Ionospheric storms, which accompany geomagnetic storms, show positive and negative phases in the ionospheric electron concentration as well as in the total electron content. This indicates that the changes are not merely a compression or expansion (see Mendillo and Klobuchar, JGR, v. *80*, p. 643, 1975).

H.Soicher, US

Recent observations have indicated that the positive and negative phases of ionospheric storms (i.e. increases and decreases of electron concentration following sudden commencements) are simultaneous in the ionospheric and plasmaspheric (protonospheric) regions. How do you interpret this?

Author's Reply

Since plasma moves freely along magnetic field lines a change of plasma density generated at ionospheric levels spreads out along the field lines into the magnetosphere.

MAGNETOSPHERE-IONOSPHERE COUPLING THROUGH THE AURORAL ACCELERATION REGION
by R.D.Sharp and E.G.Shelley

J.Forbes, US

You suggest that precipitation of hot O^+ ions can explain the increase in ion temperature during an SAR arc. Could these hot ions also contribute significantly to the 6300Å radiation by enhanced charge exchange with O_2?

Author's Reply

Possibly they could. I don't know if the relevant cross sections are available.

M.R.Torr, US

Have you a sufficient amount of data to be able to give even a rough estimate of the variation with magnetic activity of the energy flux of the precipitating fast O^+? For example, could one say that an energy flux of 0.1 ergs cm^{-2} sec^{-1} would be typical for an average storm and perhaps 0.01 ergs cm^{-2} sec^{-1} for quiet conditions?

Author's Reply

The only currently available results that I know of that include quiet time data come from the energetic ion mass spectrometer on the GEOS-1 satellite and have been published by Balsiger et al. (JGR, v. *85*, P. 1645, 1980).

A.Egeland, No

Can you explain the hydrogen emissions, in particular the ground observed Doppler shift, from the upstreaming energetic protons?

Author's Reply

No. The emissions from the upward moving energized H^+ are generally mixed with emissions from precipitating downward moving H^+ ions, and it would be very difficult to sort out the two phenomena from ground based optical observations.

THE PHYSICS OF RADIO WAVE ABSORPTION
by E.V.Thrane

T.B.Jones, UK

Could you comment on the small scale horizontal structure of absorption at high latitudes. Recent observations at Tromso with reflection point separations of $\sim$ 100 km indicate that quite different levels of absorption can occur on these paths.

Authors Reply

There is definitely a small scale structure present, both in the observed absorption and in the ionizing particle flux. The general auroral absorption has spatial scales of the order of 100 km and temporal scales of hours, but fine scales within these patches may be of order kilometers and minutes respectively.

ELECTRODYNAMICS OF THE LOWER IONOSPHERE

Hans Volland
Radioastronomical Institute
University of Bonn
53 Bonn, W.-Germany

Abstract:
The ionospheric E and lower F layers possess electric conductivities sufficiently large to carry electric current systems. This region is called the dynamo region because electric currents are generated there by the interaction between tidal winds and the ionospheric plasma in the presence of the geomagnetic field. The dynamo region also acts as a load resistance in the hydromagnetic generator system "solar wind- magnetosphere- ionosphere", and is an equalizing layer in the global electric current system of thunderstorm origin. The present state of knowledge about electric current systems flowing within the ionospheric dynamo region and their generator mechanisms is reviewed.

1. INTRODUCTION

The Earth's atmosphere is an electrically conducting medium. The main ionizing agencies are cosmic rays within the first 60 km altitude and solar XUV radiation above that height. The mobility of ions and electrons, which is influenced by their collisions with neutrals, and above 80 km by the geomagnetic field, determines the degree of conductivity. The conductivity is isotropic and increases nearly exponentially up to about 80 km (see Fig.1). It becomes anisotropic above 80 km with a conductivity $\sigma_{\parallel}$ parallel to the geomagnetic lines of force $\underline{B}_o$ that increases with altitude. There is also a Pedersen conductivity σ_p orthogonal to $\underline{B}_o$, but parallel to an electric field $\underline{E}$ that maximizes near 140 km, and a Hall conductivity σ_h, which is orthogonal to $\underline{B}_o$ and $\underline{E}$ and maximizes near 110 km altitude. Both the Pedersen and Hall conductivities vary with time, latitude and longitude in a complex manner (e.g., RISHBETH and GARRIOTT, 1969). The region of the lower ionosphere between about 80 km and 200 km where the Pedersen and Hall conductivities are of significant magnitude is called the dynamo region.

An electromagnetic field breaks down very quickly in such a conducting medium if it is not maintained continuously by an external non- electric force. At present, three sources of large scale electromagnetic fields in the atmosphere are known:

(a) Dynamo action between tidal winds and ionospheric plasma
(b) Interaction between solar wind and magnetosphere
(c) Thunderstorms

The electromagnetic field due to source (c) is basically a transverse magnetic field with a strong DC component. The fields of source (a) and (b) are basically transverse electric AC fields with a fundamental period of one solar day.

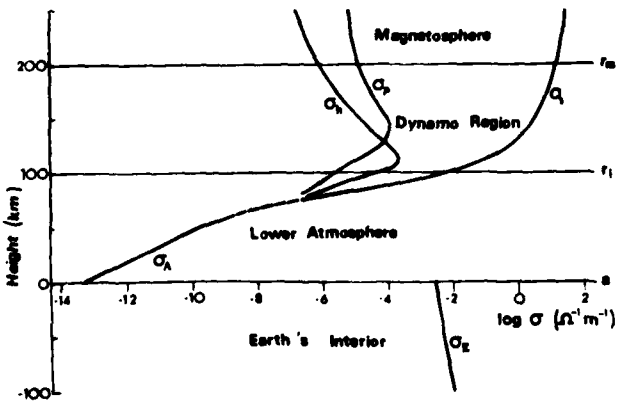


FIG. 1 MEAN ALTITUDE PROFILE OF ELECTRICAL CONDUCTIVITY. IT IS
σ_p = PEDERSEN CONDUCTIVITY;
σ_h = HALL CONDUCTIVITY;
$\sigma_{\parallel}$ = PARALLEL CONDUCTIVITY;
σ_E = CONDUCTIVITY OF THE EARTH'S INTERIOR.

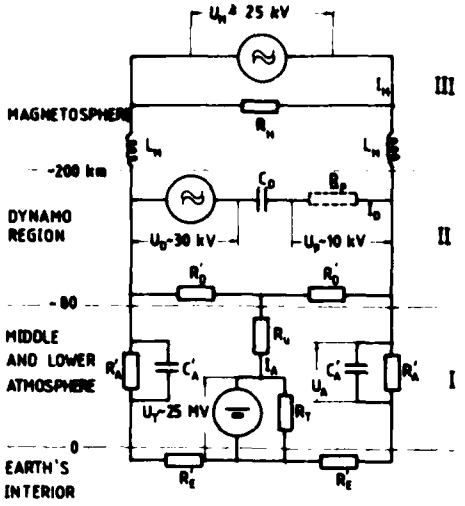


FIG. 2 SCHEMATIC EQUIVALENT CURRENT OF THE SYSTEM "EARTH- ATMOSPHERE- MAGNETOSPHERE".

$R_E = R_E'/2 \leq 1$ mΩ
$R_A = R_A'/2 \approx 250$ Ω
$R_U \approx 25$ kΩ
$R_T \approx 100$ kΩ
$|\underline{R}_P| \approx 0.1$ Ω
$R_D = R_D'/2 \approx 30$ mΩ
$R_M = 0.3$ Ω

$C_A = 2\,C_A' = 0.1$ F
$C_D = 200$ kF
$L_M = L_M'/2 = 30$ H
$I_A = 1$ kA
$I_D = 100$ kA
$I_M \geq 1$ MA
$U_A = 250$ kV

The region of interest here - the dynamo region - is the source region of the dynamo field (a). However, it also acts as a load resistance for currents which are driven by the sources (b) and (c). This is schematically shown in Fig. 2. U_D is the dynamo source driving a horizontal electric current of the order of 100 kA on one hemisphere within system II. The resistance R_D in that system is also load resistance in system III, and it is part of the global current system I.

In this paper, I want to review the generation of the dynamo current, its structure and its modulation during solar flares and solar eclipses. In addition, the role of the dynamo region as a passive region for the current systems I and III in Fig. 2 will be outlined. Emphasis is placed on electromagnetic fields of lowest periods greater than about one hour.

2. TIDAL WINDS

The agency for the in situ generation of large scale electric currents within the dynamo region are regular tidal wind systems. These tidal winds are excited either by the solar thermal heat source (solar tides) or by the lunar mechanical gravitational force (lunar tides). The Earth's atmosphere behaves like a huge wave guide in which only individual wave modes can exist.

The meridional structure of these wave modes is determined from an eigenvalue equation called Laplace's equation (CHAPMAN and LINDZEN, 1970). The wave modes are symbolized by two wave numbers (m,n) where $m = 0, 1, 2...$ is a zonal wave number indicating the number of wavelengths spanning a zonal circle. The diurnal tides with a period of one solar (lunar) day have wave number $m = 1$, the semidiurnal tides have $m = 2$, etc. n is a meridional wave number ($|n| \geq m$) related to a hierarchy of wave structures along a meridional circle. Positive n belong to waves with finite vertical wavelengths (propagating waves). Negative n belong to waves with infinite vertical wavelengths (evanescent waves).

The vertical structure of the waves is connected to their horizontal structure by a separation constant called the equivalent depth h_n^m. Positive h belong to propagating waves, negative h belong to evanescent waves. The fundamental solar diurnal tide (1,-2) is an evanescent wave within the lower and middle atmosphere. The vertical structure of forced waves is proportional to the corresponding source function (e.g., VOLLAND,1979a).

3. IONOSPHERIC DYNAMO

The basic equation of the dynamo theory is Ohm's law

$$\underline{j} = \underline{\sigma} \cdot (\underline{E} + \underline{U} \times \underline{B}_0) \qquad (1)$$

together with a condition for source free electric currents (e.g. CHAPMAN and BARTELS, 1951; FEJER, 1964). $\underline{j}$ is the electric current density; $\underline{B}_o$ is the geomagnetic field. $\underline{E}$ is an electric polarization field which is caused by charge separation between ions and electrons. It adjusts the current configuration in such a manner that $\underline{j}$ becomes source free.
$\underline{\sigma}$ is a conductivity tensor which is at middle and higher latitudes

$$\underline{\sigma} = \begin{pmatrix} \sigma_p & \sigma_h \\ -\sigma_h & \sigma_p \end{pmatrix} \tag{2}$$

with σ_p the Pedersen conductivity and σ_h the Hall conductivity (see Fig. 1). Near the geomagnetic equator, it is

$$\underline{\sigma} = \begin{pmatrix} \sigma_{\parallel} & 0 \\ 0 & \sigma_c \end{pmatrix} \qquad ; \sigma_c = \sigma_p + \sigma_h^2/\sigma_p \tag{3}$$

with $\sigma_{\parallel}$ the parallel conductivity and σ_c the Cowling conductivity (MAEDA and MATSUMOTO, 1962).

Since the conductivities σ_p and σ_h are sufficiently large only within a narrow sheet (see Fig. 1), the current density $\underline{j}$ flows essentially in that sheet, and vertical currents can be ignored in a first approximation (apart from the geomagnetic equatorial region). Introducing height integrated averaged currents and winds:

$$\tilde{\underline{j}} = \int_{z_1}^{z_2} \underline{j}\, dz \quad ; \tilde{\Sigma}_i \tilde{\underline{U}} = \int_{z_1}^{z_2} \sigma_i \underline{U}\, dz \quad ; (i = p,h) \tag{4}$$

it is evident that only those wind systems contribute significantly to the height integrated currents $\tilde{\underline{j}}$ that have vertical wavelengths large compared with the thickness of the dynamo layer. The evanescent modes with the same phase at all altitudes (i.e., with infinitely large vertical wavelengths) contribute predominantly to the dynamo currents. In particular, the solar diurnal (1,-2) mode is mainly responsible for the Sq current (KATO, 1966; STENNING, 1969; TARPLEY, 1970b). The Sq current can be observed magnetically on the ground as the solar quiet (Sq) geomagnetic variation. The main contribution to the lunar (L) variation comes from the lunar semidiurnal (2,2) mode (TARPLEY, 1970a).

Separating the height integrated Lorentz field $\underline{U}\times\underline{B}_o$ in (1) into a curl free and a source free term W, introducing a quasistatic electric potential for the secondary electric field, and using the simplified form of the conductivity tensor (2) and (4), one arrives at a current (MÖHLMANN, 1974)

$$\tilde{\underline{j}} = \tilde{\Sigma}_c \nabla \times W \underline{r} \tag{5}$$

with $\tilde{\Sigma}_c = \tilde{\Sigma}_p + \tilde{\Sigma}_h^2/\tilde{\Sigma}_p$ an average Cowling conductivity.
In this approximation, the curl free part of the Lorentz field does not contribute to the current, and the current $\tilde{\underline{j}}$ can be derived from a stream function

$$\Psi = \tilde{\Sigma}_c W \tag{6}$$

The Cowling conductivity is the effective conductivity. The lines Ψ = const are streamlines of the current system.

4. STREAMFUNCTION AND ELECTRIC FIELD OF THE Sq CURRENT

The earth's interior is a good electric conductor when compared with the conductivity of the lower atmosphere (see Fig. 1). Therefore, the system "dynamo region- Earth's interior" behaves like a huge transformer with the dynamo region the primary winding and the Earth the secondary winding. Electromagnetic induction in the secondary winding generates secondary electric currents in the Earth, the magnetic fields of which are superimposed onto the magnetic fields of the dynamo current. If one considers the primary

electromagnetic field of the Sq current as a transverse electric AC field reflected on a conducting half plane (the Earth's interior), the effect on the ground is an enhancement of the horizontal component and a reduction of the vertical component of the total magnetic field, while the electric field disappears. This is in basic agreement with more detailed calculations of electromagnetic induction within the spherical Earth (PRICE, 1967).

A plane homogeneous electric sheet current of strength I generates a horizontal magnetic field underneath which is orthogonal to the direction of the current and has the strength

$$B = \mu_o I/2 \qquad (7)$$

Electromagnetic induction within the Earth enhances the horizontal field on the ground by a factor of about 1/3. Therefore, an equivalent overhead current is related to the horizontal intensity observed on the ground as

$$I = \frac{3}{2}\frac{\Delta H}{\mu_o} \qquad (8)$$

More detailed methods allow to separate the contribution of the ionospheric current (the external part) from that of the Earth's interior (the internal part) and to relate the external magnetic variations observed on the ground to a virtual spherical sheet current within the dynamo layer derivable from a stream function Ψ. The lines Ψ = const. constitute the streamlines of that current (CHAPMAN and BARTELS, 1951).

Fig. 3a shows the streamlines of the external component of the Sq current. One notices a strong vortex in each hemisphere peaking near 35° latitude and near 11:00 local time with an overall magnitude of about 150 kA. The internal current (Fig. 3b) which has a similar structure, has a strength about 1/3 of that of the external current. The L current due to lunar tides is about 20 times weaker than the Sq current.

Given the streamlines in Fig. 3a, the electric field $\underline{E}$ and the wind system $\underline{U}$ can be derived from (1) (MAEDA, 1955; KATO, 1956). This is now regularly done in the World Data Center C2 for Geomagnetism, Kyoto University, Japan (SUZUKI, 1978).

The electric field so determined is not unique because this process ignores the irrotational component of the Lorentz field (MÖHLMANN, 1974).

The electric Sq field can also be derived from incoherent scatter

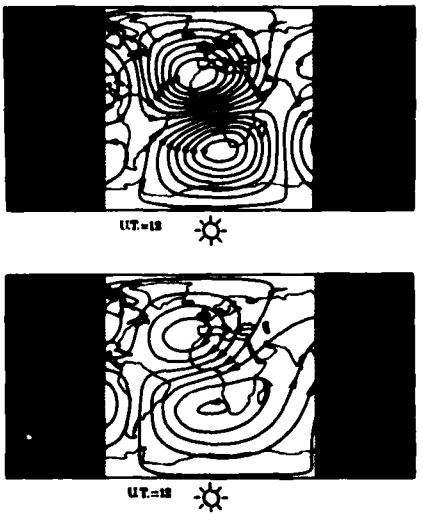


FIG. 3 EXTERNAL (ABOVE) AND INTERNAL (BELOW) PART OF THE EQUIVALENT ELECTRIC Sq CURRENT DURING EQUINOX (1957 - 1960). BETWEEN TWO STREAMS LINES FLOW 20 kA. (after MALIN, 1973).

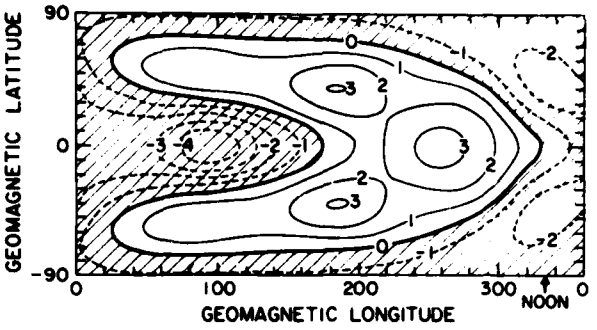


FIG. 4 EQUIPOTENTIAL LINES OF ELECTRIC POTENTIAL OF THE ELECTRICAL DYNAMO FIELD DURING QUIET CONDITIONS (NUMBERS IN kV) (after RICHMOND, 1976).

radar. Fig. 4 from RICHMOND (1976) shows equipotential lines of that field with maximum horizontal potential differences of about 7 kV. HARPER (1977) has determined the electric current density at dynamo layer heights from incoherent scatter observations of $\underline{E}$ and the ion drift velocity.

Theoretical approaches to determine $\underline{j}$ and $\underline{E}$ usually start from calculated tidal wind systems $\underline{U}$ and use these theoretical winds as "driving forces" for the electric fields and currents in (1) preassuming that the conductivity is known (e.g., STENING, 1973; RICHMOND et al., 1976; FORBES and LINDZEN, 1976a, 1977; MÖHLMANN, 1977). The result of these studies is the following:
The diurnal (1,-2) mode contributes about 80 - 90% of the Sq current. The balance of 10 to 20% is due to the diurnal (1,1) mode and the semidiurnal (2,2) and/or (2,4) modes. These propagating modes with their finite vertical wavelengths are mainly responsible for the observed variability in the Sq variations. The bulk of the Pedersen current flows in a broad region between about 120 and 170 km altitude whereas the Hall current is concentrated near 120 km. The contribution of the Hall current is smaller than that of the Pedersen current yielding $\vec{\Sigma}_h/\vec{\Sigma}_p \approx 0.7$ and $\vec{\Sigma}_c \approx 1.5\ \vec{\Sigma}_p$ in (5).

5. INFLUENCE OF CONDUCTIVITY ON THE CURRENT CONFIGURATION

The diurnal (1,-2) wind which is the dominant generator of the Sq current is a harmonic wave with period of one solar day:

$$U \approx U_1 \cos\tau$$

On the other hand, the geomagnetic Sq variation represented by the equivalent sheet current in Fig.3 contains strong harmonics. This can be seen in Fig. 5 which gives the first four Fourier components in magnitude and phase of the horizontal intensity H versus dip latitude (MATSUSHITA,1967). The magnitudes of the two first components have ratios $h_1 : h_2 \approx 2 : 1$, and the phases have the relationship $\gamma_1 \approx \gamma_2 + \pi$. This clearly indicates the influence of the daily variation of the conductivity. Developing the temporal variation of σ into a Fourier series:

$$\sigma = \sigma_o - \sigma_1 \cos\tau + \ldots \qquad (9)$$

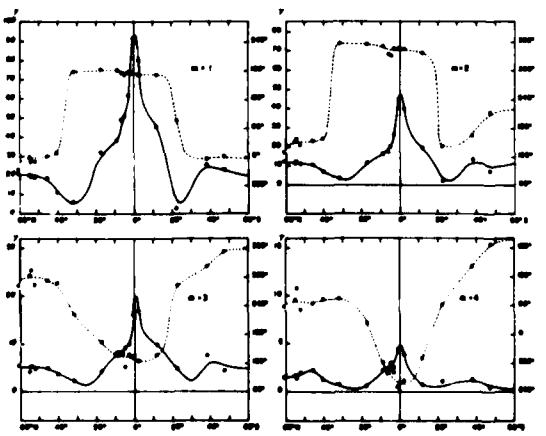


FIG. 5 DIP- LATITUDE DISTRIBUTION OF THE AMPLITUDES h_m (SOLID CIRCLES) AND THE PHASES γ_m (OPEN CIRCLES) OF THE FIRST FOURIER HARMONICS (m = 1 to 4) FOR THE HORIZONTAL INTENSITY IN THE AMERICAN ZONE DURING EQUINOX 1958. THE CURVES REPRESENT SMOOTHED VALUES (after MATSUSHITA, 1967).

one obtains the product σU, which is a measure of the current:

$$qU \approx U_1(-\frac{\sigma_1}{2} + \sigma_o \cos\tau - \frac{\sigma_1}{2}\cos 2\tau + \ldots \qquad (10)$$

Hence, with the reasonable values of $\sigma_o \approx \sigma_1$, the relationship between the two first Fourier components in Fig. 5 is verified.

6. EQUATORIAL ELECTROJET

At the geomagnetic equator, one may apply (3) to determine the height integrated zonal component of the electric conductivity

$$\Sigma_c = \int_{z_1}^{z_2} \sigma_c \, dz > \bar{\Sigma}_c \qquad (11)$$

This leads to a zone of strongly enhanced currents at the height of the maximum of the Hall conductivity (≈ 110 km). That electrojet is observed magnetically as a two fold increase in the horizontal intensity at the geomagnetic dip equator. The effect decreases to zero within about + 5° latitude (Fig. 5).

The physics underlying the derivation of (3) is that the vertical electric currents outside the thin sheet are prevented from flowing. Hence, a vertical electric polarization field is set up which drives a zonal Hall current within the sheet. That Hall current is superimposed on the zonal Pedersen current. Because of the limited meridional range of such polarization field, an assoziated current system flows in the meridional plane on both sides of the equator as indicated in Fig. 6 (UNTIEDT, 1967). That current generates a toroidal zonally directed magnetic field which cannot be observed on the ground. Fig. 7 shows a cross section of the density of the zonally flowing electrojet. Such current profiles have indeed been measured by Rockets (e.g. SHUMAN, 1970).

The influence of various tidal modes as well as the temporal variations of the conductivity is taken into account in more detailed calculations (RICHMOND, 1973a, 1973b; FAMBITAKOYE et al., 1976; FORBES and LINDZEN, 1976b).

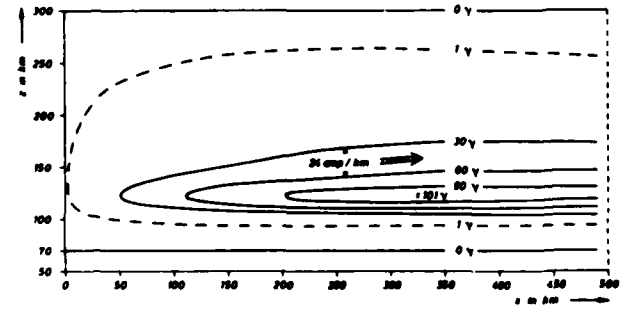


FIG. 6 ISOLINES OF THE TOROIDAL MAGNETIC FIELD OF THE EQUATORIAL ELECTROJET DURING SUNLIT HOURS IN THE MERIDIONAL PLANE. THEY ARE ALSO STREAM LINES THAT FLOW IN THE DIRECTION OF THE ARROW. THE MAGNETIC FIELD IS ANTISYMMETRIC WITH RESPECT TO THE GEOMAGNETIC EQUATOR (x = 0) (after UNTIEDT, 1967).

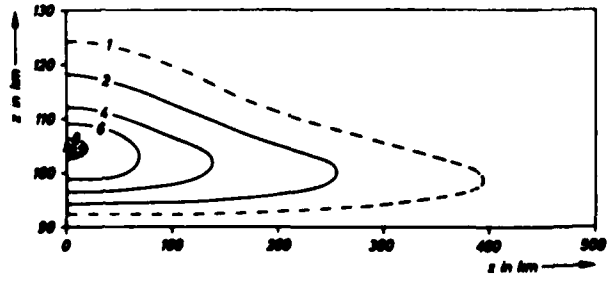


FIG. 7 ISOLINES OF THE ZONALLY FLOWING EQUATORIAL ELECTROJET CURRENT DENSITY IN THE MERIDIONAL PLANE - IN UNITS OF 10^{-6} A/m^2. x = 0 IS THE GEOMAGNETIC EQUATOR (after UNTIEDT, 1967).

7. GEOMAGNETIC SOLAR FLARE EFFECT AND SOLAR ECLIPSE EFFECT

During a solar flare, enhanced solar XUV radiation increases the electron density within the lower ionosphere on the sunlit hemisphere for a time interval of the order of one hour. Consequently, the electric conductivity in that region increases, and an enlargement of the Sq current is expected. This is observed in the geomagnetic registrations as a small deviation from the undisturbed Sq pattern and is known as crochet or geomagnetic solar flare effect (s.f.e.) (see Fig. 8). The corresponding equivalent electric current is indeed very similar in structure to that of the Sq current. However, its vortex centers are shifted to the west by about 10^{o} (VAN SABBEN, 1961). The most likely explanation for this is that the s.f.e.-current is centered at a height which is somewhat below the mean height of the Sq current. The tidal winds which shift their phase with altitude are then responsible for different locations of the vortices. More recent papers dealing with this subject are RIKITAKE and YUKUTAKE (1962), OHSHIO et al. (1963), and RICHMOND and VENKATESWARAN (1971).

During a solar eclipse, the shadow of the moon generates a conductivity hole within the dynamo region of about 6000 km in diameter which moves along the central line of the eclipse. One expects therefore a reduction of the strength of the Sq current in the environment of the central line during the eclipse. This has been observed on very quiet days (see Fig. 9) (WAGNER, 1963; BOMKE et al.,1967). The equivalent current system superimposed onto the Sq current which generates ΔX_e is shown in Fig. 10. It is directed opposite to the Sq current and moves with the shadow along the central line.

8. COUPLING BETWEEN THE IONOSPHERIC PLASMA AND THE TIDAL WIND

The tidal wind in (1) is not independent of the electric current. Feedback from the current to the wind occurs via the mechanical Ampere force $\underline{j} \times \underline{B}_o$ in the equation of horizontal momentum.

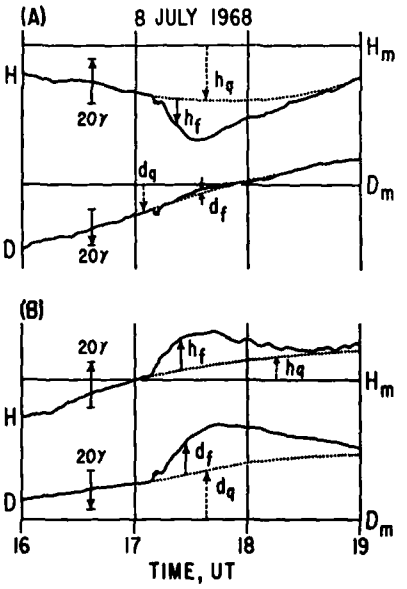


FIG. 8 GEOMAGNETIC CROCHET RECORDED AT BOULDER (A) AND FREDERICKSBURG (B) DURING THE SOLAR FLARE OF JULY 8, 1968, 1708 UT. H_m and D_m ARE THE MEAN MAGNETIC VALUES OF H AND D. h_q AND d_q ARE THE COMPONENTS OF Sq. h_f AND d_f ARE THE CROCHET AMPLITUDES (after RICHMOND and VENKATESWARAN, 1971).

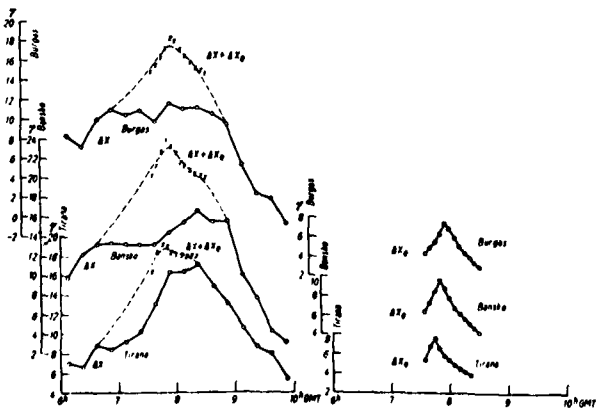


FIG. 9 NORTH COMPONENT (ΔX) OF GEOMAGNETIC VARIATION DURING SOLAR ECLIPSE OF FEBRUARY 15, 1961 (SOLID LINES; LEFT) AND EXTRAPOLATED Sq VARIATION (DASHED LINES; LEFT) AT THREE STATIONS IN SOUTH- EAST EUROPE. THE CURVES ON THE RIGHT GIVE THE SOLAR ECLIPSE EFFECT ΔX_e (after WAGNER, 1963).

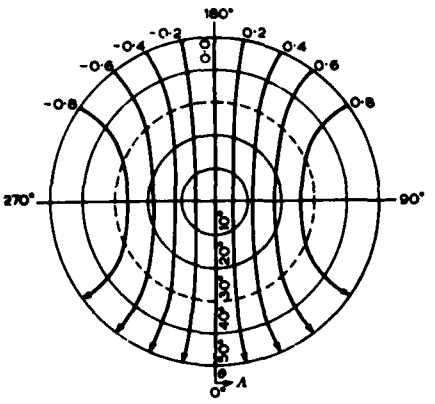


FIG. 10 EQUIVALENT STREAM LINES OF THE GEOMAGNETIC SOLAR ECLIPSE EFFECT FLOWING IN OPPOSITE DIRECTION TO Sq (after VOLLAND, 1956).

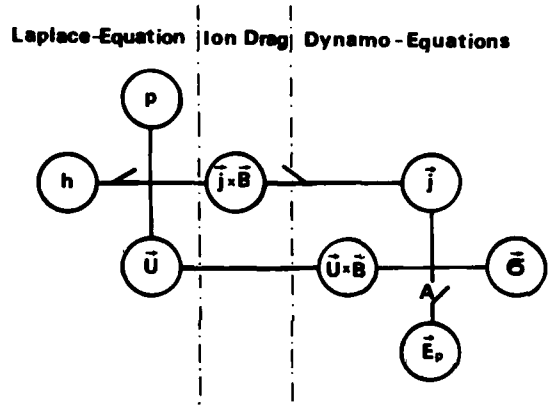


FIG. 11 BLOCK DIAGRAM INDICATING COUPLING BETWEEN HORIZONTAL WIND U, PRESSURE p, ELECTRIC CURRENT j, AND ELECTRIC FIELD E VIA AMPERE FORCE j×B AND ELECTRIC LORENTZ FIELD U×B. B IS THE GEOMAGNETIC FIELD. h IS THE EQUIVALENT HEIGHT, AND σ IS THE ELECTRIC CONDUCTIVITY TENSOR (after VOLLAND, 1976b).

Feedback between wind and current is schematically shown in Fig. 11. Gate B in Fig. 11 is open in the conventional dynamo theories, gates A and C are open in F- layer wind calculations (e.g. KOHL and KING, 1967).

A numerial treatment of the complete set of the hydrodynamic and electrodynamic equations is possible in the case of a simplified conductivity tensor such as in (2) (VOLLAND and GRELLMANN, 1978). Approximate analytic solutions have been found for the diurnal (1,-2) and the semidiurnal (2,2) modes (VOLLAND, 1976a). The solution for the stream function Ψ in (6) for the (1,-2) mode is

$$\Psi = -\frac{3\,\tilde{\Sigma}_c}{1 + 4i\delta/3}\,\sin\phi\,\cos\phi\,\frac{\tilde{U}_o B_o r_o}{5}\,e^{i\tau} \qquad (12)$$

with

$$\delta = \frac{\tilde{\Sigma}_c B_o^2}{\Omega \rho_o\,\Delta r} \quad \text{a feedback factor} \qquad (13)$$

Here $B_o = 3\times10^{-5}$ T is the geomagnetic dipole field at the equator, τ is the local time, r_o is the radial distance of the dynamo layer, and $\hat{U}_o$ is a height averaged wind amplitude. The current strength within one vortex of the Sq current is given by

$$\Psi_{max} = |\Psi|_{\phi = 45^o}$$

With the values

$r_o - a = 130$ km ; ρ_o (130 km) $\approx 8\times10^{-9}$ kg/m^3 ; $\Delta r \approx 5$ km ; $|\tilde{U}_o| \approx 80$ m/s ; $\tilde{\Sigma}_c \approx 30$ S

one arrives at

$$\delta \approx 0.9 \qquad \text{and} \qquad \Psi_{max} \approx 90 \text{ kA}$$

The value of Ψ_{max} would increase to 140 kA with no feedback ($\delta = 0$). Hence, the efficiency of the wind to produce the Sq current is reduced by a factor of about 0.65 due to feedback.

With no feedback ($\delta = 0$; gate B in Fig. 11 open), the wind is considered as an "external driving force". Evidently, the current would increase unlimited with increasing conductivity in that case. With feedback, however, the short circuit current obtains the finite value:

$$\Psi_{max} \Big|_{\tilde{\Sigma}_c \to \infty} = \frac{9\, \Delta r \Omega \rho_o |\tilde{U}_o| r_o}{40\, B_o} = 115 \text{ kA} \qquad (14)$$

Therefore, the observed Sq current reaches about 80% of its short circuit value.

The ionospheric dynamo with feedback between wind and plasma is called a hydromagnetic dynamo. It can be simulated by the equivalent electric circuit II in Fig. 2. After integrating over a meridional circle, the meridional components of the wind and the current in (1) disappear. The parameters in Fig. 2 are the meriodionally averaged zonal components (VOLLAND, 1976b). $\bar{U}_D$ corresponds to the driving force (the solar heat) and is an external source voltage, U corresponds to the Lorentz field $\underline{U} \times \underline{B}_o$, I_D is analogous to the height integrated electric current, U_p is the polarization potential, and

$$R_D = \frac{1}{\tilde{\Sigma}_c} \; ; \; R_i = \frac{1}{i\Omega C_D} \; ; \; C_D = \frac{\rho_o \Delta r}{4B_o^2} \; ; \; \underline{R}_p = -\frac{13}{3} R_D + \frac{1}{9i\Omega C_D} \qquad (15)$$

R_D is the load resistance, R_i an internal resistance which does not depend on the plasma parameters, and $\underline{R}_p$ is a complex impedance relating the polarization potential with the current.

The circuit II in Fig. 2 is analogous to a technical dynamo with $\underline{R}_p$ corresponding to the self inductance of the coil which is space charge in the case of the ionospheric dynamo.

The Sq current produces Joule heating which is transferred to the neutral gas. This Joule heating averaged over the sphere is

$$\frac{1}{4\,\bar{\Sigma}c} \int_0^{\pi} \underline{\tilde{j}} \cdot \underline{\tilde{j}}^* \sin\theta \, d\theta = \frac{36\, \Sigma_c B_o^2 \tilde{U}_o^2}{250\{1+(4\delta/3)^2\}} \approx 10^{-5} \text{ W/m}^2 \qquad (16)$$

with $\underline{\tilde{j}}$ from (5). This is about two orders of magnitude smaller than the solar XUV heating above 120 km altitude. The number in (16) is in basic agreement with more detailed calculations (ROBLE and MATSUSHITA, 1975; FORBES and LINDZEN, 1976a).

9. PENETRATION OF Sq CURRENTS AND FIELDS INTO MAGNETOSPHERE AND LOWER ATMOSPHERE

The electric conductivity is very large within and above the dynamo region along the geometric lines of force. Therefore, one expects field-aligned electric currents to flow during solstice conditions when the tidal winds create potential differences between summer and winter hemisphere. Estimates carried out by MAEDA and MURATA (1965) suggest that not more than about 10% of the Sq current flows through the magnetosphere with field -aligned current densities of the order of 10^{-10} A/m^2 as compared with the horizontal current densities of the order of 10^{-7} A/m^2. The magnetic effect of these field aligned currents on the ground is negligible.

The small but nevertheless finite electric conductivity of the middle and lower atmosphere allows part of the Sq current to flow through those regions. From the ratio between the load resistance R_D of the dynamo layer to the overall resistance R_A of the lower and middle atmosphere (see Fig. 2) one estimates that not more than a factor of 10^{-4} of the Sq current flows through the lower and middle atmosphere.

The electric polarization field of the Sq current can map down into the middle and lower atmosphere. Its vertical component on the ground has been estimated to be of the order of 1 V/m. This is much smaller than the thunderstorm field of about 100 V/m and is therefore barely detectable (ROBLE and HAYS, 1979).

The Sq field can also be communicated upward into the magnetosphere along the geomagnetic field lines without appreciable attenuation. One expects this field to cause an outward drift of the magnetospheric plasma during daytime and an inward flow during the night (MATSUSHITA and TARPLEY, 1970; MAEDA and KAMEI, 1975)

10. THE DYNAMO REGION AS A LOAD RESISTANCE FOR CURRENTS OF MAGNETOSPHERIC ORIGIN

The interaction between the solar wind and the magnetospheric plasma acts like a hydromagnetic generator in which the kinetic energy of the solar wind is transferred into electric energy of the magnetosphere. Large-scale electric potentials develop on the dawn and dusk sides at higher latitudes along the open field lines (STERN, 1977). The dynamo layer connects these two regions electrically so that field-aligned electric currents flow from the magnetosphere into the ionosphere on the dawn side and out of the ionosphere during dusk (IIJIMA and POTEMRA, 1976). A secondary system of field-aligned currents flowing in the opposite direction builds up within the regions of closed field lines equatorward of the auroral zones. These fields and currents are enhanced and shifted to lower latitudes during disturbed conditions. They also depend on the polarity of the interplanetary magnetic field.

The ionospheric current component in that current system is called the S_q^P current during magnetically quiet conditions, and the DP2 current during disturbed conditions (NISHIDA and KOKUBUN, 1971). Since the electric conductivity within the auroral zones increases during disturbed conditions (BANKS, 1977), a narrow band of currents develops there called the polar electrojet or DP1.

Fig. 12 shows the equipotential lines of the magnetospheric electric field mapped down to ionospheric heights during moderately disturbed conditions. A potential difference between dawn and dusk of about 70 kV can be observed. During very quiet conditions, that potential difference is only about 25 kV (HEPPNER, 1977).

The equipotential lines in Fig. 12 are also the streamlines of the Hall component of the ionospheric currents. FUKUSHIMA (1971) has shown that the equivalent ionospheric current system derived from ground based magnetic measurements is predominantly the Hall component.

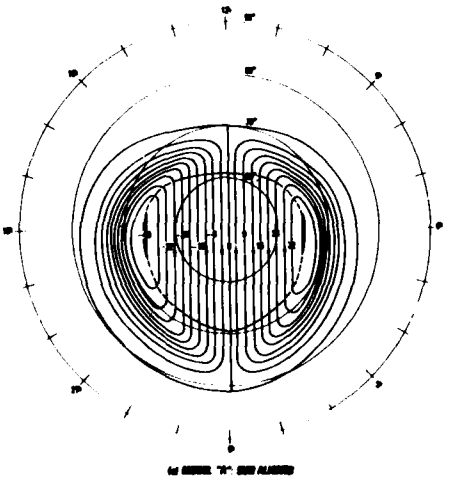


FIG. 12 SEMIEMPIRICAL MODEL OF THE MAGNETOSPHERIC ELECTRIC CONVECTION FIELD AT IONOSPHERIC HEIGHTS. EQUIPOTENTIAL LINES (IN kV) DURING MODERATELY DISTURBED CONDITIONS ON THE NORTHERN HEMISPHERE (after HEPPNER, 1977).

Joule heating due to these currents is a significant fraction of solar XUV heating (≈ 10%) during very quiet conditions, and even exceeds XUV heating during disturbed conditions (VOLLAND, 1979a).

Model calculations of these currents and fields are due to KAMIDE and MATSUSHITA (1979), VOLLAND (1979b) and others.

11. THE DYNAMO LAYER AS AN EQUALIZING LAYER FOR THE GLOBAL THUNDERSTORM FIELD

Electric charges separate within a thundercloud as a result of the combined action of convection and gravity with positive charge stored in the upper part of the cloud and negative charge stored in the lower part. Voltage differences of 10 to 100 MV can develop between the upper and lower parts of the cloud. Discharging currents flow outside the cloud. Intracloud lightning flashes are short circuit currents within the cloud.

The exponentially increasing conductivity of the atmosphere is responsible for a substantial current of the order of 1A flowing from the top of the cloud into the ionosphere. With about 1200 thunderstorms acting simultaneously on Earth, a total current of about 1 kA flows into the ionosphere. A smaller current of about 200 A flows directly to the Earth in the immediate environment of the thunderstorms, i.e. along R_T in Fig. 2 (KASIMIR, 1959). The total resistance between the cloud tops and the ionosphere is of the order of 25 kΩ (R_u in Fig. 2). The total resistance between Earth and ionosphere in fair weather regions is about $R_A \approx 250\ \Omega$. The overall ionospheric resistance $R_D \approx 1/\tilde{\Sigma} \approx 30$ mΩ is negligible compared with R_u and R_A. Therefore, the ionosphere behaves like an equipotential layer for the DC currents of thunderstorm origin. The same is true for the Earth (R_E in Fig.2). Hence, the currents flow from the cloud tops via the ionosphere down to the Earth in the fair weather regions. The link between Earth and cloud bottom is provided mainly by cloud to ground lightning discharges (not shown in Fig. 2).

The atmosphere with its low conductivity imbedded within two spherical shells of relatively high conductivity (the Earth and the ionosphere) resembles a huge spherical capacitor with capacitance C_A.

The globally averaged electric potential between ionosphere and Earth is (HAYS and ROBLE, 1979)

$$U_A = \frac{R_A R_T}{(R_T + R_u)} N I_T \qquad (17)$$

where I_T is the source current of one thunderstorm and N the total number of storms. U_A does not depend on the ionospheric resistance.

However, the thunderstorm field directly above the cloud which does penetrate into the ionosphere, can locally modulate the ionospheric electric field (PARK and DEJNAKARINTRA, 1973).

12. REFERENCES

BANKS, P. M., 1977, Observations of joule and particle heating in the auroral zone, J. Atm. Terr. Phys. 15, 179

BOMKE, H. A., BLAKE, H. A., HARRIS, A. K., HULSE, W. H., STEPPARD, D. J., GIESECKE, A. A., and PANTOGA, A., 1967, Recombination coefficient and coronal contribution to E- layer ionization from magnetic observations of a solar eclipse, J. Geophys. Res. 72, 5913

CHAPMAN, S., and BARTELS, J., 1951, Geomagnetism, Clarendon Press, Oxford

CHAPMAN, S., and LINDZEN, R. S., 1970, Atmospheric Tides, D. Reidel, Dordrecht

FAMBITAKOYE, O., MAYAUD, P. N., and RICHMOND, A. D., 1976, Equatorial electrojet and regular daily variation S_R- III, J. Atm. Terr. Phys. 38, 113

FEJER, J. A., 1964, Atmospheric tides and associated magnetic effects, Rev. Geophys. 2, 275

FORBES, J. M., and LINDZEN, R. S., 1976a, Atmospheric solar tides and their electrodynamic effects-I, J. Atm. Terr. Phys. 38, 897

FORBES, J. M., and LINDZEN, R. S., 1976b, Atmospheric tides and their electrodynamic effects-II, J. Atm. Terr. Phys. 38, 911

FORBES, J. M., and LINDZEN, R. S., 1977, Atmospheric tides and their electrodynamic effects-III, J. Atm. Terr. Phys. 39, 1369

FUKUSHIMA, N., 1971, Electric current systems for polar substorms and their magnetic effect below and above the ionosphere, Radio Sci. 6, 269

HARPER, R. M., 1977, A comparison of ionospheric currents, magnetic variations, and electric fields at Arecibo, J. Geophys. Res. 82, 3233

HAYS, P. B., and ROBLE, R. G., 1979, A quasi- static model of global atmospheric electricity-I, J. Geophys. Res. 84, 3291

HEPPNER, J. P., 1977, Empirical models of high latitude electric fields, J. Geophys. Res. 82, 1115

IIJIMA, T., and POTEMRA, T. A., 1976, The amplitude distribution of field- aligned currents at northern high latitudes observed by Triad, J. Geophys. Res. 81, 2165

KAMIDE, Y., and MATSUSHITA, S., 1979, Simulation studies of ionospheric electric fields and currents in relation to field- aligned currents, J. Geophys. Res. 84, 4083

KASIMIR, H. W., 1959, The thunderstorm as a generator of the global electric current (in german), Z. Geophys. 25, 33

KATO, S., 1956, Horizontal wind systems in the ionospheric E region deduced from the dynamo theory of the geomagnetic Sq variations, J. Geomagn. Geoelectr. 8, 24

KATO, S., 1966, Diurnal atmospheric oscillation, J. Geophys. Res. 71, 3211

KOHL, H., and KING, J. W., 1967, Atmospheric winds between 100 km and 700 km and their effects on the ionosphere, J. Atm. Terr. Phys. 29, 1045

MAEDA, H., 1955, Horizontal wind systems in the ionospheric E region deduced from the dynamo theory of the geomagnetic Sq variations, J. Geomagn. Geoelectr. 7, 121

MAEDA, H., and KAMEI, T., 1975, Tides in the Earth's far environment, Rep. Ionosph. Space Res. Japan 29, 177

MAEDA, K., and MATSUMOTO, H., 1962, Conductivity of the ionosphere and current system, Rep. Ionosph. Space Res. Japan 1, 1

MAEDA, K., and MURATA, H., 1965, Ionospheric dynamo theory with consideration for magnetospheric current along the geomagnetic lines of force, Rep. Ionosph. Space Res. Japan 19, 272

MALIN, S. R. C., 1973, Worldwide distribution of geomagnetic tides, Phil. Trans. Roy. Soc. London 274, 551

MATSUSHITA, S., 1967, Solar quiet and lunar daily variation fields, in Physics of Geomagnetic Phenomena, Matsushita, S., and Campbell, W. H., eds., Academic Press, New York

MATSUSHITA, S., and TARPLEY, J. D., 1970, Effects of dynamo- region electric fields on the magnetosphere, J. Geophys. Res. 75, 5433

MÖHLMANN, D., 1974, Non- uniqueness of ionospheric winds and electric fields, deduced from geomagnetic Sq variations, Gerland's Beitr. Geophys. 83, 16

MÖHLMANN, D., 1977, Ionospheric electrostatic fields, J. Atm. Terr. Phys. 39, 1325

NISHIDA, A., and KOKUBIN, S., 1971, New polar magnetic disturbances S_q^P, SP, DPC and DP2, Rev. Geophys. Space Phys. 9, 417

OHSHIO, M., FUKUSHIMA, N., and NAGATA, T., 1963, Solar flare effect on geomagnetic variation, Rep. Ionosph. Space Res. Japan 17, 77

PARK, C. G., and DEJNAKARINTRA, M., 1973, Penetration of thunderstorm electric fields into the ionosphere and magnetosphere, J. Geophys. Res. 78, 6623

PRICE, A. T., 1967, Electrodynamic induction within the earth, in Physics of Geomagnetic Phenomena, Matsushita, S., and Campbell, W.H., eds., Academic Press, New York

RICHMOND, A. D., 1973a, Equatorial electrojet- I, J. Atm. Terr. Phys. 35, 1083

RICHMOND, A. D., 1973b, Equatorial electrojet- II, J. Atm. Terr. Phys. 35, 1105

RICHMOND, A. D., 1976, Electric field in the ionosphere and plasmasphere during quiet days, J. Geophys. Res. 81, 1447

RICHMOND, A. D., MATSUSHITA, S., and TARPLEY, J.D., 1976, On the production mechanism of electric currents and fields in the ionosphere, J. Geophys. Res. 81, 547

RICHMOND, A. D., and VENKATESWARAN, S. V., 1971, Geomagnetic crochets and associated ionospheric current systems, Radio Sci. 6, 139

RIKITAKE, T., and YUKUTAKE, T., 1962,A theory of s.f.e. current systems, J. Atm. Terr. Phys. 24, 93

RISHBETH, H., and GARRIOTT, O. K., 1969, Introduction to Ionospheric Physics, Academic Press, New York

ROBLE, R. G., and HAYS, P. B., 1979, A quasi- static model of global atmospheric electricity- II, J. Geophys. Res. 84, 7247

ROBLE, R. G., and MATSUSHITA, S., 1975, An estimate of the global- scale joule heating rates in the thermosphere due to time mean currents, Radio Sci. 10, 389

SHUMAN, B. M., 1970, Rocket measurements of the equatorial electrojet, J. Geophys. Res. 75, 3889

STENING, R. J., 1969, An assessment of the contributions of various tidal winds to the Sq current system, Planet. Space Sci. 17, 889

STENING, R. J., 1973, The electrostatic field in the ionosphere, Planet. Space Sci. 21, 1897

STERN, D. D., 1977, Large scale electric fields in the Earth's magnetosphere, Rev. Geophys. Space Phys. 15, 156

SUZUKI, A., 1978, Geomagnetic Sq fields at successive universal times, J. Atm. Terr. Phys. 40, 449

TARPLEY, J. D., 1970a, The ionospheric wind dynamo- I, Planet. Space Sci. 18, 1075

TARPLEY, J. D., 1970b, The ionospheric wind dynamo- II, Planet. Space Sci. 18, 1091

UNTIEDT, J., 1967, A model of the equatorial electrojet involving meridional currents, J. Geophys. Res. 72, 5799

VAN SABBEN, D., 1961, Ionospheric current systems of ten IGY- solar flares effects, J. Atm. Terr. Phys. 22, 32

VOLLAND, H., 1956, On the theory of the solar eclipe effect in the geomagnetic field (in german), J. Atm. Terr. Phys. 9, 131

VOLLAND, H., 1976a, Coupling between the neutral tidal wind and the ionospheric dynamo current, J. Geophys. Res. 81, 1621

VOLLAND, H., 1976b, The atmospheric dynamo, J. Atm. Terr. Phys. 38, 869

VOLLAND, H., 1979a, Magnetospheric electric fields and currents and their influence on large scale thermospheric circulation and composition, J. Atm. Terr. Phys. 41, 853

VOLLAND, H., 1979b, Semiempirical models of magnetospheric electric fields, in Quantitative Modeling of Magnetospheric Processes, Olson, W. P., ed., Geophys. Monogr. 21, AGU, Washington D.C.

VOLLAND, H., and GRELLMANN, L., 1978, A hydromagnetic dynamo of the atmosphere, J.Geophys. Res. 83, 3699

WAGNER, C. U., 1963, Measurements of the geomagnetic solar eclipse effect (in german), J. Atm. Terr. Phys. 25, 529

THE EFFECTS OF AURORAL ACTIVITY ON THE MIDLATITUDE IONOSPHERE

Michel BLANC
CNET/RPE
38-40, rue du Général Leclerc
92131 ISSY-LES-MOULINEAUX, FRANCE

SUMMARY

We review the various physical processes through which part of the solar wind energy impinging on the magnetospheric cavity is finally dissipated in the midlatitude ionosphere, and the present knowledge of the corresponding energy budget.

The solar wind energy and matter entering the magnetosphere is first partly transferred into the auroral zone ionosphere by field-aligned currents, and partly stored into the geomagnetic tail. It is then released from the tail into the ionosphere (field-aligned currents and particle precipitation) and into the ring current (plasma injection). The relative amounts of energy flowing through these various channels are still very imperfectly known.

Energy deposition into the midlatitude ionosphere and thermosphere thus results in part from charge-exchange neutralization of ring current particles, out of which about ten per cent impact on the mid- and low- latitude thermosphere, and in part from global redistribution of the energy entering the auroral zones.

This redistribution takes places either via the neutral atmosphere, or via electrodynamic coupling processes in the plasma.

The neutral atmosphere channel involves the dynamical response of the thermosphere to auroral particle and Joule heating. We discuss the total amount of energy deposited, its altitude and latitude distribution, as well as the relative importance of the various types of dynamical responses of the neutral air (essentially gravity waves and Hadley cells).

The electrodynamic channel involves the midlatitude extension of magnetospheric convection electric fields during magnetic events. We review the present observational evidence for this extension, and our over-all understanding of the global distribution of midlatitude convection electric fields.

I. INTRODUCTION : AN OUTLINE OF THE MAIN ENERGY CHANNELS IN THE MAGNETOSPHERE

In any attempt to establish the energy budget of the various middle and low latitude effects of magnetospheric and auroral activity, one must first describe the complicated geometry of the various energy channels through which solar wind energy is transported within the magnetospheric cavity from its source to the various dissipation regions.

Figure 1 presents a schematic diagram of this energy flow. The main large-scale ionospheric/magnetospheric regions to which the solar-wind/magnetosphere dynamo transfers energy are listed from top to bottom. If we order them radially inwards across the field lines, we find first the solar wind, then the magnetotail, the auroral zone, the ring current, and finally the middle and low latitude ionosphere. The energy storage regions (the magnetotail and the ring current) are shown on the right-hand side, and the energy dissipation regions on the left-hand side. The energy flows along channels involving either the plasma and electromagnetic interactions (continuous arrows) or only neutral particles (dashed arrows). Let us first comment each of them briefly, before concentrating on the problem of energy transfer to midlatitudes.

The solar wind actually exchanges energy in two intricated ways with the magnetosphere. The existence of these two basically different though simultaneously operating ways is probably one of the main causes of the complexity and variability of magnetospheric responses to solar wind parameter changes :

1- The solar wind partly controls the relative amount of open and closed terrestrial magnetic fluxes. The open magnetic flux corresponds at ionospheric heights to the polar caps, and in the magnetosphere to the tail lobes. There is considerable evidence that its total amount increases with the southward component, B_z, of the Interplanetary Magnetic Field. For instance the polar cleft, which delineates the noon sector boundary of the polar cap, is observed to move equatorward (Burch, 1973 ; Kamide et al., 1976) and the total area of the polar cap is observed to increase (Holzworth and Meng, 1975) with the southward increase of B_z.

In this first process, the energy goes essentially from the solar wind to the magnetotail, where it is stored as magnetic field energy.

2- The interplanetary electric field generates, essentially through its Y component, an electric field component parallel to the magnetopause which, projected along magnetic field lines down to the ionosphere, permits to generate a large-scale circulation of plasmas and electric currents inside the magnetosphere by means of the convection electric field. Here too the B_z component of the I.M.F., which induces the dawn-to-dusk component E_y of the interplanetary electric field, seems to mainly control the intensity of the magnetospheric dawn-to-dusk potential drop Φ_o. Gonzales and Mozer (1974) and Reiff et al. (1980) have developed semi-empirical formulas relating Φ_o to B_z.

In this second process, the energy is first transferred from the solar wind to the polar-cap/auroral-zone boundary by means of magnetic field-aligned currents which are part of Iijima and Potemra's (1976) region I currents. Indeed it is easy to see that in the zero-frequency limit an electromagnetic energy flux vector can be written as :

$$S = J \Phi$$

where J is the electric current density and Φ is the electrostatic potential. Therefore with a total region I current intensity of, say, 10^6 A, and a dawn-to-dusk potential drop of 10 to 100 kV depending on magnetic activity, a total power of 10^{10} to 10^{11} Watts, incident upon the region I area of the auroral zone, is available for redistribution within the magnetospheric cavity.

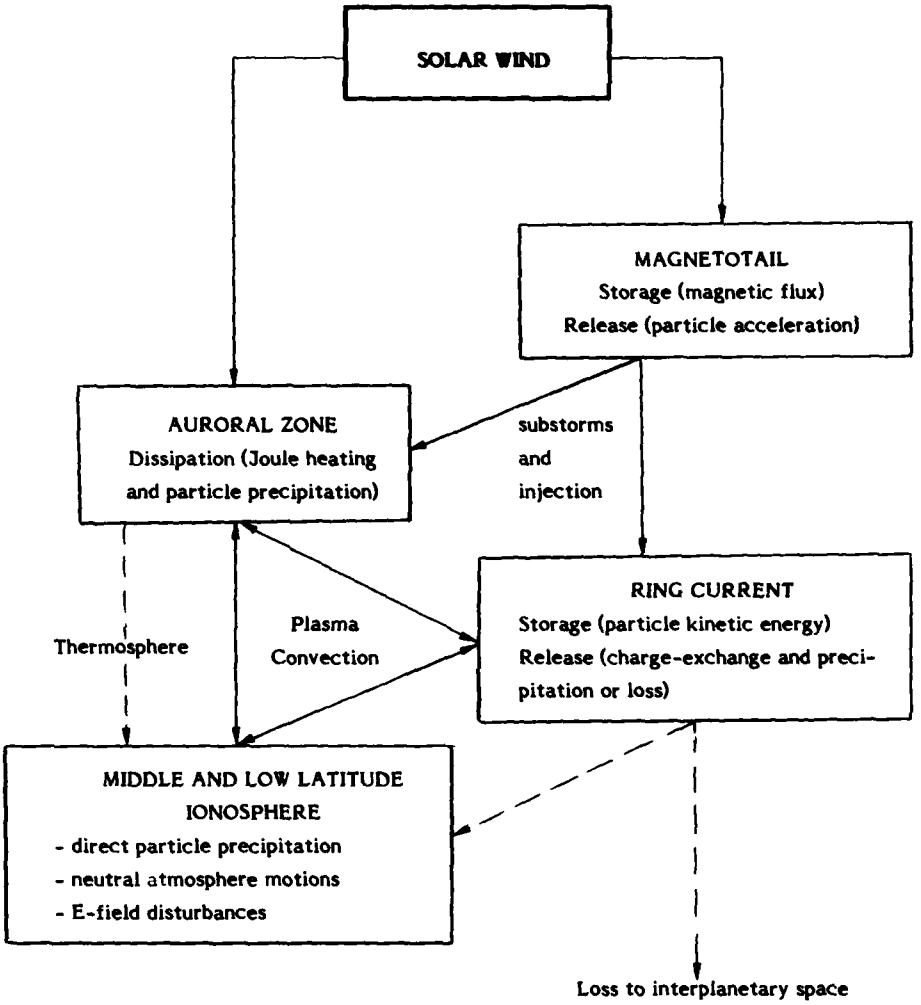


Figure 1 : Schematic diagram of the main magnetospheric regions and of the energy channels connecting them. These regions are ordered radially inwards from top to bottom ; the storage regions (magnetotail and ring current) are presented on the right-hand side, and the dissipation regions (auroral, and middle and low latitude ionosphere) on the left-hand side. The channels involving only neutral particles as energy carriers are indicated by dashed lines.

The magnetotail, which stores magnetic flux and solar wind particle energy, restitutes this energy, in part impulsively, in the course of what is called magnetospheric substorms (see Akasofu, 1977, for a comprehensive description). Whereas a first part of the substorm energy is converted locally into neutral sheet particle heating, another part is transmitted to the auroral ionosphere by field-aligned currents and their ionospheric closure (essentially the westward electrojet). The locally accelerated particles are in turn transported towards the earth, and are finally either precipitated along field lines into the auroral thermosphere (this holds for the major part of the electron population), or convected across field lines into the inner magnetosphere, where part of them are finally trapped on closed drift orbits. This holds for the major part of the ion population, which is stored in the ring current as a result of what is called the injection process.

One can see that the variety of processes responsible for earthward transfer of magnetotail energy makes it very difficult to evaluate the amount of this energy transferred to the auroral ionosphere and to the ring current respectively. As noted by Kamide (1979) in his review of storm/substorms relations, what determines the energy partition rate between these two regions is one of the main questions to which a satisfactory answer has not het been found, despite attempts to evaluate either the sole ring current energy source (Burton et al., 1975) or the auroral energy source separately, or both at the same time (Perreault and Akasofu, 1978). Let us keep in mind that this partition rate is highly variable, with a tendency for this ratio to be in favour of ring current energy storage during large magnetic storm, for which Perreault and Akasofu estimated an energy injection into the ring current about ten times larger than into the auroral ionosphere. This problem covers the whole question of the relation between storms and substorms, whose definitions based on ground magnetic signatures are biased towards either of the two systems : the substorm is defined on the basis of its auroral electrojet current signature, and the magnetic storm (main phase) on the basis of ring current inflation.

The auroral ionosphere clearly appears to be at the crossroads of the system, since it exchanges electric currents and matter with all other regions. Its dynamic interaction with the ring current determines whether magnetospheric convection is confined to the auroral and subauroral field lines (that is, outside the inner edge of the ring current) or whether it extends to the plasmasphere and to the middle and low latitude ionosphere. This interaction has been described by magnetospheric convection theory along linearized, analytical resolutions (Vasyliunas, 1970, 1972 ; Pellat and Laval, 1972 ; Southward, 1977 ; Senior, 1980) or by use of complex numerical models (Wolf, 1970 ; Jaggi and Wolf, 1973 ; Harel et al., 1979). But in addition to this direct electrodynamic coupling process, the auroral ionosphere also transfers energy to midlatitudes via the large-scale dynamic response of the thermosphere to auroral heating.

The midlatitude ionosphere is, finally, at the end point of three main energy channels, which we shall successively describe in this paper :

1- The channel of direct ring current particle precipitation.
2- The channel of the large-scale thermospheric response to auroral heating, which involves thermospheric neutral air convection.
3- The channel of magnetospheric plasma convection.

2. PRECIPITATION OF RING CURRENT PARTICLES INTO MIDDLE AND LOW LATITUDES

The main loss mechanism for the ring current is resonant charge exchange, in particular between H^+ ions and geocoronal neutral hydrogen. Prölss (1973) who calculated the spatial distribution of energetic neutral hydrogen atoms generated in this process, showed that roughly 90 per cent of them are lost into the interplanetary space, and that only 10 per cent are precipitated into the thermosphere at all latitudes. This direct ring current particle precipitation mechanism may be a substantial source of nighttime ionization, as suggested by Lyons and Richmond (1978), who calculated that neutral energetic hydrogen could provide 0.1 particle/cm^3.s during quiet time (a value comparable to production by scattered sunlight) and up to several times 10 particles/cm^3.s during the main phase of a large magnetic storm, thus representing by far the dominant contribution to E region ionization at night under such conditions. The increase of the nighttime electron density with magnetic activity has actually been observed, for instance by Shen et al. (1976) in the upper E region above Arecibo (invariant latitude 32°).

Tinsley (1979 a, b) recently stressed the importance of energetic neutrals even as an energy source for the midlatitude thermosphere. Noting that, as we previously indicated, there may be ten times more energy stored into the ring current than doposited into the auroral thermosphere during major storms, and Prölss' estimate for the proportion of energetic hydrogen atoms reaching the thermosphere, he suggested that there may be just as much energy deposited, rather uniformly, into the middle and low latitude thermosphere, as deposited locally into the auroral zones. Torr and Torr (1979) showed that neutral energetic oxygen atoms can also be a substantial energy source for the thermosphere.

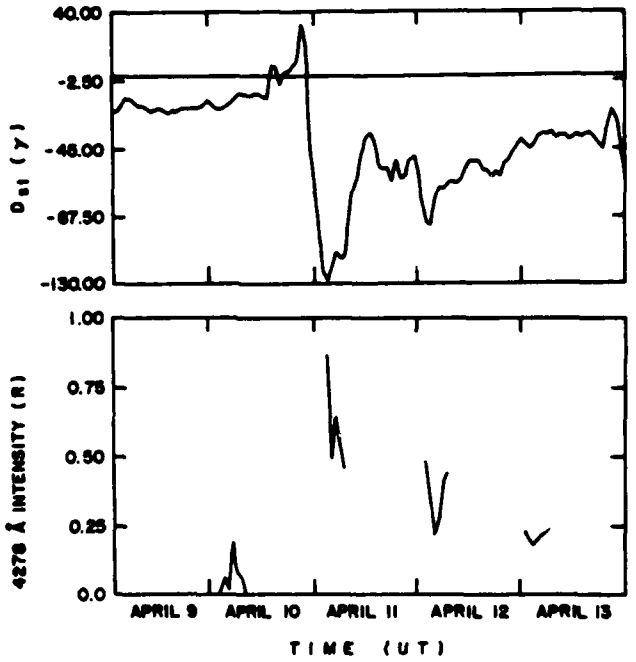


Figure 2 : Variation of the equatorial Dst index (top diagram) and of the intensity of the 4278 Å emission of N_2^+ measured above Arecibo by Meriwether and Walker (1980) during a major magnetic storm in April 1978 (bottom diagram). Despite the limited nighttime interval available for comparison, the similarity of the two curves supports the idea that the observed emission is excited by energetic particles lost from the ring current.

The flux of energetic neutrals produced by the ring current is expected, in a first order approximation neglecting time changes in the pitch angle distribution and in the composition of ring current ions, to be simply proportional to the total ring current intensity, a rough measure of which is given by the equatorial Dst index. Recent optical measurements by Meriwether and Walker (1980) provide evidence for this simple proportionality, as shown in figure 2. The intensity of the 4278 A emission of N_2^+, measured during nighttime above Arecibo, is plotted as a function of time on the lower diagram for a large magnetic storm in April 1978. It seems to follow rather closely the amplitude of Dst, plotted for comparison in the upper part of the figure. The comparison is rather suggestive, despite the limited number of nighttime hours available, and the fact, as stressed by the authors, that the nature of this emission does not permit to identify the corpuscular source responsible for it, in contrast to the previously quoted studies by Tinsley.

Though the basic idea of charge exchange losses of ring current particles, and its effect as a nighttime ionization source, appear rather well established, the chain of multiple charge exchanges and energy degradation of incident energetic neutral atoms is not so well known yet. Similarly, the real importance of this source not for ionization, but for large-scale thermospheric heating and generation of neutral air motions does not appear to me so well established. Tinsley (1979 a) gave such an interpretation of the observation by Hernandez and Roble (1978) that midlatitude storm-time thermospheric winds at the altitude of the 6300 Å emission correlate with the time rate of change of Dst. But his conclusions depend very critically upon the assumed amount of energy deposition by neutral precipitation, and upon its altitude distribution. As we shall see later, alternative explanations for Hernandez and Roble's observations

exist. Considerable experimental progress remains to be done to clarify this point. Actually, even in the auroral zone, where precipitation phenomena have been studied for years or even decades, a global description of the corpuscular energy source and its dependence upon magnetic activity is only now starting to emerge.

3. GLOBAL DYNAMIC RESPONSE OF THE THERMOSPHERE TO AURORAL ENERGY DEPOSITION

Though the main mechanisms at work are at least qualitatively understood, several questions concerning the global redistribution of auroral energy deposition by large-scale neutral air motions have not found a satisfactory answer yet.

The first question is : "What is the altitude, latitude, and magnetic local time distribution of the energy deposition rate, and its variation with magnetic activity ?"

As a preliminary point, one must first separately examine the contributions of particle precipitation and of electromagnetic energy dissipation (Joule heating and the work of the Lorentz force). Banks (1977) performed such a study on the basis of Chatanika incoherent scatter radar data. He showed on the basis of the two nights he examined that the altitude-integrated energy deposition rates by particles and by Joule heating were roughly comparable, but that their altitude distributions were very different. For a neutral wind independant of height as he assumed it to be, the Joule heating rate, which can be written as

$$Q_J = \sigma_p . E'^2$$

where E is the electric field intensity in the frame of reference of the neutral gas, is just proportional to the Pedersen conductivity σ_p of the ionosphere, which maximizes around 120 km altitude. Conversely, the heating rate due to particle precipitation maximizes 10 to 20 kilometers lower down, that is within a much denser neutral atmosphere. Therefore the energy deposition rate per neutral particle, which is the important parameter for the generation of pressure gradients and winds, is substantially larger for Joule than for particle heating.

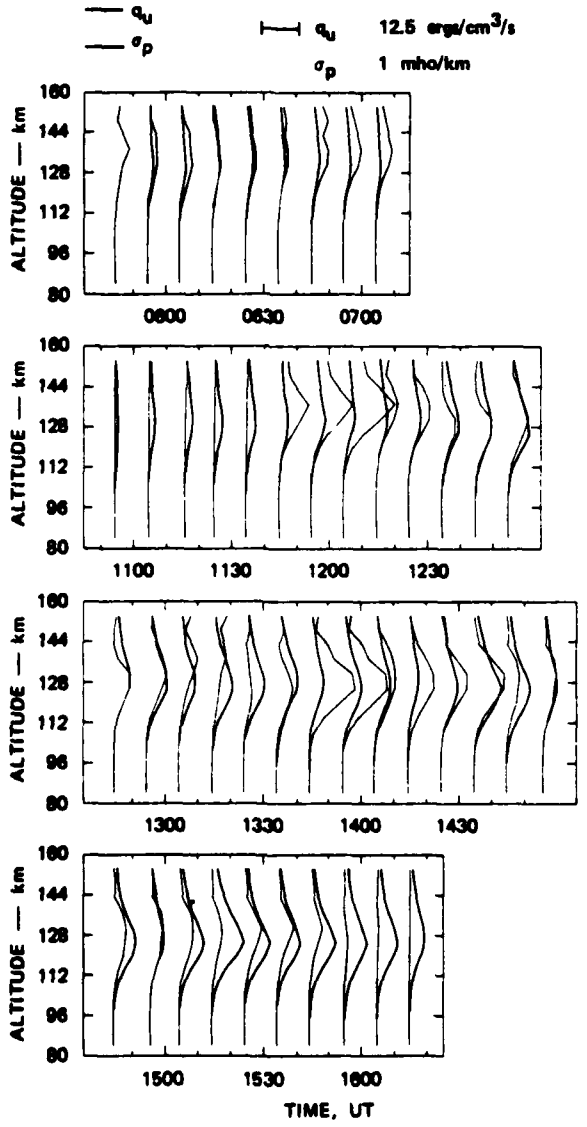


Figure 3 : Altitude profiles of the Joule energy deposition rate q_u (thin lines) and of the Pedersen conductivity σ_p (thick lines) calculated every ten minutes from multiple pulse Chatanika radar data by Brekke and Rino (1978) for an auroral night.

Banks' study, which emphasized the importance of the altitude distribution of heating, suffered itself from a poor altitude resolution in the neutral wind measurement (only 50 km) due to the use of a long pulse for the autocorrelation function measurement. Brekke and Rino (1978) used the multiple pulse technique to estimate the neutral wind profile with an improved 10 km altitude resolution. If this profile is not constant, the maximum of Joule heating may differ from the Pedersen conductivity maximum. This is actually what they found in their results, which are presented in figure 3. Joule heating profiles calculated every tenth minute from the radar data are shown as thin lines throughout an auroral night corresponding to intense magnetic activity. Thick lines corresponding to the Pedersen conductivity profile are also shown for comparison. For the period displayed here the Joule heating profile often maximizes about ten kilometers above the Pedersen profile. Such a value, which corresponds to a substantial fraction of the local atmospheric scale height, is large enough to show the critical importance of an accurate determination, taking into account its neutral wind dependence, of Joule heating. It also demonstrates the difficulty of a global theoretical approach to this problem, since the neutral wind in the auroral zone is itself in part a product of Joule heating.

The second question to be answered is "what kind of organized motion transfers the auroral energy input to middle latitudes" ? Considerable progress, both theoretical and experimental, has been accomplished in the last ten years on this subject.

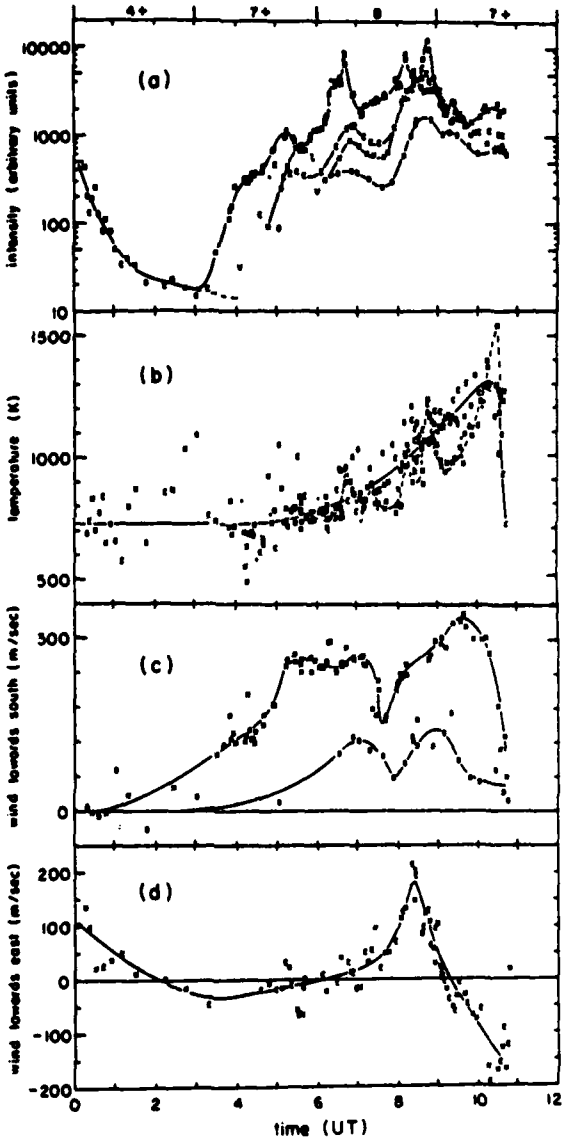


Figure 4 : Time evolution of various atmospheric parameters measured at 6300 Å with a Fabry-Perot interferometer by Sipler and Biondi (1979) at the Laurel Ridge observatory (41°N). The large increase in magnetic activity (see the Kp indices at the top of the figure) produces an intensification of the red line emission intensity (curve a), a gradual large-amplitude increase of the neutral temperature by nearly 600 K (curve b) and the development of a large southward neutral wind (curves c).

On the experimental side, optical interferometry techniques have recently become available to measure neutral temperatures and winds in the various emission layers of the thermosphere. They add to other techniques, and in particular to incoherent scatter, which has for several years provided important contributions to the study of disturbed neutral winds and gravity waves of auroral origin (e.g., Testud et al., 1975, and references therein ; Roble et al., 1978 ; Babcock and Evans, 1979). Using a Fabry-Perot interferometer to measure the width and Doppler shift of the 6300 Å atomic oxygen emission, Hernandez and Roble (1976, 1978 a, b) and Sipler and Biondi (1979) were able to detect large-scale

thermospheric winds and waves generated by auroral activity. One of Sipler and Biondi's observations at the Laurel Ridge observatory (41°N) is presented in figure 4. One can see from top to bottom the emission intensity, which increases abruptly at the time of magnetic activity increase at 03.00 U.T., then the neutral temperature, which increases, much more smoothly, by 600 K in 6 hours, and finally the southward and eastward components of the neutral wind. The southward wind has been measured by pointing the instrument alternatively to the north (measurement points plotted with an N) and to the south (points plotted with an S) of the station. The neutral wind speed appears to be two times smaller to the south than to the north, indicating a strong attenuation of the disturbance with decreasing latitude.

This observation shows rather clearly that on long-term variations (several hours) of the wind, which must correspond to a net equatorward transport of the neutral air, shorter-period wave-like fluctuations are superimposed. In the usual terminology, the first type of neutral wind variation is interpreted as the development of a meridional circulation cell of the neutral air, and the second type as a gravity wave (though there is no possibility here to check the vertical structure of the oscillations). This leads us to the third question we would like to examine : "what are the relative contributions of gravity waves and meridional circulation cells to the transfer of auroral energy to middle and low latitudes" ? This question was addressed in a series of theoretical papers by Richmond (1978 ; 1979 a, b). Starting from an adequate definition of wave energy, i.e. one which coincides with the usual definition for small-amplitude motions and an isothermal and non-dissipative medium, but which can be generalized to motions of arbitrary amplitude in the real thermosphere (see Richmond, 1979 a, for the exact definition), he established the energy budget of a long-lasting, large-amplitude auroral heating event which he simulated using the numerical model described by Richmond and Matsushita (1975). This model is zonally symmetric (i.e. all parameters are independent of longitude or local time), and does not account for the relative diffusion of different atmospheric constituents, but it solves for the U.T. variations of all parameters. One of the main results is presented in figure 5 (Richmond, 1979 b). On the righ-hand side the global energy budget of the generation region (the auroral zone, defined as that region through which magnetospheric field-aligned currents close and the auroral electrojet flows) is presented as a function of altitude in terms of the total Joule heat deposited, and of the energy used for gravity wave generation (each calculated per unit pressure scale height). On the left-hand side, the total amount of energy deposited within the 0° - 45° latitude region is separated into compressional heating and dissipation of wave energy (note the change of scale between the right and left panels). Wave energy accounts for only a very small fraction of the total energy both in the generation and in the dissipation regions. Most of the energy deposited at low latitudes is due to compressional heating of the neutral air in the descending part of the meridional "Hadley type" circulation cell driven by the auroral heat source. Richmond (1979 b) stressed the point that his simulation was intentionally designed to favor gravity wave energy generation rather than meridional circulation. The large amplitude of the heating, its 6-hour duration (rather short for a magnetic storm) and the large fluctuations in the electrojet currents assumed, all tend to bias the result towards gravity wave generation. This suggests that in a real magnetic storm the contribution of gravity waves to large-scale energy transfer must be even smaller than shown in figure 5 (the situation might be different, however, for a short duration heating event, such as an isolated magnetic substorm).

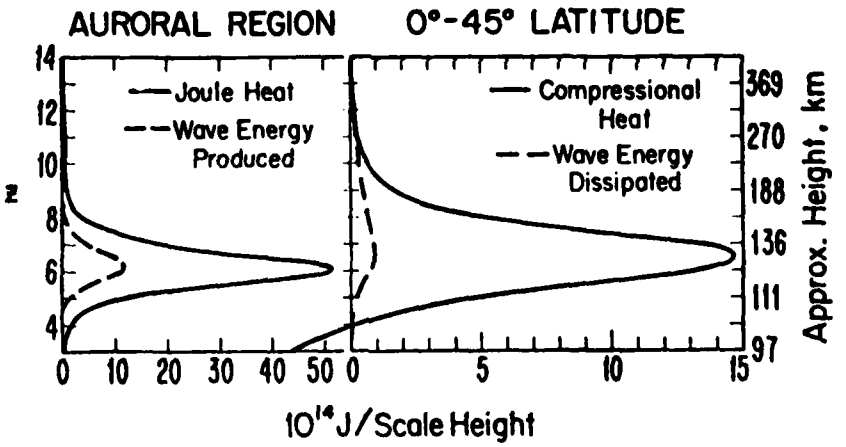


Figure 5 : Some of the main elements of the energy budget of the 6-hour duration magnetic storm simulated numerically by Richmond (1979 b). (Left) Total Joule heat and wave energy production in the auroral zone. (Right) Compressional heating and wave energy dissipated in the low latitude regions (0° - 45°). All quantities are computed per unit pressure scale height, after integration in time, longitude, and latitude over the region indicated.

The dynamical effects of auroral heating are actually not limited to neutral gas motions, but can extend to plasma motions as well. Blanc and Richmond (1980) noticed that since the equatorward wind portion of the meridional circulation cells flows in the upper part of the ionospheric dynamo layer, it may have a substantial dynamo effect, and therefore generate disturbance electric fields which could in turn induce F-region plasma motions. Using the same kind of numerical simulation as previously described, and a numerical dynamo model of the ionosphere, they made a quantitative test of this idea. The main effect of this mechanism can be described as follows : the winds blowing towards the equator in the meridional cell tend to be deviated westward by the Coriolis force, thus leading to a subrotation of the neutral air at subauroral and/or middle latitudes. The dynamo effect of this westward wind blowing at all local times is a poleward electric field which in turn drives a westward motion, or subrotation, of the F-region plasma. In "naïve" terms one could say that the neutral air in the lower thermosphere drags the field lines and forces them, as well as the F-region "frozen-in" plasma, to follow its subrotation. The U.T. and latitude distributions of the westward F-region plasma drifts produced by this mechanism are illustrated by figure 6. It shows the results of the two simulations performed. For each of them, starting from a thermosphere initially at rest, the auroral heat source was turned on in half an hour, and then maintained to a constant value for the rest of the twelve hours of total simulation time. The "case I" simulation (top diagram) corresponds to a large magnetic storm ; case II is a rather weak heating event, and may even be representative of the quiet-time level of auroral heating. The auroral heating zone is limited by the two vertical dashed lines in each panel. The calculated zonal drifts prevailing 3, 6, 9, and 12 hours after storm onset are superposed. For the case I simulation very large drifts (more than 100 m/s) are produced at midlatitudes. They progressively extend towards the equator, and maximize approximately at 45° latitude after 12 hours of heating. Conversely, case II shows much smaller

drifts (no more than 20 m/s) maximizing at the equatorward edge of the heating region, and then decreasing monotonically towards the equator. It thus appears that the extension of disturbance dynamo fields to midlatitudes is a critical function of the amount of heat deposited. Case I and case II represent rather extreme cases, and the real situation prevailing during magnetic storms must stand in between.

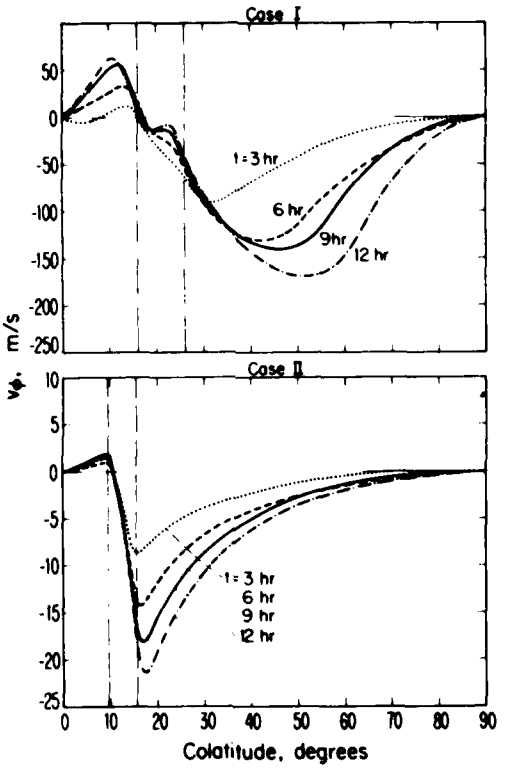


Figure 6 : Latitudinal variations of the electrodynamic east-west ExB drifts in the F region computed by Blanc and Richmond (1980) for four storm times of their simulations. These F-region westward drifts are generated by the dynamo action of the equatorward and westward neutral winds of a meridional circulation cell induced by auroral heating.

4. PENETRATION OF MAGNETOSPHERIC CONVECTION ELECTRIC FIELDS TO MIDLATITUDES

The dynamo effect of thermospheric circulation cells is not the only disturbance mechanism for the midlatitude ionospheric electric field. Incoherent scatter observations at Millstone Hill (Evans, 1972), Saint-Santin (Testud et al., 1975 ; Blanc, 1978, 1980 a, b), and also at the magnetic equator at Jicamarca (Fejer et al., 1979 ; Gonzales et al., 1979) have provided evidence for the direct penetration of magnetospheric convection electric fields to middle and low latitudes during certain types of magnetospheric events. But a clear and unambiguous identification of these events is only starting to emerge. In particular, one must determine what are the roles of solar wind parameters and of substorm activity, respectively, in the generation of these midlatitude convection electric fields. Figure 7 (from Blanc, 1980 a) is an example of such an attempt. The time variations of (a) the I.M.F. B_z component, (b) the AU and AL auroral electrojet indices, (c) the H traces of the Abisko and Leirvogur magnetograms, in the afternoon sector, and (d) the east-west ExB disturbance drift in the F region above Saint-Santin, are shown for April 29, 1976. Curve (e) is an estimate, based on the Dst time rate of change, of the ring current inflation rate (in nT/hr), which must show large negative values during periods of fast ring current growth. The auroral electrojet indices show that two substorms occurred in sequence. During the first one there is no marked ring current growth, and B_z stays close to zero, except between 13.00 and 13.30 U.T., when it briefly displays a large southward value of more than 5 nT. Exactly at that time a marked westward excursion of the Saint-Santin ExB drifts is observed. The second substorm appears to be associated with a sustained 7 nT southward B_z from 15.00 to 17.00 U.T.. A westward excursion of the Saint-Santin ExB drift is observed in good time coincidence, and at the same time the ring current grows rapidly at a rate of about 15 nT/hr.

This example, and several others in Saint-Santin measurements, have led us to the following tentative conclusion : the penetration of convection fields to low latitudes does not occur for every substorm ; rather, it occurs when the I.M.F. displays a large southward excursion. This suggests that it is not substorm activity in itself, but the intensification of the dawn-to-dusk potential drop across the magnetosphere, which generates midlatitude convection fields. The fact that the occurrence of these convection events at midlatitudes is very often associated (for long enough southward excursions of B_z) with fast ring current growth reinforces this hypothesis.

Just as in the auroral zone,midlatitude convection electric fields are strongly local-time dependent. To estimate in an average sense the corresponding plasma flow pattern, we have performed a statistical study of the Saint-Santin disturbance drifts. Averaging separately the equinox drifts for Kp above 2$_+$ and for Kp less than or equal to 2$_+$, and taking the difference between these two patterns, we have constructed the average disturbance drift model presented in figure 8 (from Blanc, 1980 b). The average drift velocities are shown in a polar representation as functions of local time (westward clockwise, and poleward to the center of the circle). The most striking feature of this pattern is the presence of a westward-directed drift at all local times, which strongly contrasts with the auroral situation. This westward drift is near zero between 06 and 12 L.T., and culminates in the evening sector with 75 m/s (about 3 mV/m) at 22 L.T.. We have shown quantitatively (Blanc, 1980 b) that this pattern can be understood as a superposition of the effects of the two electric field disturbance mechanisms mentioned in this paper : the disturbance dynamo electric fields generated by

storm-time meridional circulation cells in the thermosphere (see section 3) create a westward drift at all local times. Superposed to directly penetrating magnetospheric convection electric fields, computed from a high-latitude potential source mapping through the ionosphere, which produces disturbance drifts essentially eastward in the morning and westward in the evening, it explains reasonably well the drift distribution of figure 8.

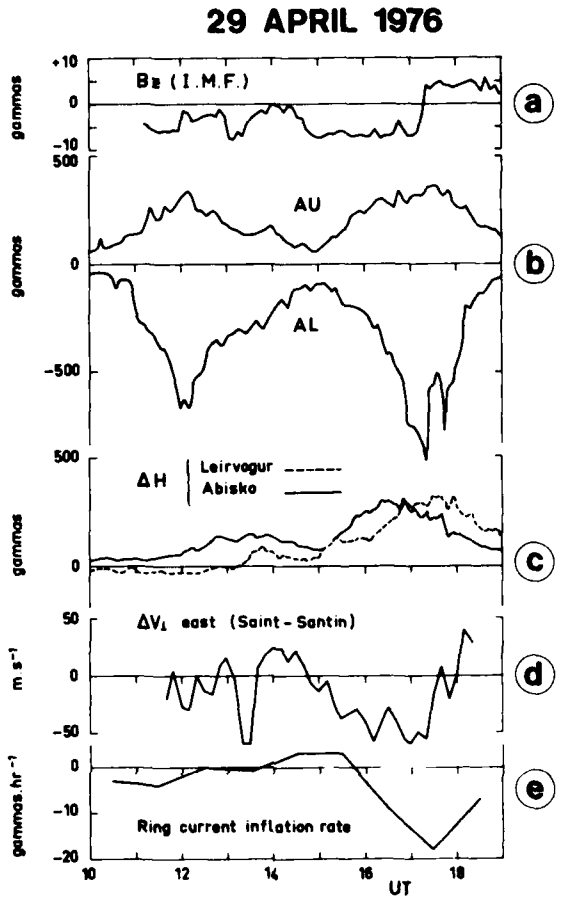


Figure 7 : The variations of the Saint-Santin (45° latitude) disturbance ExB drifts in the east-west direction (curve d) on April 29, 1976, are compared with various solar wind and magnetospheric parameters. The Saint-Santin drifts seem to be controlled more by the large southward values of the I.M.F. B_z component than by substorm activity (as seen on AU and AL) in itself (from Blanc, 1980 a).

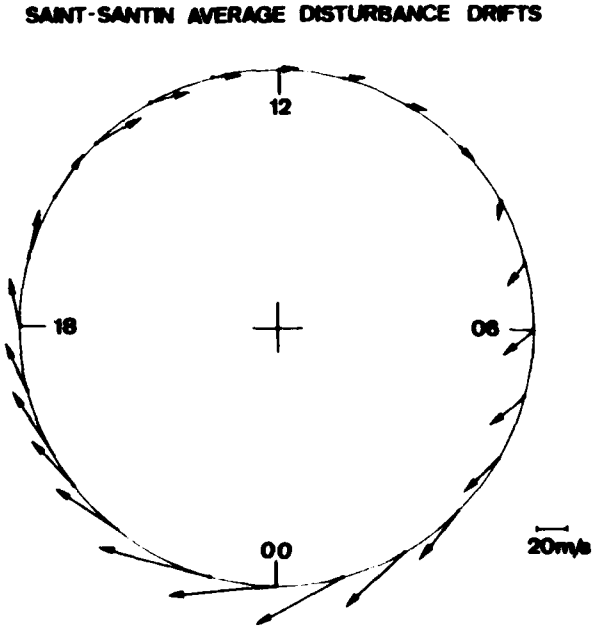


Figure 8 : Average pattern of the Saint-Santin disturbance ExB drifts, presented as a vector plot on a latitude-local time diagram of the northern ionospheric hemisphere by Blanc (1980 b). It was obtained as a harmonic fit to half-hourly values of the difference between the average drifts for Kp $\geq 3_0$ and for Kp $\leq 2+$.

5. SUMMARY

In an attempt to briefly examine the main channels through which solar wind energy is ultimately deposited to middle and low latitudes, we have met a number of unsolved problems, of poorly known physical mechanisms, but also some problems on which considerable progress has been achieved in the recent years.

1- First of all, magnetotail energy is known to be released during the substorm expansion phase, and to go partly into the auroral ionosphere, and partly into the magnetospheric ring current. But the dividing factor between these two contributions seems to depend in a complicated way, which is not yet well understood, upon solar wind parameters. The clarification of this point is a preliminary condition for an improved understanding of storm/substorm relationships.

2- There is mounting evidence that charge-exchange decay of the ring current contributes a substantial part of direct energetic particle precipitation into middle and low latitudes. This mechanism probably has an important effect on nighttime ionospheric conductivities.

3- Paradoxically, the total amount of energy deposited in the auroral zone is much more difficult to monitor than ring current energy variations, which are rather closely reflected in the Dst index. This is due to the complex shape of the three-dimensional distribution of auroral heating, and of its magnetic activity variations. Incoherent scatter observations at Chatanika have contributed to a better knowledge of at least the altitude distribution of auroral heating. They have shown that despite comparable total energy inputs, Joule heating is more efficient than particle heating in generating thermospheric motions because it culminates at higher altitudes (Banks, 1977). But an accurate determination of the neutral wind profile is necessary to determine the exact peak altitude of Joule heating (Brekke and Rino, 1978).

4- The dynamic response of the thermosphere to auroral heating is classically separated into meridional circulation and atmospheric waves. In a recent study, Richmond (1979 a, b) started from a general definition of wave energy to show that waves contribute only for a very small fraction to the budget of energy transfer from auroral to middle and low latitudes during magnetic storms. Most of the energy transfer operates through compressional heating of the neutral gas in the descending part of the meridional circulation cell induced by auroral heating.

5- Storm-induced meridional circulation also induces, via the action of the Coriolis force on an equatorward wind, a subrotation of the thermosphere above 120 km altitude, according to a study by Blanc and Richmond (1980). They showed by a numerical simulation that this neutral air subrotation in turn produces a subrotation of comparable magnitude of the F region, through generation of a disturbance in the ionospheric dynamo electric field.

6- Another important disturbance mechanism acting on midlatitude electric fields is the direct penetration of magnetospheric convection inside the plasmasphere. Incoherent scatter observations support the idea that this penetration occurs for large intensifications of the magnetospheric dawn-to-dusk potential drop produced by large southward excursions of the I.M.F. B_z component, and are generally associated with fast ring current growth.

7- When the local time distributions of these two electric field disturbance mechanisms, computed by a theoretical dynamo model, are superimposed, the observed average distribution of the disturbance drifts at 45° latitude can be satisfactorily reproduced (Blanc, 1980 b).

Among the topics just listed on which attention has been focused in the recent years, there is one particularly clear link : the concept of the magnetic storm main phase. It is within this phase that fast ring current growth, penetration of the convection electric field and direct particle precipitation are observed. I believe that further efforts should be focused on a comprehensive study of the magnetic storm main phase, of all its effects and their relationships to solar wind parameters, and of the storm/substorm relations. It is on this basis that it will be possible to improve our present knowledge of the impact of auroral and magnetospheric activity on the earth's upper atmosphere.

REFERENCES

Akasofu, S.-I., Physics of magnetospheric substorms, 206-257, D. Reidel, Dordrecht-Holland, 1977.

Babcock, R.R., Jr, and J.V. Evans, Effects of geomagnetic disturbances on neutral winds and temperatures in the thermosphere observed over Millstone Hill, J. Geophys. Res., 84, 5349-5354, 1979.

Blanc, M., Midlatitude convection electric fields and their relation to ring current development, Geophys. Res. Lett., 5, 203-206, 1978.

Blanc, M., Magnetospheric convection effects at midlatitudes. I - Saint-Santin observations, submitted to J. Geophys. Res., 1980 a.

Blanc, M., Magnetospheric convection effects at midlatitudes. III - Theoretical derivation of the disturbance convection pattern inside the plasmasphere, submitted to J. Geophys. Res., 1980 b.

Blanc, M., and A.D. Richmond, The ionospheric disturbance dynamo, J. Geophys. Res., 85, 1669-1686, 1980.

Brekke, A., and C.L. Rino, High-resolution altitude profiles of the auroral zone energy dissipation due to ionospheric currents, J. Geophys. Res., 83, 2517-2524, 1978.

Burch, J.L., Rate of erosion of dayside magnetic flux based on a quantitative study of the dependence of polar cusp latitude on the interplanetary magnetic field, Radio Sci., 8, 955, 1973.

Burton, R.K., R.L. McPherron, and C.T. Russell, An empirical relationship between interplanetary conditions and Dst, J. Geophys. Res., 80, 4204-4214, 1975.

Evans, J.V., Measurements of horizontal drifts in the E and F regions at Millstone Hill, J. Geophys. Res., 77, 2341, 1972.

Fejer, B.G., C.A. Gonzales, D.T. Farley, M.C. Kelley, and R.F. Woodman, Equatorial electric fields during magnetically disturbed conditions 1. The effect of the interplanetary magnetic field, J. Geophys. Res., 84, 5797-5802, 1979.

Gonzales, W.D., and F.S. Mozer, A quantitative model for the potential resulting from reconnection with an arbitrary interplanetary magnetic field, J. Geophys. Res., 79, 4186-4194, 1974.

Gonzales, C.A., M.C. Kelley, B.G. Fejer, J.F. Vickrey, and R.F. Woodman, Equatorial electric fields during magnetically disturbed conditions 2. Implications of simultaneous auroral and equatorial measurements, J. Geophys. Res., 84, 5803-5812, 1979.

Harel, M., R.A. Wolf, P.H. Reiff, and M. Smiddy, computer modelling of events in the inner magnetosphere, p. 499 in "Quantitative modelling of magnetospheric processes", edited by W.P. Olson, A.G.U., Washington, D.C., 1979.

Hernandez, G., and R.G. Roble, Direct measurements of nighttime thermospheric winds and temperatures 2. Geomagnetic storms, J. Geophys. Res., 81, 5173-5181, 1976.

Hernandez, G., and R.G. Roble, Observations of large-scale thermospheric waves during geomagnetic storms, J. Geophys. Res., 83, 5531-5538, 1978 a.

Hernandez, G., and R.G. Roble, Relationship between midlatitude thermospheric winds and the time rate of change of Dst, Geophys. Res. Lett., 5, 835-838, 1978 b.

Holzworth, R.H., and C.I. Meng, Mathematical representation of the auroral oval, Geophys. Res. Lett., 2, 377, 1975.

Iijima, T., and T.A. Potemra, The amplitude distribution of field-aligned currents at northern high latitudes observed by Triad, J. Geophys. Res., 81, 2165-2174, 1976.

Jaggi, R.K., and R.A. Wolf, Self-consistent calculation of the motion of a sheet of ions in the magnetosphere, J. Geophys. Res., 78, 2852, 1973.

Kamide, Y., Relationship between substorms and storms, 425-446 in "Dynamics of the magnetosphere", edited by S.-I. Akasofu, D. Reidel, Dordrecht, Holland, 1980.

Kamide, Y., J.L. Burch, J.D. Winningham, and S.-I. Akasofu, Dependence of the latitude of the cleft on the interplanetary magnetic field and substorm activity, J. Geophys. Res., 81, 698, 1976.

Lyons, L.R., and A.D. Richmond, Low-latitude E-region ionization by energetic ring current particles, J. Geophys. Res., 83, 2201-2204, 1978;

Meriwether, J.W., and J.C.G. Walker, First negative band system of nitrogen in the night sky over Arecibo during geomagnetic storms, J. Geophys. Res., 85, 1279-1284, 1980.

Pellat, R., and G. Laval, Remarks on the steady and time dependent mathematical convection models, in "Critical problems of magnetospheric physics", S.T.P. Madrid, Edited by E.R. Dyer, IUCSTP, Washington, D.C., 1972.

Perreault, P.D., and S.-I. Akasofu, A study of geomagnetic storms, Geophys. J. R. Astr. Soc., 54, 547-573, 1978.

Prölss, G.W., Radiation production and energy deposition by ring current protons dissipated by the charge exchange mechanism, Ann. Geophys., 29, 503, 1973.

Reiff, P.H., R. Spiro, and T. Hill, Prediction of the polar cap potential drop, proceedings of the Chapman Conference on high latitude electric fields, Yosemite, California, 1980.

Richmond, A.D., Gravity wave generation, propagation, and dissipation in the thermosphere, J. Geophys. Res., 83, 4131-4145, 1978.

Richmond, A.D., Large-amplitude gravity wave energy production and dissipation in the thermosphere, J. Geophys. Res., 84, 1880-1890, 1979 a.

Richmond, A.D., Thermospheric heating in a magnetic storm : dynamic transport of energy from high to low latitudes, J. Geophys. Res., 84, 5259-5266, 1979 b.

Richmond, A.D., and S. Matsushita, Thermospheric response to a magnetic substorm, J. Geophys. Res., 80, 2839-2850, 1975.

Roble, R.G., A.D. Richmond, W.L. Oliver, and R.M. Harper, Ionospheric effects of the gravity wave launched by the September 18, 1974, Sudden commencement, J. Geophys. Res., 83, 999-1010, 1978.

Senior, C., Les conductivités ionosphériques et leur rôle dans la convection magnétosphérique : une étude expérimentale et théorique, Thèse de 3ème cycle, Université Paris VI, 1980.

Shen, J.S., W.E. Swartz, D.T. Farley, and R.M. Harper, Ionization layers in the nighttime E region valley above Arecibo, J. Geophys. Res., 81, 5517-5526, 1976.

Sipler, D.P., and M.A. Biondi, Midlatitude F region neutral winds and temperatures during the geomagnetic storm of March 26, 1976, J. Geophys. Res., 84, 37-40, 1979.

Southwood, D.J., The role of hot plasma in magnetosphere convection, J. Geophys. Res., 82, 5512, 1977.

Testud, J., P. Amayenc, and M. Blanc, Middle and low latitude effects of auroral disturbances from incoherent-scatter, J. Atmos. Terr. Phys., 37, 989-1009, 1975.

Tinsley, B.A., Energetic neutral atom precipitation during magnetic storms : optical emission, ionization, and energy deposition at low and middle latitudes, J. Geophys. Res., 84, 1855-1864, 1979 a.

Tinsley, B.A., Energetic neutral atom precipitation as a possible source of midlatitude F region winds, Geophys. Res. Lett., 6, 291-293, 1979 b.

Torr, M.R., and D.G. Torr, Energetic oxygen : a direct coupling mechanism between the magnetosphere and thermosphere, Geophys. Res. Lett., 6, 700-703, 1979.

Vasyliunas, V.M., Mathematical models of magnetospheric convection and its coupling to the ionosphere, in "Particles and fields in the magnetosphere", edited by B.M. McCormac, 60, 1970.

Vasyliunas, V.M., The interrelationship of magnetospheric processes, p. 29 in "Earth's magnetospheric processes", edited by B.M. McCormac, 1972.

Wolf, R.A., effects of ionospheric conductivity on convective flow of plasma in the magnetosphere, J. Geophys. Res., 75, 4677-4698, 1970.

CURRENTS IN THE AURORAL ZONE IONOSPHERE

Asgeir Brekke
The Auroral Observatory
P.O. Box 953, N-9001 Tromsø, Norway

SUMMARY

A short outline of the evolution of the model current system at polar latitudes is given. The more recent observations of the Birkeland currents together with the ionospheric currents are interpreted in terms of a simplified current model, which is a current induced by the motion of plasma across the magnetic field lines in the magnetosphere. The amount of Joule heating released in the upper auroral atmosphere during strong auroral electrojets is related to the magnetic field perturbations on ground and it is shown that the amount of heat released will be larger for a positive magnetic bay than for a negative bay of similar strength.

1. INTRODUCTION

The presence of electric currents in the auroral zone and polar cap ionosphere was realized at the turn of this century by Birkeland (1902) who erected an observatory in the northern part of Norway to investigate this phenomenon. His early concept of the auroral zone current system was rather primitive, but he stated two fundamental principles (Figure 1):

1) The particles penetrating the atmosphere and create the aurora are responsible for the geomagnetic disturbances at high latitudes.

2) The particles penetrating the atmosphere are guided by the geomagnetic field, forming a current wedge which is aligned along the magnetic field lines. The current consists of an upgoing and a downgoing field-aligned branch and a horizontal closure current at auroral altitudes. This horizontal current branch is responsible for the large geomagnetic fluctuations seen at high latitudes and the smaller excursions observed at lower latitudes during auroral disturbances. The horizontal current closes in the ionosphere via the low latitude and the polar cap.

The field aligned current branches have later become known as Birkeland currents.

Harang (1946) based on a similar analysis as Birkeland concluded that the intense horizontal current in the auroral zone consists of two branches, one eastward directed in the afternoon and evening hours, and one westward directed in the morning hours. The demarcation region between the two current branches on the night side has later been named the Harang discontinuity. A similar lesser noted demarcation region is also present in the forenoon sector (Figure 2).

2. MODERN OBSERVATIONS OF THE HIGH LATITUDE CURRENT SYSTEM

The presence of the field aligned current branches could never be proven by Birkeland, and their existence was disputed for many years. Not until 1970 were their presence definitely confirmed by Armstrong and Zmuda (1970) using satellite data of the geomagnetic field (Figure 3).

From more recent satellite data of Zmuda and Armstrong (1974) and Iijima and Potemra (1978) it has been shown that the field aligned currents are not only present during auroral disturbances but at any time, disturbed or quiet (Figure 4). Furthermore, the Birkeland currents appear to form a consistent pattern where the downgoing current is on the equatorward side of the auroral oval in the evening sector, and on the poleward side of the oval in the morning sector. The upgoing current on the other hand is on the poleward side of the oval in the evening sector and on the equatorward side of the oval in the morning sector. Thus the field aligned currents are not line currents but rather sheet currents forming two pairs of oppositely directed sheets in the evening and morning sector.

Based on measurements of the electron density and theoretical consideration of electrical conductivity in the auroral zone ionosphere, Baker and Martyn (1953) concluded that the ionospheric currents forming the so-called auroral electrojets were Hall currents. This result has now been confirmed by incoherent scatter measurements at Chatanika, Alaska. From this experiment it is possible to derive the height integrated ionospheric conductivities and the electric field with a time resolution of a few minutes (Figure 5).

Assuming a vertical magnetic field and neglecting any parallel electric field the height integrated overhead ionospheric current is given by

$$\underline{J} = \Sigma_P \underline{E}' - \Sigma_H \underline{E}' \times \underline{B}/B^2 \qquad (1)$$

where Σ_P and Σ_H are the height integrated Pedersen and Hall conductivities respectively. The electric field $\underline{E}' = \underline{E} + \underline{u} \times \underline{B}$ is assumed constant in the height region of the ionospheric currents and $\underline{u}$ is the neutral wind. These estimates of the overhead current density can be compared at any time with the geomagnetic perturbations measured by a nearby magnetometer. The relationship between this current and the magnetic field perturbations on ground is given by

$$\Delta H = + K_1(\Sigma_P E_E' + \Sigma_H E_N') \qquad (2)$$

and

$$\Delta D = - K_2(\Sigma_P E_N' - \Sigma_H E_E') \qquad (3)$$

where K_1 and K_2 are constants taking the earth induction and the geographic extent of the currents into account. (ΔH and ΔD are measured positive northward and eastward respectively.)

Results of such calculations are shown in Figure 6. A good agreement between the H-component of the geomagnetic perturbation vector and the eastward ionospheric current density is evident. Since the electric field is basically directed in the meridional plane this observed relationship implies that a polarization effect is present in the auroral ionosphere and that the Hall current is the major source of the deflection in the H-component.

When a similar comparison is made between the D-component of the geomagnetic perturbation vector and the northward component of the current, the agreement is less apparent (Figure 6). The reason for this is not easy to find but some likely sources could be:

- Distant currents in the polar cap magnetosphere (magnetopause)
- Induction effects in the ground
- Currents in the ionospheric D-region
- Unbalance between upgoing and downgoing current sheets above the site.

Particularly during auroral precipitation events a rather dense D-layer is formed, and because the electric field can penetrate to D-region heights (see Goldberg, paper 15) it is not unlikely that a considerable current might flow. It has lately been possible to derive D-region densities down to altitudes of 70 km at Chatanika which do indicate that such currents actually are flowing (Banks, 1979).

3. A SIMPLIFIED CONCEPT OF THE BIRKELAND CURRENTS

The results from the Triad data obtained by Potemra (1978) hold together with the Chatanika results indicate that the Birkeland currents are basically closed by Pedersen currents in the meridional direction in the E-region ionosphere. The presence of this current system at all time and particularly in quiet periods suggests a simplified current model as illustrated in Figure 7. The Birkeland current pairs are closed by Pedersen currents in the morning and evening side auroral ionosphere and by radial currents directed from dusk to dawn in the equatorial plane. Since the electric field in the equatorial plane is directed from dawn to dusk the radial current in the magnetosphere will represent an energy source. In the ionosphere, however, the electric field and the Pedersen currents will be parallel and therefore represent a load in terms of Joule heating in the atmosphere (Figure 8). The Lorenz force which will result in the equatorial plane ($j \times \underline{B}$) will be directed towards the plasma flow. In this simplified concent one can therefore imagine the high latitude current system as a result of Lenz's law where the currents are induced in order to sustain the momentum of the plasma flow across the magnetic field in the equatorial plane.

4. FINE HIGH RESOLUTION MEASUREMENTS OF THE AURORAL ZONE CURRENTS

Recently it has also been possible to derive altitude profiles of the ionospheric currents by the incoherent scatter radar at Chatanika, Alaska (Rino et al., 1977). These observations (Figure 9) clearly show how the currents change direction by altitude from being mainly a Pedersen current at the uppermost altitudes to being a Hall current at the lower altitudes. In Figure 9 the electric field is basically northward between 0600 and 0700 UT and southward from 1500 to 1600 UT.

One also notices from these observations that the maximum auroral electrojet is situated slightly higher during an eastward electrojet than during a westward electrojet. It is not unlikely that this difference can have some influence on the creation of gravity waves in the auroral zone. There are indications that the source of these waves are related to the eastward electrojet.

5. JOULE HEATING IN THE AURORAL IONOSPHERE

From the observations of the currents and the electric field one can estimate the amount of Joule heat which is deposited in the auroral atmosphere. This heat is given by

$$Q = \int_{h_1}^{h_2} \underline{j} \cdot \underline{E}' \cdot dh = \Sigma_p \, \underline{E}'^2$$

where h_1 and h_2 are the lower and upper borders of the region where the heating takes place, and Σ_p is the height integrated Pedersen conductivity assuming the electric field is constant in the actual height region. In Figure 10 calculations of this heating is shown together with the magnetic fluctuations close to the site of the radar observations. One can see that the amount of Joule heating is much larger during the negative bay than during the positive bay. Cole (1971) has shown theoretically that the amount of Joule heating can be expressed as

$$Q = k \cdot (\Delta H)^2 \tag{4}$$

where ΔH is the magnetic field fluctuations and k is a scaling factor. Now the magnetic field fluctuations are basically caused by Hall currents and can approximately be given by

$$\Delta H = k' \cdot \Sigma_H \, E' \tag{5}$$

where Σ_H is the height integrated Hall conductivity and k' is a proportionality factor, and on the average equal to 0.17 and 0.29 in a positive and negative bay respectively (Kamide and Brekke, 1975). Therefore

$$Q = kk' \cdot \Sigma_H^2 \cdot E'^2 \tag{6}$$

and since $Q = \Sigma_p \cdot E'^2$ one finds that

$$k = \begin{cases} 5.89 \cdot \Sigma_p/\Sigma_H^2 & \text{in a positive bay} \\ 3.50 \cdot \Sigma_p/\Sigma_H^2 & \text{in a negative bay} \end{cases} \quad (7)$$

This means that the scaling factor k which expresses the efficiency of the magnetic fluctuations in terms of Joule heating, will vary with the ratio of the conductivities. During an auroral substorm the Hall-conductivity will increase more than Σ_p and therefore k will decrease during such events. This is shown in Figure 10, where it can be seen that during a positive bay when there is less precipitation the amount of Joule heating is relatively larger than during a similar negative bay. In Figure 11 this effect on the k value is summarized for several events of Joule heating in association with positive and negative bays.

The energy released by the precipitating particles during the ionization process is also of considerable importance to the total amount of heat released in the polar ionosphere during auroral events. In order to make numerical models of the dynamics of the upper atmosphere on a global scale, it will be of interest to obtain quantitative values for the heat deposited in total by auroral substorm phenomena. Vondrak and Baron (1977) have developed a method for deriving the energy distribution of auroral electrons from incoherent scatter radar measurements. By assuming equilibrium and neglecting any transport terms in the continuity equation for the electron density, the production rate is given by

$$q = \alpha_{rec} N_e^2 \quad (8)$$

From measurements of N_e at each altitude one can derive the production rate q as function of altitude. It is usually assumed that 35 eV is released on the average during every ionization event, therefore one can find an average amount of the particle energy deposited in the auroral ionosphere at every height. In Figure 12 are shown the total amount of heat deposited, both the Joule heating rate and the particle heating rate added together. On the abscissa is the simultaneously measured $(\Delta H)^2$ parameter from a nearby magnetometer. The relationship is shown both for negative and positive bays. Again one can see that relative more heat is deposited for positive magnetic bays than for negative. In these calculations, we have accounted for the fact that only a fraction ($\sim$ 30%) of the total particle energy goes into heating the neutral atmosphere. On the average Figure 12 indicates that the Joule heating rate is the most important heating source in the polar ionosphere.

6. CONCLUSION

The current system in the auroral ionosphere appears to be fairly well understood in a coarse sense. The observed field aligned current sheets are most likely closed by meridional Pedersen currents in the auroral oval. The electric field driving the ionospheric currents is strongly polarized forming an eastward and a westward auroral Hall current electrojet, which are slightly shifted in altitude. The eastward electrojet being situated a few kilometers higher. The amount of Joule heating released in the upper atmosphere appears to be relatively stronger with respect to the strength of positive magnetic bay events than to the strength of negative bay events. By scaling magnetograms in order to derive a quantitative estimate of the amount of Joule heating released at any time, one has to be aware of this lesser efficiency of the negative bays in terms of Joule heating. In addition the Joule heat also appears to be deposited at slightly higher altitudes during positive bays than during negative bays (Brekke and Rino, 1978; and Blanc paper 12).

REFERENCES

ARMSTRONG, J.C. and A.J. ZMUDA, 1970, "Field-aligned currents at 1100 km in the auroral region measured by satellite", J. Geophys. Res., 75, 7122.

BAKER, W.G. and D.F. MARTYN, 1953, "Electric currents in the ionosphere, I. The Conductivity", Phil. Trans. R. Soc., Vol. 246, A913, 281-294.

BANKS, P., 1979, "Joule heating in the high-altitude mesosphere", J. Geophys. Res., 84, A11, 6709-6712.

BIRKELAND, K., 1908, "The Norwegian Aurora Polaris Expedition 1902-03", Christiania, Norway.

BREKKE, A. and C.L. RINO, 1978, "High resolution profiles of the auroral zone energy dissipation due to ionospheric currents", J. Geophys. Res., 83, 2517-2524.

COLE, K.D., 1971, "Electrodynamic heating and movement of the thermosphere", Planet. Space Sci., 19, 59-75.

HARANG, L.M., 1951, "The Aurora", Chapman and Hall, LTD, London.

IIJIMA, T. and T.A. POTEMRA, 1978, "Large-scale characteristics of field-aligned currents associated with substorms", J. Geophys. Res., 83, A2, 599-615.

KAMIDE, Y. and A. BREKKE, 1975, "Auroral Electrojet Current Density Deduced from the Chatanika Radar and from the Alaska Meridian Chain of Magnetic Observations", J. Geophys. Res., 80, 587-594.

MURDIN, J. and A. BREKKE, 1980, "Chatanika Incoherent Scatter Radar Measurements: A three day experiment", Internal Report at Kiruna Geophysical Institute.

POTEMRA, T.A., 1977, "Large-scale characteristics of field-aligned currents determined from the TRIAD magnetometer experiment", in Dynamical and Chemical Coupling Bewteen the Neutral and Ionized Atmosphere", ed. B. Grandal and J.A. Holtet, D. Reidel Publ. Comp., pp. 337-352.

RINO, C.L., A. BREKKE, and M.J. BARON, 1977, "High resolution auroral zone E-region neutral wind and current measurements by incoherent scatter radar", J. Geophys. Res., 82, 2295-2304.

VONDRAK, R.R. and M.J. BARON, 1977, "A method of obtaining the energy distribution of auroral electrons from incoherent scatter radar measurements", in Radar probing of the auroral plasma, ed. A. Brekke, Universitetsforlaget, Oslo and Tromsø, Norway, pp. 315-330.

ZMUDA, A.J. and J.C. ARMSTRONG, 1974, "The diurnal flow pattern of field-aligned currents", J. Geophys. Res., 79, 4611.

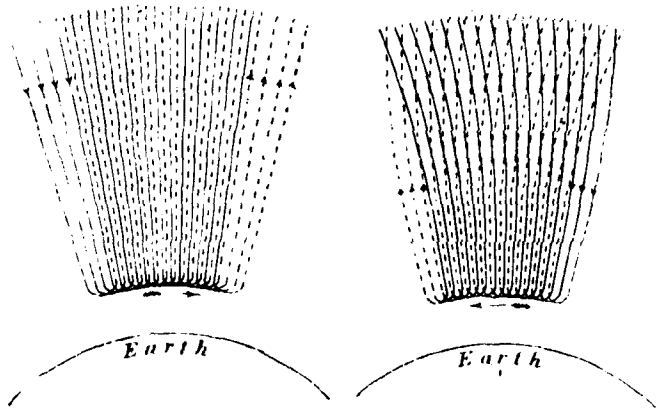


Figure 1 An early model of the current system in the auroral zone (Birkeland, 1908).

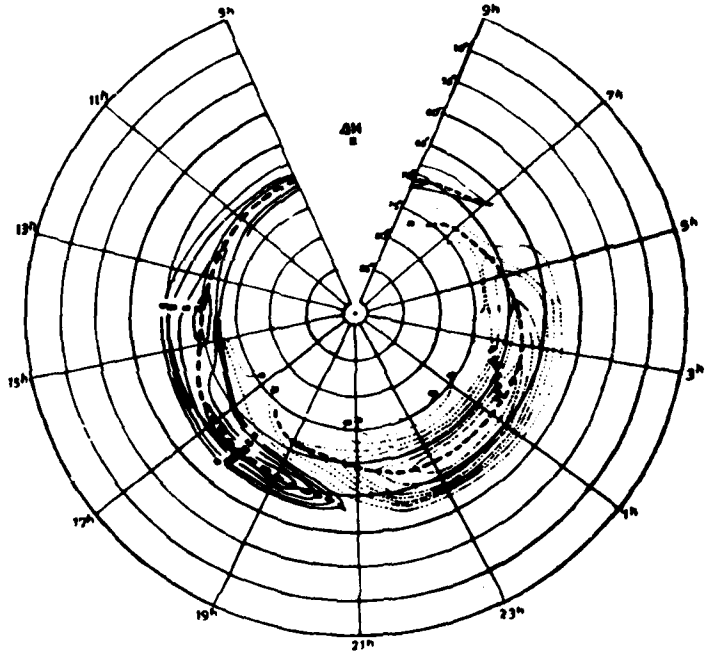


Figure 2 Diurnal variation of the geomagnetic field as function of time of day and geomagnetic latitude observed in Scandinavia by Harang. Full lines indicate ΔH positive northward (Harang, 1946).

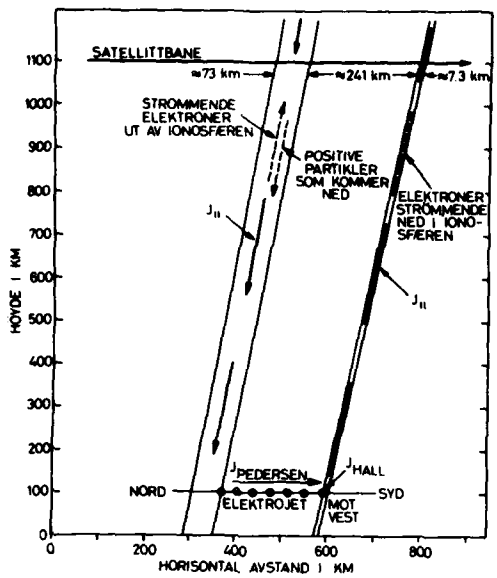


Figure 3 The current system from the satellite altitude downward. (After Armstrong and Zmuda, 1970).

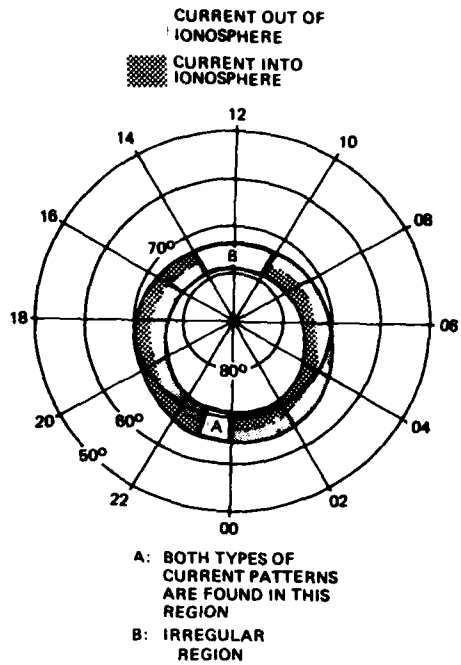


Figure 4 The diurnal flow pattern along the auroral oval marked by transverse magnetic disturbances. The coordinates are dipole latitudes and hours. (After Zmuda and Armstrong, 1974).

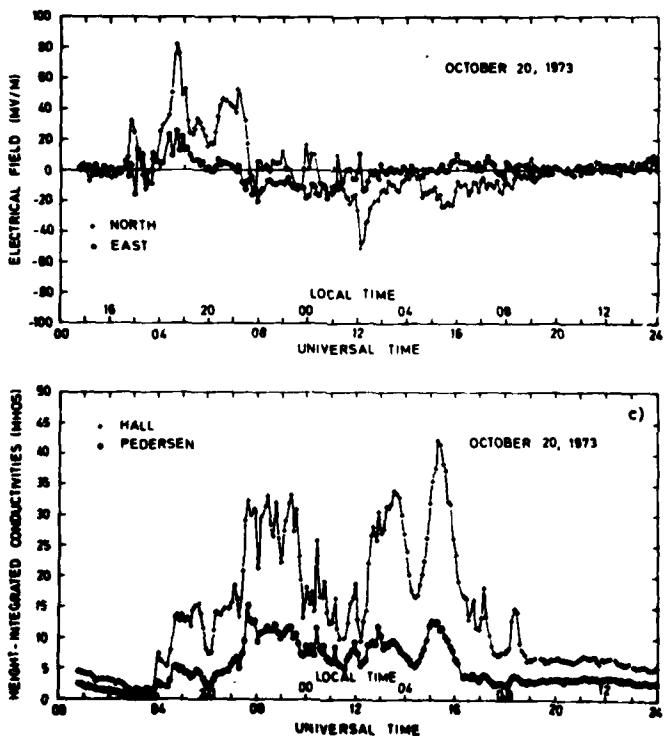


Figure 5 Time variation in the electric field and height integrated conductivities derived from the Chatanika radar.

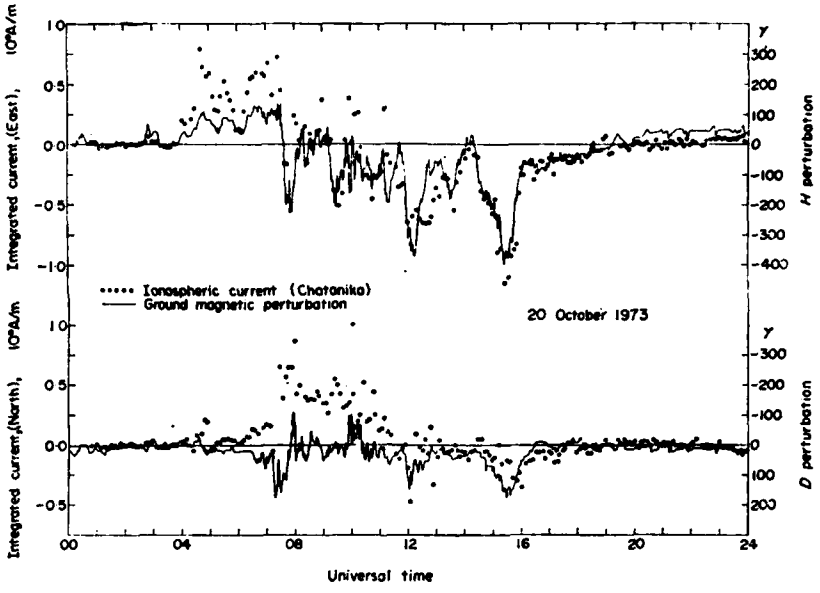


Figure 6 A comparison between the height integrated currents derived by data from the Chatanika Radar and simultaneous magnetic field fluctuations measured by a nearby magnetometer.

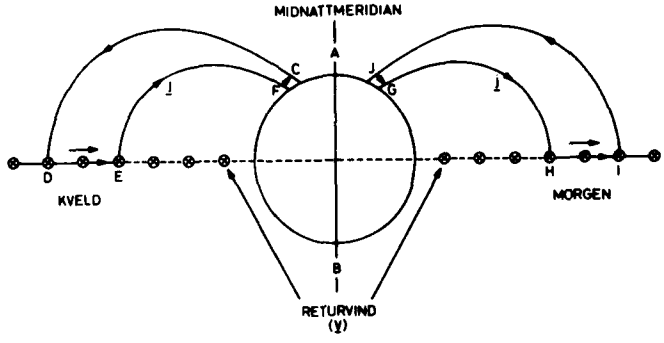


Figure 7 A simplified model of the quiet time current system agreeable to satellite measurements of field aligned currents and radar measurements of the horizontal ionospheric current.

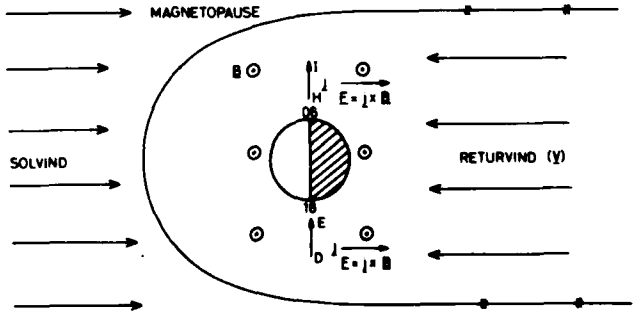


Figure 8 An illustration of the closure current in the equatorial plane resulting in a Lorenz force opposing the plasma flow in the magnetosphere.

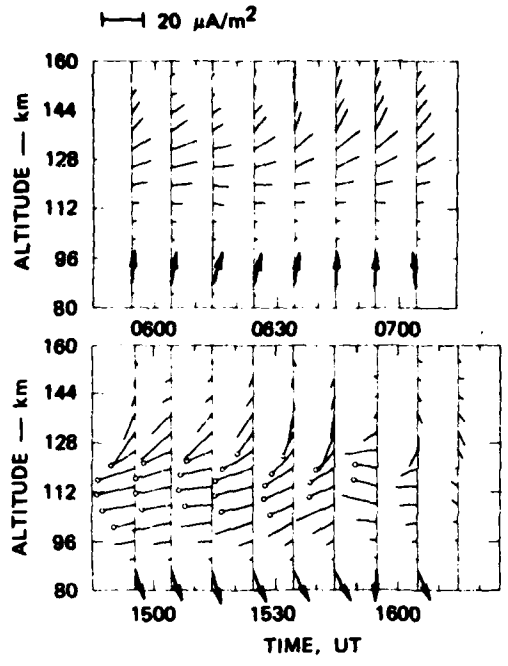


Figure 9 Fine height resolution measurements of the currents in the auroral ionosphere. The direction of the current arrows is northward upwards and eastward towards the right, and the electric field is indicated by a black arrow. (After Rino et al., 1977).

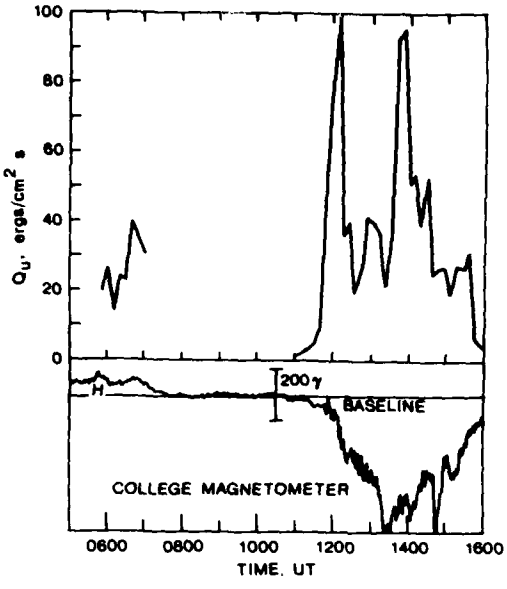


Figure 10 The height integrated Joule heating deposited in the auroral upper atmosphere during a substorm event. (After Brekke and Rino, 1978).

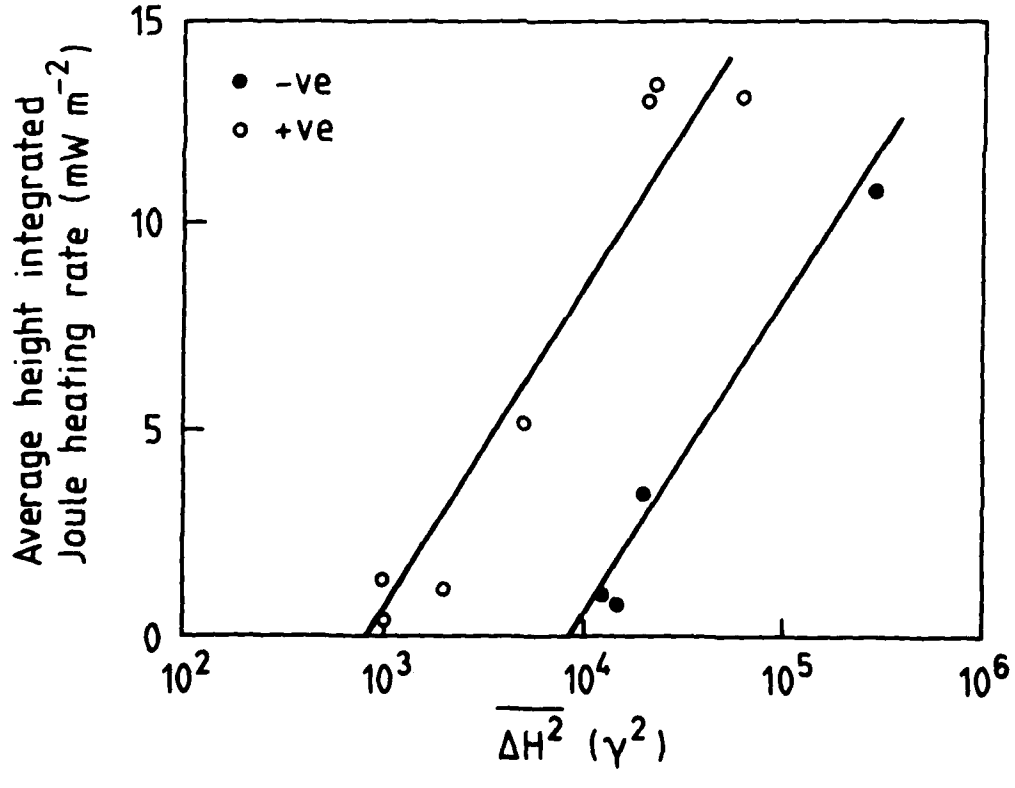


Figure 11 The average height integrated Joule heating rate versus the square of the geomagnetic perturbation $(\Delta H)^2$, as measured by the Chatanika radar and a nearby magnetometer, respectively. (After Murdin and Brekke, 1980).

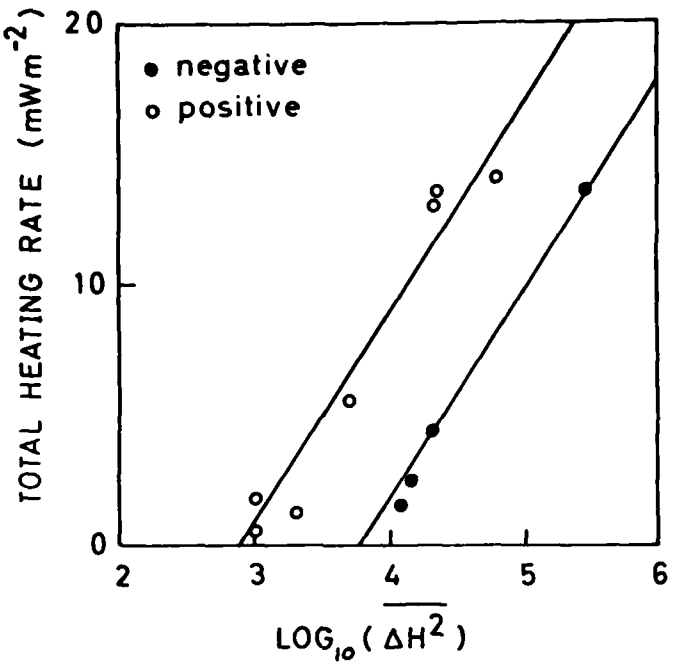


Figure 12 The average height integrated Joule heating rate and particle energy dissipation rate versus the square of the magnetic perturbation $(\Delta H)^2$ measured by the Chatanika radar and a nearby magnetometer. (After Murdin and Brekke, 1980).

NEUTRAL WINDS IN THE POLAR IONOSPHERE

Asgeir Brekke
The Auroral Observatory
P.O. Box 953, N-9001 Tromsø, Norway

SUMMARY

During the last ten years we have witnessed progress in the understanding of the neutral air dynamics in the polar ionosphere. In this report we will present some of the most recent observations of this important geophysical phenomenon. A comparison with some of the present models of the neutral air global wind system is also made. The need for more research in this field is stressed.

1. INTRODUCTION

The motion of the neutral air in the upper atmosphere, say above 90 km altitude is not easy to investigate. There are only a few techniques available for direct observations of the neutral air dynamics. Until about 10 years ago most of our knowledge about the neutral winds above 90 km was based on observations of meteor trails. In the beginning of the 1970's the first rocket releases of chemiluminescent trails were successfully launched into the upper atmosphere, and later on we have got observations of the neutral wind in the thermosphere by incoherent scatter radars and Fabry-Perot interferometers. Except for the incoherent scatter radar techniques the other methods are limited to twilight or complete darkness, therefore can all techniques give a complete daily pattern of the neutral wind only in limited regions at high latitudes. During the last years we have experienced an upswing in the interest of the thermospheric neutral wind system at high latitudes and realized its importance to the global system of neutral air dynamics.

2. EQUATIONS OF MOTION FOR THE NEUTRAL AIR

To investigate the relative importance of the different forces acting on the neutral air mass, the momentum equation of the neutrals can be written:

$$\frac{d\underline{u}}{dt} = -2\underline{\Omega} \times \underline{u} + \underline{g} - \frac{1}{\rho}\nabla p + \nabla\phi + \frac{\mu}{\rho}\nabla^2\underline{u} - \nu_{ni}(\underline{u} - \underline{v}) \qquad (1)$$

In this equation the following symbols are used:

- $\underline{u}$ - the neutral air velocity
- $\underline{\Omega}$ - the angular velocity of the earth ($= 7.3 \cdot 10^{-5}$ sec^{-1})
- ρ - the neutral gas density
- p - the neutral pressure
- $\nabla\phi$ - represents a tidal force
- μ - the coefficient of molecular viscosity
- ν_{ni} - the neutral-ion collision frequency ($\sim 7 \cdot 10^{-16} \cdot N_i$)
- $\underline{v}$ - the ion gas velocity
- N_i - the ion number density

Since the neutral gas motion is so closely linked to the ion gas through the collision term one should also solve simultaneously the ion mobility equation. This equation is given by

$$\frac{d\underline{v}}{dt} = -2\underline{\Omega} \times \underline{v} + \underline{g} - \frac{1}{\rho_i}\nabla p_i + \frac{e}{m_i}(\underline{E} + \underline{v} \times \underline{B}) - \nu_{in}(\underline{v} - \underline{u}) \qquad (2)$$

where the subscript i refers to the ions and

- m_i - the ion mass
- ν_{in} - the ion-neutral collision frequency ($\nu_{in} \sim 7 \cdot 10^{-16} \cdot N$)
- N - the neutral gas number density
- $\underline{E}$ - the electric field
- $\underline{B}$ - the geomagnetic field

In this figure (Figure 1) from Maeda and Fujiwara (1967) the different forces are compared on a normalized scale and one can see that in the E-region between 90 and 150 km the pressure force, the Coriolis force and the ion drag are dominating forces, while the viscous drag term is of minor importance and the tidal force $\nabla\phi$ can be completely neglected together with the gravity force for horizontal motion of the air.

One complicating result of this combination of forces is that an instantaneous observation of the neutral wind at a particular point in space reflects the effects of forces some other places at some other time. In order to get a comprehensive view of the neutral air dynamics one would need observations based on a global scale. This is not easy to organize and so far most of the measurements of the neutral winds in the polar ionosphere have been performed as single cases or event studies. Recently, however, one has seen long time series of the neutral wind based on radar observations and Fabry-Perot measurements. When

such long time series are performed one will eventually get a better understanding of the behaviour of the wind both during more regular quiet conditions and during geomagnetic disturbances.

3. EARLIER MODELS OF THE THERMOSPHERIC WINDS

In the earlier models of the thermospheric winds, several assumptions were made with respect to the relative importance of the different forces. In the well known work by Kohl and King (1967) it was assumed that the atmosphere was static below 110 km and that the only driving force was the global pressure asymmetry caused by the daily variation of the solar heat input. Thus Kohl and King (1967) neglected any effect of electric fields. In Figure 2 are shown the model of Kohl and King (1967) when an ion density of 10^{12} m^{-3} is assumed. From the model which applies for the 300 km height region the wind direction across the pole is mainly aligned along the 1500-0300 LT meridian or approximately parallel to the driving force. For model calculations with a third the ion density (3×10^{11} m^{-3}) the wind amplitude is increased by a factor of 3 and the direction is turned approximately 30° towards the right.

This relationship can be seen by a simple analysis of the neutral mobility equation. By assuming balance between the driving force, the Coriolis force and the ion drag term one gets the following reduced mobility equation for the neutrals in steady state:

$$+ 2\underline{\Omega} \times \underline{u} + \frac{1}{\rho} \nabla p + \nu_{ni} (\underline{u} - \underline{v}) = 0 \tag{3}$$

In this equation we have neglected the viscosity force, and according Kohl and King (1967) this assumption is valid between 200 and 400 km and below 150 km altitude for electron densities larger than 3×10^{11} m^{-3}. Solving Eq. (3) for $\underline{u}$ gives

$$\underline{u} = K_1 \underline{v}_\perp' - K_2 \underline{\Omega} \times \underline{v}_\perp'/\Omega + \underline{v}_\parallel' \tag{4}$$

where

$$\underline{v}' = \underline{v} - \frac{1}{\rho \nu_{ni}} \nabla p \tag{5}$$

and

$$K_1 = \left[1 + \left(\frac{2\Omega}{\nu_{ni}}\right)^2\right]^{-1} \tag{6a}$$

$$K_2 = \frac{2\Omega}{\nu_{ni}} \left[1 + \left(\frac{2\Omega}{\nu_{ni}}\right)^2\right]^{-1} \tag{6b}$$

At high latitudes we can assume that $\underline{\Omega}$ is vertical. Furthermore we put $\underline{v}_\perp = 0$ such that the expression for the horizontal neutral wind reduces to

$$\underline{u}_\perp = K_1 \underline{f} - K_2 \underline{\Omega} \times \underline{f}/\Omega \tag{7}$$

where

$$\underline{f} = - \frac{1}{\rho} \nabla p \tag{8}$$

is the driving force due to pressure gradients. The angle between the horizontal wind and the driving force will be given by (see Figure 3)

$$\mathrm{tg}\theta = \frac{2\Omega}{\nu_{ni}} \tag{9}$$

The direction of $\underline{u}$ will be to the right of the driving force at northern latitudes. As $\nu_{ni} \approx N_i$ this angle will decrease by increasing N_i (ion density). For an ion density of 10^{12} m^{-3} the angle will be $\sim 12°$. For an ion density of 3×10^{11} m^{-3} the angle will be $\sim 35°$. For densities of the order 10^{10} m^{-3} or less the angle will be close to 90° or the wind is approximately geostrophic. For such low ion densities, however, the deletion of the viscous drag term is more critical. The magnitude of the horizontal neutral wind is in this approximation given by

$$|\underline{u}_\perp| = \frac{1}{(\nu_{ni}^2 + 4\Omega^2)^{\frac{1}{2}}} |\underline{f}_\perp| \tag{10}$$

which shows that $\underline{u}_\perp$ increases toward the limit

$$|\underline{u}_\perp|_{as} \approx \frac{1}{2\Omega} \left(\frac{1}{\rho} \nabla p\right) \tag{11}$$

when the collision frequency or the ion density decreases. The maximum value of $\frac{1}{\rho} \nabla p$ in the ionosphere below 300 km is about $4 \cdot 10^{-2}$ m/s^2 (Kohl and King, 1967) and therefore the maximum asymptotic velocity will be approximately 300 m/s.

Most of the earlier models like the one by Kohl and King (1967) are usually limited to lower latitudes and take little or no account of the special situation in the polar regions such as very strong electric fields and excessive heat input due to the auroral electrojet and particle precipitation.

4. EARLY MEASUREMENTS OF THE POLAR IONOSPHERE NEUTRAL WIND

When Stoffregen (1972) made his barium releases above Kiruna the neutral cloud did not move in accord with what the theories predicted, and he inferred that in addition to the global pressure system there must be a local pressure system in the auroral zones, modifying the neutral wind motion at high latitudes. In an attempt to explain his observations he modified the model of Jacchia (1965) to include a high pressure cell at nighttime in the auroral zone (Figure 4).

Rees (1971) (Figure 5) showed from a sample of rocket releases above Kiruna that the east-west component of the neutral wind velocity vector was very well correlated with the mean value of the north-south component of the magnetic perturbation vector for the 2 hours prior to launch. This result strongly indicates that the neutral motion is affected by the auroral electrojet through the ion drag coupling. Rees (1971) also inferred electric fields from his measurements and concluded that the ions accelerated by the ionospheric electric field will represent high enough momentum to push the neutral air with a time constant of one to two hours.

The pioneer results by Rees (1971) and Stoffregen (1972) disagree in the sence that they explain their measurements in terms of two different driving forces, ion drag and local pressure gradients respectively. On the other hand they complement each other as both these two effects most likely play an important part in the upper polar ionosphere dynamics.

One more important conclusion that could be drawn out of the work by Rees (1971) and Stoffregen (1972) was the fact that the neutral air is not at all stationary at E-region heights in the polar ionosphere.

5. LONG TIME SERIES OF NEUTRAL WIND MEASUREMENTS

To interprete single rocket measurements of the neutral wind at auroral heights is of course very difficult since the forces that control the neutral air motion are present over very long distances and during long periods. Therefore a measurement of the neutral wind at one point in space is related to the time history of the neutral air dynamics in a large area during periods of hours. The ideal way to master this problem would be to observe the neutral wind at many different places over long periods. This is of course not economically feasible but a step in the right direction is to measure the wind over a long period at one place. This was made possible when an incoherent scatter radar was moved to auroral zone latitudes in 1972. From the Chatanika radar measurements one has been able to derive a pseudo neutral wind which is a weighted height average of the E-region neutral wind (Brekke et al., 1973).

The neutral wind is actually derived as a residual by solving a simplified ion mobility equation where a balance is assumed between the Lorenz force and the collision term

$$\frac{e}{m_i}(\underline{E} + \underline{v} \times \underline{B}) - \nu_{in}(\underline{v} - \underline{u}) = 0 \tag{12}$$

As the incoherent radar measures a height independent E-field and a weighted height-average value of the ion velocity $\underline{v}$ one has to account for this in the measurement and therefore the residual is a weighted average of the neutral wind in the E-region (Brekke et al., 1973).

A summary of such measurements is shown in Figure 6 where one can see that the flow pattern in the daytime is directed antisunward across the polar cap as predicted by the Kohl and King (1967) model. In the night sector the wind appears to deviate towards west before local midnight and towards east after midnight as if the ion flow is pushing the neutrals along approximately in the $\underline{E} \times \underline{B}$ direction. Especially around midnight there are details in the flow pattern that does not agree with a simple ion drag component to the neutral wind, particularly is this the case for the strong northward component before midnight.

If one now returns to the simplified expression for $\underline{u}$ and neglect the pressure gradients with respect to the ion velocity the equation for the horizontal neutral wind reduces to

$$\underline{u}_\perp = K_1 \underline{v}_\perp - K_2 \underline{\Omega} \times \underline{v}_\perp/\Omega \tag{13}$$

The angle between the neutral wind velocity and the ion velocity will now be given by (Figure 4)

$$tg\theta' = \frac{2\Omega}{\nu_{ni}} \tag{14}$$

where $\underline{u}_\perp$ is directed to the right of $\underline{v}_\perp$. During the evening hours when the electric field is very strong and northward directed the ion velocity has a strong westward component. If one then neglects the effect of the pressure gradients the neutral wind should have a northward component. In the morning hours the precipitation is usually very strong and therefore the neutral-ion collision frequency will be larger and reducing this effect. Some of the effects seen around midnight from the Chatanika data in the neutral wind therefore may be due to the effect of the Coriolis force turning the ion velocity into a northward component which couples to the neutrals through the ion drag term.

6. RECENT MEASUREMENTS OF THE NEUTRAL WIND IN THE AURORAL AND POLAR CAP IONOSPHERE

The incoherent scatter technique has recently been improved to give neutral wind measurements with a height resolution of a few kilometers. One of the first results from this technique is shown in Figure 7 (Rino et al., 1977) where the neutral wind is shown for 3 different heights in the E-region. One will notice that there is a strong meridional flow antisunward before midnight and sunward after midnight in this particular experiment. This flow pattern is not well understood but tentatively it is explained as being due to strong local heat sources in the auroral atmosphere. The amount of Joule heat input was as high

as 100 ergs/cm^2/sec during periods of this event. One inherent difficulty in the derivation of the neutral wind is the choice of a proper model neutral atmosphere. So far we have used only static models (Rino et al., 1977) but with the large amount of energy released during auroral substorms in the upper atmosphere it is likely that the neutral air density and collision frequency can be modified.

Kelly et al. (1977) have made measurements of the ion drift and the neutral wind in the F-region polar cap by barium and strontium releases. They found that the neutral wind velocity observed are too large to be explained by the global thermospheric pressure gradients. Therefore they infer that either local heat sources are present in the auroral oval, or that high ion velocities coupled with enhanced plasma density due to particle precipitation in the polar cap put the neutral air in motion at higher latitudes. Further they claim that the inertia of this motion could sustain even when the plasma density has decayed sufficiently that ions can no longer affect the neutrals.

Smith and Sweeney (1980) has derived neutral wind patterns in the polar cap F-region by the Fabry-Perot technique. Their results which are shown in Figure 8 are in remarkable good agreement with the Kohl and King (1967) model, showing a very uniform antisunward flow across the polar cap.

7. RECENT MODELS OF THE GLOBAL WIND PATTERN

Recently several authors have published extensive model calculations of the global neutral air wind pattern at thermospheric heights.

The basic content of these calculations is that the auroral zone heat input plays an important part in the global wind system in the winter hemisphere, such as in the model published by Roble et al. (1979) (Figure 9). Above 120 km a local wind cell is set up which modifies the symmetric wind system down to middle and low latitudes.

The validity of such models must be tested by future experiments, and today we stand on the threshold of a challenging future in this field. We are now mastering a few good techniques which make it possible to monitor the wind from hour to hour and within a few years time we most likely will experience important progress in this research.

8. CONCLUSION

The behaviour of the neutral wind in the polar ionosphere is not well understood. A few measurements do exist of this phenomenon which on one hand indicates the presence of local heat sources in the auroral zone and on the other a strong influence on the ion drag force. A better understanding of the neutral dynamics in the polar ionosphere is needed since it is generally accepted that the polar cap and auroral zone disturbances influence the neutral atmosphere dynamics at lower latitudes on a global scale. Furthermore the neutral dynamics at auroral heights are coupled to the magnetospheric processes and to the interplanetary space via the magnetic field. It is not unrealistic to believe that a better understanding of the sun-earth relationship can be obtained by studying the neutral atmosphere dynamics at high latitudes.

REFERENCES

BREKKE, A., J.R. DOUPNIK, and P.M. BANKS, 1973, "A preliminary study of the neutral wind in the auroral E-region", J. Geophys. Res., Vol. 78, 34, pp. 8235-8250.

JACCHIA, L.G., 1965, "The temperature above the thermosphere", Space Research V, p. 1152, North-Holland, Amsterdam.

KELLY, M.S., T.S. JØRGENSEN, and I.S. MIKKELSEN, 1977, "Thermospheric wind measurements in the polar region", J. Atmos. Terr. Phys., Vol. 39, pp. 211-219.

KOHL, H. and J.W. KING, 1967, "Atmospheric winds between 100 and 700 km and their effects on the ionosphere", J. Atmos. Terr. Phys., Vol. 29, pp. 1045-1082.

MAEDA, H. and M. FUJIWARA, 1967, "Lunar ionospheric winds deduced from the dynamo theory of geomagnetic variations", J. Atmos. Terr. Phys., Vol. 29, pp. 917-936.

REES, D., 1971, "Ionospheric winds in the Auroral Zone", J. Brit. Interpl. Soc., Vol. 24, pp. 233-246.

RINO, C.L., A. BREKKE, and M.J. BARON, 1977, "High resolution auroral zone E-region neutral wind and current measurements by incoherent scatter radar", J. Geophys. Res., Vol. 82, 16, pp. 2295-2304.

ROBLE, R.G., R.E. DICKINSON, E.C. RIDLEY, and Y. KAMIDE, 1979, "Thermospheric Response to the November 8-9, 1969, Magnetic Disturbances", J. Geophys. Res., Vol. 84, 8, pp. 4207-4216.

SMITH, R.W. and P.J. SWEENEY, 1980, "Winds in the thermosphere of the northern polar cap", Nature, Vol. 284, pp. 437-438.

STOFFREGEN, W., 1972, "Anomaly of the neutral wind at ∝ 200 km height at high latitudes", in Magnetosphere-Ionosphere Interactions (ed. K. Folkestad), Universitetsforlaget, Oslo, pp. 83-86.

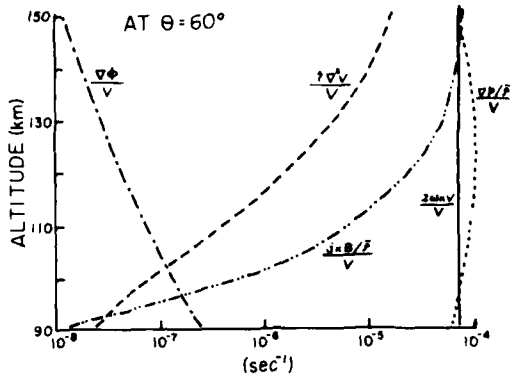


Figure 1 A comparison of pressure ($\nabla p/\rho$), tidal ($\nabla\phi$), Coriolis $2(\underline{\omega} \times \underline{v})$, ion drag ($j \times \underline{B}$) and viscous ($\eta\nabla^2 v$) drag terms in the equation of motion normalized by the velocity v. (After Maeda and Fujiwara, 1967).

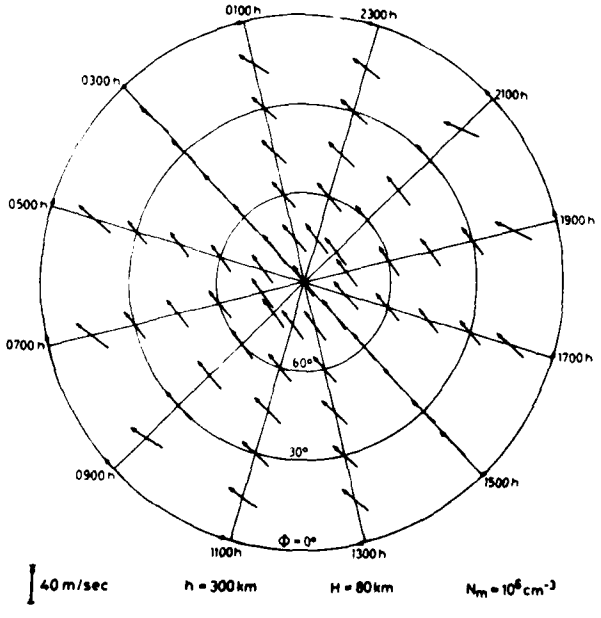


Figure 2 The atmospheric wind system in the northern hemisphere calculated for an altitude of 300 km when the peak electron density is 10^{12} m^{-3}. (After Kohl and King, 1967).

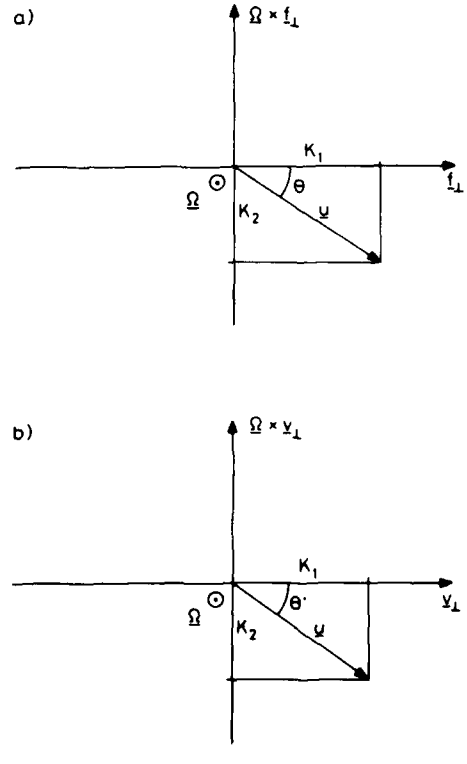


Figure 3 Schematic sketches showing the relationship between a) the neutral wind velocity and the pressure gradient force when the ion velocity is neglected, b) the neutral wind velocity and the ion velocity when the pressure gradient is neglected. The relationship is derived when a steady state is obtained between the pressure force, the Coriolis force and the ion drag term.

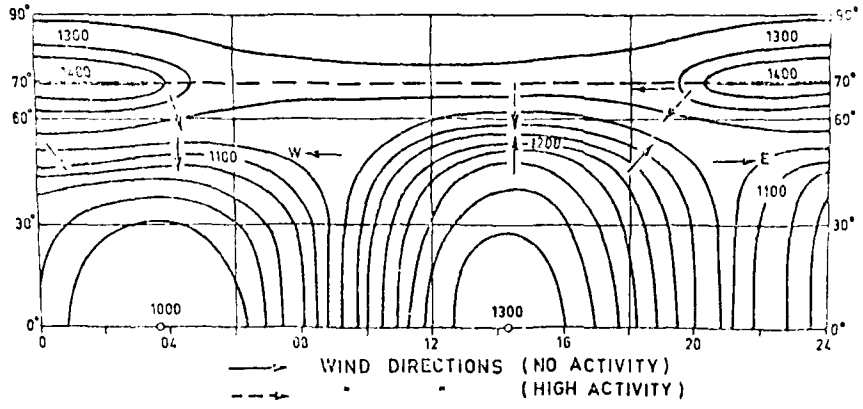


Figure 4 Temperature distribution, when an additional heat source at the nightside of the auroral zone is introduced. The change of wind directions is indicated by arrows. (After Stoffregen, 1972).

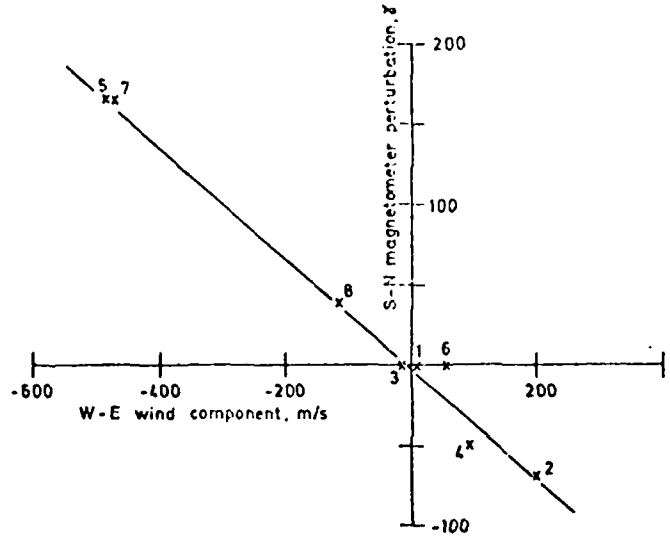


Figure 5 W-E neutral wind component (Above 150 km generally) plotted against S-N ground magnetometer perturbation. (Rees, 1971).

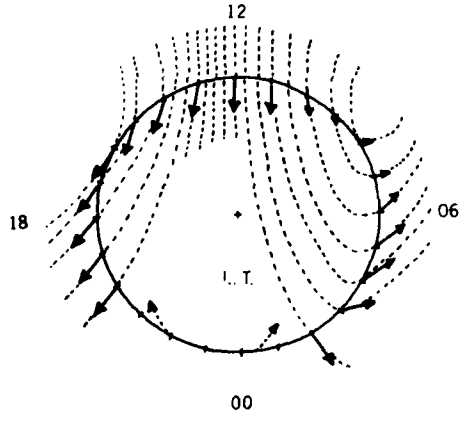


Figure 6 Horizontal wind velocities derived at Chatanika. The geographic north pole is located at the center of the circle, which represents the geographic latitude of Chatanika. Local time being as indicated. An arrow toward the center represents a geographic northward motion, and a counterclockwise one is geographic eastward. (Brekke et al., 1973).

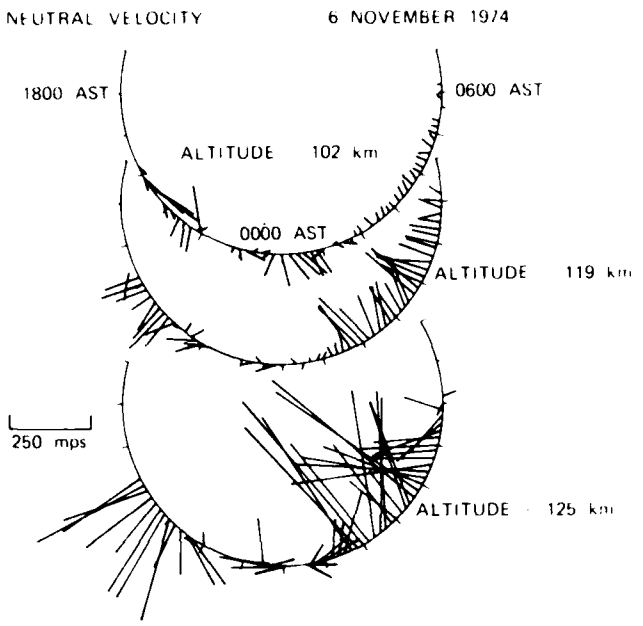


Figure 7 Neutral wind velocity vectors at 3 different altitudes measured by the Chatanika radar November 6, 1974. The format is as explained in Figure 6 except the scale of the wind vectors is given on the side. (Rino et al., 1977).

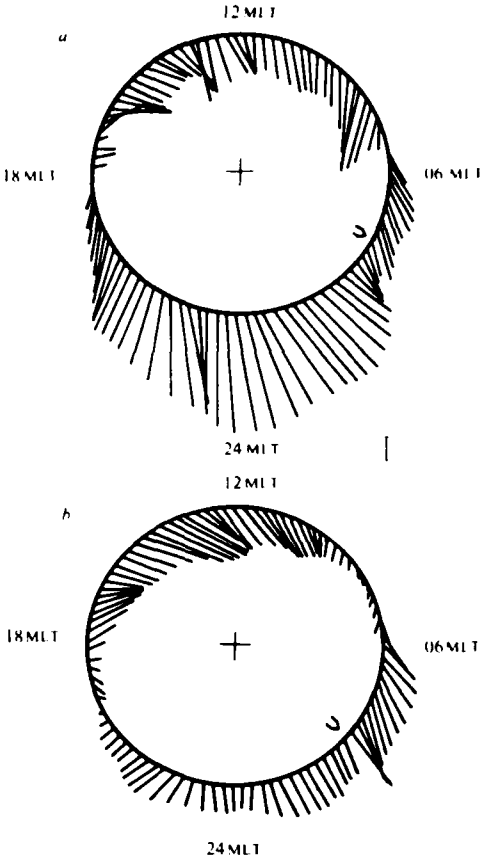


Figure 8 A 24-h plot in geomagnetic coordinates showing vectors representing the neutral wind in the polar thermosphere during a) 27 January 1979, b) 21 January 1979. The circle is at a geomagnetic latitude of 73.1°N, traced out by the observing site during the period of observation. The vectors are drawn with their tails on the circle. Scale bar, 100 m/s. (Smith and Sweeney, 1980).

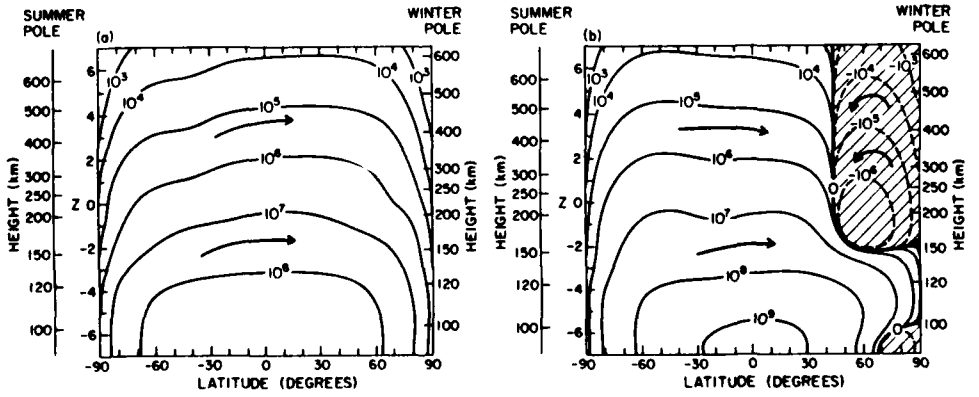


Figure 9 The calculated contours of mass flow stream function (g s^{-1}) for a) solar heating alone, b) solar heating plus the OGO 6 inferred momentum source, c) solar heating plus the OGO 6 inferred momentum source and symmetrical high-latitude heat source, and d) solar heating plus the OGO 6 momentum source and the asymmetrical high-latitude heat source. (After Dickinson et al., 1977).

ELECTRICAL COUPLING BETWEEN THE LOWER ATMOSPHERE AND THE IONOSPHERE

Richard A. Goldberg
NASA/Goddard Space Flight Center
Laboratory for Planetary Atmospheres
Greenbelt, Maryland 20771
USA

ABSTRACT

The classical model for the global electrical circuit depicts a current flow generated by the estimated 1500-2000 electrical storms in progress at any instant of time. This system is often thought to be self-contained, neither affecting nor affected by changes in the electrical properties of the magnetosphere and ionosphere, because of a highly conductive electrical shield (the electrosphere) beginning near the base of the ionosphere. Mounting experimental evidence requires reevalution of these traditional concepts. For example, ionospheric signatures induced by tornadoes and hurricanes are well established, and probably occur through the propagation of gravity waves. VLF spherics initiated by thunderstorm lightning have also been associated with the subsequent stimulation of magnetospheric particle precipitation. In the reverse direction, modulations in the lower atmospheric electrical structure (e.g. fair weather electric field, ionospheric potential, air-earth current density, and possibly thunderstorm activity) appear to correlate with solar activity and geomagnetic events. This evidence for electrical coupling between the upper and lower atmosphere is reviewed, including theoretical implications regarding the possibility, magnitude, and importance of such effects.

1. INTRODUCTION

Traditionally, the upper and lower regions of the Earth's atmosphere have been studied on an independent basis. Within each domain, the various competing physical and chemical processes have been treated as a self-contained system governed externally by radiations of solar induced origin. These views are now under refinement, as we begin to learn more about the various and often subtle interactions which must occur between the upper and lower atmosphere to explain many of the correlative observations between the two regions.

An important subset of upper-lower atmospheric interactions are those coupling the ionosphere and troposphere. For nearly 30 years, various experimentalists have accumulated evidence to indicate ionospheric response to meteorological influences. Now, new results suggest possible interactions in the reverse direction, i.e., those where ionospheric events may influence tropospheric electrical and meteorological structure. Of particular interest is the role that atmospheric electrical structure (particularly in the middle atmosphere, ~15-90 km) plays in the transmittance of stimuli for triggering remote responses in either direction. Electrical coupling (between the ionosphere and troposphere) is appealing because it occurs nearly instantaneously and does not require the exchange of large quantities of energy. However, in addressing this concept, it has been necessary to reevaluate some of the classical concepts regarding atmospheric electricity. Several recent documents including Dolezalek, 1978; Herman and Goldberg, 1978a,b; an important workshop report edited by Maynard (1979); and papers by Markson (1978) and Markson and Muir (1980) have discussed the significance and implications of electrical coupling between the ionosphere and troposphere, and have suggested methods to proceed toward a better insight in this provocative field.

In this short review, we discuss the classical concept of atmospheric electricity, and then proceed to survey the evidence for electrical coupling; first for ionospheric response to meteorological phenomena (although most of these probably involve energy transfer through a non-electrical process such as gravity waves) and next, for tropospheric response to ionospheric phenomena. Violations to the classical concept of atmospheric electricity are mentioned where appropriate. New findings regarding the middle atmosphere and the possible effect it may have on regulating the above exchanges are also considered. Finally, recent theoretical and experimental findings concerning coupling mechanisms are reviewed.

2. THE CLASSICAL ATMOSPHERIC ELECTRICAL CIRCUIT

In most major texts on atmospheric electricity (e.g. Israel 1970, 1973), it is specified that the global electrical current is driven by global thunderstorm activity (about 1500 to 2000 storms at any given instance of time), which provides a current of approximately 0.5 to 1 amp per storm. Figure 1 depicts the general circuit characteristics. The current is thought to flow upward to the lower ionosphere (electrosphere) where the relatively high conductivity causes the current flow to diverge horizontally and return through the fair-weather regions. The most popular parameters to define the return flow include the return current (air-Earth current density, J_c), the atmospheric vertical potential gradient (or vertical electric field, E), and the total ionospheric potential (potential difference between the ionospheric and ground, V_i). The atmospheric conductivity is often made to increase exponentially with altitude which implies that most of the potential drop (V_i) occurs in the lowest 5-10 km of the atmosphere. Hence, within this conceptical framework, mesospheric and stratospheric fields are small and can have little influence on the circuit characteristics. This classical model may now be subject to some modification based on recent findings to be discussed within this text.

The thunderstorm driver concept is largely deduced from indirect measurements (e.g. Figure 2) wherein the observed daily (UT) variation in vertical potential gradient appears as the envelope of the peaked activity for the three major continental thunderstorm regions. This assumption has not been verified with a suitable global mapping study since, at present, there is no adequate technique available for establishing the variation of thunderstorm activity on a spatial or temporal basis. Both of these are required to evaluate the total electrical budget. In favor of the thunderstorm driver concept are calculations which imply a complete decay of the global electrical circuit within 1 hour following the sudden disappearance of all thunderstorms (e.g. Markson and Muir, 1980).

In the classical sense, the basic circuit cannot be subject to perturbations from external sources, since the ionosphere acts as an "infinitely" conducting spherical shield. However, a wealth of correlative information exists to suggest a direct link between tropospheric electric field structure and solar activity related phenomena. In addition, recent calculations by Sugiura (cf. Goldberg, 1979) show that magnetospheric and atmospheric electric fields are of similar magnitude. Since the ionosphere is not an infinite conductor in actuality, we might suspect penetration of the external magnetospheric electric fields into the middle and lower atmosphere, and vice versa. These ideas provide the foundation for newly emerging concepts which argue the existence of significant electrical coupling between the ionosphere and troposphere, including the existence of competitive external drivers for the global circuit.

3. IONOSPHERIC RESPONSE TO THE TROPOSPHERE

We first consider influences of hurricanes, tornadoes, and other tropospheric meteorological events on the ionosphere. These do not directly involve the atmospheric electrical circuit, but offer competition with those processes which do. This upward coupling is usually understood in terms of direct energy transfer provided by the energy rich tropospheric phenomena. For example, Gherzi (1950) and Bauer (1958) demonstrated ionospheric F region ionization enhancements above hurricane passages. Apparent ionospheric responses to cold fronts were also noted by Gherzi (1950), Bauer (1957), and Arendt and Frisby (1968). To explain the observational results, Bauer (1957) hypothesized a dynamic coupling process between the troposphere and ionosphere.

Acoustic waves generated by several local storms have been detected in the F region by Georges (1968), Baker and Davies (1969), Hung et al. (1975), and Hung and Kuo (1978). The waves, whose periods in the ionosphere are in the range 2-5 minutes, are detected by CW Doppler techniques. In more recent work, Hung et al. (1979 a,b) have found a gravity wave relationship between tornadoes and Doppler soundings. They have also demonstrated the importance of using simultaneous GOES IR satellite data to detect low temperature cloud regions where high convective cloud growth rates and tornado touchdowns are likely to occur (Hung et al., 1980).

Evidence for gravity waves above thunderstorms has been detected by experiments with the Arecibo Radar Observatory in Puerto Rico (Larsen et al, 1980). Figure 3 illustrates a sample of three sequential profiles taken above a cell on September 14, 1979. The region up to 15 km (tropopause) is thought to include the convective cell. Above the tropopause, particularly between 18 and 21 km, are indications of waves with a wave length of 6 km, and a period of approximately 8.6 minutes. These are shown to fit theoretical predictions for locally produced gravity waves.

Tropospheric gravity waves may do more than simply perturb the ionosphere. Hines and Halevy (1977) have proposed a gravity wave feedback mechanism, whereby energy produced in the troposphere by meteorological phenomena propagates upward via gravity waves. Under certain assumed conditions which could be modulated by solar activity, the upper atmosphere will reflect the waves downward to interfere constructively or destructively.

Lightning can also induce effects in the middle atmosphere and above. For example, whistlers, which are VLF waves thought to be induced by lightning discharges, are guided along the Earth's magnetic field between conjugate points for path lengths exceeding 30,000 to 50,000 km (e.g. Davies, 1965). Recently, Bering et al. (1980) have measured microbursts of energetic electron precipitation from the magnetosphere, stimulated by VLF spherics generated by lightning in a thunderstorm cell. Since the precipitating particles can ionize and dissociate molecules in the upper atmosphere, a lightning induced feedback mechanism for perturbation of the upper atmosphere is created.

4. DOWNWARD ELECTRICAL COUPLING

This section samples the various correlations which exist to provide evidence for downward mapping of electrical influences. The coupling may occur through changes in localized ionization, which can alter stratospheric and mesospheric chemical processes. The ionization can also affect the production of aerosols and other macroscopic or submacroscopic particles. The ionizing radiations (both electromagnetic and corpuscular) can affect ion size and mobility, thereby modifying ion-neutral drag and other transport processes. In addition, deviations of the vertical and horizontal atmospheric electric fields from the average quiet norm (fair weather electric field) may also play an important role in processes governing interactions between regions.

The ultimate goal, to determine how changes in atmospheric electrical structure can alter meteorological conditions in the troposphere, remains to be demonstrated. However, section 4.2 briefly considers one aspect of this problem, e.g. the relationship of thunderstorms with solar activity. Finally, section 4.3 addresses some recent theoretical and experimental findings which may have bearing on the physical processes involved.

4.1. Electric Circuit Response to External Sources

The mapping of solar related phenomena to the atmospheric fair-weather electric circuit has been observed through correlative studies relating on short time scales, to solar (interplanetary magnetic field) sector boundary passages and solar flares, and on the long-term to the 11-year sunspot cycle. The parameters usually studied are J_c, E and V_i.

Fair-weather measurements of E and J_c over an 11-year period at Zugspitze, Germany, exhibit response to sector boundary crossings (Reiter, 1976; 1977). Reiter's superposed epoch analysis results (Figure 4) show that both E and J_c increased by 20% or more on the 2 days following a -/+ (+ implies outward bound interplanetary magnetic field) sector boundary crossing in maximum solar activity years, but on the same day as +/- crossings. A similar analysis (Figure 5) of a 10-month period by Park (1976) indicates that the potential gradient at Vostok, Antarctica, increased sharply by 20-30% beginning 3 days after boundary

passages, and the response was similar for +/- and -/+ boundary crossings. The difference in response time for the Vostok and Zugspitze electric fields remains unclear, although new modeling results by Roble and Hays (1979) may account for geographical differences in temporal lag.

For solar flares, ground measurements at Mauna Loa, Hawaii have shown that E and J_c increase significantly after solar flare eruptions. The potential gradient maximized 3-4 days after the flare day, while the air-earth current density reached a maximum after 1 day; both quantities remained above normal for several days thereafter (Cobb, 1967). Reiter (1969, 1971, 1972) has shown that E and J_c measured at Zugspitze also increased following solar flare eruptions. Balloon measurements in the Arctic (Holzworth and Mozer, 1979) and the Antarctic (Cobb, 1978) have revealed that the potential gradient at 30 km decreases severalfold following solar flares.

Potential gradient enhancements have been observed to be correlated with solar radio noise bursts (associated with flares) at Zugspitze (Reiter, 1972) and at stations within the Arctic Circle (Sao, 1967). Marcz (1976) found enhancements in E of 30-50% in Poland (approx. 51°N) associated with geomagnetic storms (which generally follow solar flare eruptions) and Tanaka et al. (1977) saw similar enhancements in the Antarctic associated with auroral substorms. Recently Markson and Muir (1980) have correlated ionospheric potential with solar wind velocity to imply a closer connection with the sun (rather than the magnetosphere) as the primary source for such interactions.

Early long-term measurements of the fair-weather electric field suggested a stronger average field strength in years of maximum sunspot activity compared to minimum years at some European stations, but no apparent influence at others (c.f., the review by Herman and Goldberg, 1978b, p. 139). According to Muhleisen (1971), the total ionospheric potential measured over Germany for a complete solar cycle exhibits a positive correlation with annual mean sunspot number. To the extent that the total potential is proportional to global thunderstorm activity, one might therefore expect the latter quantity to be correlated with the 11-yr. sunspot cycle.

4.2. Thunderstorm Responses

In contrast to the classical electrical picture that thunderstorms drive and control the fair weather electrical circuit, there is some emerging evidence that external influences may help modulate thunderstorms. At this time, the results are largely statistical and in many cases, questionable or controversial (Pittock, 1978; Herman and Goldberg, 1978b).

Markson (1971) was the first to observe interplanetary magnetic sector boundary response and based his study on U.S. thunderstorm data collected during solar cycle minimum period covering approximately 52 sector boundary crossings. There seemed to be a definite preference for thunderstorms to occur from about 1 day before to one day after a +/- crossing, but no statistically significant response to -/+ crossings. A more detailed analysis was made by Lethbridge (1979), who used a daily index derived from 30 years (1947-1976) of thunderstorm data from 102 weather stations in the United States. In a superposed epoch analysis of three separate 10 year sets of data, she found the strongest solar signal in the thunderstorm index for the winter months (Nov.-Mar.) in the latitude band 40-45°N, with peak activity occurring 1 day after +/- boundary crossings. With the index combined for all seasons and latitudes, and for both polarity crossings, Lethbridge could find no discernible response to sector boundaries.

Several different measures of thunderstorm activity have been utilized in studies of the response to solar flare eruptions, and all seem to show a positive response but with varying lag times. For example, Reiter (1969) used spherics counts in Germany (indicative of lightning flashes in thunderstorms within a 300-mile radius), and found a 57% increase in count rate peaking about 4 days after the eruption of solar flares with importance ≥ 2. Other results include Bossolasco et al. (1973), who used the number of long-range radio direction-finding fixes recorded in the Mediterranean area as a measure of thunderstorm activity in that area for the 1961-1971 period, and showed a 50% increase in the daily number of fixes beginning one day and peaking 4 days after solar flare occurrences. VLF whistler counts were used by Holzworth and Mozer (1979) as an indicator of high-latitude thunderstorm activity during the August 1972 solar-terrestrial events. They found a strong increase in count rate beginning about 12 hours after the class 3 flares of August 4 and August 7.

Finally, long-term studies of the variability of the annual number of thunderstorm days in different regions with the 11-yr solar cycle are replete with contradictions. The analysis of Siberian thunderstorms by Septer (1926) has been cited often as proof that middle-to-high latitude thunderstorm occurrence is directly correlated with annual mean sunspot number (correlation coefficient = +0.9). This result, however, has been severely criticized by Pittock (1978). Brooks' (1934) analysis from 57 years of England and Wales data suggests that annual mean sunspot numbers and annual number of thunderstorm days are uncorrelated, but Stringfellow (1974) found a strong positive correlation for the same region (Figure 6) based on English thunderstorm occurrence in the years 1930-1973.

4.3. Theoretical Considerations and Recent Findings

Several theories seeking to explain the thunderstorm (lightning) response to solar activity have been proposed recently, but none bridge the final gap to explain how modification in the electrical structure of the atmosphere can affect thunderstorm occurrence. Presumably, this might occur through electrical influence on cloud growth, etc. (e.g., Dolezalek, 1978). Ney (1959) suggested that thunderstorm activity may be modulated by solar variability through alteration of the electrical state of the middle and lower atmosphere. Markson (1971, 1975, 1978) amplified this idea to evolve a theory which its present state (1978) assumes that the electrical resistance of the atmosphere above thundercloud tops (the charging resistor, see Figure 7) will be lowered by enhanced ionization associated with incoming solar particles. The charging current is thereby increased leading to an enhanced ionospheric potential and fair-weather electric field which must adjust globally to the increased charging current.

D'Angelo (1978) theoretically modified the global circuit by introducing a variable emf source representing the ionospheric potential at high latitudes. Fluctuations in this second driver caused readjustments in the global circuit including the fair-weather electric field. He argued that the ionospheric potential is sensitive to solar wind electric fields, magnetospheric fields, etc., thereby introducing a coupling link within the solar-terrestrial system. These results are also consistent with the recent analysis of Sugiura (e.g. Goldberg, 1979) which demonstrates the equivalent magnitudes of the fair-weather and magnetospheric electric fields, and help justify the concept of magnetospheric electric field mapping into the stratosphere (Figure 8). Dejnakarintra (1974) has theoretically mapped magnetospheric fields into the middle atmosphere and found that fields with effective scale sizes of 500 km or more can penetrate to tropospheric depths (Figure 9). New rocket borne experimental results of the atmospheric electrical field (Hale and Croskey, 1979) appear to confirm D'Angelo's concept that middle atmospheric electric fields in the auroral zone are sensitive to auroral phenomena. They find evidence for the existence of large mesospheric electric fields ($\sim$3-4V/m near 60 km) which appear to disappear (or short out) at high latitudes during aurorally disturbed conditions (Figure 10) when strong ionizing radiations due to precipitating particles are present. If this currently unexplained electric field is a reasonably permanent feature on a global basis (Hale, private communication) then the implications for its effect as a modulator or competitive driver of the global circuit are significant.

Roble and Hays (1979) have provided the most quantitative and sophisticated modeling effort to date. They have developed a global model for electrical parameters, which includes as input parameters orographic effects from the Earth's surface, the global thunderstorm distribution (limited to dawn and dusk) as observed from the DMSP Satellite, and the latitudinal distribution of cosmic ray flux. They too, have calculated high latitude effects induced by solar active and/or magnetospheric phenomena, and have found significant perturbation on the global circuit properties.

Finally, Herman and Goldberg (1978a) have considered how cosmic rays and solar protons affect the local environment near thunderstorms, and if modifications in the local conductivity and electric fields can assist lightning generation. For the case of cosmic rays, the changes appear quantitatively reasonable based on the school of thought (c.f., Chalmers, 1967) that an increase in the fair-weather field enhances the probability of thunderstorm formation under appropriate meteorological conditions. Thus, solar-controlled variation in cosmic ray intensity, especially over the 11-yr. cycle, may modulate the fair-weather field and hence the rate of thunderstorm occurrence. The more difficult question concerns how solar protons, which typically are absorbed above 20-25 km, could affect tropospheric electrical structure.

5. CONCLUSIONS

We have reviewed the evidence for interactions between the ionosphere and troposphere as a first step in establishing the role of electrical coupling betwen two regions. Upward processes appear to include the transfer of energy through gravity waves, or through the transmittance of VLF radio waves induced by lightning. In each case, the excitation of the upper atmosphere by tropospheric sources can stimulate feedback processes returning to the troposphere.

Downward coupling involves changes in the global and local electrical structure of the atmosphere, and is often induced by solar activity. These effects are more direct and may possibly bypass the stratosphere. The responses in the atmospheric circuit to local changes are nearly instantaneous, and therefore are looked on with promise for possible influences on weather systems. Newly emerging experimental evidence indicates that atmospheric electric fields, especially at high latitudes and in the upper atmosphere, are responsive to solar and geomagnetic phenomena. Corpuscular radiations (protons, relativistic electrons, bremsstrahlung X-rays, etc.) are known to strongly enhance the local atmospheric electrical conductivity and ionizations at stratospheric and mesospheric altitudes. However, the electric field perturbations are not easily explained by conventional atmospheric electricity considerations.

Theories now postulate that the electric field perturbations regulated by solar activity may be responsible for the observed statistical correlations between solar activity and thunderstorms. They include the concept that modulations in stratospheric electric fields induced by cosmic rays and/or solar protons can in turn affect the local field near thunderstorms, and change the rate of lightning formation. This is offered as an example of how atmospheric coupling may occur electrically.

The final scenario, what happens to the meteorology by changes in atmospheric electrification, is open to speculation. Here, various suggestions have postulated improved conditions for hail formation, water droplet growth and enhanced rain, etc. (cf. for example, Dolezalek, 1978). These concepts await physical validation.

REFERENCES

Arendt, P. R. and E. M. Frisby, 1968, "Possible relation of a specific ionospheric event to simultaneous meteorological data," Nature, 219, 475.

Baker, D. M. and K. Davies, 1969, "F2-region accoustic waves from severe weather," J. Atmos. Terr. Phys., 31, 1345.

Bauer, S. J., 1957, "A possible tropospheric-ionospheric relationship," J. Geophys. Res., 62, 425.

Bauer, S. J., 1958, "An apparent ionospheric response to the passage of hurricanes," J. Geophys. Res., 63, 265.

Bering, E. A., T. J. Rosenberg, J. R. Benbrook, D. Detrick, D. L. Matthews, M. J. Rycroft, M. A. Saunders and W. R. Sheldon, 1980, "Electric fields, electron precipitation and VLF radiation during a simultaneous magnetospheric substorm and atmospheric thunderstorm," J. Geophys. Res., 85, 55.

Bossolasco, M., I. Dagnino, A. Elena and G. Flocchini, 1973, "The thunderstorm activity over the Mediterranean area," Revista Italiana di Geofisica, 12, 21.

Brooks, C. E. P., 1934, "The variation of the annual frequency of thunderstorms in relation to sunspots," Quart. J. R. Meteorol. Soc. 60, 153.

Chalmers, J. A., 1967, Atmospheric Electricity, 2nd edition, Pergamon Press, Oxford.

Cobb, W. E. 1967, "Evidence of a solar influence on the atmospheric electric elementsat Mauna Loa Obsservatory, Monthly Weather Rev., 95, 12.

Cobb, W. E., 1978, "Balloon measurements of the air-earth current density at the South Pole before and after a solar flare," preprint from AMS Conference on Cloud Physics and Atmospheric Electricity, Issaquah, Washington, (p.552).

D'Angelo, N., 1978, "Thunderstorms, ionosphere to ground potential drop, and the solar wind sector structure, U. of Iowa Report, 78-24.

Davies, K., 1965, Ionospheric Radio Propagation, National Bureau of Standards Monograph #80.

Dejnakarintra, J., 1974, "A theoretical study of electrical coupling between the troposphere, ionospheric and magnetosphere, Report #SEL 74-051, Radioscience Laboratory, Stanford University, California.

Dolezalek, H. (Editor), 1978, The Application of Atmospheric Electricity Concepts and Methods to Other Parts of Meteorology, Technical Note #162, World Meteorological Organization No. 507.

Georges, T. M., 1968, "HF Doppler studies of travelling ionospheric disturbances," J. Atmos. Terr. Phys., 30, 725.

Gherzi, E., 1950, "Ionosphere and weather," Nature, 165, 38.

Goldberg, R. A., 1979, "Areas where solar terrestrial coupling may influence or be influenced by the middle atmosphere," in Middle Atmospheric Electrodynamics, ed. by N. C. Maynard, NASA CP-2090.

Hale, L. C. and C. L. Croskey, 1979, "An auroral effect on the fair weather electric field, Nature, 278, 239.

Herman, J. R. and R. A. Goldberg, 1978a, "Initiation of non-tropical thunderstorms by solar activity, J. Atmos. Terr. Phys. 40, 121.

Herman, J. R. and R. A. Goldberg, 1978b, Sun, Weather and Climate, NASA SP426.

Hines, C. O. and I. Halevy, 1977, "On the reality and nature of a certain sun-weather correlation," J. Atmos. Sci., 3, 382.

Holzworth, R. H. and F. S. Mozer, 1979, "Direct evidence of solar flare modification of stratospheric electric fields," J. Geophys. Res., 84, 363.

Hung, R. J. and J. P. Kuo, 1978, "Ionospheric observation of gravity-waves associated with hurricane Eloise," J. Geophys. Res., 45, 67.

Hung, R. J., T. Phan and R. E. Smith, 1979a, "Coupling of ionosphere and troposphere during the occurrence of isolated tornadoes on November 20, 1973, J. Geophys. Res., 84, 1261.

Hung, R. J., T. Phan and R. E. Smith, 1979b, "Case studies of gravity waves associated with isolated tornadic storms on 13 January 1976, J. Appl. Met., 18, 460.

Hung, R. J., T. Phan, D. C. Lin, R. E. Smith, R. R. Jayroe and G. S. West, 1980, Gravity waves and GOES IR data study of an isolated tornadic storm on 29 May 1977, Monthly Weather Review, 108, 70.

Hung, R. J., G. L. Rao, R. E. Smith, G. S. West and B. B. Hensen, 1975, Ionospheric disturbances during severe weather activity, in Proc. Symp. Effects of the Ionosphere on Space Systems and Communications, Naval Research Laboratory.

Israel, H., 1970, Atmospheric Electricity, Volume 1: Fundamentals, Conductivity, Ions, Israel Program for Scientific Translations, Jerusalem NTIS Doc. TT-67-51394/1.

Israel, H., 1973, Atmospheric Electricity, Volume 2: Fields, Charges, Currents. Israel Program for Scientific Translations, Jerusalem, NTIS Doc TT-67-51394/2.

Larsen, M. F., W. E. Swartz and R. F. Woodman, 1980, "Radar observations of the development of tropical thunderstorms and convection cells using Arecibo radar, in "Proceedings of the 19th Conference on Radar Meteorology, Miami Beach, April 1980.

Lethbridge, M., 1979, "Thunderstorm frequency and solar sector boundaries," in Solar Terrestrial Influences on Weather and Climate, July 1978, Columbus, Ohio, ed. by B. M. McCormac and T. Selega.

Märcz, F., 1976, "Links between atmospheric electricity and ionospheric absorption due to extraterrestrial influences, J. Geophys. Res., 81, 4566.

Markson, R., 1971, "Considerations regarding solar and lunar modulation of geophysical parameters, atmospheric electricity, and thunderstorms," Pure Appl. Geophys., 84, 161.

Markson, R., 1975, "Solar modulation of atmospheric electrification through variation of the conductivity over thunderstorms," in NASA SP-366, ed by W. R. Bandeen and S. P. Maran.

Markson, R., 1978, "Solar modulation of atmospheric electrification and possible implications for the sun-weather relationship," Nature, 273, 103.

Markson and M.Muir, 1980, "Solar wind control of the Earth's Field," Science, 208, 979.

Maynard, N. C., (Editor) 1979, Middle Atmospheric Electrodynamics, NASA CP 2090.

Mühleisen, R., 1971, "Neue Ergebnisse und Probleme in der Luftelektrizitat, Zs. Geophysik, 37, 759.

Ney, E. P., 1959, "Cosmic radiation and the weather," Nature, 183, 451.

Park, C. G., 1976, "Solar magnetic sector effects on the vertical atmospheric electric field at Vostok, Antarctica," Geophys. Res. Let., 3, 475.

Pittock, A. B., 1978, "A critical look at long term sun-weather relationships, Rev. of Geophys. and Space Sci., 16, 400.

Reiter, R., 1969, "Solar flares and their impact on potential gradient and air-earth current characteristics at high mountain stations," Pure Appl. Geophys., 72, 259.

Reiter, R., 1971, "Further evidence for impact of solar flares on potential gradient and air-earth current characteristics at high mountain stations," Pure Appl. Geophys., 86, 142.

Reiter, R., 1972, "Case study concerning the impact of solar activity upon potential gradient and air-earth current in the lower troposphere," Pure Appl. Geophys., 94, 218.

Reiter, R., 1976, "The electrical potential of the ionosphere as controlled by the solar magnetic sector structure," Die Naturwissenschaften, 63, Part 4, 192.

Reiter, R., 1977, "The electrical potential of the ionosphere as contolled by the solar magnetic sector structure, result of a study over the period of a solar cycle," J. Atmos. Terr. Phys., 39, 95.

Roble, R. G. and P. B. Hays, 1979, "Electrical coupling between the upper and lower atmosphere," in Solar Terrestrial Influences on Weather and Climate, July, 1978. Columbus, Ohio, ed. by B. M. McCormac and T. Selega.

Sao, K., 1967, "Correlations between solar activity and the atmospheric potential gradient at the Earth's surface in the polar regions, J. Atmos. Terr. Phys., 29, 213.

Septer, E., 1926, "Sonnenflecken and Gewitter in Siberien, "Meteorol. Z., 43, 229.

Stringfellow, M. F., 1974, "Lightning incidence in Britain and the solar cycle, Nature, 249, 332.

Tanaka, Y., T.Ogawa and M.Kodama, 1977, "Stratospheric electric fields and currents measured at Syoma Station, Antarctica – 1. The vertical component." J.Atmos. Terr. Phys., 39, 523.

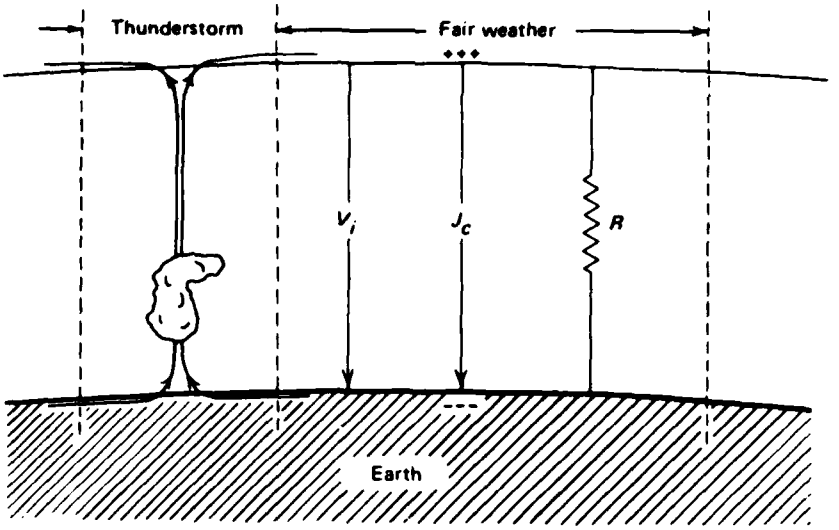


Figure 1. Qualitative model of the global electric circuit, illustrating the thunderstorm generator, ionospheric potential (V_i), air-Earth current density (J_c), and total atmospheric resistance (R) (after Herman and Goldberg, 1978b).

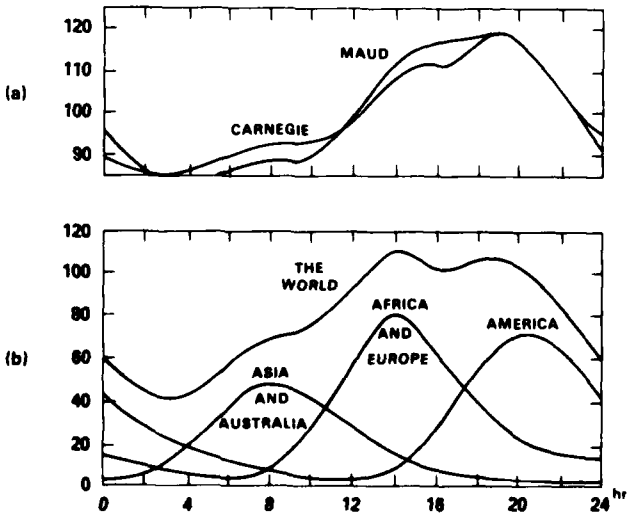


Figure 2. (a) Diurnal variation of the atmospheric vertical potential gradient. (b) Diurnal variation of thunderstorm expectation for individual continents and for the world envelope. The areas covered at any time by thunderstorms is in units of 10^4 km^2 (after Israel, 1973, p. 366).

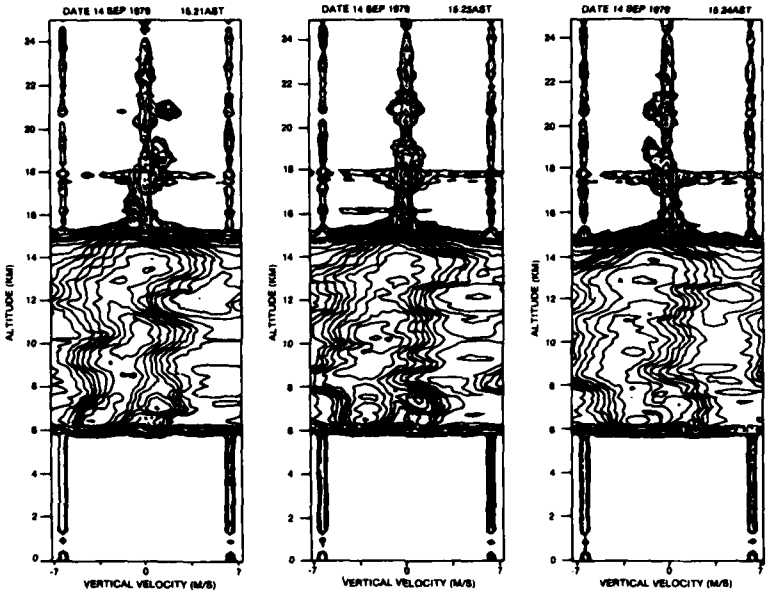


Figure 3. Three consecutive profiles of reflected power as a function of altitude and Doppler shift. The profiles are 82 seconds apart (after Larsen et al., 1980).

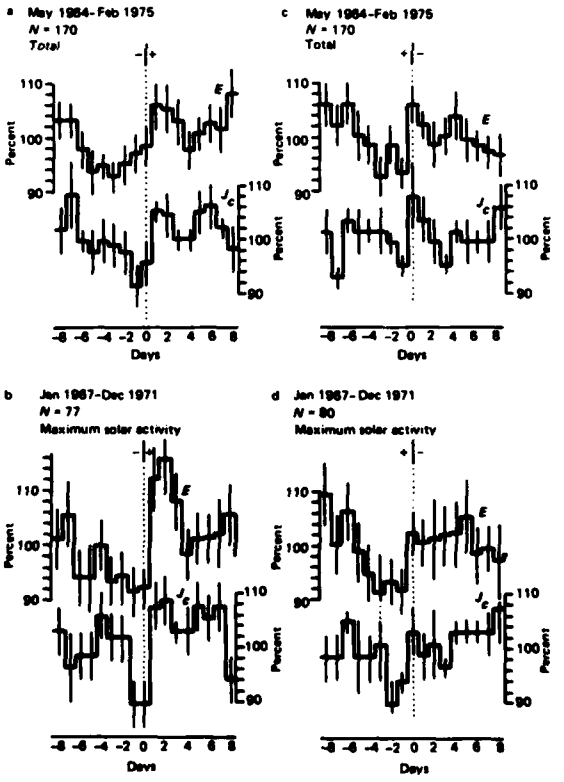


Figure 4. Superposed epoch analysis of fair weather atmospheric electric field (E) and air-Earth current density (J_c) recorded at Zugspitze on days surrounding solar secter boundary crossings on day 0. (a) Data for total solar cycle with negative to positive boundary crossings. (b) -/+ crossings for maximum solar years only. (c) +/-, all years. (d) +/-, maximum years only (after Reiter, 1977).

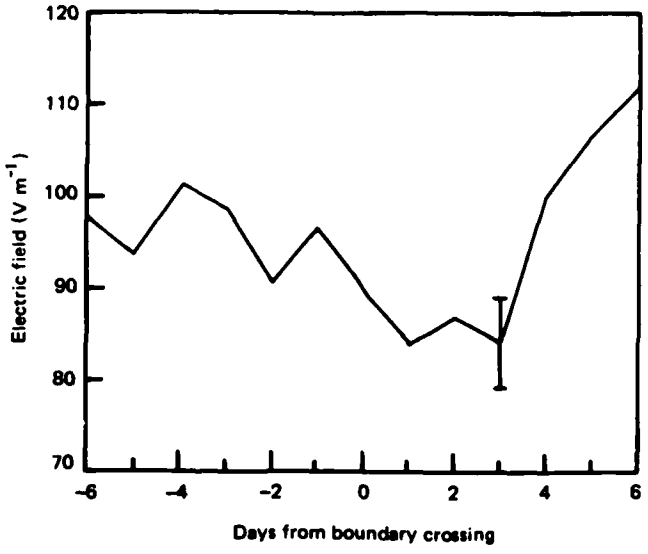


Figure 5. Average behavior of Vostok vertical atmospheric electric field for 17 solar magnetic sector boundary crossings. The standard error is indicated by vertical bar at day 3. (after Park, 1976).

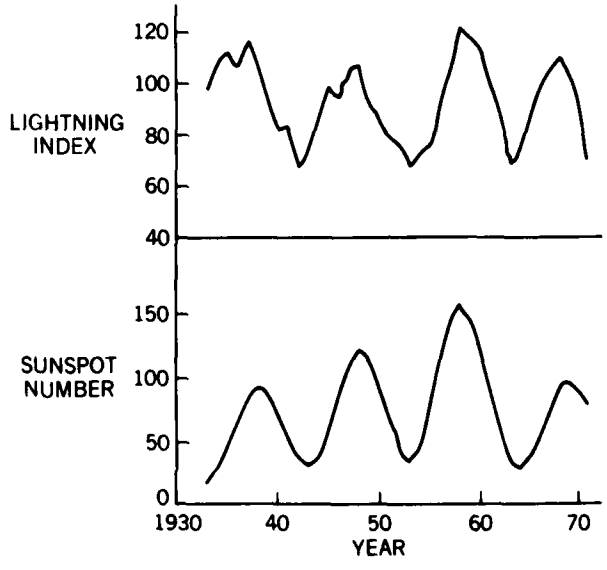


Figure 6. Five year running means of average annual sunspot number and average annual lightning index (after Stringfellow, 1974).

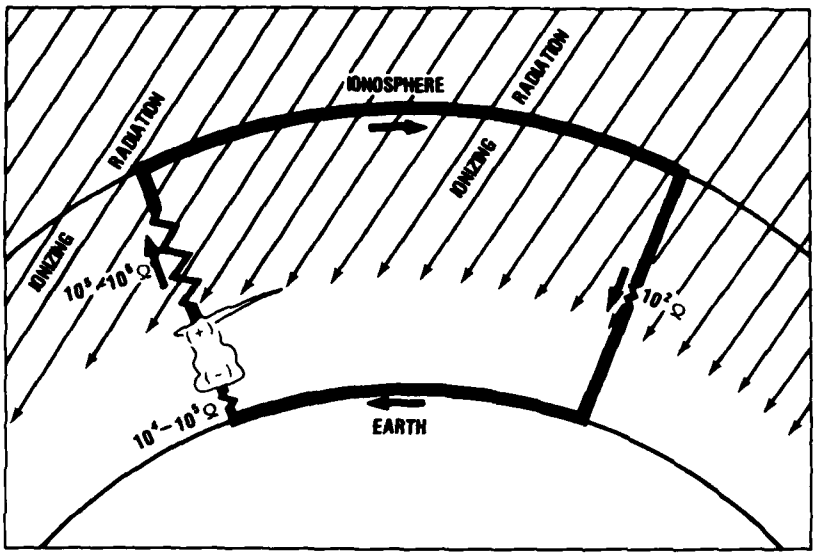


Figure 7. The atmospheric global electrical circuit. Large arrows indicate the flow of positive charge. Estimated resistances of circuit elements are given. The thunderstorm depicted represents the global electrical generator, i.e. the totality of all thunderstorms, and sends current through the charging resistor (10^{8} - $10^{9}\Omega$) to the ionosphere. The cumulative effect of the global return current passes through the load ($10^{2}\Omega$) resistor (after Markson, 1978).

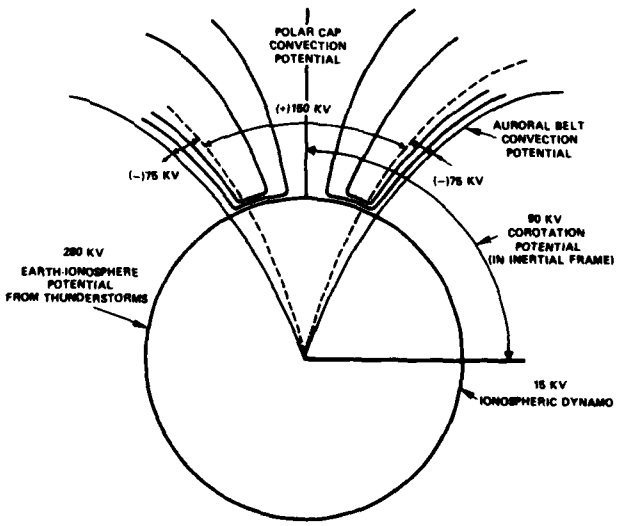


Figure 8. Schematic illustrating sources and magnitudes of electric fields in near Earth space. (Sugiura, pvt. communication).

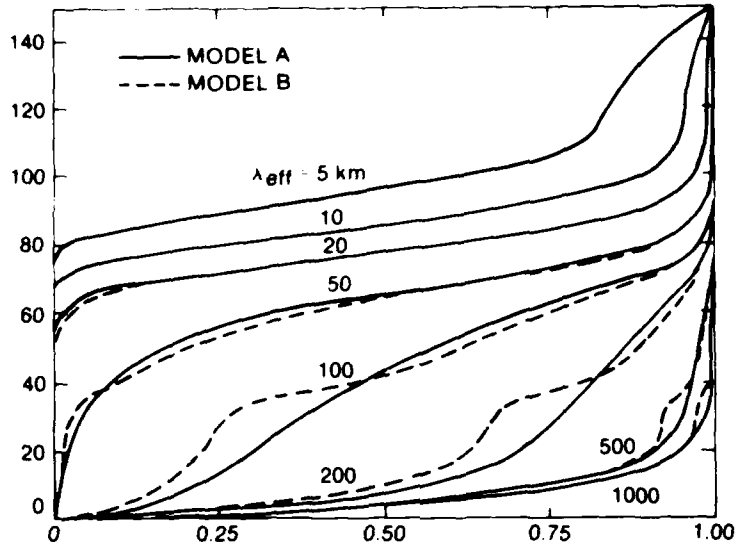


Figure 9. The mapping factor for horizontal electric fields of various scale sizes (λ_{eff}) imposed on the ionosphere at 150 km. Two different conductivity models have been used. The dashed curves correspond to a model with an enhanced conductivity near 40 km. (after Dejankarintra, 1974).

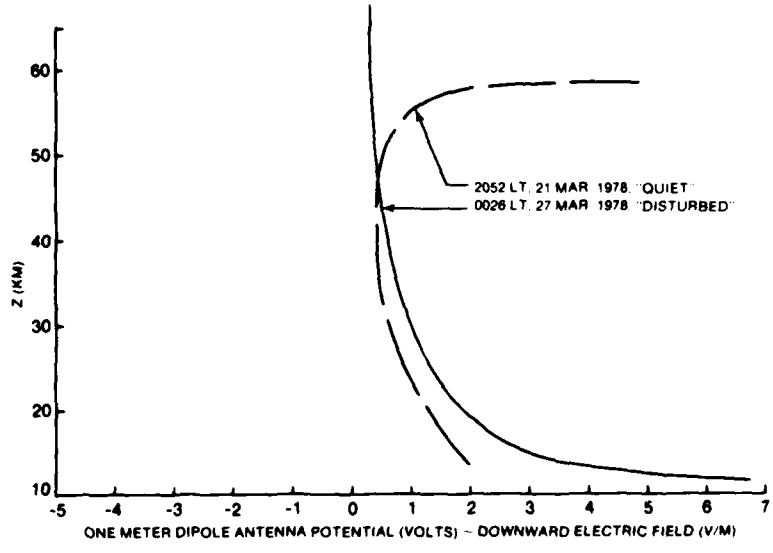


Figure 10. The vertical electric field versus altitude for two parachuted descents under quiet and disturbed conditions. The antenna potential can be approximately equated to the downward electric field (V/m), (after Hale and Croskey, 1978).

FIRST RESULTS FROM THE TROMSØ IONOSPHERIC MODIFICATION FACILITY

P. Stubbe and H. Kopka

Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, Germany

A. Brekke, T. Hansen, O. Holt

Institute of Mathematical and Physical Sciences, University of Tromsø
P.O. Box 953, 9001 Tromsø, Norway

R.L. Dowden

Department of Physics, University of Otago
P.O. Box 56, Dunedin, New Zealand

T.B. Jones and T. Robinson

Department of Physics, University of Leicester
Leicester LE1 7 RH, Great Britain

H.-J. Lotz and J. Watermann

Institut für Geophysik, Universität Göttingen
Herzberger Landstr. 180, 3400 Göttingen, Germany

SUMMARY

The Heating facility at Ramfjordmoen near Tromsø, Norway, has recently been completed. Preliminary experiments have been performed during the construction phase, partly with much less than full power. The present paper gives a brief survey of the first experimental results, relating to D region modification, generation of ELF and VLF waves by polar electrojet modulation, excitation of micropulsations, F region deformation, and anomalous absorption.

1. INTRODUCTION

The Heating facility at Ramfjordmoen near Tromsø, Norway, built by the Max-Planck-Institut für Aeronomie in cooperation with the University of Tromsø, has recently been completed. The facility generates a maximum CW power of 1.5 MW in the frequency range 2.5 - 8.0 MHz. The antenna gain is 24 dBi, corresponding to a beam width of 14.5 deg. The effective radiated power (ERP), i.e. the product of transmitted power and antenna gain, reaches up to 360 MW. For further details see the report by STUBBE and KOPKA (1979).

Preliminary experiments with less, or even much less than full power have been performed since August 1979. These first experiments will be described in the following.

2. D REGION MODIFICATION

The Ohmic dissipation rate of the heating wave maximizes in the D region. Correspondingly, we expect large electron temperature enhancements to occur in the D region. Theoretical estimates are shown in Fig. 1. We notice that x-mode heating leads to higher T_e-values, but in a narrower height range. The maximum occurs around 80 to 85 km, roughly at the altitude where $f_o \pm f_L \cos\Theta = \nu_{en}/2\pi$ (f_o = heating frequency, f_L = electron gyrofrequency, Θ = angle between $\underline{k}$-vector of heating wave and geomagnetic field, ν_{en} = electron-neutral momentum transfer collision frequency, + or - for o- or x-mode, respectively). For smaller heating powers, the collision frequency ν_{en}, which is approximately proportional to T_e, becomes smaller. Thus, the T_e maximum occurs at a lower altitude. The characteristic time for T_e to reach steady state is about 1 msec at 90 km, less below, more above that altitude.

In the upper D region, where the electron chemistry is governed by recombination, the electron density, N_e, is expected to increase with increasing T_e, with a characteristic time of the order minutes. In the lower D region, where electron attachment and detachment dominate, N_e is expected to decrease with increasing T_e, with a characteristic time of the order seconds (TOMKO, FERRARO, LEE and MITRA, 1980).

The very first experiment performed with our Heating facility was on Aug. 29, 1979. Two transmitters, fed into one row of 6 crossed dipoles, generated an ERP of 8 MW at 4.04 MHz, corresponding to 1/45th of full power. The PRE partial reflection facility of the University of Tromsø (HOLT, BREKKE and HANSEN, 1980) was used to investigate D region modification effects. Despite the comparatively low power, very pronounced changes in the PRE amplitudes were found, as shown by Fig. 2. The significant features, which show up in

all of the data, is the strong decrease of the A*/A ratio above 60 km and the tendency to form a minimum between 70 and 80 km. This minimum comes out much more clearly when higher heating powers are applied, as displayed by Fig. 3.

D region modification is a difficult experiment in the sense that a well developed and quiet D region is required, which is rarely found at auroral latitudes. So far we were not able to perform experiments under favourable conditions. Correspondingly, the data showed a high amount of variability, and we had to refrain from drawing conclusions concerning the ion chemistry. Our plans are to combine the partial reflection, wave interaction, and incoherent backscatter techniques to measure the T_e- and N_e-changes after heater switch-on and switch-off in order to determine chemical reaction rates in different altitude regimes as a function of T_e.

One problem concerning the partial reflection mechanism is still not conclusively solved, i.e. the contribution of electron collision frequency fluctuations to the partial reflection coefficient (e.g. JONES and KOPKA, 1978). The partial reflection coefficient can be formally expressed as the sum of two terms, one stemming from electron density fluctuations, the other from collision frequency fluctuations. Their relative importance is unknown, and an assumption is required in order to be able to interpret partial reflection measurements. Both contributions depend on T_e, but in a widely different fashion. Thus, by varying T_e by means of varying the heating power, one should be able to separate the two terms and estimate their relative weight.

3. GENERATION OF VLF AND ELF WAVES BY POLAR ELECTROJET MODULATION

Due to the short time constants for T_e-changes in the lower ionosphere, oscillations of T_e are imposed if the heating wave is amplitude modulated, provided the modulation frequency, f, does not exceed, say, 6 kHz. An oscillation of T_e with frequency f gives rise to corresponding oscillations of the elements of the conductivity tensor, since the electron collision frequency depends on T_e. Thus, if a dc current is flowing in the lower ionosphere, an ac current will be superimposed upon it, and the ionospheric portion illuminated by the heating wave will be turned into a huge antenna, radiating at the modulation frequency. We have found that this method of generating VLF and ELF waves is sensitive down to frequencies in the 100 Hz range. It is hard to imagine that such long electromagnetic waves could efficiently be generated by any other ground-based technique.

The ELF and VLF waves generated within the lower ionosphere were received with a vertical loop antenna, located at a distance of 17 km from the Heating station. The received broad band signal was transmitted, via an UHF link, to the Heating station, processed through a narrow lock-in amplifier, and both the amplitude and phase, relative to the modulation phase, were recorded. In order to obtain the amplitude, B, and the phase, Φ, as a function of f, we swept the modulation frequency from 200 Hz to 5.5 kHz within 2 minutes.

Fig. 4 shows a typical B(f) curve. The typical features are that B(f) possesses several maxima, that the highest maximum lies between 1.5 and 2.5 kHz, and that B strongly declines beyond about 5 kHz. Curves of this type prevail during most of the time. Sometimes, however, B(f) has only one or two very pronounced maxima, and it also occurs that the highest maximum lies at very low frequencies (down to 400 Hz) or at very high frequencies (up to 4 kHz). For more details, see STUBBE, KOPKA and DOWDEN (1980).

The amplitude in Fig. 4 is given in relative units, because we had not yet thoroughly calibrated our receiver. In absolute terms, B falls into the range 10^{-5} - 10^{-4} nT, depending on the strength of the polar electrojet current, for heating at 4.04 MHz, o-mode, ERP = 32 MW. This is clearly above the background noise. Much stronger effects are expected (and have been found in the meantime) for higher powers, lower HF frequencies, and x-mode heating.

From the phase as a function of frequency, $\Phi(f)$, a delay time can be defined by $\tau = d\Phi/df$ which is a measure of the apparent height of the ELF-VLF source, $h' = \tau c/2$. The delay time is the sum of the travel times of the HF wave upwards, the ELF-VLF wave downwards, and the UHF wave from the receiver to the Heating station, plus receiver delays.

A typical $\tau(f)$ curve is shown in Fig. 5. The typical features are: Strong increase of τ towards small frequencies, maxima of τ in the vicinity of maxima of B, and modest increase of τ towards high frequencies. While the increase of τ towards small frequencies can be understood in terms of the frequency dependence of the whistler mode refractive index, the increase towards high frequencies cannot readily be understood.

Our present interest is focused on the properties of the ELF-VLF generation mechanism within the lower ionosphere. Further measurements, with higher power and improved diagnostics, will be performed along this line. Eventually, however, we intend to use these ELF-VLF waves to initiate magnetospheric wave-particle interaction experiments (DOWDEN and RYCROFT, 1976). The wide accessible frequency range allows much freedom in designing such experiments.

4. GENERATION OF MICROPULSATIONS BY POLAR ELECTROJET MODULATION

If the modulation period, $T = 1/f$, becomes comparable with the characteristic time for electron density changes to occur, which is of the order 1 min in the upper D region, the electron density will be modulated, too. When this happens, the conductivity tensor will be much more affected than by a sole electron temperature modulation, and correspondingly much stronger effects are expected than in the ELF-VLF range.

Using the flux-gate magnetometer at Ramfjordmoen, operated by the University of Tromsø, we were able to detect artificially generated micropulsations with an amplitude of 10 nT at T = 10 min and 5 nT at T = 5 min (STUBBE and KOPKA, 1980). The heater operated on 4.04 MHz, x-mode, ERP = 70 MW.

Recently, using an induction magnetometer set up by the University of Göttingen at a distance of 20 km from the Heating station, we were able to record pulsations down to a period of T = 10 sec, with the heater operating on 4.04 MHz, x-mode, ERP = 115 MW. A sample is shown in Fig. 6, for T = 10, 20, 40 sec. We clearly notice the occurrence of pulsations at the wanted periods, particularly on the H-trace (i.e. the N-S component). The amplitude increases with increasing period, as theoretically expected. No pulsations are visible on the Z-trace.

The possibility of generating ULF waves with a precisely known source location, a choosable frequency, and, to some extent, a controllable amplitude gives the unique opportunity of performing active experiments in a field which so far was open to passive observations only. Certain theoretical concepts can now be directly tested and, possibly, be progessed: Guidance of Alfven waves along magnetic field lines (e.g. FEJER and LEE, 1967), magnetic field line resonances (e.g. SOUTHWOOD, 1974), ULF wave amplification and proton precipitation by wave-particle interaction (e.g. GENDRIN, 1975).

5. HF DOPPLER AND A3-TYPE F REGION ABSORPTION AND PHASE MEASUREMENTS

Recently, in a joint campaign with the University of Leicester group, we performed experiments aiming at measuring F layer deformation and anomalous absorption, caused by the heating wave. The experimental setup is sketched in Fig. 7. The HF Doppler system consists of three transmitters, at the Northern station (f = 3.498 MHz + 2.0 Hz), the Western station (f = 3.498 MHz + 5.0 Hz) and the Eastern station (f = 3.498 MHz + 8.9 Hz) and two receivers with 100 Hz bandwidth, one at Tromsø, the other at the Southern station. The measured quantity is $\Delta f = -(1/\lambda)\, dP/dt$, with λ the free space wave length and $P = \int n\, ds$ (n = refractive index). For amplitude and phase measurements, a transmitter at the Northern station (f = 3.778 MHz) and a receiver at the Southern station are used.

Fig. 8 shows a sample of the Doppler results, taken at the Southern station. A positive frequency shift is obtained during heater-on, and a negative shift during heater-off. The total frequency change within one on-off-cycle amounts to 2 - 3 Hz. Furthermore, we observe that during heater-on the traces become fainter. This absorption effect is strongest for the North-South-connexion, for which the ionospheric reflection point is almost above the Heating station.

Fig. 9 shows the corresponding phase results. The phase changes are negative during heater-on, and positive during heater-off. The Doppler and phase results can consistently be interpreted by a local enhancement of the electron density after heater switch-on, i.e. by a downward motion of the reflection point. The ionospheric conditions during the experiments were such that reflection occurred in the lower portion of the F region where NO^+ and O_2^+ are supposedly the dominating ions. Thus, our experimental findings can be understood in terms of a decrease of the recombination coefficient due to an electron temperature increase.

The absorption phenomenon which showed up in Fig. 8 is quantitatively depicted by Fig. 10. We alternately transmitted 2 min o-mode, 2 min x-mode, seperated by equally long heater-off periods. What we see from Fig. 10 is that o-mode heating leads to a very strong absorption of the diagnostic wave, of the order 15 dB, while x-mode heating virtually causes no effect. Since the experimental setup was such that the diagnostic wave could not penetrate the modified D region, there is no doubt that the observed absorption occurs in the F region. Most likely it is caused by scattering of the diagnostic wave into Langmuir waves due to the existence of short-scale field aligned irregularities (so-called striations) which are the result of a parametric instability. Theory (DAS and FEJER, 1979) predicts that this instability is excited in the vicinity of the altitude where the heating frequency equals the upper hybrid frequency. An x-mode wave cannot reach this altitude and, thus, cannot cause anomalous absorption.

ACKNOWLEDGEMENTS

The Heating project has been supported by the Deutsche Forschungsgemeinschaft, for which we express our gratitude. Thanks are due to all who contributed to the construction and operation of the Heating facility and to the performance of the experiments described above.

REFERENCES

DAS, A.C., and FEJER, J.A., 1979, "Resonance instability of small-scale field-aligned irregularities", J. Geophys. Res. 84, 6701-6704.

DOWDEN, R.L., and RYCROFT, M.J., 1976, "Wave injection experiments and wave-particle interactions", Space Res. XVI, Akademie-Verlag Berlin, 613-629.

FEJER, J.A., and LEE, K.F., 1967, "Guided propagation of Alfven waves in the magnetosphere", J. Plasma Phys. 1, 387-406.

GENDRIN, R., 1975, "Waves and wave-particle interactions in the magnetosphere: A review", Space Sci. Rev. 18, 145-200

HOLT, O., BREKKE, A. and HANSEN, T., 1980, "The PRE (Partial Refelction Experiment) facility at Ramfjordmoen, Tromsø", Proc. Vth ESA-PAC Symposium, Bournemouth (UK), April 1980, 387-392.

JONES, R.M., and KOPKA, H., 1978, "The sensitivity of D-region partial reflections to irregularity composition", J. Atmosph. Terr. Phys. 40, 723-731.

SOUTHWOOD, D.J., 1974, "Some features of field line resonances in the magnetosphere", Planet. Space Sci. 22, 483-491.

STUBBE, P., and KOPKA, H., 1979, "Ionospheric modification experiments in northern scandinavia - A description of the Heating project", Report MPAE-W-02-79-04, Max-Planck-Institut für Aeronomie, 3411 Katlenburg-Lindau 3, W. Germany.

STUBBE, P., and KOPKA, H., 1980, "Generation of Pc5 pulsations by polar electrojet modulation: First experimental evidence", accepted for publication by J. Geophys. Res.

STUBBE, P., KOPKA, H., and DOWDEN, R.L., 1980, "Generation of ELF and VLF waves by polar electrojet modulation: Experimental results", submitted to J. Geophys. Res.

TOMKO, A.A., FERRARO, A.J., LEE, H.S., and MITRA, A.P., 1980, "A theoretical model of D-region ion chemistry modifications during high power radio wave heating", J. Atmosph. Terr. Phys. 42, 275-285.

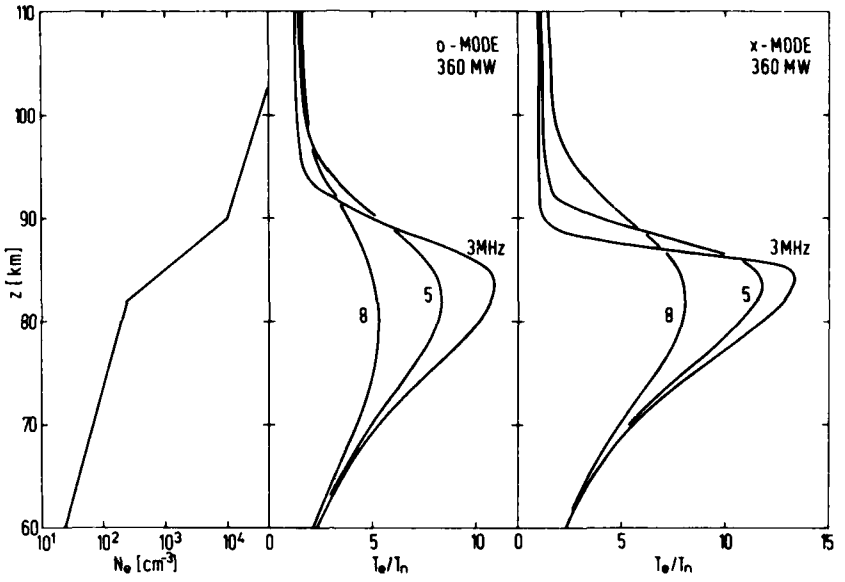


Fig. 1

Electron to neutral temperature ratio, T_e/T_n, as a function of altitude (theoretical estimate). In the unmodified ionosphere we would have $T_e/T_n = 1$. Left: Assumed electron density profile. Middle: T_e/T_n for o-mode heating with ERP = 360 MW and three heating frequencies. Right: Same for x-mode heating.

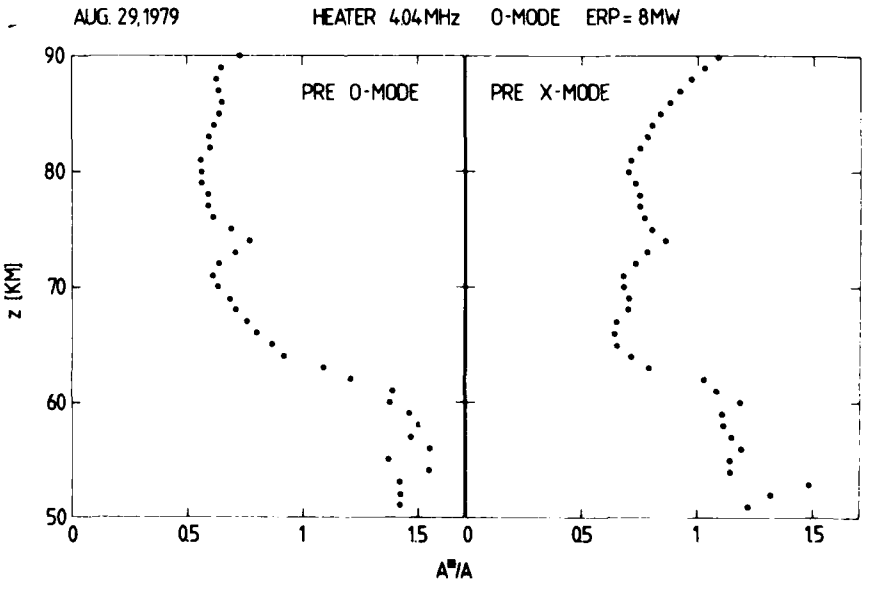


Fig. 2

Ratio of PRE amplitude with heater on, A*, to PRE amplitude with heater off, A, as a function of altitude. Both A* and A are 2 min averages. The heater operated on 4.04 MHz, o-mode, ERP = 8 MW.

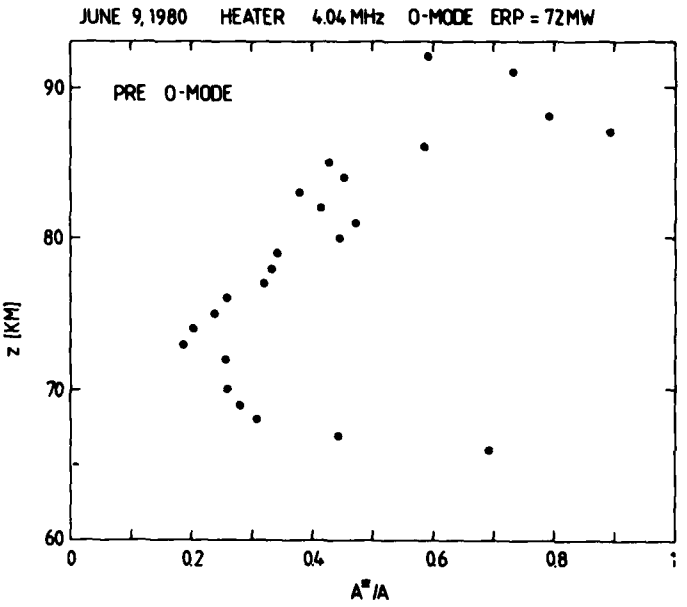


Fig. 3

A*/A vs. altitude. In this sample, A* and A are 1 min averages. The heater operated on 4.04 MHz, o-mode, ERP = 72 MW.

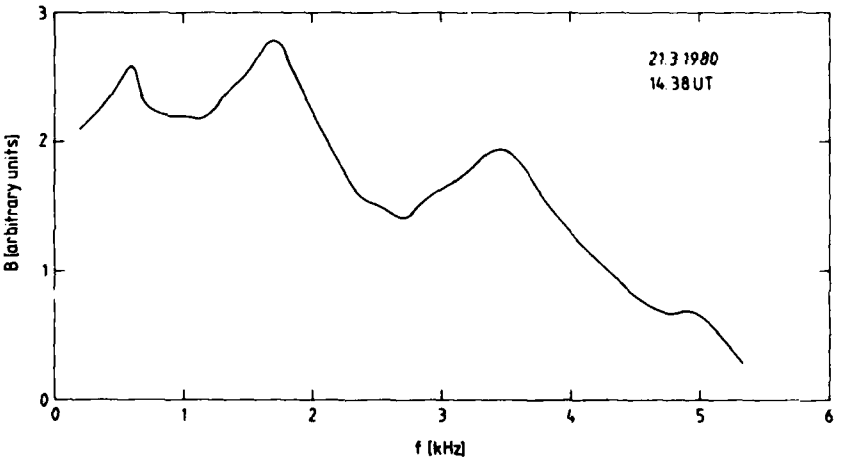


Fig. 4

ELF-VLF-amplitude, B, in arbitrary units, vs. frequency, f. The heater operated on 4.04 MHz, o-mode, ERP = 32 MW. March 21, 1980, 14.38 UT.

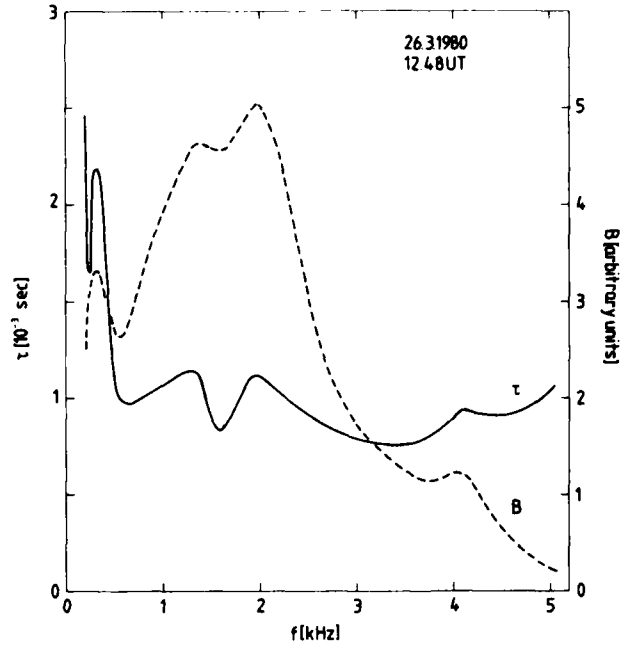


Fig. 5

Delay time, τ (solid curve), and amplitude, B (dashed curve), vs. frequency, f. The heater operated on 4.04 MHz, o-mode, ERP = 32 MW. March 26, 1980, 12.48 UT

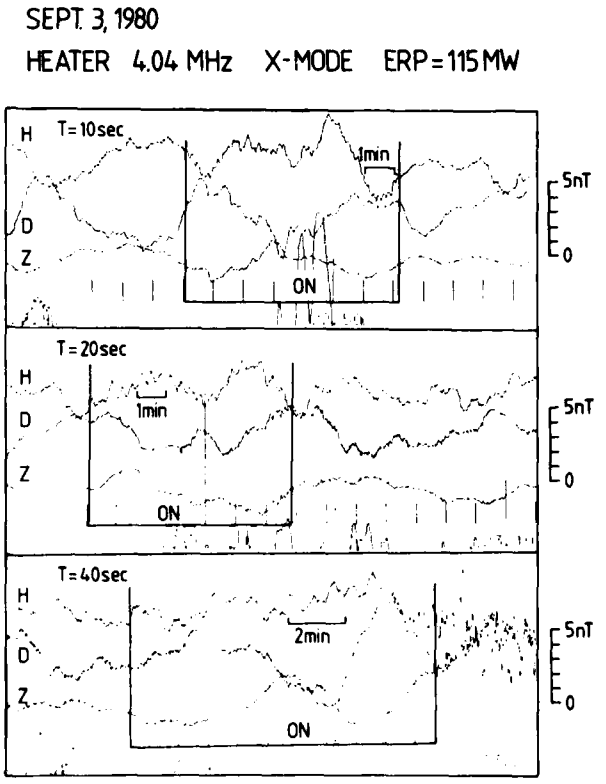


Fig. 6

Magnetometer recording of artificially generated micropulsations. The heater-on times are indicated by bars. The modulation periods are 10 sec (upper panel), 20 sec (middle panel), and 40 sec (lower panel). The heater operated on 4.04 MHz, x-mode, ERP = 115 MW. Sept. 3, 1980, 22.39 - 23.21 UT.

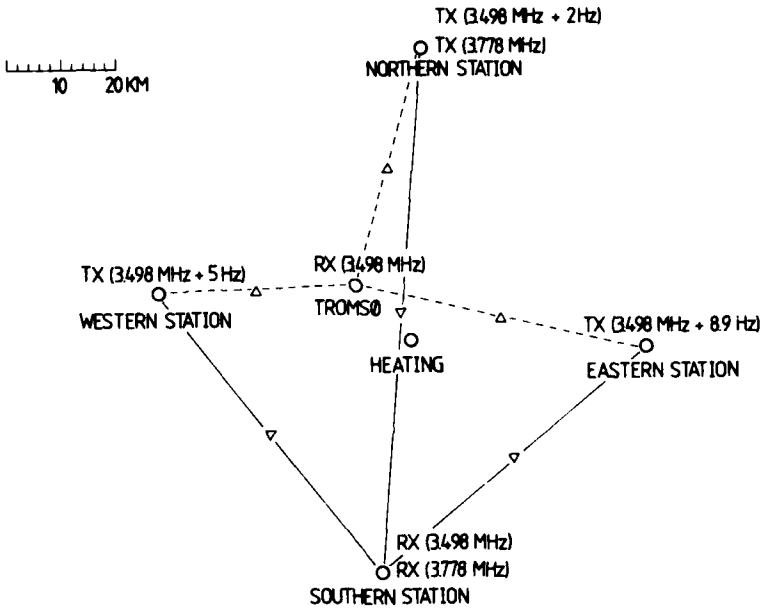


Fig. 7

Sketch illustrating the experimental setup for HF Doppler and A3-type F region absorption and phase measurements. The triangles mark the approximate locations of the ionospheric reflection points (Δ for the Tromsø receiver, ∇ for the receivers at the Southern station).

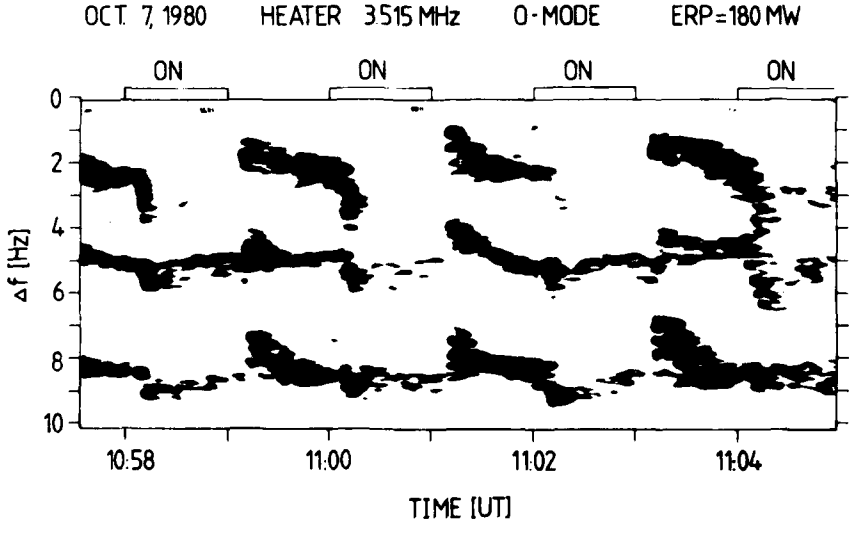


Fig. 8

Spectrograms, taken from data obtained at the Southern station. The three Doppler traces relate, from above, to the Northern, Western and Eastern transmitter, respectively (see Fig. 7). The heater-on times are indicated by bars. The heater operated on 3.515 MHz, o-mode, ERP = 180 MW. Oct. 7, 1980, 10.58 - 11.04 UT.

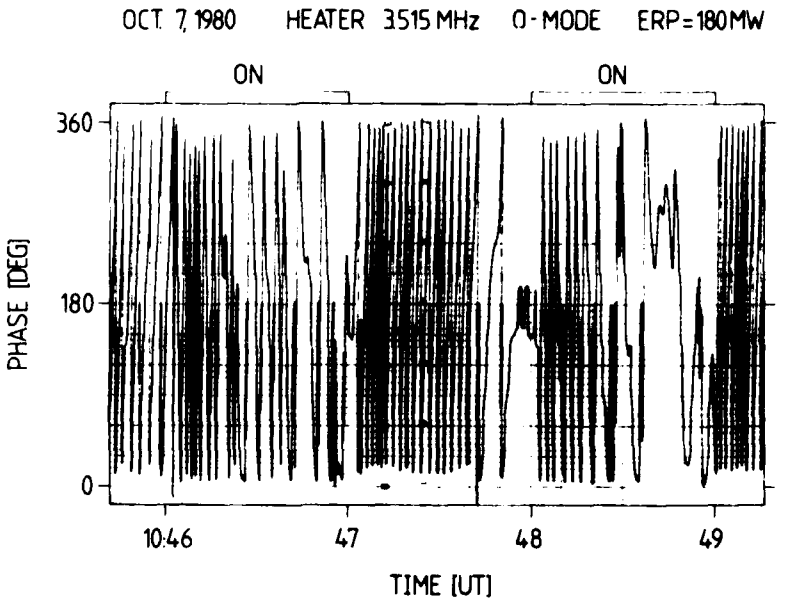


Fig. 9

Phase vs. time, measured at the Southern station with the transmitter at the Northern station (see Fig. 7). The heater-on times are indicated by bars. The heater operated on 3.515 MHz, o-mode, ERP = 180 MW. Oct. 7, 1980, 10.46 - 10.49 UT.

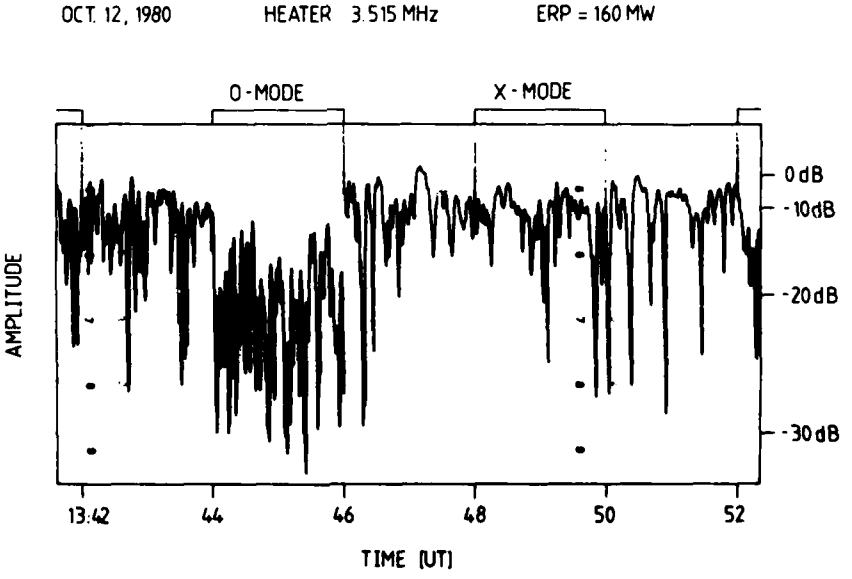


Fig. 10

Amplitude vs. time, measured at the Southern station with the transmitter at the Northern station (see Fig. 7). The heater-on times are indicated by bars. The heater operated on 3.515 MHz, alternately o- and x-mode, ERP = 160 MW. Oct. 12, 1980, 13.42 - 13.52 UT.

A SUMMARY OF ROCKET-INDUCED IONOSPHERIC PERTURBATIONS

by

Michael Mendillo
Graduate School of Arts and Scie
Boston University
Boston, Massachusetts 02215 U.

and

Jeffrey Forbes
Space Data Analysis Laboratory
Boston College
Chestnut Hill, Massachusetts 02167 U.S.A.

ABSTRACT

While rocket exhaust gases can result in perturbations of the neutral atmosphere, their major effect is realized in the upper atmosphere's tenuous ionized component. Since the earliest day of the U.S. Space Program, scattered reports of ionospheric depletions (or "holes") attempted to document the F-region effect ($h \geq 200$ km) while plasma enhancements were occasionally offered as the E-region effect ($h \simeq 150$ km). A major re-appraisal of rocket exhaust/ionospheric coupling occurred following observations of the large-scale F-region hole caused by the Saturn V launch of SKYLAB in 1973. The new theoretical basis for the effect (i.e., rapid diffusion of a rocket's highly reactive molecular exhaust cloud) was verified during the recent launch of NASA's HEAO-C satellite by an Atlas/Centaur rocket on 20 September, 1979. A new generation of active-plasma experiments will be carried out when the Space Shuttle executes a series of "dedicated engine burns" for a program of plasma depletion experiments for environmental and radioastronomical studies in the early 1980's.

I. INTRODUCTION

Since the earliest days of space exploration, concerns have been raised regarding the atmospheric perturbations that might be caused by the exhaust products of large rocket engines. In 1964, a comprehensive review of the subject was published by W.W. Kellogg in which the space science community was alerted to the many possible ways that rocket "pollutants" could have environmental impacts. The final impression left by Kellogg's paper was that the terrestrial atmosphere was sufficiently dense to absorb any conceivable shock that an aerospace technology might reasonably be expected to build.

In the 16 years that have passed since Kellogg's assessment, very little evidence was found to suggest that rocket effluents were in fact perilous to our environment. With the coming of the Space Shuttle as the U.S. Space Transportation System for the 1980's, and with increased discussions of various "large space structures" for the 1990's, the question of environmental assessments for new aerospace systems is once again an area of public concern. The Space Shuttle era also initiates a period of "active experiments" in the near-Earth environment. The many in-situ, laboratory-in-space, plasma research experiments proposed for the 1980's amount to controlled atmospheric modification experiments that will be capable of addressing a wide range of physical and chemical processes of aeronomic interest. The purpose of this review is to summarize the known theoretical and observational aspects of "inadvertant" rocket-exhaust effects upon the ionosphere. An understanding of these effects, their causes and their consequences, will be a valuable data-base from which second-generation "chemical release experiments" may be developed; in addition, these "experiments of opportunity" are very useful case studies for atmospheric assessment concerns associated with large space structures.

II. ATMOSPHERIC REGIONS SUSCEPTIBLE TO ROCKET EXHAUST EFFECTS.

Figure 1 presents a summary of relevant terminology in which the Earth's atmosphere is divided into regions based upon properties of the neutral atmosphere as well as its ionized component (the ionosphere). While a rocket launch can obviously affect any atmospheric region traversed by a vehicle, the major perturbations occur in those regions where the sun's radiation causes partial ionization to occur - that is, in the so-called D, E, F1 and F2 regions of the ionosphere. These regions contain a small fraction of the total mass of the atmosphere, and yet they are important because their electrically charged components govern many aspects of radio communications. The positive ions found in the lower three regions are molecular ions (e.g., NO^+, O_2^+), while in the F2 region the dominant ion is atomic (O^+). Overall charge neutrality implies that in any given volume there are equal numbers of ions and electrons (e^-). As will be shown in later sections, the dominant cause of atmospheric perturbations due to rocket exhaust rests in the variety of chemical reactions that can occur between the exhaust material (usually molecular species, e.g., H_2O, H_2, CO_2) and the neutral and ionized components of the atmosphere.

Figure 2 offers a schematic summary of the various regions where typical rocket launch activity might cause atmospheric effects. While the spatial extent of possible perturbations spans tropospheric to

magnetospheric domains, the scope of this review will be limited to ionospheric heights.

In the so-called F2-region of the ionosphere (h > 200 km), the dominant neutral species is atomic oxygen (O) and the major plasma components are O^+ and e^-. In this tenuous region the neutral concentrations are $\sim 10^9$ cm^{-3} or less and the ionized concentrations $\sim 10^6$ cm^{-3} or less. The introduction of large quantities of rocket exhaust molecules can thus represent a significant departure from ambient conditions and the resultant atomic - molecular chemistry can lead to dramatic effects. Figure 3 offers a summary of the types of physical processes (and possible consequences) initiated by a large engine burn in the F-region. The bulk of this report will deal with observations and theory related to the scenario depicted in Figure 3.

Below 200 km, the neutral atmosphere becomes increasingly more dense and it is composed almost entirely of molecular species. The ionospheric densities in the so-called D and E regions are orders of magnitude smaller than those found at F-region heights, and the ions are predominantly molecular. In these lower portions of the ionosphere, molecular chemistry is already the dominant ionospheric process, and thus truly enormous amounts of rocket exhaust are required to over-power ambient processes --- even for small spatial regions. As a result, very little data exists showing how artificially induced perturbations can be created in these regions, if at all. The little evidence that does exist, as well as a new theoretical treatment of perturbations in these regions, have recently been reviewed by Forbes (1980). The spectrum of possible effects to be considered from "low-altitude" engine burns is given in Figure 4.

III. OBSERVATIONAL EVIDENCE FOR ARTIFICIALLY-INDUCED IONOSPHERIC MODIFICATIONS.

The literature on artificially-induced ionospheric modifications (excluding RF heating effects) falls into two distinct categories: (1) effects due to missle transits close to an ionospheric observing station (the experiment may or may not have been planned beforehand) and (2) effects due to the deliberate release of a known quantity of highly reactive molecules into the ionosphere. The latter topic was reviewed recently by Pongratz (1980).

While the discussion of rocket exhaust effects upon the Earth's upper atmosphere is not a new topic, past documentations of plume-associated disturbances have not been very great in number. Table 1 presents a chronological summary of published reports that suggest a causitive link between observed ionospheric variations and rocket launches. It is surprising, at first, to consider that so few accounts should exist given the many hundreds of rocket launches that have occurred since 1957. (On occasion of the 20th Anniversary of the U.S. Space Program, NASA announced that 402 rockets were launched from its facilities during the period 1958-1977). Of the reports on rocket-induced perturbations that do exist, most of the effects described in the early literature were termed "localized" (Booker, 1961, Jackson et al. 1962, Stone et al., 1964). In retrospect, the reasons for this are clear: (1) The vast majority of rockets launched in the last two decades were relatively small ones and thus the exhaust emissions amounted to no more than very minor additions of essentially trace species. (2) Of the large rockets (Saturns, Atlas/Centaurs, SL-4/SOYUZ), the overwhelming majority of launches carried payloads into low-earth-orbit (h < 200 km) where the typical exhaust products (H_2O, H_2, CO_2, N_2, O_2) were again relatively inconspicuous additions to ambient conditions. (3) Large rocket launches made by U.S. agencies generally occur at the Kennedy Space Flight Center (KSFC), thereby insuring that rocket ascent trajectories occur over water where atmospheric monitoring capabilities are extremely limited.

The main conclusion to be drawn from these points is that virtually all past rocket launches offered little reason, whether scientific or technological, to search systematically for the atmospheric perturbations caused by rocket effluents. Morevover, the theoretical understanding of the processes involved was limited to discussions of rocket exhaust "snow-plow effects" upon the atmosphere, that is, to a physical displacement of the ionosphere by the plume --- effects that by their very nature would be limited to the regions immediately surrounding a moving rocket (Booker, 1961, Barnes, 1961).

A new class of ionospheric disturbance --- the large-scale F-region hole - became apparent during the launch of Skylab in 1973 (Mendillo et al., 1975a, b). The Saturn V rocket that launched NASA's Skylab Workshop was the last Saturn V used in the U.S. Space Program and the only one ever to have its main engines burning above 200 km. The resultant deposition of approximately 1000 kg/sec of H_2 and H_2O exhaust molecules into the 200-440 km altitude region initiated a rapid and large-scale depetion of the ionosphere to an extent never seen before. The artificially-created "ionospheric hole" amounted to nearly a 50% decrease in the total electron content (TEC) of the ionosphere over an area of approximately a million square kilometers. Mendillo et al. attributed the effect to the rapid expansion of an exhaust cloud of highly reactive molecules that initiate a rapid recombination of the ionospheric plasma. The details of this mechanism were first presented to the AGARD community by Mendillo and Forbes (1976) and Bernhardt (1976). Subsequent theoretical model studies by Mendillo and Forbes (1978), Anderson and Bernhardt (1978) and Zinn and Sutherland (1980) have described various aspects of the overall F-region modification scenario given in Figure 3.

IV. SCIENTIFIC AND TECHNOLOGICAL INTEREST IN ARTIFICIALLY-CREATED IONOSPHERIC HOLES

The Skylab event of 13 May 1973 rekindled an interest in molecular release experiments that had languished since the mid-1960's. The effects caused by highly reactive molecules introduced into the F-region offer the opportunity to study a variety of atmospheric processes (e.g., molecular diffusion, neutral/plasma chemistry, airglow emissions, thermal energy balance, plasma transport, and electron density irregularity generation). In addition, the large-scale disruption of the ionosphere's vertical density structure, $N_e(h)$, has implications for radio propagation. The discussion of these issues led to the very successful set of Lagopedo modification experiments of September 1977 in which rocket-borne payloads of H_2O and CO_2 were released into the F-region over Hawaii (Pongratz and Smith, 1978). A more extensive series of F-region modification experiments are to be performed via "dedicated burns" of the Space Shuttle engines as the orbiter passes near five ionospheric and radio astronomical observatories. The experiments are part of the Spacelab-2 mission currently scheduled for November, 1983 (Mendillo et al. (1978), Rosendhal (1978), Bernhardt et al. (1978)).

Quite separate from the scientific objectives of the Lagopedo and Spacelab-2 plasma depletion experiments is a concern that the new generation of space transportation systems being developed and/or considered for the future require routine engine firings in the upper ionosphere and therefore the routine creation of large-scale ionospheric holes. The soon to be operational Space Shuttle program will include orbital configurations usually in the 250-450 km altitude range --- at precisely the heights where the ionospheric plasma densities (O^+ and e^-) reach their maximum values. The Space Shuttle's engine exhaust rates are considerably smaller than for a Saturn V rocket and thus the very large spatial extent associated with the SKYLAB effect will not be found for the Space Shuttle induced holes.

The heavy-lift launch vehicles (HLLV's) and personnel orbital transfer vehicles (POTV's) required for the proposed Solar Power Satellite (SPS) system represent a very substantial increase over "conventional" Space Shuttle cargo and support-staff transportation modes. The regular transfer of material and personnel to low-earth-orbit (LEO) and from LEO to geostationary-earth-orbit (GEO) suggests that a routine modification of the ionosphere will be a consequence of the fully implemented SPS concept (Rote, 1979). The ultimate spatial and temporal extent of the ionospheric holes associated with any large space structure will depend on the specifics of the launch vehicles to be designed and the orbital flight plans that evolve. Ample opportunity exists to influence these decisions by model/simulation studies, as suggested by some of the preliminary SPS environmental assessment results presented by Zinn and Sutherland (1980), and Mendillo et al. (1980a), or by actual observations of perturbations associated with "experiments of opportunity." The best documented case of a large-scale ionospheric hole occurred during the "opportunity" created by NASA's launch of the satellite HEAO-C by an Atlas/Centaur rocket in September, 1979 (Mendillo et al. 1979, Bernhardt et al., 1979). The results of that study, as described during an environmental assessment Workshop/Symposium related to SPS concerns (Mendillo, 1980), may be summarized as follows (Mendillo et al., 1980b):

(1) The "HEAO-Hole" was monitored by a network of 12 satellite radio beacon observatories, airborne and ground-based optical observatories, 2 incoherent scatter radars, and over 150 radio propagation monitoring sites.

(2) The ionospheric depletion caused by the rocket exhaust cloud spanned a region of approximately 2 million square kilometers. The total electron content depletions near the rocket trajectory showed more than an 80% reduction from ambient conditions.

(3) The dominant airglow emissions associated with the hole were 6300Å observations of 8300 Rayleighs (viewed vertically), or approximately 20,000 Rayleighs measured along a slanted path from the launch site.

(4) HF radio propagation experiments showed that no severe long-term effects were associated with the F-region hole. Some propagation paths showed short-term fades or increased fading rates in the 7-21 MHZ range that occurred at times consistent with a rocket-associated effect. However, similar disturbances of natural origin were noted before and after the launch.

In summary, the series of experiments that provided diagnostics for the HEAO-Hole effects represent the best available description of a rocket-induced atmospheric perturbation. The data collected during that event will provide the space science community with vital information needed for (1) on-going plans for future, active experiments in space, (2) test cases for computer simulation codes used for ionospheric perturbation studies, and (3) timely imput to environmental assessment programs concerned with the construction scenarios for large structures in space.

Acknowledgement. This work was supported in part by NASA contract #NASC-8-32844 to Boston University.

REFERENCES

1. Anderson, D.A. and P.A. Bernhardt, Modeling the Effects of an H_2 Gas Release on the Equatorial Ionosphere, J. Geophys. Res., 83, 1978.

2. Arendt, P.R., Ionospheric Undulations Following Apollo 14 Launching, Nature, 231, 438, 1971.

3. Arendt, P.R., Ionospheric Shock Front from Apollo 15 Launching, Nature, 236, 8, 1972.

4. Bakai, A.S., Kuperov, L.P. and G.K. Solodovnikov, Excitation and Propagation of Solitary Drift Waves in the Ionosphere, Sov.J. Plasma Phys., 3 (3), 323, 1977, Amer. Inst. Phys., 1978.

5. Barnes, Cecil, Jr., Comment on Paper by Henry G. Booker, "A Local Reduction of F-region Ionization Due to Missile Transit", J. Geophys. Res., 66, 2589, 1961.

6. Bernhardt, P.A., da Rosa, A.V. and M. Mendillo, High Altitude Vapor Releases from the Space Shuttle, Trans. Amer. Geophys. Union, 59, 1162, 1978.

7. Bernhardt, P.A., da Rosa, A.V. and C.G. Park, Chemical Depletion of the Ionosphere, NATO/AGARD specilists' Meeting on "Artificial Modification of Propagation Media," 26-29 April, 1976, Brussels, Belgium, AGARD-CP-192 (H.J. Albrecht, ed.), 1977.

8. Bernhardt, P.A., Klobuchar, J.A., Villard, O.G., Simpson, R., Troster, J.G., Mendillo, M. and J.H. Reisert, The Great Ionospheric Hole Experiment, QST, LXIII, 22, 1979.

9. Booker, H.G., A Local Reduction of F-region Ionization Due to Missile Transit, J. Geophys. Res., 66, 1073, 1961.

10. Felker, J.K. and W.T. Roberts, Ionospheric Rarefaction Following Rocket Transit, J. Geophys. Res., 71, 4692, 1966.

11. Forbes, J.M., Upper Atmosphere Modifications due to Chronic Discharges of Vapor from Space Launch Vehicle Exhausts, in Space Systems and Their Interactions with the Earth's Space Environment (H.B. Garrett and C.P. Pile, editors), AIAA, New York (in press, 1980).

12. Jackson, J.E., Whale, H.A. and S.J. Bauer, Local Ionospheric Distrubance Created by a Burning Rocket, J. Geophys. Res., 67, 2059, 1962.

13. Kellogg, W.W., Pollution of the Upper Atmosphere by Rockets, Space Sci, Rev., 3, 275, 1964.

14. Mendillo, M., (ed.) Proc. Workshop/Symp. on Preliminary Evaluation of the Ionospheric Disturbances Associated with the HEAO-C Launch, with Applications to the SPS Environmental Assessment, Boston Univ, Nov. 1979 (ANL/DOE Tech. Rept. in press, 1980).

15. Mendillo, M. and J.M. Forbes, Spatial-Temporal Development of Molecular Releases Capable of Creating Large-Scale F-Region Holes, NATO/AGARD Specialists' Meeting on "Artificial Modification of Propagation Media," 26-29 April, 1976, Brussels, Belgium, AGARD-CP-192 (H.J. Albrecht, ed.), 1977.

16. Mendillo, M., Baumgardner, J. and J.A. Klobuchar, Opportunity to Observe a Large-Scale Hole in the Ionosphere, EOS, Trans. Amer. Geophys. Union, 60, 513, 1979.

17. Mendillo, M., da Rosa, A.V. and P.A. Bernhardt, Spacelab-2 Plasma Depletions Experiments, Trans. Amer. Geophys. Union, 59, 334, 1978.

18. Mendillo, M., G.S. Hawkins and J.A. Klobuchar, A Large-Scale Hole in the Ionosphere Caused oy the Launch of Skylab., Science, 187, 343, 1975a.

19. Mendillo, M. and J.M. Forbes, Artificially Created Holes in the Ionosphere, J. Geophys. Res., 83, 151, 1978.

20. Mendillo, M., G.S. Hawkins and J.A. Klobuchar, A Sudden Vanishing of the Ionospheric F-region Due to the Launch of Skylab., J. Geophys. Res., 80, 2217, 1975b.

21. Pongratz, M.B. and G.M. Smith, The Lagopedo Experiments - an Overview, Trans. Amer. Geophys. Union, 59, 334, 1978.

22. Pongratz, M.B., Large Scientific Releases, Proceedings of COSPAR XXIII Plenary Meeting, Budapest, 2-14 June, 1980, Space Research (in press, 1980).

23. Reirisch, B.W., Burnt-out Rocket Punches Hole into Ionosphere. Space Research XIII. 503, 1973.

24. Rosendhal, J.D., The Spacelab-2 Mission, Sky Telesc., 55, 1978.

25. Stone, M.L., Bird, L.E. and M. Balser, A Faraday Rotation Measurement on the Ionospheric Perturbation Produced by a Burning Rocket, J. Geophys. Res., 69, 971, 1964.

26. Zinn, J. and C.D. Sutherland, Effects of Rocket Exhaust Products in the Thermosphere and Ionosphere, Space Solar Power Review, 1, 109, 1980.

TABLE 1: OBSERVATIONS OF IONOSPHERIC DISTURBANCES CAUSED BY ROCKET LAUNCHINGS

Rocket	Date	Altitude Engine Shut-off (km)	Effect	Observation Technique	Observer
Vanguard 2*	1959	F-region	F-region depletion	Vertical soundings	Booker (1961)
ScoutΔ	1961	300	F-region depletion; E-region enhancement	Vertical soundings Faraday rotation from rocket beacon	Jackson et al. (1962)
Atlas (5 cases)*	1961 1962	350	F-region depletion; E-region enhancement	Faraday rotation from rocket beacon	Stone et al. (1964)
Saturn SA-9/ Pegasus A*	1965	500	F-region depletion; E-region enhancement	Vertical soundings	Felker & Roberts (1966)
Saturn 5 (Apollo 14)*	1971	$\approx$190	F-region travelling ionospheric disturbances	Vertical soundings	Arendt (1971)
(Apollo 15)*	1971	$\approx$190		Vertical soundings	Arendt (1972)
Black Brant+	1971	35	F-region depletion	Vertical soundings	Reinisch (1973)
Saturn 5*	1973	442	Large-scale F-region depletion ('iono-spheric hole')	Total electron measurements using geosta-tionary satellite beacons	Mendillo et al. (1975a,b)
Saturn 1B (Apollo)$^{\circ}$ (Soyuz-19)$^{\circ}$	1975	200	F-region travelling ionospheric disturbance E-region enhancement	Vertical soundings	Bakai et al. (1978)

*Kennedy Space Flight Center ΔWallops Island +Eglin, Fla. $^{\circ}$Baykonur Cosmodrone

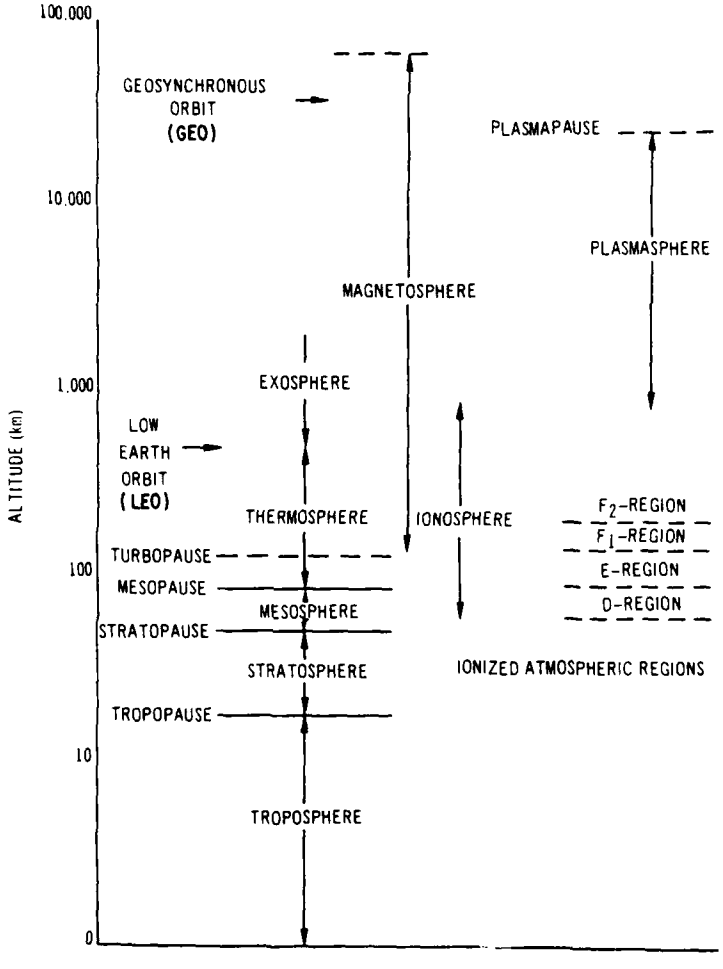


Figure 1. Regions of the Earth's atmosphere (after Rote (1979)).

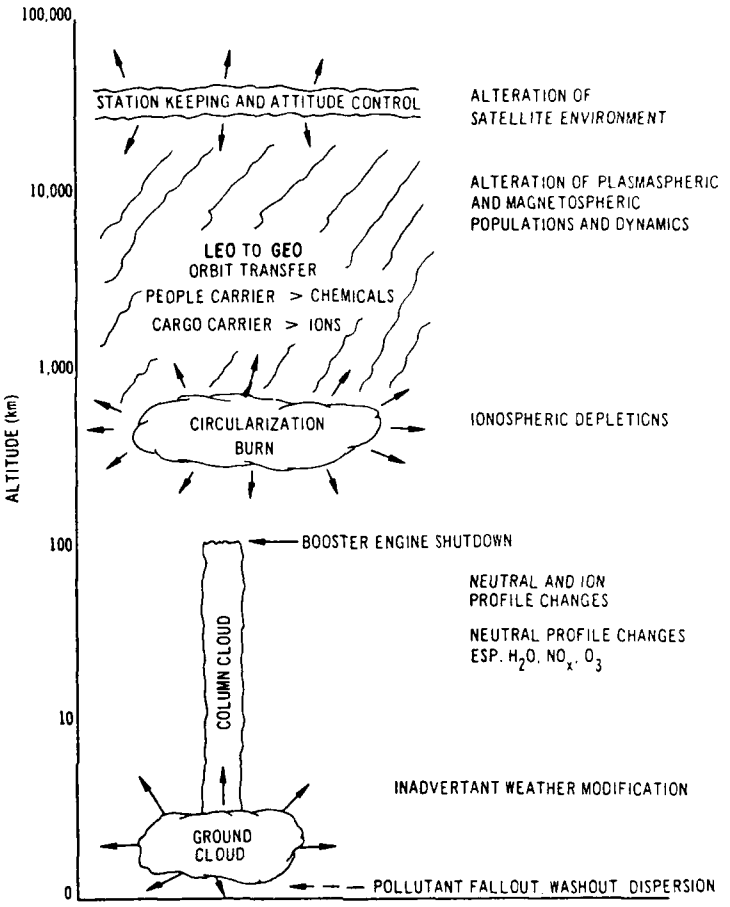


Figure 2. Profile of possible rocket-induced perturbations in the atmosphere (after Rote (1979)).

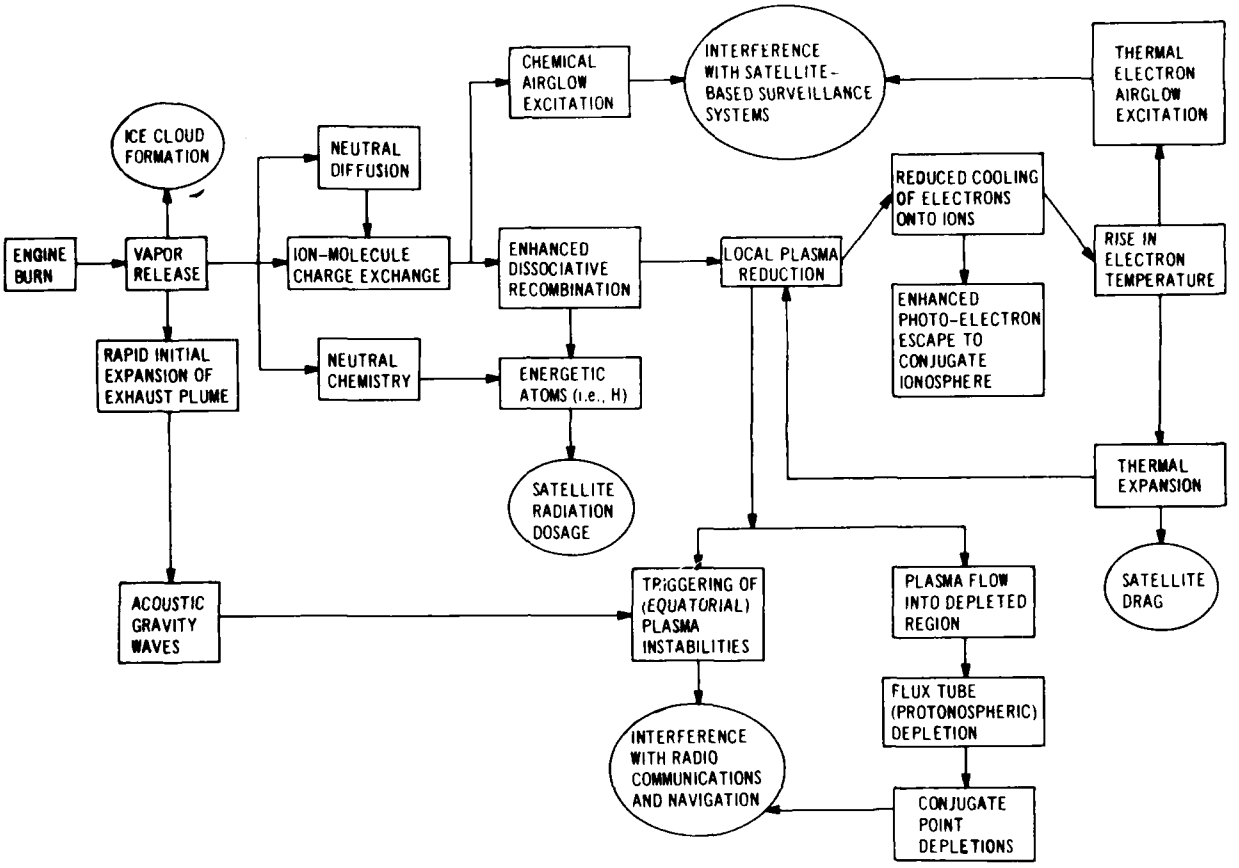


Figure 3. Summary of rocket effluent effects in the F-region of the ionosphere ($h > 200$ km) (after Rote (1979)).

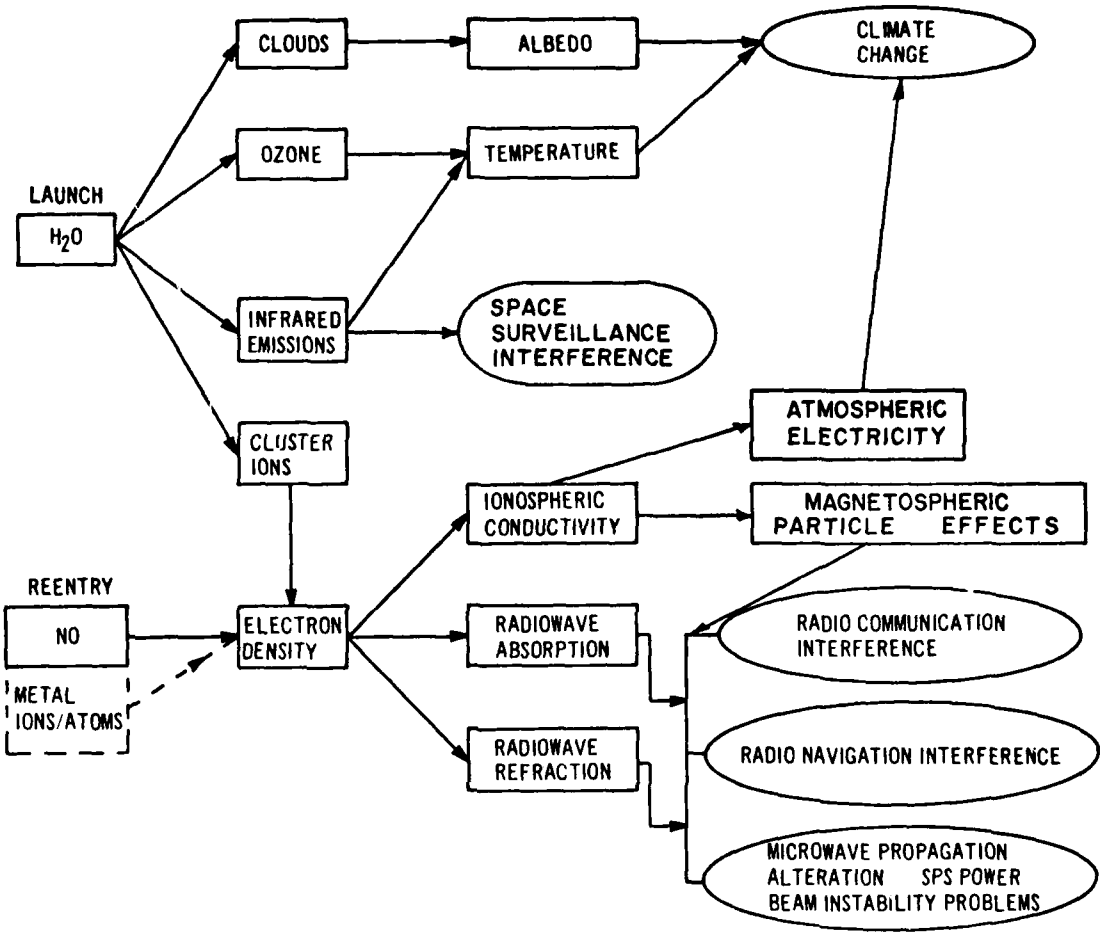


Figure 4. Summary of rocket effluent effects in the D and E-regions of the ionosphere (h ≃ 70-150 km) (after Rote (1979)).

SESSION DISCUSSION

SESSION III DYNAMICAL PHENOMENA

Chairman and Editor Prof. J.S.Nisbet
Code 960
Goddard Space Flight Center
Greenbelt, MD 20771, USA

ELECTRODYNAMICS OF THE LOWER IONOSPHERE
by H.Volland

J.S.Nisbet, US

Would you care to comment on how the mesospheric electric fields discussed by Dr Goldberg yesterday affect your electric circuit model.

Author's Reply

If the large electric fields measured at upper stratospheric heights should turn out to be real, they must be maintained by an additional source, e.g. a space charge layer at these heights. Such a source would then have to be added in Figure 2.

S.H.Gross, US

Your equivalent circuit arrangement appears to be balanced except for the dynamo generator. Why do you use this balanced arrangement and why is the dynamo generator unbalanced?

Author's Reply

The symmetric arrangement of the circuit elements in Systems I and III of my Figure 2 has been made in order to allow the current from System II to flow within regions I and III outside the source ranges.

H.Poeverlein, Ge

Which is the energy source in the case of currents originating from phenomena at the magnetopause? It may be the solar wind or the rotary motion of earth and atmosphere.

Author's Reply

The magnetospheric voltage source in my Figure 2 illustrates electric fields related to magnetospheric convection. Its origin may be reconnection processes and/or viscoid-like interactions between solar wind and magnetopause.

NEUTRAL WINDS IN THE POLAR IONOSPHERE
by A.Brekke

K.Rawer, Ge

Stolarsky and Roble also used energy deposition at high latitude in their (more empirically based) model computations. It might be interesting to compare your results with theirs.

Author's Reply

As a matter of fact, Roble has partly used our estimates of the relationship between magnetic bay intensities and Joule heating rate to estimate the Joule heat on a global scale in the auroral zone.

M.Blanc, Fr

You made an interesting point about the way to relate a local magnetometer H trace to a local estimate of auroral zone Joule heating. But I have two further questions: 1. Is it possible using the available network of observations, to extend this local estimate to an estimate of the latitude distribution of the heat deposition? 2. How well (or poorly) do we presently know the amount of heat deposited on the dayside (where heating takes place at very high latitudes)?

Author's Reply

One possible use of extending this local estimate to an estimate of the latitude distribution would be to use the existing network of ionosondes and magnetometers, together with models of the global electric field. The ionosondes would give a global view of the electron densities which again could be used to derive the height integrated conductivities. The amount of Joule heat deposited in the dayside and high latitudes is even less well known than at low latitudes, due to the lack of a well-developed network of observatories in these areas. From the Chatanika radar and from the EISCAT radar it will be possible to do some relevant experiments in this regard.

H.Kohl, Ge

Why did you relate the Joule heat to the variation of the horizontal component of the magnetic field and not to the total radiation?

Author's Reply

By the use of the incoherent scatter radar at Chatanika, we can only estimate the horizontal ionospheric current. In this comparison it is inherent that the horizontal current is assumed to be a sheet current of finite length. In this sense there should be no vertical component of the magnetic field unless the current sheet was at a distance from the observing site. Since we have only scaled maximum bays we try to avoid this complexity of the geometry of the currents.

G.Prolss, Ge

You indicated in your talk that the AU index corresponding to positive bay activity is associated with larger energy deposition than the AL index corresponding to negative bay activity. For the atmosphere response to this energy injection it is very important to know at what height the energy is dissipated in both cases. Can you comment on this aspect?

Author's Reply

It seems likely that the Joule heat is on the average deposited at about five kilometer higher altitudes during positive magnetic bays than during negative negative bays. (See Brekke, A. & R. L. Rino, "High-Resolution Altitude Profiles of the Auroral Zone Energy Deposition Due to Atmospheric Currents", Jnl Geophysics Res, 83pp 2517-2524, 1978.

J.Aarons, US

What do you predict will take place at F-Layer heights particularly in the polar regions vis-a-vis electron densities and F-Layer irregularities?

Author's Reply

It is difficult for me to predict much about the F-Layer since my basic interest has been in the E.-Layer. The difference in height of the ionospheric currents and Joule heat deposition during positive and negative bay, indicates that there probably is a tendency for auroral zone gravity waves to be more easily heated in a positive bay.

A SUMMARY OF ROCKET-INDUCED IONOSPHERIC PERTURBATIONS

by M.Mendillo and J.Forbes

D.G.Torr, US

You mentioned that 8 kR of red line emission were observed. What is the mechanism which produces this emission? Secondly, what intensity was observed for the visible (3090Å?) OH emissions that was also monitored. Presumably the expected source is recombination of H_2O^+.

Author's Reply

For a release consisting of H_2O and H_2 the relevant reactions producing excited oxygen (O^*) and Hydroxyl (OH^*) are:

$$H_2O + O^+ \rightarrow H_2O^+ + O$$

$$H_2 + O^+ \rightarrow OH^+ + H$$

$$H_2O^+ + e^- \rightarrow OH^* + H$$

$$H_2O^+ + e^- \rightarrow H_2 + O^*$$

$$OH^+ + e^- \rightarrow O^* + H$$

The photometrics/AFGL group (Kofsky, Villanucci and Huppi) monitored visible emissions from OH 6700Å, but no effect was observed. However, they did detect a 30% increase in the 2.9μm emission from OH following rocket passage.

S.H.Gross, US

Was there an effort to determine whether waves spread from the depleted region? Were any of the wave characteristics measured?

Author's Reply

Yes, Behnke and Duncan at Arecibo reported electron density fluctuations with a period of 20 30 minutes, which appeared at a time after rocket passage consistent with a horizontal phase speed of 200m sec^{-1}. Of course, from these single station data one cannot exclude the possibility that the density fluctuations are of local origin.

K.Rawer, Ge

(1) Concerning the ion chemistry, has the possibility of attachment and negative ion formation been considered?

(2) Has the existence of H_2O^+ ions as the major ionic constituent in the depleted region been checked in situ?

Author's Reply

(1) No, electron attachment has not been considered. However, our current theoretical simulations model the data quite well without consideration of this effect.

(2) Yes, in the Lagopedo rocket experiment where a cannister of H_2O was ruptured in the F-region, rocket-borne ion mass spectrometers identified H_2O^+ as the major molecular ion produced.

IONOSPHERIC COMPOSITION: THE SEASONAL ANOMALY EXPLAINED

D. G. Torr*, P. G. Richards, M. R. Torr*
Center for Atmospheric and Space Sciences
Utah State University
Logan, Utah 84322

* Also affiliated with the
National Institute for Telecommunications Research
of the C.S.I.R., Johannesburg, South Africa.

ABSTRACT

Despite years of study, the seasonal behavior of the F_2 layer of the ionosphere has never been satisfactorily explained. We demonstrate that the main reason for this failure can be attributed to serious errors in our basic understanding of the photochemistry of the ionosphere. In this paper we reanalyze the main chemical processes in the light of recent laboratory measurements of rate coefficients, using the Atmosphere Explorer data. Major changes to the chemistry include the transfer of nearly all metastable O^+ ions to N_2^+ via charge exchange with N_2. The N_2^+ ions become vibrationally excited by resonant fluorescence of solar near UV and near infrared radiation, leading to a return transfer of N_2^+ ionization to O^+ by charge exchange of vibrationally excited N_2 with atomic oxygen. With this chemistry the seasonal variations in the peak electron densities are then shown to be caused primarily by anomalous seasonal variations in neutral composition. The required neutral composition variations are empirically reproduced by the MSIS model atmosphere. The circulation derived from recent 3D models of the global thermosphere qualitatively accounts for the seasonal variations in neutral composition predicted by the MSIS model. In addition to the composition effect, vibrationally excited N_2 is found to contribute a $\sim$ 20% effect to the anomalous seasonal behavior at solar maximum.

1. Introduction

Despite the great wealth of data that has been acquired on the ionosphere and thermosphere over the last two decades, several first order phenomena have not been satisfactorily explained. Amongst these is the phenomenon known as the F-region winter anomaly. Its existence was identified at an early stage in the history of ionospheric physics. Ground based ionosonde measurements of the critical frequency of the F_2 layer, f_oF_2, revealed an anomalous seasonal variation in this parameter. Early theories established that the ionosphere at middle and low latitudes is produced mainly by photoionization of the neutral atmosphere. These theories predicted that production of ionization would maximize in summer yielding larger peak electron densities in that season. This kind of behavior was observed in the E and F_1 layers of the ionosphere, but not in the case of the F_2 layer which exhibits larger peak densities in winter.

Although the F_2 layer anomaly historically has also been referred to as the winter anomaly, this is really a misnomer, because the anomalous seasonal behavior becomes most pronounced at the equinoxes at most locations. It also varies strongly with the epoch of the solar cycle and geographic location.

2. Summary of the Morphology of the Phenomenon

Numerous workers have studied the F region anomaly, but the most comprehensive morphological analysis to date was carried out by M. Torr and Torr (1973) who reviewed the earlier work in detail. Figures 1a and 1b from M. Torr and Torr (1973) show the long term behavior of the f_oF_2 at four stations, one in the northern and three in the southern hemisphere. The curves represent plots of the median noon and midnight values of f_oF_2 for each month of the year, and serve to illustrate the anamolous seasonal behavior referred to above.

The data illustrate that the seasonal variations are quite different on the continents of Europe, Africa and Australia. At the South African stations, Johannesburg and Cape Town, the noon values of f_oF_2 exhibit a clear semi-annual variation superimposed on a solar cyclic variation with seasonal peaks at the equinoxes. The summer values of f_oF_2 are higher than the winter values except in the rising phase of the solar cycle when the rapid change in f_oF_2 with solar activity obscures the summer-winter difference. The amplitude of the seasonal variation does not change significantly with the epoch of the solar cycle.

The midnight values of f_oF_2 at Johannesburg and Cape Town exhibit different behavior from that at noon. The seasonal variation is predominantly annual peaking in summer. An odd phenomenon is the fact that the winter midnight values show no solar cycle variation.

The behavior at Slough in the northern hemisphere is different again. The seasonal variation is predominantly annual peaking in winter, with a pronounced summer minimum. It is data such as these which lead to the term "winter anomaly". In this case the amplitude of the seasonal variation increases strongly with solar activity. There is some suggestion that the winter peak is actually a composite of two separate overlapping peaks. The bifurcation of these peaks is clearly evident at solar maximum in the late 1960's, where the phase shifts from December toward the fall equinox. The midnight values do not show any anomalous behavior, i.e. f_oF_2 peaks in summer. This seasonal variation is superimposed on a normal solar cycle variation.

The behavior of f_oF_2 at Christchurch in the Australian sector of the globe exhibits a mixture of the seasonal behavior observed at both the previously discussed locations. At noon there is evidence of a semi-annual variation with peaks at the equinoxes with the vernal equinox exceeding the autumnal. In this case the summer minimum is lower than the winter minimum, resulting in an anomalous summer to winter variation. The midnight values exhibit no anomalous seasonal behavior, although there is the suggestion of two peaks which lie very close together in summer. The data clearly indicate the presence of several competing phenomena, whose characteristics not only depend on season and phase of the solar cycle, but also on geographical location.

Figure 2 from M. Torr and Torr (1973) illustrates the global morphology of the winter anomaly. These plots were drawn from data taken by 140 ionosonde stations. We see that with the exception of a region over the South Pacific which extends over South America into Central Africa, the winter anomaly occurred over most of the globe in the solar maximum of 1958. The anomaly in the northern hemisphere has its largest amplitude between latitudes 45°N and 60°N, and where it occurs in the southern hemisphere it reaches its maximum amplitude between the same latitudes. The amplitude of the variation in the southern hemisphere, which maximizes over the Australian sector, is not as great as in the northern hemisphere.

In Figure 3, we have used the equinoctial value of f_oF_2 as a reference value and plotted contours of the difference between summer or winter values exceeding this reference. This figure therefore illustrates the geographical regions in which each of the three seasonal variations predominates (i.e. summer, winter, or equinoctial peaks). It is evident from Figure 3a that for 1958 (smoothed sunspot number, SSN = 200) the f_oF_2 peak occurred in winter over most of the northern hemisphere, while it occurred in the equinox months over most of the southern hemisphere. At no station did the peak f_oF_2 occur in summer.

Figure 3b shows a similar plot for 1969 (SSN = 110) for which the average sunspot number was approximately half that during 1958. During 1969 the peak electron density occurred in the equinox months over a much larger portion of the globe, and the winter anomaly region was greatly reduced in extent. In this year a few southern hemisphere stations reached peak f_oF_2 in summer.

Figure 3c shows the situation at solar minimum (SSN = 10). In this case summer peaks in f_oF_2 occur over 30% of the globe.

We do not claim to be capable of explaining all these variations yet. However, these characteristics represent the broad aspects which any theoretical interpretation of the phenomenon must explain. In this paper we show how accurate quantitative modelling of the neutral thermosphere, and ionosphere are essential to gaining even a basic understanding of the above effects. We concentrate primarily on the physics of the ionosphere in this paper, using the empirical MSIS model of Hedin et al. (1977a,b) to represent variations in the neutral thermosphere.

3. Previous Theoretical Work

It was suspected for many years that the main cause of the anomalous summer-winter behavior might be related to seasonal changes in atmospheric composition. This is readily understood in terms of the processes which produce and destroy F region ionization. The primary source of production is photoionization of atomic oxygen, i.e.

$$O + h\nu \xrightarrow{J_1} O^+ + e$$

where J_1 represents the attenuated frequency for photoionization of atomic oxygen by solar EUV radiation.

Removal of O^+ ionization occurs via the process:

$$O^+ + N_2 \xrightarrow{k_2} NO^+ + O \qquad (2)$$

and

$$O^+ + O_2 \xrightarrow{k_3} O_2 + O \qquad (3)$$

which is followed by rapid recombination of the molecular ions with electrons. Reactions (2) and (3) constitute the rate limiting processes (Bates and Massey, 1947).

Early models of the neutral thermosphere predicted that the concentrations of O, O_2 and N_2 would be larger in summer than winter. These models were based on simple hydrostatic formulations in which the exospheric temperature in summer is larger than that in winter due to increased heat input in summer. Since the atomic oxygen scale height is larger than that of N_2 and O_2, it was reasoned that the $[O]/[N_2]$ and $[O]/[O_2]$ ratios would increase from summer to winter resulting in a smaller effective O^+ loss rate in winter.

This mechanism was unable to account for the observed seasonal variations however, because the actual decrease in loss only partially balanced the corresponding decrease in production. If seasonal variations in the atomic to molecular concentration ratios were to account for the anomalous summer-winter behavior, it was clear that anomalous seasonal variations in all or some of the neutral constituents was required. The possibility of composition changes being a cause of the winter anomaly was first suggested by Rishbeth and Setty (1961). Johnsen (1964) suggested that the composition change would arise as a result of the summer hemisphere being heated and the lighter atomic oxygen being convected to the winter hemisphere. In 1968, mass spectrometer measurements by Krankowsky et al. (1968) indicated that the $[O]/[N_2]$ ratio was a factor of 2 larger in winter than in summer.

Cox and Evans (1970) were the first to obtain quantitative information on the potential role played by seasonal variations in the concentrations of O, O_2 and N_2 in producing the anomalous seasonal behavior of O^+. Using incoherent scatter radar measurements of F region ionospheric parameters they were able to independently determine values for the O concentration, and found that it increased from summer to winter by about 50%. When combined with the normal seasonal variation in the molecular concentrations, i.e. larger densities in summer than winter of N_2 and O_2, they were able to obtain qualitative theoretical agreement with the measured seasonal variation in electron density at the peak of the F_2 layer. Cox and Evans (1970) concluded that the winter anomaly was therefore produced by enhanced production and smaller destruction of O^+ ions in winter. Strobel and McElroy (1970) subsequently carried out more quantitative calculations and concluded that the seasonal variation in the neutral concentrations could only partially account for the observed seasonal variations in electron density.

The radar measurements did, however, provide additional information on the nature of the phenomenon. It

was observed that the anomalous behavior is restricted mainly to the vicinity of the F_2 layer, and does not extend into the protonosphere in harmony with topside satellite measurements (c.f. Fatkullin, 1970). The radar data also precluded a dynamical effect as a cause of the additional anomalous ionization, since the F_2 peak occurs at greater altitudes in summer than in winter. To decrease the summer peak densities sufficiently required that the peak occur at a considerably lower height (Strobel and McElroy, 1970) than that observed. Vasseur (1970), Wu and Newell (1972) and Yonezawa (1971) drew similar conclusions. Eccles et al. (1971) showed that while the inclusion of winds results in a better fit to the height of the layer, the winds could not explain the variation in N_{max}. Winds cause downward drift at noon and are larger in winter than in summer. Their effect is therefore to counteract the winter anomaly.

Several workers have suggested enhanced vibrationally excited nitrogen concentrations in summer as a possible cause of the winter anomaly. The rationale is based on a measurement by Schmeltekopf et al. (1967) of the rate coefficient for the process

$$O^+ + N_2^* \rightarrow NO^+ + e \qquad (4)$$

which increased dramatically with N_2 vibrational temperature. If sufficient N_2^* is present, reaction (4) could result in a significant loss of O^+ ions.

Recently Vlasov and Izakova (1978) presented a paper in which they modelled the seasonal variation of the effective vibrational temperature of N_2 in the thermosphere. They claimed that the seasonal variation in this parameter is large enough to affect the destruction of O^+ by N_2, and that the seasonal variations are consistent with the observed variation in the O^+ concentration. Subsequently Richards and Torr (1979) repeated these calculations for solar minimum conditions and found that reaction (4) is only effective if the N_2^* diffusion coefficient is significantly smaller than that for ground state N_2. Subsequent unpublished attempts to theoretically justify the use of a small diffusion coefficient were unsuccessful since the cross-sections required were physically unrealistic. The conclusion was drawn that seasonal changes in the N_2 vibrational population would not enhance the compositional effect significantly at solar minimum.

D. Torr et al. (1980) also identified a third process which would enhance the anomalous seasonal variation in peak electron density. This mechanism involved additional production by metastable O^+ ions. Using the best rate coefficients available at that time, they demonstrated that the inclusion of metastable O^+ production results in larger O^+ concentrations in winter due to enhanced electron quenching. In summer the metastable O^+ ions charge transfer with N_2 to form N_2^+, because of the higher summer N_2 densities. However, this mechanism has been ruled out by recent laboratory measurements of the charge transfer rate coefficient (Johnsen and Biondi, 1980, Rowe et al., 1980) which yielded values an order of magnitude larger than those deduced aeronomically (c.f. D. Torr and Orsini, 1978; D. Torr and Torr, 1979; M. Torr et al., 1980d; M. Torr and Torr, 1980).

At this point the winter anomaly faded into insignificance, because almost the entire structure of thermospheric chemistry that had been built up over seven years of careful analysis of Atmosphere Explorer (AE) satellite data was threatened with collapse. The new laboratory measurement produces discrepancies in the concentrations of certain constituents which considerably exceed those of the anomalous seasonal behavior.

The major problems that arise are listed below:

1.) Theoretical values of the N_2^+ concentration exceed observed values by up to a factor of 4.

2.) The theory is no longer able to account for the magnitude of the observed O^+ concentration, because all the metastable O^+ ionization is channeled into N_2^+.

3.) The theoretical concentration of O_2^+ is too low because the main production mechanism of O_2^+ ions via charge exchange of O^+ with O_2 decreases correspondingly.

In this paper we shall describe the work which led to the resolution of these problems, and which also solved the basic problem of the F region winter anomaly, but not necessarily the details of the geographical variations.

4. Re-analysis of Thermospheric Chemistry

The parameter which is most severely affected by the new laboratory results of Johnsen and Biondi (1980) and Rowe et al. (1980) is the N_2^+ concentration, since the discrepancy in this case is so large (up to a factor of 4) that it is clear that something is radically wrong with our current understanding of the thermospheric chemistry. We list below the sources and sinks of N_2^+ and examine these in detail in order to evaluate where earlier work may have gone wrong.

N_2^+ sources:

$$N_2 + h\upsilon \xrightarrow{J_5} N_2^+ + e \qquad (5)$$

$$O^+(^2D) + N_2 \xrightarrow{k_6} N_2^+ + O \qquad (6)$$

N_2^+ sinks:

$$N_2^+ + O \xrightarrow{k_7} NO^+ + N \qquad (7)$$

$$N_2^+ + e \xrightarrow{k_8} N + N \qquad (8)$$

It was suggested by Biondi (1978) and Johnsen and Biondi (1980) that k_7 may be larger in the ionosphere than in the laboratory, since N_2^+ is known to be vibrationally excited in the atmosphere (Broadfoot, 1967, 1971) ($T_v \sim 4000^oK$). There might be a strong dependence of the rate coefficient on vibrational excitation. The vibrational dependence of this rate has never been studied in the laboratory, and Johnsen and Biondi (1980) point out that the kinetic temperature of the reactants (which are probably in thermal equilibrium) in the laboratory measurement of k_7 (McFarland et al. 1974) did not exceed $\sim 600^oK$, which is low compared to the 4000^oK which occurs in the atmosphere. This hypothesis raises two serious questions from the aeronomical standpoint:

1.) The problem will simply be transferred to NO^+, producing not only excess NO^+ ionization, but height profiles with the wrong scale height.

2.) The O^+ and O_2^+ problems will not be solved.

The NO^+ problem is serious, because the only loss mechanism for this species is by dissociative recombination. An adjustment of the NO^+ recombination rate coefficient at this time is considered untenable for the following reasons. NO^+ has been more extensively studied with the AE data than any other ion species. The determination of the NO^+ rate coefficient was made at night when no N_2^+ ions were measured (D. Torr et al., 1976; M. Torr et al., 1977; M. Torr and Torr, 1979a,b) and hence could not have been affected by uncertainties k_6 and in k_7. The only alternative consideration to date is that a source of NO^+ exists at night, which has escaped detection thus far. The associative ionization reaction

$$N(^2D) + O \xrightarrow{k_9} NO^+ + e \qquad (9)$$

where $N(^2D)$ is hot, has been suggested by M. Biondi (private communication) and by Zipf (1978) as a nocturnal source of NO^+ ions. This possibility was discounted by M. Torr and Torr (1979b). In any event, it is generally conceded that at most, the contribution due to reaction (9) might amount to 30%. We therefore conclude that the problem cannot be solved by increasing k_7.

The next most obvious possibility is that N_2^+ recombines more rapidly in high vibrational levels, an idea originally proposed by Orsini et al. (1977a) and discounted by Biondi (1978). This conclusion was recently confirmed by Zipf (1980) who, in an exceptionally well conducted experiment, measured k_8 as a function of N_2^+ vibrational excitation. Zipf (1980) found no significant change in k_8 with increasing vibrational levels. The evidence therefore is strong that an additional sink for N_2^+ ions is required. If we assume that a single missing process is the cause of our problems (i.e. including the problems with O^+, O_2^+, and the associated seasonal behavior of the F_2 layer), then the conclusion must be drawn that the required process must simultaneously result in the production of $O^+(^4S)$ ions. Careful study of all possible candidates indicates that the only viable single step scheme is:

$$N_2^{+*} + O \xrightarrow{k_{10}} O^+ + N_2 \qquad (10)$$

i.e. charge exchange of vibrationally excited N_2^+ with O. We have therefore adopted the basic chemical scheme for the production of N_2^{+*} ions used by Orsini et al. (1977a), and which is described in detail by Orsini (1977), to calculate the vibrational distribution of N_2^+ ions in the thermosphere. We determine k_{10} as a function of vibrational level. By weighting the value obtained for each level by the relative population, an effective value for k_{10} is derived.

Briefly, the Orsini (1977) model for calculating vibrationally excited N_2^+ ions involves excitation of the A and B states of N_2^+ via resonance absorption of solar UV radiation in the Meinel and first negative bands respectively. Subsequent fluorescence via A - X and B - X transitions results in the production of vibrational quanta in the ground electronic state. The situation is schematically illustrated in Figure 4 for the first three vibrational levels of the X - A transition. All permitted excitation and de-excitation channels for the first five vibrational levels of the X, A, and B states are used in the calculations. The X state ions are destroyed via ion atom interchange (7), recombination (8) as well as charge exchange (10), where k_7 is assumed to be independent of vibrational excitation. An additional process of significance that has been included is exchange of N_2^+ vibrational quanta, with N_2, i.e.

$$N_2^+(X)\ (v=i) + N_2(X)\ (v=j) \rightleftarrows N_2^+(X)\ (v=i\pm1) + N_2(X)\ (v=j\mp1) \qquad (11)$$

The rate coefficients for (11) are somewhat arbitrary, but nevertheless realistic. A precise knowledge of these coefficients is not required, since the main effect of (11) is to produce strong quenching of N_2^{+*} below about 200 km. The necessity for this process is indicated by the studies of D. Torr et al. (1977) and D. Torr (1979) which showed that below 200 km the N_2^+ concentration is controlled by photoionization and loss via (7) (ion atom interchange, see Figure 6). These calculations were found to be quite insensitive to values used for k_6 (i.e. $O^+(^2D) + N_2$), and hence are not affected by the new value for k_6. D. Torr (1979) determined values for k_7 as a function of kinetic temperature using AE data taken below 200 km, and obtained excellent agreement with the laboratory measurements of McFarland et al. (1974). Furthermore, the O^+ concentration could be explained satisfactorily with the conventional chemistry at these altitudes. Below 200 km altitude O^+ is near chemical equilibrium, and there was no indication that an additional source of ionization is required. The inclusion of reaction (11) is also a logical requirement based on the physics of such reactions (D. Albritton, private communication, 1980), and the analogous process is known to be rapid for $O_2^+ - O_2$ vibrational exchange processes.

The above chemical scheme results in an effective N_2^+ vibrational temperature which is in kinetic thermal equilibrium below 200 km, and which is in radiative equilibrium with sunlight at $\sim$ 300 km. Reaction (11) produces a Boltzmann N_2^+ distribution at altitudes above 200 km.

In addition to the above changes, account had to be taken of a new laboratory measurement of k_3 (i.e. for the $O^+ + O_2$ reaction) (Chen et al., 1980), which yielded values about 50% lower at ~800°K with a different temperature dependence than was obtained from the results of Albritton et al. (1977) by St-Maurice and Torr (1978). Since k_3 was also used to aeronomically determine the O_2^+ recombination rate coefficient, this too had to be re-evaluated. The new values have been reported by M. Torr et al. (1980c).

The effect of the new value for k_6 on the chemistry of $O^+(^2D)$ requires a little discussion. The introduction of the large value for k_6 effectively eliminates $O^+(^2D)$ as an important parameter in the chemistry of $O^+(^4S)$. Therefore, further analysis of this species is not pertinent to this paper. However, it might be pointed out that Orsini et al. (1977b) and D. Torr and Orsini (1977) demonstrated that the main effect of a value of 5×10^{-10} cm^3 s^{-1} for k_6 is to increase the upper limit on the rate coefficient for atomic oxygen quenching of $O^+(^2D)$ from ~1×10^{-11} to ~3×10^{-11} cm^3 s^{-1}. The electron quenching rate coefficient is not significantly affected. The impact of the O quenching on $O^+(^4S)$ is marginal in the presence of a large value for k_6, i.e. $\gtrsim 5 \times 10^{-10}$ cm^3 s^{-1}.

At this point we may begin to realize what little hope there was of quantitatively explaining the behavior of the F_2 layer with earlier models. All these changes were incorporated into a comprehensive computer code (Young et al., 1980) which solves the ionospheric continuity, momentum, energy, and photoelectron flux equations. The solution is carried out from the foot of a field line located in the E region of the northern hemisphere, to the E region of the conjugate ionosphere, thereby incorporating the effects of inhemispheric flow of heat, ionization, momentum and photoelectrons on the ionosphere. The basic elements of the code are schematically illustrated in Figure 5. This figure indicates that the code also simultaneously solves for the concentration of the first five vibrational levels of N_2 and evaluates the effects of the vibrational excitation on k_2. We stress that no ad hoc manipulation of parameters was made in order to produce the anomalous seasonal variation of the F_2 layer. As will be seen from the results presented in the next section, the basically correct seasonal behavior emerges self-consistently from the model.

In its final form therefore, the model incorporates all the latest available data on cross-sections, rate coefficients, and solar fluxes. The MSIS model atmosphere (Hedin et al, 1977 a,b) is used with a correction applied for the solar cycle variation of O_2 as was discussed by Kayser et al. (1980). We have established that the model reproduces both incoherent scatter radar and AE measurements of electron ion temperatures and concentrations when measured input data are available (Young et al., 1980). In the results presented below we compare our results mainly with mean values in order to simulate the general features of the anomalous seasonal behavior, because the day to day variations of the F_2 layer are sporadic as a result of dynamical effects. It should be noted that these calculations cannot be expected to reproduce the detailed longitudinal variations discussed earlier, because the MSIS model atmosphere represents zonally averaged values.

Figure 6 schematically summarizes the revised concepts of the main chemical processes which govern the behavior of ionization in F_2 layer. The dashed lines indicate processes of minor significance. In the new scheme, production of metastable O^+ ions indirectly results in an equivalent production of N_2^+ ions. Therefore the chemistry of the metastable O^+ ions is not discussed further. The rate coefficients for the remaining reactions are given in Table 1.

5. Results

In Figure 7 we show the effect of the new process reaction 10; i.e.

$$N_2^{+*} + O \xrightarrow{k_{10}} O^+(^4S) + N_2$$

on the N_2^+ concentration. The solid line shows the results obtained from the earlier model which reproduced the AE N_2^+ measurements to within ± 30%. The measurements were made over the range geophysical conditions encountered by the spacecraft over the last seven years. Hence these results can effectively be used to represent the N_2^+ measurements parametrically, even if the physics of the model is invalid. We have selected an occasion of low O^+ concentration here, because this maximizes the effect of the large charge exchange rate coefficient (k_6) on the N_2^+ concentration. We have chosen to adopt a value of 5×10^{-10} cm^3 s^{-1} for k_6, which lies close to the lower error limit of the experimental results. The reason for this choice is that larger values for k_6 require correspondingly larger values for k_{10}, which begin to look unreasonable, i.e. in excess of 5×10^{-10} cm^3 s^{-1}. The curve labelled (1) shows the effect of increasing k_6 from the aeronomically deduced value of 5×10^{-11} cm^3 s^{-1} to 7×10^{-10} cm^3 s^{-1}. This results in an increase in $[N_2^+]$ of nearly a factor of 2 at 325 km. The dashed curve shows the effect of using an effective value of 5×10^{-10} cm^3 s^{-1} for k_{10}. This value yields values for the N_2^+ concentration which agree with measurements to within 30% over a wide range of O^+ concentrations.

In Figure 8 we show the O^+ production rates for Millstone Hill in summer for solar maximum of 1979. The significance of reaction 10 is evident. We see that this source is comparable with direct photoionization of atomic oxygen at heights below ~250 km. Figure 9 shows the corresponding ion concentrations. Also shown are typical peak electron densities (N_{max}) observed at Millstone Hill during the solar maximum of 1969. Measurements are not available for 1979, and input data required by the model are not available for 1969. The Millstone Hill values are about 50% lower than the peak theoretical O^+ concentrations shown. Figure 10 shows the effect of including vibrationally excited N_2 on the $O^+ + N_2$ loss rate. The corresponding densities are shown in Figure 11. The effect is to decrease N_{max} by ~30%. The effect on NO^+ is quite significant. Although the effect on N_{max} is fairly small, it is sufficient to reduce the disagreement between experiment and theory from ~50% to less than 30%, that is if the 1969 measurements can be used to represent the 1979 epoch. In any event, the results demonstrate the principle of the mechanisms involved.

We do not address the question of the height of the layer in this paper, since this is affected by day-to-day variations in the thermospheric winds. The profiles shown in Figures 9 and 11 were obtained

using realistic values for thermospheric winds, which are numerically inserted into the code. Figure 12 shows the winter theoretical results for Millstone Hill in January, 1979, compared with 1969 measurements. In this case the theoretical $N_{max} = 1.4 \times 10^6$ cm^{-3} which agrees with typical results observed at Millstone Hill at solar maximum in January. These theoretical values yield a net seasonal variation in N_{max} of a factor of 2.3, compared with observed variations of $\sim$ 2.5. Our calculations do indicate, however, that the effect of the vibrationally excited N_2 is quite sensitive to the $[N_2]/[O_2]$ ratio, and to absolute concentrations of N_2 and O_2. If the ratio increases, loss of O^+ with N_2 becomes relatively more important and this enhances the seasonal anomaly, because vibrationally excited N_2^* does not have a significant effect in winter. When the neutral concentrations increase, diffusion of N_2^* molecules is inhibited, which results in a build up of N_2^* in the F_2 region. Day-to-day variations in $[N_2]$ at 300 km can amount to a factor of two, which gives rise to the well known sporadic nature of occurrence of the seasonal anomaly. It is not possible to reproduce this kind of variability with our code in its present form, since this would require detailed in situ altitude information on the input parameters, which are not available. Furthermore, uncertainties still exist concerning the variation of the $[N_2]/[O_2]$ ratio as a function of solar cycle (Kayser, 1980). While our results indicate that the winter anomaly can be explained largely in terms of seasonal variations in the concentrations of the neutral constituents, it is clear that an additional mechanism is required at solar maximum. This mechanism could very well be vibrationally excited N_2.

Figure 13 shows the seasonal variations in N_{max} at solar minimum. In this case the anomalous variations are explained quite satisfactorily by seasonal changes in neutral composition alone. The effect of vibrationally excited N_2 is negligible. The early results of Strobel and McElroy (1970) were based on incomplete knowledge of thermospheric chemistry, and are therefore no longer valid. The seasonal variation obtained for Millstone Hill is consistent with the characteristic northern hemisphere behavior, i.e. a peak in winter. This variation is attributable entirely to the variations in neutral composition depicted by the MSIS model atmosphere.

The corresponding results for Cape Town (33° S, 18° E) which showed no winter maximum are presented in Figure 14. In this case the seasonal anomaly is greatly reduced. We reiterate that the vintage of the MSIS model that we have used represents longitudinally averaged data. Hence we would expect a mean variation which includes that observed in the Australian, African and South American sectors. The Australian sector exhibits anomalous seasonal behavior whereas the remaining southern hemisphere sectors do not.

Finally in Figure 15 we show calculations of Dickenson et al. (1980) of the zonally averaged meridional circulation at the solstices at sunspot maximum. These calculations show very clearly a general flow from the summer (southern) hemisphere to the winter hemisphere, except in the winter hemisphere above 200 km, where the high latitude joule heat source drives a counter circulation to about 40° latitude. At approximately 40° N latitude there is a downward convergence of air from both hemispheres. Dickenson et al (1980) point out that their calculations are based on a pre-AE determination of the solar EUV heating efficiency of $\sim$ 33% (Stolarski et al., 1975). The recent results of M. Torr et al. (1980a,b,e) have shown that the AE data yield a value of $\sim$ 55%. The net effect will be to drive the subsidence zone to higher latitudes, placing the region of the largest anomalous high $[O]/[N_2]$ ratio between 40° and 60°, in harmony with the observed regions of occurrence of the winter anomaly.

A final point worth mentioning is that the model of Dickensen et al. (1980) assumes hemispheric seasonal symmetry in the high latitude heat sources, which would not yield the asymmetric hemispheric characteristics of the anomalous seasonal behavior in N_{max}. M. Torr and Torr (1973) argued that hemispheric symmetry should not be expected, because of the asymmetrical configuration of the earth's geomagnetic field. It is well known that the magnetic intensities are anomalously low over South America and the South Atlantic sectors. The implication is that auroral particle precipitation should be enhanced in these regions as has been demonstrated by Gledhill et al. (1967) and M. Torr and Torr (1972). The larger energy influx over these regions in summer may well generate a longitudinally asymmetric meridional wind field in the southern hemisphere. Recent results by Berger and Barlier (1980) indicate that an asymmetry in the neutral concentrations does indeed exist over the South Atlantic, and they re-iterate the arguments of M. Torr and Torr (1973) in explaining their measurements. Further quantitative analysis of these hypotheses will have to await the next generation of three dimensional thermospheric global models which take longitudinal asymmetries of heat sources into account.

6. Conclusions

In this paper we identify erroneous chemistry as the basic cause of difficulties encountered in the past in explaining the behavior of the F_2 layer of the ionosphere. Another prerequisite to successful theoretical interpretation of ionospheric behavior is accurate knowledge of controlling geophysical parameters, such as the concentration and temperature of the neutral atmosphere, and the solar EUV flux beyond the atmosphere. In this analysis we find that with the MSIS model atmosphere (Hedin et al., 1977a,b), the Atmosphere Exlorer measurements of Hinteregger et al. (1973) of the solar EUV flux (M. Torr et al., 1979), and the revised chemistry presented in this paper, we are able to reproduce the basic global and solar cyclic variations of the seasonal behavior of the F_2 layer, i.e. including the anomalous seasonal variations.

The new chemistry involves charge exchange of nearly all metastable O^+ ions to N_2^+ via N_2. This is followed by vibrational excitation of N_2^+ by a process of resonance fluorescence, and a second charge exchange reaction of vibrationally excited N_2^+ with O restoring the original O^+ ion. In addition at solar maximum the production of N_2 vibrational excitation is sufficiently enhanced to produce an additional 30% reduction in the peak electron density in summer. The effect is smaller in winter and negligible at solar minimum. Hence given the correct chemistry, the seasonal anomaly can be explained largely by seasonal and geographical variations in neutral thermospheric composition as represented by the MSIS model atmosphere. The MSIS concentrations and temperatures are qualtitatively consistent with the convection of atomic constituents from the summer to winter hemisphere as described by a three dimensional general circulation model of Dickensen et al. (1980). Therefore the basic characteristics of the F region winter anomaly are explained. In future when the longitudinal variation of neutral constituents are parameterized, the

geographic variations of the seasonal anomaly can be investigated in finer detail.

7. Acknowledgements

This work was supported at Utah State University by NASA Grant NAS5-24331, NSF Grant ATM-8005088, and a grant of computer resources by the National Center for Atmospheric Research which is sponsored by the National Science Foundation.

8. References

ALBRITTON, D. L., I DOTAN, W. LINDINGER, M. McFARLAND, J. TELLINGHUISER, and F. C. FEHSENFELD, 1977, "Effects of Ion Speed Distributions in Flow Drift Tube Studies on Ion-Neutral Reactions", J. Chem. Phys., 66, 410.

BATES, D. R., H W. S. MASSEY, 1947, "The Basic Reactions in the Upper Atmosphere. II. The theory of Recombination in Ionized Layers", Proc. Roy. Soc., (London), A192, 1-16.

BERGER, C., and F. BARLIER, 1980, "Response of the Equatorial Thermosphere to Magnetic Activity Analysed with Accelerometer Total Density Data. Asymmetrical Structure", J. Atmos. Terr. Phys., in press.

BIONDI, M. A., 1978, "Objections to $N_2^+ + e^-$ Dissociative Recombination Coefficients Inferred from Analysis of atmosphere Explorer Measurements", Geophys. Res. Lett., 5, 661.

BROADFOOT, A. L., 1967, "Resonance Scattering by N_2^+", Planet. Space Sci., 15, 1801.

BROADFOOT, A. L., 1971, "Dayglow Nitrogen Band Systems", The Radiating Atmosphere, ed. B. M. McCormac, D. Reidel Publishing Co., Dordrecht, Holland, pp. 33-34.

CHEN, A., R. JOHNSEN, and M. A. BIONDI, 1980, "Measurements of the $O^+ + N_2$ and O+ + O_2 Reaction Rates from 300 to 900 K", J. Chem. Phys., 69, 2688.

COX, L. P., and J. V. EVANS, 1970, "Seasonal Variation of the O/N_2 Ratio in the F_1 Region", J. Geophys. Res., 75, 6271.

DICKENSON, R. E., E. C. RIDLEY, and R. G. ROBLE, 1980, "A Three Dimensional Circulation Model of the Thermosphere", submitted to J. Geophys. Res.

ECCLES, D., J. W. KING, H. KOHL, and R. J. PRATT, 1971, "The Semi-Annual Variation in the Height of the F2-Layer Peak", J. Atmos. Terr. Phys., 33, 1641.

EVANS, J. E., 1975, "A Study of F2 Region Daytime Vertical Ionization Fluxes at Millstone Hill During 1969", Planet. Space Sci., 23, 1461.

FATKULLIN, M. N., 1970, "The Seasonal Anomaly in Electron Density in the Topside F-2 Region", J. Atmos. Terr. Phys., 32, 1067.

GLEDHILL, J. A., D. G. TORR, and M. R. TORR, 1967, "Ionospheric Disturbance and Electron Precipitation from the Outer Radiation",J. Geophys. Res., 72, 208.

HEDIN, A. E., J. E. SALAH, J. V. EVANS, C. A. REBER, G. P. NEWTON, N. W. SPENCER, D. C. KAYSER, D. ALCAYDE, P. BAUER, L. COGGER, and J. P. McCLURE, 1977a, "A Global Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data, MSIS 1, N_2 Density and Temperature", J. Geophys. Res., 82, 2139.

HEDIN, A. E., C. A. REBER, G. P. NEWTON, N. W. SPENCER, H. C. BRINTON, H. G. MAYR, and W. E. POTTER, 1977b, "A Global Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data: MSIS 2 - Composition", J. Geophys. Res., 82, 2148.

HINTEREGGER, H. E., D. E. BEDO, and J. E. MANSON, 1973, "The EUV Spectrophotometer on Atmosphere Explorer", Radio Sci., 8, 349.

JOHNSEN, R., and M. A. BIONDI, 1980, "Laborataory Measurements of the $O^+(^2D) + N_2$ and $O^+(^2D) + O_2$ Reaction Rate Coefficients and Their Ionospheric Implications, Geophys. Res. Lett., 7, 401.

JOHNSON, F. S., 1964, Electron Density Distribution in Ionosphere and Exosphere, (ed. E. Thrane), North-Holland, Amsterdam, p. 81.

KAYSER, D. C., 1980, "Solar Flux Variation of the Thermospheric Molecular Oxygen Density", J. Geophys. Res., 85, 695.

KRANKOWSKY, D., W. T. KASPRAZAK, and A. O. NIER, 1968, "Mass Spectrometric Studies of the Composition of the Lower Thermosphere During Summer 1967", J. Geophys. Res., 73, 7291.

McFARLAND, M., D. L. ALBRITTON, F. C. FEHSENFELD, E. E. FERGUSON, and A. L. SCHMELTEKOPF, 1974, "Energy Dependence and Branching Ratio of the $N_2^+ + O$ Reaction", J. Geophys. Res., 79, 2925.

MEHR, F. J., and M. A. BIONDI, 1969, "Electron Temperature Dependence of Recombination of O_2^+ and N_2^+ Ions with Electrons", Physical Review, 181, 264.

ORSINI, N., 1977, "The Vibrationally Excited N_2^+ Ion and the Twilight 4278 Å Emission Band in the Earth;s Mid-Latitude Thermosphere", Ph.D. Dissertation, Space Physics Research Laboratory, Department of

Atmospheric and Oceanic Sciences, University of Michigan, Ann Arbor, Michigan, Technical Report 010554/013716, September.

ORSINI, N., D. G. TORR, H. C. BRINTON, L. H. BRACE, W. B. HANSON, J. H. HOFFMAN, and A. O. NIER, 1977a, "Determination of the N_2^+ Recombination Rate Coefficient in the Ionosphere", J. Geophys. Res., 4, 431.

ORSINI, N., D. G. TORR, M. R. TORR, H. C. BRINTON, L. H. BRACE, A. O. NIER, and J. C. G. WALKER, 1977b, "Quenching of Metastable ^2D Oxygen Ions in the Thermosphere by Atomic Oxygen", J. Geophys. Res., 82, 4829.

RICHARDS, P. G., and D. G. TORR, 1979, "An Experimental and Theoretical Study of Vibrationally Excited N_2 Molecules in the Thermosphere", to be submitted to J. Geophys. Res.; paper presented at IAGA Symposium 11-4, Thermospheric Photochemistry, Canberra, Australia.

RISHBETH, H., and C. S. G. K. SETTY, 1961, "The F-Layer at Sunrise", J. Atmos. Terr. Phys., 20, 263.

ROWE, B. R., D. W. FAHEY, F. C. FEHSENFELD, and D. L. ALBRITTON, 1980, "Rate Constants for the Reactions of Metastable O^- Ions with N_2 and O_2 at Collision Energies 0.04 to 0.2 eV and the Mobilities of These Ions at 300 K", submitted to J. Chem. Phys.

ST. MAURICE, J. -P., and D. G. TORR, 1978, "Non-Thermal Rate Coefficients in the Ionosphere: The Reactions of O^+ with N_2, O_2 and NO", J. Geophys. Res., 83, 969.

SCHMELTEKOPF, A. L., F. C. FEHSENFELD, G. I. GILMAN, and E. E. FERGUSON, 1967, "Reaction of atomic Oxygen Ions with Vibrationally Excited Nitrogen Molecules", Planet. Space Sci., 15, 401.

STOLARSKI, R. S., P. B. HAYS, and P. B. HAYS, 1975, "Atmospheric Heating by Solar EUV Radiation", J. Geophys. Res., 80, 2266.

STROBEL, D. F., and M. B. McELROY, 1970, "The F_2 Layer at Middle Latitudes", Planet. Space Sci., 18, 1181.

TORR, D. G., 1979, "Refinement to the Aeronomic Determination of the Rate Coefficient for the $N_2^+ + O$ Reaction", J. Geophys. Res., 84, 1939.

TORR, D. G., and N. ORSINI, 1977, "Charge Exchange of Metastable ^2D Oxygen Ions with N_2 in the Thermosphere", Planet. Space Sci., 25, 1171.

TORR, D. G., and N. ORSINI, 1978, "The Effect of N_2^+ Recombination on the Aeronomic Determination of the Charge Exchange Rate Coefficient of $O^+(^2D)$ with N_2", Geophys. Res. Lett., 5, 657.

TORR, D. G., and M. R. TORR, 1980, "Determination of the Thermal Rate Coefficient, Products, and Branching Ratios for the Reaction of $O^+(^2D)$ with N_2", J. Geophys. Res., 85, 783.

TORR, D. G., J. C. G. WALKER, L. H. BRACE, H. C. BRINTON, W. B. HANSON, J. H. HOFFMAN, A. O. NIER, and M. OPPENHEIMER, 1976, "Recombination of NO^+ in the Ionosphere", Geophys. Res. Lett., 3, 209.

TORR, D. G., N. ORSINI, M. R. TORR, W. B. HANSON, J. H. HOFFMAN, and J. C. G. WALKER, 1977, "Determination of the Rate Coefficient for the $N_2^+ + O$ Reaction in the Ionosphere", J. Geophys. Res., 82, 1631.

TORR, D. G., M. R. TORR, and P. G. RICHARDS, 1980, "Causes of the F Region Winter Anomaly", Geophys. Res. Lett., 7, 301.

TORR, M. R., and D. G. TORR, 1972, "Ionization Enhancements in the Post-Midnight Hours at Some Antarctic Stations and the Diurnal Excursion of the Auroral Oval", South African J. of Antarctic Res., 1, 32.

TORR, M. R., and D. G. TORR, 1973, "The Seasonal Behavior of the F_2-Layer of the Ionosphere", J. Atmos. Terr. Phys., 35, 2237.

TORR, M. R., and D. G. TORR, 1979a, "Recombination of NO^+ in the Midlatitude Trough and the Polar Ionization Hole", J. Geophys. Res., 84, 4316.

TORR, M. R., and D. G. TORR, 1979b, "Associative Recombination of $N(^2D)$ and O", Planet. Space Sci., 27, 1233.

TORR, M. R., and D. G. TORR, 1980, "Temperature Dependence of the Rate Coefficient for Charge Exchange of Metastable $O^+(^2D)$ with N_2", Geophys. Res. Lett., 7, 103.

TORR, M. R., J.- P. ST. MAURICE, D. G. TORR, 1977, "The Rate Coefficient for the $O^+ + N_2$ Reaction in the Ionosphere", J. Geophys. Res., 22, 3287.

TORR, M. R., D. G. TORR, R. A. ONG, and H. E. HINTEREGGER, 1979, "Ionization Frequencies for Major Thermospheric Constituents as a Function of Solar Cycle 21", Geophys. Res. Lett., 6, 771.

TORR, M. R., P. G. RICHARDS, and D. G. TORR, 1980a, "Solar EUV Energy Budget of the Thermosphere", Space Research, 21, in press.

TORR, M. R., P. G. RICHARDS, and D. G. TORR, 1980b, "A New Determination of the Ultraviolet Heating Efficiency of the Thermosphere", J. Geophys. Res., in press.

TORR, M. R., D. G. TORR, and D. KAYSER, 1980c, "The Dissociative Recombination of O_2^+ in the Ionosphere", to be submitted to J. Geophys. Res., August.

TORR, M. R., D. G. TORR, and R. A. ONG, 1980d, "Comparison of the N_2^+ Photochemistry at Different Phases of the Solar Cycle", J. Geophys. Res., 85, 2171.

TORR, M. R., D. G. TORR, and P. G. RICHARDS, 1980e, "The Solar Ultraviolet Heating Efficiency of the Midlatitude Thermosphere", Geophys. Res. Lett., 7, 373.

VASSEUR, G., 1970, "Dynamics of the F-Region Observed with Thomson Scatter - II. Influence of Neutral Air Winds on the Ionospheric F-Region", J. Atmos. Terr. Phys., 32, 775.

VLASOV, M. N., and T. M. IZAKOVA, 1978, "Vibrationally Excited Nitrogen and F_2 Winter Anomaly", paper presented at the International Symposium on Solar Terrestrial Physics, held in Innsbruck, Austria, May-June.

WU, M. -F., and R. E. NEWELL, 1972, "Computer Simulation of the F-Region Seasonal Anomaly", J. Atmos. Terr. Phys., 34, 1635.

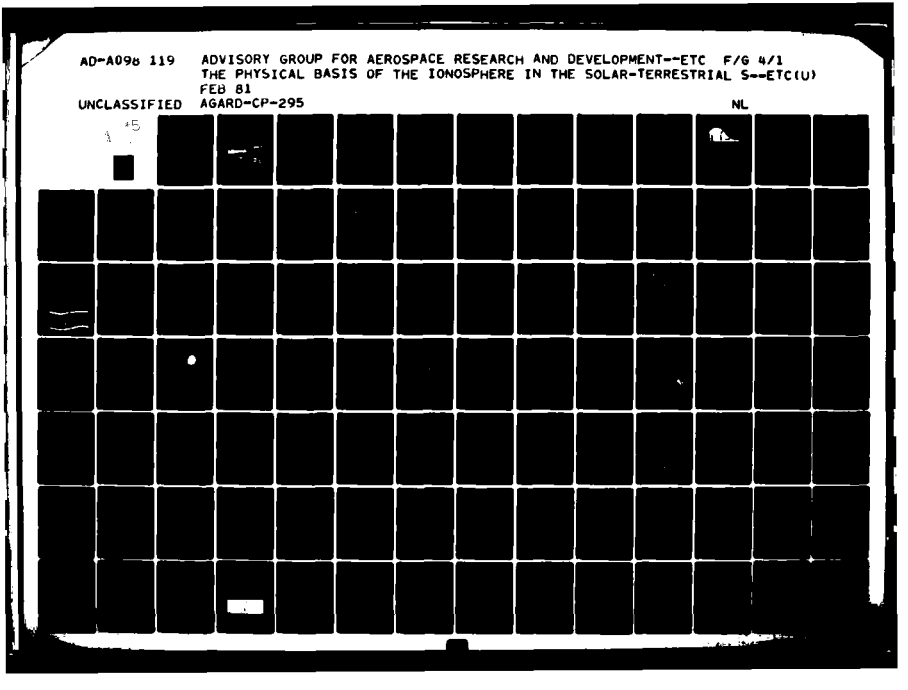
YONEZAWA, T., 1971, "The Solar-Activity and Latitudinal Characteristics of the Seasonal, Non-Seasonal and Semi-Annual Variations in the Peak Electron Densities of the F2-Layer at Noon and at Midnight in Middle and Low Latitudes", J. Atmos. Terr. Phys., 33, 889.

YOUNG, E. R., D. G. TORR, P. G. RICHARDS, and A. F. NAGY, 1980, "A Computer Simulation of the Midlatitude Plasmasphere and Ionosphere", Planet. Space Sci., in press.

ZIPF, E., 1978, "N(^{2}P) and N(^{2}D) Atoms: Their Production by e-impact of N_2 and Destruction by Associative Ionization", paper presented at the Spring Meeting of the AGU, Miami, Florida, Abstract in EOS, 59, 336.

ZIPF, E. C., 1980, "The Dissociative Recombination of Vibrationally Excited N_2^+ Ions", Geophys. Res. Lett., in press.

AD-A098 119 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 4/1
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL S--ETC(U)
FEB 81
UNCLASSIFIED AGARD-CP-295 NL



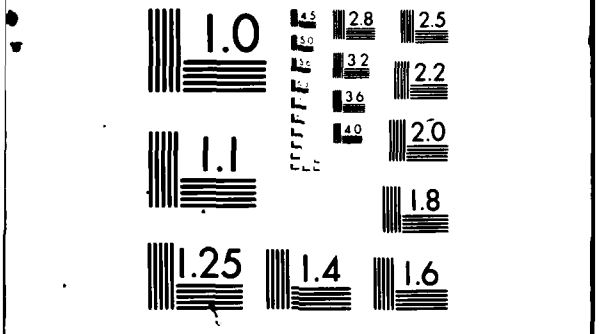
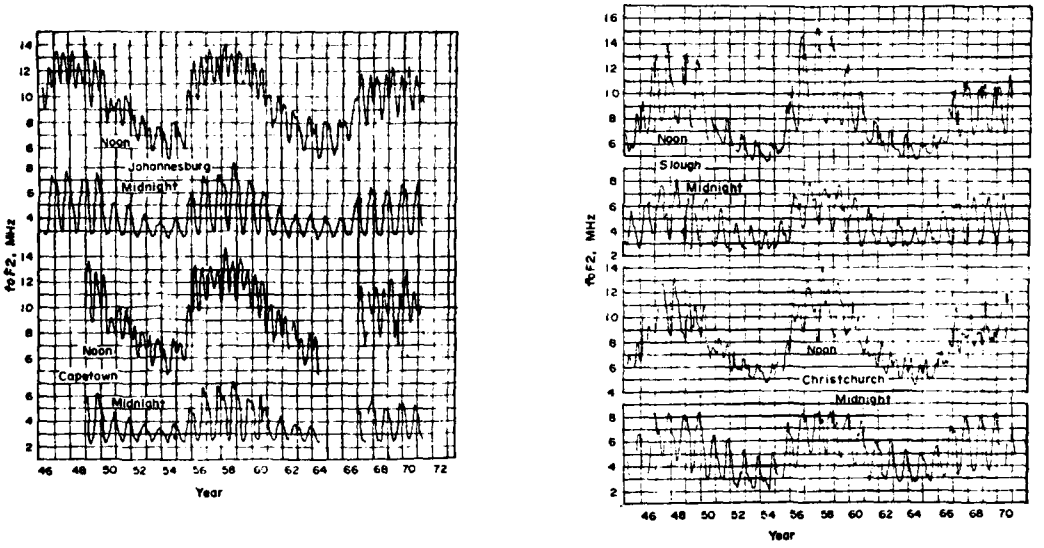


TABLE 1

Tabulation of Reactions and Rate Coefficients

		Reference
1) $O + h\nu \xrightarrow{J_1} O^+(^4S) + e$	(J_1 varies with solar cycle)	M. Torr et a. (1979)
2) $O^+ + N_2 \underset{k_2}{\rightarrow} NO^+ + O$	$k_2 = 1.533 \times 10^{-12} - 5.92 \times 10^{-13} \times (T_{eff}/300) + 8.6 \times 10^{-14} (T_{eff}/300)^2$ for 300 K $\leq T_{eff} \leq$ 1700 K and $k_2 = 2.73 \times 10^{-12} - 1.155 \times 10^{-12} (T_{eff}/300) + 1.483 \times 10^{-13} \times (T_{eff}/300)^2$ for 1700 $< T_{eff} <$ 6000 K	Albritton et al. (1977); St. Maurice and Torr (1978); Chen et al. (1980)
3) $O^+ + O_2 \underset{k_3}{\rightarrow} O_2^+ + O$	$k_3 = 1.25 \times 10^{-17} T_{eff}^2 - 3.7 \times 10^{-14} T_{eff} + 3.0975 \times 10^{-11}$	Chen et al. (1980); parameterization by M. Torr et al. (1980c)
4) $O^+ + N_2^* \underset{k_4}{\rightarrow} NO^+ + e$	k_4 varies with T_v	The rate coefficient is computed as a function of vibrational excitation using the results of Schmeltekopf et al. (1967).
5) $N_2 + h\nu \xrightarrow{J_5} N_2^+ + e$	(J_5 varies with solar cycle)	M. Torr et al. (1979)
6) $O^+(^2D) + N_2 \underset{k_6}{\rightarrow} N_2^+ + O$	$k_6 \approx 5 \times 10^{-10}$ cm^3 s^{-1}	Lower limit of values from Johnsen and Biondi (1980); Rowe et al. (1980)
7) $N_2^+ + O \underset{k_7}{\rightarrow} NO^+ + N$	$k_7 = 1.4 \times 10^{-10} (T_i/300)^{-0.44}$ for $T_i < 1500$ K	McFarland et al. (1974); D. Torr et al. (1977); D. Torr (1979)
8) $N_2^+ + e \underset{k_8}{\rightarrow} N + N$	$k_8 = 1.8 \times 10^{-7} (T_e/300)^{-0.39}$	Mehr and Biondi (1969); Zipf (1980)
9) $N(^2D) + O \underset{k_9}{\rightarrow} NO^+ + N$	$k_9 \approx 10^{-11}$ cm^3 s^{-1}	Zipf (1978). See discussion by M. Torr and Torr (1979b) on k_9.
10) $N_2^{+*} + O \underset{k_{10}}{\rightarrow} O^+ + N_2$	$k_{10} = 5 \times 10^{-10}$ cm^3 s^{-1}	This paper.
11) $N_2^+(X)(v=i) + N_2(X)(v=j) \rightleftarrows N_2^+(X)(v=i\pm1) + N_2(X)(v=j\mp1)$	$k_{r11} = k_{f11} \approx 1 \times 10^{-9}$ cm s^{-1}	This paper.



1a 1b

Figure 1. Variation of noon and midnight monthly median values of f_oF_2 over a number of years at a) Johannesburg and Cape Town, b) Slough and Christchurch. Vertical lines are drawn at 1 January each year (from M. Torr and Torr, 1973).

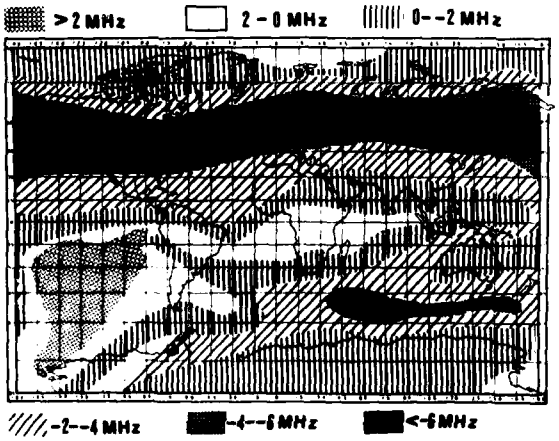


Figure 2. Global distribution of noon $(\overline{f_oF_2}$ summer $- \overline{f_oF_2}$ winter$)$ for 1958 where the bar represents the monthly median values (from M. Torr and Torr, 1973).

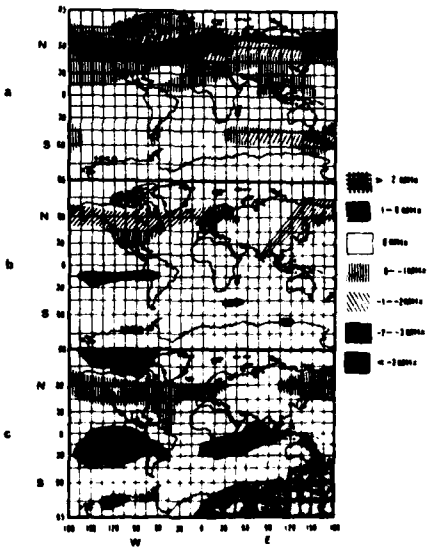


Figure 3. Global contours of $\overline{f_oF_2}$ summer $- \overline{f_oF_2}$ equinox). Cross-hatched codes 1a and 1b on the right hand side indicate areas where the seasonal maximum in f_oF_2 occurs in summer. The symbols a and b designate the following ranges: a) > 1 MHz; b) 0-1 MHz. Areas where greatest f_oF_2 occurs at equinox are indicated by the unhatched area (2). Areas where the anomalous seasonal behavior peaks in winter are coded by cross-hatched areas 3a to 3d, where the symbols a to d indicate the following ranges: a) 0-1 MHz, b) 1-2 MHz, c) 2-3 MHz, d) >3 MHz (from M. Torr and Torr, 1973).

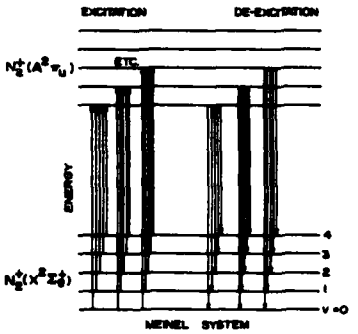


Figure 4. A schematic illustration of N_2^+ X - A excitation and decay transitions which lead to the production of N_2^+ vibrational quanta in the ground electronic state.

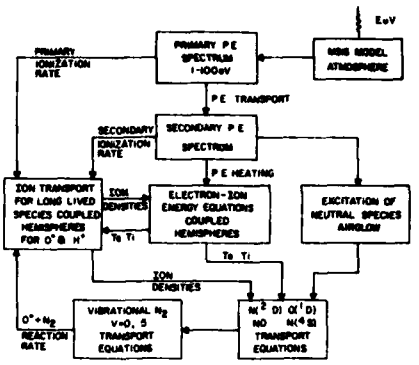


Figure 5. Block diagram illustrating the main processes used in the code used to solve for the F region electron density.

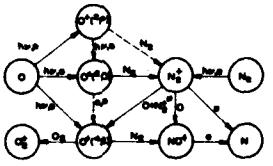


Figure 6. Schematic summary of the revised thermospheric chemistry used in the present calculations.

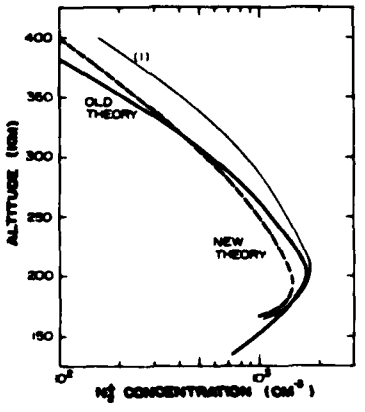


Figure 7. Theoretical values of the N_2^+ concentration as a function of altitude for low O^+ concentrations: ——— old theory these values represent the measurements; ——— old theory with the new laboratory result for k_6; ····· new theory including the reaction $N_2^+ + O \xrightarrow{k_{10}} O^+ + N_2$.

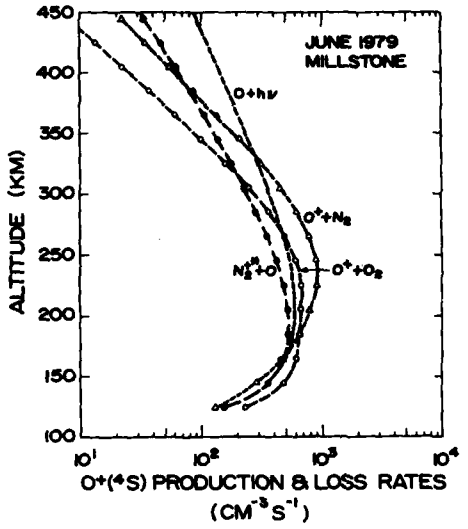


Figure 8. O^+ production and loss rates at Millstone Hill as a function of altitude for solar maximum summer conditions in 1979. The effect of vibrationally excited nitrogen is not included.

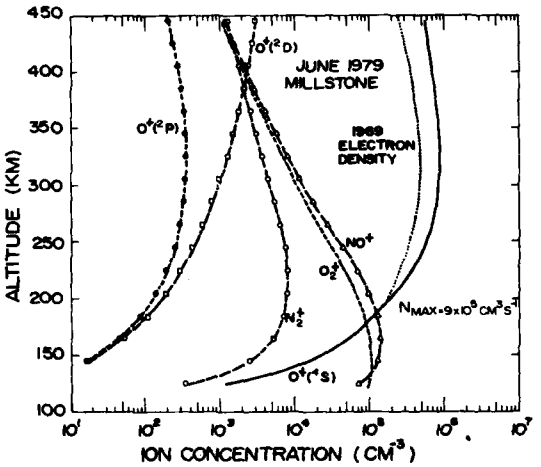


Figure 9. Ion concentrations as a function of altitude at Millstone Hill summer solar maximum. These concentrations correspond to the O^+ sources and sinks shown in Figure 8 (electron densities from Evans, 1975).

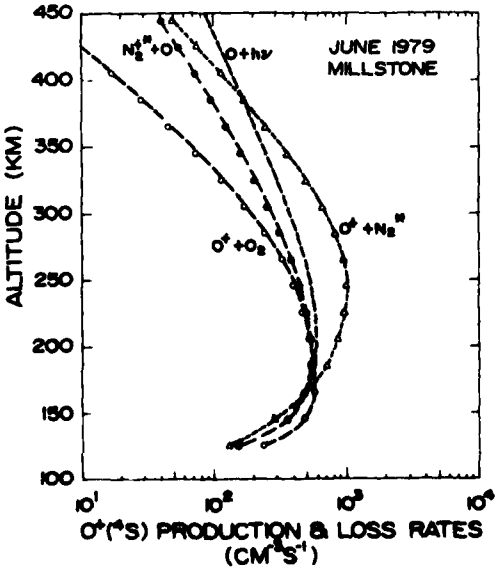


Figure 10. Same as Figure 8, but with vibrationally excited N_2 included.

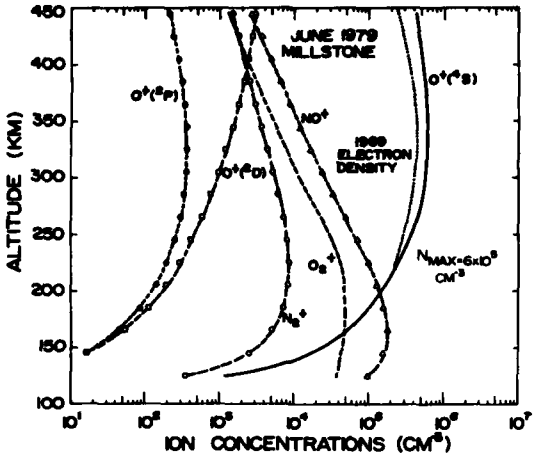


Figure 11. Ion concentrations which correspond to the sources and sinks shown in Figure 10 (electron densities from Evans, 1975).

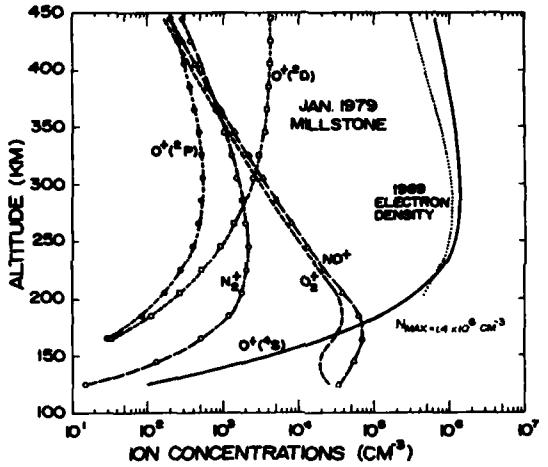


Figure 12. Ion concentrations as a function of altitude at Millstone Hill for solar maximum winter conditions in 1979 (electron densities from Evans, 1975).

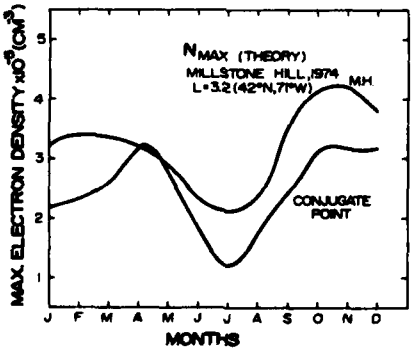


Figure 13. Seasonal variation of the maximum electron density at Millstone Hill and the southern hemisphere conjugate point in 1974.

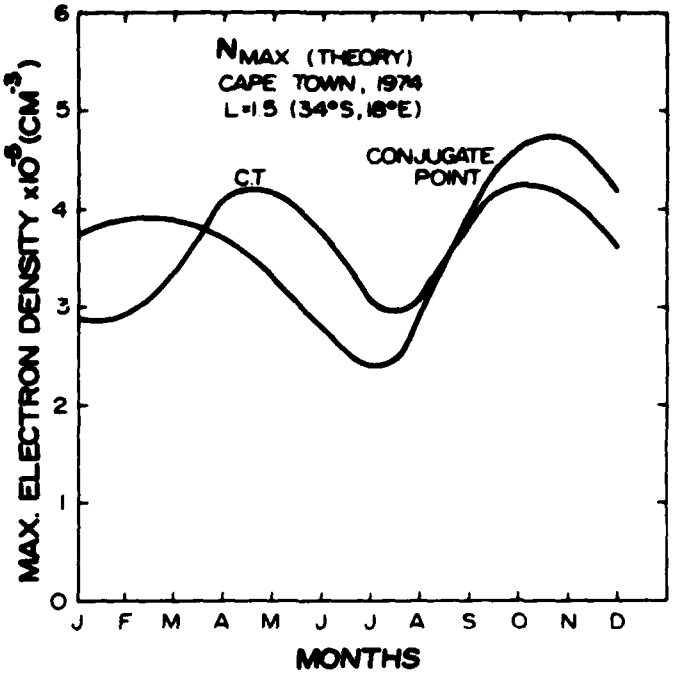


Figure 14. Seasonal variation in N_{max} at Cape Town and the northern hemisphere conjugate point in 1974.

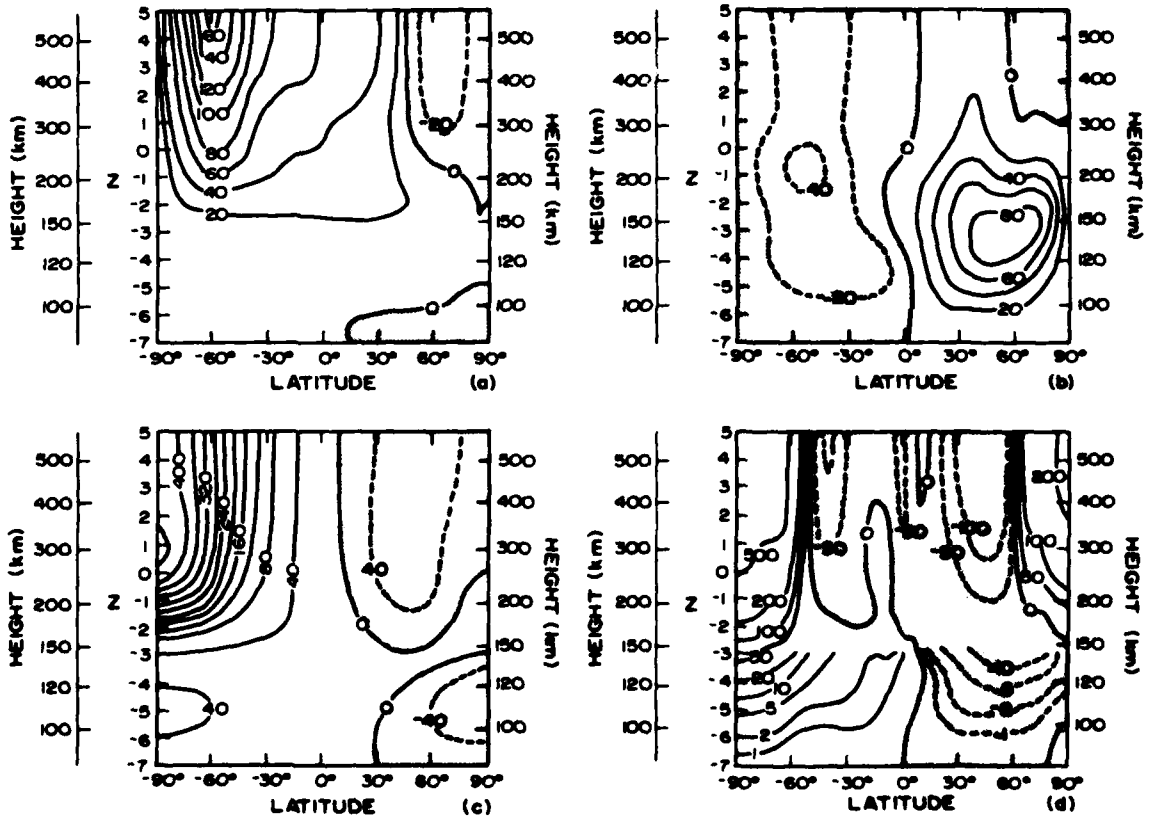


Figure 15. Calculated contours of the perturbation zonal mean temperature and winds from Dickenson et al. (1980). a) mean meridional winds (ms^{-1}), b) mean zonal winds (ms^{-1}), c) perturbation temperature (k), d) vertical winds.

THERMOSPHERIC NEUTRAL COMPOSITION CHANGES AND THEIR CAUSES

P. BAUER
CNET/CNRS
3 Av. de la République 92131 Issy-les-Moulineaux FRANCE

SUMMARY

Since 1961 the so called Winter F region anomaly has drawn the attention of scientists to possible seasonal changes in thermosphere neutral composition other than those simply explained in terms of thermal expansion. Similarly changes in composition associated with magnetic storms have been observed in many instances. Two mechanisms are likely candidates for inducing the composition changes : variations in the turbulent state of the lower thermosphere and atmospheric circulation cells. Due to observational difficulties, there is a lack of definite experimental evidence of variations in the turbulent state of the lower thermosphere. On another hand, both experimental and theoretical works indicate the existence of selective transports of the various thermospheric gases according to their atomic or molecular masses on a planetary scale, due either to solar heating or to auroral heating. The combined effects of Photochemestry and transport appear also to be responsible of seasonal and solar cycle changes in the molecular oxygen concentration. This last type of variations seems to be one of the major causes of the F region Winter anomaly.

I. INTRODUCTION

The Winter F region anomaly has since 1961 drawn the attention of the scientific community to the existence of seasonal thermospheric composition changes. Indeed, a detailed study conducted by Rishbeth and Setty (1961) has shown that the observed effects, in particular the winter enhancement of the F region ionospheric layer at low altitude, could only be explained by neutral composition changes other than those resulting from the thermal breathing of the atmosphere. Another characteristic which was identified, was the fact that the anomaly is particularly developed around the solar cycle maximum. The subsequent flow of in situ measurements obtained from rocket flights or from satellites has allowed to rapidly verify the existence of seasonal composition changes and to discover composition changes induced by the magnetic activity. The purpose of this paper is to present various types of thermospheric composition changes, to analyse the various mechanisms taking place and to show the role played by molecular oxygen.

II. SEASONAL COMPOSITION CHANGES

2.1 Fig. 1 extracted from a paper by Alcaydé et al. (1974) shows the seasonal variations of the ratio of the Atomic oxygen concentration to a weighed sum of the molecular concentrations (nitrogen and oxygen) observed at 200 km over St-Santin (45°N) during the years 1969-1970. These observations are inferred from incoherent scatter measurements of the electron density and the ionic composition through the use of the O^+ ion continuity equation. The ratio varies by a factor of 5 between winter (maximum) and summer. Such a change is considerably larger than the one expected from simple thermal expansion effects on the atmosphere between winter and summer.

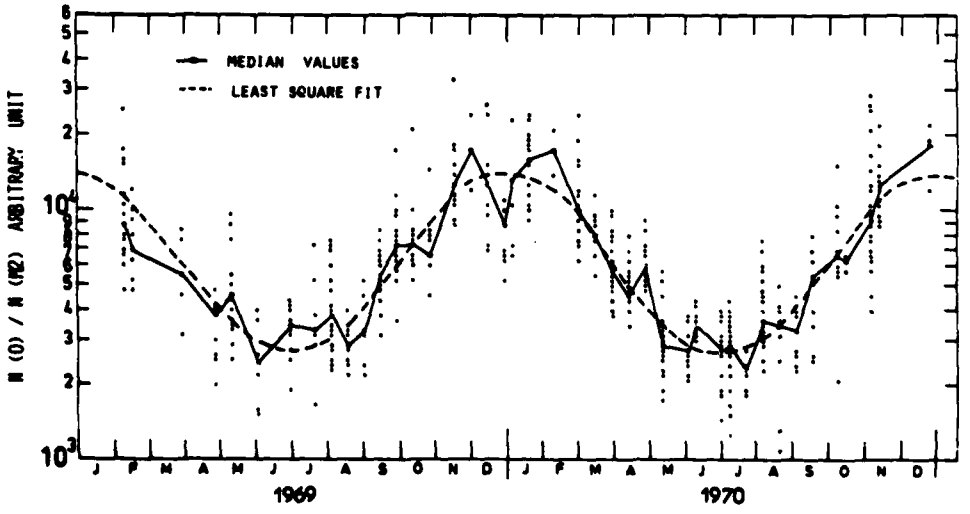


Fig. 1 : Variation of the $N(O)/N(M_2)$ ratio at 200 km in 1969 and 1970. The dashed line corresponds to the least squares fit to an annual component (After Alcaydé et al., 1974).

2.2 Fig. 2 extracted from the same paper shows the seasonal variation of atomic oxygen at 200 km also derived from incoherent scatter data over the same period of time. It is characterized by a large maximum in winter and a minimum in summer. This trend is oppositive to the one expected from the summer heating of the thermosphere in a static atmosphere. With respect to a static atmosphere model this observation calls for a winter to summer modulation of 20 %. Such variations in atomic oxygen are now well established and have been modelled by various authors (e.g. Hedin et al. 1977a, Jacchia 1977, Thuillier et al. 1977, Alcaydé and Bauer 1977, Alcaydé et al. 1978). This evolution permits to explain partly the observations shown on fig. 1. However most of the seasonal change in the ratio of the concentration of atomic oxygen to be weighed sum of the moleculars is still to be explained.

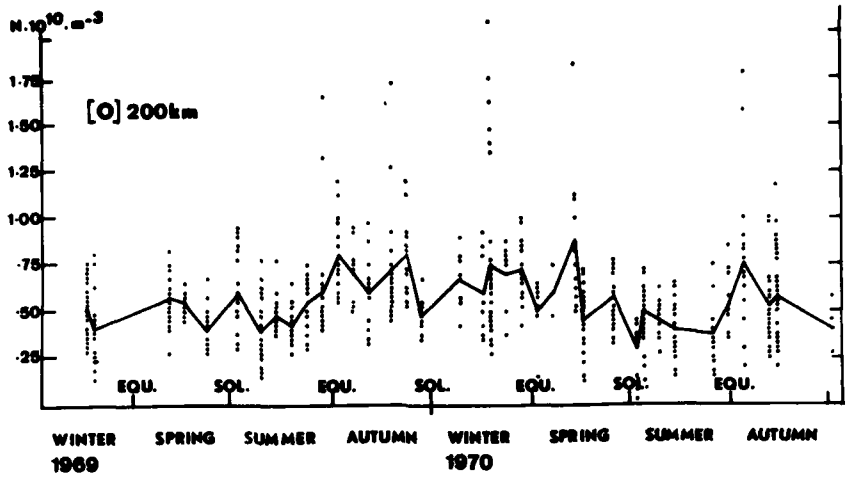


Fig. 2 : Atomic oxygen number density observed at 200 km for the period 1969-1970. The continuous line joins the daily median points (After Alcaydé et al., 1974).

2.3 A quantitative estimate of the respective contributions of molecular nitrogen and of molecular oxygen calls for an increase by a factor of 6 of the molecular oxygen concentration between winter and summer. The considerable change in molecular concentration at 200 km implied by these data corrobors the findings of Roble and Norton (1972), Noxon and Johanson (1972) and Scialom (1974). This finding is however at large from the direct observations of Nier et al. (1976) and Potter et al. (1979) performed on board the Atmospheric explorer satellite around 1975, i.e. during the last solar cycle minimum. This discrepancy will be dealt with in section V.

2.4 It appears therefore that for the observing period, i.e. around the maximum of the solar cycle, atomic oxygen tends to accumulate in winter and molecular oxygen exhibits a considerable increase in summer around 200 km.

III. COMPOSITION CHANGES ASSOCIATED TO MAGNETIC ACTIVITY

3.1 Fig. 3 extracted from a paper by Hedin et al. (1977b) shows the response of several thermospheric species to magnetic activity. The data obtained with the Atmospheric explorer C satellite exhibit departures of the concentrations versus latitude with respect to their quiet time values represented through the MSIS model (Hedin et al. 1977a). A large increase of the heavy gases accompanied by a decrease of the light constituants is observed at high latitudes. Once more the heating the thermosphere associated with the magnetic activity cannot solely explain the behaviour of the different gases. These observations illustrate again the "non static" characteristics of the atmosphere. Such changes have also been observed in many instances by other authors, e.g. Taeusch et al. (1971) , Prölss and von Zahn (1974) , Prölss and Fricke (1975) , Trinks et al. (1975) , Jacchia et al. (1976) .

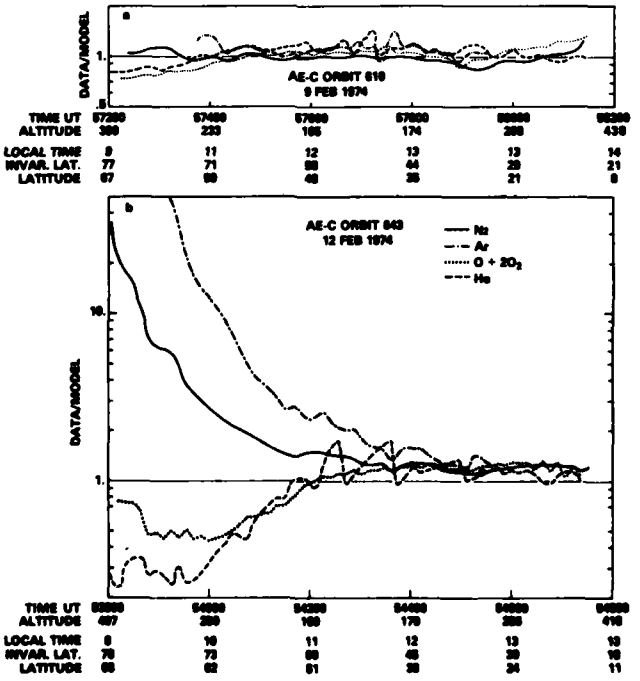


Fig. 3 : a) Data to model density ratios for quiet conditions on February 9 1974, obtained on board the Atmospheric Explorer C satellite. b) Data to model density ratios for disturbed conditions on February 12, 1974 (After Hedin et al., 1977b).

3.2 The response of the ionosphere to these composition changes is shown on fig. 4. The variations of the ratio of the atomic oxygen concentration to the molecular nitrogen concentration together with the electron density and the atomic oxygen ion concentration at 57°N and 200 km of altitude have been plotted versus time during the period of the storm. The variations are strongly correlated illustrating the major role played by the thermospheric composition in controlling the ionosphere. The apparent 24 h modulation corresponds to the periodic change in invariant latitude from orbit to orbit.

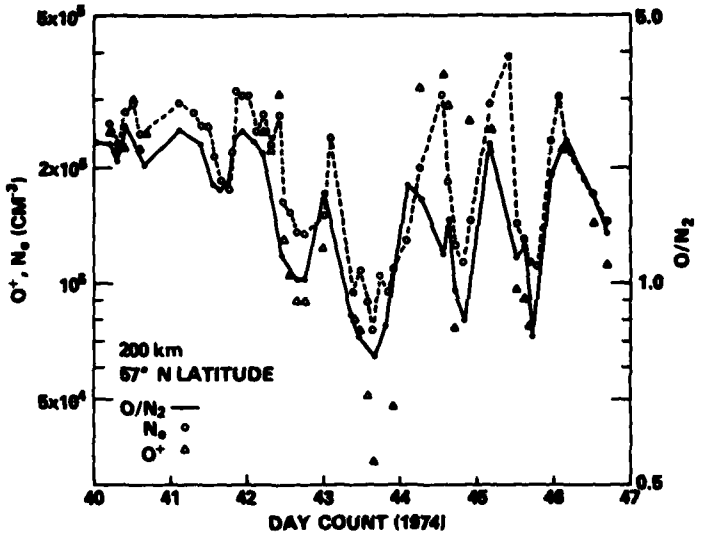


Fig. 4 : Temporal variation of the O/N_2 ratio, O^+ and N_e at 200 km (57°N) from February 9, 1974 to February 16, 1974 observed with the satellite Atmospheric Explorer C (After Hedin et al., 1977b).

IV. PRINCIPAL MECHANISMS CAPABLE OF INDUCING COMPOSITION CHANGES

4.1 Two main processes have been proposed in order to explain the non static behaviour of the thermosphere. The first consists in changes in the turbulent state of the thermosphere inducing changes in the vertical distribution of the various atmospheric species. The second consists in the redistribution of the atmospheric species through atmospheric circulation cells (Johnson and Gottlieb, 1970).

4.2 The data giving access to the turbulent state of the thermosphere is still very scarce and therefore it is difficult to infer seasonal trends. An indirect determination of such seasonal changes is given on fig. 5 extracted from a paper by Alcaydé et al. (1979). Vertical profiles of the eddy diffusion coefficient derived from the thermal structure of lower thermosphere obtained from Incoherent scatter observations above St-Santin (45°N) are presented for different seasons. The weakness of the eddy diffusion coefficient in winter corresponds to a lowering of the turbopause and should drive an enhancement of the atomic oxygen concentration at that time which corresponds to the observed trend. However the knowledge of the turbulent state of the thermosphere calls for more data before any definite conclusion can be drawn.

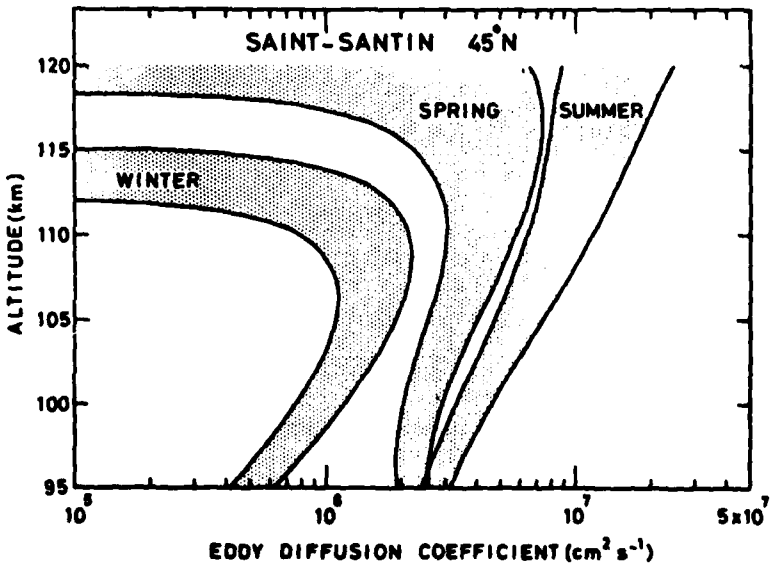


Fig. 5 : Vertical distribution of eddy diffusion coefficients for winter, spring and summer conditions (After Alcaydé et al., 1979).

4.3 The existence circulation meridional cells have been suggested by Johnson and Gottlieb (1970) and Johnson and Gottlieb (1973) in order to explain selective the transport of the various gases. Meridional transports have indeed been observed experimentally (e.g. Amayenc (1973), Reddy (1974)) and this mechanism appears consequently as an efficient

one. Circulation cells tend to establish between cold and warm regions ; they are characterized by an upwelling of the air over the warm regions, a flow at high latitude between the warm and cold regions and a return flow at low altitude. They therefore develop between either the summer and the winter hemispheres or between the high latitude region and the middle and low latitude region during period of large magnetic activity as illustrated on fig. 6 extracted from a paper by Mayr and Volland (1972). The circulation cells tend to accumulate the light constituants (with respect to the major specie) over the cold regions and the heavy ones over the warm regions. This mechanism can account for the observed behaviour of the atomic oxygen concentration.

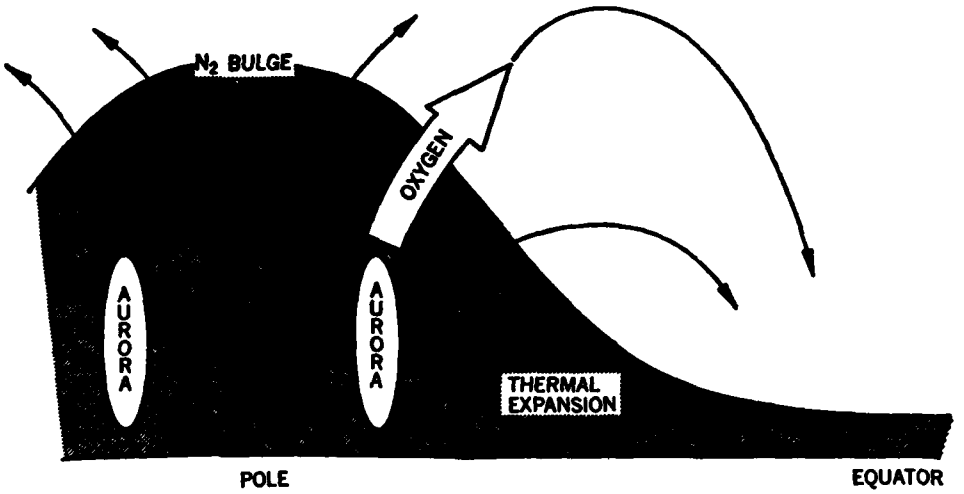


Fig. 6 : Schematic illustration of the circulation cell set up by auroral heating during storm time conditions (After Mayr and Volland, 1972).

V. THE CASE OF MOLECULAR OXYGEN

5.1 Fig. 1 illustrated the changes in composition associated to the winter anomaly during the maximum of the solar cycle. The situation is however significantly different around the minimum of the solar cycle as shown on fig. 7 extracted from a paper by Alcaydé et al. (1980). The amplitude of the variation is now reduced to 2.5 between summer and winter and can essentially be explained by the changes in atomic oxygen concentration and the thermally induced molecular nitrogen concentration changes. The required extra variation in molecular oxygen concentration is now reduced to about 1.5. As a consequence there is no more essential disagreement, for the period considered, with the Atmospheric Explorer in situ data (Nier et al. 1976, Potter et al. 1979).

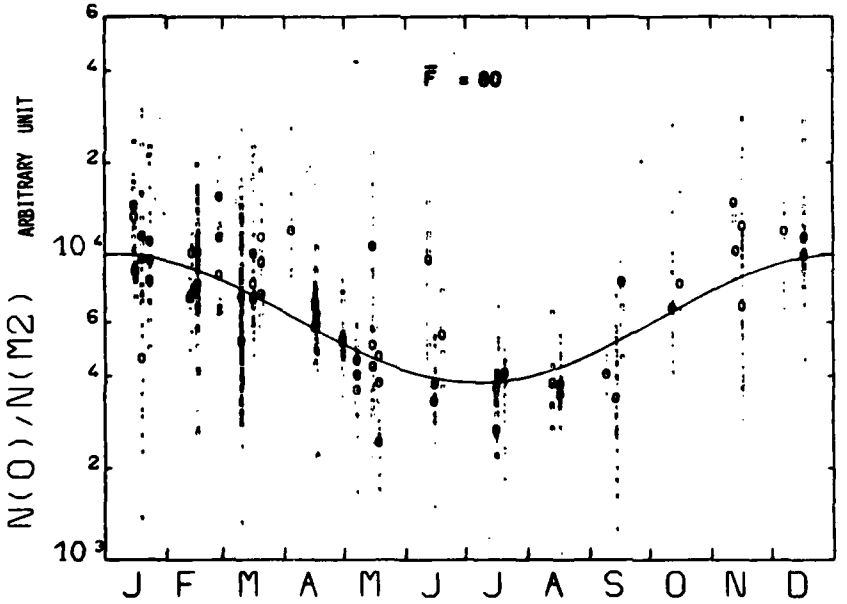


Fig. 7 : Variation of the $N(0)/N(M_2)$ ratio at 200 km for low solar activity ($\bar{F} = 80 \times 10^{-22}$ W m^{-2} Hz^{-1}) (After Alcaydé et al., 1980).

5.2 The different behaviour of molecular oxygen at time of high and low solar activity can be explained by the combination of two mechanisms :

- The direct and indirect photodissociation of O_2 (von Zahn, private communication, 1973 ; Alcaydé et al. 1974 ; Oran and Strobel 1977) leads to a destruction of molecular nitrogen in the upper thermosphere which tends to be compensated by a vertical upward diffusion flux.
- The above mentioned circulation cells established between the summer and the winter hemispheres reinforce the upward diffusion flux of molecular oxygen on the summer side and inhibit the same flux on the winter side creating an unbalance.

In short the molecular oxygen tends to behave as an heavy gas since photodissociation tends to reduce its scale height.

5.3 However the intensity of the circulation is significantly larger around the solar cycle maximum than around the minimum leading to the observed different behaviours.
The result of a simulation of this effect performed by Alcaydé et al. (1980) is given in table 1. The essential trends are indeed fairly well reproduced.

VI. CONCLUSION

It appears clearly that the winter F region anomaly is associated with changes in thermospheric composition. In addition to purely dynamical effects driving an increase in atomic oxygen concentration over the winter hemisphere, it has been shown that a large part of the anomaly seemed to arise from molecular oxygen concentration changes induced by a combination of dynamical and photochemical effects.

TABLE 1 (After Alcaydé et al., 1980)

Solar flux	Mean atmospheric vertical velocity at 200 km in Winter	Mean atmospheric vertical velocity at 200 km in Summer	Summer to Winter ratio of departures from *diffusive equilibrium*
125×10^{-22} $Wm^{-2}Hz^{-1}$	- 1.70 m/s	+ 1.30 m/s	3.2
70×10^{-22} $Wm^{-2}Hz^{-1}$	- .70 m/s	+ .60 m/s	1.4

REFERENCES

Alcaydé D., P. Bauer and J. Fontanari, Long-term variations of thermospheric temperature and composition, J. Geophys. Res., 79, 629-637, 1974.

Alcaydé D. and P. Bauer, Modélisation des concentrations d'oxygène atomique observées par diffusion incohérente, Ann. Géophys., 33, 305-320, 1977.

Alcaydé D., P. Bauer, A.E. Hedin and J.E. Salah, "Compatibility of seasonal variations in midlatitude thermospheric models at solar maximum and low geomagnetic activity", J. Geophys. Res., 83, 1141-1144, 1978.

Alcaydé D., J. Fontanari, G. Kockarts, P. Bauer, R. Bernard, Temperature, molecular nitrogen concentration and turbulence in the lower thermosphere inferred from incoherent scatter data, Ann. Géophys., 35, 41-51, 1979.

Alcaydé D., P. Bauer, J. Fontanari, Solar flux variations in thermospheric molecular oxygen seasonal changes, in preparation, 1980.

Amayenc P., Mouvements thermosphériques de grande échelle ; Etude par diffusion incohérente à moyenne latitude. Thèse de Doctorat d'Etat ès Sciences Physiques, Université Pierre et Marie Curie, Paris VI, 1975.

Hedin A.E., C.A. Reber, G.P. Newton, N.W. Spencer, H.C. Brinton, H.G. Mayr and W.E. Potter, A global thermospheric model based on mass spectrometer and incoherent scatter data, MSIS 2, Composition, J. Geophys. Res., 82, 2148-2156, 1977a.

Hedin A.E., P. Bauer, H.G. Mayr, G.R. Carignan, L.H. Brace, H.C. Brinton, A.D. Parks and D.T. Pelz, Observations of Neutral Composition and related ionospheric variations during a magnetic storm in February 1974, J. Geophys. Res., 82, 3183-3189, 1977b.

Jacchia, L.G., J.W. Slowey, and U. von Zahn, Latitudinal changes of composition in the disturbed thermosphere from Esro 4 measurements, J. Geophys. Res., 81, 36-42, 1976.

Jacchia L.G., "Thermospheric temperature, density, and composition : new models", Smith. Astrophys. Obs. Special Report n° 375, 1977.

Johnson F.S. and B. Gottlieb, Eddy mixing and circulation at ionospheric levels, Planet. Space. Sci., 18, 1707-1718, 1970.

Johnson F.S. and B. Gottlieb, Atomic oxygen transport in the thermosphere, Planet. Space Sci., 21, 1001-1010, 1973.

Mayr H.G. and H. Volland, Magnetic storm effect in the neutral composition, Planet. Space Sci., 20, 379-394, 1972.

Nier A.O., W.E. Potter and D.C. Kayser, Atomic and molecular oxygen densities in the lower thermosphere, J. Geophys. Res., 81, 17-24, 1976.

Noxon, J.F. and A.E. Johanson, Changes in thermospheric molecular oxygen abundance inferred from twilight 6300 Å airglow, Planet. Space Sci., 20, 2125, 1972.

Oran E. and D. Strobel, Photochemical deviations of thermospheric 0 and 0_2 densities from diffusive equilibrium, J. Geophys. Res., 82, 696-698, 1977.

Potter W.E., D.C. Kayser and A.O. Nier, J. Geophys. Res., 84, 10-16, 1979.

Prölss G.W. and U. von Zahn, Esro 4 gas analyzer results, 2, Direct measurements of changes in the neutral composition during an ionospheric storm, J. Geophys Res., 79, 2535-2539, 1974.

Prölss G.W. and K.H. Fricke, Neutral composition changes during a period of increasing magnetic activity, Planet. Space Sci., 24, 61-67, 1975.

Reddy C.A., Evidence of a meridional circulation cell in the lower thermosphere during a magnetic storm, J. Atm. Terr. Phys., 36, 1561-1564, 1974.

Rishbeth H. and C.S.G.K. Setty, The F layer at sunrise, J. Atmos. Terr. Phys., 20, 263, 1961.

Roble R.G. and R.B. Norton, Thermospheric molecular oxygen from solar extreme ultraviolet occultation measurements, J. Geophys. Res., 77, 3524, 1972.

Scialom G., Neutral composition in the lower thermosphere, Radio Sci., 9, 253-262, 1974.

Taeusch D.R., G.R. Garignan and C.A. Reber, Neutral composition variation above 400 kilometers during a magnetic storm, J. Geophys. Res., 76, 8318-8325, 1971.

Thuillier G., J.L. Falin and C. Wachtel, Experimental global model of the exospheric temperature based on measurements from the Fabry-Perot interferometer on board Ogo E satellite, J. Atmos. Terr. Phys., 39, 399, 1977.

Trinks H., K.H. Fricke, U. Laux, G.W. Prölss, and U. von Zahn, Esro 4 gas analyzer results, 3, Spatial and temporal structure of the mid-latitude atmosphere during a geomagnetic storm, J. Geophys. Res., 80, 4571-4575, 1975.

RELATIONS BETWEEN SOLAR FLUX AND E-REGION PARAMETERS

Marcel NICOLET and Lucien BOSSY
Institut Royal Météorologique
3, Avenue Circulaire
1180 Bruxelles

ABSTRACT

A detailed analysis of the behaviour of the critical frequency of the E-layer with the variation of the solar flux at 10.7 cm leads to the possibility of an explanation of the simultaneous ionization by ultraviolet radiation (Lyman-β) and soft X-rays (30 to 100 A). The ratio of the X-radiation to the UV-radiation varying strongly with solar activity changes the ratio of the principal ions concentrations.

1. INTRODUCTION

The investigation of the structure of the E-layer, and the way in which it varies with solar activity, differs from that of the layers below and above it : the D- and F-layers. The E-layer cannot be studied with the aid of satellites, and it lies near the turbopause where there are variations in the concentrations of molecular and atomic oxygen. Moreover, photoionization in the E-layer is due not only to the absorption by O_2 of UV-radiation at wavelengths greater than 800 A, but also of X-rays in the whole range between 10 A and 100 A.

Finally since, for many years, ionospheric soundings have provided data relating to the E-layer that are not available for the D-layer, it is interesting to examine the variations of foE, the critical frequency. With this in mind, the ionospheric data obtained at the station at Dourbes (50.1°N, 04.6°E) of the Royal Meteorological Institute of Belgium have been compared with the flux of solar UV-radiation (provided by H. Hinteregger) and of X-radiation (provided by R. Kreplin).

2. IONOSPHERIC AND SOLAR DATA

The variations in the critical frequency (f) of the E-layer, and hence of the maximum electron concentration (n_e), are controlled mainly by the solar zenith angle (χ), but also by the level of solar activity which influences the degree of ionization at heights near 100 km (Minnis, 1964 a,b). For a constant level of solar activity, the annual variation of f at local noon for the station at Dourbes can be written as follows, using the nomenclature of Minnis (1964b) :

$$f_{12} = f_a \, (\cos \chi_{12})^{0.263} \qquad (1)$$

while the diurnal variation can be written :

$$f_t = f_d \, (\cos \chi_t)^{0.33} \qquad (2)$$

In these equations, f_a and f_d are regarded as constants, for a station at a given latitude, while the suffixes t and 12 identify the values of f at time t and at local noon, respectively. The application of these equations to 1500 observations of foE at Dourbes, for the period early 1975 to June 1980, makes it possible to derive the critical frequency for an overhead sun F_o^E. Figure 1 shows the relation between F_o^E(MHz) and the solar radio flux Φ at 10.7 cm, observed by the National Research Council of Canada; the relation can be expressed by the following equation with a coefficient of correlation of 0.94 :

$$F_o^E = 3.22 + 0.0063 \, (\Phi - 65) \qquad (3)$$

As can be seen, the value of F_o^E for a completely quiet sun ($\Phi = 65$) is 3.2 MHz, with a fluctuation of about ± 5%. In other words, a value of 3 MHz can be considered as the lower limit; this corresponds to a maximum electron concentration of 1.1×10^5 cm^{-3}, while a frequency of 2.94 MHz corresponds to 10^5 cm^{-3}.

After the first studies made by Appleton and Naismith (1939), who compared monthly mean values of critical frequency with sunspot numbers for the period 1931-1938; Nicolet (1945) made a similar study for the years 1935-1940, but also used calcium plages as the solar index for the period 1939-1940 for the three stations : Washington D.C., Huancayo, Peru and Watheroo, Australia.

Since the initial comparisons made by Denisse and Kundu (1957), using the 10.7 cm radio flux and the Freiburg data, this index of solar activity has continued to be widely used. A full analysis of this question has been given by Minnis (1964a).

A comparison (Figure 2) of the values of Φ with 600 observations of the intensity of Lyman-β, made between June 1977 and June 1980, leads to the following relation, with a correlation coefficient of 0.9 :

$$q_\infty \, (\text{Ly-}\beta) = 5.8 \times 10^9 + 0.036 \, (\Phi - 65) \times 10^9 \qquad (4)$$

where q_∞ (Ly-β) represents the number of photons cm^{-2} sec^{-1} at the top of the atmosphere. This equation is not applicable to a completely quiet sun, for it gives a value of 6×10^9 photons cm^{-2} sec^{-1}, whereas Hinteregger (1980) has obtained values of 3×10^9 and 4.4×10^9 photons cm^{-2} sec^{-1} for 1974 and 1976 respectively. As there is a fluctuation of ± 30%, the solar flux is not known to better than ± 30% and it seems best to accept the value of the Lyman-β flux as being $(4.5 \pm 1.5) \times 10^9$ photons cm^{-2} sec^{-1} at the top of the earth's atmosphere.

A similar comparison has been made with data for X-radiation in the range 44 A - 60 A for the years 1976-1978, as shown in Figure 3. The relation can be expressed as follows with a correlation coefficient of 0.9 :

$$X_{40-60} = 3.8 \times 10^{-3} + 1.055\ (\Phi - 65) \times 10^{-3} \quad (5)$$

where X_{40-60} is the solar flux in erg cm^{-2} sec^{-1} at the top of the atmosphere. If X_{40-60} is expressed as the number of photons, equation 5 becomes :

$$X_{40-60} = 1 \times 10^{7} + 2.78\ (\Phi - 65) \times 10^{6} \quad (6)$$

Equations 3 to 6 ought to make it possible to study the ionization in the E-region as a function of the variations in solar activity.

3. ABSORPTION AND IONIZATION PROCESSES

The principal constituents of the lower thermosphere are N_2, O_2 and O and, when their absorption spectra are examined, it is clear that X-radiation at wavelengths less than 100 A can play a role in the ionization process. The absorption cross sections shown in Figure 4 show that unit optical depth corresponding to 10^{19} to 10^{18} molecules cm^{-2}, occurs in fact in the E-region between 95 and 110 km.

Moreover, the photoionization of O_2 which begins at 1027.8 A (Samson and Cairns, 1964) depends on the Ly-β line at 1025.7 A and on the CIII line at 977 A both of which are strong in the solar spectrum. At 1025.7 A the absorption cross section $\sigma_A = 1.5 \times 10^{-18}$ cm^2 and the ionization cross section $\sigma_I = 1 \times 10^{-18}$ cm^2; the corresponding values for CIII are $\sigma_A = 4 \times 10^{-18}$ cm^2 and $\sigma_I = 2.5 \times 10^{-18}$ cm^2 respectively.

For the spectral range 910 A - 800 A, with its characteristic lines and continuum, the very approximate values are $\sigma_A(O_2) = 7.5 \times 10^{-18}$ cm^2 and $\sigma_I = 5 \times 10^{-18}$ cm^2 and these indicate that such an absorption occurs only in the upper part of the E-region.

It is necessary to take also account of the supplementary absorption due to the N_2 bands in this spectral range. But the detailed structure of these bands is so pronounced that it is not possible to base calculations on average values; with high resolution it is found that the absorption cross section varies between 10^{-19} and 10^{-17} cm^2. Although this spectral range can be neglected when dealing with the ionization at the peak of the E-region, it must be retained for the part of the E-region where there is photoionization of the atomic oxygen,which has its first ionization continuum in the Lyman continuum at 910 A. The ionization cross section of OI is $\sigma_I = 4 \times 10^{-18}$ cm^2 (Taylor and Burke, 1976; Kohl et al., 1978). However, it should be remembered that its effect is only partial near the maximum of the E-region where Lyman-β and CIII play the leading role alongside X-radiation.

Since X-rays of wavelengths less than 100 A correspond to energies between 125 eV and 450 eV, the photoelectrons resulting from primary photoionization have energies sufficient to ionize certain atmospheric constituents. The use of the conventional value of 34 eV for the production of an ion-electron pair implies a high rate of production of electrons.

Considering first the branching ratios for the different electronic states of N_2^+ (Hamnet et al., 1976; Plummer et al., 1977) and of O_2^+ (Samson et al., 1977), it is possible to determine the kinetic energies of the electrons released during photoionization. However, it should be remembered that there is some uncertainty because measurements of most of the branching ratios have been made only up to 303 A. On the other hand, the changes would not be very great near 100 A, at least for N_2 (cf. Plummer et al., 1977). Hence we can adopt a mean value for O_2 and N_2, and this leads to the following numbers of secondary electrons for each photoionization : 4.5 at 70 A, 6.5 at 50 A and 9.5 at 35 A.

It is important to remember also that the absorption at these short wavelengths leads not only to direct ionization of molecules, but also to dissociative ionization of N_2 and O_2 giving $N^+ + N$ and $O^+ + O$, respectively. No experimental studies have been carried out at short wavelengths but, considering that the values obtained by Wight et al. (1976), between 55 eV and 60 eV are nearly constant, we shall adopt, for the ratios N^+/N_2^+ and O^+/O_2^+, identical values with the branching ratios for N^+ and O^+, namely 1/3 and hence 2/3 for N_2^+ and O_2^+.

It must be added that, for atomic oxygen, the photoionization transitions beginning at 435 A and 310 A are followed by radiation emitted at 833 A and at 581 A - 534 A respectively (Dalgarno et al., 1963). The radiation at 833 A is absorbed either by O_2 with an effective cross section of $\sigma_A = 2.4 \times 10^{-17}$ cm^2 and $\sigma_I = 6.5 \times 10^{-18}$ cm^2 or by O with an effective cross section of $\sigma_O = 4 \times 10^{-18}$ cm^2. As for the other two radiations at 534 A and 581 A, they are absorbed by N_2 and O_2 with an ionization cross section of the order of $\sigma_I = 2.4 \times 10^{-17}$ cm^2, and by O with a cross section of $\sigma_I = 1.2 \times 10^{-18}$ cm^2. In the light of the theoretical results of Henry (1967), it can be estimated that this second ionization occurs in 30% of cases with respect to the global photoionization.

Finally we must bear in mind the proportion of dissociative ionization resulting from electronic collisions. Following Rapp et al. (1966) we shall adopt the maximum branching ratios of 0.3 for O^+ and of 0.2 for N^+.

To summarize, during photoionization at wavelengths below 100 A, the ionization of a molecule, XY^+, leads to the appearance of an atomic ion $X^+ + Y$ in a certain proportion of cases and to the generation of a photoelectron with a kinetic energy which is sufficient to ionize several times the other atmospheric constituents in proportion to their abundance. The result is the appearance of both atomic and molecular ions.

Some experimental and theoretical determinations of absorption cross sections have been made in the range

100 A to 300 A by Messner (1933), Denne (1970), Cole and Dexter (1978) and Mehlman et al. (1978) and these can be used to calculate the various cross sections represented in Figure 5. The calculations are based on the following formulae which seem to represent best the results for wavelengths less than 100 A :

$$\sigma_I(N_2) = 1.25 \times 10^{-18} \ (\lambda/100)^{2.5} \ cm^2 \qquad \text{for} \qquad 31 \leq \lambda \leq 100 \ A$$

$$\sigma_I(O_2) = 2.25 \times 10^{-18} \ (\lambda/100)^{2.5} \ cm^2 \qquad \text{for} \qquad 24 \leq \lambda \leq 100 \ A$$

and

$$\sigma_I(O) = 1.1 \times 10^{-18} \ (\lambda/100)^{2.5} \ cm^2 \qquad \text{for} \qquad 24 \leq \lambda \leq 100 \ A.$$

Since the accuracy of the data given by Cole and Dexter (1978) is ± 10%, the values of the cross sections given by the above formulae are sufficiently accurate for a study of the E-region.

4. PRODUCTION OF IONS

It is necessary, first of all, to select the basic data on solar fluxes in the different spectral ranges. These of Hinteregger (1980) are approximate for the ionization of O_2 by Lyman-β and CIII; they refer to July 1976, when the sun was quiet (ϕ = 70 units), or to 23 April 1974, when the solar flux was about 10% less, even though ϕ was about 75 units. Thus we have adopted :

$$q_\infty(\text{Lyman-}\beta) = 3.5 \times 10^9 \text{ photons cm}^{-2} \text{ sec}^{-1} \tag{7}$$

$$q_\infty(\text{CIII - 977 A}) = 4.4 \times 10^9 \text{ photons cm}^{-2} \text{ sec}^{-1} \tag{8}$$

$$I_{O_2}(\text{Lyman-}\beta) = 3.5 \times 10^9 \ e^{-1.5 \times 10^{-18} N(O_2)} \ sec^{-1} \tag{9}$$

and

$$I_{O_2}(\text{CIII - 977 A}) = 1.1 \times 10^{-8} \ e^{-4 \times 10^{-18} N(O_2)} \ sec^{-1} \tag{10}$$

$N(O_2)$ is the number of absorbing oxygen molecules. For the photoionization rates due to the Lyman continuum at wavelengths less than 910 A, we would have, for quiet solar conditions :

$$I_{O_2} = 1.75 \times 10^{-8} \ e^{-4 \times 10^{-18} N(O_2)} e^{- \times N(N_2)} \ sec^{-1} \tag{11}$$

$$I_O = 1.4 \times 10^{-8} \ e^{-4 \times 10^{-18} N(O_2)} e^{- \times N(N_2)} \ sec^{-1} \tag{12}$$

We shall not consider these ionization processes in our discussion because their role is not of primary importance in the behaviour of the peak of the E-layer.

It is not possible to determine either the number of photons in all the solar emission lines between 100 A and 30 A, or their variations with solar activity. For this reason it is necessary to deal with several ranges of the spectrum, determined partly in the light of an average absorption, and partly by the availability of observational data which make it possible to follow the trend of solar activity.

In view of the various data that have been published, and in particular those of Kreplin and Gregory (1966) and Manson (1967, 1968), we shall adopt three spectral ranges : 70 ± 10 A, 50 ± 10 A and 35 ± 5 A and the following minimum values respectively : 0.75×10^{-2} erg cm^{-2} sec^{-1} corresponding to 2.5×10^7 photons cm^{-2} sec^{-1}; 1.0×10^{-2} erg cm^{-2} sec^{-1} to 2.5×10^7 photons cm^{-2} sec^{-1} and 0.425×10^{-2} erg cm^{-2} sec^{-1} to 0.75×10^7 photons cm^{-2} sec^{-1}. The results of the calculations are given in Figure 6 for overhead sun; this shows the distribution of the different ions, with their respective contributions to the total ionization. For a recombination coefficient of 3×10^{-7} cm^3 sec^{-1}, the electron concentration at the peak of the layer is about 1.1×10^5 cm^{-3} corresponding to a critical frequency of 3 MHz.

The results of a typical calculation are illustrated in Figures 4, 7, 8 and 9 for ϕ = 100 units. Figure 7 shows the production of N_2^+ ions for the three spectral ranges mentioned and it can be seen that the 50 A region is the most important at the maximum of the E-layer. Figure 8 shows the respective roles of the ionization of O, O_2 and N_2 in the production of N_2^+. Figure 9 illustrates, for the global X-radiation, the roles in the total ion production of each of the ions N_2^+, N^+, O^+ and O_2^+; it is clear that N_2^+ is the most important of the group.

Finally, the direct production of O_2^+ by UV-radiation is presented in Figure 10 for solar zenith angles 0° and 60°. Comparison with the results in Figure 6 shows at once that the action of UV-radiation in the ionization of O_2 is more important than the action on all constituents of X-radiation when the sun is very quiet. Figure 11 compares the contributions of UV- and X-radiation in the E-region for conditions of very quiet and disturbed sun. Although at minimum solar activity, the ionization of O_2^+ is greater than the general ionization by X-radiation, at maximum solar activity the two sources of ionization are practically of the same order of magnitude. With an increase in X-radiation by a factor of 9, and in Lyman-β by a factor of 3, the electron concentration at the peak increases from 1.1×10^5 cm^{-3} to 2.3×10^5 cm^{-3}, which corresponds to an increase of the critical frequency from 3 MHz to 4.3 MHz. These results are in conformity with the observed variations in the critical frequency of the E-layer during the solar activity cycle.

SUMMARY AND CONCLUSIONS

The combined study of the variations in the critical frequency of the E-layer, in the UV-radiation ionizing the oxygen molecule, and in soft X-radiation leads to the conclusion that the relative importance of the two spectral regions varies with the solar cycle.

It will be necessary to compare the results of the present study with those obtained by other authors; for example : Keneshea et al. (1970), Bomke et al. (1970), Ivanov-Kholodnyi and Nusimov (1976) and Donnelly (1976). However, it is important to draw attention to the variations in the vertical distribution and in the concentration of O_2 during the solar cycle (Kayser, 1980), and to the lack of agreement between the values of the atomic oxygen concentration observed by Howlett et al. (1980) and Dickinson et al. (1980).

In addition, it will be necessary to determine with precision the various branching ratios leading to O_2^+ and NO^+, the ions which take part in the final dissociative recombination in the E-layer. Besides the ionization of O_2 by UV- and X-radiation, it is appropriate to add the contributions made by several other ionic reactions :

$$N_2^+ + O_2 \rightarrow N_2 + O_2^+$$
$$O^+ + O_2 \rightarrow O + O_2^+$$
$$N^+ + O_2 \rightarrow N + O_2^+$$

After determining the production of NO^+ by the reactions :

$$N_2^+ + O \rightarrow N + NO^+$$
$$O^+ + N_2 \rightarrow N + NO^+$$
$$N^+ + O_2 \rightarrow O + NO^+$$

and perhaps also

$$O_2^+ + NO \rightarrow O_2 + NO^+$$

it will be possible to determine the average electronic recombination coefficient :

$$\alpha_E = \frac{\alpha_{NO^+}\, n^+(NO) + \alpha_{O_2^+}\, n^+(O_2)}{n^+(NO) + n^+(O_2)}$$

in which α_{NO^+} and $\alpha_{O_2^+}$ represent respectively the approximate values 4×10^{-7} and 2×10^{-7} $cm^3\ sec^{-1}$. The resulting value for α_E is $(3 \pm 1) \times 10^{-7}$ $cm^3\ sec^{-1}$ if it is assumed that the action of the dissociative recombination of N_2^+ is less than that of the ionic reactions.

REFERENCES

Appleton, E.V. and R. Naismith. The variation of the ultra-violet radiation during the sunspot cycle, Phil. Mag., 27, 144, 1939.

Bomke, H.A., H.A. Blake, A.K. Harris and D.J. Sheppard. An eclipse study of soft-X-ray distribution over the sun and of the relative contributions of X-rays and UV to the E-layer formation, J. Geophys. Res., 75, 6980, 1970.

Cole, B.E. and R.N. Dexter. Photoabsorption and photoionization measurements on some atmospheric gases in the wavelength region 50 - 340 A, J. Phys. B., Atm. Molec. Phys., 11, 1011, 1978.

Dalgarno, A., M.B. McElroy and R.J. Moffett. Electron temperatures in the ionosphere, Planet.Space Sci., 11, 463, 1963.

Denisse, J.F. and M.R. Kundu. Relation entre l'ionisation de la couche E de l'ionosphère et le rayonnement solaire radioélectrique, C.R. Acad. Sc. Paris, 244, 45, 1957.

Denne, D.R., Measurements of the ultrasoft X-ray absorption of Ar, Ne, N_2, O_2, CH_4, He and H_2, J. Phys. D, Appl. Phys., 3, 1392, 1970.

Dickinson, P.H.G., W.C. Bain, L. Thomas, E.R. Williams, D.B. Jenkins and N.D. Twiddy. The determination of the atomic oxygen concentration and associated parameters in the lower ionosphere, Proc.R.Soc.London, A 369, 379, 1980.

Donnelly, R.F., Empirical models of solar flare X-ray and EUV-emission for use in studying their E- and F-region effects, J. Geophys. Res., 81, 4745, 1976.

Hamnett, A., W. Stoll and C.E. Brion. Photoelectron branching rations and partial ionization cross sections for CO and N_2 in the energy range 18 - 50 eV, J. Elect. Spect., 8, 367, 1976.

Henry, R.J., Photoionization cross sections for atomic oxygen, Planet. Space Sci., 15, 1747, 1967.

Hinteregger, H., Representation of solar EUV-fluxes for aeronomical applications, Space Research, to be published , 1980.

Nowlet, L.C., K.D. Baker, L.R. Megill, A.W. Shaw and W.R. Pendleton. Measurement of a structured profile of atomic oxygen in the mesosphere and lower thermosphere, J. Geophys. Res., 85, 1291, 1980.

Ivanov-Kholodnyi, G.S. and A.A. Nusinov. Relative contribution of ultraviolet and X-radiations to E-region ionization. Geom. and Aeron., 16, 43, 1976.

Kayser, D.C., Solar flux variation of the thermospheric oxygen density, J. Geophys. Res., 85, 695, 1980.

Keneshea, T.J., R.S. Narcisi and W. Swider. Diurnal model of E-region, J. Geophys. Res., 75, 845, 1970.

Kohl, J.L., G.P. Lagyatis, N.P. Palenius and W.M. Parkinson. Absolute cross sections for photoionization of atomic oxygen, Phys. Rev., A, 18, 571, 1978.

Kreplin, R.W. and B.N. Gregory. Solar monitoring during the IQSY, Space Research 6, 1011, 1966.

Manson, J.E. The spectrum of the quiet sun between 30 A and 128 A for November 1965, Astrophys. J., 147, 703, 1967.

Manson, J.E. Instrumental recalibration and refinement of solar ultrasoft X-ray intensities, Astrophys. J., 153, L 191, 1968.

Mahlman, G., D.L. Ederer and E.B. Saloman. The photoabsorption cross section of O_2 from 55 - 350 A, J. Chem. Phys., 68, 1862, 1978.

Messner, R.H. Der Einfluss der chemischen Bindung auf den Absorptions-koeffizienten leichter Elemente in Gebiete ultraweicher Röntgenstrahlen, Z. Physik, 85, 727, 1933.

Minnis, C.M. Annual variation in E-layer ionization, Nature, 202, 170, 1964.

Minnis, C.M. Ionospheric indices, in Adv. in Radio Res., Vol. 2, pp. 1 - 36, ed. J.A. Saxton, Acad. Press, New York, 1964.

Nicolet, M. Contribution à l'étude de la structure de l'ionosphère, Inst. Roy. Météor. Belgium, Mém. n°19, pp. 1 - 163, 1945.

Plummer, E.W., T. Gustafsson, W. Gudat and D.E. Eastman. Partial photoionization cross sections of N_2 and CO using synchrotron radiation, Phys. Rev., 15, 2339, 1977.

Rapp, D., D.D. Briglia and P. Englander-Golden. Cross sections for dissociative ionization of molecules by electron impact, J. Chem. Phys., 42, 4081, 1965.

Samson, J.A.R. and R.B. Cairns. Absorption and photoionization cross sections of O_2 and N_2 at intense solar emission lines, J. Geophys. Res., 69, 4583, 1964.

Samson, J.A.R. and J.L. Gardner and G.N. Haddas. Total and partial photoionization cross sections of N_2 from 100 to 800 A, J. Elect. Spect., 12, 281, 1977.

Taylor, K.T. and P.G. Burke. Photoionization of ground-state carbon and oxygen atoms, J. Phys. B., Atm. Molec. Phys., 9, L 353, 1976.

Wight, G.R., M.J. Van der Wiel and C.E. Brion. Dipole excitation, ionization and fragmentation of N_2 and CO in the 10 - 60 eV region, J. Phys. B., Atm. Molec. Phys., 9, 675, 1976.

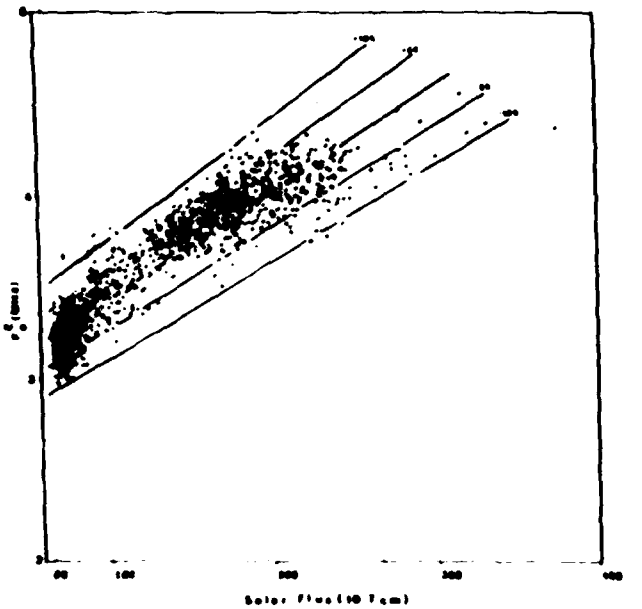


Fig. 1.- Relationship between the critical frequency for an overhead sun as observed in Dourbes and the solar radio flux at 10.7 cm.

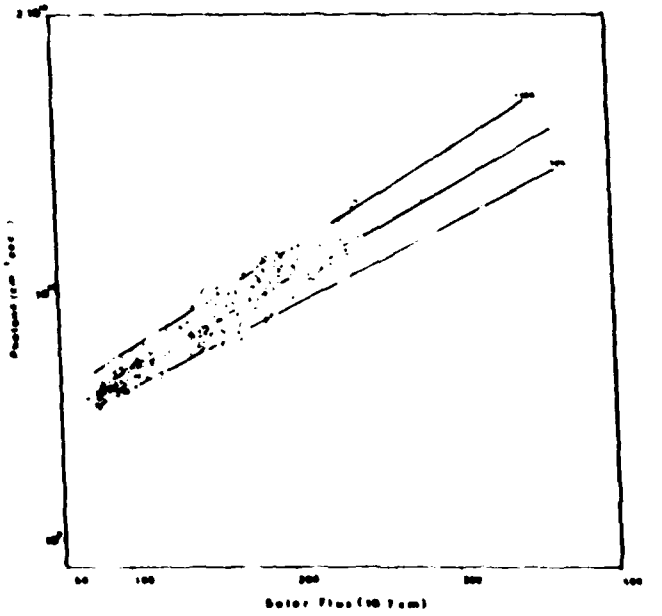


Fig. 2.- Relationship between the solar intensities of Lyman- β and 10.7 cm.

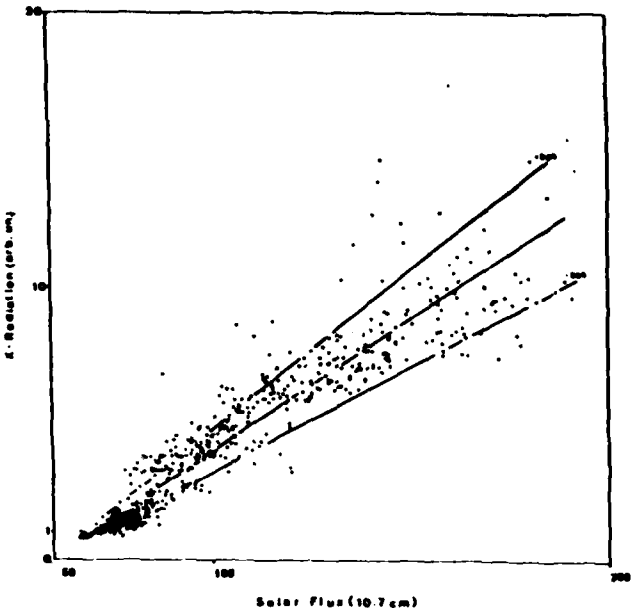


Fig. 3.- Relationship between the solar intensities of X-radiation at 44-60 A and 10.7 cm.

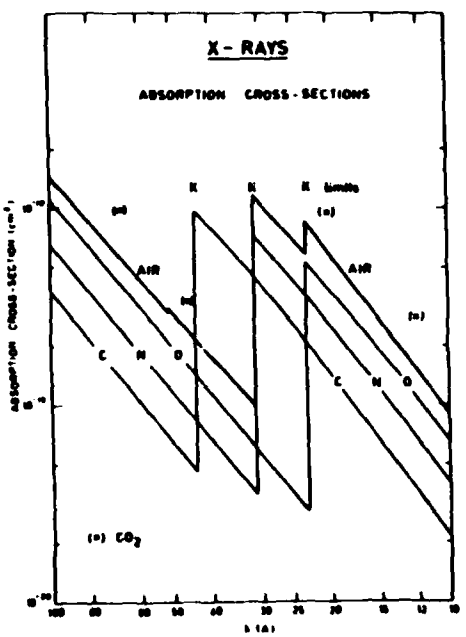


Fig. 4.- Absorption cross section of O and N between 100 and 10 A.

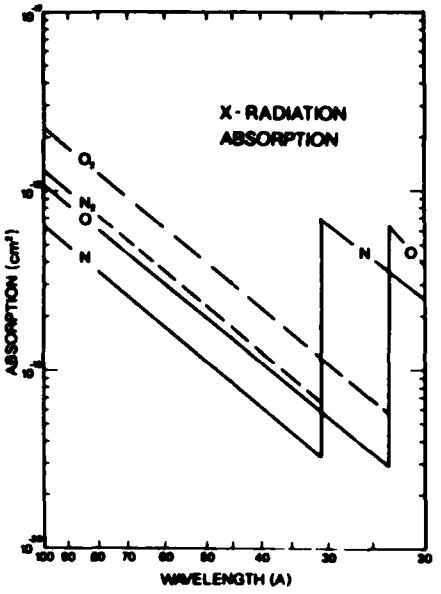


Fig.5.- Absorption cross section between 100 A and 20 A of atomic and molecular oxygen and nitrogen.

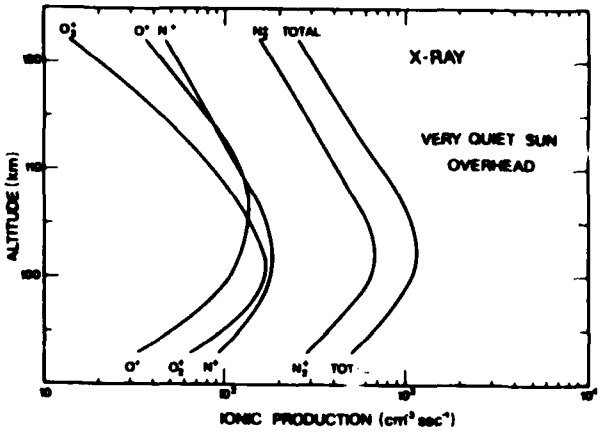


Fig.6.- Ionic production (N_2^+, N^+, O^+ and O_2^+) by X-radiation in the E-layer for an overhead sun and very quiet solar conditions.

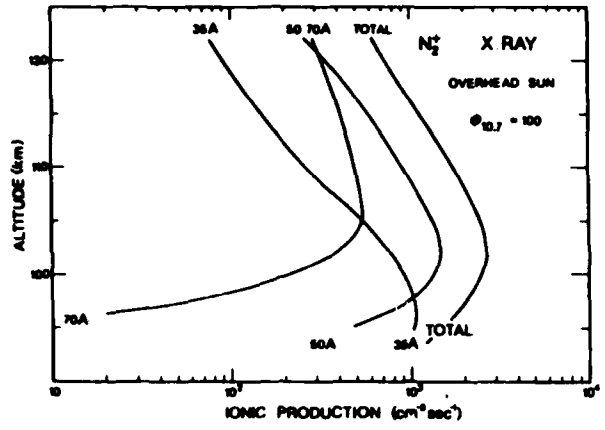


Fig. 7.- N_2^+ production in the 70, 50 and 35 A spectral ranges for an overhead sun and $\Phi_{10.7}$ = 100.

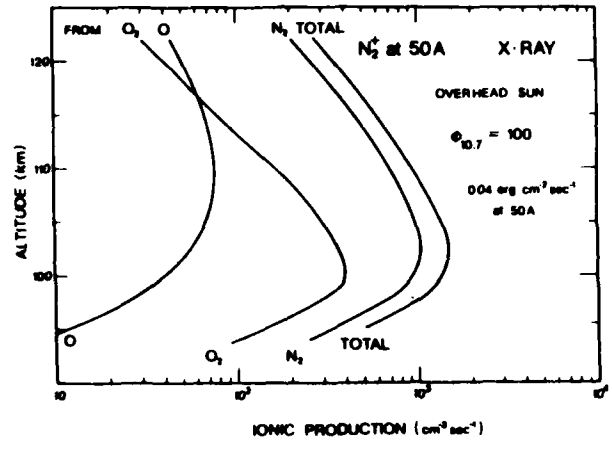


Fig.8.- N_2^+ production due to the ionization of O_2, O and N_2 in the 50 A spectral range for an energy of 0.04 erg cm^{-2} sec^{-1} at the top of the earth's atmosphere. Overhead sun conditions.

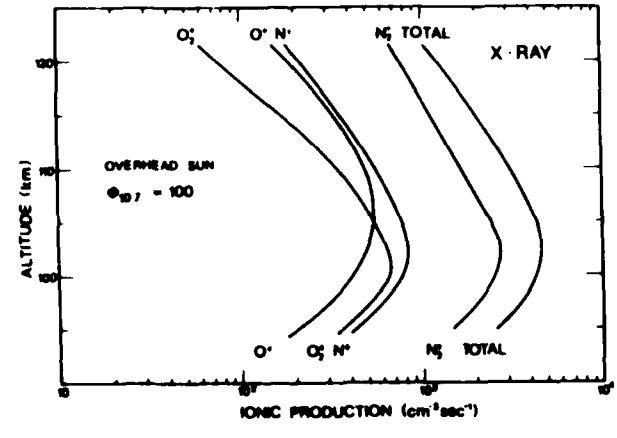


Fig. 9.- Ionic production by the E-layer by X-radiation for solar flux condition represented by $\Phi_{10.7}$ = 100. N_2^+, N^+, O^+ and O_2^+ production for overhead sun conditions.

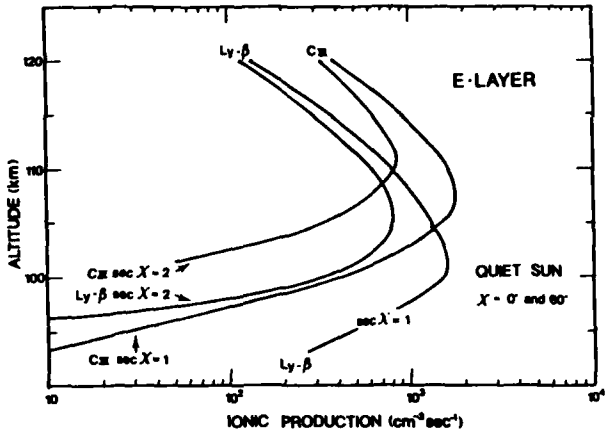


Fig. 10.- O_2^+ production by ultraviolet radiation by Lyman - β and CIII lines. Sun overhead and at 60° ; relatively quiet solar conditions.

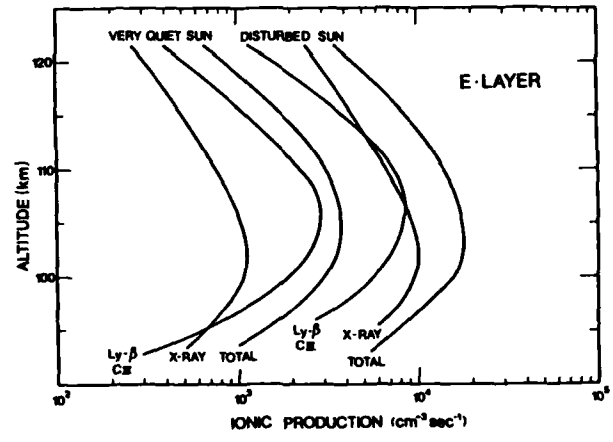


Fig.11.- Calculated ionic production for overhead sun. Comparison between very quiet and disturbed sun conditions.

ENERGETIC PARTICLE PRECIPITATION AT MIDDLE AND LOW LATITUDES

*L. G. Smith and H. D. Voss**

Aeronomy Laboratory, Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, U.S.A.

SUMMARY

Rocket observations of the nighttime electron density profile in the upper E region (120 to 200 km) over Wallops Island, Virginia, have demonstrated the importance of atmospheric tides in the formation of the intermediate layer near 150 km. These observations have also shown that, except perhaps on geomagnetically quiet nights, energetic particles are the dominant ionization source in the layer. Using rocket, satellite and ground-based observations it has been possible to study the global morphology of particle precipitation. It is found that, just as there are northern and southern auroral zones, there are also mid-latitude, low-latitude and equatorial zones of precipitation.

1. INTRODUCTION

The electron density in the nighttime E region at midlatitudes is generally less than 10^4 cm^{-3}. Because of the low density the lifetime of the ions, even those that are non-metallic, is long enough that vertical redistribution under the influence of the neutral winds is an important factor in determining the electron density profile [*Geller et al.*, 1975]. In rocket observations at Wallops Island, Virginia (38°N, 75°W) it produces, near midnight, the intermediate layer [*Smith*, 1970]. As shown in Figure 1 the layer is most clearly evident for magnetically quiet conditions.

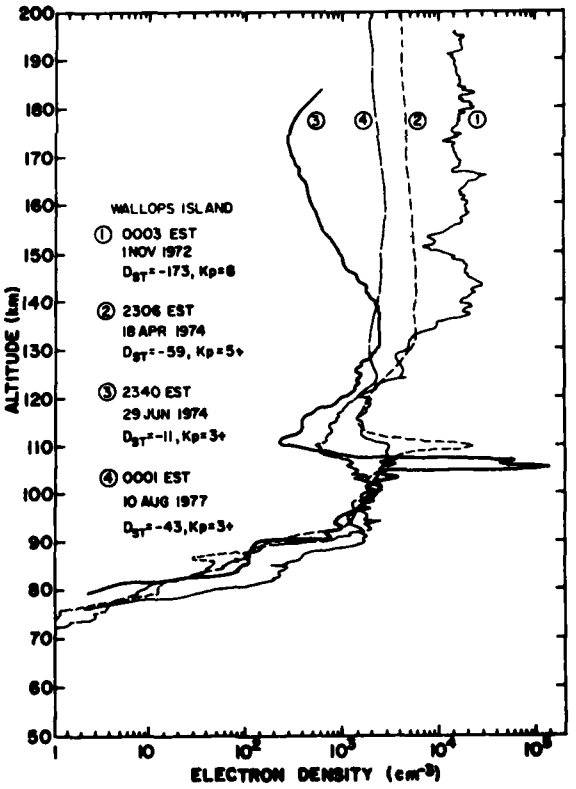


Fig. 1. Nighttime electron density profiles from rockets launched at Wallops Island [*Voss and Smith*, 1979].

The lifetime of the ions is not long enough, however, to maintain the nighttime E region without an additional source of ionization. Solar Lyman α and Lyman β scattered in the geocorona are ionizing sources in the lower E region [*Strobel et al.*, 1974] but are not sufficient for the upper E region. As may also be seen in Figure 1 the electron density increases with increasing magnetic activity. *Smith et al.*, [1974] have shown that the ionization rate obtained from the electron density profile shows a strong dependence on magnetic activity. This is illustrated in Figure 2, which includes additional data from recent rocket flights [*Voss and Smith*, 1979]. A new magnetic index (I), based on both the Kp and Dst indices, is de-

*Present address: Lockheed Palo Alto Research Laboratory, 3251 Hanover Street, Palo Alto, CA 94304, U.S.A.

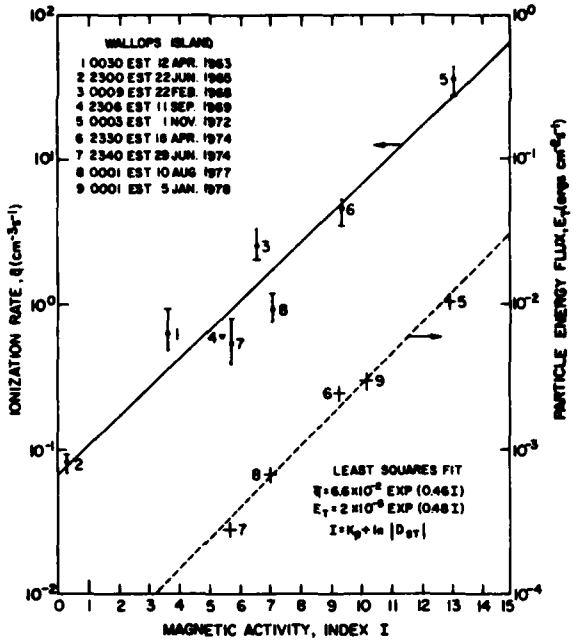


Fig. 2. The variation with magnetic activity of the nighttime ionization rate in the upper E region and of the particle energy flux obtained from rocket experiments [*Voss and Smith*, 1979].

fined, by I ≡ Kp + 1n |Dst|, for Dst ≠ 0. This is chosen to give the best fit to the data and can be related to the physical processes involved.

The best-fit relation between the ionization production rate (q) and the magnetic activity index (I) is given by

$$q = 6.6 \times 10^{-2} \exp(0.46I) \text{ cm}^{-3}\text{s}^{-1}$$

Inclusion of detectors on later rocket flights has revealed that the energy flux, also shown in Figure 2, has the same dependence on magnetic activity as does the ionization rate. A surprising result of these flights is that the energetic particles at Wallops Island (L = 2.6) are protons, presumably orginating in the ring current and precipitating as a result of wave-particle interactions.

The observations at Wallops Island provided the stimulus for an examination of the global morphology of particle precipitation at middle and low latitudes. It has been found that the observations at Wallops Island are characteristic of a narrow zone, the midlatitude zone of proton precipitation, near L = 2.6, but, in addition to the now well-established equatorial zone of proton (and energetic neutral hydrogen) precipitation there is another zone, the low latitude zone of electron precipitation. In the following sections of this paper we summarize the evidence for these zones and the properties of the precipitating particles.

2. THE MIDLATITUDE ZONE

Rocket observations of energetic particles at Wallops Island, Virginia, Figure 3, show that the flux is effectively zero below 120 km and increases linearly from that altitude to at least 190 km.

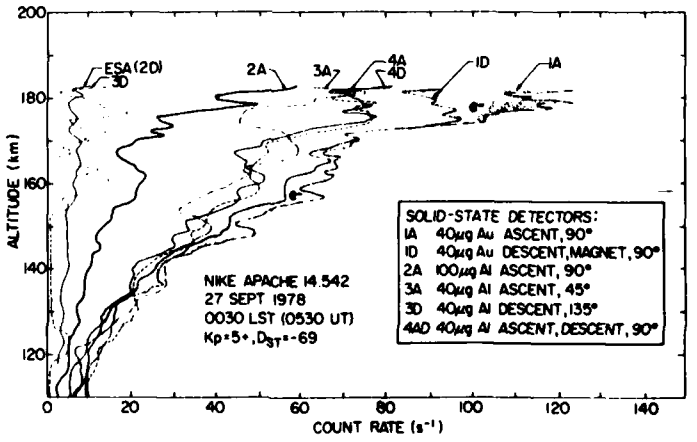


Fig. 3. Detector count rates as a function of altitude [*Voss et al.*, 1980].

In Figure 3 the count rates are obtained from six solid-state detectors and an electrostatic analyzer (ESA). The solid-state detectors are sensitive to particles with E > 15 keV; the ESA is used for particles with 1 < E < 10 keV. The solid-state detectors are ruggedized surface-barrier silicon devices with a sensitive area of 50 mm^2 and a depletion depth of 300 μm. Data on this flight are obtained on ascent (A) or descent (D) or both (AD).

As indicated in Figure 3 the solid-state detectors have front surfaces of gold or aluminum of different thicknesss. Comparison between detector signals permits the identification of the particles. Also one detector is equipped with a magnet, which excludes electrons with E < 150 keV.

Comparison of profiles 1A and 1D immediately indicates that the energetic particles are predominantly protons (and heavier ions), with electrons contributing about 10 percent of the total flux.

The largest fluxes are observed by the detectors oriented at 90° to the rocket axis. These are detecting particles near 90° pitch angles. Detector 3D shows that only a very small flux is backscattered.

The energy flux (E > 20 keV) deduced from these measurements is 2×10^{-3} erg cm^{-2} s^{-1}. This is a significant ionization source for the region 120 to 200 km altitude.

The latitudinal extent of this zone shows up clearly in satellite measurements which are plotted in Figure 4. The midlatitude zone is clearly separated from the auroral zone by the plasmapause. It can also be noted that the midlatitude Stable Auroral Red (SAR) arcs usually occurs on the auroral side of the midlatitude zone during disturbed conditions.

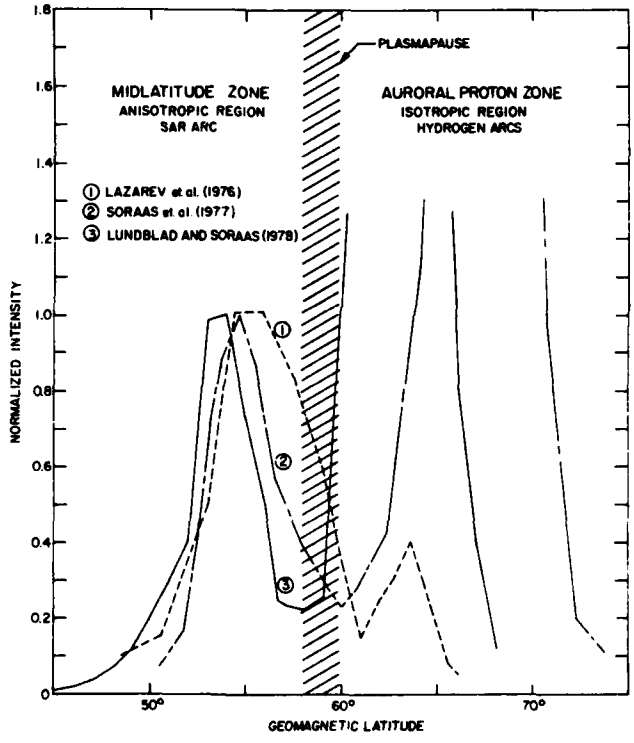


Fig. 4. Satellite observations of the midlatitude zone and auroral proton zone [*Voss and Smith*, 1980b].

3. THE EQUATORIAL ZONE

The existance of an equatorial zone of precipitation of protons and energetic hydrogen atoms has been established by particle and airglow measurements. The latitudinal extent of the zone as it appears in satellite observations of particle flux and of airglow is shown in Figure 5. The zone is seen to be limited to ±20 degrees in latitude, centered on the geomagnetic equator. The existance of the zone is also supported by rocket measurements and theoretical studies. See *Voss and Smith* [1980b] for detailed references.

Rocket and satellite measurements indicate that the particles are concentrated at a pitch angle of 90°. The intensity increases linearly from essentially zero at 180 km to 260 km, above which it becomes nearly independent of altitude.

The energy spectrum at low altitudes is similar to that of the ring current, with most particles having energies in the range 10 < E < 100 keV. Also the variation in intensity is similar to that of the ring current (as measured by Dst).

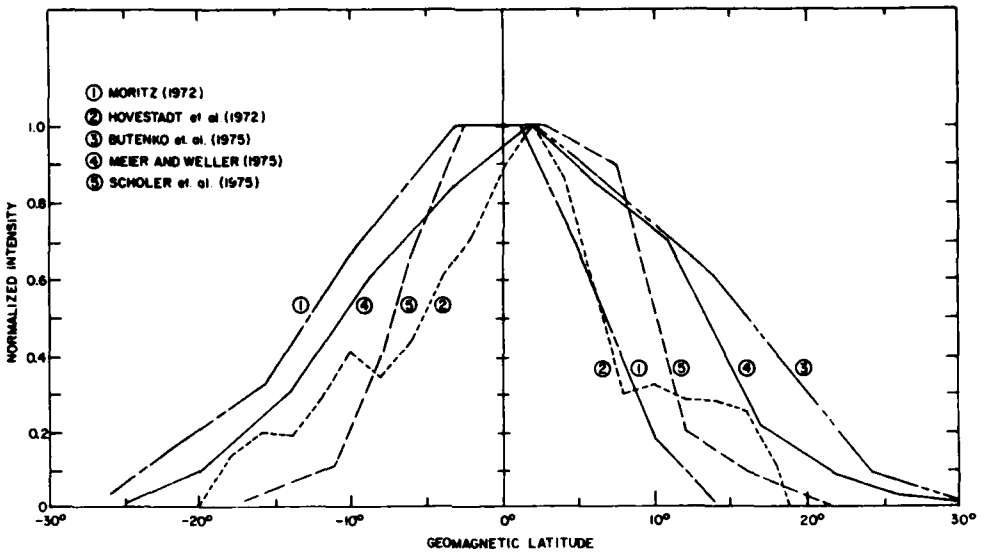


Fig. 5. Satellite observations of the equatorial zone [*Voss and Smith*, 1980b].

The explanation of this equatorial precipitation is believed to be a double-charge exchange process of protons originating in the ring current. The charge-exchange process is consistant with the altitude profile of the flux, with the energy spectrum and with the pitch-angle distribution. The extent of the zone in altitude has, however, not yet been satisfactorily explained: present theories indicate a larger range of latitude than is observed.

An interesting feature of the particle flux at the equator is an east-west asymmetry in flux. Figure 6 shows a rocket observation of this effect. The asymmetry has been interpreted by *Voss and Smith* [1980a] as a consequence of the large radius of gyration of energetic protons about the (horizontal) magnetic field.

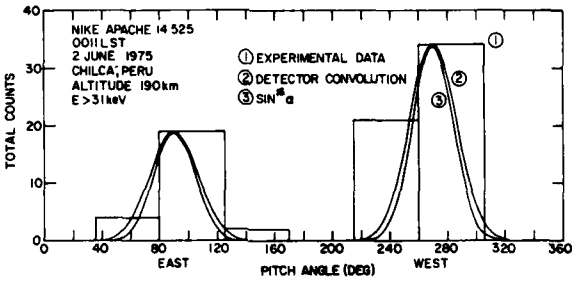


Fig. 6. East-west asymmetry in particle flux at the equator [*Voss and Smith*, 1980a].

4. THE LOW-LATITUDE ZONE

Evidence for the low-latitude zone of energetic particle precipitation is, perhaps, more circumstantial than for either the mid-latitude or equatorial zones. *Voss and Smith* [1980b, which should be consulted for the references contained there] conclude that there is a zone near L = 1.4 in which the precipitating particles are predominantly energetic electrons, with a small contribution of energetic protons.

Ionospheric effects in the form of enhanced ionization at night at Arecibo (L = 1.4) have been attributed to particle precipitation. This is supported by observations of plasma lines and of 427.8 nm nightglow, also at Arecibo.

The global extent of this zone is indicated in 341.4 nm airglow enhancements and ionospheric particle effects observed in Japan (at L = 1.38), in satellite measurements of low energy electrons (E < 26 keV) and by rocket measurements at low latitudes by Japanese, Russian and American investigators.

The precipitated particles of the low-latitude zone are predominantly energetic electrons. This is shown most definitely in the measurements of *Turiel and MacGregor* [1970] and of *Seward et al.* [1973]. Using magnetic rejection and window transmission characteristics in rocket experiments at 125° longitude (and L = 1.45) they find the particles are electrons (E ≥ 2 keV) which have pitch angles near 90°.

An energy flux of 10^{-3} erg cm^{-2} s^{-1} is typical of this zone during moderately disturbed conditions. This is consistant with airglow measurements and the ionospheric enhancements of the upper E region.

There appears to be a minimum of energetic particle precipitation between L values of 1.6 and 2.2., i.e. between the midlatitude and low-latitude zones. This is shown by very low intensities of 391.4 nm radiation, by low electron densities in the upper E region and by direct measurements of particle fluxes.

There is no corresponding well-defined minimum of particle precipitation between the low-latitude and equatorial zones. Here the separation is characterized by the transition of the predominant energetic particles from electrons to protons.

5. GLOBAL MORPHOLOGY

The latitude extent and relative intensity of the zones of particle precipitation are indicated in Figure 7, for moderately disturbed conditions. At high latitudes the overlapping regions of electron and proton precipitation are indicated. This is separated from the midlatitude proton zone by the plasmapause. Between the midlatitude and low-latitude zones is a gap (2.2 > L > 1.6) of low particle intensity. The low-latitude and equatorial zones overlap but represent a transition from electron to proton precipitation.

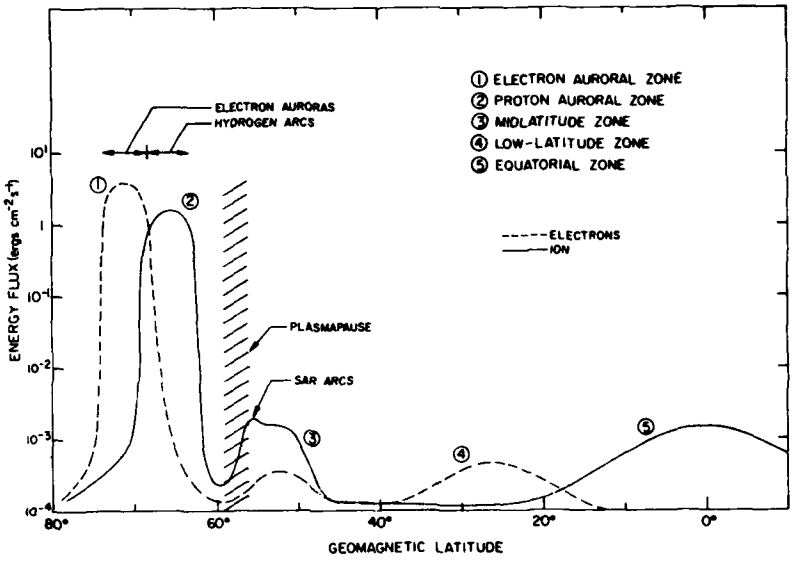


Fig. 7. Particle precipitation as a function of latitude [*Voss and Smith*, 1980b].

The latitude variation is modified on a global scale by the South Atlantic anomaly in the magnetic field. As indicated in Figure 8 the major effects are an intensification, with maxima at L = 1.4 and L = 2.6, in the region of the anomaly, and by a depletion of the intensity of the low-latitude zone in the northern hemisphere extending from about 45°W to 75°E.

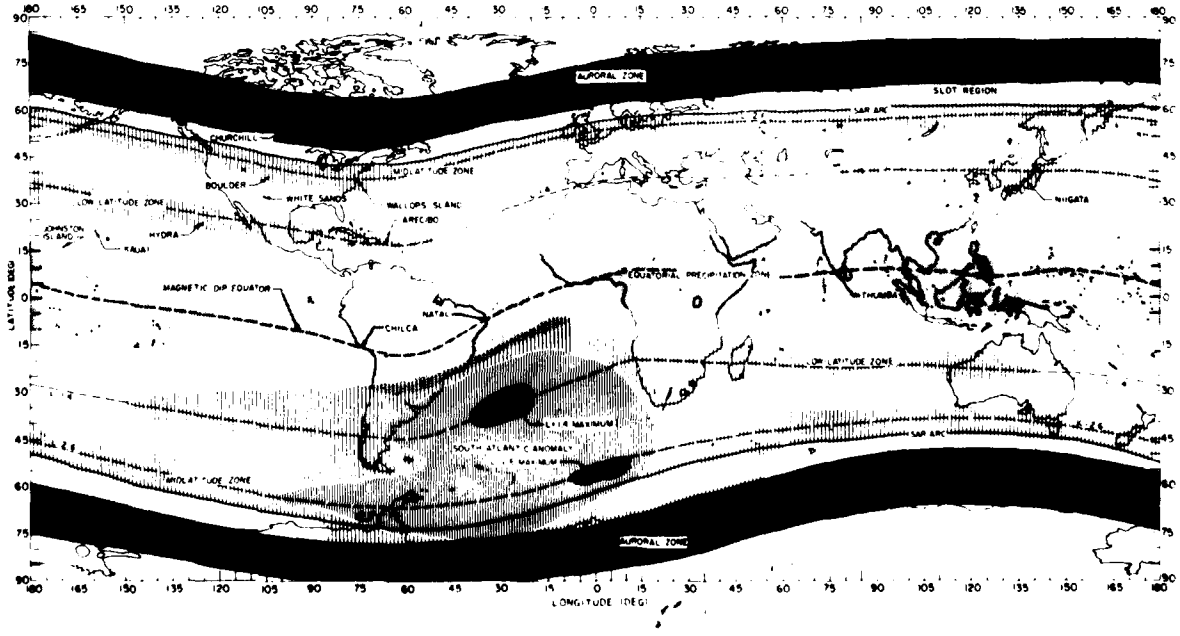


Fig. 8. Global zones of particle precipitation [*Voss and Smith*, 1980b].

6. CONCLUSIONS

Analysis of data from rocket, satellite and ground-based measurements shows a complex picture of particle precipitation at middle and low latitudes. The precipitation has zonal maxima and some meridional variation. The interpretations appears to be broadly consistant with our present, rather limited knowledge of the trapped radiation in the earth's upper atmosphere, but, as yet, no detailed explanation of the global morphology exists, particularly with respect to the limited extent of the zones in latitude.

ACKNOWLEDGEMENT

The work has been supported by the National Aeronautics and Space Administration under grant NGR 14-005-181.

REFERENCES

Butenko, V. D., et al., 1975, Proton currents with Ep > 70 keV at low altitudes in the equatorial region. *Cosmic Res.*, *13*, 457-460.

Geller, M. A., et al., 1975, Analysis of the nighttime E-region winds and ionization production. *Radio Sci.*, *10*, 335-345.

Hovestadt, D., et al., 1972, Observation of energetic particles at very low altitudes near the geomagnetic equator. *Phys. Rev. Lett.*, *78*, 1340-1344.

Lazarev, V. I., et al., 1976, Features of the 2-20 keV electron and proton flux distribution in the magnetosphere according to the measurements from Molniya I. *Space Res.*, *16*, 523-527.

Lundblad, J. A., and Soraas, F., 1978, Proton observations supporting the ion cyclotron wave heating theory of SAR arc formation. *Planet. Space Sci.*, *26*, 245-254.

Meier, R. R., and Weller, C. S., 1975, Observations of equatorial EUV bands: Evidence for low-altitude precipitation of ring current helium. *J. Geophys. Res.* *80*, 2813-2818.

Moritz, J., 1972, Energetic protons at low equatorial altitudes: A newly discovered radiation belt phenomenon and its explanation. *Z. Geophys.*, *38*, 701.

Scholer, M., et al., 1975, Energetic He^+ ions from the radiation belt at low altitudes near the geomagnetic equator. *J. Geophys. Res.*, *80*, 80-85.

Seward, F. D., et al., 1973, Electrons at low altitudes: A difficult background problem for soft X-ray astronomy. *Report UCRL-51470*, Lawrence Livermore Laboratory, Univ. Cal., Livermore, Cal.

Smith, L. G., 1970, A sequence of rocket observations of nighttime sporadic-E. *J. Atoms. Terr. Phys.*, *32*, 1247-1257.

Smith, L. G., et al., 1974, Energetic electrons in the midlatitude nighttime E-region. *J. Atmos. Terr. Phys.*, *36*, 1601-1612.

Soraas, F., et al., 1977, On the energy dependence of the ring current proton precipitation. *Planet. Space Sci.*, *25*, 757-763.

Strobel, D. F. et al., 1974, The nighttime lower ionosphere. *J. Geophys. Res.*, *79*, 3171-3178.

Turiel, I. and MacGregor, A., 1970, An observation of mirroring electrons at low altitudes. *Report UCRL-72615*, Lawrence Radiation Laboratory, Univ. Cal., Livermore, Cal.

Voss, H. D., and Smith, L. G., 1977, Energetic particles and ionization in the nighttime middle and low latitude ionosphere. *Aeron. Report 78*, Aeron. Lab., Dep. Elec. Eng., Univ. Ill., Urbana, Ill.

Voss, H. D., and Smith, L. G., 1979, Nighttime ionization by energetic particles at Wallops Island in the altitude region 120 to 200 km. *Geophys. Res. Lett.*, *6*, 93-96.

Voss, H. D., and Smith, L. G., 1980a, Rocket observations of energetic ions in the nighttime equatorial precipitation zone. *Low Latitude Aeronomic Processes (ed. A. P. Mitra)* 131-134, Pergamon Press, Oxford.

Voss, H. D., and Smith, L. G., 1980b, Global zone of energetic particle precipitation. *J. Atmos. Terr. Phys.*, *42*, 227-239.

Voss, H. D., et al., 1980, Rocket measurements of energetic particles in the midlatitude precipitation zone. *Space Res.*, *20*, 149-152.

ENERGETIC O^+ PRECIPITATION: A SIGNIFICANT ENERGY SOURCE FOR THE IONOSPHERE

Marsha R. Torr and D. G. Torr
Center for Atmospheric and Space Sciences, Department of Physics
Utah State University, Logan, Utah 84322

Also affiliated with: National Institute for Telecommunications
Research of the C.S.I.R., Johannesburg, South Africa

R. Roble
National Center for Atmospheric Research
PO Box 3000 Boulder Colorado 80307

SUMMARY

The energy flux associated with the precipitation of energetic O^+ ions during magnetic storms constitutes a significant source of heating for the F region. The precipitation events have been observed to occur during every magnetic storm that was monitored during the course of a year, and also during non-storm conditions at reduced intensities. Ion fluxes with corresponding energy fluxes as large as 0.4 ergs cm^{-2} sec^{-1} sr^{-1} have been measured. The duration of these events is of the order of a day, and they occur at subauroral/auroral latitudes. The associated energy influx results in an energy deposition rate at 250-300 km on the nightside that is comparable with the corresponding solar UV deposition rate on the dayside. As such the energetic oxygen events constitute a direct coupling between the magnetosphere and thermosphere. In this paper we describe the interaction of the O^+ ions with the thermosphere and their subsequent history. The large scale effects of this precipitation are examined using a global thermospheric circulation model. A mechanism is suggested by which the energetic O^+ fluxes may be self-sustaining.

1. INTRODUCTION

During the early 1970's, the discovery of the precipitation into the atmosphere of large fluxes of energetic oxygen ions was reported by Shelley et al. (1972)[1]. In a series of subsequent papers (Shelley et al., 1974 [2]; Sharp et al., 1974[3]; Sharp et al., 1976a,b[4,5]) they proceeded to further characterize the O^+ fluxes. The measurements were made at an altitude of 800 km over an energy range of 0.7 to 12 keV, and at a pitch angle near to the edge of the loss cone, 55^o - 58^o. The precipitating O^+ fluxes were found to occur over a wide latitudinal range, $2 \lesssim L \lesssim 9$, and were observed in every storm studied over a one year period. They were also found to occur at reduced intensities during non-storm conditions.

While exhibiting larger fluxes on the nightside, the energetic oxygen ions were found at all local times. The dayside fluxes were of the order of 5 - 10 times lower than those on the nightside. The average particle energy during a two day magnetic storm period (16/17 December 1971) was 4.5 keV. However, energy spectra have been observed in which the intensity was still increasing at the highest energy observed, 12 keV. The energy fluxes that have been observed are considerable, reaching 0.4 ergs cm^{-2} sec^{-1} sr^{-1} on the nightside.

As the bulk of the observations were made at a pitch angle of $55\text{-}58^o$ (i.e. primarily precipitating particles near the edge of the loss cone), little information exists concerning the pitch angle distribution. However, preliminary measurements made with an earlier instrument at pitch angles of $17\text{-}35^o$ and a co-incident ISIS crossing of the same region, resulted in an inference of an isotropic distribution.

Sharp et al. (1976a)[4] examined the relationship of the O^+ fluxes to the geomagnetic disturbance indices, D_{st} and AE. They found a strong correlation in both cases, but differing in detail. They pointed out that the correlation with D_{st} is suggestive of a ring current source, while the AE correlation suggests the possibility of a source associated with magnetic substorms.

In view of the magnitude of the energy influx to the ionosphere associated with these events, Torr et al (1974)[6] investigated the potential effects on the atmosphere. They extended the analysis in a later paper (Torr and Torr, 1979 [7]). In these two studies it was found that the two main energy loss processes for O^+ ions in the 0.7 to 12 keV range both involve collisions with atomic oxygen, the major atmospheric constituent at these altitudes. The processes are

charge exchange

$$O_f^+ + O \rightarrow O_f + O^+ \qquad (1)$$

and momentum transfer

$$O_f^+ + O \rightarrow O_f^+ + O_f \qquad (2)$$

where the subscript f denotes a fast particle.

The first of these processes, charge exchange, results in the production of an oxygen atom with almost the energy of the incident O^+ ion. Thus the energy of the O^+ ion is transferred to an O atom. In the case of the second process, the collision shares the energy of the incident O^+ ion between the ion and an O atom with which it collides, thus producing an O^+ ion and an O atom each with half the energy of the

incident O^+ ion. As a result of these O^+–O interactions, in less than ∿12 collisions, a 12 keV O^+ ion transfers all of its energy to neutral O atoms. The cross-sections for collision processes (1) and (2) are large over the energies considered, $\sim 10^{-15}$ cm^2 (Lo and Fite, 1970 [8]; Amdur et al., 1957 [9]) and so we neglect other processes such as ionization and excitation.

The energetic oxygen atoms thus produced, in turn thermalize by colliding with oxygen atoms. It is this process which we shall show has a significant impact on the atmosphere. The energy loss of the fast oxygen atoms is not simply analogous with the auroral case of protons and hydrogen atoms. Most proton auroral studies are done over an energy range of 5 to 100 keV. In this energy range the major energy loss processes are charge exchange, ionization and excitation. For proton and hydrogen atoms, momentum transfer only becomes important below 1 keV. However, the cross-sections are velocity dependent and a 1 keV H atom is equivalent to a 16 keV O atom. Therefore, over the energy range that we are considering here, the cross-section for momentum transfer is significantly larger than the cross-section for ionization.

Torr et al. (1974)[6] developed a model to trace the thermalization of the energetic oxygen atoms. The model uses the methods of radiative transfer to keep track of the oxygen atoms as they scatter in height, energy and angle while approaching thermal energies. Preliminary results from this model showed three major effects. The first of these effects is that almost all of the energy of the initial O^+ flux is transferred to the neutral atmosphere at F-region altitudes (250-350 km) in the form of heating. This constitutes a heating source on the nightside that is larger in magnitude than the solar UV heat deposition at the same altitudes.

A second major outcome of the earlier studies is a result of the fact that each energetic O^+ ion generates a large number of low energy O atoms. A large flux of these low energy O atoms is generated at 600 km moving upwards. An estimate was made of the flux entering ballistic orbits and the subsequent ionization fraction (via photoionization and charge exchange with H^+). It was found that these atoms would result in an O^+ flux 10 times larger than the incoming O^+ flux in the 0.7-12 keV range. The results thus suggest that these atoms constitute the O^+ fluxes in the ∿30 eV range measured by the ISEE satellite (Chappell, private communication, 1979 [10]).

Subsequent acceleration of a fraction of these ions by some unidentified mechanism may also provide a source for the keV O^+ fluxes, thus making the heavy ion events self-sustaining.

Torr et al. (1974 [6]) showed that there is an upward moving flux of oxygen atoms with energies greater than the escape velocity from the earth. An average global escape flux was estimated and found to be of a magnitude that may be sufficient to dissipate the atmospheric oxygen over the lifetime of the earth.

Lack of accurate cross-section data for the relevant collision processes and the simplifications imposed by the models that we have applied, mean that the results drawn from such studies are strictly first order indicators of the effects. However, the energetic oxygen atom population that is produced is so large (e.g. a single 1 keV O atom will ultimately produce 500 2 eV O atoms after multiple elastic scattering collisions) that even if the assumed parameters are found to be somewhat in error, the basic conclusions will not be affected.

In the study reported here, we further refine the earlier models and extend the investigation to consider the impact of the heating source on global thermospheric circulation models. We also investigate the ionization resulting from the multiple O-O collisions.

Because of limited computer resources, the earlier studies treated the idealized case of a single fast O atom incident at the top of the atmosphere with a given energy, e.g. 2000 eV. The subsequent thermalization "cloud" of oxygen atoms was modelled and all conclusions were based on the extrapolation of these results to the observed O^+ flux spectra. Here for the first time we treat the conversion of an actual observed O^+ flux spectrum to fast neutral atoms and model the resulting thermalization process. In the calculations reported below, we shall use the example O^+ flux spectra measured by Shelley et al. (1972) [1] during the magnetic storm of 16/17 December 1971. The spectra observed at three different latitudes are reproduced in Figure 1.

2. RESULTS

2.1. Conversion of energetic O^+ to energetic O.

In determining the conversion of O_f^+ to O_f, we assume that the only significant processes are charge exchange and momentum transfer i.e. equations (1) and (2) above. The energy loss from the O^+ beam is computed using the expressions given by Dalgarno and Griffing (1955)[11]. We assume that the incident O^+ flux is injected isotropically into the downward hemisphere. We further assume that the magnetic field line is perpendicular to the surface of the earth, and we do not consider angular scattering in the O^+ - O collisions.

In Figure 2 we show the production rates of fast oxygen atoms resulting from mono-energetic O^+ particles incident at ∿800 km with energies of 2000 eV and 10,000 eV respectively. For these illustrations the mono-energetic particles are injected with unit fluxes isotropic over the downward hemisphere. It can be seen that the conversion from energetic O^+ to energetic O peaks near 500 km, and this will be the main modifying factor over our earlier highly simplified injection of mono-energetic O near 800 km.

In Figure 3 we show the energetic O production rate profiles that result from the three O^+ spectra shown in Figure 1. Of these we shall use the L ∿ 3.4 example in the calculations reported below.

As all of the modelling is done in terms of an optical depth parameter, τ, (see Torr et al., 1974) where

$$\tau = \int_{h}^{\infty} n\,dh \qquad (3)$$

it is necessary to assume a model atmosphere in order to convert from τ to altitude. For this purpose we have used the MSIS model appropriate to magnetic storm conditions and midlatitudes at night (Hedin, 1979 [12]).

2.2. Thermalization of energetic oxygen atoms

We now apply the thermalization model described in our earlier papers (Torr et al., 1974 [6]; Torr and Torr, 1979 [7]) to the energetic oxygen atoms produced via the O_2^+-O collisions discussed above. In Figures 4 and 5 we show the distribution that the energetic oxygen fluxes assume as a function of optical depth and elevation angle for various energies. The initial input flux produces a large low energy flux which tends to an isotropic angular distribution after the particles have penetrated some depth into the atmosphere. However, because of the exponential increase in the number of collisions with decreasing altitude, the angular distribution at very small optical depth shows a larger flux of upward moving particles.

We shall address these large fluxes of low energy oxygen atoms further in the sections below.

2.3. Atmospheric heating by the thermalization of energetic oxygen atoms

Once the energetic oxygen atoms have degraded in energy down to below 2 eV, we consider them thermal and their kinetic energy is effectively transferred to the ambient neutral atmosphere as heat. The heating rate can be determined by computing the production rate of oxygen atoms with energies less than 2 eV and multiplying by the energy of the particle. Torr and Torr (1979) [7] computed this heating rate for the 2 keV and 10 keV unit flux cases. It was found that almost the entire (~96%) incoming O^+ flux is transferred to the neutral atmosphere as heat.

In Figure 6 we show the heating rate that would result from the O^+ spectra measured by Shelley et al (1972) [1] during the December 1971 storm. These results show, as did the preliminary study, that the bulk of the input energy is deposited as heat in the F region. There are, however, two updates to the earlier work. The heating rate profile peaks at a higher altitude than before, namely ~340 km rather than 270 km. We find also that the fraction of input energy converted into heat is ~60% rather than ~90% obtained from the preliminary results, and a larger fraction escapes the atmosphere.

2.4. Effect of energetic oxygen heating source on global thermospheric circulation

For O^+ flux events of the magnitude observed by Shelley et al. (1972) [1] during the December 1971 magnetic storm, i.e. 0.4 ergs cm^{-2} sec^{-1} sr^{-1}, the incoming energy flux to the nocturnal F-region (250-350 km) can be comparable to the solar UV energy influx at these altitudes on the dayside. We have shown in the previous section that this O^+ energy is coupled directly into the neutral atmosphere as heat. The energy input to a fairly restricted latitudinal range is thus considerable. It thus becomes of interest to investigate the perturbation that this localized heat source would produce in a global thermospheric circulation model. We have therefore parameterized the heating rate as a function of latitude, altitude and local time, and have made a preliminary study of the effects of this energy input on the neutral atmosphere, using such a model. At the time of writing this paper, we had only investigated the effects of a heat source deduced from the thermalization of a 2000 eV particle flux and extrapolated to an O^+ spectrum such as we have now computed and shown in Figure 6. The actual heating rate results differ from the extrapolated results in certain respects, but the initial conclusions drawn from the global model results are illustrative of the effects. We present these results in this section, and the calculations will be extended and improved in the ongoing analysis. The global circulation model has been described by Dickinson et al. (1980) [13].

In our initial introduction of the energetic oxygen heating source to the global thermospheric circulation model, certain program instabilities were encountered due to the magnitude of the heating rate. For the preliminary investigation we have therefore considered an energy flux an order of magnitude smaller than that observed by Shelley et al., (1972) [1] during the December 17, 1971 storm. It is therefore important to bear in mind in considering the results shown in this section that we have used an input flux of 0.04 ergs cm^{-2} sec^{-1} sr^{-1} rather than 0.4 ergs cm^{-2} sec^{-1} sr^{-1}. The results may thus be more typical of average storms but the effects are less than would be encountered during larger storms.

We have examined the effects of this heavy ion heating on such global circulation aspects as velocity distributions, perturbations to the neutral atmospheric temperature, and zonal mass flow stream functions. The computations have been made using solar heating as the only other heat source, i.e. no auroral joule heating effects are included in this initial comparison.

In Figure 7 we show the effect on the vertical velocities. These are pronounced. The eastward and northward velocities show less pronounced effects (Figures 8 and 9 respectively). In Figure 10 we show the effect on zonal average mass stream flow, and in Figure 11 we show the perturbations to the global mean temperature. Even at the reduced energy influx used here, the effects are significant and worthy of further study. In future models we plan on optimizing the heating results from Section 2.3 and running more extensive comparisons.

2.5. Escape of energetic oxygen from the Earth's atmosphere

We have shown in Section 2.2 that a large flux of upward moving O atoms is generated at the top of the atmosphere. Those particles moving upward near $\tau = 0$ with energies larger than 10 eV (the escape velocity) are potentially capable of escaping from the earth. From the thermalization model we can evaluate the escape flux of atomic oxygen. For the incoming O^+ flux example measured by Shelley et al (1972) [1] for L ~ 3.4 (Figure 1) the escape flux would amount to 30% of the total incoming energy.

because of the avalanche effect as higher energy particles elastically collide and produce lower energy particles, this 90% in energy corresponds to a large escape number flux. In the case of this particular observed O^+ flux spectrum which corresponds to a total particle flux of $\sim 3.5 \times 10^8$ particles cm^{-2} sec^{-1}, a flux of $\sim 4 \times 10^{10}$ cm^{-2} sec^{-1} is generated that is potentially capable of escaping from the atmosphere.

It is an interesting exercise to use this data to estimate whether the escape mechanism could result in a significant loss of atmospheric oxygen over the lifetime of the earth. To do this we must estimate an average escape flux. With the data available any such estimate must necessarily be crude, but does yield an indication of the potential magnitude of the effect.

We have estimated the average escape flux as follows: If the O^+ precipitation events occur largely in a latitudinal zone between 50^0 and 70^0, they would occur over ~17% of the globe. If we assume a nightside energy influx of 0.1 ergs cm^{-2} sec^{-1} sr^{-1}, and a dayside influx of 10 times less, the local time average would be 0.05 ergs cm^{-2} sec^{-1} sr^{-1}.

The Lockheed group have reported that the energetic O^+ events were observed in every magnetic storm during the course of a year and at reduced intensities during non-storm times. We are thus probably being conservative if we assume that fluxes with a zonal average of 0.05 ergs cm^{-2} sec^{-1} sr^{-1} occur whenever Kp $\geq$ 4. An examination of Kp indices over 8 year period 1973-1980 shows that Kp $\geq$ 4 occured for 22% of the time.

Thus if we scale the escape flux computed for the December 1971 storm case to the lower 0.05 ergs cm^{-2} sec^{-1} sr^{-1} energy influx and multiply by the 17% global factor and the 22% storm time factor, we obtain an average global escape flux of 6×10^7 cm^{-2} sec^{-1} sr^{-1}.

The total column number density of oxygen is 8×10^{24} atoms cm^{-2}. Thus the flux estimated above would yield an escape time of 4×10^9 years, which is comparable with the lifetime of the earth.

2.6 Reionization of the energetic oxygen splash

If we consider the production of low energy oxygen atoms near $\chi = 0$, a considerable upward flux is generated. In addition to the escaping atoms, those moving upwards near 800km with energies between ~2 eV and 10 eV will travel in fairly long trajectory ballistic orbits. If we consider a representative energy of 6 eV, an upward moving oxygen atom with this energy can reach an altitude of 1.2×10^5 km in 8.1×10^3 seconds. Thus, during the course of its 1.6×10^4 sec trajectory, there is the possibility for the atom to be converted to an oxygen ion, either through photoionization or charge exchange with H^+ ions. The lifetime for photoionization is 2.4×10^6 sec, and if we assume a plasmaspheric H^+ concentration of ~500 cm^{-3} sec, the lifetime for charge exchange is 4×10^6 sec.

Using an average ballistic trajectory energy of 6 eV, 0.64 of the atoms will photoionize and 0.44 will charge exchange to form O^+ ions. Similarly, if we assume a plasmapause at $4R_E$, calculate the time of travel for escaping 10 eV atoms within the plasmasphere, and apply the same arguments, we find that 0.34 of the escaping flux will be converted into O^+ ions. If we consider the O^+ spectrum measured by Shelley et al. (1972) [1] at L ~ 3.4 as an example, we can estimate the escaping and ballistic trajectory fluxes. We find that for this spectrum, the escape flux would be 4×10^{10} atoms cm^{-2} sec^{-1}, and the flux capable of entering ballistic orbits would be 2.8×10^{11} atoms cm^{-2} sec^{-1}. These atoms would result in an O^+ flux of 2.8×10^9 cm^{-2} sec^{-1}, which is 10 times larger than the incoming O^+ flux in the 0.7 - 12 keV range. As the energetic O^+ ions were only measured to 12 keV, the actual low energy O^+ flux produced may be larger than this.

2.7. Ionization produced by energetic O atoms

According to Massey (1949) [14] direct ionization and excitation of O by energetic O will only be important when the velocity of the projectile exceeds the velocity of the orbital electron. For the energies considered here ($\lesssim$12 keV), the O atoms will have velocities of $\lesssim 4 \times 10^7$ cm sec^{-1}, while the velocity of the orbital electron is $\sim 10^9$ cm sec^{-1}. For this reason, we did not consider reionization of the beam in our earlier studies. However, the total number of O-O collisions resulting from an energetic O^+ precipitation event is so large, that an estimation of the ionization source is in order. An effective ionization rate can be computed by determining the total flux of fast oxygen atoms at each altitude with energies greater than 50 eV

$$P = \sigma_i [O] \int_{50}^{1200eV} F(E)dE \qquad (4)$$

where F(E) is the flux of oxygen atoms with a given energy E and σ_i is the mean cross section for ionization from an O-O collision.

Cross-section information for various O-O collisions, whether momentum transfer, ionization or excitation are difficult to find. Solov'ev et al. (1972) [15] have given values for σ_i down to energies of 11 keV. If we make a rough extrapolation of their data down to, say, 1000 eV, we obtain an estimate for σ_i of $\sim 2 \times 10^{-17}$ cm^2, which is a large ionization cross-section.

Lacking any additional information of σ_i at the lower energies, we would have abandoned the estimation of ionization effects. However, some support for a relatively large ionization cross-section has been provided by Drawin (1968, 1969) [16,17] who has published formulas for σ_i as a function of energy. Using these expressions we obtain σ_i(1000 eV) of 1.08×10^{-16} cm^2, a value even larger than that inferred from the Solov'ev et al (1972) [15] data. Thus, while the uncertainties in σ_i over the energy range in which we are interested, are large, the indications are that the cross-section is large enough to warrant an estimate of the ionization rate.

Again, the estimation is fairly crude but illustrative of the potential significance of the source. In equation (4) we have simply used a constant σ_i of 2×10^{-17} cm^2, and in Figure 12 show the ionization profile that would result from the observed O^+ flux spectrum for L ~ 3.4 (Figure 1). Near the peak of the

F_2 layer, the resulting ionization would be $\sim 60\ cm^{-3}sec^{-1}$.

At night in the topside F region there are no comparable sources of ionization. The important question is thus whether this ionization rate is competitive with the F region loss rate. Without having modelled this ionization source, it is difficult to predict the exact effects. However, we might reasonably speculate that the impact will be to apparently slow down the recombination process in the F region.

3. DISCUSSION AND CONCLUSIONS

The results presented in this paper, while an extension of earlier work, still represent a highly simplified and approximate picture of the interaction of the observed O^+ fluxes with the atmosphere and ionosphere. However, despite uncertainties in the model and the various assumptions that have been made, several conclusions can be drawn.

The bulk of the energy carried by the incoming energetic ions is transferred to the neutral atmosphere in the form of heat. This heating occurs at F-region altitudes and is sufficient to markedly perturb the global wind fields and the global temperature distributions. As such, the mechanism represents a direct coupling between the magnetosphere and the neutral atmosphere.

In addition, the O^+ precipitation events generate a large splash of backscattered oxygen atoms. A significant number of these have energies in excess of the escape velocity from the earth, and it is interesting to estimate the possible effects on the atmospheric composition over the lifetime of the earth. The indications are that the escape flux might be comparable with that of hydrogen.

A larger component of the backscattered atomic oxygen flux will, however, enter ballistic orbit trajectories with very long paths (and travel times) through the plasmasphere. We have estimated the fraction of this flux that would be reionized by photoionization and by charge exchange with H^+. This process would result in a low energy O^+ flux that is larger than the incoming O^+ flux observed at the higher energies (0.7 - 12 keV). We suggest that this mechanism (reionization of the splash atoms) produces the low energy O^+ ions observed by the ISEE satellite in the $\sim$30 eV energy range.

The first measurements of trapped ring current ions (made at energies < 50 keV and L < 4), were reported by Johnson et al (1977) [18]. It was found that O^+ and H^+ were approximately equal during the main phase of three magnetic storms, with He^+ an order of magnitude less. During a recovery phase they found O^+ to dominate, He^+ to follow in magnitude and H^+ to be the smallest. Recently evidence has been presented (Lundin et al, 1980 [19]) from the PROGNOZ-7 satellite, which shows that a large fraction of the ring current ($\sim$82% on occasion) is O^+ ions for L < 4. In all cases analysed during quiet and recovery phases, they found O^+ and He^+ more abundant than H^+ for L < 4. Earlier, Lyons and Evans (1976) [20] presented theoretical arguments as to why the ring current ions at energies < 50 keV and L < 4 would not be dominated by protons during the recovery phase of large geomagnetic storms. Both O^+ and He^+ would meet the requirements for a longer lifetime for charge exchange with hydrogen than that of protons. Tinsley (1976) [21] also argued for a decrease of H^+ following a storm main phase.

Thus the evidence is that O^+ is a major, if not the dominant component of the ring current at all times. It is interesting to speculate on the relationship between this finding and the results reported here. A possible scenario might be that the energetic ions have their source in the ring current and are dumped into the atmosphere during storms. The precipitation produces the large splash of energetic O predicted by our theory, and the subsequent reionization of this splash produces very large fluxes of lower energy (<30 eV) O^+ ions. These ions return to the ring current where they are accelerated to the keV energies. Thus the entire process might be self sustaining.

The O^+ precipitation events produce F-region ionization that is significant under certain conditions. The potential excitation effects are unknown at this time.

4. REFERENCES

1. SHELLEY, E. G., R. G. JOHNSON and R. D. SHARP,(1972), Satellite observations of energetic heavy ions during a geomagnetic storm, J. Geophys. Res., 77, 6104.

2. SHELLEY, E. G., R. G. JOHNSON and R. D. SHARP,(1974), Morphology of energetic O^+ in the magnetosphere, in Magnetospheric Physics, p. 135-139. Reidel Publishing Company, Dordrecht-Holland.

3. SHARP, R. D., R. G. JOHNSON, E. G. SHELLEY, and K. K. HARRIS, (1974), Energetic O^+ ions in the magnetosphere, J. Geophys. Res., 79, 144.

4. SHARP, R. D., R. G. JOHNSON and E. G. SHELLEY,(1976a) The morphology of energetic O^+ ions during two magnetic storms: temporal variations, J. Geophys. Res. 81, 3283.

5. SHARP, R. D., R. G. JOHNSON, and E. G. SHELLEY, (1976b), The morphology of energetic O^+ ions during two magnetic storms: latitudinal variations, J. Geophys. Res. 81, 3292.

6. TORR, Marsha, R. J. C. G. WALKER and D. G. TORR,(1974), Escape of fast oxygen fromthe atmosphere during geomagnetic storms, J. Geophys. Res. 79, 5267,.

7. TORR, Marsha R. and D. G. TORR,(1979), Energetic oxygen: a direct coupling mechanism between the magnetosphere and thermosphere, Geophys. Res. Lett. 6, 700.

8. LO, H. H., and W. L. FITE, (1970), Electron-capture and loss cross sections for fast heavy particles passing through gases, Atomic Data, 1, 305.

9. AMDUR, I., E. A. MASON, and J. E. JORDAN, (1957), Scattering of high velocity neutral particles, 10, He-N_2; A-N_2: The N_2-N_2 interaction, J. Chem. Phys., 27, 527.

10. CHAPPELL, C. R., (1979), private communication.

11. DALGARNO, A. and G. W. GRIFFING, (1955), Energy loss of protons passing through hydrogen, Proc. Roy. Soc.Ser. A., 232, 423.

12. HEDIN, A. E., (January 1979), Tables of thermospheric temperature, density and composition derived from satellite and ground based measurements, Ap = 100, Goddard Space Flight Center Report.

13. DICKINSON, R. E., E. C. RIDLEY, and R. G. ROBLE, (July 1980) A three-dimensional general circulation model of the thermosphere, submitted to J. Geophys. Res.

14. MASSEY, H. S. W., (1949), Collisions between atoms and molecules at ordinary temperatures, Rep. Progr. Phys., 12, 248.

15. SOLOV'EV, E. S., R. N. IL'IN, V. A. OPARIN, I. T. SERENKOV and N. V. FEDORENKO, (1972), Capture and loss of electrons by fast oxygen atoms and ions in air, nitrogen and oxygen atoms and ions in air, nitrogen and oxygen, Sov. Phys. Tech. Phys., 17, 267.

16. DRAWIN, H. W.(1968) Zur formelmässigen Darstellung des Ionisierungsquerschnitts für den Atom-Atomstoss und über die Ionen-Elektronen-Rekombination im dichten Neutralgas, Z. Physik, 211, 404.

17. DRAWIN, H. W., (1969), Influence of atom-atom collisions on the collisional-radioactive ionization and recombination coeffiecients of hydrogen plasmas, Z. Physik. 225, 483.

18. JOHNSON, R. G., R. D. SHARP, and E. G. SHELLEY, (1977), Observations of ions of ionospheric origin in the storm time ring current, Geophys. Res. Lett, 4, 403.

19. LUNDIN, R., L. R. LYONS, and N. PISSARENKO, (1980), Observations of the ring current composition at L-values less than 4, Geophys. Res. Lett. 7, 425.

20. LYONS, L. R. and D. S. EVANS, (1976) The inconsistency between proton charge exchange and the observed ring current decay, J. Geophys Res. 81, 6197.

21. TINSLEY, B. A., (1976), Evidence that the recovery phase ring current consists of helium ions, J. Geophys. Res., 81, 6193.

ACKNOWLEDGEMENTS

This work was supported by NASA grant NAS5-24331 and NASA contract NAS8-33992. The calculations reported in this paper were carried out under a grant of computer resources from the National Center for Atmospheric Research which is sponsored by the National Science Foundation.

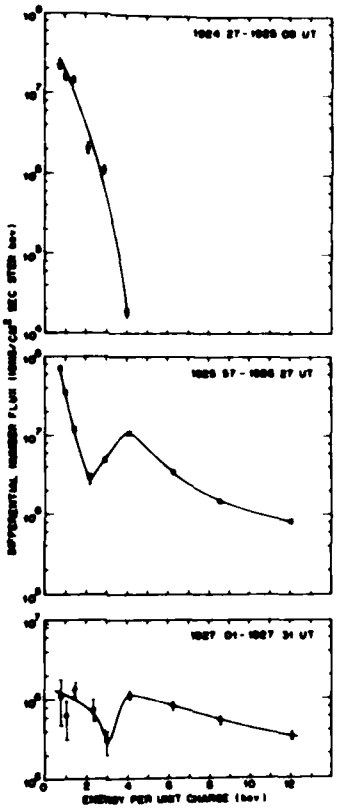


Figure 1. Flux energy spectra observed during the December 16/17, 1971 magnetic storm. From top to bottom the spectra correspond to L values of ~2.7, 3.4 and 4.2 respectively, (from Shelley et al., 1972).

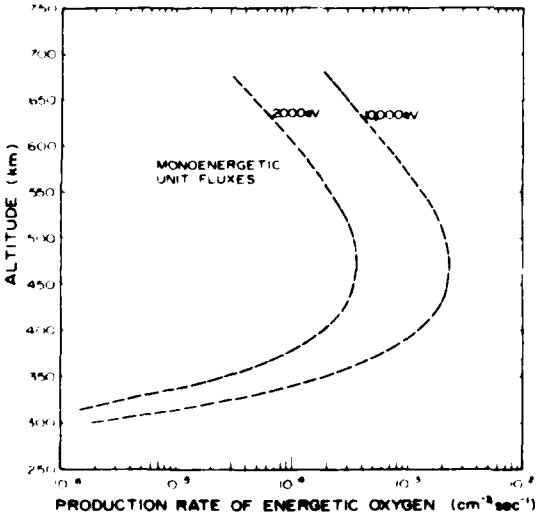


Figure 2. Production rates of energetic oxygen resulting from monoenergetic fluxes of O^+ ions at 2000 eV and 10,000 eV respectively. The monoenergetic O^+ fluxes are 1 $cm^{-2}sec^{-1}sr^{-1}$ into the downward hemisphere.

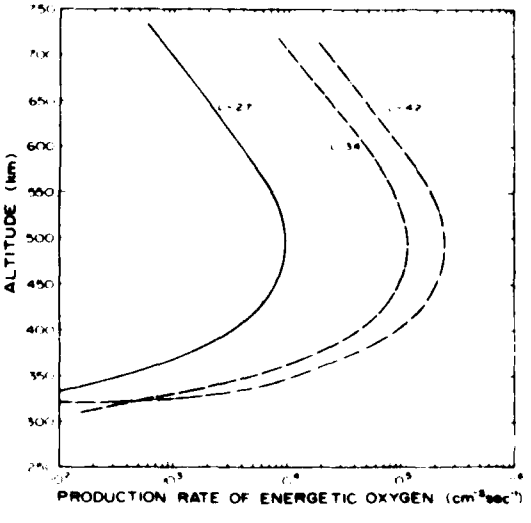


Figure 3. Production rates of energetic oxygen from the three observed O^+ spectra shown in Figure 1

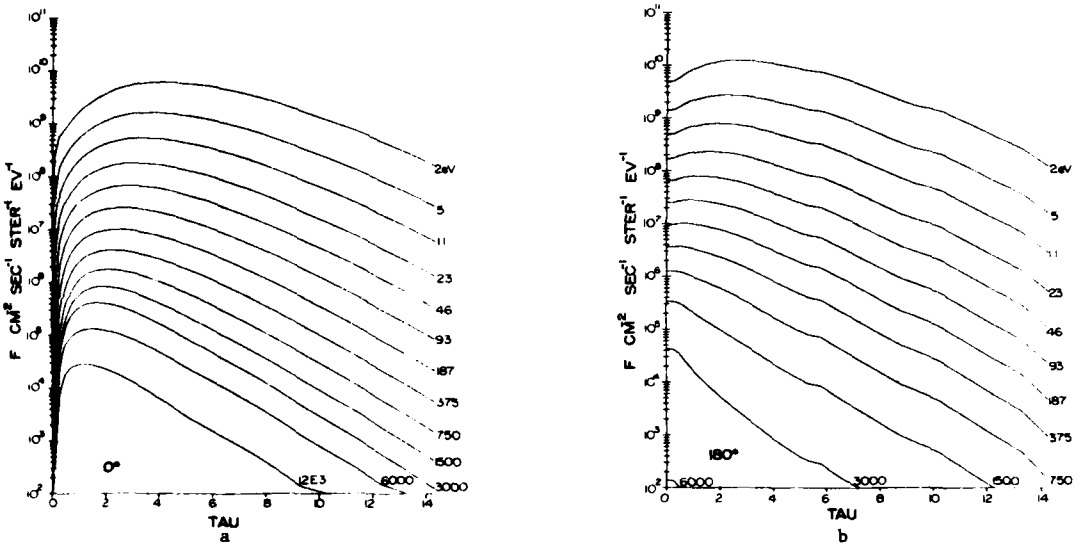


Figure 4. Flux of oxygen atoms vs optical depth (at selected energies) resulting from the thermalization of energetic oxygen atoms produced by the O^+ spectrum for L ∿3.4 in Figure 1 (a) $\theta = 0^o$ is directly downward moving flux while (b) $\theta = 180^o$ is the upward moving flux.

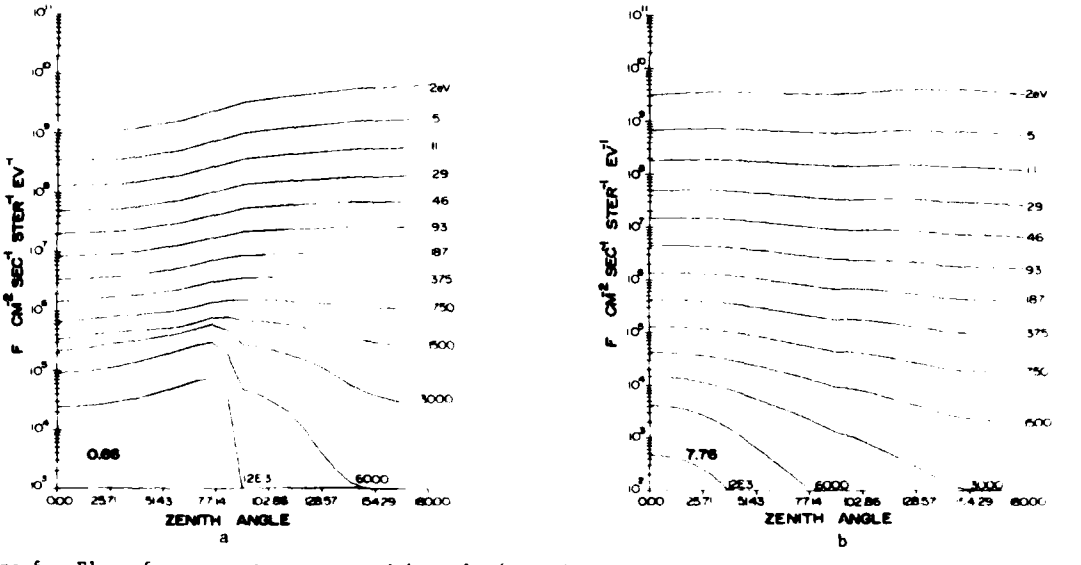


Figure 5. Flux of oxygen atoms vs. zenith angle (at selected energies) produced by the thermalization of the energetic oxygen atoms resulting from the O^+ spectrum for L ∿3.4 shown in Figure 1. Two selected optical depths are shown (a) $\tau = 0.66$ and (b) $\tau = 7.6$.

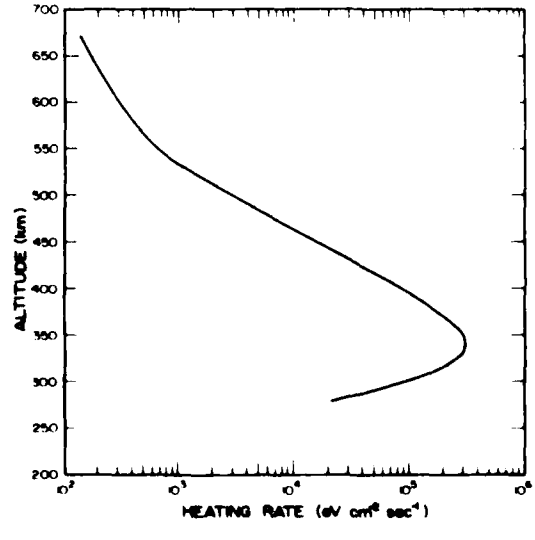


Figure 6. Heating rate corresponding to the O^+ flux spectrum for L ∿3.4 in Figure 1.

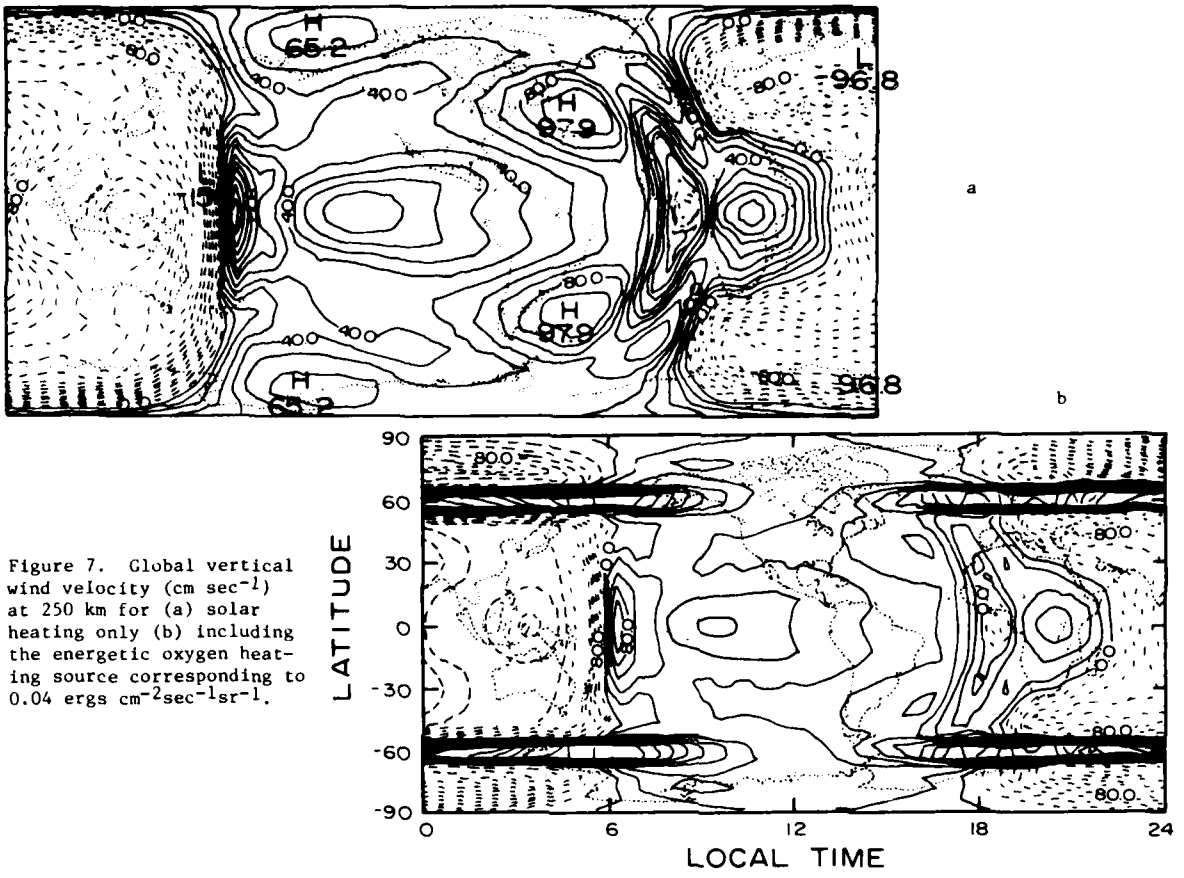


Figure 7. Global vertical wind velocity (cm sec^{-1}) at 250 km for (a) solar heating only (b) including the energetic oxygen heating source corresponding to 0.04 ergs $cm^{-2}sec^{-1}sr^{-1}$.

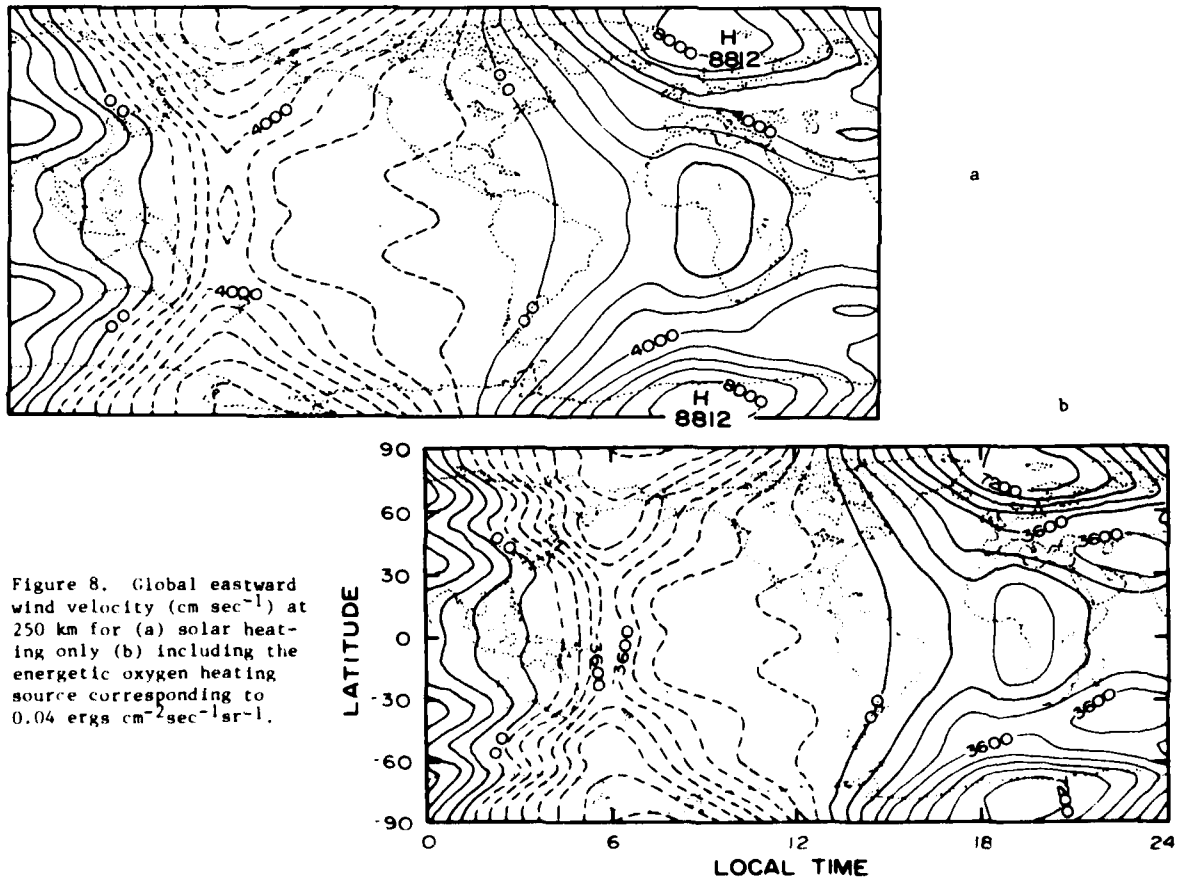


Figure 8. Global eastward wind velocity (cm sec^{-1}) at 250 km for (a) solar heating only (b) including the energetic oxygen heating source corresponding to 0.04 ergs $cm^{-2}sec^{-1}sr^{-1}$.

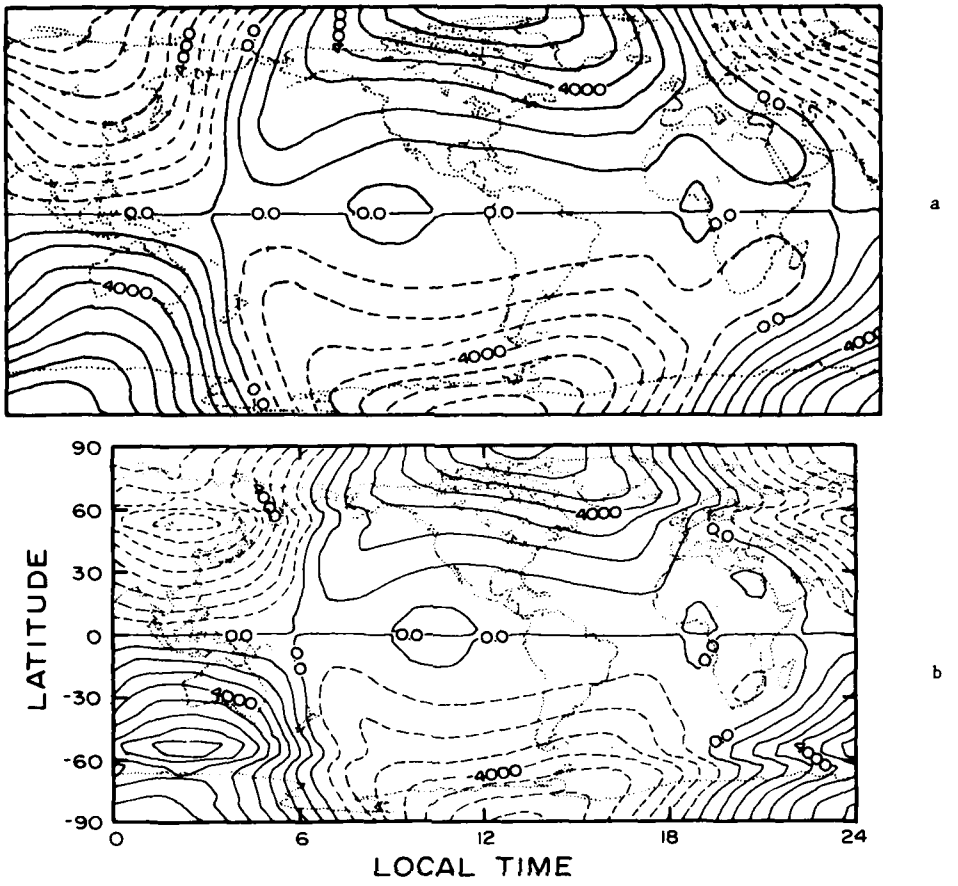


Figure 9. Global northward wind velocity (cm sec^{-1}) at 250 km for (a) solar heating only (b) including the energetic oxygen heating sources corresponding to 0.04 ergs $cm^{-2}sec^{-1}sr^{-1}$.

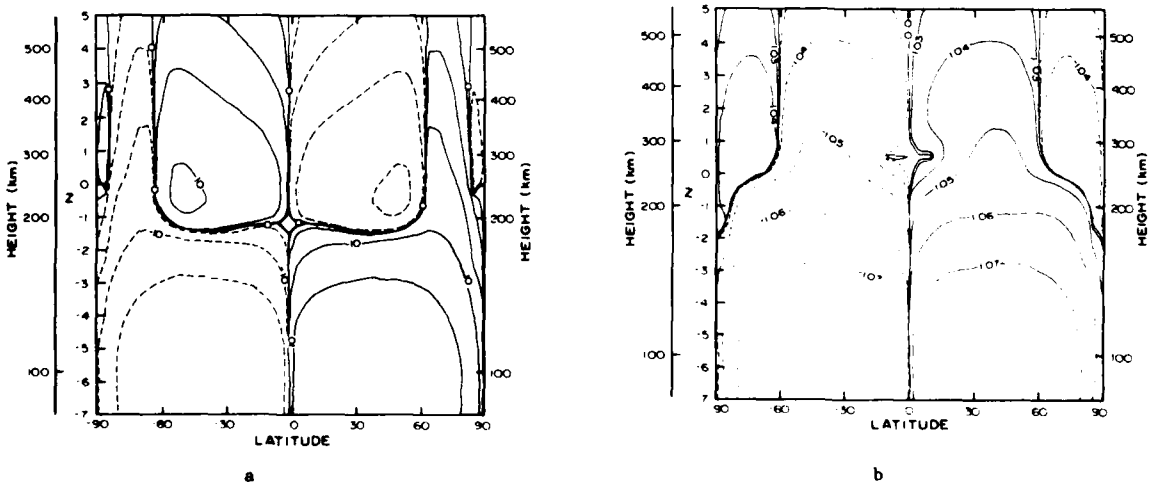
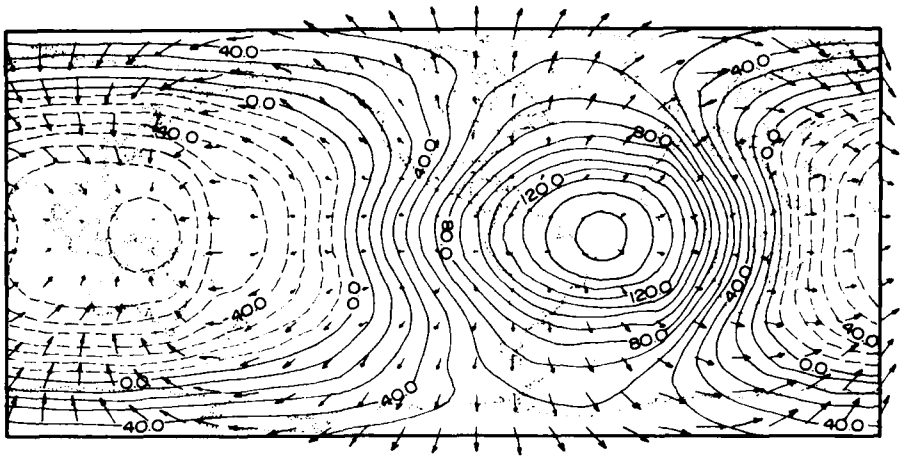
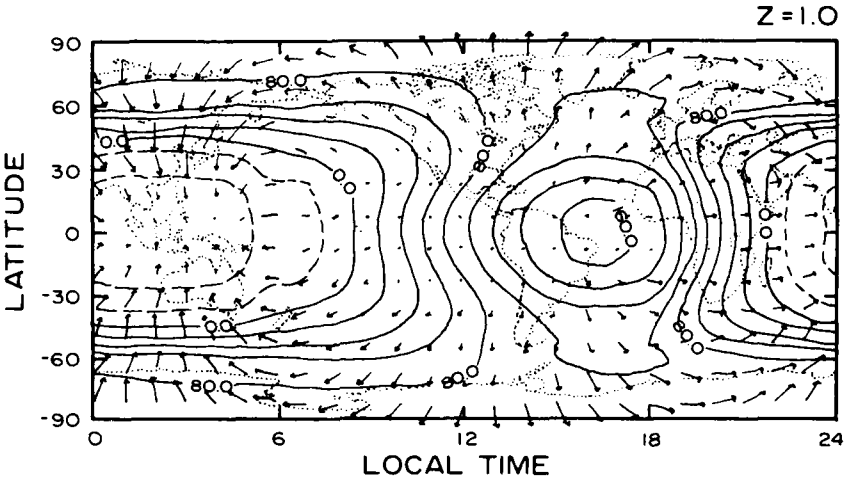


Figure 10. Zonal mass flow function (gm sec^{-1}) at 250 km for (a) solar heating only and (b) including the energetic oxygen heating source corresponding to 0.04 ergs $cm^{-2}sec^{-1}sr^{-1}$.



a



b

Figure 11. Perturbation to the global mean temperature at 250 km for (a) solar heating only and (b) including the energetic oxygen heating source corresponding to 0.04 ergs $cm^{-2}sec^{-1}sr^{-1}$.

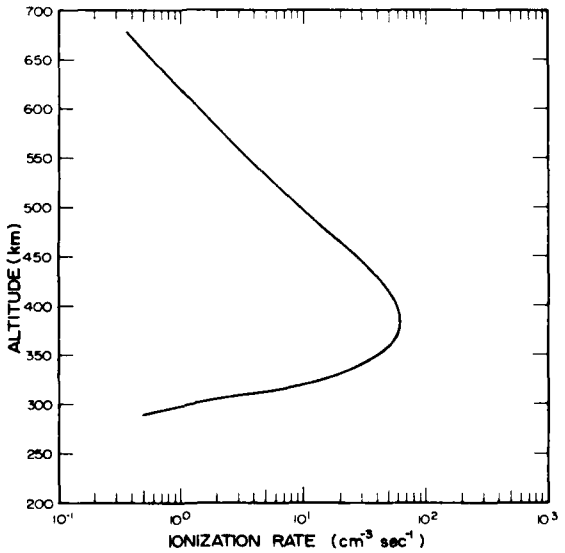


Figure 12. Ionization profile resulting from O-O collisions produced by O^{+} spectrum for L ∿ 3.4 in

ON THE MORPHOLOGY OF THE POLAR THERMOSPHERE

G. W. Prölss
Institut für Astrophysik und extraterrestrische Forschung
Auf dem Hügel 71, D 5300 Bonn, F.R. Germany

ABSTRACT

The dissipation of solar wind energy in the upper atmosphere generates a disturbance whose morphology depends on local time, season, and magnetic activity. A statistical analysis of ESRO 4 data reveals that (1) in the afternoon/evening sector the disturbance boundary coincides with the region of electric current dissipation along the auroral oval; (2) in the midnight/early morning sector dynamical effects extend the disturbance zone to lower latitudes and this expansion is strongly dependent on season and magnetic activity; and (3) in the late morning sector direct heating effects along the auroral oval are superimposed on the residuals of the early morning disturbance. These results are consistent with previous observations and provide new boundary conditions for upper atmospheric models.

1. INTRODUCTION

There are two principle modes by which solar energy is brought into the earth's upper atmosphere: (1) direct dissipation of solar radiation energy; (2) indirect injection of solar wind kinetic energy via the magnetosphere (Fig. 1). Whereas the first mode has long been recognized to be the dominant one, the significance of the solar wind/magnetospheric energy source is more and more appreciated. This study is concerned with the imprint the solar wind energy dissipation leaves in the upper atmosphere. Specifically, the latitudinal structure and extension of the thermospheric disturbance zone generated by this energy source is investigated using density data obtained by the polar orbiting satellite ESRO 4 (von Zahn, 1975).

2. DATA ANALYSIS AND RESULTS

The basic latitudinal structure of a thermospheric perturbation observed on a winter afternoon during weakly disturbed conditions is illustrated in Fig. 2. Relative variations in the molecular nitrogen density, in the atomic oxygen density, and in the helium density as measured above the Asian continent are plotted. The density data have been adjusted to a common altitude of 260 km using standard hydrostatic techniques. In addition they have been normalized to suitable density values observed outside the disturbance region (for example, R_O = ([O]260 km)observed/([O]260 km)normalization value, with R_O = 1 serving as a reference value). Note that whereas molecular nitrogen is plotted on a linear scale, the oxygen and helium data are presented on a logarithmic scale. For comparison, the upper panel also shows the exospheric temperature inferred from the nitrogen data.

The essential feature illustrated in Fig. 2 is a localized disturbance zone at high invariant latitudes. It is characterized by a fairly abrupt increase in the nitrogen density (and accordingly in the inferred gas temperature), a small decrease in the oxygen density, and a significant V-shaped drop in the helium density. Theoretical studies (e.g. Mayr and Volland, 1972; Shimazaki, 1972; Hays et al., 1973; Bates, 1974) indicate that these features are the typical thermospheric signature of energy deposition in the lower thermosphere. The most effective energy source at these heights and latitudes are Pedersen currents which heat the ambient neutral gas through Joule dissipation (Cole, 1975; Banks, 1977). This is schematically illustrated in Fig. 3. Whereas it is difficult to measure the Pedersen currents (I_P) directly, a sizeable set of data is available on the magnetic signature of field-aligned Birkeland currents (I_B) which close the current circuit in the magnetosphere. Therefore, to support our interpretation the location of the equatorward portion of the field-aligned currents (= region 2 currents according to Iijima and Potemra, 1976) has been compared to the location of the equatorward boundary of the atmospheric disturbance zone. Figure 4 demonstrates that in the afternoon local time sector (14-16 MLT) and during moderately disturbed conditions (AE $\leq$ 300 nT for at least 4 hours prior to measurement) both boundaries are approximately collocated. We conclude that the atmospheric disturbances observed at high latitudes are a direct signature of electric field and current dissipation by the magnetospheric energy source.

A very different situation is encountered in the early morning sector where the disturbance boundary is located at middle latitudes and far beyond the electric current dissipation region (Fig. 4). It is suggested that the broad disturbance zone observed at these times and illustrated in Fig. 5 is generated by strong winds blowing away from the polar regions and carrying along heated air of perturbed composition. This interpretation is supported by measurements which establish the existence of strong equatorward directed winds commencing around midnight (e.g. Bates and Roberts, 1977; Babcock and Evans, 1979). Whereas at high latitudes this "midnight surge" is observed even during moderately disturbed conditions (e.g. Hays et al., 1979), at middle latitudes it is primarily recorded during magnetic storms (e.g. Hernandez and Roble, 1976). This expansion of the disturbance to lower latitudes during increased magnetic activity is also seen in the density data. Figure 6 shows the median boundary latitude for three different levels of magnetic activity and separately for summer and winter conditions. As can be seen, the boundary moves to lower latitudes with increasing disturbance intensity. This shift appears to be larger during summer than during winter. Even more impressive, however, is the large change with season. This suggests that the range of the midnight surge should also be strongly dependent on season.

Whereas the midnight surge is a transient phenomenon which may last only a few hours, the disturbance generated by it is a more persistent feature. Thus residuals of this perturbation are clearly evident in the late morning data. This is illustrated in Fig. 7, which shows the basic latitudinal structure of the atmospheric disturbance in the forenoon local time sector. Here the broad zone of enhanced argon and nitrogen densities and slightly decreased oxygen and helium densities extending all the way to middle latitudes is identified with the remnants of the midnight surge perturbation. In contrast, the additional disturbance increase observed at higher latitudes is again attributed directly to the dissipation of electric currents in this region. A comparison shows that the boundary of this high latitude disturbance is approximately collocated with region 2 Birkeland currents (see Fig. 8) and this is in agreement with our expectations.

3. SUMMARY

The local time variations in the extension of the atmospheric disturbance zone are summarized in Fig. 8. For each time interval the median value and the upper and lower quartiles of the observed boundary latitudes are given. The position of region 2 Birkeland currents also indicated in this figure serves as a reference. A comparison shows that in the afternoon/evening sector the disturbance boundary is coupled to the location of field-aligned currents and therefore to the region of electric energy dissipation along the auroral oval. This also applies to the high latitude heating zone observed during the morning hours. In the midnight/early morning sector the perturbation limit is shifted toward lower latitudes; this is attributed to the disturbing action of the midnight surge. Once disturbed the thermosphere recovers only slowly, and residuals of the nighttime perturbation are frequently seen in the late morning sector. This interpretation is schematically summarized in Fig. 9. Although not indicated in this figure, the midnight surge disturbance is strongly dependent on season and magnetic activity.

4. DISCUSSION

Previous Studies. Heating effects in the polar region during low levels of magnetic activity were first documented by Hedin and Reber (1972). A more detailed analysis of these disturbance effects was subsequently performed by Taeusch and Hinton (1975). Using statistical methods they established the main regions of molecular nitrogen enhancements in an invariant latitude/magnetic local time coordinate system. Although the averaging and normalization procedure employed precludes a more accurate determination of the disturbance zone limits, the general trend of the local time variation documented in this study (see their Fig. 4) is in good agreement with the present results. Whereas these authors do not consider seasonal changes, a subsequent study (Taeusch, 1977) indicates that an increase in magnetic activity leads primarily to an expansion of the disturbance zone in the nighttime sector, which is again in accord with the present results. There are, however, differences in the interpretation of the data. Whereas these authors also consider local heating by particle precipitation to be an important disturbance source, this study assumes that electric energy dissipation and dynamical effects are primarily responsible for the observed perturbation structure.

Variations of the atmospheric disturbance morphology with local time, season, and magnetic activity have also been reported in studies based on the analysis of individual storm events (see review by Prölss, 1980). The form of these variations is consistent with the phenomenological model (Fig. 9) suggested in the present study.

Models. A comparison of our data with empirical atmospheric models (e.g. Köhnlein, 1980, and references therein) is not meaningful at present since none of these algorithms allows for changes in the extension of the polar atmospheric disturbance. To improve these models both the distribution and intensity of the geomagnetic activity effect should be considered functions of local time, season, and disturbance intensity. Also, the time delay between a magnetic and atmospheric disturbance should be treated as a function of latitude and local time. According to our model this time lag should vary between tens of minutes at high latitudes and 6 to 12 hours at middle latitudes in the afternoon sector.

The data presented in this study should also provide new boundary conditions for theoretical models. Although the existing circulation models (e.g. Volland, 1979, and references therein) are capable of reproducing certain aspects of the disturbance morphology, the results obtained are not sufficient for a more detailed comparison.

Ionospheric Data. It is well known that depletions of the ionospheric plasma density (negative ionospheric storms) during magnetically active periods are caused by changes in the neutral gas composition. Given this cause-effect relationship, the ionospheric storm morphology should reflect the properties of the atmospheric disturbance. This also implies that any model of the disturbed upper atmosphere should be consistent with ionospheric storm data. For our model this is indeed the case. As detailed in a recent review (Prölss, 1980), the latitudinal extent of negative ionospheric storms exhibits significant variations with local time, season, and disturbance intensity, and these changes are consistent with those predicted here. Indeed, some aspects of our model (e.g. perturbation of the midlatitude atmosphere during nighttime and subsequent recovery during the daytime) were first discussed in connection with ionospheric storm data (King, 1967).

However, there are also points of disagreement. Thus our model is not consistent with a meridional wind circulation set up in the dayside upper atmosphere as has been proposed in a number of ionospheric storm models (see, for example, Park and Meng, 1976; Miller et al., 1979; Mayr et al., 1978).

Energy Considerations. Finally, we note that the midnight surge causing the early morning disturbance should carry a considerable amount of energy towards lower latitudes. Thus our model implies a significant energy coupling between the polar and the midlatitude region even during moderately disturbed conditions. This also suggests that the thermospheric energy budget at middle latitudes is not only controlled by solar UV radiation but also by solar wind/magnetospheric energy dissipation. The timing of the additional energy influx is such that it should induce a strong semidiurnal variation. A more detailed discussion of these energy balance considerations is given, for example, in the work of Glenar et al. (1978).

ACKNOWLEDGMENTS

I am very grateful to U. von Zahn who kindly provided all the neutral composition data used in this study. Thanks are also due to K.H. Fricke for helpful discussions and to M. Roemer for his continued support. This research was financed by the Deutsche Forschungsgemeinschaft under grant Ro 446/3.

5. REFERENCES

Babcock, R.R., and J.V. Evans, 1979, "Effects of geomagnetic disturbances on neutral winds and temperatures in the thermosphere observed over Millstone Hill", J. Geophys. Res., 84, 5349 - 5354.

Banks, P.M., 1977, "Observations of joule and particle heating in the auroral zone", J. Atmos. Terr. Phys., 39, 179 - 193.

Bates, H.F., 1974, "Thermospheric changes shortly after the onset of daytime joule heating", Planet. Space Sci., 22, 1625 - 1636.

Bates, F.B., and T.D. Roberts, 1977, "The southward midnight surge in F-layer wind observed with the Chatanika incoherent scatter radar", J. Atmos. Terr. Phys., 39, 87 - 93.

Cole, K.D., 1975, "Energy deposition in the thermosphere caused by the solar wind", J. Atmos. Terr. Phys., 37, 939 - 949.

Glenar, D.A., E. Bleuler, and J.S. Nisbet, 1978, "The energy balance of the nighttime thermosphere", J. Geophys. Res., 83, 5550 - 5562.

Hays, P.B., R.A. Jones, and M.H. Rees, 1973, "Auroral heating and the composition of the neutral atmosphere", Planet. Space Sci., 21, 559 - 573.

Hays, P.B., J.W. Meriwether, and R.G. Roble, 1979, "Nighttime thermospheric winds at high latitudes", J. Geophys. Res., 84, 1905 - 1913.

Hedin, A.E., and C.A. Reber, 1972, "Longitudinal variations of thermospheric composition indicating magnetic control of polar heat input", J. Geophys. Res., 77, 2871 - 2879.

Hernandez, G., and R.G. Roble, 1976, "Direct measurements of nighttime thermospheric winds and temperatures, 2. Geomagnetic storms", J. Geophys. Res., 81, 5173 - 5181.

Iijima, T., and T.A. Potemra, 1976, "The amplitude distribution of field-aligned currents at northern high latitudes observed by Triad", J. Geophys. Res., 81, 2165 - 2174.

King, G.A.M., 1967, "The ionospheric disturbance and atmospheric waves II - The F-region at Campbell Island", J. Atmos. Terr. Phys., 29, 161 - 168.

Köhnlein, W., 1980, "A model of thermospheric temperature and composition", Planet. Space Sci., 28, 225 - 243.

Mayr, H.G., and H. Volland, 1972, "Magnetic storm effects in the neutral composition", Planet. Space Sci., 20, 379 - 393.

Mayr, H.G., I. Harris, and N.W. Spencer, 1978, "Some properties of upper atmospheric dynamics", Rev. Geophys. Space Phys., 16, 539 - 565.

Miller, N.J., J.M. Grebowsky, H.G. Mayr, and I. Harris, 1979, "F layer positive response to a geomagnetic storm - June 1972", J. Geophys. Res., 84, 6493 - 6500.

Park, C.G., and C.-I. Meng, 1976, "Aftereffects of isolated magnetospheric substorm activity on the mid-latitude ionosphere: Localized depression in F layer densities", J. Geophys. Res., 81, 4571 - 4578.

Prölss, G.W., 1980, "Magnetic storm associated perturbations of the upper atmosphere: Recent results obtained by satellite-borne gas analyzers", Rev. Geophys. Space Phys., 18, 183 - 202.

Shimazaki, T., 1972, "Effects of vertical mass motions on the composition structure in the thermosphere", Space Res., 12, 1039 - 1045.

Taeusch, D.R., 1977, "Structure of electrodynamic and particle heating in the disturbed polar thermosphere", J. Geophys. Res., 82, 455 - 460.

Taeusch, D.R., and B.B. Hinton, 1975, "Structure of electrodynamic and particle heating in the undisturbed polar thermosphere", J. Geophys. Res., 80, 4346 - 4350.

Volland, H., 1979, "Magnetospheric electric fields and currents and their influence on large scale thermospheric circulation and composition", J. Atmos. Terr. Phys., 41, 853 - 866.

von Zahn, U., 1975, "Early aeronomy results from the satellite Esro 4", in Atmospheres of Earth and the Planets, B.M. McCormac (ed.), Reidel Publishing Co., Dordrecht, Holland, 133 - 157.

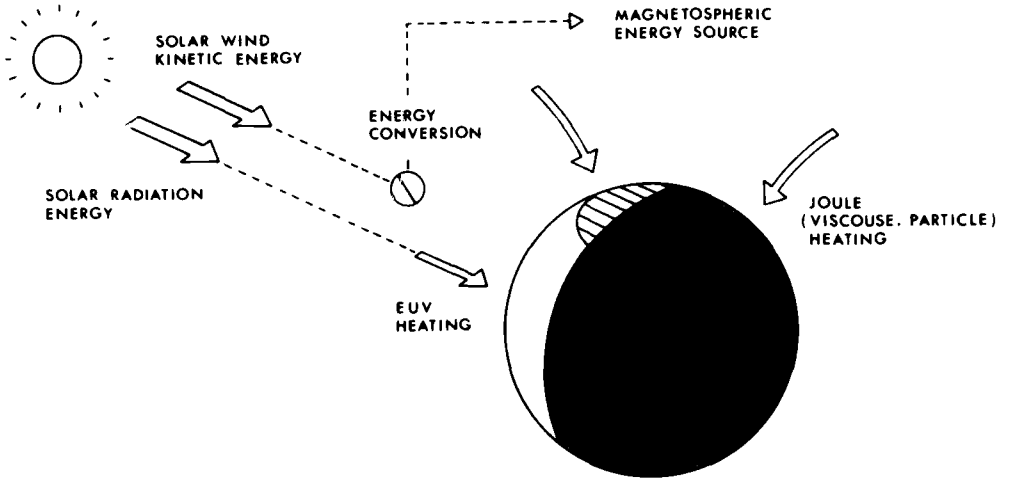


Fig. 1 Principle modes of solar energy transfer to the terrestrial upper atmosphere.

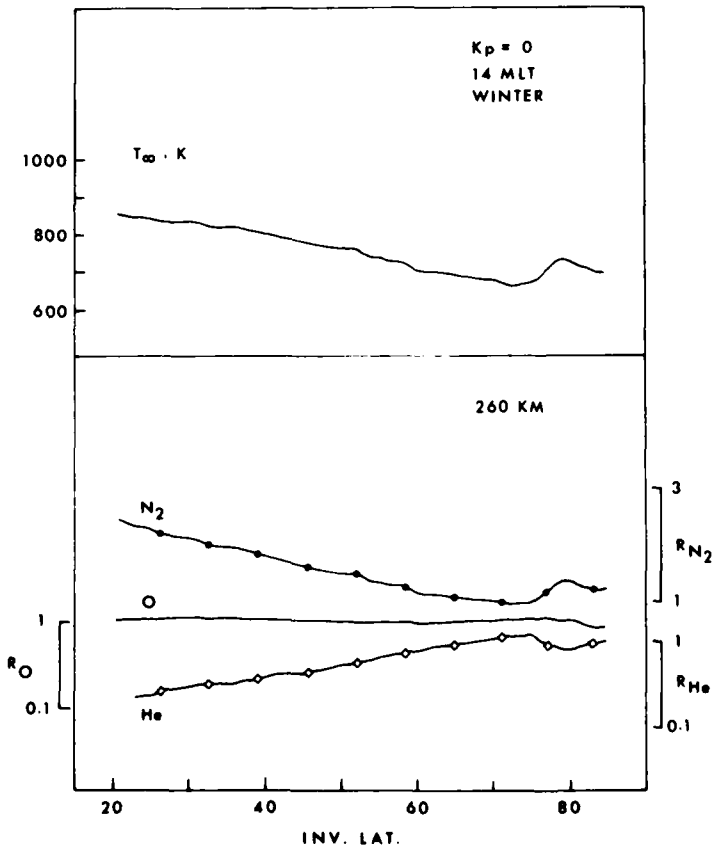


Fig. 2 Latitudinal structure of polar atmospheric disturbance in afternoon sector. The data were obtained in the northern hemisphere between 5:38 and 5:54 UT on 9 December 1972. At 70^0 invariant latitude the geographic longitude and solar local time were 148^0 E and 15:42, respectively.

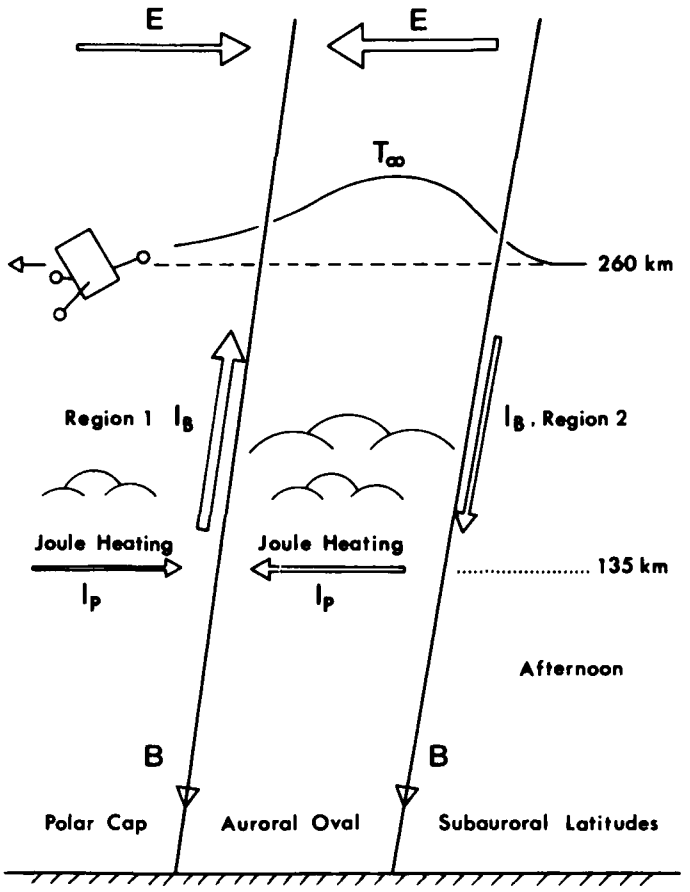


Fig. 3 Spatial association between polar atmospheric disturbance and magnetospheric processes in the afternoon sector. High-latitude electric fields (E) drive Pedersen currents (I_P) in the ionospheric E-region which heat the ambient gas. The resulting increase in gas temperature is observed by a satellite passing over this region. The current circuit closes in the magnetosphere via Birkeland currents (I_B) which flow along the magnetic field lines (B).

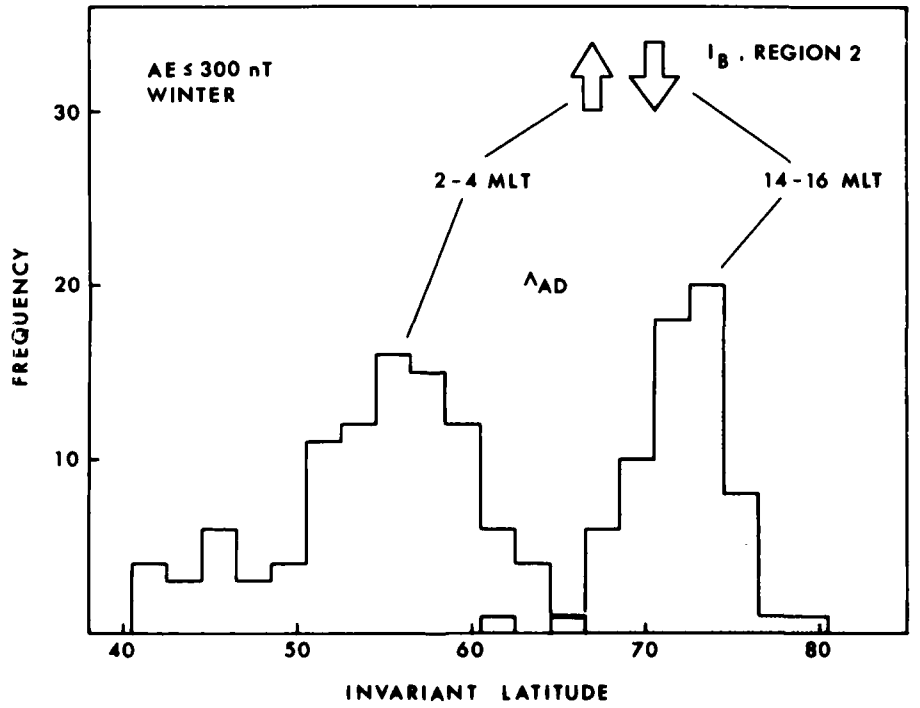


Fig. 4 Comparison between location of Birkeland currents and atmospheric disturbance boundary for two different local time sectors. Arrows indicate statistical position of region 2 Birkeland currents (Iijima and Potemra, 1976) and histograms the latitudinal distribution of the observed atmospheric disturbance boundaries (Λ_{AD}). All data were obtained during moderately disturbed conditions (current data: $2_- \leq Kp \leq 4_+$ at time of observation; atmospheric data: AE $\leq$ 300 nT for at least 4 hours prior to observation). The atmospheric data refer to winter conditions.

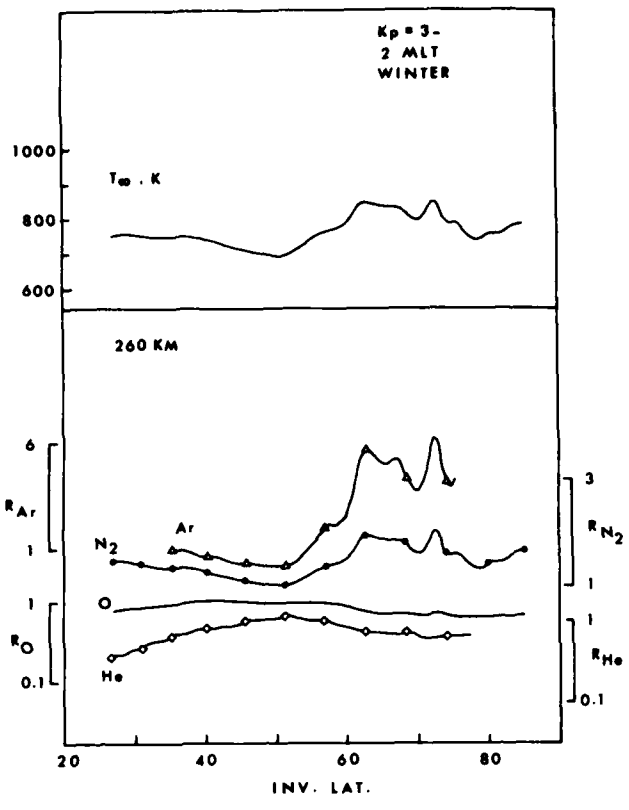


Fig. 5 Latitudinal structure of atmospheric disturbance in the early morning sector. The data were obtained in the northern hemisphere between 5:57 and 6:14 UT on 30 December 1972. At 50^0 invariant latitude the geographic longitude and solar local time were 299^0 E and 1:59, respectively.

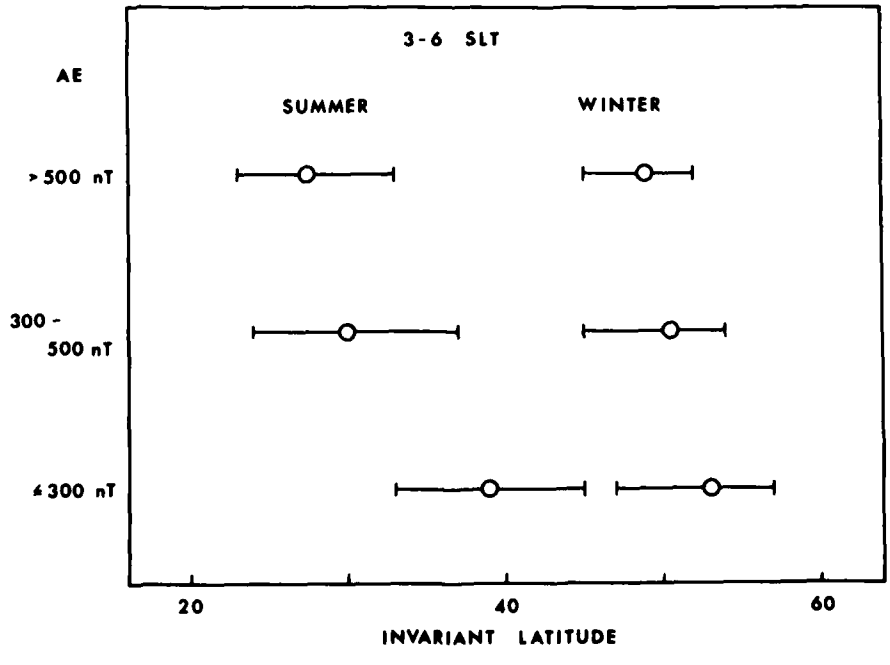


Fig. 6 Extension of atmospheric disturbance in the early morning sector. Boundary location data collected in the 3 to 6^h solar local time sector were divided according to season and disturbance intensity. The activity level was defined by the maximum hourly AE index during the 4-hour interval prior to the measurement. The median and the upper and lower quartiles have been computed for each set (containing a minimum of 55 data points) and are shown in the figure.

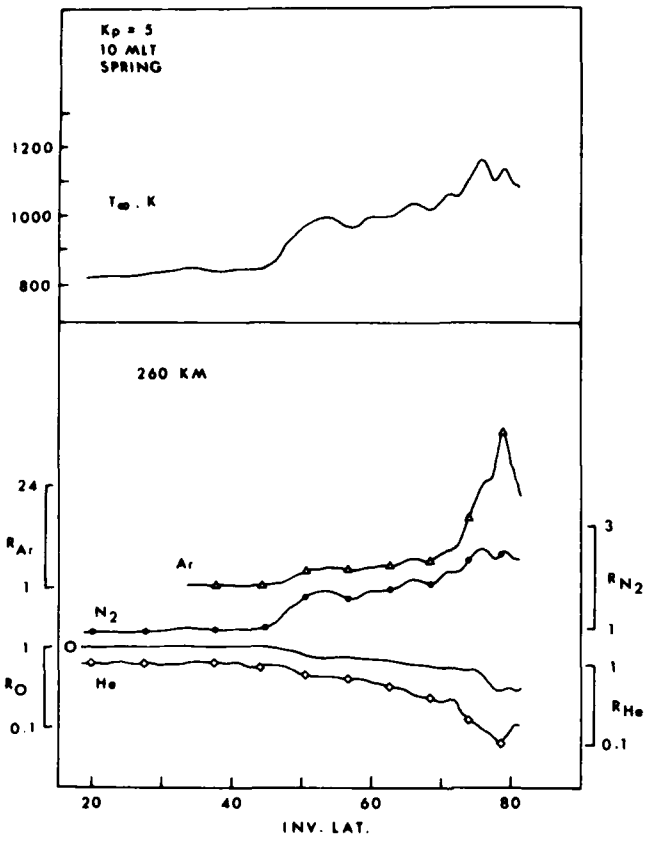


Fig. 7 Latitudinal structure of atmospheric disturbance in the late morning sector. The data were obtained in the northern hemisphere between 7:03 and 7:21 UT on 19 March 1973. At 40^0 invariant latitude the geographic longitude and local solar time were 37^0 E and 9:33, respectively.

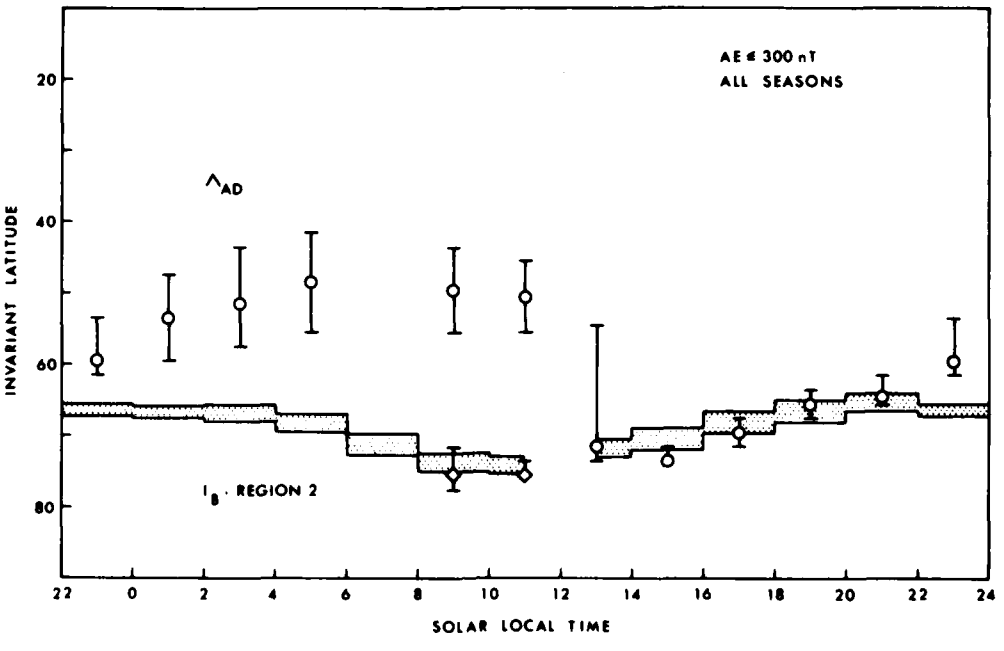


Fig. 8 Local time variations in the extension of the atmospheric disturbance zone. For each 2-hour interval the median and the upper and lower quartiles of the observed boundary distribution are given. The limit of the high latitude disturbance observed in the morning sector is also indicated for the 8 to 10^h and 10 to 12^h intervals. All data refer to moderately disturbed conditions (AE $\gtrsim$ 300 nT for at least 4 hours prior to measurement) and all seasons. For comparison the dotted area indicates the position of region 2 Birkeland currents (Iijima and Potemra, 1976). These data also apply to moderately active periods ($2_- \leq Kp \leq 4_+$) and all seasons.

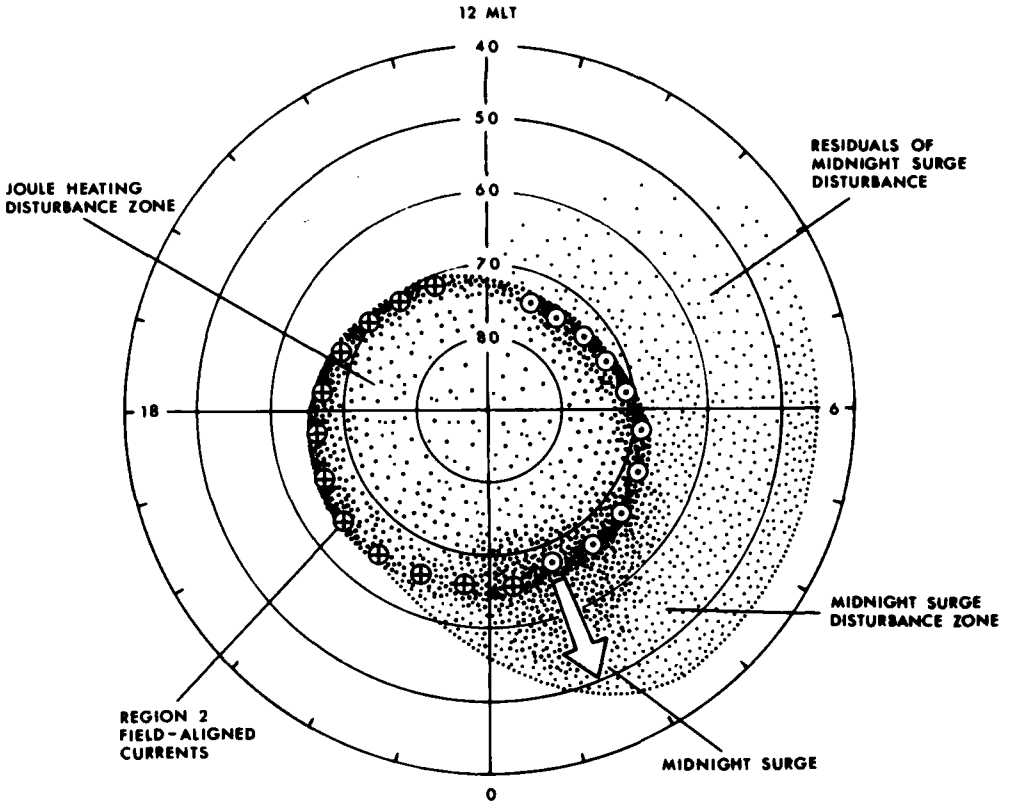


Fig. 9 Phenomenological interpretation of atmospheric disturbance morphology. The polar coordinates are invariant latitude and magnetic local time. The dotted area schematically indicates the atmospheric disturbance zone, and circles mark the position of region 2 Birkeland currents flowing into and out of the ionosphere. An arrow symbolizes the disturbing action of the midnight surge.

STRUCTURE OF ELECTRON TEMPERATURE

H. Thiemann

Rotlaubstrasse 20 D-7800 Freiburg

SUMMARY

Electron density and -temperature measurements of the Retarding Potential Analyzer on board the AEROS-B satellite are analyzed in an altitude range from about 300-700 km for daytime conditions. The average temperature always displays the same qualitative pattern as function of electron density and latitude for fixed altitudes. A model function describing this behaviour is given. By means of this function the temperature response of the corresponding experimental data on seasonal as well as on hemispherical effects is discussed.

1. INTRODUCTION

With the daytime data of electron temperature experiments of different satellites, there have been proposed several electron temperature models. Although the T_e-data of satellites - such as AEROS-A and B, TAIYO, AE-C - show a lot of variety, the individually measured temperature values could not yet be directly related to other parameters such as geomagnetic longitude and latitude, electron density, season, local time, magnetic- or sun activity. But with averaged T_e-data functional patterns could be detected. Here we present two major results:

a, With the Langmuir probe measurements of electron temperature T_e and ion density N_i on Atmosphere Explorer C there has been fashioned a model of their relationship and its variations with altitude in the daytime, non auroral ionosphere. (BRACE, H. et al., 1978.) Fifty equally spaced bins covered the N_i- (10^4-10^7/cc) as well as the height - (130-400 km) range. The model, using the averaged T_e- values within each bin, has the following major features, that (1) T_e is independent of N_i at altitudes below 200 km and (2) T_e varies inversely with N_i above perhaps 250 km. Data above 400 km were excluded, because of expected effects of heat conduction on T_e , that may introduce increasingly larger T_e-variations at higher altitudes.

b, The measurements of the planar Retarding Potential Analyzer on board the AEROS-A satellite formed the input data for an empirical model function of the global electron temperature distribution. (SPENNER, K. et al., 1979.) For 1500 LT and altitudes between 300 and 700 km all data were grouped in ranges of 10^o geomagnetic latitude, 60^o longitude, 100 km altitude and averaged within each cell. The model shows the pronounced latitudinal distribution as derived by other authors with a valley region around the equator. In addition a minor longitudinal effect occured, which might be not so significant because of the averaging procedure.

The purpose of this paper is to apply both results on the RPA data of AEROS-B and to give an empirical model function for fixed altitudes between 300 and 700 km.

2. CHARACTERISTICS OF THE ELECTRON TEMPERATURE

From summer 1974 till autumn 1975 the AEROS-B satellite was sun-synchronous in polar orbit. At low and middle latitudes the local time was almost constant around 1530 at the dayside. The same LT-criterion was applied for high latitudes. For altitudes between 300 and 700 km the data confirm the just noted T_e variations. To demonstrate this we used a representative height of h=500 km. Figure 1a shows all T_e data in relation to the logarithm of the electron density N_e with the inverse relationship between both. The latitudinal effect is presented in Figure 1b. As latitudinal coordinate we chose the modified dip (Modip) coordinate μ, which combines the magnetic dip ψ and the geographic latitude θ . (RAWER, K., 1963.) This particular coordinate is defined by

$$\tan \mu = \frac{\psi}{\sqrt{\cos \theta}}$$

It could be shown that Modip can be regarded as an appropriate latitudinal coordinate for describing plasma parameters in the ionosphere (SPENNER, K. et al., 1977). Modip essentially equals dip in the equatorial region and approaches the geographic latitude near the geographic poles. Figure 1b exhibits the latitudial temperature structure with low T_e-values around the equator and a more complicated structure for higher Modip values.

Besides the general tendency in both figures the electron temperature exhibits a remarkable scattering of the single T_e values. Till now we could not find any resolution of this scattering in terms of geophysical indices or other parameters. Therefore we investigated the statistical behaviour of T_e. To implement this, we divided the $\log_{10}(N_e)$-μ-h space into a grid system with

$10.6 \leq \log_{10}(N_e) \leq 12.0$	with an increment	$\Delta(\log_{10}(N_e))$	= .1
$0 \leq \mu \leq 70$		$\Delta \mu$	= 5
$260 \leq h \leq 740$		Δh	= 20

The daytime data of the AEROS-B mission are classified after these criterions. Within each cell the points are averaged to form the input data of the following diagrams, where we get vacant squares if no experimental data are available. Figure 2a and 2b illustrate the statistical relationship between T_e and N_e over the whole height range for a fixed Modip

coordinate. The temperature ist displayed once from the high density (Figure 2a) and then from the low density side (Figure 2b). The general inverse relationship between T_e and N_e holds true over the whole height range. As seen in both case for low and high densities the electron temperature becomes nearly independent from N_e. Figures 3a and 3b present the Modip dependency of T_e for a fixed electron density. Again the qualitative behaviour is seen for all altitudes, i.e. a T_e-valley region around the equator (compare Figure 3a from the low Modip side), a maximum of the electron temperature at Modip values around 50° and a decrease of T_e for high latitudes (compare Figure 3b from the high Modip side). Even for high latitudes the electron temperature is statistically quite well behaved. As for the temperature alteration with respect to height we cannot see a very significant structure except a slight increase of T_e with height. The T_e-values in the previous plots that don't fit very well into the general pattern, are almost characterized by a low statistical weight. Figures 4a and 4b in comparison to Figures 2a and 3a with the same independent variables show that if we exclude there averaged T_e-values with less than four contributions, we get a more harmonic pattern.

With the very pronounced temperature response on density and latitude variations we now concentrated on the T_e behaviour for fixed heights. Figures 5a-c show the experimental daytime behaviour of T_e ($\log_{10}(N_e)$, Modip) for the northern hemisphere and h=300, 500 and 700 km respectively. One main effect of the height influence is seen in shifting the location of the steepest temperature gradient in the Modip - as well as in the density direction. Another effect consists in the decrease of the electron density with height and therefore especially for low and middle latitudes an increase of T_e.

In order to get quantitative results we modelled the statistical T_e values by the following model function which reproduces for fixed altitudes the superposition of the $\log_{10}(N_e)$ and μ influences as demonstrated in Figure 1 to Figure 5.

$$T_e(\log_{10}(N_e),\ \mu,\ h=\text{const}) = \left[P_1 - P_2 \cdot \cos(P_3 \cdot \mu^2 + P_4 \cdot \mu)\right] \cdot \left[\frac{P_5}{1+\left(\frac{\log_{10}(N_e)}{P_6}\right)^{P_7}} + 1\right]$$

where T_e is in units of ^oK
N_e is in units of m^{-3}
μ is in units of radians.

For specific heights between 300 and 700 km the parameter P_1, ... P_7 are obtained by a non linear least squares method applied on the averages of T_e as received from the experimental data for the northern hemisphere. The effect of possible runaways has been reduced bei excluding averaged T_e-values with less than four contributions. A plot of the model function with the appropriate parameters at h=500 km is shown in Figure 5d and demonstrates the satisfying correspondence to the averaged experimental data of Figure 5b. The parameters P_1 ... P_7 for h=300, 500, 700 km and the whole set of data are presented in the following table:

h/km	P_1	P_2	P_3	P_4	P_5	P_6	P_7	
300	1375	133	5.6	-1.3	.58	11.74	153	
500	1327	249	1.1	2.8	.69	11.34	88	$\mu > o$
700	1572	226	2.1	5.4	.49	11.08	120	

For large electron densities, $P_1 - P_2$ and $P_1 + P_2$ are the lowest and highest temperature values in the μ-space where we simulated the behaviour by a stretching argument by means of P_3 and P_4. For $\log_{10}(N_e)$ near or lower than P_6 the second term of our model function becomes more and more influential and reaches finally $(P_5 + 1)$ for $\log(N_e) < P_6$. The last parameter P_7 characterizes the strength of the transition from low to high densities.

We applied now our optimizing procedure on the experimental data, which we separated into two seasons (April-October and November-March) without restriction of the year. The diagrams of Figures 6a and 6b show still the basic pattern. However we see a seasonal effect with in general higher temperatures in the summer than in the winter period. For the case of h=300 and 500 km we got the following parameters for the summer-winter behaviour of T_e:

h/km	season	P_1	P_2	P_3	P_4	P_5	P_6	P_7	
300	su	1440	103	8.9	-4.7	.58	11.75	286	
300	wi	802	92	4.2	.3	1.78	11.86	43	
500	su	1427	276	1.6	1.9	.62	11.31	82	$\mu > o$
500	wi	1226	265	2.1	2.4	.92	11.32	42	

For both altitudes the two seasons show remarkable differences in T_e with higher temperatures in summer than winter if the electron density is sufficiently high ($\log(N_e) \gg 11.8$ for 300 km and $\log(N_e) \gg 11.3$ for 500 km). But with decreasing N_e, P_5 becomes dominant and gradually the seasonal effect is cancelled by the density influence. $\left(P_5^{winter} > P_5^{summer}\right)$

For the same case we determined the parameters of the southern hemisphere ($\mu < o$):

h/km	season	P_1	P_2	P_3	P_4	P_5	P_6	P_7
300	su	1218	148	-.8	4.3	.59	11.75	67
300	wi	633	116	-.8	2.7	3.74	11.70	29
500	su	1304	254	-.5	3.4	1.14	11.24	68
500	wi	1302	283	1.6	1.4	.76	11.32	61

$\mu < o$

One striking point is that there is no symmetry between north and south. Moreover, while we have a very strong density influence for h=300 km in the winter period ($P_5 = 3.74$) this behaviour changes for h=500 km with $P_5^{summer} > P_5^{winter}$ just in contrary to the northern hemisphere.

3. SUMMARY

The major T_e-features of the experimental daytime data of the AEROS-B Retarding Potential Analyzer are the pronounced electron density- and latitude dependencies over a height range from 300-700 km. The latitudinal distribution starting with low Modip values is given by an initially slower then stronger increase of T_e which reaches a maximum at about $\mu \approx 50^o$. For higher Modip values T_e is decreasing. The relationship between T_e and N_e is almost inverse and can be expressed by a smoothed step transition. This general behaviour was used to define a model function T_e ($\log_{10}(N_e)$, μ, h=const) that yields its describing parameters after optimizing the corresponding data. With the separation of the data for example in both hemispheres and seasons we get more information about the T_e response in the specific case. Thus we notice that the latitudical north - south behaviour of T_e is not symmetric and that the seasonal effect in both hemispheres shows different features, especially for h=500 km the temperature displays an inverse reaction on density changes during the seasons. In general this model function is suitable for investigating experimental electron temperature data in view of the different statistical connections of T_e with other parameters such as seasons, local time, magnetic or sun activity.

REFERENCES

BRACE, L.H. and R.F. Theis, 1978, "An Emperical Model of the Interrelationship of Electron Temperature and Density in the Daytime Thermosphere at Solar Minimum", Geophys.Res.Letters, 5, 275-278

RAWER, K., 1963, in "Meteorological and Astronomical Influences on Radio Wave Propagation", Pergamon, Oxford, 221

SPENNER, K. and K. Rawer, 1977, "Relationship between Plasma Density, Temperature and Suprathermal Electron Flux in the Topside Ionosphere", Cospar Space Research XVII, 445-449

SPENNER, K. and R.Plugge, 1979, "Empirical Model of Global Electron Temperature Distribution Between 300 and 700 km Based on Data from AEROS-A", J.Geophys. 46, 43-56

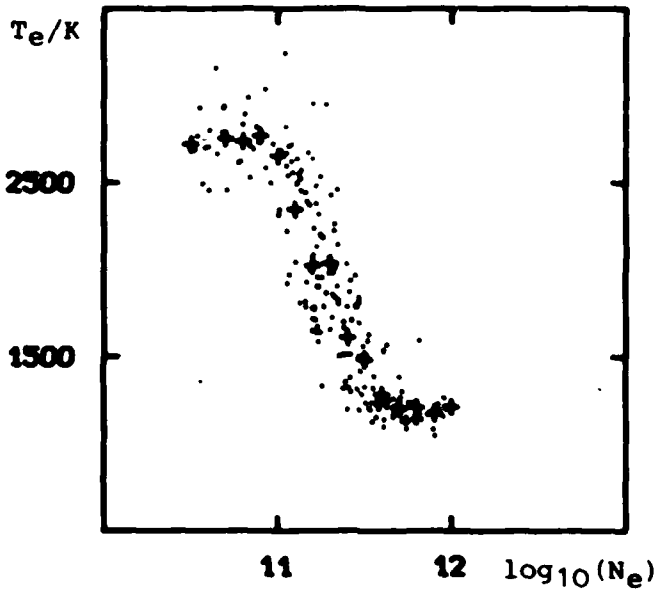


Fig.1(a) T_e ($\log_{10}(N_e)$, h = 500)

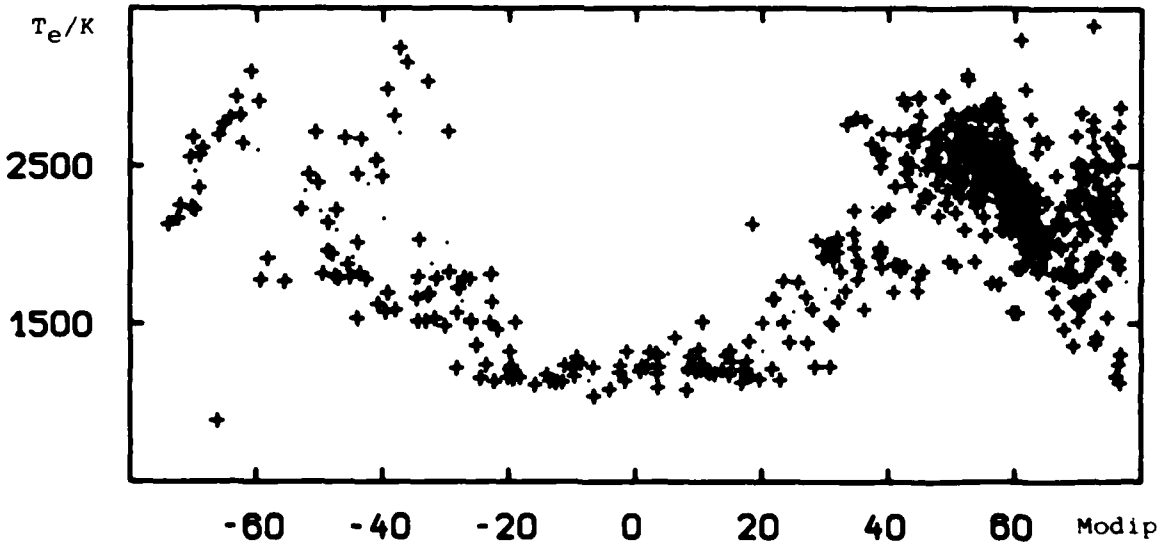


Fig.1(b) T_e (μ, h = 500)

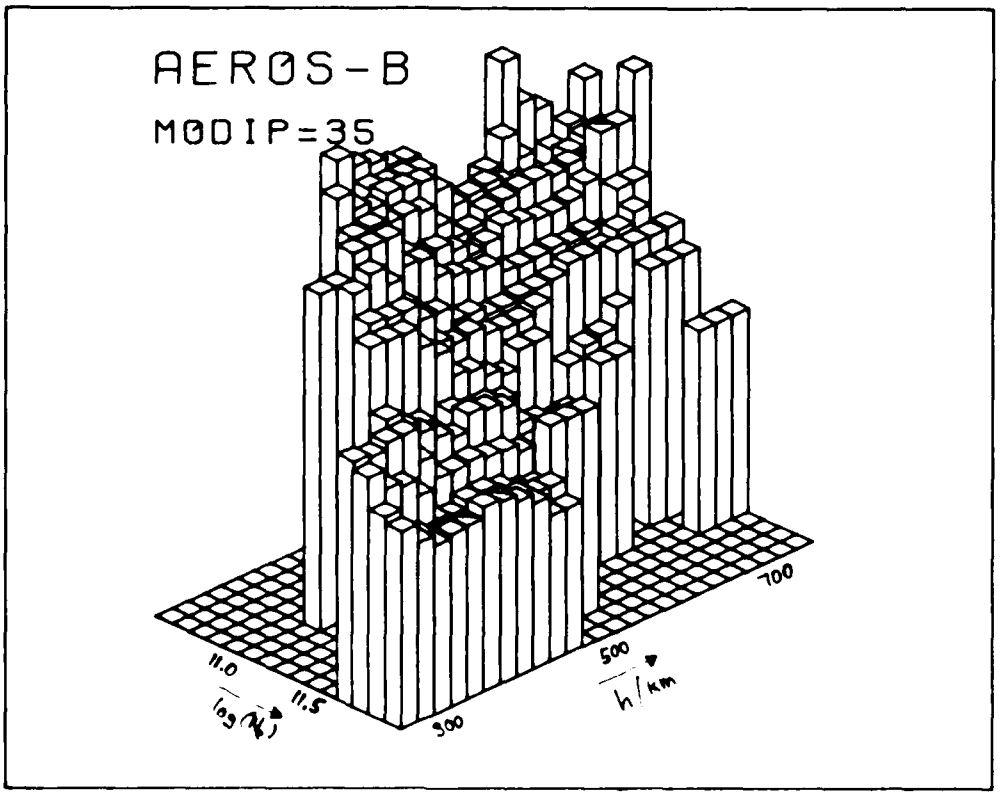


Fig.2(a) $I_e(\log_{10}(N_e), \mu, h)$ from the high density side

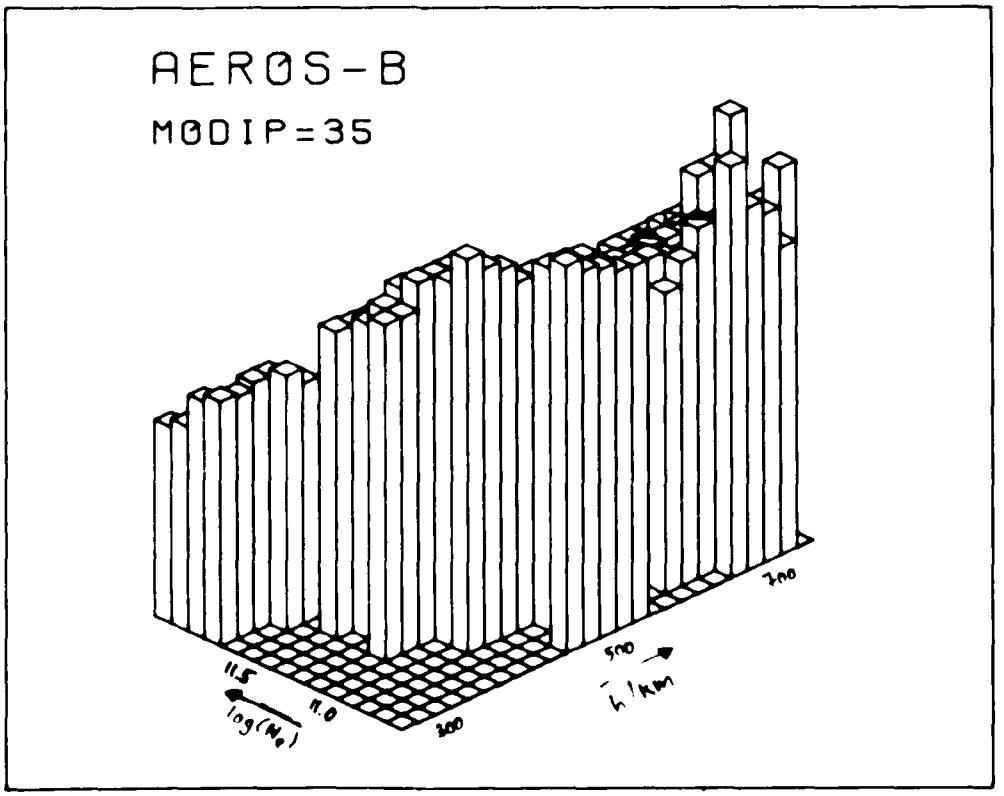


Fig.2(b) $I_e(\log_{10}(N_e), \mu, h)$ from the low density side

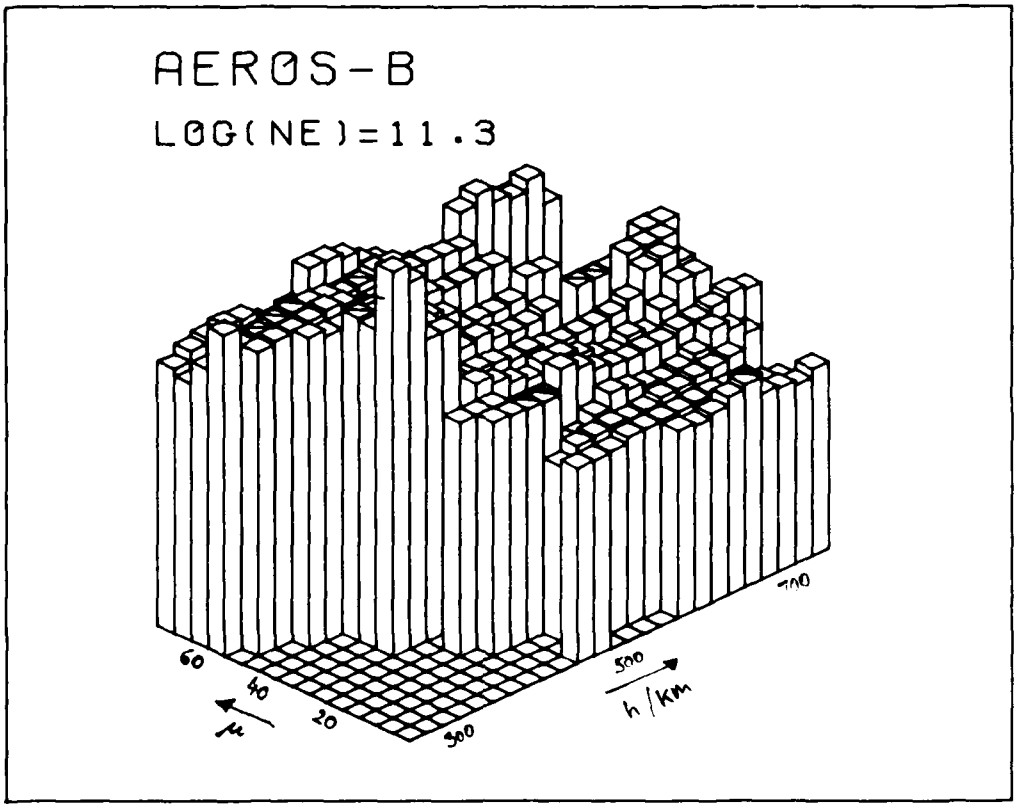


Fig.3(a) T_e ($\log_{10}(N_e) = 11.3$, μ, h) from the low Modip side

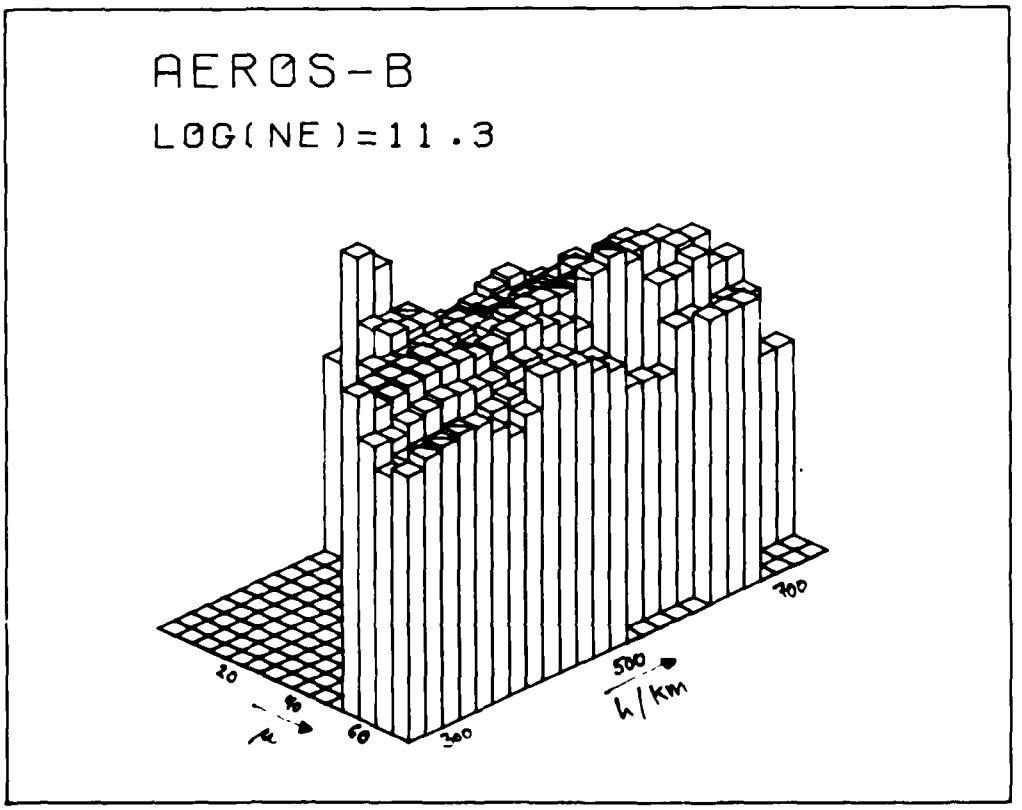


Fig.3(b) T_e ($\log_{10}(N_e) = 11.3$, μ, h) from the high Modip side

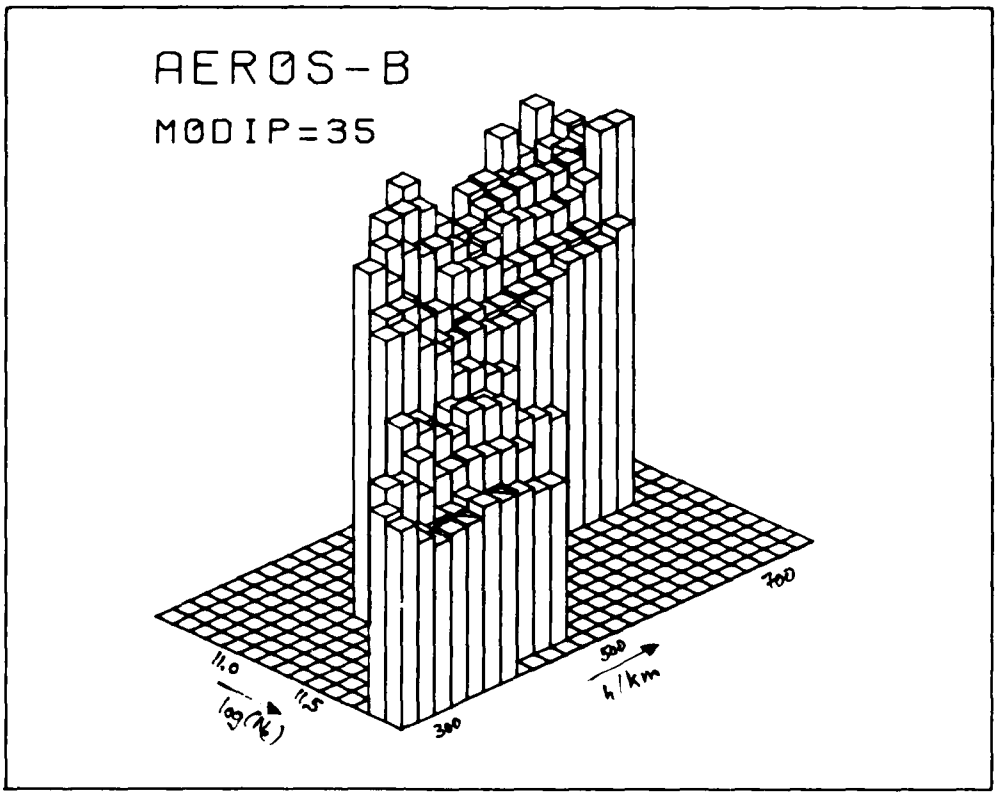


Fig.4(a) T_e ($\log_{10}(N_e)$, μ = 35, h) compare Fig.2(a) but now only T_e averages with more than three contributions

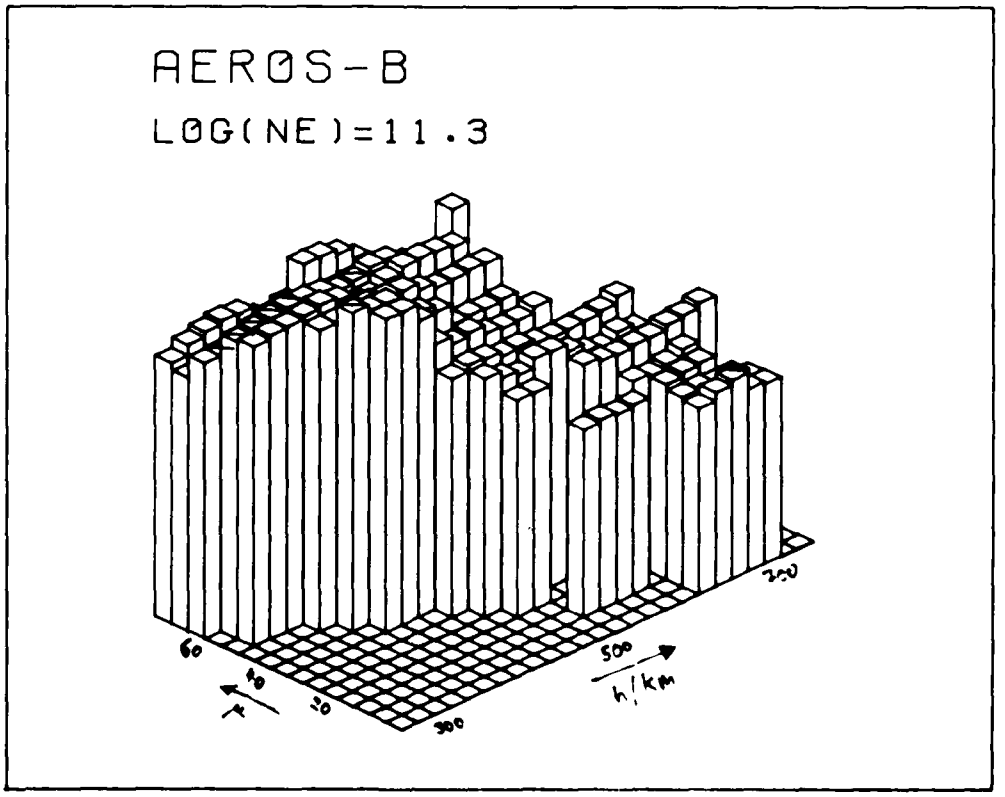
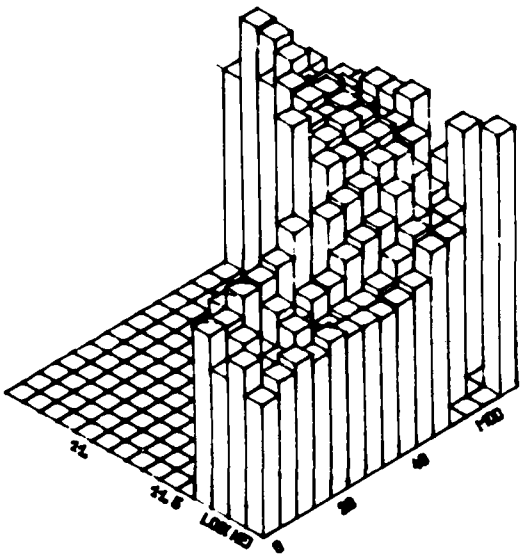


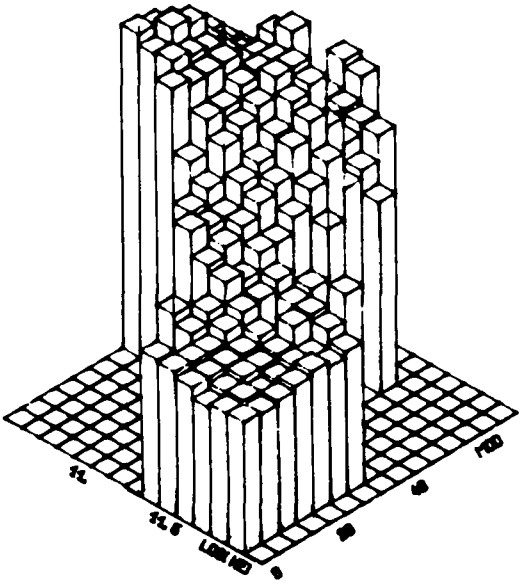
Fig.4(b) T_e ($\log_{10}(N_e)$ = 11.3, μ, h) compare Fig.3(a) but now only T_e averages with more than three contributions

AEROS-B
TE(LOG(NE) , MODIP, H=300KM)



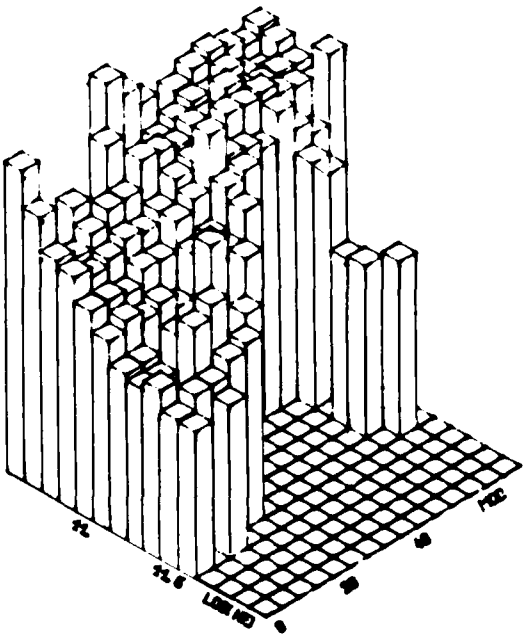
(a)

AEROS-B
TE(LOG(NE) , MODIP, H=500KM)



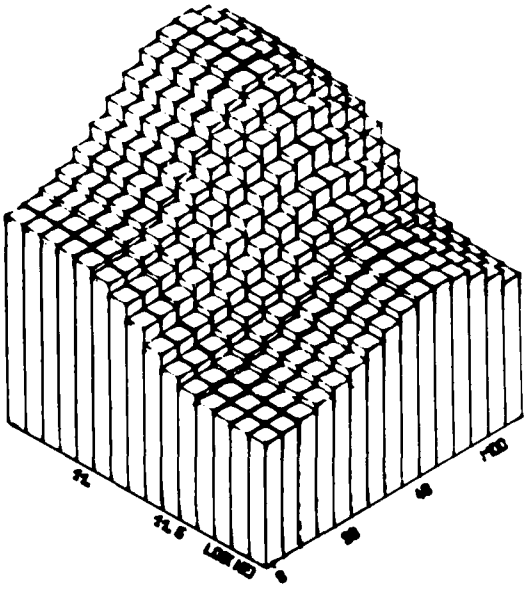
(b)

AEROS-B
TE(LOG(NE) , MODIP, H=700KM)



(c)

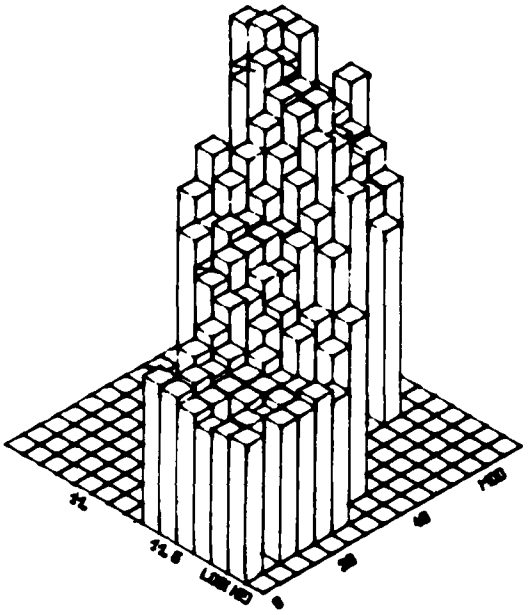
TE-MODEL
TE(LOG(NE) , MODIP, H=500KM)



(d)

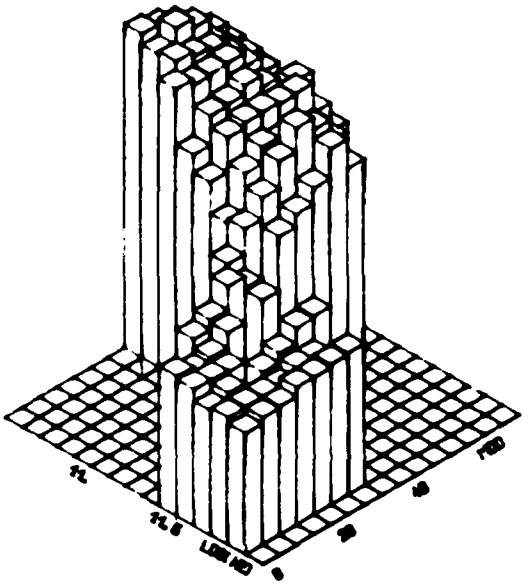
Figure 5

AEROS-B APRIL - OCTOBER
TE(LOG(NE) , MODIP, H=500KM)



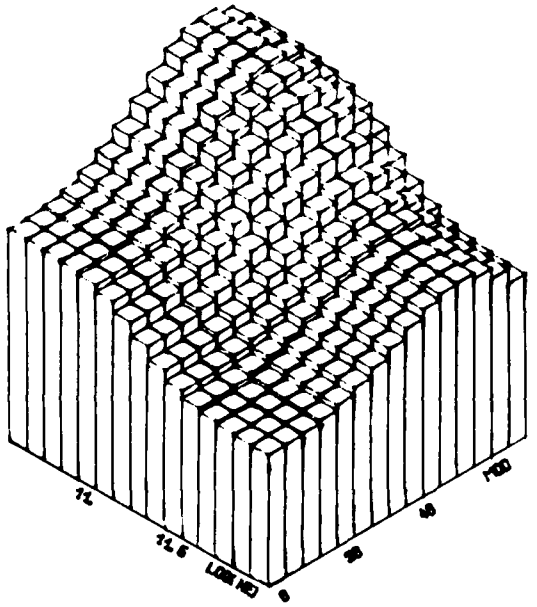
(a)

AEROS-B NOVEMBER - MARCH
TE(LOG(NE) , MODIP, H=500KM)



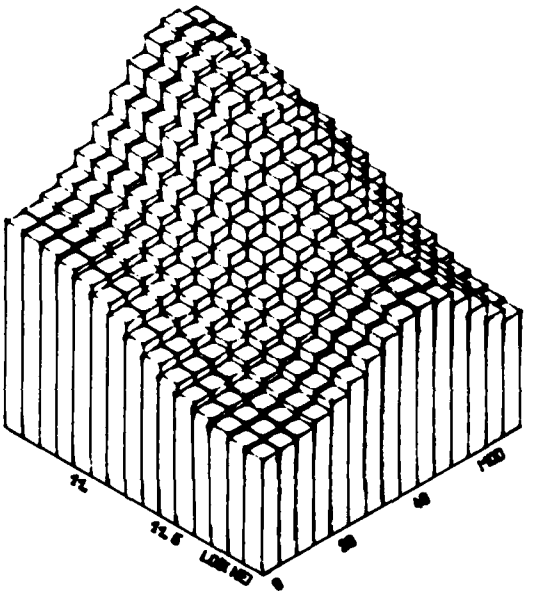
(b)

TE-MODEL APRIL - OCTOBER
TE(LOG(NE) , MODIP, H=500KM)



(c)

TE-MODEL NOVEMBER - MARCH
TE(LOG(NE) , MODIP, H=500KM)



(d)

Figure 6

SESSION DISCUSSION

SESSION IV – IONOSPHERIC PREDICTION AND MORPHOLOGY

Chairman and Editor Prof. T.B.Jones
Physics Dept.
The University
Leicester, UK

IONOSPHERIC COMPOSIITON: THE SEASONAL ANOMALY EXPLAINED
by D.G.Torr, P.G.Richards and M.R.Torr

H.E.Hinteregger, US

Would you expect that the altitudes of both the production peak of energetic oxygen atoms and associated peak of the resulting energy deposition rate density about 100 km further below in your slides, would show a fairly strong variation between solar maximum and minimum, where, for the present solar cycle, we have a density ratio of max/min of the order of 20 for heights around 400 km?

Author's Reply

The only storm that we have investigated so far is that of 16/17 December, 1971. It will be very interesting to look at the effects of changing neutral atmospheres. Our results apply to an oxygen model appropriate for the December 1971 conditions.

P.Bauer, Fr

How do you infer seasonal and long term variations of N_2^+ vibrational state?

Author's Reply

These arise directly from the effect of quenching of N_2^+ vibrational excitations by N_2 via vibration-vibration exchange. In summer, when the N_2 density is large, the N_2^+ is quenched to greater altitudes than in winter, thereby reducing the vibrational temperature and hence charge exchange of N_2^+ with O.

R.A.Goldberg, US

It is interesting to note from your Figure 2, that the strongest winter anomaly effect aligns best in a magnetic coordinate system. Could this imply that your nitrogen excited states are induced in some way be magnetically oriented sources such as corpuscular radiation?

Author's Reply

The model is reflecting largely the dependence of variations in the neutral atmosphere on magnetic coordinates. These probably arise from longitudinal and latitudinal variations which arise as a result of asymmetries in the earth's geomagnetic field which in time may give rise to asymmetries in Joule heating and particle energy input.

***Comment* by W.J.Raitt**, US

I believe that there is a dynamical effect associated with the winter anomaly related to the diminution of this effect with increasing latitude. In our model, which I presented yesterday, we did not see the seasonal anomaly for steady state solutions at 80° latitude. We attributed this to downward diffusion in winter when the lower F-region is in darkness at these latitudes. Confirmation that we could get a winter anomaly was obtained by repeating our calculations for the latitude of Millstone Hill when the expected variation of F_2 peak density from summer to winter was observed.

J.Forbes, US

What aspect of your mechanism accounts for the strength of the anomaly at solar maximum and its virtual absence during solar minimum?

Author's Reply

The primary effect is due to variations in neutral composition reproduced by the MSIS semi-empirical model. A secondary effect is due to an enhanced N_2 vibrational temperature which increases the $O^+ + N_2$ loss rate. Thirdly the seasonal change in the N_2^+ vibrational temperature is enhanced due to a larger seasonal change in the N_2 concentrations.

H.Volland, Ge

With a maximum heat input near 300 km, only the thermosphere near and above that height can be moved. The much greater mass of the lower thermosphere probably will remain at rest.

Author's Reply

Yes, we agree with this.

Comment by E.R.Schmerling, US

We now have about half the explanation of the seasonal anomaly. Your revised chemistry plus the MSIS model gives a reasonably good representation of the ionosphere, but the MSIS model itself is semi-empirical, and contains the features that make the northern and southern hemispheres so different. I await the full theory, which starts with a physical basis for the *neutral* atmosphere in which the chemical processes you have discussed today occur.

Author's Reply

I agree.

Comment by P.Stubbe, Ge

I do not want to doubt the merits of your mechanism. On the other hand, however, I do not share your view that dynamical processes do not contribute to the F-region seasonal anomaly. I have, some years ago, looked into the seasonal variation of the F-layer height. The experimental findings are such that one is simply forced to conclude that neutral winds play a significant role in producing the seasonal F-region anomaly. If I remember correctly, the winds are blowing such that for high solar activity, they enhance the seasonal F-region anomaly as caused by neutral composition change while they weaken the anomaly during sunspot minimum.

THERMOSPHERIC NEUTRAL COMPOSITION CHANGES AND THEIR CAUSES
by P.Bauer

J.Rottger, Ge

(1) How have you measured the profile of eddy diffusion coefficient up to 120 km altitude?

(2) I accept that it is also ultimately necessary to include into the calculations a profile of the horizontal wind shear to improve the estimate of the eddy diffusion profile.

Author's Reply

(1) The eddy diffusion coefficient is obtained *indirectly* through the use of the neutral heat balance equation in the lower thermosphere and taking account of the temperature profiles obtained from incoherent scatter ion temperature measurements. (See paper Alcayde et al, 1979).

(2) The temperature profiles are seasonal averages. Furthermore, large scale dynamics seems to play a minor role in the heat budget of the lower thermosphere.

K.Rawer, Ge

Etant donné que les idées concernant les processes chimiques et leur importance ont variés au cours du temps, il me semblent utile de préciser la theorie chimique utilisée a la déduction de paramètres neutres a partir de mesures qui ne penvent donner une information directe que pour les constituents changés. On ne doit pas oublier le changer que des lecteurs qui ne sont pas familiers avec ces methodes risquent de prendre les valeurs neutres déduites indirectement comme des mesures directes. Quelle est la théorie que vous avez utilisée pour déduire vos premières figures?

Response d'auteur

Le détail des calculs est donne dans l'article Alcayde et coll (1974). Il est certain que l'équation de continuité de l'ion O^+ utilise dans cet article ne prend pas en compte les termes nouveaux proposés par D.Torr dans le papier No 18. Conscient du travail en cours seu la question, j'ai cherché à evaluer l'influence du terme propose par D.Torr et coll (Geophys. Res. Lett., 5, 301, 1980); il s'est avéré que le mécanisme en cause est inopérant à l'altitude de 200 km et que de plus ce travail est dejà depassé. Sans sousestimer l'effet possible des termes proposés dans le papier no 18, il est à noter que les conclusions auxquelles nous avons abouti sur O_2 ont mis en lumière les roles conjugués de la photodissociation de O_2 et de la circulation thermospherique. Ces 2 processes devant être pres en compte dans les essais de modelisation.

RELATIONS BETWEEN SOLAR FLUX AND E-REGION PARAMETERS
by M.Nicolet and L.Bossy

J.S.Nisbet, US

Have you any comments about the accuracy of current models of the neutral atmosphere in the region of 100 km as regards E-region ion chemistry?

Author's Reply

The photoionisation rate of atomic and molecular oxygen must change with the solar cycle since the photo-dissociation frequency of O_2 increases by a factor of 2 to 2.5 from the minimum to maximum. Furthermore, it is known that the O_2 concentration at 110 km decreases with solar activity. As far as atomic oxygen is concerned, the various observations are not yet in good agreement. Thus, if there is no difficulty for N_2, there is still a problem for O_2 and O. Thus, the variation of the ratio $n(O_2)|n(O)$ may affect the ion chemistry, namely the value of the ratio of the ions, $n(NO^+)|n(O_2^+)$.

T.B.Jones, UK

Could you please comment on the likely effects of a solar flare. One would expect the X-ray spectrum to harden and the flux to increase substantially.

Author's Reply

Yes, there is a solar flare effect since the solar spectrum in the soft X-ray region is composed of coronal lines such as Fe XI-XVI in the region of 90 40 A and even Fe XVII-XXV in the region of 20 10 A. Other coronal lines (Si, Mg, Ni, . . .) are also involved. The solar flare effect leads to an increase in the photoionization rate particularly below the peak of the normal E-layer.

K.Rawer, Ge

I suppose that your model computation was made initially for explaining the E-peak, and the E-region below it. At altitudes above the peak other radiations must probably be taken into account such that the real production profile would not show a true peak with a valley above, but increases steadily up to at least 160 180 km. Do you agree?

Author's Reply

Yes, above the E-layer peak, there is an effect of the solar radiation emitted in the H-Lyman continuum; it ionizes atomic oxygen at wavelengths less that 910 A. In addition, the action of the solar radiation between 100 A and 250 A (where the absorption cross-section is still small enough) is particularly important. However, if there is no clear indication of a valley between the E and F_1 peak for high solar activity conditions and overhead sun, there is a transition region leading to a valley which becomes particularly pronounced for high solar zenith angles, particularly near sunset and sunrise, and for relatively quiet solar conditions. This can be explained by the differences in the optical depths which increase with the solar zenith angle.

ENERGETIC O^+ PRECIPITATION: A SIGNIFICANT ENERGY SOURCE FOR THE IONOSPHERE
by M.R.Torr, D.G.Torr and R.Roble

G.Prölss, Ge

(1) How well known are the cross sections needed in your calculation of the energy deposition profile and the ionization profile?

(2) How sensitive are your results to the accurate knowledge of these cross sections?

Author's Reply

In determining the conversion of energetic O^+ to energetic O the two primary processes are momentum exchange and charge exchange collisions. By comparison with the less important processes such as excitation, these cross sections are reasonably well known. The cross-sections could change significantly but only result in a modification to the altitude profiles for production of just O and the heating rate.

ON THE MORPHOLOGY OF THE POLAR THERMOSPHERE
by G.W.Prölss

A.Brekke, No

Is it possible to see any difference in energy densities deposited in the evening and the morning side of the auroral oval?

Author's Reply

So far, only the latitudinal structure and extension of the polar atmospheric disturbance has been investigated. I am afraid, however, that even a study of the disturbance intensity will not be very helpful in identifying variations in energy deposition morphology because transport effects are very important. Therefore the atmospheric disturbance intensity may not reflect the local energy deposition.

H.Volland, Ge

(1) Superposition of the diurnal (King-Kohl) wind and of the wind due to Joule heating reduces the meridional wind during day-time and enhances the equatorward wind during the night. This may explain the equatorward change of the disturbance zone during the night.

(2) A momentum force generates mainly horizontal (rotational) winds. These winds do not affect composition. It is the solenoidal wind generated by Joule heating or EUV heating which is mainly responsible for wind induced diffusion.

Comment by K.Rawer, Ge

I feel this statement is too general. In the case, for example, where a certain minor constituent is concentrated at certain latitude, the situation would be changed by a cell of horizontal winds transporting this constituent to another latitude. By the way, this is how Professor Nicolet thinks NO could be transport to mid latitudes from the auroral region where it is produced.

Reply by H.Volland

A cloud of minor constituents transported only horizontally within the middle and upper thermosphere would disappear very quickly due to diffusion. Effective transport of minor constituents from higher to lower latitudes occur via circulation cells of planetary scale which reach from the lower to the upper thermosphere and include significant vertical wind components.

STRUCTURE OF ELECTRON TEMPERATURE
by H.Thiemann

Comment by K.Rawer, Ge

In order to understand the relations empirically found here, one should note that the charged constituents are only minor ones in the F-region, further that there is no thermal equilibrium in daytime. In this situation the electron temperature can be changed by a change in thermal conductivity between the constituents without external energy in-or output, just by heat flow from the electrons to the main neutral population, for example. Now since the heat transfer goes with the square of electron density, an increase of this latter lowers the electron temperature. This is valid around the F-region peak where such local heat transfer is most important. At higher altitudes heat transport upwards or downwards along the lines of magnetic force, can become important. This also is more efficient the higher the electron density. Therefore, the inverse relation between electron density and temperature can well be understood.

MIDLATITUDE SPORADIC-E LAYERS

L. G. Smith and K. L. Miller*

Aeronomy Laboratory, Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801 U.S.A.

SUMMARY

Rocket-borne probes and incoherent-scatter radar have been demonstrated to be effective methods of studying the structure of midlatitude sporadic-E layers. Layers are formed when metal ions are converged vertically in a wind shear to produce a local enhancement of electron density. Rocket and radar observations show that the layers may occasionally have complex structure produced by an unstable wind shear. The partial transparency, to radio waves, of sporadic-E layers is shown to be due to localized regions of high electron density.

1. INTRODUCTION

Observations of sporadic-E layers by ground-based radio experiments, such as the ionosonde, show strong echoes from an altitude of about 100 km. The virtual height is constant with frequency. Also there is usually a range of frequencies, from the blanketing frequency to the top (or critical) frequency, for which the layer is partially transparent.

Rocket-borne probe experiments confirm the implication of the radio experiments: that there is a layer of enhanced electron density having a total thickness usually about 1 km, but occasionally having greater thickness.

The appearance of the layers in electron density profiles obtained by rocket experiments is illustrated in Figure 1. These are intense layers and are relatively rare. In the daytime enhancements of a factor of 3 in electron density above the ambient value are observed. At night, when the ambient electron density is much reduced, the enhancement can exceed a factor of 10.

It can also be noted in Figure 1 that the daytime sporadic-E layer has a simple structure whereas the nighttime layer has a more complicated, double-peaked structure. Similar double-peaked layers and more complicated structures have been observed in daytime profiles [*Smith and Mechtly*, 1972].

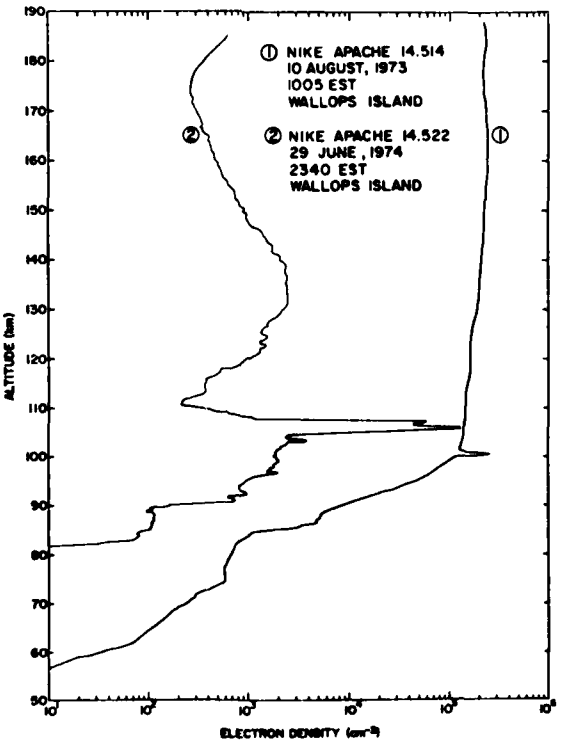


Fig. 1. Examples of intense sporadic-E layers in daytime and nighttime electron-density profiles.

*Present address: Lockheed Palo Alto Research Laboratory, 3521 Hanover Street, Palo Alto, CA 94304 U.S.A.

Rocket data also show that most layers occur in the altitude range of 95 to 120 km and that the layers are horizontal over distances of several hundred km [*Smith and Mechtly*, 1972]. Radio and radar observations indicate wave-like structure in the layer.

Rocket-borne positive ion mass spectrometers show that the mid-latitude sporadic-E layers are produced by thin layers of metal ions; an example is shown in Figure 2. The presence of metallic ions has also been inferred from incoherent scatter observations.

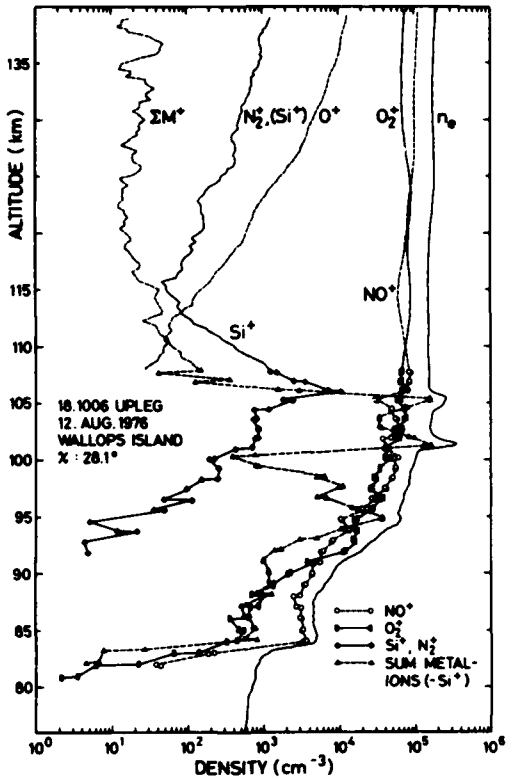


Fig. 2. Positive ion concentrations observed during the occurrence of intense sporadic E [*Hermann et al.*, 1978].

The metallic ions are undoubtedly of meteoritic origin but many questions remain regarding their deposition rates and ionization rates. This is, in part, attributable to missing information on chemical reaction rates but also to lack of reliable information on the concentrations of metal atoms in the neutral state.

The mechanism by which the metal ions are converged to form sporadic-E layers is now well established. It is the wind-shear mechansim, a manifestation of the atmospheric dynamo. However, the structure of the wind system near 100 km is complicated by gravity waves, tides and turbulence, and their interactions, and cannot be predicted. Thus the sporadic-E layers provide a method of studying, in some detail, the neutral wind system of the dynamo region. However, sporadic E is likely to remain unpredictable.

2. FORMATION OF SPORADIC-E LAYERS

The motion of a charged particle in the dynamo region of the atmosphere is determined mainly by the effects of its collisions with neutral atoms and molecules, by electric fields and by the geomagnetic field. The equation of motion is [*Macleod et al.*, 1975].

$$m\frac{d\vec{v}}{dt} = m\nu(\vec{\omega}-\vec{v}) + q\vec{v}\times\vec{B} + q\vec{E} \qquad (1)$$

where q, m and $\vec{v}$ are the charge, mass and velocity of the charged particle, $\vec{\omega}$ is the velocity of the neutral wind, ν is an effective collision frequency, $\vec{E}$ is the electric field and $\vec{B}$ is the geomagnetic field. For the steady state, equation (1) can be solved for $\vec{v}$ giving

$$\vec{v} = \frac{1}{1+\rho^2}\left[\rho^2\vec{u} + \rho\vec{u}\times\hat{\Gamma} + (u\cdot\hat{\Gamma})\hat{\Gamma}\right] \qquad (2)$$

where $\rho \equiv \nu/\omega_H$, the gyrofrequency $\omega_H = qB_O/m$, $\hat{\Gamma}$ is a unit vector in the direction of $\vec{B}$ and B_O is the magnitude of $\vec{B}$. Terms containing $\vec{\omega}$ and $\vec{E}$ have been combined into $\vec{u} \equiv \vec{\omega} + \vec{E}/\rho B_O$. This is the most general formulation.

In the absence of electric fields the vertical component of ion velocity is, from equation (2),

$$v_z = \frac{\rho}{1+\rho^2}\cos\phi\, u_E - \frac{1}{1+\rho^2}\cos\phi\,\sin\phi\, u_N \qquad (3)$$

The effectiveness of the eastward (u_E) and northward (u_N) components of the wind vary with altitude (since ρ is proportional to ν) and with latitude (through the dip angle, ϕ). As shown in Figure 3 the eastward wind has a maximum effect near 120 km, where it equals the effectiveness of the northward wind. However, in the region of maximum occurrence of sporadic E (100 to 110 km) the eastward wind is the more important.

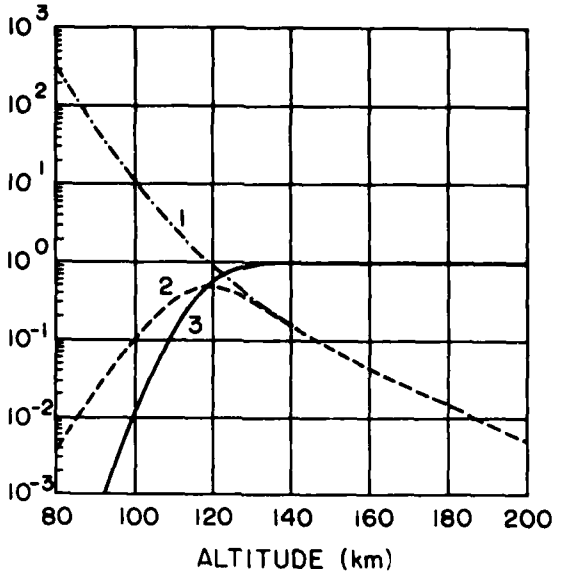


Fig. 3. Variation of (1) ρ, (2) $\rho/(1+\rho^2)$ and (3) $1/(1+\rho^2)$ *with altitude [adapted from MacLeod, 1966].*

The effect of changing latitude is illustrated in Figure 4. This shows the dip-angle factors of equation (3). The northward wind is most effective at a dip angle of 45° but the eastward wind increases continuously in effectiveness from the pole to the equator. Detailed consideration of the role of electrons shows that a polarization electric field develops at the geomagnetic equator, inhibiting the vertical ion motion. Except near the equator the motions of electrons may be ignored in considering the convergence of ions.

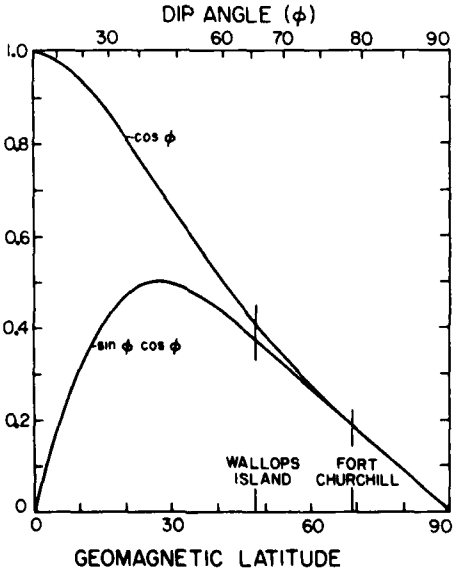


Fig. 4. Variation of the dip-angle factors with latitude.

One of the major problems encountered in the early development of the wind-shear theory was the relatively short lifetimes of the molecular ions of this region. *Axford and Cunnold* [1966] subsequently suggested that metal ions, with their very long lifetimes (hours, or days) could be converged to form the layer. This is

supported by experimental observations of metal ions in sporadic-E layers and provides a generally satisfactory explanation.

Since recombination of the metal ions is unimportant the shape of the layer, in the steady state, is determined by diffusion. Retaining only the vertical transport and diffusion terms in the continuity equation leads to

$$\frac{d}{dz}(nv_z) = \frac{d}{dz}(D\frac{dn}{dz}) \tag{4}$$

where n is the concentration of metal ions and D is a diffusion coefficient. Integration gives

$$v_z = D\frac{d}{dz}(\ln n) \tag{5}$$

When the wind profile is known v_z may be calculated from equation (3). Then equation (5) may be used to obtain the steady state profile of metallic ions.

3. EFFECTS OF UNSTABLE WIND SHEARS

Equation (5) may also be used, when the ion profile is known, to obtain a profile of v_z. Since v_z depends on both u_E and u_N, through equation (3), it is not possible to solve uniquely for either u_E or u_N. However when layers with altitudes near 100 km are considered u_E is the more important component of the wind.

Using this approach it has been shown [*Smith and Miller*, 1980] that sporadic-E layers of complex structure are consistent with the vertical convergence of metallic ions in a shear in the neutral wind, as is illustrated in Figure 5. The linear wind shear gives a simple Gaussian shape but shears greater than 50 m s^{-1} km^{-1}, which corresponds to a Richardson number of 1/4 at 100 km, develop instabilities, resulting in the complex structures in the ion (and electron) density profile. The actual shape that results has not yet been explained.

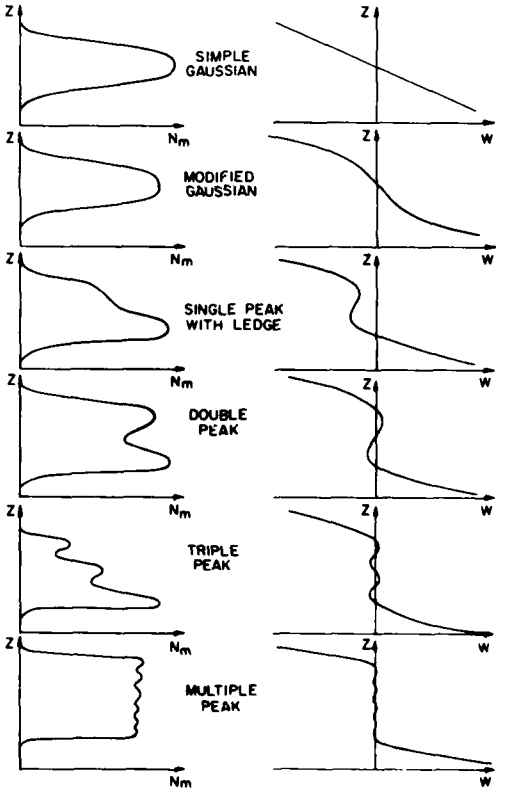


Fig. 5. Observed shapes of sporadic-E layers (left column) and the wind shear (right column) required to maintain the layers against diffusion [*Smith and Miller*, 1980].

4. PARTIAL TRANSPARENCY OF SPORADIC-E ECHOES

Various models have been proposed to explain the partially transparent echo which is characteristic of sporadic-E layers recorded by ionosondes. There are two main groups of theories: those that interpret the echo as a partial reflection at the sharp upper and lower boundaries and those that postulate small region of high electron density embedded in the layer.

A recent study [*Miller and Smith*, 1977] of the partial reflection at layers observed by rockets shows conclusively that partial reflection is inadequate to explain the large frequency range often observed for the transparent echo.

The other theory is, however, supported by a study of sporadic-E layers using the incoherent scatter radar at Arecibo [*Miller and Smith*, 1978]. One example of irregular structure containing regions of high electron density is shown in Figure 6, where each scan of the radar covers a horizontal distance of about 20 km at this altitude.

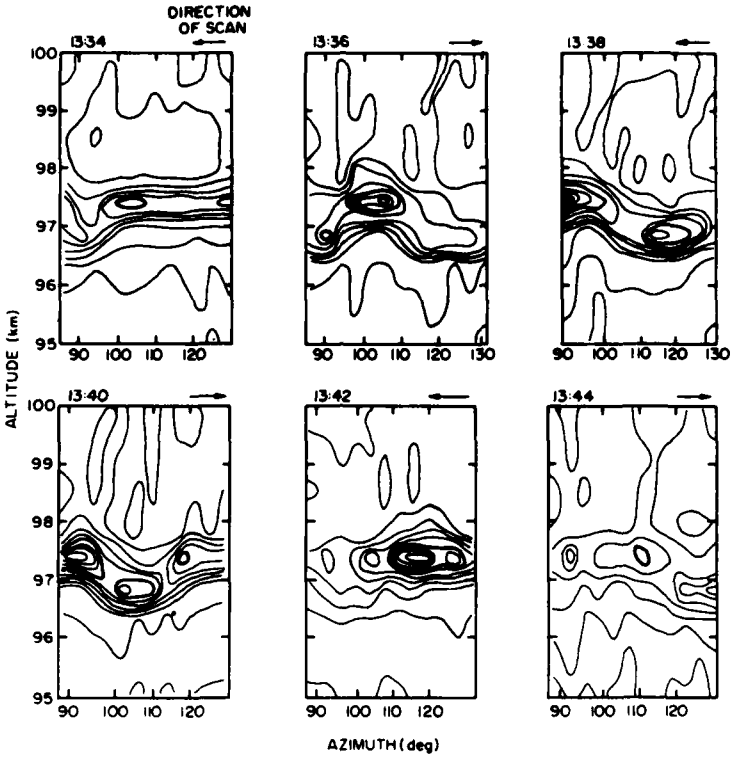


Fig. 6. Contours of electron density in a sporadic-E layer showing regions of enhancement [*Miller and Smith*, 1978].

The data at Arecibo show even more complex structure than had been anticipated, including some having the appearance of instabilities. A particularly striking example of another kind, Figure 7, shows the edge of a layer. This implies boundaries exist in the E region between air masses of different properties, as in tropospheric frontal systems.

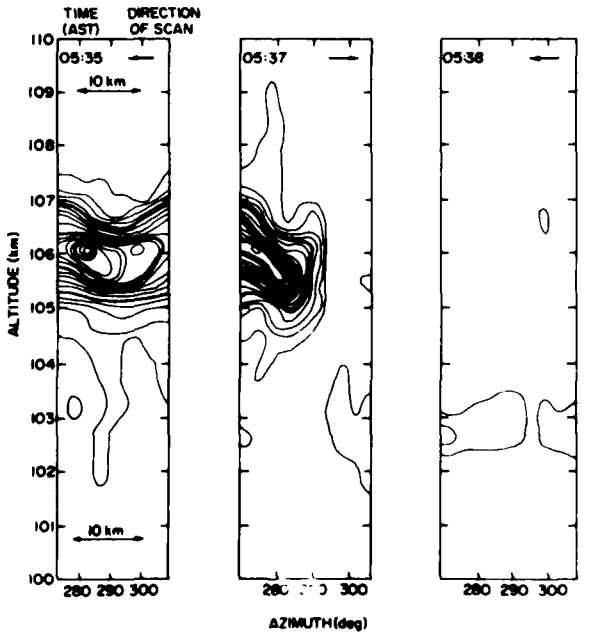


Fig. 7. Contours of electron density showing the edge of a sporadic-E layer [*Miller and Smith*, 1978].

5. CONCLUSIONS

Investigation of midlatitude sporadic-E layers show that they are formed by the vertical convergence of metal ions under the influence of a shear in the neutral wind system. The metal ions are of meteoritic origin. The prediction of occurrence of sporadic-E, if ever possible, requires a much greater knowledge of the neutral wind system in the dynamo region than is now available.

ACKNOWLEDGEMENT

This work has been supported by the National Aeronautics and Space Administration under grant NGR 14-005-181.

REFERENCES

Axford, W. I., and Cunnold, D. M., 1966, The wind-shear theory of temperate zone sporadic E, *Radio Sci.*, *1*, 191-198.

Hermann, U., et al., 1978, Metal ions and isotopes in sporadic-E layers during the Perseid meteor shower, *Space Res.*, *18*, 249-252.

MacLeod, M. A., 1966, Sporadic E theory. I. Collision-geomagnetic equilibrium, *J. Atmos. Sci.*, *23*, 96-109.

MacLeod, M. A., et al., 1975, Numerical modelling of a metallic ion sporadic-E layer, *Radio Sci.*, *10*, 371-388.

Miller, K. L., and Smith, L. G., 1977, Reflection of radio waves by sporadic-E layers, *J. Atmos. Terr. Phys.*, *39*, 899-911.

Miller, K. L., and Smith, L. G., 1978, Incoherent scatter radar observations of irregular structure in mid-latitude sporadic-E layers, *J. Geophys. Res.*, *83*, 3761-3775.

Smith, L. G., and Mechtly, E. A., 1972, Rocket observations of sporadic-E layers, *Radio Sci.*, *7*, 367-376.

Smith, L. G., and Miller, K. L., 1980, Sporadic-E layers and unstable wind shears, *J. Atmos. Terr. Phys.*, *42*, 45-50.

THE SEASONAL AND GEOGRAPHICAL VARIATION OF EQUATORIAL SPREAD-F IRREGULARITIES INFLUENCED BY ATMOSPHERIC GRAVITY WAVES AND ELECTRIC FIELDS DUE TO THUNDERSTORMS

J. Röttger

Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, Fed. Rep. Germany

SUMMARY

It is proposed that thunderstorm generated electric fields superimpose on the ionospheric electric field and therefore could control the occurrence of equatorial spread-F irregularities. It is more likely, however, that steepened travelling ionospheric disturbances due to gravity waves generated by penetrative thunderstorm convection are a determining factor for large-scale irregularity generation. An empirical model, taking into account the global activity of tropospheric convection, is presented which can explain some typical features of equatorial range spread-F occurrence.

1. INTRODUCTION

There is accepted evidence that atmospheric gravity waves (AGWs) generated in the troposphere and stratosphere can propagate into the thermosphere where they manifest themselves as travelling ionospheric disturbances (TIDs). It has been shown that a connection between ionization irregularities of the equatorial spread-F and atmospheric gravity waves exists (KLOSTERMEYER, 1978; RÖTTGER, 1978a; BOOKER, 1979). The coupling mechanism is the spatial resonance between ionization drift and travelling ionospheric disturbances. Recent reviews (e.g. BASU and KELLEY, 1979; OSSAKOW, 1979; FEJER and KELLEY, 1980) call attention to this mechanism.

By means of this resonance, which sets in if the phase velocity of TIDs equals the ionization ($\underline{E} \times \underline{B}$) drift velocity, amplitudes of TIDs may be non-linearly steepened so that they "break" into turbulence (i.e. ionization density irregularities). Observations of spread-F irregularities (viz. radar, ionosonde, HF-Doppler and transequatorial propagation) can be discussed in terms of this mechanism.

Since the spatial resonance mechanism can be accepted as a physical basis of distinct irregularity features (viz. macroscale structure) and the microscale structure (viz. small-scale plasma irregularities) is also governed by $\underline{E} \times \underline{B}$ forces, we approach the position to explain temporal and spatial variations of equatorial spread-F by applying knowledge of ionization drift and TIDs. There are two reasons why we expect a connection between drift, electric fields and TIDs on the one side, and lower atmospheric phenomena on the other. This incidentally leads to a connection between special features of the equatorial spread-F and lower atmospheric phenomena.

(1) Superimposed on the ionization drift at the equator, caused by the dynamo effect, is a drift component due to electric fields ($\underline{E}$) induced by thunderstorms (e.g. HAYS and ROBLE, 1979). It follows that a correlation between the variation of drift velocities, and consequently the occurrence of spread-F irregularities, and the thunderstorm activity is expected.

(2) It was shown (RÖTTGER, 1977) that penetrative cumulus convection, which is connected to thunderstorm activity (RÖTTGER, 1980), is an effective source of gravity waves propagating from their generation region in the troposphere up to the thermosphere. Here they are observed as TIDs which represent a basic condition for the spatial resonance mechanism. Again a correlation between thunderstorm activity and spread-F irregularities is expected.

A more elaborate description of these mechanisms will appear elsewhere. In this paper, experimental proof for these processes is presented by pointing to a correlation between the diurnal, seasonal and geographical variation of equatorial spread-F irregularities and convective (thunderstorm) activity. Originating mechanisms are dynamical processes (convection) in the lower atmosphere, the coupling mechanisms are electric fields and atmospheric gravity waves, and one generating mechanism (in connection with plasma instabilities) of equatorial spread-F structures is the spatial resonance mechanism.

2. AN EMPIRICAL MODEL OF EQUATORIAL SPREAD-F MORPHOLOGY

Assuming that gravity waves play a role as a seeding mechanism of equatorial spread-F (ESF), a probability of the occurrence of large-scale ESF can be defined. To generate ESF irregularities by means of Rayleigh-Taylor respectively $\underline{E} \times \underline{B}$ instability, a background plasma drift $\underline{u}$ corresponding to an electric field $\underline{E}$ is required. To generate wavelike ESF disturbances, TIDs are necessary. Then the conditional probability for large-scale wavelike ESF irregularities can be defined by the product of the occurrence frequency of TIDs and an assumed normalized occurrence frequency of ionospheric plasma drift velocities. Admitting that steepened TID structures are caused by the spatial resonance mechanism, this conditional probability is a fundamental of the spatial resonance effect which needs TIDs as well as plasma drifts. The spatial resonance condition (RÖTTGER, 1978a) then selects those TIDs with proper phase velocities matching the plasma drift velocities. It is evident that a downward plasma drift, corresponding to a westward electric field, is required for spatial resonance. This vertical drift direction is observed most regularly in the post-sunset/premidnight hours when the horizontal drift direction mostly is towards the east. Correspondingly, TIDs with an eastward phase velocity component can excite the resonance mechanism in the post-sunset equatorial F-region (e.g. RÖTTGER, 1978a).

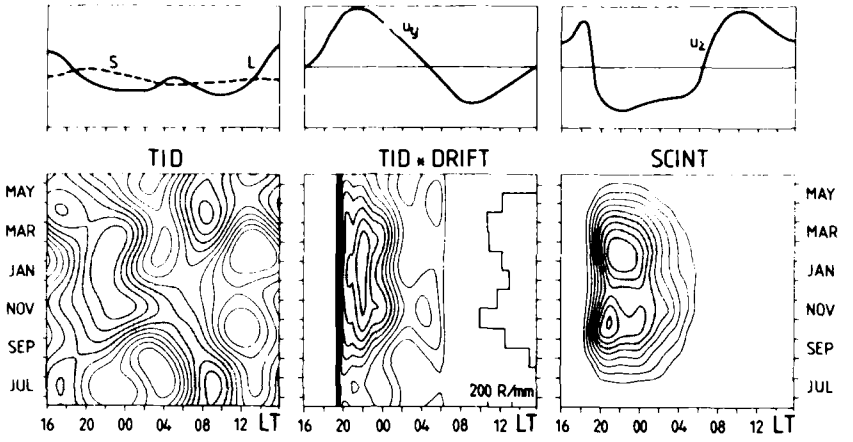


Figure 1

Relative TID occurrence, conditional probability distribution TID * DRIFT and VHF scintillation occurrence SCINT (after WHITNEY and AARONS, 1976).

In Figure 1 the occurrence frequency of TIDs as function of day and season is shown, which was observed by RÖTTGER (1978b) during the years 1969, 1970 and 1973 in the equatorial region of Africa. A broad TID occurrence maximum is observed in the post-sunset hours during the southern summer months from October to March. These TIDs are most likely generated by gravity waves due to penetrative cumulus convection (RÖTTGER, 1977), which is intimated in Figure 1 by the correlation of TID occurrence with the seasonal variation of rainfall rate R in the observation area. The maximum of TID occurrence is connected with late afternoon/ early evening maxima of cumulonimbus rainfall activity over land (L). A smaller rainfall maximum is observed in the morning hours, connected with a secondary maximum of TIDs. The normalized occurrence frequency of TIDs is assumed to represent the probability distribution (TID) for a specific region. Knowing an average diurnal variation of the drift velocities u_y and u_z, which are normalized to their maximum values (Figure 1), we can multiply it with (TID) to obtain (TID * DRIFT). The corresponding diurnal and seasonal probability distribution is shown in Figure 1. We find a diurnal variation of the probability (TID * DRIFT) which fits to the observed average diurnal variation of ESF occurrence. The seasonal variation shows a broad maximum from October to March with some post-midnight widening during the equinoxes. This can be caused only by the variation of TID occurrence since the seasonal variation of the drift velocities was not considered in Figure 1.

The model calculations can be compared with observations of scintillation activity which is related to ESF activity. The scintillation distribution shown in Figure 1 was deduced by WHITNEY and AARONS (1976) from observations of 254-MHz LES-6 satellite signals at Huancayo in 1972. We find a general similarity of

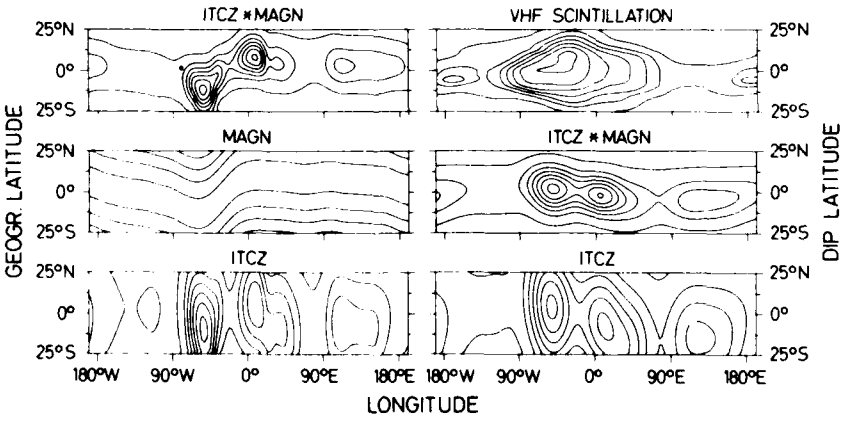


Figure 2

Probability distributions: ITCZ = intertropical convergence zone, MAGN = magnetic field controlled drift velocity, ITCZ * MAGN = conditional probability of travelling ionospheric disturbances in the intertropical convergence zone and ionospheric drift velocity. VHF scintillation observations after BASU et al. (1976).

the model distribution and the observed scintillation distribution. By accepting that TIDs seed the ESF bubbles, which cause the most intense scintillations (BASU and KELLEY, 1979), the fitting of model and observations turns out to be not an unexpected result.

This result is supported by further computations of the geographical variation of ESF occurrence. The $\underline{E} \times \underline{B}$ drift velocities in the equatorial ionosphere are controlled by the geometry of the earth's magnetic field $\underline{B}$. We assume in a simple approach that the probability to find adequate drift velocities fulfilling the resonance condition has a maximum at the dip equator, and its latitudinal variation has a Gaussian shape with 30° half width. The diagram, labelled MAGN in Figure 2, indicates the lines of constant probability for this model. By taking into account the mean position of the ITCZ and the secondary convergence areas in January (NIEUWOLT, 1977) as well as the yearly average of thunderstorm days (LILJEQUIST, 1974), a model of TID occurrence can be found. The obtained probability distribution is depicted in Figure 2 in the diagrams labelled ITCZ. A model distribution is then computed, which is shown in the diagrams of Figure 2 labelled ITCZ * MAGN. The right-hand diagrams show the distributions as function of dip latitude to enable a comparison with scintillation observations (OGO-6 data on 140 MHz, 19-23 LMT, Nov.-Dec. 1969, 1970, deduced by BASU et al. (1976)). In spite of the fact that the scintillation data are for November-December only and the ITCZ activity is a yearly average, a general correlation of scintillation and model distributions is recognized. Largest occurrence maxima are located over the continents of Africa and South America; a steep decrease west of the South American continent, low occurrence over the Pacific Ocean, and a minor maximum in the East Asia regions with gradual increase towards the African continent are found. It generally follows that the described empirical model which basically relies on the distribution of convective storms generating TIDs can give an appropriate explanation of the seasonal and geographical variation of equatorial spread-F irregularities.

3. CONCLUSION

An empirical model is described which can explain features of large-scale equatorial spread-F irregularities. A correlation between the diurnal, seasonal and geographical variation of spread-F irregularities and convective thunderstorm activity is found. Originating mechanisms for this correlation are dynamical processes (convection) in the lower atmosphere; the coupling mechanisms are electric fields and atmospheric gravity waves. The empirical model, which bases on worldwide distribution of convective regions and empirical assumptions about ionospheric electric fields, does not claim to be exhaustive but can describe several typical features of the morphology of equatorial spread-F irregularities.

References

BASU, S. and M.C. KELLEY, "A review of recent observations of equatorial scintillations and their relationship to current theories of F region irregularity generation", Radio Sci. 14, 471-485, 1979.

BASU, S., S. BASU and B.K. KHAN, "Model of equatorial scintillations from in-situ measurements", Radio Sci. 11, 821-832, 1976.

BOOKER, H.G., "The role of acoustic gravity waves in the generation of spread F and ionospheric scintillation", J. Atmos. Terr. Phys. 41, 501-515, 1979.

FEJER, B.G. and M.C. KELLEY, "Ionospheric irregularities", Rev. Geophys. Space Phys. 18, 401-454, 1980.

HAYS, P.B. and R.G. ROBLE, "A quasi-static model of global atmospheric electricity. 1. The lower atmosphere", J. Geophys. Res. 84, 3291-3305, 1979.

KLOSTERMEYER, J., "Nonlinear investigation of the spatial resonance effect in the nighttime equatorial F region", J. Geophys. Res. 83, 3753-3760, 1978.

Liljequist, G.H., "Allgemeine Meteorologie", Friedr. Vieweg + Sohn, Braunschweig, 1974.

NIEUWOLT, S., "Tropical Climatology", John Wiley + Sons, London, 1977.

OSSAKOW, S.L., "Ionospheric irregularities", Rev. Geophys. Space Phys. 17, 521-533, 1979.

RÖTTGER, J., "Travelling disturbances in the equatorial ionosphere and their association with penetrative cumulus convection", J. Atmos. Terr. Phys. 39, 987-998, 1977.

RÖTTGER, J., "Drifting patches of equatorial spread-F irregularities - Experimental support for the spatial resonance mechanism in the ionosphere", J. Atmos. Terr. Phys. 40, 1103-1112, 1978a.

RÖTTGER, J., "Modelling the diurnal and seasonal variation of medium-scale travelling ionospheric disturbances", AGARD Conference Proc. 238, paper 19, 1978b.

RÖTTGER, J., "Structure and dynamics of the stratosphere and mesosphere revealed by VHF radar investigations", Pure Appl. Geophys. 118, 494-527, 1980.

WHITNEY, H.E. and J. AARONS, "Amplitude scintillation observations and systems application", AGARD Conference Proc. 173, paper 2, 1976.

M. Crochet
Laboratoire de Sondages Electromagnétiques de l'Environnement Terrestre
(L.S.E.E.T) - Université de Toulon et du Var
La Giponne - Boulevard des Armaris - Toulon 83100 - France

PLASMA INSTABILITIES IN THE ELECTROJETS
INSTABILITES DE PLASMA DANS LES ELECTROJETS

ABSTRACT

The electron density irregularities embedded in the equatorial and auroral electrojets have been associated with plasma instability phenomena (two-stream and cross-field instabilities). These irregularities extensively investigated using VHF and UHF radars and more recently HF radars, at the equator and in the auroral zone are often used as tracers of the electron velocity and the electric field in the electrojets. A number of recent experimental results obtained in the equatorial electrojet in Africa with HF radar are presented : long wavelengths measurements, counter-electrojet observations. At the light of these experimental results a new classification of the electrojet plasma instabilities is proposed. Threshold conditions and saturation of the instabilities are discussed for different equatorial and auroral configurations as the validity of electric field measurements.

RESUME

Les irrégularités de densité électronique des électrojets équatorial et auroraux ont été associées à deux types d'instabilités de plasma (instabilité à deux faisceaux, instabilité en champs croisés). Ces irrégularités étudiées de façon intensive à l'aide de radars VHF et UHF et plus récemment de radars HF, en zones équatoriale et aurorale, ont souvent été utilisées comme traceurs du mouvement des électrons et moyen de mesure des champs électriques dans les électrojets. Des résultats expérimentaux obtenus récemment par radar HF en Afrique sont présentés (influence de la longueur d'onde, observations du contre-électrojet) et une nouvelle classification des instabilités de l'électrojet est proposée à la lumière de ces résultats. Les conditions d'apparition des instabilités et de leur saturation sont discutées dans les différentes configurations, équatoriale et aurorale, ainsi que la validité des mesures de champ électrique.

1. INTRODUCTION

Depuis la première association d'une instabilité de plasma à des irrégularités ionosphériques dans l'électrojet équatorial (BOWLES K. et al, 1963; FARLEY D., 1963) les études tant expérimentales que théoriques des instabilités de plasma dans l'ionosphère ont connu un développement considérable au cours des deux dernières décades.

Récemment, plusieurs revues (FEJER, J.A., 1979 ; OSSAKOW S.L., 1979 ; FARLEY D., 1979 ; FEJER B.G. and KELLEY M.C., 1980) ont couvert les aspects théoriques et expérimentaux, les phénomènes équatoriaux et auroraux et les irrégularités des couches ionosphériques E et F.

Il convient de noter deux orientations principales dans le cadre des études des irrégularités des électrojets :

a) l'étude de la physique des instabilités de plasma et de la turbulence associée qui connait actuellement un développement important en particulier par les méthodes de simulation numérique du milieu (KESKINEN M. et al., 1979 ; OSSAKOW S.L., 1979).

b) l'étude de la dynamique (champs et courants) du milieu ionosphérique considérant les irrégularités comme traceurs des champs électriques et des courants. C'est ainsi que dans certaines conditions les irrégularités de l'électrojet peuvent être utilisées pour mesurer les champs électriques à l'équateur (BALSLEY B.B., 1973) et en zone aurorale (GREENWALD R.A., 1977).

Un dernier aspect de l'étude des instabilités de plasma dans les électrojets qui n'a été abordé que très partiellement (RICHMOND A.D., 1973) est l'influence des instabilités sur la résistivité du milieu et les sources qui lui donnent naissance (effet inhibiteur).

Alors qu'une large part des résultats initiaux sur les irrégularités a été obtenue à l'aide de radars Doppler VHF et UHF, des expériences récentes effectuées en Afrique à l'aide de radars Doppler HF ont permis d'étendre le champ d'investigation des irrégularités de l'électrojet équatorial et d'obtenir des résultats importants pour l'interprétation des processus d'instabilités donnant naissance à ces irrégularités (CROCHET M., 1980). Avec un objectif similaire, des études expérimentales HF (projet S.A.F.A.R.I. : Swedish and French Auroral Investigation) viennent d'être entreprises en zone aurorale en coordination avec les projets STARE et EISCAT (CROCHET M. et al., 1980).

Le but du présent article est de présenter brièvement l'état de la connaissance des instabilités dans les électrojets équatorial et auroraux en incluant les plus récents résultats des radars HF et en comparant les conditions équatoriales et aurorales.

2. CONDITIONS D'OCCURRENCE DES INSTABILITES

L'étude du spectre Doppler des échos radars observés à JICAMARCA à 50 MHz a permis d'identifier deux types de spectres appelés historiquement "Type 1" et "Type 2" (BALSLEY B.B., 1969) et associés respectivement à l'instabilité à deux faisceaux (FARLEY D., 1963) et en champs croisés (ROGISTER R.A. and N. D'ANGELO, 1970). Ce n'est que plus tard (FARLEY D. and FEJER B., 1975) qu'il est apparu clairement qu'il n'y avait qu'une seule équation de dispersion où les deux termes correspondant aux deux instabilités précédentes pouvaient alternativement dominer le spectre selon les conditions expérimentales (valeur du champ électrique et du gradient de densité électronique, longueur d'onde observée) comme il apparaîtra clairement par la suite.

La théorie linéaire des instabilités dans l'électrojet a été développée, d'une part à l'aide de la théorie cinétique (FARLEY D., 1963), d'autre part à partir des équations fluides (BUNEMAN O., 1963 ; WALDTEUFEL P., 1965 ; ROGISTER A. and N. D'ANGELO, 1970 ; FEJER B. et al, 1975). Pour des ondes se propageant perpendiculairement au champ magnétique et en négligeant le déplacement des ions et l'inertie des électrons les termes réels et imaginaires de l'équation de dispersion s'écrivent respectivement (FEJER B. and M. KELLEY, 1980) :

$$\omega_r = \frac{\vec{k} \cdot \vec{V}_{oe}}{1 + \Psi} \tag{1}$$

$$\Gamma = \frac{1}{1+\Psi} \left\{ \frac{\Psi}{\nu_i} \left[\omega_r^2 - k^2 C_S \right] + \frac{1}{L_N k^2} \frac{\omega_r \nu_i k_y}{\Omega_i} \right\} \tag{2}$$

ω_r étant la pulsation de l'instabilité sélectionnée, Γ son taux de croissance, $\vec{V}_{oe}$ la vitesse de dérive des électrons, $\vec{k}$ le vecteur d'onde, $\Psi = \nu_e \nu_i / \Omega_e \Omega_i$ avec ν_e, ν_i, Ω_e, Ω_i les fréquences de collision et gyrofréquences des électrons et des ions, $L_N = N (\partial N/\partial z)^{-1}$ la grandeur d'échelle du gradient de densité électronique (positive vers le haut à l'équateur), C_S la vitesse acoustique des ions, k_y la composante du vecteur d'onde (vers l'ouest à l'équateur).

Avec Θ angle du vecteur d'onde et de la vitesse électronique, $V_{ph}=(\omega_r/k)$ vitesse de phase des instabilités, on peut écrire :

$$V_{ph} = \frac{V_{oe} \cos \Theta}{1 + \Psi} \tag{3}$$

La vitesse de phase des instabilités donnée par la théorie linéaire est donc proportionnelle à la vitesse électronique et à cos Θ .

$$\Gamma = \frac{\Psi}{\nu_i (1+\Psi)} \left[\omega_r^2 - k^2 C_S^2 + 2 F C_S k_y \omega_r \right] \tag{4}$$

en posant

$$F = \nu_i \Omega_e / 2 \nu_e k^2 L_N C_S \tag{5}$$

avec

$$F \simeq 10^{-2} \text{ à } 50 \text{ MHz}, \quad F \simeq 0{,}25 \text{ à } 10 \text{ MHz}$$

pour L_N = 10 Km et les valeurs usuelles des paramètres ionosphériques.

Le premier terme entre crochets correspond à l'instabilité à deux faisceaux et le troisième à l'instabilité en champs croisés. Le milieu deviendra instable pour des vitesses de phase supérieures à la vitesse de seuil V_S donnée par l'équation (6) (FARLEY D. and FEJER B., 1975) :

$$V_s = C_S \left[(1 + F^2)^{1/2} - F \right] \tag{6}$$

avec $V_S = C_S$ pour les fréquences radar VHF et UHF correspondant aux grandes valeurs de k. Le terme en champs croisés qui est donc pratiquement négligeable en VHF et UHF (sauf pour de très forts gradients de densité électronique de type E sporadique) sera très important lors des observations de la gamme HF et sera stabilisateur ou déstabilisateur selon la direction respective du gradient de densité électronique et du vecteur $\vec{E} \times \vec{B}$. L'équation (6) donnant la vitesse de seuil a été représentée sur la Figure 1 en fonction du nombre d'onde et du gradient. On constate :

- pour les gammes VHF et UHF ($k > 1{,}2$), ou en l'absence de gradients importants de densité ($L_N > 50$ Km) le terme en champs croisés est négligeable et la vitesse de seuil est égale à la vitesse acoustique des ions C_S.
- pour la gamme radar HF et des gradients de densité habituels dans la région E ($L_N = 7$ Km), le terme en champs croisés modifiera très largement la vitesse de seuil pour les grandes longueurs d'ondes observées par radar HF. Ce terme sera déstabilisateur dans l'électrojet équatorial de jour où il abaissera alors largement la vitesse de seuil. Dans l'électrojet de nuit pour des gradients irréguliers (alternativement vers le haut et vers le bas) le terme sera respectivement stabilisateur et déstabilisateur. En période de contre-électrojet, le terme en champs croisés sera déstabilisateur et les grandes longueurs d'onde ne seront excitées que pour des vitesses électroniques très élevées.

En zone aurorale (Figure 2) la composante de gradient de densité parallèle au terme en $\vec{E} \times \vec{B}$ sera une composition du gradient latitudinal et de la projection dans cette direction du gradient vertical ce qui, compte-tenu de l'inclinaison du champ, conduira à une longueur d'échelle équivalente de l'ordre de 50 Km. L'effet en champs croisés déduit de la théorie linéaire sera ainsi moins important en zone aurorale qu'en zone équatoriale.

3. OBSERVATIONS DE L'ELECTROJET EQUATORIAL

En période d'électrojet de jour la théorie linéaire prévoit que dès que la vitesse de phase est supérieure à la vitesse de seuil, des instabilités doivent apparaître dont la vitesse de phase est donnée par la relation :

$$V_{ph} = \frac{V_{oe} \cos \Theta}{1 + \Psi} \qquad (7)$$

En réalité, deux types de spectres dont les caractéristiques contredisent ces prévisions sont observés :

a) les spectres de "type 1" apparaissent bien en VHF pour des vitesses de seuil égales à C_s, mais la vitesse de phase V_{ph} ne suit pas la relation (7) et reste constante et égale à C_s quels que soient V_{oe} et cos Θ (BOWLES K. et al., 1963),

b) les spectres de "type 2" sont observés en VHF alors que l'électrojet est linéairement stable. Ils correspondent à des vitesses de phase qui suivent paradoxalement la relation (7) et ils peuvent donc être utilisés pour mesurer les champs électriques (BALSLEY B.B., 1969).

Ces résultats ont été confirmés et étendus à l'aide d'observations récentes par radar HF :

c) des spectres de "type 1" apparaissent en HF à des vitesses de seuil égales à V_s (équation 7 - HANUISE C. et CROCHET M., 1979), leur vitesse de phase reste également limitée à la vitesse de seuil quels que soient V_{oe} et Θ. Les résultats multifréquence de la Figure 3 illustrent le bon accord entre l'expérience et la théorie pour la valeur des vitesses de phase des spectres de "type 1" égales à cette vitesse de seuil. Ils permettent donc de généraliser la notion de type 1.

d) les spectres de "type 2" observés en HF aux grandes longueurs d'onde sont similaires à ceux observés en VHF (HANUISE C. et CROCHET M., 1980).

Les caractéristiques des spectres Doppler de "type 1" et de "type 2" sont illustrées sur la Figure 4 pour des résultats obtenus en Afrique à 14 MHz.

On peut donc conclure que l'ensemble des résultats obtenus par radar VHF et HF dans l'électrojet équatorial conduisent à deux paradoxes qui ont été la source de très nombreux travaux théoriques:

- les ondes linéairement instables ("type 1") ne suivent pas l'équation de dispersion !
- les ondes linéairement stables ("type 2") suivent l'équation de dispersion !

4. OBSERVATIONS DU CONTRE-ELECTROJET ET DE L'ELECTROJET AURORAL

4.1 - Contre-électrojet

Lorsque le champ électrique se renverse en période de jour à l'équateur, le terme en champs croisés devient stabilisateur et le milieu ne devient instable que pour des valeurs très élevées de la vitesse électronique. Malheureusement, cette situation est exceptionnelle (un phénomène tous les cinq ans) comme on peut le déterminer à l'aide des observations magnétiques des stations équatoriales (HANUISE C. et al., 1979).

Cependant, après de nombreuses tentatives infructueuses un tel évènement a été détecté par radar en Ethiopie (CROCHET M. et al., 1979) (Figure 5). Les spectres Doppler observés illustrés sur la Figure 4 présentent les caractéristiques suivantes :

a) les spectres de "type 2" n'apparaissent plus,

b) les spectres observés ressemblent par leur largeur aux spectres de type 1 mais cette fois leur vitesse de phase suit la relation (7) (proportionnalité à la vitesse électronique et à cosΘ) comme illustré sur la Figure 6. Ces spectres ont été qualifiés de spectres de "type 0" pour prendre en compte l'adéquation de leurs caractéristiques avec les résultats de la théorie linéaire.

4.2 - Electrojet auroral

De nombreuses mesures radar ont été effectuées en Alaska (ECKLUND W.L. et al., 1973 ; TSUNODA R.T et al., 1974), Nouvelle Zélande (UNWIN R.L. and J. KEYS, 1975), Canada (HALDOUPIS G. and C. SOFKO, 1979) et en Scandinavie (GREENWALD et al., 1978), à l'aide de radar VHF et UHF. Les caractéristiques spectrales sont assez complexes et parfois difficiles à identifier à cause de la valeur élevée des vitesses entrainant des repliements du spectre aux fréquences radar élevées. Cependant les caractéristiques principales sont les suivantes (FEJER B. and M. KELLEY, 1980) :

- la vitesse peut dépasser largement la vitesse de seuil contrairement aux spectres de "type 1" équatoriaux

- la vitesse est proportionnelle à cos ϕ et à la vitesse électronique.

Ces caractéristiques semblent similaires à celles des spectres de "type 0" observés lors du contre-électrojet équatorial. Cependant, des résultats complémentaires indiquent que les phénomènes sont de nature différente :

- les échos sont observés dans la direction perpendiculaire au courant,

- deux composantes spectrales (large et étroite) peuvent coexister,

- lorsque le vecteur d'onde s'éloigne de la direction perpendiculaire au courant, un plateau est observé dans la vitesse de phase par les radars UHF.

5. BILAN DES RESULTATS ET TENTATIVES D'INTERPRETATION

Les résultats obtenus tant à l'équateur qu'en zone aurorale peuvent être classés en deux grandes catégories :

5.1 - Electrojet présentant une turbulence à grande échelle due à l'effet déstabilisateur de l'instabilité en champs croisés. C'est la situation qui prévaut à l'équateur avec deux types de spectres :

- les spectres de "type 2" linéairement stables en présence de grandes longueurs d'ondes linéairement instables qui peuvent dégénérer par cascade,

- les spectres de "type 1" linéairement instables mais saturés par un processus non linéaire.

5.2 - Electrojet sans turbulence grande échelle : c'est la situation du contre-électrojet et peut être de l'électrojet auroral avec un type de spectre ("type 0") correspondant à des ondes linéairement instables.

En réalité, malgré des similitudes apparentes, la situation est certainement plus complexe dans l'électrojet auroral que dans le contre-électrojet. En particulier, la présence d'ondes se propageant perpendiculairement au courant, alors que la théorie linéaire ne prévoit pas l'existence de ces ondes, a conduit certains auteurs (GREENWALD R., 1974) à suggérer un processus de cascade linéaire-linéaire, identique à celui proposé par SUDAN et al. (1973) dans l'électrojet équatorial. Dans ces conditions des ondes primaires instables sont tout d'abord générées dans un cône centré dans la direction du courant. Ces ondes primaires ne seront pas détectées par les radars VHF et UHF, seuls des radars HF pourraient éventuellement les détecter par réfraction lors d'une observation dans le plan du courant.Avec SUDAN et al. (1973), GREENWALD (1974) fait l'hypothèse que des irrégularités secondaires sont générées dans la direction perpendiculaire au courant si l'on prend en compte comme nouveau gradient local, le gradient de densité des irrégularités primaires. Dans ces conditions, la vitesse de seuil secondaire s'écrira :

$$V_{S2} = \frac{2\Omega_i \nu_e C_S^2}{\Omega_e \nu_i^2 A^2} \frac{k_S^2}{k_p} \qquad (8)$$

avec A amplitude de modulation des ondes primaires, k_S nombre d'ondes secondaires, k_p nombre d'ondes primaires. GREENWALD peut ainsi rendre compte des instabilités qui se propagent dans des directions perpendiculaires au courant.

L'essentiel des théories a été élaboré pour interpréter le paradoxe des résultats des spectres de "types 1 et 2" observés à l'équateur. Différents processus ont été successivement ou simultanément considérés pour interpréter ces paradoxes :

1) un effet quasi-linéaire (ROGISTER A., 1971 ; KAMENETSKAYA G., 1971) prenant en compte la diffusion des électrons, réduirait le champ de polarisation et la vitesse de phase à leur vitesse de seuil. Cette théorie particulièrement séduisante lorsque la vitesse de seuil observée en VHF égale à C_S semblait unique, ne peut en aucun cas rendre compte de la courbe de variation de la vitesse de phase des spectres de "type 1" observés en HF (Figure 3).

2) la prise en considération de la convection des ondes de plasma et de leur réfraction dans le profil de vitesse électronique (KAW P.K., 1972) laisse supposer que les radars ne détecteraient que les ondes de plus forte amplitude qui sont marginalement stables à la sortie de la zone d'instabilité. Deux considérations vont à l'encontre de cette théorie :

   a) les résultats lors du contre-électrojet devraient être de même nature que les résultats sur les spectres de "type 1",

   b) le taux de croissance linéaire d'ondes de type 1 convectives est trop important (LEE K. and C.F. KENNEL, 1973) et ces ondes doivent être saturées par un autre processus,

3) les théories non-linéaires et les simulations numériques ont par contre apporté des résultats fructueux pour l'interprétation des spectres de "type 2". En particulier les théories de la turbulence (SUDAN et al., 1973 ; KESKINEN M., 1978) rendent bien compte des caractéristiques de ces spectres.

Malheureusement en ce qui concerne les spectres de "type 1", les théories les plus sophistiquées (LEE K. et al., 1974) ont été élaborées en l'absence de turbulence grande échelle ; conditions dans lesquelles l'expérience a montré récemment que les spectres de "type 1" n'apparaissent pas mais où, par contre, des spectres de "type 0" doivent être observés.

En conclusion, si les interprétations théoriques et les simulations numériques sont satisfaisantes pour les spectres de "type 2", la situation des conditions de "type 1" et de "type 0" est beaucoup moins claire et des simulations numériques sont envisagées pour simuler ces différents résultats (KESKINEN M., communication privée). D'un point de vue physique, il est peu probable qu'un processus de limitation quasi-linéaire ou non-linéaire permette d'interpréter la variation de la vitesse de seuil des spectres de "type 1" en fonction du nombre d'ondes, par contre un processus convectif dans un milieu turbulent à grande échelle pourrait rendre compte de cette dépendance tout en s'affranchissant de la trop forte croissance des ondes de "type 1" convectives qui se rencontre dans le milieu laminaire. Par ailleurs, une meilleure connaissance théorique des processus auroraux semble nécessaire.

6. DETERMINATION DES CHAMPS ELECTRIQUES

La vitesse électronique et le champ électrique ont fréquemment été déduits de la mesure de la dérive des instabilités dans les électrojets. Divers cas doivent être considérés :

1) "Type 0" - En zone équatoriale, il semble que les caractéristiques des spectres de "type 0" observés lors d'un fort contre-électrojet soient proches de celles prévues par la théorie linéaire. Elles devraient donc permettre de déterminer le champ électrique ; malheureusement, ces situations ne sont qu'exceptionnelles.

2) "Type 2" - Les irrégularités de "type 2" ont des caractéristiques de vitesse de phase voisines de $V_e \cos \Theta$ comme les simulations numériques l'ont indiqué récemment mais ne suivent qu'une loi approchée (KESKINEN et al., 1979) et ne sont utilisables que pour des vitesses inférieures à 200 m/s

3) "Type 1" - Pour des vitesses élevées apparaissent des irrégularités de "type 1" dont la vitesse de phase est constante et dont seul l'angle d'ouverture pourrait être utilisé pour déterminer la vitesse électronique par la loi :

$$V_{eo} = C_s \, (1 + \Psi) \, / \cos \Theta$$

4) Zone aurorale - Les résultats de mesures radar et fusées simultanées (CAHILL et al., 1978) permettent d'accroître le degré de confiance dans les mesures radar alors que la compréhension physique et théorique des processus reste encore très limitée. Par ailleurs la validité de la méthode double impulsion telle qu'elle est utilisée dans l'expérience STARE impose des contraintes supplémentaires sur la forme et la pureté du spectre.

Des études complémentaires sont donc nécessaires sur la physique de la mesure par différentes méthodes, sur la nature des instabilités et sur les processus non-linéaires qui limitent la croissance des instabilités. Des comparaisons ultérieures sont envisagées entre radars VHF (STARE) et HF (SAFARI) et entre radars cohérents (STARE - SAFARI) et incohérents (EISCAT) en zone aurorale. En zone équatoriale, une expérience coordonnée est prévue en 1982 au Pérou mettant en oeuvre, radars (VHF et UF) et mesures in-situ par fusées.

7. CONTRIBUTION DES RADARS COHERENTS HF

Les radars cohérents HF ont dès à présent permis d'obtenir un certain nombre de résultats sur les instabilités de l'électrojet équatorial :

- exploration des grandes longueurs d'onde du spectre d'instabilité (5 à 40 m) (HANUISE C. and CROCHET M., 1978),
- vérification de la variation de la vitesse de seuil en fonction de la longueur d'onde (HANUISE C. and CROCHET M., 1979),
- généralisation de la notion de "type 1"
- détermination des caractéristiques des irrégularités de "type 2" aux grandes longueurs (plus aisément simulables numériquement qu'aux courtes longueurs d'onde),
- observation d'un contre-électrojet et identification des irrégularités de "type 0" (cette observation aurait pu être également effectuée par un radar VHF opérationnel en de telles circonstances),
- observation des ondes dans la direction du courant (CROCHET M. et al., 1979).

Des résultats analogues devraient pouvoir être obtenus en zone aurorale :

- exploration des grandes longueurs d'onde (déjà effectuée de 5 à 10 m ; CROCHET M. and HANUISE C., 1979),
- étude des instabilités dans la direction du courant par réfraction des ondes radioélectriques,
- identification d'autres instabilités (ion cyclotron, instabilités associées aux courants parallèles ...).

8. CONCLUSION

Les mesures récentes par radar HF et en période de contre-électrojet ont permis d'approfondir la connaissance des irrégularités dans l'électrojet équatorial. La contribution de la turbulence grande échelle due au terme d'instabilité en champs croisés déstabilisateur en temps normal, semble prépondérante aussi bien pour interpréter les caractéristiques des spectres de "type 2" que de "type 1". En l'absence de cette turbulence grande échelle, les spectres de "type 0" ont les caractéristiques prévues par la théorie linéaire.

En zone aurorale, la situation est encore obscure et un panorama global expérimental et théorique n'est pas encore disponible. Des travaux complémentaires, aussi bien expérimentaux que théoriques, sont donc nécessaires dans l'avenir pour résoudre les problèmes auroraux. Ces travaux devraient permettre d'apporter des prévisions sur la validité des mesures des champs électriques à l'aide des irrégularités de l'électrojet et sur ses fondements physiques.

REMERCIEMENTS

Les études par radar HF de l'électrojet équatorial en Afrique ont été rendues possible grâce au support du C.N.R.S. (E.R.A. 668), de l'I.N.A.G. (opération radar cohérent), de l'Université de Toulon et du Var, du Ministère des Affaires Etrangères, de la D.G.R.S.T. (Aide n° 75.7.1433) et de l'OTAN (Contrat de recherche multinationale n° 684).

REFERENCES

BALSLEY, B.B., 1969, "Some characteristics of non-two-stream irregularities in the equatorial electrojet", J. Geophys. Res., 74, p. 2333.

BALSLEY, B.B., 1973, "Electric fields in the equatorial ionosphere : A review of techniques and measurements" J. Atmos. Terr. Phys., 35, p. 1035.

BOWLES, K.L., BALSLEY, B.B., and COHEN, E., 1963, "Field-aligned E region irregularities identified with ion acoustic waves, J. Geophys. Res., 68, p. 2485.

BUNEMAN, O., 1963, "Excitation of field aligned sound waves by electron streams, Phys. Rev. Lett., 10, p. 285.

CAHILL, L.J., GREENWALD, R.A., and NIELSEN, E., 1978, "Auroral radar and rocket double probe observations of electric field across the Harang discontinuity", Geophys. Res. Lett., 5, p. 687.

CROCHET, M., July 1980, "Review of the equatorial electrojet instability in light of recent developments in HF radar measurements", 6th Symposium of Equatorial Aeronomy, Aguadilla, Puerto-Rico, submitted to J.A.T.P.

CROCHET, M., and HANUISE, C., September 1979, "Objectives and preliminary results of auroral multifrequency HF radar experiments", 3rd I.M.S. Workshop, Skokloster (Sweden).

CROCHET, M., HANUISE, C., and BROCHE, P., 1979, "HF radar studies of two-stream instability during an equatorial counter-electrojet", J. Geophys. Res., 84, p. 5223.

CROCHET, M., HANUISE, C., and HEDBERG, Å., September 1980, "SAFARI - A program for HF radar studies of plasma instabilities and electric fields in the auroral electrojet", 4th I.M.S. Workshop, Paris.

ECKLUND, W.L., BALSLEY, B.B., and GREENWALD, R.A., 1973, "Doppler spectra of diffuse radar auroras", J. Geophys. Res., 78, p. 4797.

FARLEY, D.T., 1963, "A plasma instability resulting in field-aligned irregularities in the ionosphere", J. Geophys. Res., 68, p. 6083.

FARLEY, D.T., 1979, "The ionospheric plasma", Ch III-1-7, in Solar System Plasma Physics, edited by C.F. KENNEL, L.J. LANZEROTTI, and E.N. PARKER, North-Holland, Amsterdam.

FARLEY, D.T. and FEJER, B.G., 1975, "The effect of the gradient drift term on the type 1 electrojet irregularities", J. Geophys. Res., 80, p. 3087.

FEJER, B.G., and KELLEY, M.C., 1980, "Ionospheric Irregularities", Reviews of Geophysics and Space Physics, 18, n° 2.

FEJER, B.G., FARLEY, D.T., BALSLEY, B.B., and WOODMAN, R.F., 1975, "Vertical structure of the VHF backscattering region in the equatorial electrojet and the gradient drift instability", J. Geophys. Res., 80, p. 1313.

FEJER, J., 1979, "Ionospheric instabilities and fine structure", J. Atmos. Terr. Phys. 41, p. 895.

GREENWALD, R.A., 1974, "Diffuse radar auroral and the gradient drift instability, J. Geophys. Res., 79, p. 4807.

GREENWALD, R.A., 1977, "Recent advances in the use of auroral backscatter to measure ionospheric electric fields, in Dynamical and Chemical Coupling, edited by B. GRAMAL and J.A. HOLTET, D. REIDEL, HINGHAM, Mass.

GREENWALD, R.A., WEISS, W., NIELSEN, E., and THOMSON, N.R., 1978, "STARE : A new radar auroral backscatter experiment in northern Scandinavia, Radio Sci., 13, p. 1021.

HALDOUPIS, C., and SOFKO, G., 1976, "Doppler spectrum of 42 MHz CW auroral backscatter, Can. J. Phys., 54, p. 1571.

HANUISE, C., and CROCHET, M., 1978 "Oblique HF radar studies of plasma instabilities in the equatorial electrojet in Africa, J. Atmos. Terr. Phys., 40, p. 49.

HANUISE, C., and CROCHET, M., 1979, "Marginal plasma waves in the equatorial electrojet observed by HF coherent radar techniques, in Wave Instabilities in Space Plasmas, edited by P.J. PALMALADESSO and K. PAPADOPOULOS, D. REIDEL, HINGHAM, Mass.

HANUISE, C., GHEBREBRHAM, O., GOUIN, P., and CROCHET, M., "1979, "Observations of the equatorial counter-electrojet in Ethiopia".

HANUISE, C., and CROCHET, M., July 1980, "5-50 meters wavelength plasma instabilities in the equatorial electrojet .1. - Cross-field conditions", 6th Symposium of equatorial aeronomy, Aguadilla, Puerto-Rico.

KAMENETSKAYA, G. Kh., 1971, "Quasi-linear theory of the formation of inhomogeneities in the equatorial electrojet", Geomagn. Aeron., 11, p. 71.

KAW, P.K., 1972, "Wave propagation effects on observation of irregularities in the equatorial electrojet", J. Geophys. Res., 77, p. 1323.

KESKINEN, M.J., 1978, "Numerical simulation and theory of strong ionospheric gradient drift turbulence", Ph. D. Thesis, Cornell Univ. Ithaca, New York.

KESKINEN, M.J., SUDAN, R.N., and FERCH, R.L., 1979, "Temporal and spatial power spectrum studies of numerical simulations of type 2 gradient drift irregularities in the equatorial electrojet, J. Geophys. Res., 84, p. 1419.

LEE, K, and KENNEL, C.F., 1973, "Effects of propagation parallel to the magnetic field on the type 1 electrojet irregularity instability, Planet. Space Sci., 21, p. 1339.

LEE, K., KENNEL, C.F., and CORONITI, F.V., 1974, "On the marginally stable saturation spectrum of unstable type 1 equatorial electrojet irregularities, J. Geophys., Res., 79, p. 249.

OSSAKOW, S.L., 1979, "Ionospheric irregularities", Reviews of Geophys. and Space Phys., 17, n° 4, p. 521.

RICHEMOND, A.D., 1973, "Equatorial electrojet, I, Development of a model including winds and instabilities, J. Atmos. Terr. Phys., 35, p. 1083.

ROGISTER, A., 1971, "Nonlinear theory of type 1 irregularities in the equatorial electrojet", J. Geophys. Res., 76, p. 7754.

ROGISTER, A., and D'ANGELO, N., 1970, "Type 2 irregularities in the equatorial electrojet", J. Geophys. Res., 75, p. 3819.

SUDAN, R.N., AKINRIMISI, J., and FARLEY, D.T., 1973, "Generation of small-scale irregularities in the equatorial electrojet, J. Geophys. Res., 78, p. 240.

TSUNODA, R.T., PRESNELL, R.I., and LEADABRAND, R.L., 1974, "Radar auroral echo characteristics as seen by a 398-MHz phased array radar operated at Homer, Alaska, J. Geophys., Res., 79, p. 4709.

UNWIN, R.S., and KEYS, J.G., 1975, "Characteristics of radio aurora during the expansive phase of polar substorms, J. Atmos. Terr. Phys., 37, p. 55.

WALDTEUFEL, P., 1965, "Etude de l'instabilité liée à la couche E sporadique équatoriale, Ann. Geophys., 21, p. 579.

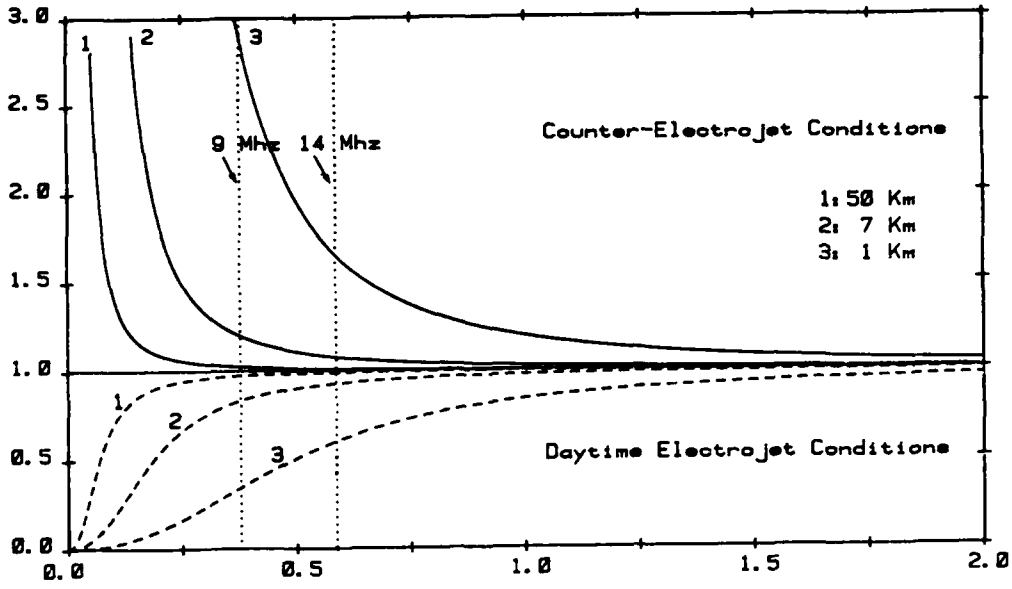


Figure 1 - Condition de seuil prévue par la théorie linéaire. Représentation de la vitesse de phase normalisée à la vitesse acoustique ionique C_S en fonction du nombre d'onde k et de la longueur d'échelle L_N du gradient de densité électronique pour un terme d'instabilité en champ croisé déstabilisateur (courbes du bas en pointillé) et pour un terme stabilisateur (courbes du haut en continu).

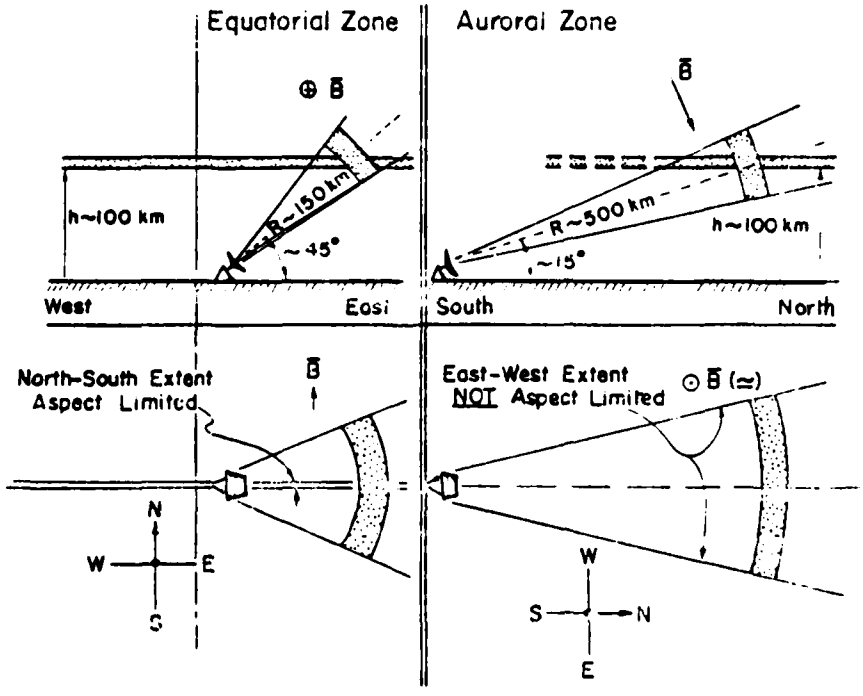


Figure 2 - Comparaison de la géométrie des champs en région équatoriale (gauche) et aurorale (droite) (d'après BALSLEY et ECKLUND, 1972).

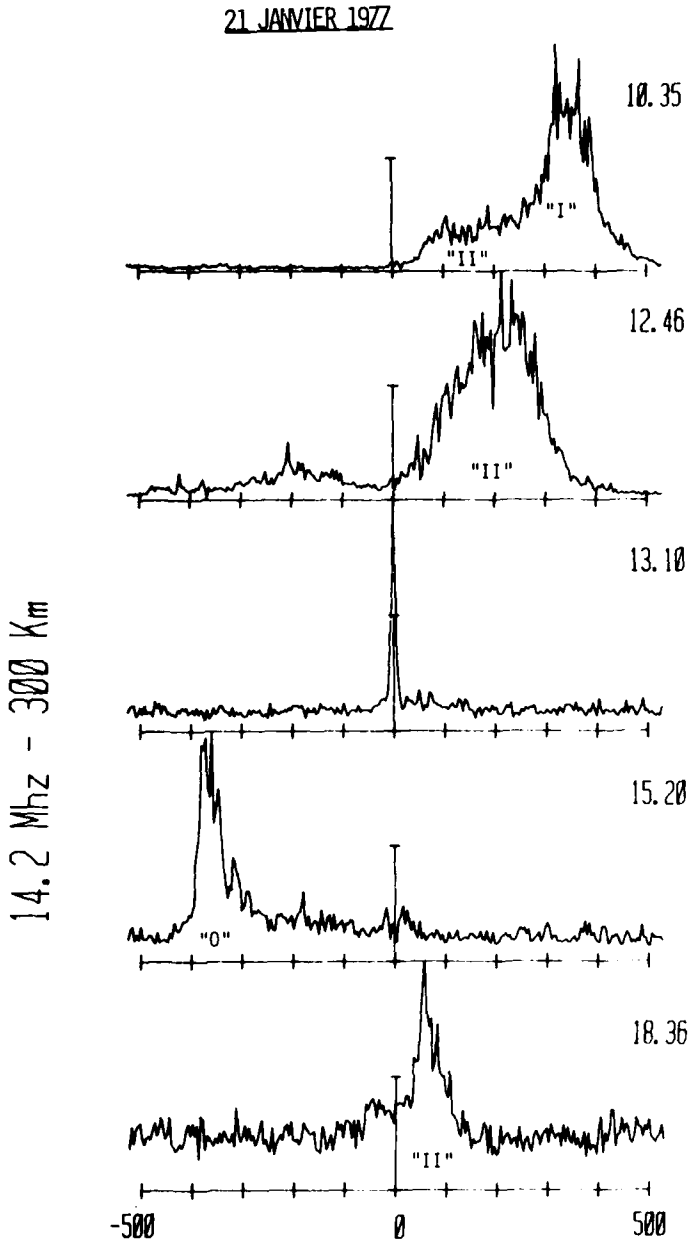


Figure 4 - Exemples des différents spectres Doppler observés à 14,2 MHz à Addis-Abeba (Ethiopie)

- électrojet fort : 10.35 LMT - type 1 + type 2
- électrojet faible : 12.46 LMT - type 2 seul
  18.36 LMT - type 2 seul
- contre-électrojet faible : 13.10 LMT pas d'écho
- contre-électrojet fort : 15.20 LMT - type 0

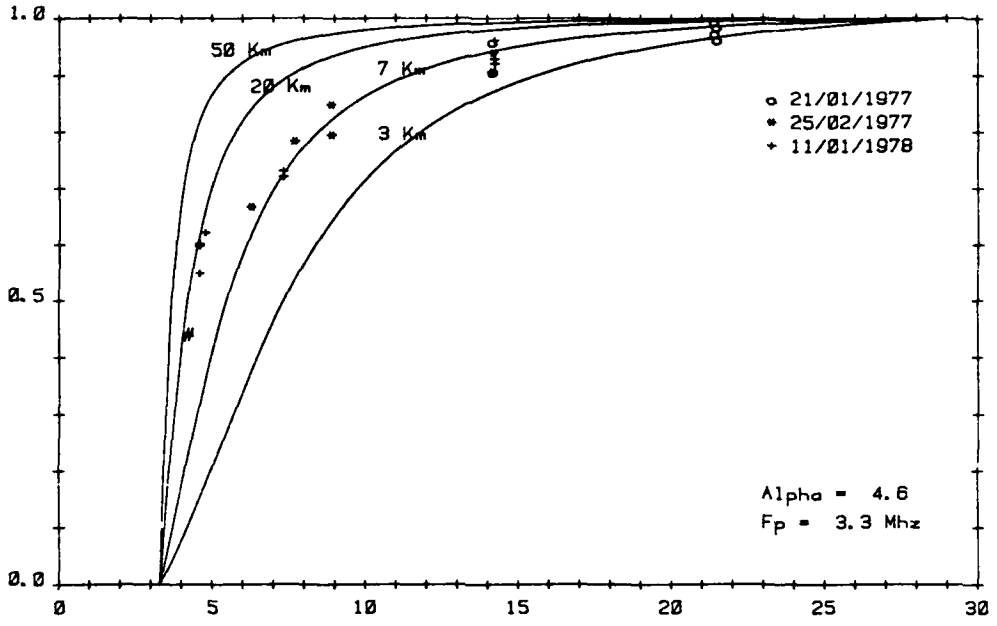


Figure 3 - Comparaison de la vitesse de phase (normalisée à 29 MHz) des irrégularités de type 1 mesurée par radar HF à la vitesse de seuil prévue par la théorie linéaire en fonction de la fréquence radar.

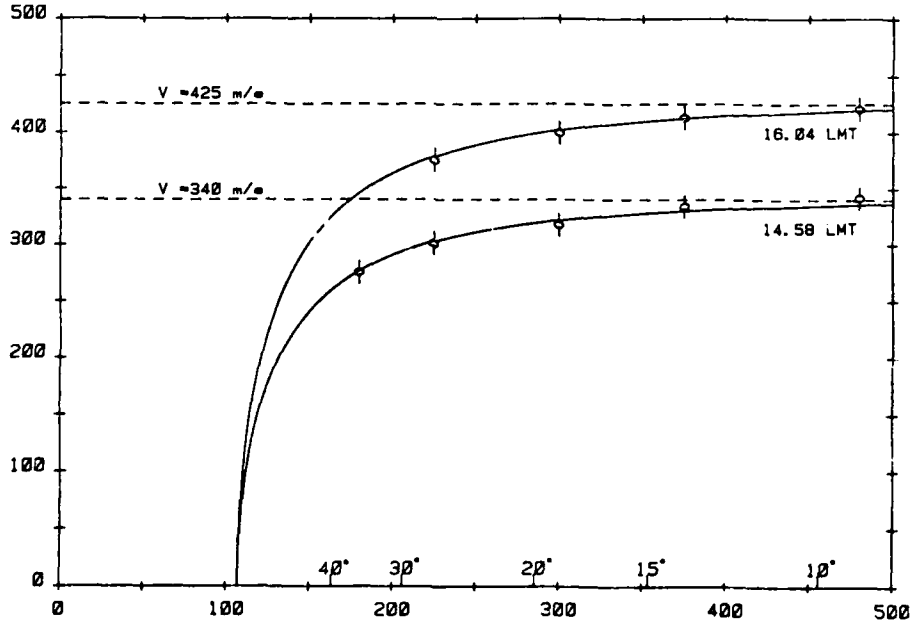


Figure 6 - Variation de la vitesse de phase des irrégularités de type 0 comparée à la loi en cos θ prévue par la théorie linéaire en fonction de l'angle θ et de la distance de groupe (en km) en abscisses.

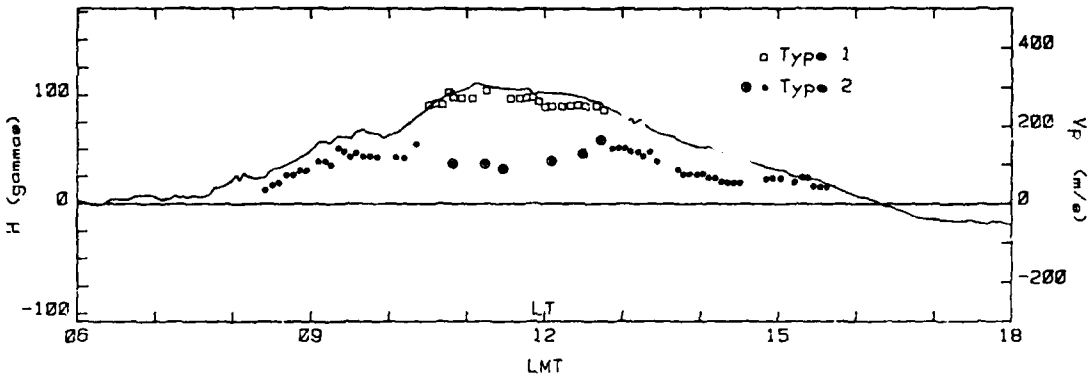


Figure 5 a) - Comparaison des mesures magnétiques et radar (type 1 + type 2) un jour d'électrojet normal à Addis-Abeba - Ethiopie (21.7.1977).

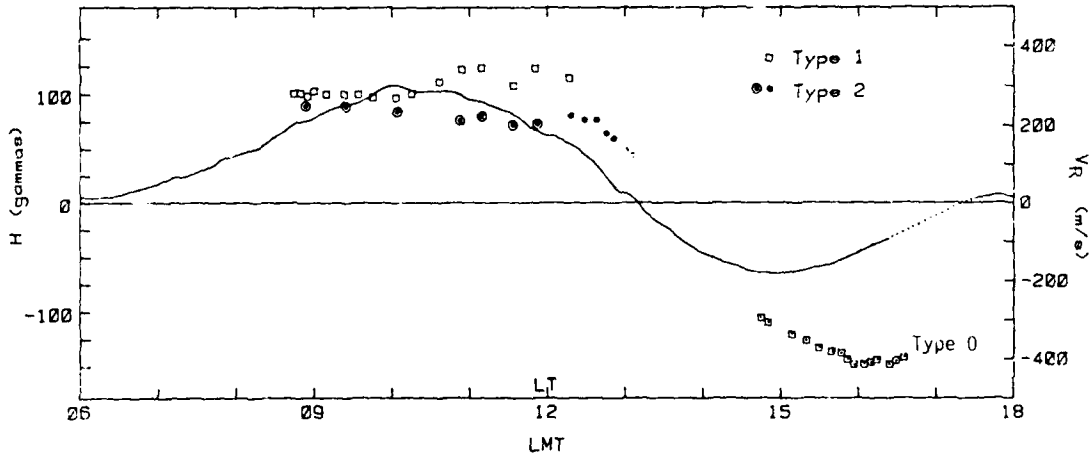


Figure 5 b) - Comparaison des mesures magnétiques et radar (type 0, type 1, type 2) le 21.1.1977 à Addis-Abeba - Ethiopie-lors d'un renversement de l'électrojet.

DYNAMICS OF THE DAYSIDE CUSP AURORA

A. Egeland
University of Oslo, Norway

P.E. Sandholt & K. Henriksen,
University of Tromsø. Norway

and

C.S. Deehr & G.G. Sivjee
University of Alaska, Fairbanks, Alaska

ABSTRACT

A multi-national auroral project for investigations of the dayside cusp aurora from Svalbard has been carried out since the winter of 1978/79. The characteristics of the cusp emission have been studied and the results are summarized. The occurrence and duration of discrete auroral forms superimposed on the weak, permanent 630 nm belt have hitherto not always been separated from the clean cusp aurora. Special attention is given to the dynamics of the dayside cusp aurora and the simultaneous occurring aurora in the night sector in relation to the day- and night-side magnetic activity. The conclusion is that a day-side current system exists and that it is associated with the position of the cusp aurora. The intensity and movements of this current system are correlated with the nightside electrojet. It is suggested that a discontinuity similar to the Harang discontinuity exists in the dayside cusp around magnetic noon.

1. INTRODUCTION

Only ground observations allow continous monitoring of location and dynamics of aurora with high resolution in both time and space. A multinational project for optical observations of the dayside cusp aurora from Svalbard - the only accessible region in the northern hemisphere which satisfies the observational conditions - was started during the winter 1978/79 (cf. Deehr et al., 1980). The project was continued for another observing season last winter and plans are being laid for three additional years.

The possibility of carrying out simultaneous, coordinated optical measurements from two sites (Longyearbyen, 78.2°N, 15.6°E; Ny Ålesund, 78.9°N, 11.9°E), on the same geomagnetic meridian and separated by 110 km, is another advantage at Svalbard. In addition, the stations are in the trans-polar meridian chain of auroral observatories extending from Alaska through Canada and into Europe through Scandinavia. This latter situation allows simultaneous day- and night-side observations, which are needed in order to understand and explain the results.

In this paper we will first briefly summarize some characteristics of the dayside cusp aurora based on recent observations. For a more detailed review the reader is referred to Shepherd (1979) and references mentioned therein. Special attention will be given to the dynamics of the cusp aurora in relation to simultaneous optical observations at night and geomagnetic measurements both day and night.

2. OPTICAL SIGNATURES OF DAYSIDE AURORA

The results presented in this section are obtained by combining the observations from the multi-channel, meridian-scanning photometers, the spectrometers, and the all-sky cameras. An overview of the instrumental set-up, the geographic and geomagnetic coordinates of sites from where data are used, and the participating research institutions for the expedition 1978/79 are listed in Table I. For a more detailed description of the first results, the reader is referred to Deehr et al. (1980), Sandholt et al. (1980), and Sivjee et al. (1980).

The dayside cusp aurora is characterized as a more or less stationary, subvisual belt with the dominant emission at 630 nm, centered around magnetic noon. This enhanced red auroral OI emission is a consequence of the soft particle precipitation (typically 100 eV) (cf. Winningham et al., 1973; Doering et al., 1976; Potemra et al., 1977).

The half-width of this red luminosity is normally between 0.5 and 3 degrees invariant latitude, with an average value close to 1 degree (cf. Fig. 1). Thus, the aurora seems to be more narrow in structure than satellite particle-data indicate. During quiet conditions around magnetic noon the 630 nm intensity is typically between 0.2 and 1 kR, and therefore well below our visible threshold.

An unusually intense (above 30 kR) red-dominaed aurora was observed 30th November 1980 (cf. Fig. 2). This spectacular cusp aurora was associated with a main geomagnetic substorm following a sudden commencement (cf. Sect. 4).

Generally, the intensity ratio 630 nm/557.7 nm is well above 1, and particularly for a short period around geomagnetic noon (i.e. approximately 09 UT), this ratio approaches 10 or occasionally even 30. This is associated with the "midday-gap", when the auroral green line at 557.7 nm is very weak (Cogger et al., 1977).

The accurate location and brightness of this red auroral belt can best be studied by coordinated multi-channel, meridian-scanning photometers. Based on such observations, isophot plots of the type shown in Fig. 3 are obtained. Here the red emission is shown as function of time and latitude. (In this drawing the average height was assumed to be 225 km for the red line.) The position of the cusp is shown by the heavy black line in Fig. 3 and the intensity by the various shadings.

During magnetic quiet conditions, the stationary or quasi-stationary red belt is located around 76 to 80 degrees invariant latitudes, and it is normally recognized for about 8 to 10 hours centered around geomagnetic noon (cf. Deehr et al., 1980).

Occasionally, visible, rayed arcs and bands - with lifetimes of 5 min or less - are observed within the cusp. These transient intensity enhancements are most frequent in the local magnetic morning and early afternoon. Several of these events are observed simultaneously with nightside magnetic activity (cf. Figs. 1 and 4). In these visual, transient forms the intensity of 557.7 nm (OI) and 427.8 (N_2^+ 1NG) rises to several kR, while the hydrogen emissions seldom exceed 50 R.

For a discussion of the spectral characteristics of cusp aurora, cf. Shepherd (1979) and Sivjee et al. (1980). Here we will only mention a few general features. Because the aurora is produced by low energy electrons interacting with atmospheric constituents in the F-region and above, enhancements of atomic emissions (i.e. the dominating constituents at the stopping altitude) and a redistribution of vibrational and rotational molecular emissions are typical for dayside aurora. Some of these interactions leave atmospheric constituents in metastable states. Since collision frequencies in the F-region are much smaller than in the E-region, some of these excited species in metastable states have a change to shed the excess energy through radiative decay rather than losing it via collisional quenching. Hence, in the cusp auroras, unlike in the nighttime auroras, the intensities of light from forbidden transitions are relatively high and easily detectable.

In general, the hydrogen emission lines are very weak (typical 20 R) implying low energy proton precipitations.

By combining the Ny Ålesund and Longyearbyen measurements, it is possible to determine the height of auroral emissions by triangulation. Because the 630 nm emission is diffuse in character and typically covers one degree in latitude, it is difficult to obtain reliable heights for the red cusp aurora. A few events of more short-lived discrete arcs with intensities above 1 kR have been triangulated. Based on these calculations the following conclusions can be drawn: The height of maximum 630 nm emission varies between roughly 200 and 250 km, while the peak of the N_2^+ 1NG and the green oxygen line at 557.7 nm are normally found between 140 and 160 km. The average widths of these discrete forms are less than 50 km.

It should also be pointed out that sunlit cusp aurora above 500 km have been found in a few events. One example of auroral emissions from a partly sunlit atmosphere is shown in Fig. 5. As illustrated in this drawing the auroral form is divided in a sunlit upper part and a lower part emitted from the dark atmosphere.

The spatial relationship of 427.8 nm, 557.7 nm, and 630 nm emissions is different from that found on the nightside. The main 427.8 nm emission region is often at higher altitudes than the 557.7 nm region. This is due mostly to resonance-scattered sunlight. However, impact ionization cannot account for all of the N_2^+ population and the process $O^+(^2D)+N_2 \rightarrow N_2^+(A^2\Sigma)+O$ is also important for the production of N_2^+ in the quiet, midday cusp.

3. SIMULTANEOUS AURORAL ACTIVITY IN THE DAY AND NIGHT SECTORS

The size, shape and location of the auroras reflect the forces acting on the auroral particles on their way into the earth's atmosphere. Unfortunately, it is difficult to accumulate enough coordinated auroral observations from a large enough density of ground stations to study simultaneous day and night variations along the oval.

The "trans-polar" chain of stations (cf. Sect. 1) allows observations of auroral activity in the day and night sectors, simultaneously. Cusp data from Svalbard have recently been compared with optical measurements from Alaska by Deehr et al. (1980b), and one example is shown in Fig. 6. The photometer scans of 630 nm for December 13, 1979 from Longyearbyen (LYR) and Poker Flat (PKR) have been aligned to show north-south (N-S) motion of auroral forms - from a single station - as function of local time. The N-S scans are "stacked" in Universal Time with magnetic midnight and noon displaced by 2 hrs because the two stations are not 180° apart either geographically or geomagnetically.

The most striking feature of the simultaneous observations is the simultanity and similarity of the auroras. The slow curve of the maximum through noon is the passage of the station under the oval, but the poleward expansion on the nightside coincides with equatorward expansion on the dayside. Other examples show that the quiet period can occur $\pm$ 3 hrs on either side of magnetic noon.

These observations indicate a more intimate connection between the night- and dayside aurora than one would anticipate assuming the separate and direct access of magnetosheath particles to the atmosphere provided by the magnetospheric cleft.

4. LATITUDINAL MOVEMENTS OF CUSP AURORA IN RELATION TO DAY- AND NIGHT-SIDE MAGNETIC SUBSTORM ACTIVITY

In this section we will first go through one event in fairly great details, while the general findings will be summarized in the next subsection. A more detailed description is found in an article by Sandholt et al. (1980) and in a manuscript by Sandholt (1980).

4.1. The January 15, 1979 substorm event

In Fig. 7 we have drawn some of the key curves for this event. The H-components from the midnight sector (the upper two panels) show a substorm event with the main phase occurring between 11 and 1145 UT. This event started with a less marked growth signature (characterized by a slow negative trend in the H-component) while the recovery phase follows the general substorm pattern.

The simultaneous dayside magnetic disturbances are presented in the middle of the figure. This is a post-noon event (magnetic noon is at 09 UT). We notice that the signature of the dayside-substorm is different from the nightside and no marked onset of an expansion phase is observed on the dayside.

The lower panel shows the width and the movements of the dayside cusp aurora as seen from our sites at Svalbard. The latitudes of the magnetic observatories on the dayside are marked by dotted lines in the figure. During the growth phase the disturbance appears to the north of Ny Ålesund ($\Delta H < 0$ and $\Delta Z < 0$) but moves equatorward and passes over our station close to 11 UT. A westward directed equivalent current characterize the main phase. Notice that the center of this current system is on the poleward side of the dayside cusp aurora.

After 1130 UT a positive bay in the H-component starts at Tromsø while the Z-component there is close to zero. From this we infer the growth of an equivalent current directed toward east on the equatorward side of the red auroral arc. When the arc returns to higher latitudes during the phase of recovery it crosses the latitude of Bjørnøya at about 1245 UT. At this time the direction of the disturbance vector detected at Bjørnøya changes from southward to northward (negative to positive bay). (The corresponding change at Hornsund takes place shortly after.) We can thus follow the movement and direction of the dayside current system fairly accurately.

At Ny Ålesund an increase in the parameter $\Delta Z/\Delta H$ is recognized. These signatures indicate that the whole system of equivalent currents moves northward during the recovery phase. Following this interpretation, the eastward equivalent current to the south of the dayside aurora moves into the field of view of the magnetometer at Bjørnøya from 1245 UT.

Notice that the negative and positive bays recorded at Bjørnøya are of similar magnitude ($\sim$ 100 nT). Furthermore, the magnetic bays at Ny Ålesund and Tromsø in the post-noon sector are also close to 100 nT, while the simultaneous nightside disturbance is typical 500 nT.

As Fig. 7 shows the equatorward movements of the cusp started before the substorm onset as seen in the midnight and post-midnight sector. Furthermore, it should be mentioned that discrete auroral forms are frequently seen in the vicinity of the sector where the magnetic disturbance vector reverses.

4.2. Dynamics of the cusp aurora in relation to magnetic activity

The magnetic nightside recordings used here are carried out at two latitudinal chains of stations - one in Alaska (operated by University of Alaska) and the other in Canada (operated by University of Alberta) - separated by roughly 30 degree in longitude. The coordinates for some main stations together with the three dayside magnetic observatories are listed in Table I. Even if we don't have a complete global magnetic coverage, the number of stations used is better than in earlier studies if this type.

The magnetic disturbances observed are not only affected by the ionospheric electrojet, but effects of Birkeland and induced currents are also superimposed. Thus, the conclusions regarding ionospheric currents - based on ground observations only - are questionable. However, a combination of several ground stations gives valuable information about direction and position of equivalent currents systems. With these limitations in mind, the following general findings of cusp, auroral dynamics in relation to magnetic activity should be stressed.

1. An ionospheric current system exists in the noon and post-noon sector of the dayside cusp. This current system is closely connected to the nightside auroral dynamics and the movements of the dayside cusp aurora.

2. The magnetic substorm signatures observed within the cusp are typically a factor of five to ten less intense than the simultaneous activity in the night sector.

3. The intensity of the dayside magnetic disturbance is enhanced in association with magnetic substorms in the night. The dayside current system moves in latitude with the optical aurora - equatorward during the growth and main phase and back poleward during the recovery phase. This motion is correlated in time with the dynamic of the nightside electrojet. An approximately one degree southward shift per 100-150 nT negative bay at the position of maximum electrojet current on the nightside is observed.

4. The horizontal perturbation vector reverses from negative (southward) to positive (northward) when we move along a meridian chain from north to south across the cusp. This magnetic discontinuity seems to occur near the latitude of maximum auroral luminosity. The current reversal is therefore associated with the latitudinal position of the cusp aurora.

5. The equatorward shift of the cusp aurora starts at the early phase of the nightside substorm; typically 0.5 - 1 hours before the main expansion.

6. During the recovery phase the cusp aurora contracts to higher latitudes. The westward current on the poleward side of the oval decreases in intensity when it moves northward again in the late recovery phase. The eastward current on the equatorward side of the red auroral belt is moving poleward during this phase.

7. The observed dayside magnetic discontinuity resembles the Harang discontinuity occurring in the evening and midnight sector (cf. e.g. Heppner, 1972; Schoufield and Nielsen, 1980). However, the occurrence and the latitudinal profile of this discontinuity seems to be somewhat different from night to day.

5. DISCUSSION AND CONCLUSION

The dynamics of the cusp aurora together with the intimate connection between night- and dayside aurora can neither be explained only as due to magnetosheat plasma extending through the magnetospheric cleft into the F-region - nor as a result of longitudinal particle drift from the nightside. (This conclusion does not reflect the more energetic particles causing auroral absorption along the auroral zone; cf. e.g. Berkey et al., 1974. An expansion from Alaska to Svalbard (roughly 5.000 km) would give time delays of the order of one hour, while our observations suggest that the dayside emissions occur simultaneously (within a few minutes) with the substorm features on the nightside.)

The latitudinal motion of dayside cusp aurora in relation to simultaneous nightside magnetic activity was also pointed out by Eather et al. (1979). However, their model did not include dayside magnetic activity. Our data indicate that the observed dayside current system and its close relation to the nightside substorm must be crusial in any realistic model. Also the connection between ionospheric currents and auroral dynamics in the dayside cleft must be included.

There is some evidence for the existence of another source region - the plasma mantle - responsible for structurized dayside aurora and the associated Harang-type discontinuity. Deehr et al. (1980b) have suggested a current system - patterned after a model by Yasuhara et al. (1975) - which provides equatorward motion on the dayside simultaneously with the poleward expansion on the nightside. However, more detailed studies - both experimental and theoretical - are needed in order to explain the complex dynamics of the dayside cusp and its intimate connection with the nightside substorm activity. Because a high-latitude two-cell convection pattern is known to exist (cf. e.g. Heppner, 1977), one would probably expect a Harang type discontinuity also in the morning sector.

REFERENCES

1. Berkey, F.T., V.M. Driatskiy, K. Henriksen, B. Hultqvist, D.H. Jelley, T.I. Shchuka, A. Theander, and J. Yliniemi, 1974: A synoptic investigation of particle precipitation dynamics for 60 substorms in IQSY (1964-1965) and IASY (1969). Planet. Space Sci., 22. 255.

2. Cogger, L.L., J.S. Murphree, S. Ismail, and C.D. Anger, 1977: Characteristics of dayside 5577 Å and 3914 Å aurora. Geophys. Res. Lett. 10, 413.

3. Deehr, C.S., G.G. Sivjee, A. Egeland, K. Henriksen, P.E. Sandholt, R. Smith, P. Sweeney, C. Duncan, and J. Gilmer, 1980: Ground-based observations of F-region aurora associated with the magnetospheric Cusp, J. Geophys. Res. 85, 2185.

4. Deehr, C.S., G.J. Romick, and G.G. Sivjee, 1980b: Auroral Morphology: in Exploration of the polar upper atmosphere (eds. Deehr and Holtet). 259-266- Reidel Publishing Comp., Dordrecht-Holland.

5. Doering, J.P., T.A. Potemra, W.K. Peterson, and C.O. Boström, 1976: Characteristic energy spectra of 1- to 500-eV electrons observed in high latitude ionosphere from Atmosphere Explorer C., J. Geophys. Res. 81, 5507.

6. Eather, R.H., S.B. Mende, and E.J. Weber, 1979: Dayside aurora and relevance to substorm current system and dayside merging, J. Geophys. Res., 84, 3339.

7. Heppner, J.P., 1972: The Harang discontinuity in auroral belt ionospheric currents, in Geofysiske publikasjoner, 29. 105.

8. Heppner, J.P., 1977: Empirical models of high-latitude electric fields, J. Geophys. Res. 82, 1115.

9. Potemra, T.A., W.K. Peterson, J.P. Doering, C.O. Boström, R.W. McEntire, and R.A. Hoffman, 1977: Low-energy particle observations in the quiet dayside cusp from AE-C and AE-D, J. Geophys. Res. 82, 4765.

10. Sandholt, P.E., 1980: Priv. communication.

11. Sandholt, P.E., K. Henriksen, C.S. Deehr, G.G. Sivjee. G.J. Romick, and A. Egeland. 1980: Dayside cusp auroral morphology related to nightside magnetic activity, J. Geophys. Res., in press.

12. Shepherd, G.G., 1979: Dayside cleft aurora and its ionospheric effects, Rev. Geophys. and Space Phys., 17, 2017.

13. Sivjee, G.G., C.S. Deehr, and K. Henriksen, Ortholhelium emissions at 3889 Å and 5876 Å in the polar upper atmoshere, J. Geophys. Res., in press.

14. Winningham, J.D., S.-I. Akasofu, F. Yashuhara, and W.J. Heikkila, 1973: Simultaneous observations of auroras from the south pole station and of precipitating electrons by Isis 1, J. Geophys. Res. 78, 6579.

15. Yasuhara, F., Kamide, Y., and Akasofu, S.I., 1975: Planet. Space Sci., 23. 575.

TABLE 1. Instrument Type, Deployment, and Purpose for the Svalbard Expedition 1978-1979.

Instrument	Institution	Location	Purpose
All-sky camera (35 mm, black and white)	GI	LYR	Location and orientation of auroral arcs and bands.
All-sky camera (16 mm, color)	UT	NYA	
Meridian-scanning photometer	GI	LYR	Mapping of local time position of red (6300 Å) cusp and discrete bands and associated characteristic energy of precipitating particles.
Meridian-scanning photometer	US	LYR	
Meridian-scanning photometer	NIKF	NYÅ	
Meridian-scanning photometer	GI	PKR, FYU	
Zenith LLLTV	NIKF	LYR	Morphology of fast-moving bands
Fabry-Perot interferometer (6300 Å)	UP	LYR	Doppler wind and temperature of neutral *F* region.
1-m spectrophotometer (3000-8700 Å)	GI	LYR	Spectroscopic studies of various atmospheric emissions
½-m spectrophotometer	GI	LYR	
Magnetometer	UT	NYÅ	Determination of substorm-related effects
Magnetometer	UT	BJØ	
Magnetometer	UT	TOS	
Magnetometer chain (seven instruments in Alaskan Chain)	GI	AK	

		Location Coordinates			
		Geographic		Geomagnetic	
	Symbol	Latitudes	Longitudes	Latitudes	Longitudes
Geophysical Institute	GI				
Norwegian Institute of Cosmic Physics	NIKF				
Ulster Polytechnic	UP				
University Saskatchewan	US				
University of Tromsø	UT				
Ny Ålesund	NYÅ	78.9	15.6	75.9	114.7
Longyearbyen	LYR	78.2	11.9	73.1	129.4
Bjørnøya	BJØ	74.5	19.2	71.3	110.5
Tromsø	TOS	69.7	19.0	66.0	105.2
Alaskan Chain	AK	63-80	~150	62.9-89	~260
Poker Flat	PKR	64.9	148.0	64.9	260.3
Fort Yukon	FYU	66.6	145.3	67.0	260.7

All meridian-scanning photometers include 4278-A N_2^+, 6300-A O I, 5577-A O I, and 4861-A H beta.

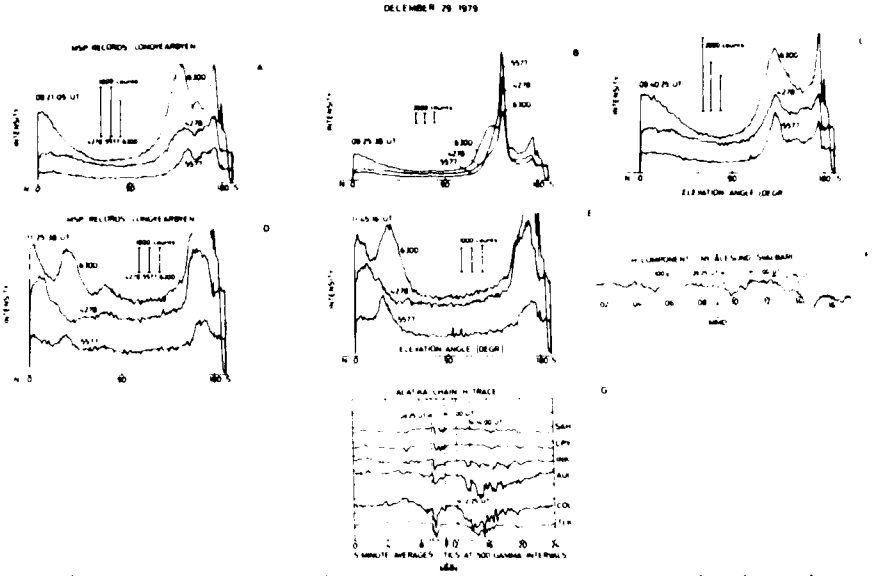


Figure 1: Dayside photometer observations from Longyearbyen. The relative intensity vs. elevation is shown for three of the channels. Panels A, C, D, and E represent the red-dominated cusp aurora. The corresponding magnetic activity is shown in curves F and G for Ny Ålesund and the nightside stations, respectively. While curves D and E refer to quiet conditions with the cusp poleward of Longyearbyen, the red-belt is south of the site in the first 3 panels when the magnetic activity is high. In scan B a strong auroral burst is recorded superimposed on the cusp, and the intensity ratio is more similar to night aurora. This is likely due to keV-electron precipitation in connection with distinct auroral forms. Universal time (UT) is used together with abbreviation MMD=magnetic local midday and MMN=magnetic local midnight (for the Alaskain chain of stations). The calibration factors are: 630 nm=0.93 R/count, 427,8 nm=0.3 R/count, 557.7 nm=0.9 R/count, and H_β(486.1nm)=1.4 R/count.

Figure 2: All-sky pictures of dayside cusp aurora from the Ny Ålesund station.

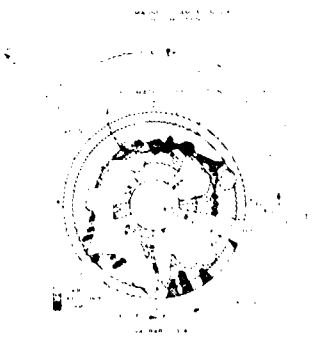


Figure 3: Isophot plot of 630 nm (OI) emission as function of time and latitude. The height of the emission was assumed to be 225 km. The positions of the cusp is shown by the heavy black line (cf. also Deehr et al., 1980).

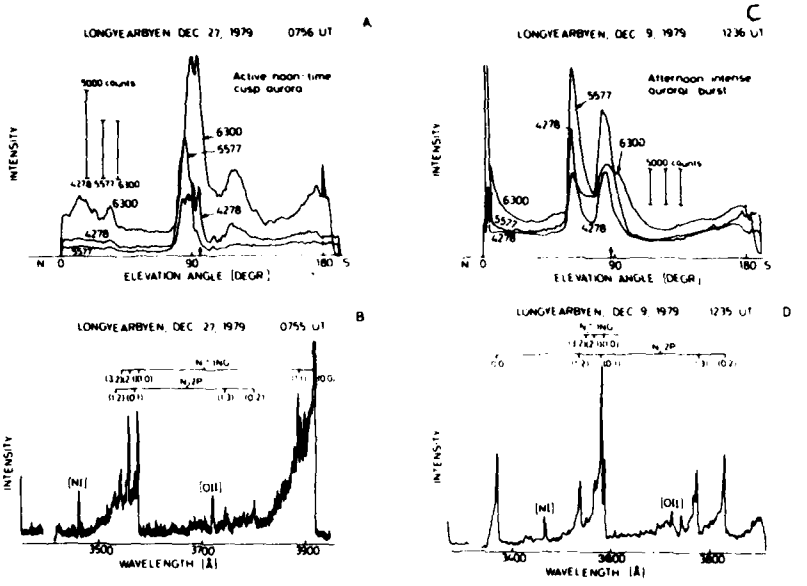


Figure 4: Photometer and spectrometer observations of a strong noon-time aurora (panel A) and an intense post-noon burst (panel C). The UV-spectrum (340-400 nm) is plotted in curve B and D. We recognize the vibrational enhancement of the N_2^+ 1NG (2,1) and (3,2) bands in Spectrum B compared to D. The spectrum of the post-noon structures (D) are more like that of nighttime aurora. The N_2 second positive band is enhanced and dominate the N_2^+ 1NG bands in the cusp rayed arcs and bands. In both B and D the spectrometer was pointing towards zenith and the integration time was approximately 8 min.

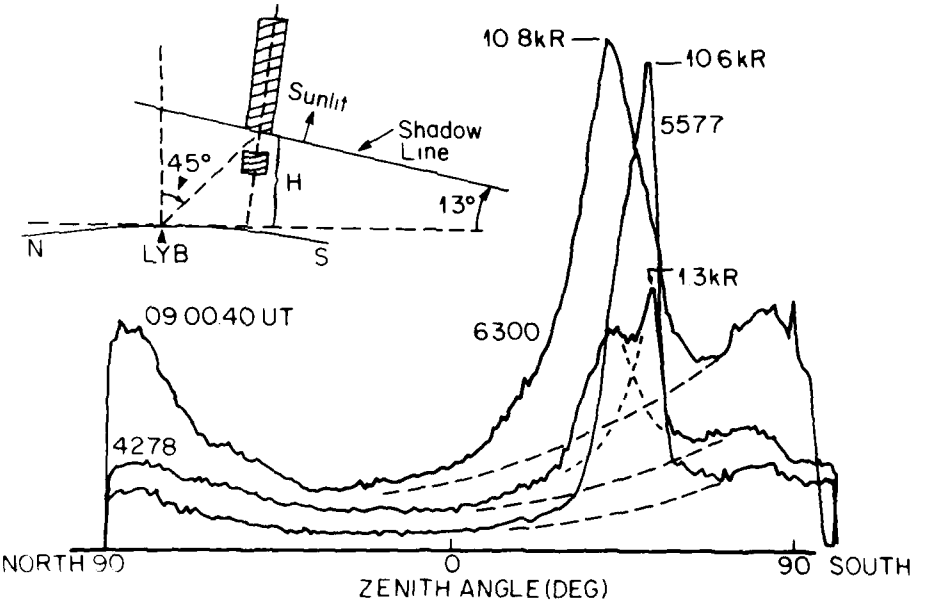


Figure 5: Photometer observations of dayside cusp aurora where three different emissions are plotted vs. elevation angle. Peak intensities are also listed. In the upper left part of this figure a sketch of the partly sunlit aurora is plotted.

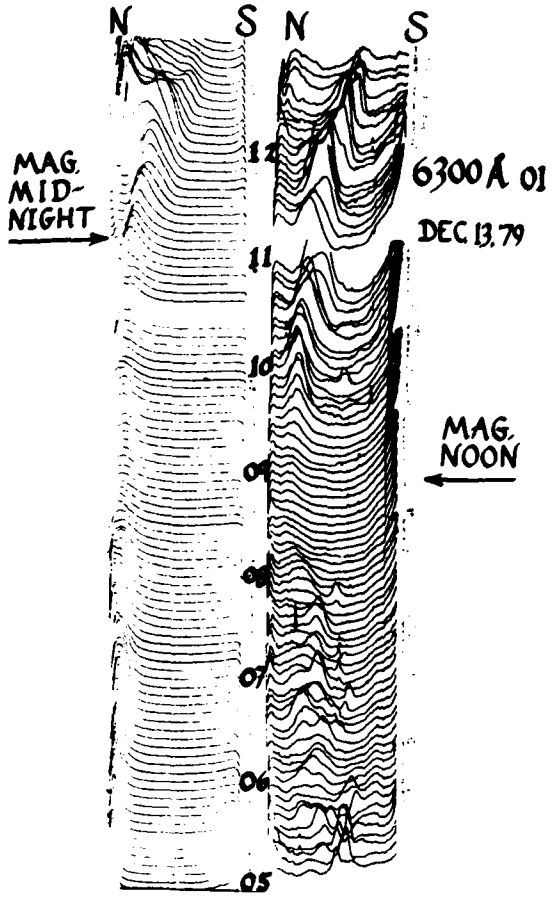


Figure 6: A comparison of simultaneous meridian photometer scans from dayside (LYR) and nightside (PKR) stations. The vertical scale is hours in UT. Each scan represents the slant intensity of 630 nm (OI) emission in the N-S meridian of the station.

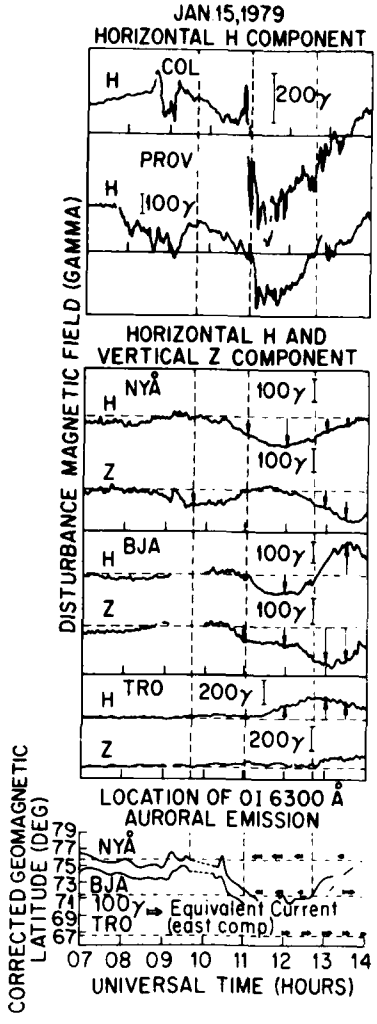


Figure 7: This figure illustrates the latitudinal movements of the cusp aurora in relation to night and daytime magnetic activity. The two upper curves show the H-component as observed from the nightside covering an area of 40 degrees in longitude. The middle panel shows the Z and H-component from the dayside stations. The latitude position of the dayside cusp aurora vs. UT is shown in the bottom curve. The transformation from elevation angle (MSP-records) to geomagnetic latitude in this plot is based on the assumption that the height of maximum volume emission rate of the red cusp aurora (630 nm) is 200 km.

LARGE-SCALE WAVES IN THE IONOSPHERE OBSERVED BY THE AE SATELLITES

S.H. Gross
Polytechnic Institute of New York
Farmingdale, New York 11735, U.S.A.

C.A. Reber
NASA, Goddard Space Flight Center
Greenbelt, Maryland 20771, U.S.A.

F. Huang
Computer Sciences Corp.
Silver Spring, Maryland 20910, U.S.A.

ABSTRACT

We report here on efforts utilizing Atmospheric Explorer (AE) satellite data to establish whether coherent waves in the gravity wave range are present in both neutral and ionized media in the thermosphere. AE-C data in particular are shown (altitude 250km, circular orbit inclined 68°). Data consists of in situ argon, helium, nitrogen and oxygen densities, plasma density and ion and electron temperatures. Filtering provides the fluctuation signals for each which are spectrum analyzed for power and cross spectra. The observed frequencies are essentially proportional to the spatial wavenumbers along the satellite track. Scale sizes range from thousands to tens of kilometers.

Modeling efforts to determine the temporal characteristics, direction of propagation and other wave parameters from the spectra were made with little success. Unfortunately, efforts to find near simultaneous ground measurements to corroborate wave characteristics have been fruitless so far. On reviewing the data it was found that turbulence-like structure over the full range of wavelengths is present with different spectral indices for each measured density and temperature. These appear to be consistent in all data studied to date. Wave packets are present superimposed on top of this turbulence structure in all data channels, (densities and temperatures), many of which exhibit excellent correlation with each other.

1. INTRODUCTION

Since the hypothesis of Martyn (1950) 30 years ago, it has been generally accepted that many, if not all traveling ionospheric disturbances (TID's) originate from gravity waves in the neutral atmosphere. Ions are presumed to be dragged by the neutrals, and in this way exhibit the wave characteristics of the neutrals, but with contraints due to the earth's magnetic field. In spite of many measurements of meteor trails, rocket experiments and radar neutral scattering experiments, there has been almost no direct evidence to confirm this postulate from ground measurements, because of the difficulty of remotely measuring the neutral medium in the upper atmosphere. The wide acceptance resulted from indirect evidence, such as the measured velocity, scale lengths, periods and the direction of phase propagation (Hines, 1960).

Direct evidence is more likely in space, since spacecraft provide a suitable platform for making observations of both neutrals and ionization. Yet the simultaneous observations of both media has only become available in more recent years, so that published results are few. Two examples are Dyson et al (1970) who use Explorer 32 data and Reber et al (1975) using Atmosphere Explorer C (AE-C) data. In the first reference the fluctuation in the total neutral density is compared with the fluctuation in the electron density. In the second, two examples are given; one compares the fluctuations in the densities of argon, nitrogen and helium, and the other compares the fluctuations in the densities of argon, nitrogen and the ions and the fluctuation in the electron temperature. Both examples were for elliptic orbits for which the altitude changes rapidly. These limited cases depend on visual examination of the records to convince oneself that correlated motion exists.

Nevertheless, it is apparent in the examples of the latter paper that the fluctuations of helium are quite different from those of nitrogen or argon in that helium appears at times to be about 180° out of phase with the other two gases. Also, the amplitudes are different with the amplitude of argon about twice that of N_2 and that of helium about half that of N_2. This behavior has been found to be consistent with the theory of gravity waves involving more than one species for thermospheric conditions (Gross and Eun, 1976 and 1978a; Dudis and Reber, 1977; Del Genio et al, 1979).

A more recent example involving just neutrals (Hoegy et al, 1979) utilized AE-E data. The interpretation and analysis was more sophisticated than a visual treatment. In that work a wave is identified and all wave characteristics are derived. However, the theory of their modeling is not fully explained. Further examples for neutrals were given by the authors of this paper at various meetings of the American Geophysical Union and URSI (Gross and Reber, 1979; Gross et al, 1979; Huang et al, 1980) based on AE-C data. Some of these results will be presented in this paper along with newer material.

We have been studying data from many orbits of the AE-C satellites, those that are near circular and at an altitude of about 250 km. Circular orbits were chosen to eliminate the effects of altitude variations on the observed fluctuations. There is evidence of correlation in the fluctuations of all neutral species and ionization in every case examined. Since the satellite moves very rapidly (~ 7.8 km/sec) through the medium, the measurements essentially provide the spatial distribution of the measured quantities. The variations exhibit scale lengths typical of gravity waves and are presumed to be such waves. Both large scale (thousands of kilometers) and smaller scale (hundreds of kilometers) waves are observed. The scales of the waves are determined by spectral analysis using power spectra and cross spectra derived by the

maximum entropy method (Ulryck and Bishop, 1975). Generally, the stronger spectra are for the larger scale waves.

The AE satellites provide in situ measurements of the density and drift velocity of argon, helium, nitrogen and oxygen as measured by the Neutral Atmospheric Composition Experiment (NACE); neutral temperature for these species from the Neutral Atmospheric Temperature Experiment (NATE); plasma density and ion temperature by means of the Retarding Potential Analyzer (RPA) experiment; plasma density and electron temperature by means of the Cylindrical Electron Probe, as well as other experiments (Dalgarno et al, 1973). Many of these measurements are essentially simultaneous, making direct comparison of the measured properties and their fluctuations meaningful. Data utilized are available in two different sample intervals: 15 second averaged data from the so-called "UA" files and approximately 1.5 second samples from the so-called "GU" files.

Our goal was to determine whether observed wave structures are auroral in origin, so we emphasized AE-C circular orbit data because that satellite is in a high inclination (~ 68°) orbit. Though AE-E has yielded much data, its low inclination orbit is more suitable for non-auroral-connected events, which are of interest as well.

2. EXAMPLES OF DATA AND SPECTRA

Examples shown in this paper are illustrative. They are taken from data measured during two orbits of AE-C, orbit 5133 on January 20, 1975, and orbit 5359 on February 3, 1975. In both cases the altitude was approximately 250 km. Data for complete orbits are shown based upon 15 second averaged data samples, as well as the higher data rate 1.5 second samples for the study of regions in each of the hemispheres.

Complete orbit data are shown in Figure 1. This figure is a composite made up of the time variation in orbit of a number of data channels for orbit 5133, consisting of nitrogen particle density, electron density and ion and electron temperatures. Universal time in seconds is along the abscissa and ranges from 28000 to 34000. The ordinates for all the channels are logarithmic. Fluctuations are evident in all cases, but correlations between these fluctuations is not clear from this record. (Vertical lines in the N_2 and N_e (CEP) graphs are due to spurious computer effects.)

One obtains the fluctuations by extracting the low frequency or background content from the data shown in Figure 1. This is done by passing the data through a lowpass filter and subtracting its output from the original data. Such filtering is readily done digitally. We utilize the relative fluctuation rather than the fluctuation. The relative fluctuation is obtained by dividing the fluctuation by the background, the output of the lowpass filter as a function of the time in orbit. Figure 2 shows the relative density fluctuations for the complete orbit 5359 for nitrogen, argon, helium and the ionization. The figure also provides the fluctuation in the electron and ion temperatures. In contrast with the variation in total quantities in Figure 1, the fluctuations in Figure 2 exhibit, on visual examination, some degree of correlation.

Stronger evidence of correlation may be observed in the power spectra for these same data channels, not shown here. These are plots against the Doppler frequency in Hz along the abscissa. Since the satellite moves so much faster than the phase velocity of a wave, this Doppler spectrum is essentially $kv\cos\theta$, where k is the wave number, v is the satellite velocity and θ is the angle between the wave normal and the satellite path. Therefore, for fixed v it represents the component of the spatial distribution as given by $k\cos\theta$. Such spectra are calculated at evenly spaced frequencies by the computer and shown as lines. There are many common lines in the spectra of the various data channels, particularly for the stronger lines. Strong lines are found for the Doppler range from about 0.003 Hz to about 0.008 Hz. For a satellite velocity of 7.8 km/sec this frequency range corresponds to a horizontal wavelength range of about 2600 km to 975 km., assuming the satellite track and wavenormal are coincident. Since θ may not be zero, the actual wavelength range may be smaller. Apparently such long wavelength waves are present on a worldwide (full orbit) scale. There are also common spectra in the hundreds of kilometer range, though of much smaller intensity.

The shorter wavelengths (hundreds of km) are much more apparent in the spectra of 1.5 second data samples. An example of power spectra for orbit 5133 for the northern latitude range 55° to 0° is shown in Figure 3. The Nyquist frequency is 1/3 Hz, and the strongest lines are evident from about .025Hz to about .040 Hz, corresponding to horizontal wavelengths of about 300 km down to about 200 km. Smaller scale structure are also evident.

Further evidence for correlation of the various data channels for orbit 5133 in the same latitude region (55° to 0°) is shown in Figure 4. Here the co- and quadrature spectra are shown for each data channel relative to nitrogen. (The power spectrum of a data channel x(t) of total sample time T is $C_{xx}(f) = 1/T|X(f)|^2$, where X(f) is the Fourier transform and f is the frequency. The cross spectrum between two data channels x(t) and y(t) for total data samples of time T is given by $C_{xy}(f) = X^*(f)Y(f)/T$, where X(f) and Y(f) are the Fourier transforms of x and y and the asterisk means complex conjugate. The real part of C_{xy} is the co-spectra, and the imaginary part is the quadrature spectra (Jenkins and Watts, 1961). It is evident from the strong lines of these spectra that the various channels are not in phase, phaseshift being given by the arctangent of the ratio of the quadrature spectrum to the co-spectrum for the same Doppler frequency.

The amplitude and phase spectra are derivable from the co- and quadrature spectra, the first by taking the square root of the sum of the squares of the two spectra and the second as explained above to obtain the phaseshift. These spectra are of use in modeling the wave perturbations and their characteristics. By dividing the amplitude spectrum by the power spectrum of nitrogen at each of the major cross-spectral frequencies, the amplitude of the relative response of the particular geophysical parameter analyzed (density of neutrals or plasma, or temperature) to that of nitrogen is obtained; that is, $|Y(f)/X(f)|$, if x is the nitrogen fluctuation and y is the particular parameter of interest. Of course, this division only makes sense if we are dealing with a coherent wave. The phase spectrum of a channel provides the phase response relative to that of nitrogen. For the most significant frequencies, one may deduce the missing characteristics of the wave by appropriate modeling using the relative phase and amplitude responses of

the various channels. The Doppler frequency only provides the spatial wavelength along the orbital track. It is not necessarily the horizontal wavelength of the wave. Thus, by modeling one, in principle, is able to deduce the wave's temporal frequency, horizontal wavelength, direction of travel relative to the orbital path, the velocity of propagation and the vertical wavelength. A knowledge of these parameters together with the wave amplitudes is of great importance, since it would permit some deductions of the energy and momentum transported by the waves.

Bands of strong frequencies are common. Modeling of such adjacent lines should provide redundant information as though all are from a common source. However, it is a characteristic of the maximum entropy method (Ulryck and Bishop, 1975) to cause spreading of these bands and to introduce uncertainty as to their structure. As a consequence one should not take each line as exact. Instead, the average over several lines may be preferable. More work in spectral data analysis in this regard is needed.

Data were analyzed using the modeling techniques as described below. The techniques were only applied to neutrals, because the theory of modeling the ionization requires further study incorporating the magnetic field.

3. FIXED LAYER MODELING ANALYSIS

A full scheme of modeling wave solutions should include localized sources, multiple constituents and at least an altitude varying background atmosphere. In addition, winds and horizontal non-uniformity may play important roles. A problem of this complexity has not as yet been solved. However, parts of it have been treated. In the literature solutions have been given for two constituent atmospheres with vertical non-uniformity in density (Del Genio et al, 1979), and for vertical non-uniform density and temperature and mean mass and height dependent collision frequencies (Gross and Eun, 1976, 1978a and b). Solutions have also been given for two constituents with all these non-uniformities and localized sources (Gross, 1980). None of these solutions exhibit the degree of variation present in the measured data. Gross (1980) has shown that the spectra derived in the theoretical problem for two species is quite smooth and slowly varying with wavelength and temporal frequency. It is not clear then whether the degree of variation exhibited in spectra as in Figure 4 is due to the stochastic nature of the source or to propagation effects caused by variations in the medium.

Modeling was attempted using a locally homogeneous medium about the satellite's altitude to avoid these difficulties. We call this method the fixed layer analysis. In this analysis one assumes a gravity wave moving through an atmosphere with given background characteristics at an altitude of 250 km: temperature, temperature vertical gradient, mean mass, density for each constituent and collision frequencies between species. These values are obtained either from the measured data or else from models of the atmosphere for the particular time of year, location, solar cycle year, etc. The MSIS models (Hedin et al, 1977a and b) were used for this purpose. Densities were obtained by averaging over the measured data for the portion of orbit being modeled. Mean mass was also obtained from the average densities or else from the MSIS models. An atmosphere of more than two species is treated with two fluid theory by taking one (atomic oxygen), or a combination (atomic oxygen and molecular nitrogen), as a major species and each of the others (helium, argon), in turn, as minor. The gravity wave's characteristics are determined by the major constituent, and the response of helium or argon is relatable to that of the main species from the two fluid theory (Gross and Eun, 1976; Dudis and Reber, 1977; or Gross, 1980, for example). When the main constituent is a mixture of atomic oxygen and nitrogen, the same two fluid theory provides the eigenvalues appropriate to the gravity wave-like solutions for this mixture as well.

Thus, for orbit 5359 at southern mid-latitudes one obtains the complex relative responses of argon to nitrogen and of helium to nitrogen, as explained in the discussion on the cross spectra, for each Doppler frequency. These relative responses consist of amplitude and phase information as shown in Figures 5 and 6, respectively. Here the Doppler frequency is plotted as a harmonic number rather than frequency, since the latter is discrete. The fundamental is 1.3021 mHz. The lines in these figures are for the most significant responses, all others being weaker, noisy or uncertain due to possible noise.

With these relative responses for each Doppler Frequency, one has sufficient information for a model, neglecting wind effects and assuming horizontal uniformity, to find the corresponding wave period, horizontal wavelength, direction relative to the orbital path and the vertical wavelength. The method of defining the wave consists of scanning an appropriate range of wave periods and horizontal wavelengths to find that wave which comes closest to matching the argon and helium relative responses. A minimum mean square error based on the in phase and quadrature components of the responses was used as the criteria; that is, the differences between the calculated and the measured in phase and quadrature components were squared and summed for both argon and helium, and the sum was minimized. There are two solutions for each frequency, one for an upgoing wave, the second for a downgoing wave. Table 1 exhibits the wave solutions for the data of Figures 5 and 6 for phase-wise downgoing waves. Wave solutions for 15 Doppler frequencies, shown as harmonic numbers, are given. These frequencies are adjacent to each other, and one would expect waves of similar characteristics. Instead, the table shows a complete scatter in the periods and horizontal wavelengths, (shown as λ_x). Furthermore, one cannot determine the wave direction relative to the satellite path, θ, for most cases, since the best solutions give $\cos\theta > 1$. The measured and calculated amplitudes for the argon/nitrogen and helium/nitrogen responses, shown as A/N_2 and He/N_2 in the table, are considerably different. The fit is poor; the errors are comparable to the measured values. Similar results are obtained for phase-wise upgoing waves, for the other parts of the same orbit and for other orbits. If one averages adjacent lines to compensate for the peculiarities of the maximum entropy method and models the resulting values, similar results are obtained averaging three at a time, four at a time, etc. Varying the temperature or the collision frequencies by $\pm$ 10% yields no improvement.

Thus, the modeling attempts were unsuccessful. The medium may be far too complex for the assumptions. Data for thousands of kilometers must be used to analyze waves of several hundred kilometers wavelength. Undoubtedly, winds, wind shears and horizontal non-uniformity are not inconsequential over such distances. Further, the greater the distance from the source, the more is the change in wave characteristics (Chimonas and Hines, 1970). To model all these effects requires information not readily available, and, even if available, difficult to incorporate. We cannot rule out noise in the data, as well.

When we started this program we were optimistic that we could determine the characteristics of the waves from the satellite measurements because of the number of simultaneously measured quantities. These negative results, however, lead us to believe that it is essential to correlate spacecraft and ground measurements to find all the characteristics of the waves, temporal and spatial. Unfortunately, to date we have not been able to find suitable ground experiments functioning during a satellite pass over the station when events occurred.

Another approach has been to review the data in a different fashion. This has led to the discovery of new, possibly important, characteristics that had not originally been appreciated, nor possibly known at all before, as far as these authors can determine. Though these new results are preliminary, they may hopefully lead to a much better understanding of the mechanisms at work and their relationships to important geophysical aspects, such as magnetic storms, solar cycle variations, season, etc. These findings will now be explained.

4. LARGE SCALE WAVE DISTRIBUTIONS AND TURBULENCE

By a very simple process, that is, by replotting the spectral data shown in Figures 3 or 4 on a log-log graph, an unappreciated feature became evident. In this paper we show this only for the power spectra. All are for the northern latitude range 55° to 0° of orbit 5133. We see similar features in other parts of the same orbit and for other orbits.

The power spectra of N_2, O and electron density n_e are plotted versus the harmonic frequency number in Figure 7 through 9. They are shown as dots on the graphs. Those below about harmonic number 20 are in the lowpass filter range and should be ignored; the fluctuations of interest are those from harmonic number 20 and up. Harmonic number 20 corresponds to a wavelength of 300 km along the orbital path. Also shown on each graph is a best logarithmic straight line fit. Such a fit is expressible as An^{-a}, where A is an amplitude constant, n is the harmonic number and a is the spectral index characterizing the wave number decay rate. The fits for Figures 7 - 9 are based on the spectrum from n = 25 to n = 100. The deviation of the spectra in these figures is, except for some discrete values of n, relatively small. The values of a for these three fits are -5.3719 for nitrogen, -3.14128 for atomic oxygen and -2.1401 for the electron density. Curiously, the ratio of the value of a for nitrogen to that for oxygen is very close to the ratio of their molecular weights. Not enough data have been analyzed to establish whether this is real or a coincidence.

Do the fits in Figures 7 - 9 imply a turbulence-like spectrum for these large scale waves? The graphs are really wave number plots, and as such they are reminiscent of eddy distributions (Monin and Yaglom, 1971), but for much larger scale lengths. It is also possible that such spectra are representative of a stochastic source and therefore not turbulence. It may originate from non-linear interactions in the region in and about the source. In any event, it is believed that strong non-linear interactions must be involved, whether turbulence or source derived.

Plots similar to those of Figures 7 through 9 were made for helium and the electron temperature. These are not shown here. The fit for helium for the spectrum between n = 25 and n = 100, as in Figures 7 - 9, has the exponent a = -0.1645. This line yields a power spectral density almost independent of wavenumber, like white noise, and corroborates a noise-like distribution evident in data, particularly beyond n = 50. If a fit is made between n = 25 and n = 50, one obtains a second line with an exponent a = -1.45. A close examination of the data shows that there is a decreasing spectrum in this range of n. Apparently at shorter wavelengths noise dominates over the decreasing spectrum suggested by the second fit. This latter fit is probably not very accurate due to the influence of noise. Though not shown here as well, the fluctuation vs. time data for helium in this portion of the orbit, unlike that in Figure 2 for helium, appears quite noisy, implying that the NACE instrument was limited by the noise level in this measurement of helium. Nitrogen and oxygen are more abundant, so that noise does not appear to be a problem. The electron density measurements are from another instrument, the cylindrical electron probe.

The spectrum for the electron temperature is even noisier than that of helium. Two fits are also suggested. One is for n = 25 to n = 50 for which the exponent is positive and small (a = 0.342), as is expected for white noise. The other fit is for n = 25 to n = 40. This fit was computed because there seems to be a decreasing spectrum in the range. Its exponent is a = -2.31. What is interesting is that the elimination of just 10 frequencies from n = 41 to n = 50 produces this change. In contrast, the computation of fits for ranges other than n = 25 to 100 for nitrogen, oxygen and the electron density causes minor modification of their exponents. As for helium, the correct fit for the electron temperature spectrum is uncertain. At best, one can only conclude there is some hint of a decreasing spectrum that becomes overwhelmed by noise as the scale size decreases. As far as these authors know, such an electron temperature spectrum is the first of its kind ever deduced in natural plasma. Unfortunately, room does not permit its inclusion.

Over the range n = 25 to n = 50 there appear to be significant departures in the various spectra from their straight line fit at or near particular values of n. These departures or deviations may imply a wave system sitting on a turbulence-like spectrum. The features are made evident by plotting the difference, or residuals, vs. n on linear plots. Figure 10 shows these differences for nitrogen, oxygen and the electron density. The match of the peaks in these differences, between those of nitrogen and oxygen, is excellent, and that for the electron density is quite good. We find this match even though the maximum entropy method tends to broaden any peaks. The first large peak in all three of the residuals is at a wavelength along the orbit of about 272 km. The second peak (missing in the electron density spectrum) is at about 240 km. The third, fourth and fifth peaks are at about 190km, 150km, and 133 km, respectively. Though speculative, this group could be indicative of a wave packet moving through the medium.

The deviations for helium and the electron temperature are shown in Figure 11, based upon the n = 25 to 50 fit for helium and the n = 25 to 40 fit for the electron temperature. Though there is some degree of overlap with the spectra in Figure 10, the wide excursions are not surprising in view of the noisiness of the data for these two channels. These spectra are not considered too reliable.

5. CONCLUSIONS

The negative results that were obtained in attempting to model the fluctuations in the neutrals as gravity waves is not surprising in view of the complexity of the wave structure seen in Figures 7 to 11. Apparently, non-linear processes, probably associated with strong stochastic sources in the auroral region, and possibly from meteorological sources, are at work producing a turbulence-like spectrum in the gravity wave range stretching from wavelengths of many hundreds of kilometers down to the smallest scales observable before noise dominates, on the order of tens of kilometers. Work with complete orbit data also suggests that the turbulence-like structure extends to waves of thousand of kilometers in wavelength. The exponents of the spectra for each of the neutral constituents, the plasma and electron temperature are of different values in the limited examples given here and the few others that have been investigated. Why this is the case is not clear, but might arise from non-linear cascading processes and loss mechanisms associated with diffusion. There is much theory to be investigated here. It is not clear how or if these waves are propagating. Are the waves stationary, are they propagating in all directions or from the sources outwards? The deviations of Figures 10 and 11 suggest that wave packets are propagating on top of turbulence. It is conceivable that such wave packets would be observed as a traveling ionospheric disturbance on the ground.

A most interesting future investigation will consist of the statistical study of these characteristics and how they may relate to geophysical aspects, such as the year in the solar cycle, seasonal effects, magnetic storms, solar sector boundary crossings, etc. There is also a need to study similar characteristics in the equatorial region, as is possible with AE-E data, for relationships to the equatorial electrojet and meteorological events.

REFERENCES

CHIMONAS, G. and C.O. Hines (1970), Atmospheric Gravity Waves Launched by Auroral Currents, Planet. Sp.Sci.18, 565-582.

DALGARNO, A., W.B. Hanson, N.W. Spencer, E.R. Schmerling (1973), The Atmospheric Explorer Mission, Radio Science 8, 263-266.

DEL GENIO, A.D., G.Schubert, J.M.Strauss (1979), Characteristics of Acoustic-Gravity Waves in a Diffusively Separated Atmosphere. J.Geophys. Res. 84, 1865-1879.

DUDIS, J.J. and C.A.Reber (1977), Composition Effects in Thermospheric Gravity Waves, Geophys. Res. Lett. 3, 727-730.

DYSON, P.L., G.P. Newton and L.H.Brace (1970), In Situ Measurements of Neutral and Electron Density Wave Structure from the Explorer 32 Satellite. J.Geophys. Res. 75, 3200-3210.

GROSS, S.H. (1980), Model Analysis of Large Scale Wave Structure as Observed by AE Satellites, paper SA100, Spring Meeting of the American Geophysical Union, May 22-27, Toronto, Canada, EOS 61, 322-323.

GROSS,S.H. and H.Eun (1976), Traveling Neutral Disturbances. GRL 3, 257-260.
(1978a), Traveling Neutral Disburbances and Minor Constituent Particle Mass. J.Atm.Terr.Phys.40, 183-193.
(1978b), Relative Phase and Amplitude Response to Gravity Waves of Minor Species in the Thermosphere. Geophys., Astrophys. Fl.Dyn. 11, 131-140.

GROSS, S.H. and C.A.Reber (1979), Wave Structure in the Neutral Species Observed by Atmospheric Explorer Satellites, paper SA-29, Spring Meeting of the American Geophysical Union, Washington, D.C., May 28 - June 1, EOS 60, 331-332.

GROSS, S.H., C.A. Reber and F.Huang (1979), Large Scale Waves in the Ionosphere Observed by the AE Satellites, Session H-1, paper 5, National Radio Science Meeting, 5-8 November, Boulder, Colorado.

HEDIN, A.E., J.E.Salah, J.V.Evans, C.A.Reber, G.P.Newton, N.W.Spencer, D.C.Kayser, D.Alcaydé, P.Bauer, L.Cogger and J.P.McClure (1977a), A Global Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data. MSIS 1. N_2 Density and Temperature. J.Geophys. Res. 82, 2139-2147.

HEDIN, A.E., C.A.Reber, G.P.Newton, N.W.Spencer, H.C.Brinton and H.G.Mayr (1977b), A Global Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data, MSIS 2. Composition. J.Geophys. Res.82, 2138-2156.

HINES, C.O. (1960), Internal Atmospheric Gravity Waves at Ionospheric Heights, Can. J.Phys. 38, 1441-1481.

HOEGY, W.R., P.Dyson, L.E.Wharton and N.Spencer (1979), Neutral Atmospheric Waves, GRL 6, 187-190.

HUANG, F., C.A.Reber and S.H.Gross (1980), Large Scale Waves in the Atmosphere and Ionosphere Observed by the AE Satellites. Paper SA101, Spring Meeting of the American Geophysical Union, May 22-27, Toronto, Canada, EOS 61, 323.

JENKINS, G.M. and D.G.Watts (1968), Spectral Analysis and Its Applications. Holden-Day, San Francisco, California, pp.210, 216-222, 341-348.

MARTYN, D.F. (1950), Cellular Atmospheric Waves in the Ionosphere and Troposphere. Proc. Roy. Soc. A201, 216-234.

MONIN, A.S. and A.M.Yaglom (1971), Statistical Fluid Mechanics: Mechanics of Turbulence. Vol. 2. The MIT Press, Cambridge, Mass., p.201.

REBER, C.A., A.E.Hedin, D.T.Pelz, N.E.Potter and L.H.Brace (1975), Phase and Amplitude Relationships of Wave Structure Observed in the Lower Thermosphere. J.Geophys. Res. 80, 4576-4580.

ULRYCK, T.J. and T.N.Bishop (1975), Maximum Entropy Spectral Analysis and Autoregressive Decomposition. Rev. Geophys. Sp.Phys. 13, 183-200.

TABLE I

FIXED LAYER ANALYSIS

250 km Southern Mid-Latitudes
AE-C Orbit 5359
Downgoing Phasewise Wave

Harmonic Number	Period min	λ_x km	ψ	Measured		Calculated	
				$\lvert A/N_2 \rvert$	$\lvert He/N_2 \rvert$	$\lvert A/N_2 \rvert$	$\lvert He/N_2 \rvert$
16	106	2600	x	6.0	12.0	2.9	6.7
17	625	1200	x	1.6	2.3	3.8	1.0
18	286	5400	x	3.8	9.0	10.9	10.4
19	10	120	67°	1.1	18.2	8.7	25.7
20	628	13000	x	2.1	5.8	5.9	7.2
21	84	1100	x	1.2	12.4	4.8	8.8
22	18	170	50°	9.1	24.3	18.8	8.6
23	11	88	70°	6.7	13.5	14.2	7.8
24	11	60	76°	5.0	4.5	2.1	3.3
25	200	1460	x	8.6	12.9	7.8	10.5
26	10	315	x	3.5	16.9	2.1	24.5
27	57	1190	x	2.6	18.8	7.9	25.3
28	70	1570	x	2.8	15.1	7.1	21.7
29	188	5400	x	3.0	8.6	6.4	12.7
30	10	38	72°	13.1	12.3	1.3	5.2

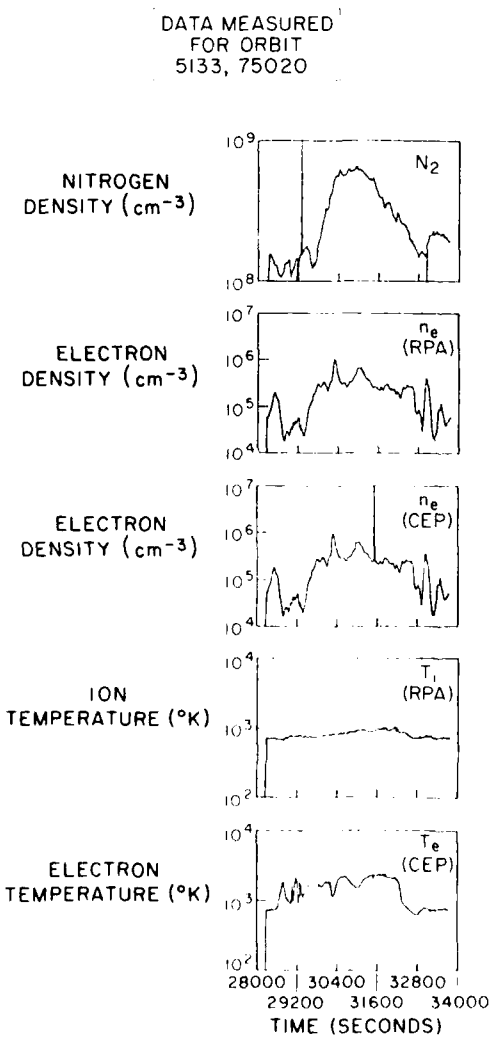


Figure 1 Densities and temperatures for one entire orbit, orbit 5133.

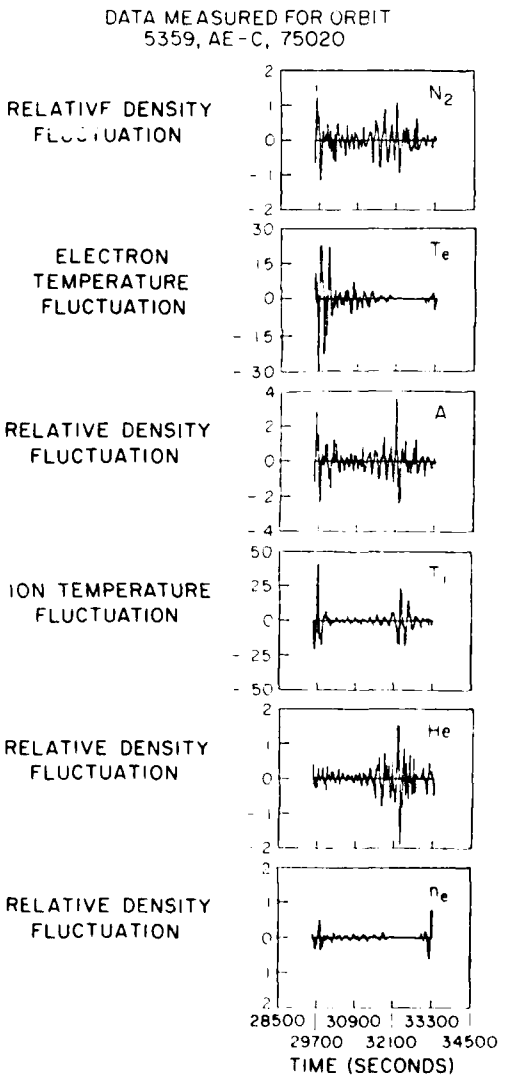
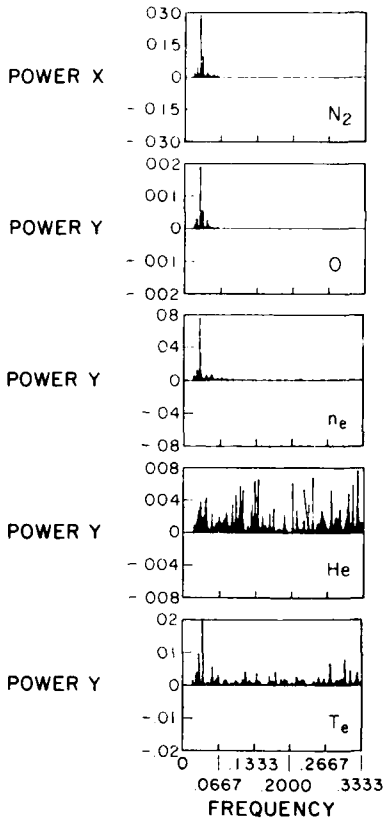


Figure 2 Relative fluctuations of densities and temperatures for orbit 5359.



5133 MEM/GU 29800

Figure 3 Power spectra for a part of orbit 5133, 55° to 0° latitude.

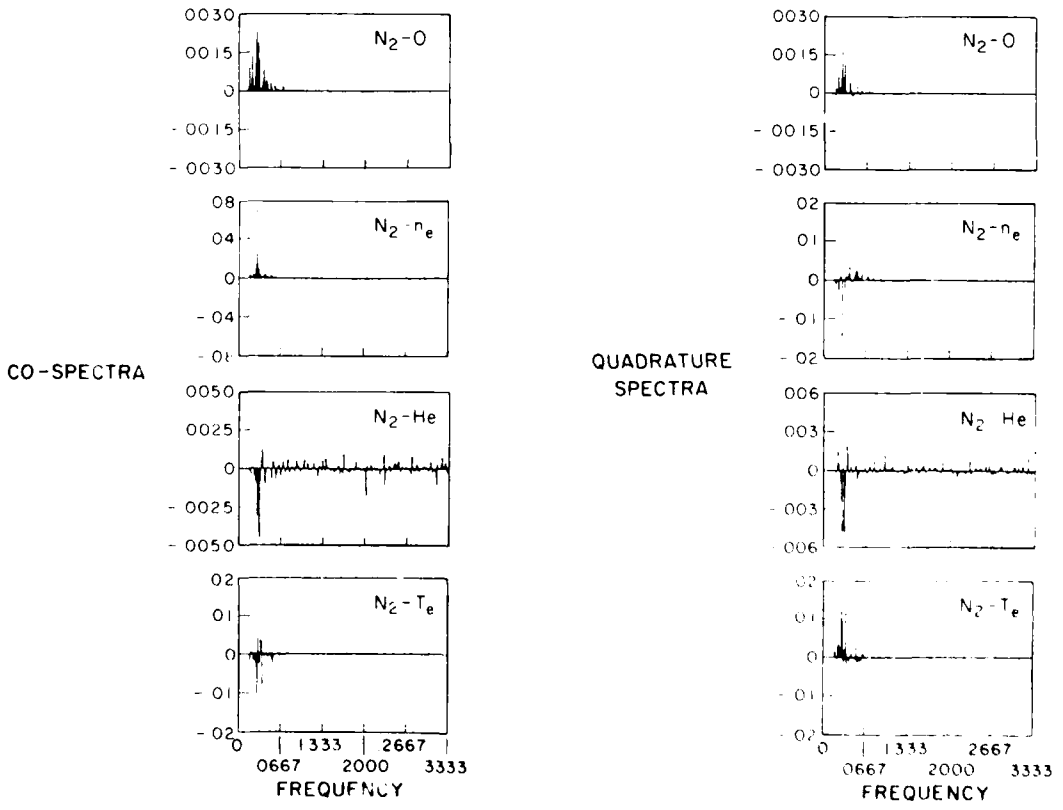
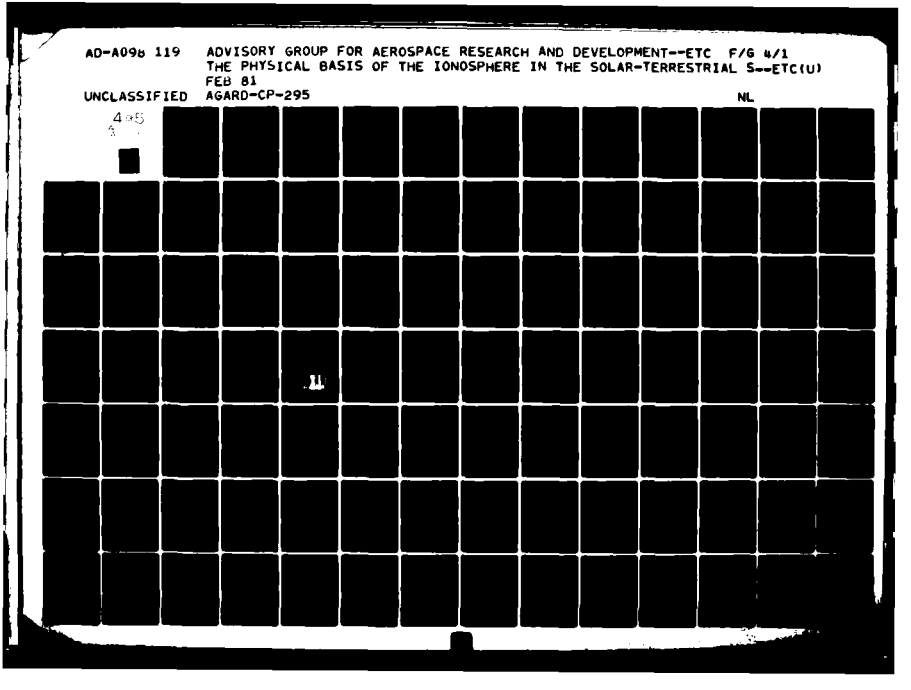
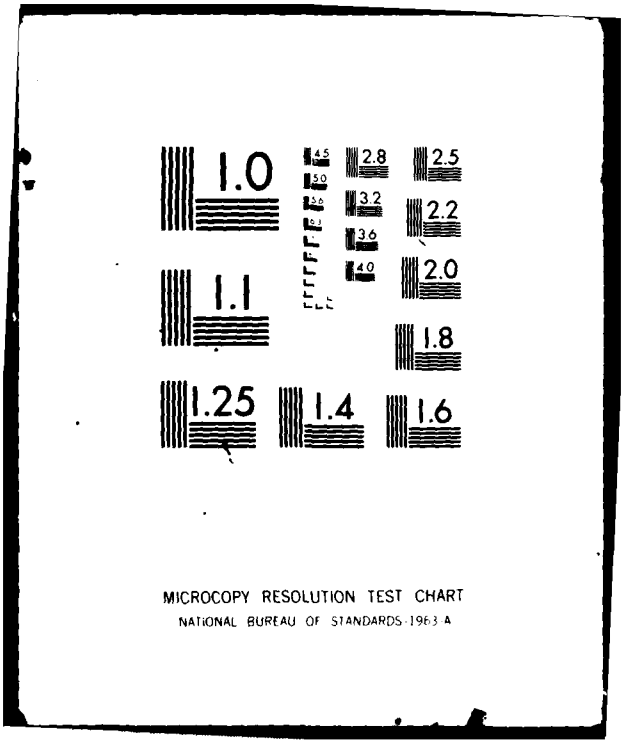


Figure 4 Co- and quadrature spectra for the data channels of Figure 3 relative to nitrogen.


AD-A098 119 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 4/1
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL S--ETC(U)
FEB 81
UNCLASSIFIED AGARD-CP-295
NL
4 of 5


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

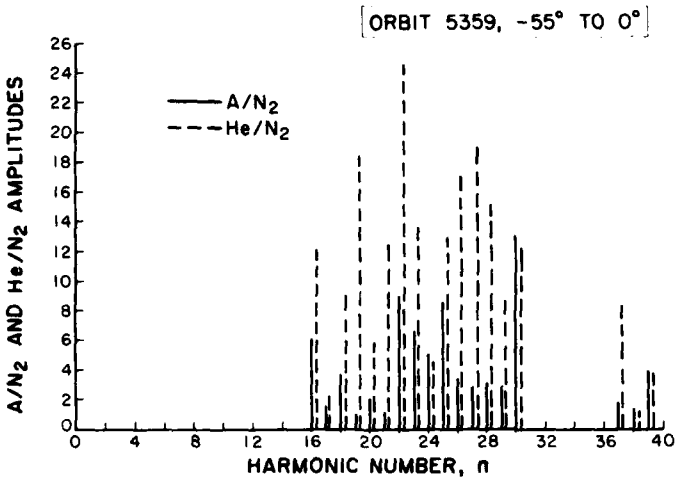


Figure 5 Amplitude ratios of argon to nitrogen and helium to nitrogen for significant lines of the cross spectra, orbit 5359, southern mid-latitudes.

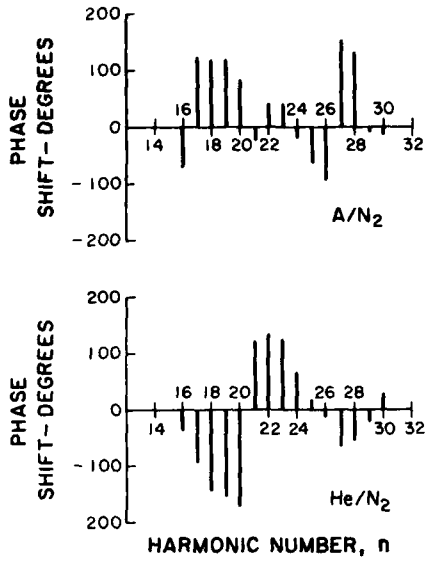


Figure 6 Phase spectra associated with the amplitude spectra of Figure 5.

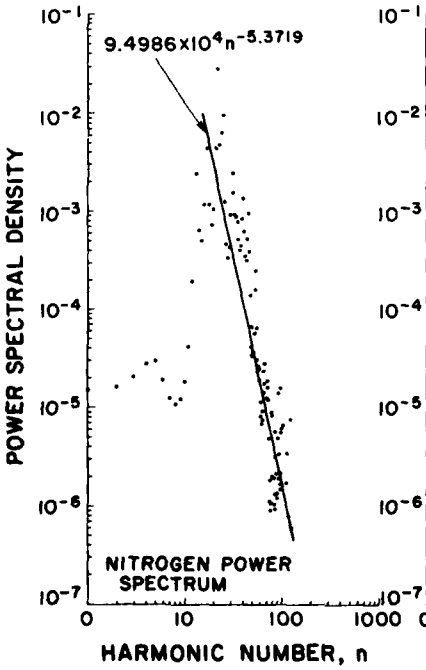


Figure 7 Power spectrum of nitrogen vs. the harmonic number n for data of Figure 3. Best fit line based on the spectra from n = 25 to 100 is shown.

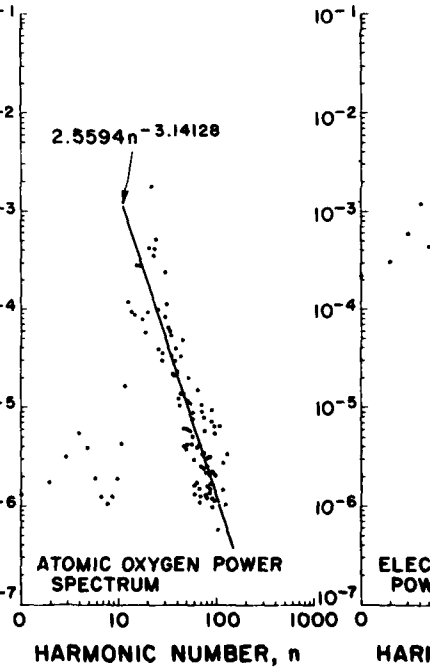


Figure 8 The same as Figure 7 but for atomic oxygen.

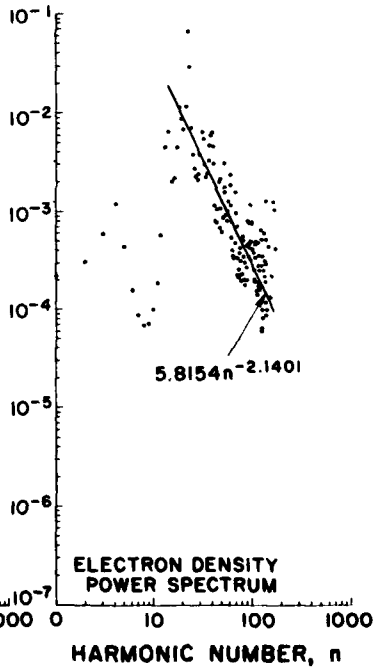


Figure 9 The same as Figure 7 but for the electron density.

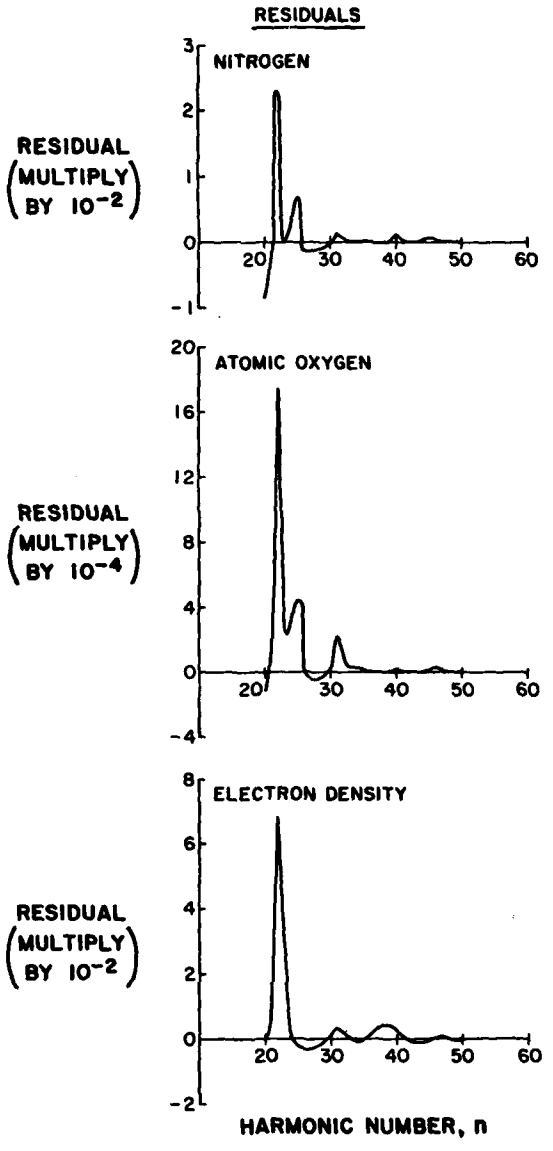


Figure 10 Plots of residuals of spectra in Figure 7 - 9 in the range n = 20 to n = 50.

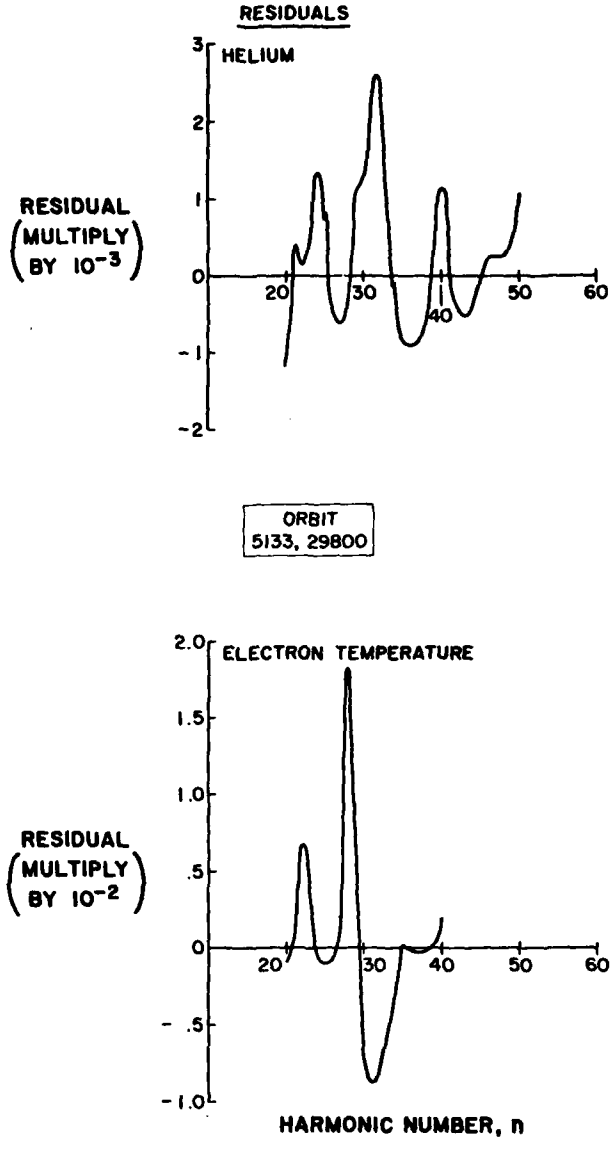


Figure 11 The same as Figure 10 but for helium and the electron temperature.

THE GENERATION AND PROPAGATION OF ATMOSPHERIC GRAVITY WAVES WITH SPECIAL REFERENCE TO RADIO PROPAGATION

T.B. Jones
Department of Physics
University of Leicester
Leicester, U.K.

SUMMARY

Atmospheric gravity waves are known to propagate at ionospheric heights where they produce localized distortions in the electron density distribution known as travelling ionospheric disturbances (TIDs). These electron density perturbations can modify the characteristics of ionospherically propagated radio waves and hence the performance of any radio system which utilizes this type of propagation.

It is generally assumed that the auroral zone is the major source location of atmospheric gravity waves since, at these latitudes, particle precipitation events can inject appreciable amounts of energy into the atmosphere. Furthermore, perturbation of the atmosphere can take place due to Joule heating and Lorentz forces which are characteristic features within the auroral oval. Recent experimental evidence suggests however that important source locations other than those in the auroral zone can exist.

The propagation of atmospheric gravity waves has been studied for many years and the propagation characteristics are well established. Ray tracing methods have been evolved which enable the passage of the gravity wave to be traced through any given atmospheric model. The wave characteristics are however critically dependent on atmospheric parameters such as the neutral wind, temperature etc. The usefulness of these analytical techniques is thus constrained by the validity of the atmospheric model adopted. At present, the lack of "real time" models is a major constraint in predicting the influence of a gravity wave as it propagates away from its source. Because of these difficulties, recent attempts to correct for the influence of these irregularities on radio direction finding systems have involved real time monitoring of the ionosphere in order to detect the extent of the TID activity.

The object of this paper is to review our knowledge of the source processes and their locations and to outline the well established theory of gravity wave propagation. Attention is also given to recent work aimed at minimizing the influence of these disturbances on radio systems.

1. INTRODUCTION

The ionosphere is frequently perturbed by wave-like fluctuations which travel for appreciable distances and are generally referred to as travelling ionospheric disturbances(TIDs). These irregularities are a consequence of the propagation of internal gravity waves in the upper atmosphere which induce electron and ion motion through neutral-ion collisions.

Experimental evidence (e.g. Georges, 1967) indicates that two types of TIDs exist which are referred to as medium and large scale respectively. The large scale events travel at high speeds (300 to 1000 ms^{-1}) and usually consist of a periodic wave train containing 1 to 3 cycles. Their occurrence is generally associated with magnetic storm activity and they tend to be observed most frequently for a few hours after sunset at temperate latitudes. Their sources lie in the auroral zone and they propagate equatorwards, being dissipated before reaching the opposite hemisphere.

Medium scale TIDs are frequently observed and very few days occur when no medium scale TIDs are present. They travel at horizontal speeds of 100 to 300 ms^{-1} and are not strongly correlated with magnetic activity. This type of TID sometimes exists as a non-periodic disturbance and sometimes as a wave train consisting of several cycles. The wave train features are frequently associated with tropospheric disturbances such as those due to strong weather fronts. The source of medium scale TIDs is not well known and tropospheric, mesospheric and electrojet sources have been suggested. The waves are dispersive and can change appreciably as they propagate over distances of a few hundred kilometers. The periods of the gravity waves range from about 20 min to 1 or 2 hours.

In addition to the gravity waves a number of short period ($\tau < 3$ min) waves are also present (Davies & Jones, 1972). These are acoustic waves but are not considered further in this review although many of their properties are related to those of the gravity waves.

The object of this paper is to discuss the propagation of gravity waves in the upper atmosphere and their influence on the ionosphere. Attention will also be given to possible wave sources and their likely locations. Finally, the influence of the resulting ionospheric disturbances on a number of radio wave systems will be considered.

2. PROPAGATION OF GRAVITY WAVES IN THE ATMOSPHERE

The equations that govern the dynamics of the atmosphere can be derived from the conservation laws and written as:

$$\frac{\partial \rho^1}{\partial t} + \nabla . (\rho_o \underline{v}^1) = 0 \qquad \text{conservation of mass} \qquad .... \quad 1)$$

$$\rho_o \frac{\partial \underline{v}^1}{\partial t} = -\nabla p^1 + \underline{g}\rho^1 \qquad \text{conservation of momentum} \qquad .... \quad 2)$$

$$\frac{\partial p^1}{\partial t} + \underline{v}^1 . \nabla p_o = \frac{\gamma p_o}{\rho_o}\left(\frac{\partial \rho^1}{\partial t} + \underline{v}^1 . \nabla \rho_o\right) \qquad \text{adiabatic state} \qquad .... \quad 3)$$

The variables in the equations are the mass density $\rho = \rho_0 + \rho^1$, temperature $t = t_0 + t^1$, pressure $p = p_0 + p^1$ and fluid velocity $\underline{v}^1$. The ambient values have zero subscripts and the primes denote the fluctuations. g is the acceleration due to gravity and γ is the ratio of specific heats.

Hines (1960) has considered the fluctuation of the isothermal atmosphere to have space and time variations of the form (p^1/p), (ρ^1/ρ_0), $\underline{v}^1/t^1$, proportional to $\exp(i(\omega t - k_z z - k_x x) + \frac{z}{2H})$ where x, y and z define a rectangular coordinate system. Using this representation solution of equations 1 to 3 leads to the dispersion equation

$$k_z^2 = \left(\frac{\omega_b^2}{\omega^2} - 1\right) k_x^2 + \frac{\omega^2 - \omega_a^2}{c^2} \quad \ldots\ 4)$$

where c is the speed of sound and can be expressed as

$$c^2 = \frac{\gamma RT}{M} = \gamma g H \quad \ldots\ 5)$$

ω_b is the Brunt frequency

$$\omega_b^2 = \frac{(\gamma - 1)g^2}{c^2} \quad \ldots\ 6)$$

and ω_a the acoustic cut-off

$$\omega_a^2 = \frac{\gamma^2 g^2}{4c^2} = \frac{c^2}{4H^2} \quad \ldots\ 7)$$

These equations together yield a description of a plane acoustic-gravity wave propagating in an isothermal atmosphere.

The spatial dependence of the wave involves oscillations as functions of x and z coupled with an exponential dependence on height, $\exp\left(\frac{z}{2H}\right)$. Thus there is an exponential growth of the wave as it moves upwards through the atmosphere to regions of small p and ρ. This can be physically interpreted as a consequence of the conservation of energy since the available energy is distributed between fewer molecules as the wave propagates upwards.

A wave packet composed of these waves travels with group velocity V_g which can be determined since

$$V_{g_x} = \left(\frac{\partial \omega}{\partial k_x}\right)_{k_z} \quad ; \quad V_{g_z} = \left(\frac{\partial \omega}{\partial k_z}\right)_{k_x} \quad \ldots\ 8)$$

so that the angle between the group velocity and the horizontal ϕ can be written as

$$\tan \phi = \frac{V_{g_z}}{V_{g_x}} = -\frac{k_z}{k_x\left(\frac{\omega_b^2}{\omega^2} - 1\right)} \quad \ldots\ 9)$$

The dispersion relationship is illustrated in Fig. 1 in which the normalized horizontal phase velocity $\left(\omega/ck_x\right)$ is plotted against the normalized period $\frac{\omega_b}{\omega}$ for each ascent angle ϕ. For real k_x and ω the value of k_z is either purely real or purely imaginary. Imaginary k_z corresponds to an evanescent wave whose amplitude either grows or decays exponentially with z and for which the energy flow is horizontal.

For real values of k_z there are two regions of the spectrum in which real waves exist. When $\omega/\omega_b \gg 1$ the waves are compressional isotropic sound waves, (acoustic waves), which satisfy the relationship

$$\omega^2 = c^2(k_x^2 + k_z^2) \quad \ldots\ 10)$$

If $\omega < \omega_a$ no acoustic waves can propagate and ω_a is the acoustic "cut-off" frequency. For $\omega_a > \omega > \omega_b$ no internal waves exist but for $\omega < \omega_b$ internal gravity waves can propagate. If $\omega \sim \omega_b$ both buoyancy and compressional forces determine the characteristics of the waves, however, if $\omega \ll \omega_b$ the buoyancy forces dominate.

In the limit $\omega \ll \omega_b$ and $\omega/ck_x \ll 1$ the dispersion relationship simplifies to give

$$k_z = \pm k_x \frac{\omega_b}{\omega} \quad \ldots\ 11)$$

and $$\frac{V_{g_z}}{V_{g_x}} = \tan \phi = \mp \frac{\omega}{\omega_b} \quad \ldots\ 12)$$

Moreover $$\frac{V_x^1}{V_z^1} = -\frac{k_z}{k_x} \quad \ldots\ 13)$$

so that $$\underline{k} \cdot \underline{V}^1 = 0 \quad \ldots\ 14)$$

For this limiting case the gravity wave consists of a wave packet propagating obliquely upwards containing within it transverse velocity oscillations which grow exponentially as the wave propagates. The phase fronts propagate downwards, perpendicular to the motion of the wave packet as a whole. This situation is illustrated in Fig. 2 after Hines (1960).

3. PROPAGATION IN A REAL ATMOSPHERE

The real atmosphere is not isotropic since temperature and mean molecular weight vary with height. Furthermore the dissipative effects of viscosity, thermal conductivity and ion drag will also modify the propagation characteristics of propagating acoustic-gravity waves. The temperature-height variation will produce a corresponding variation in the sound speed, the scale height and in ω_b. The change in molecular weight will affect the acoustic cut-off ω_a.

Propagation in this type of atmosphere is usually considered by dividing the atmosphere into a number of slabs in each of which the temperature is assumed constant (e.g. Pierce, 1966). The reflection and transmission coefficients of the wave at each slab boundary are determined and a summation for all the slabs is undertaken to derive the total effect. The existence of multiple layers normally requires the application of iterative techniques such as those described by Hines and Reddy (1967). The reflection conditions for the case of an atmosphere composed of two isothermal regions is illustrated in Fig. 3. Waves in the shaded area propagate in the lower region (region 1) but are evanescent in the upper region (region 2). This implies total reflection at the boundary between the two regions for waves in the shaded area of the figure. Waves above the shaded area will be refracted at the boundary according to Snell's law.

As already stated, the reflected wave will be considerably influenced by dissipative processes. The wave equation can be generalized to include the effects of viscosity and thermal conductivity to give

$$\frac{\partial \rho'}{\partial t} + \nabla.(\rho_o \underline{v}') = 0 \qquad \ldots. \quad 15)$$

$$\rho_o \frac{\partial \underline{v}'_i}{\partial t} = -\frac{\partial p'}{\partial x_i} + \rho' g_i + \frac{\partial}{\partial x_j}\left(\mu\left(\frac{\partial v'_i}{\partial x_j} + \frac{\partial v'_j}{\partial x_i} - \tfrac{2}{3}\delta_{ij}\nabla.\underline{v}'\right)\right) \qquad \ldots. \quad 16)$$

$$\frac{\rho_o R}{(\gamma-1)M}\frac{\partial T'}{\partial t} = \nabla.(\lambda \nabla T') - \rho_o \nabla.\underline{v}'$$

$$p' = \frac{\rho' R T_o}{M} + \frac{\rho_o R T'}{M} \qquad \ldots. \quad 17)$$

where μ is the coefficient of viscosity and λ is that of thermal conductivity, assumed to be slowly varying with height.

The dissipative effects of both viscosity and thermal conductivity generally increase with height. At low altitudes the effects of these loss processes can generally be neglected, but as the wave moves upwards through the atmosphere their influence increases rapidly. The wave loses its energy at a rate depending on the atmospheric model, wave period and frequency. This effect has been considered by Francis (1973, 1975) and his results are illustrated in Fig. 4 in which the horizontal velocity perturbation due to the wave is shown with and without dissipation. The solid curve shows the instantaneous phase structure, while the broken curve represents the envelope within which the phase structure propagates as a function of x and t. The dash-dot curve indicates how the envelope would appear in the absence of dissipation. The two curves in Fig. 4 begin to deviate from each other at about 150 km illustrating the importance of including dissipative effects in calculations of gravity wave propagation in the ionosphere.

4. THE INFLUENCE OF GRAVITY WAVES IN THE IONOSPHERE

Movements of the neutral atmosphere will induce changes in the electron and ion densities of the ionosphere since the neutral and charged particles are coupled by collision processes with time constants ≤ 1 sec. This time is much shorter than the gravity wave period ($\sim$ 20 min) thus the ions and electrons would have the same fluid velocity as the neutrals if other parameters such as the earth's magnetic field did not influence their behaviour. Since the ion gyro frequency ($> 100\ s^{-1}$) is much larger than the ion-neutral collision frequency ($\leq 10\ s^{-1}$) in the F-region the electrons and ions are constrained to move along the $\underline{B}$ field with a velocity

$$v'_e = \frac{\underline{B}\,\underline{B}.\underline{v}'}{B^2} \qquad \ldots. \quad 18)$$

The compressions and rarefactions of the electron gas in the magnetic field direction can be represented by the continuity equation

$$\frac{\partial n'_e}{\partial t} + \nabla.(n_{co}\,.\,\underline{v}'_e) = 0 \qquad \ldots. \quad 19)$$

where $n_e = n_{co} + n_e^1$ is the electron density.

The disturbance produced by the gravity wave of Fig. 4 in a Chapman layer where $Z_{max} = 300$ km and $H_{max} = 50$ km has been calculated by Francis (1973). The resulting changes in the electron density distribution are shown in Fig. 5 which indicates considerable structure in the height variation. This occurs because the electron density fluctuation at a given height is approximately proportional to the vertical gradient of the ambient electron density at that height. The discussion outlined above holds near the F-layer peak (200-400 km) but breaks down at lower heights due to the effect of electron loss processes.

5. SOURCES OF GRAVITY WAVES

5.1 Medium Scale TIDs

These disturbances can be produced by a variety of sources. Among the naturally occurring sources are changes in the auroral electrojet, tropospheric weather systems and earthquakes (e.g. Brekke (1979), Baker and Davies (1969), Davies and Baker (1965)). Man-made sources such as nuclear explosions, have also been identified (e.g. Baker and Davies , 1968).

Consider the generation of gravity waves by the electrojet. Sudden movements of the electrojet current $\underline{J}$ produce a Lorentz force $\underline{J} \times \underline{B}$ which couples to the neutral atmosphere by collisions to produce gravity waves. Joule heating by the electrojet can also provide source processes for gravity waves.

The gravity wave generation from an impulsive source has been discussed by Chimonas and Hines (1970), Francis (1973) and others. Fig. 6 illustrates the medium scale period response in the atmosphere to an upper atmosphere source. In this example due to Francis (1973) a current surge of 10^5 amp in the auroral electrojet has been assumed. Two families of waves are emitted from the source corresponding to those which (a) propagate directly upwards into the F-region and (b) those which initially travel downward, are reflected at the earth's surface and then travel upwards into the F-region.

At a given height and time the disturbance due to the earth reflected waves appears as a wave packet whose wavelengths are relatively constant. The average horizontal wavelength (λ_x) increases almost linearly with distance from the source. The disturbance always appears as a wave train, the number of cycles being approximately constant. The gravity waves which propagate directly into the F-layer without being reflected at the earth's surface produce a quite different type of disturbance. In this case they appear as isolated fluctuations rather than as conventional wave packets. Both direct and earth reflected waves are dispersive, but for the earth reflected waves the propagation path into the F-layer is considerably greater than that of the direct waves. The earth reflected waves appear in the F-layer as a relatively monochromatic wave packet. The direct waves have less time to disperse and to separate the various spectral components into different regions of space. As a result, the response approximates more to a single pulse than to an elongated wave packet. Both types of response are observed experimentally in the ionosphere.

5.2 Large Scale TIDs

In addition to the medium scale disturbances described above, a localized atmospheric source will also give rise to a discrete spectrum composed of several families of guided modes, one of which produces the so-called large scale TIDs. The features of these modes and their ducting mechanisms have been discussed by several authors, e.g. Pitteway and Hines (1965), Friedman (1966). Ducting will occur at heights in the atmosphere where there is a marked gradient in the sound speed e.g. when there is a steep temperature gradient. For a simple atmospheric model consisting of two isothermal regions but different in temperature, the long period waves are non-dispersive and the kinetic energy will maximize at the temperature discontinuity and decrease exponentially above and below this level. The well-known Lamb mode will be excited at the atmosphere-ground boundary. However, the kinetic energy of this mode decreases exponentially with height and at F-region levels it is greatly attenuated and its influence can be neglected.

The long period gravity waves propagate into the upper atmosphere and disturb the E and F layers of the ionosphere. These modes travel along the steep temperature gradient at the base of the thermosphere and can leak energy to regions above and below this level. The energy flow is nearly horizontal and the waves are dissipated by viscosity and thermal conductivity effects. The magnitude of these loss processes decreases with increasing wave period and the energy flow becomes more horizontal. The propagation of the modes is critically dependent on the thermal structure of the atmosphere and major changes can occur, for example, between day and night conditions.

The period of large scale TIDs ranges from 30 min to 3 hours and they travel at speeds ranging from 300 ms^{-1} to 1000 ms^{-1}. Their phase surfaces tilt forward in the direction of propagation with an upward component of the energy flux. These large scale events are well correlated with magnetic disturbances which suggests auroral zone sources. These properties of large scale TIDs are consistent with the postulation that they are generated by long period ducted gravity waves as discussed above. These ducted modes are dissipated over distances of less than one quarter of the earth's circumference. The longer the period of the mode and the higher its speed at a given period, the further it can propagate. Thus, large scale TIDs tend to be long period, high speed waves which travel great distances. A typical example of the distortions of the isoionic contours produced by ducted gravity waves is reproduced in Fig. 7 from the work of Francis (1973).

It should be noted that the propagation of both long and medium scale TIDs are influenced by neutral atmospheric winds. There is evidence (e.g. Morton and Essex, 1978) that the neutral wind can act as a gravity wave filter. As a consequence medium scale TIDs tend to be observed only when propagating in directions which are opposed to that of the neutral thermospheric wind.

6. EFFECTS OF TIDs ON RADIO WAVE PROPAGATION

Gravity waves travelling in the upper atmosphere produce time varying distortions in the isoionic contours. These in turn can affect the characteristics of radio waves propagating through the ionosphere and consequently, the performance of radio systems utilizing this type of propagation can be modified. It is not possible to discuss every type of radio system but a number of examples are now considered to illustrate these effects of TID activity.

(a) Changes in layer reflection heights. The redistribution of the electron density produced by the TIDs can be observed on vertical incidence ionograms. A time sequence of ionograms illustrating the change of reflection height caused by a TID is reproduced in Fig. 8. By reducing the ionograms to true height profiles the time variation of the levels of constant electron density can be investigated.

A plot of the isoionic contours derived in this way is reproduced in Fig. 9, which clearly shows the oscillatory nature of the TID and the increasing amplitude of the disturbance with height in the ionosphere. These changes can affect the propagation characteristics for an oblique incidence signal, particularly if the frequency is close to the path MUF.

(b) Fading and focussing. The TID disturbance gives rise to concave and convex distortions of the isionic contours as illustrated in Fig. 9. These features will generally produce focussing and defocussing respectively in a reflected signal. Fading of the signal is observed, the period and depth of which depends on the path geometry relative to the TID. Short paths are particularly affected, though marked TID effects on long transequatorial paths have been reported by Rottger (1972).

(c) Doppler frequency shifts. During a travelling disturbance the reflection point of a radio wave is displaced and consequently the optical path between transmitter and receiver changes during the event. This change of ray path produces a frequency shift Δf in the reflected signal of magnitude

$$\Delta f = \frac{-1}{\lambda} \frac{dp}{dt}$$

where λ is the free space wavelength and $\frac{dp}{dt}$ is the time rate of change of phase path. The travelling disturbances produce characteristic quasi-sinusoidal signatures in frequency-time recordings as indicated in Figs 10(a) and (b) for medium and large scale TIDs respectively. The frequency variations in the signals received from four transmitters separated by approximately 100 km are reproduced in the figure. Note the time delays between the various features from which (and a knowledge of the transmitter-receiver geometry) the TID vector velocity can be determined by triangulation. The magnitudes of the frequency shifts produced are quite small but can be significant for some applications, particularly where phase stability is important.

(d) HF direction finding. It is now possible to construct an HF radio direction finding system with an instrumental accuracy of about 0.1 degree. When a distant transmitter is observed by means of the ionospherically reflected signal, accuracies of only about $\pm 2^{\circ}$ are obtained. This degradation in bearing accuracy results from the presence of TIDs and static tilts in the ionosphere. During these events strong gradients exist in the isoionic contours and reflection takes place in an off great circle path direction. Thus an incorrect, i.e. non great circle, bearing is recorded by the direction finder (Jones and Reynolds, 1975). The errors produced by TIDs are particularly troublesome since they are of appreciable magnitude and their occurrence is difficult to predict. A typical recording of the effects of TIDs on HF bearings over a 850 km path is reproduced in Fig. 11. It is difficult to correct for these errors although several techniques are currently being developed (Jones and Spracklen, 1978).

(e) Trans-ionospheric propagation. Radio transmissions received on the ground from satellites are influenced by the electron distribution in the ionosphere through which they must propagate. Changes in electron content along the propagation path will modify the characteristics of the signals even though the frequency greatly exceeds the critical frequency of the F-layer. The major effects on a 100 MHz signal are:

(1) a reduction in propagation speed to less than that of the free space velocity which results in time delays;

(2) polarization rotation (Faraday effect) of linearly polarized waves;

(3) angular refraction of the ray path;

(4) phase advance of the carrier phase with respect to its free space value.

A TID will produce a time varying modulation of the total electron content along the ray path and consequently the parameters referred to above will also vary with time. This has important consequences for satellite navigation systems and in some communication applications. As typical examples of TID induced disturbances the Faraday rotation of the plane of polarization of the 137 MHz signal received from the ATS satellite during a TID is reproduced in Fig. 12 (Yeh, 1972). The sinusoidal nature of the TID is evident as are the time delays between the observation of similar features at the spaced receiving sites, indicating the travelling nature of the disturbance. The TIDs produce a rotation of about 90° which at 137 MHz corresponds to a change of approximately 10^{16} el/m^2 in total electron content. A change of this magnitude will induce a time delay of about 80 nano sec into the signal which corresponds to a range error of 25 meters.

The Faraday rotation measurements tend to be more sensitive to TID effects at or above the F-layer maximum, whereas wave refraction appears to be most sensitive to changes near the bottom of the ionosphere. The variations in phase, amplitude, angle of arrival and time of flight induced by TIDs are generally small but are nevertheless important in most types of precision satellite radio systems.

(f) UHF radars.

A signal propagating upwards from a ground based radar will be subject to disturbances induced by TID activity in a similar manner to those signals received on the ground from a satellite. The TID perturations generally oscillate through a few cycles so that the integrated effect of the TID train on the radar signal will be diminished. One can envisage that the effects produced by each half of a TID cycle will cancel each other so that the net time integrated influence of the TID is small. The cumulative effects of the TID could be large if the phase of the wave is nearly constant along the radar ray path e.g. for TIDs with wavelengths so large that $\lambda/2$ approximate to the thickness of the ionosphere. A similar effect is produced if the ray path of the radar lies nearly parallel to the crests and troughs of the TID. The order of magnitude of the disturbance has been calculated by Badura (1973) for a 400 MHz radar using a ray tracing analysis and a model TID. The range and elevation angle errors determined as a function of apparent elevation are reproduced in Fig. 13. The broken curves indicate the error

produced by the background ionosphere and the full lines those due to the TID. The errors are greatest at low elevation angles and fall off with increasing elevation angle becoming negligible above about 50° elevation. As in the case of transmission from a satellite to the ground the effects produced by the TIDs are relatively small in magnitude. They can however be important when high precision and stability are prime requirements in the radar system.

7. SUMMARY

The characteristics of the propagation of internal gravity waves in the upper atmosphere can be derived from a consideration of the hydrostatic equations. These are shown to be consistent with the wave-like disturbances in the ionosphere known as travelling ionospheric disturbances (TIDs). The two main classes of TIDs observed experimentally can be identified with two independent solutions of the gravity wave equations. Possible sources of the gravity waves have been considered and the disturbances resulting from a point impulsive source in the atmosphere are consistent with the measured parameters of TIDs.

The changes in electron density distributions in the ionosphere during TID activity have been described in detail and their influence on radio systems ranging from HF to UHF considered. It is evident that travelling disturbances can exert an important influence on many types of radio systems, generally producing a reduction in system performance. In general it is not possible to correct for the errors induced by TIDs since it is not yet possible to predict their magnitude or their time of occurrence.

REFERENCES

Badura, D.C., 1973, "Simulation of e-m wave propagation by ray tracing, methods and applications", Bell Laboratories/Lincoln Laboratories. Joint Radar Propagation Study.

Baker, D.M. and Davies, K., 1968, "Waves in the ionosphere produced by nuclear explosions", J. Geophys Res. 73, 448.

Baker, D.M. and Davies, K., 1969, "F2 region acoustic waves from severe weather", J. At. Terr. Phys., 31, 1345.

Brekke, A., 1979, The relative importance of Joule heating and the Lorentz force in generating atmospheric gravity waves and infra sonic waves in the auroral electrojets", J. At. Terr. Phys., 41, 475.

Chimonas, G. and Hines, C.O., 1970, "Atmospheric gravity waves launched by auroral currents", Planetary and Space Science, 18, 565.

Davies, K. and Baker, D.M., 1965, "Ionospheric effects observed around the time of the Alaskan earthquake of March 28 1964", J. Geophys. Res., 70, 2251.

Davies, K. and Jones, J.E., 1972, "Ionospheric disturbances produced by severe thunderstorms", NOAA professional paper No. 6, U.S. Dept. of Commerce.

Francis, S.H., 1973, "Theory and models of atmospheric acoustic-gravity waves and TIDs", Bell Laboratories/ Lincoln Laboratories, Joint Radar Propagation Study.

Francis, S.H., 1975, "Global propagation of atmospheric gravity waves. A review", J. At. Terr. Phys., 37, 1011.

Friedman, J.P., 1966, "Propagation of internal gravity waves in a thermally stratified atmosphere", J. Geophys. Res., 71, 1033.

Goerges, T.M., 1967, "Ionospheric effect of atmospheric waves", ESSA Tech. Rept., IER 57 ITSA 54.

Hines, C.O., 1960, "Internal atmospheric gravity waves at ionospheric heights", Canadian Journal of Physics, 38, 1441.

Hines, C.O. and Reddy, C.A., 1967, "On the propagation of gravity waves through a wind shear", J. Geophys. Res., 72, 1015.

Jones, T.B. and Reynolds, J.S.B., 1975, "Ionospheric perturbations and their effects on the accuracy of HF direction finders", The Radio & Electronic Engineer (J. IERE), 45, 63.

Jones, T.B. and Spracklen, C.T., 1978, "A technique for assessing the bearing accuracy of distant HF transmitters", Proc. Inst. Elec. Engers., 125, 469.

Morton, F.W. and Essex, E.A., 1978, "Gravity wave observations at a southern hemisphere mid latitude station using the total electron content technique", J. At. Terr. Phys., 40, 1113.

Pierce, A.D., 1966, "Justification of the use of multiple layer as an approximation to the real atmosphere for acoustic gravity wave propagation", Radio Science, 1, 265.

Pitteway, M.L.V. and Hines, C.O., 1965, "The reflection and ducting of atmospheric acoustic gravity waves", Canadian Journal of Physics, 43, 2222.

Rottger, J., 1972, "Some effects of atmospheric gravity waves observed on a transequatorial radio path", AGARD Conf. Proc. CP 115, Ch. 33.

Yeh, K.C., 1972, "Travelling ionospheric disturbances as a tool for thermospheric dynamics", J. Geophys. Res., 77, 709.

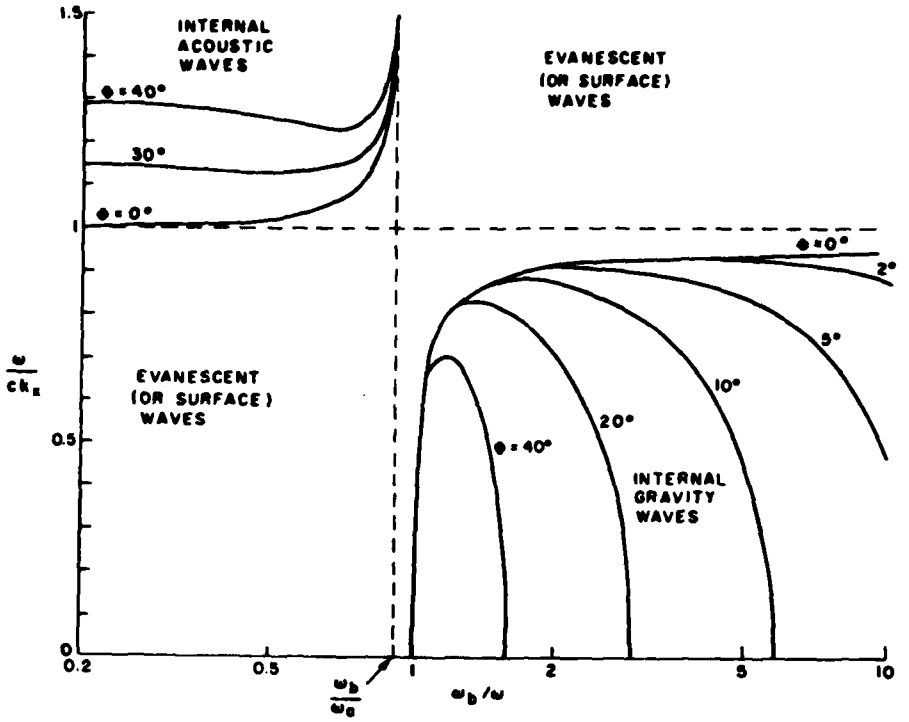


Fig.1 Curves relating normalized horizontal phase velocity (ω/ck_x) to normalized wave period (ω_b/ω) for various values of Φ, the angle between the group velocity and the horizontal (Francis, 1973)

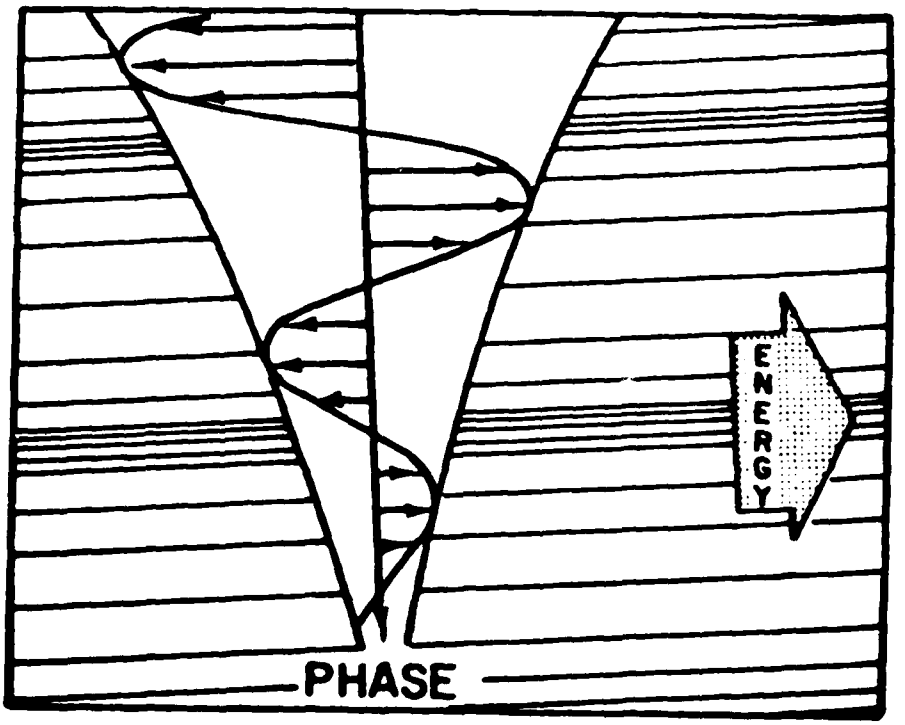


Fig.2 Pictorial representation of internal atmospheric gravity waves. Instantaneous velocity vectors are shown, together with their instantaneous and over-all envelopes. Density variations are depicted by a background of parallel lines lying in surfaces of constant phase. Phase progression is essentially downwards in this case, and energy propagation obliquely upwards; gravity is directed vertically downwards (Hines, 1964)

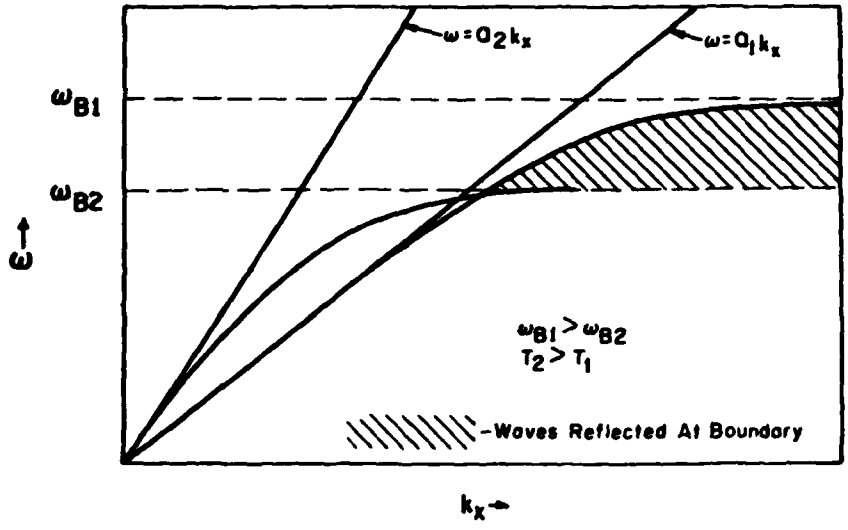


Fig.3 Reflection conditions for internal gravity waves in a two-layer model (Georges, 1967)

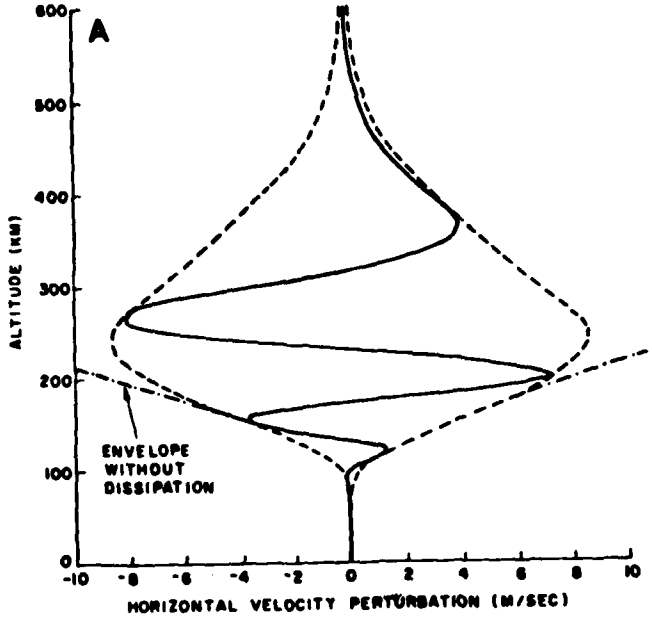


Fig.4 Horizontal velocity perturbation vs height for a gravity wave with parameters typical of medium-scale TIDs ($2\pi/\omega$ = 25 min, ω/k_x = 200 m/sec). Solid curve – instantaneous phase structure; Broken curve – envelope within which the phase structure propagates as a function of x and t; dash-dot curve – envelope without dissipation (Francis, 1973)

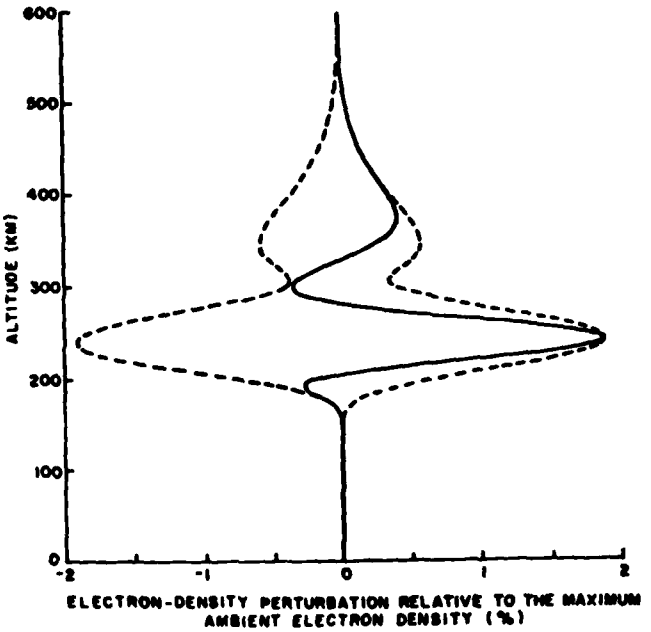


Fig.5 Electron-density perturbation induced by the gravity wave of Fig.4 (Francis, 1973)

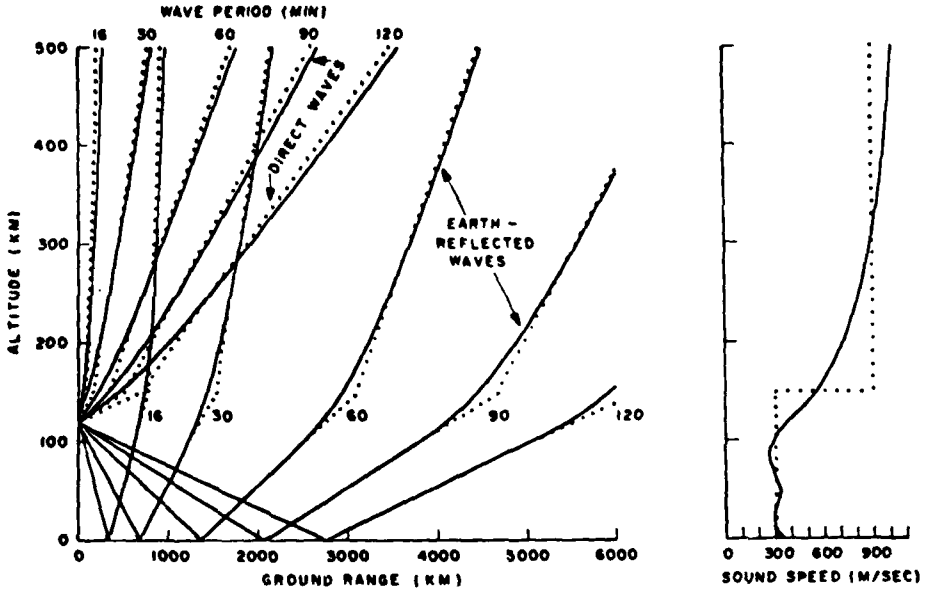


Fig.6 Asymptotic gravity-wave ray paths for direct and Earth-reflected waves, plotted using the CIRA 1965 model atmosphere (solid curves) and a simple step-function atmosphere (broken curves) (Francis, 1973)

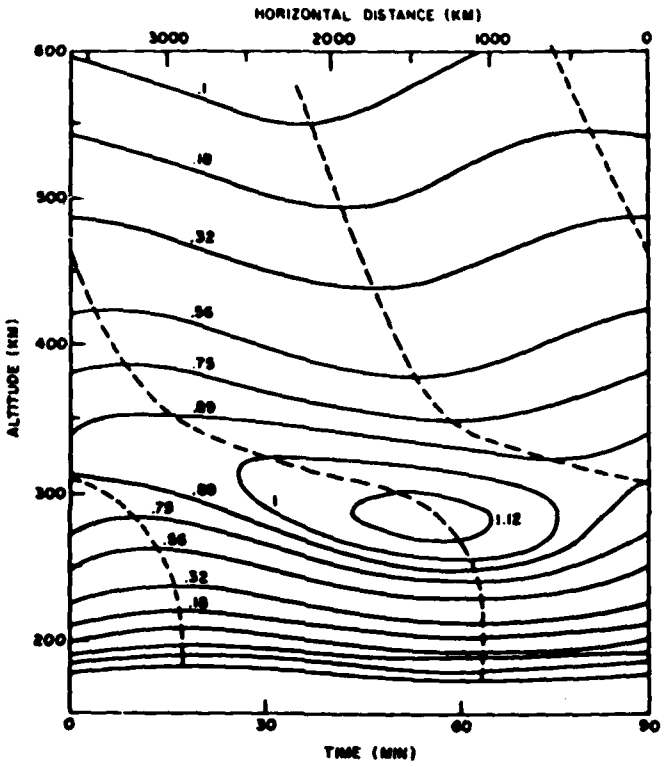


Fig.7 Electron-density contours of a TID caused by the fundamental gravity mode (Gó) with a period of 90 minutes and a horizontal phase speed of 677 m/sec (Francis, 1973)

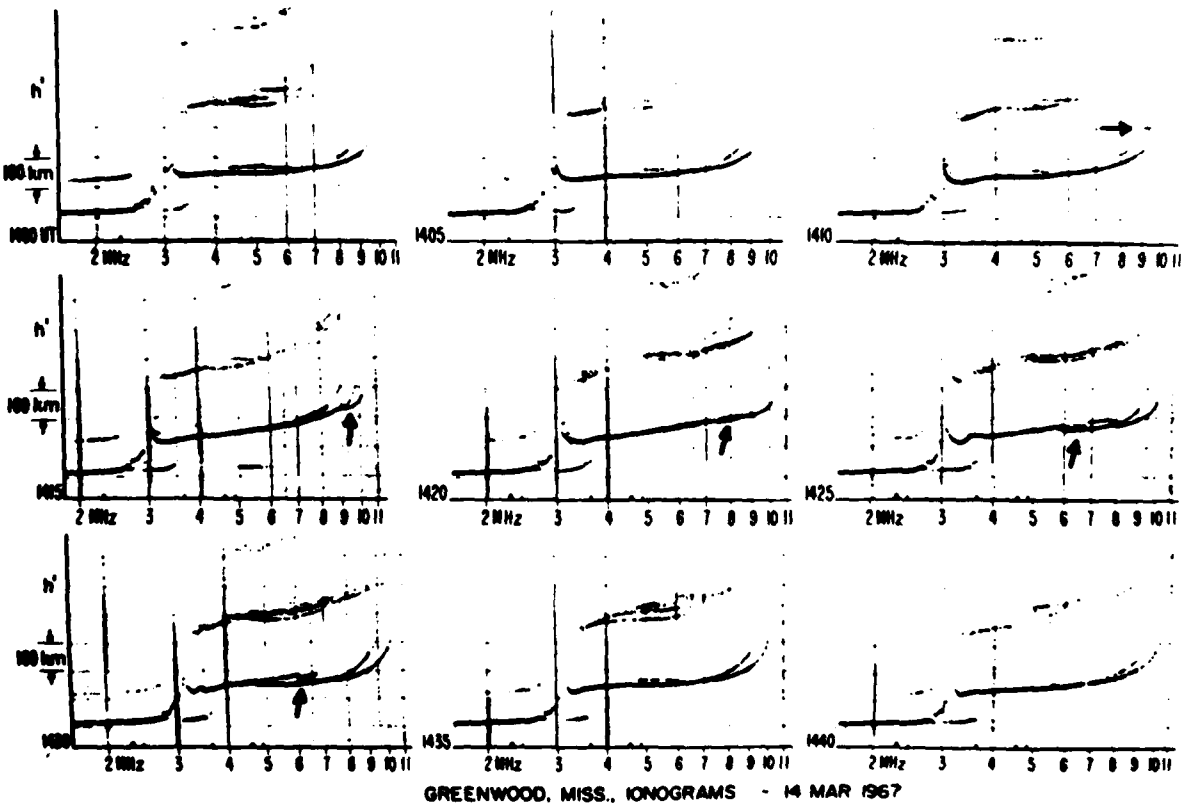


Fig.8 Ionograms showing a "medium-sized", localized travelling disturbance (Georges, 1967)

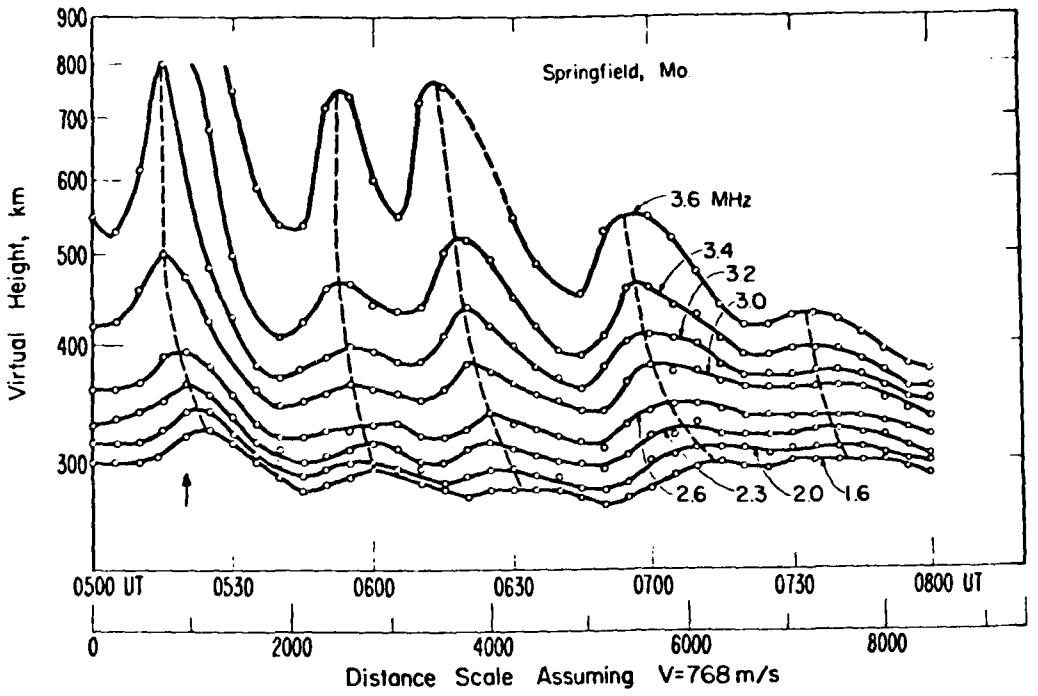


Fig.9 Virtual-height contours for the disturbance of 14 December, 1966 (Georges, 1967)

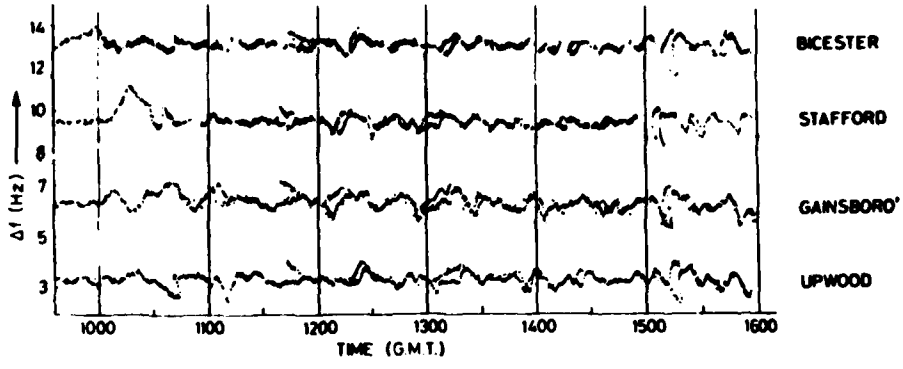


Fig.10(a) Typical daytime TID activity observed by the Leicester four-station Doppler

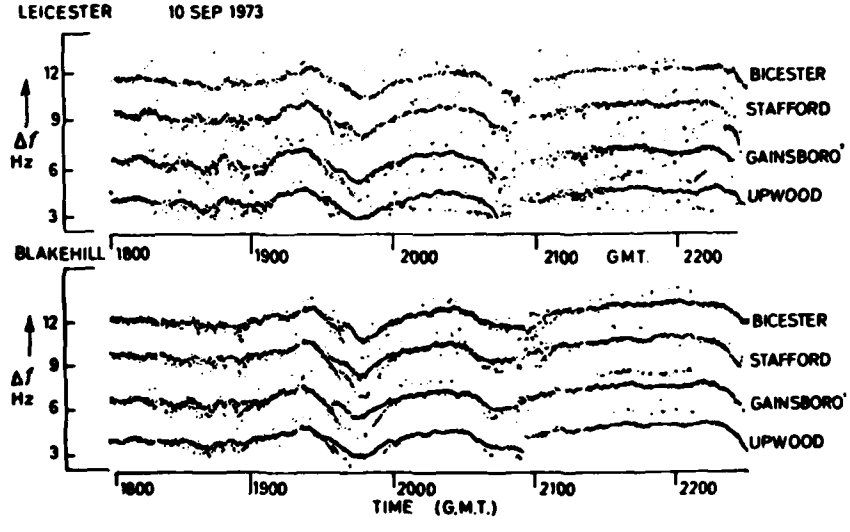


Fig.10(b) Examples of large-scale travelling waves observed during night-time at Leicester and Blakehill (near Cheltenham)

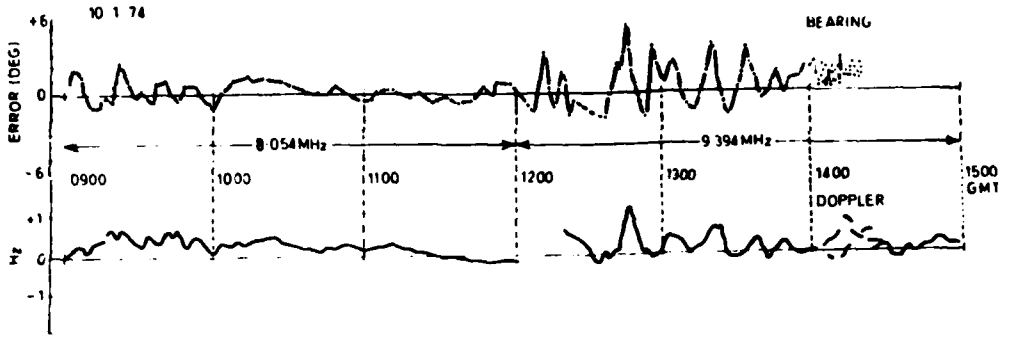


Fig. 11 Bearing error and Doppler frequency shifts recorded on transmissions from Breisach, West Germany

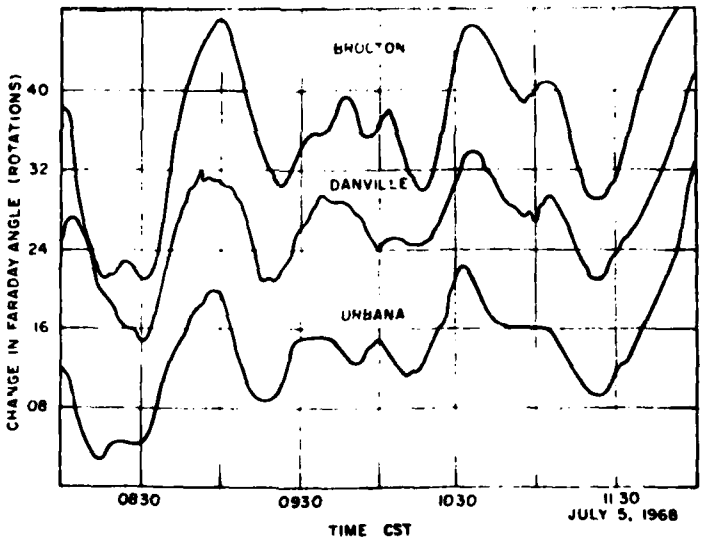


Fig. 12 Relative Faraday rotation of signals transmitted by ATS 3 and recorded at Urbana, Danville, and Brocton on July 5, 1968 (Yeh, 1972)

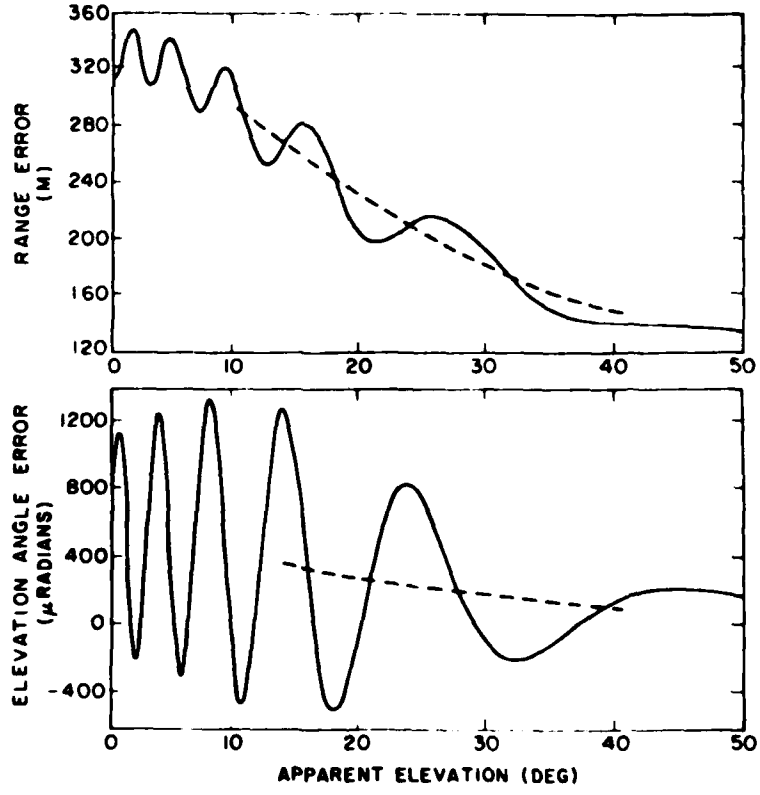


Fig.13 The range and elevation angle errors induced by the ambient and TID-perturbed ionospheres (broken and solid curves respectively) for a radar tracking a satellite at 1000 km altitude passing overhead towards geomagnetic north (Francis, 1973)

SESSION DISCUSSION

SESSION V - IRREGULARITIES AND WAVES

Chairman and Editor Dr G.M.Brown
Dept. of Physics
University College of Wales
Aberystwyth, Dyfed, UK

MID LATITUDE SPORADIC-E LAYERS
by L.G.Smith and K.L.Miller

K.Rawer, Ge

The question of whether Es-layers are cloudy or homogeneous was, in fact, discussed twenty five years ago by analysing the transparency range between 'blanketing' and 'top' frequencies appearing on ionograms. A statistical study, published in the book "Sporadic E" by E.K.Smith and S.Matsushita, led to the result that, at midlatitudes, the majority of all observed layers (say 80%, and almost all cases of exceptionally high top frequency) could not be explained with the transparency of a homogeneous thin layer. The point I wish to make is that there is an important latitude dependence of the structure of E_s, equatorial layers being almost always of the cloudy type. It would therefore be interesting to have incoherent scatter observations such as those from Arecibo from midlatitude stations e.g. St. Santin.

Author's Reply

I did not attempt to review the long history of radio studies of sporadic-E layers. Indeed, many of the properties of these layers have been anticipated by radio experiments. The rocket observations have the advantage of good height resolution but are very limited in number, even for a particular launch site. Similarly, the incoherent-scatter radar data are available from very few locations though the temporal variation of a layer can be observed with resolution of one minute. For global morphology we still rely mainly on ionosondes, though these are not ideal experiments for observing sporadic-E layers.

M.Crochet, Fr

I wish to ask a general question about the origin of sporadic-E at mid-latitudes. We know, as indicated in your paper, that an important source of sporadic-E at mid-latitudes is wind shear acting on metallic ions, but we know also from theory that another possible source operating at such latitudes is the gradient-drift instability. Do you know of any experimental proof from radar or rockets that this mechanism might be operating at mid-latitudes?

Author's Reply

I am not aware of any observations which show the gradient-drift instability in mid-latitude sporadic-E layers. It would be expected that this mechanism would produce small-scale irregularities and would not explain the major complex structure described in the paper. Our rocket instrumentation has the ability to detect features down to 1m, such as are found in equatorial sporadic-E, but has not recorded them in mid-latitude sporadic-E layers.

G.M.Brown, UK

A number of the rocket measurements of winds have indicated that sporadic-E sometimes forms at places where there is a wind shear of the wrong sign, i.e. where one would expect ion divergence rather than convergence. Has there been any explanation of this?

Author's Reply

The conventional explanation is that there is an electric field of external origin (not the polarization electric field). As indicated by equation (2) in the paper, the electric field has the effect of a systematic change in the local wind and can easily be incorporated in the analysis. Although fields of reasonable magnitude are postulated to explain the discrepancies referred to, there have not been simultaneous direct measurements of electric field to confirm this explanation.

W.I.Axford, Ge

(a) Is it usual, as shown in your figure, that the Si^+ ions tend to be absent in the lower layers but quite prominent in the higher layers?

(b) It seems to me that it is possible that complex layer structure could be a result of time-dependent layer formation as well as complex wind profiles. But in any case it is interesting to consider the possibility of such wind profiles and I would like to ask if there have been observations with sufficient resolution to detect these.

Author's Reply

(a) There is a problem with identification of the Si^+ in ion mass spectrometer data since it is indistinguishable from N_2^+ (both 28 a.m.u.). It is unlikely that there is a substantial concentration of Si^+ at altitudes above 120 km.

(b) The analysis of the complex layer shapes was based on a steady-state solution of the continuity equation. This assumption clearly needs to be tested. We initially considered the possibility that the double-peaked layers represented the partial merging of single-peaked layers but could not obtain a satisfactory explanation on that basis. Further, in the case of two sequential rocket observations mentioned in Smith and Miller (1980), there is support for the evolution of a single-peaked layer into a double- peaked layer.

There have been many observations of neutral wind by the vapour-trail technique. These show that shears of the required magnitude exist. The height resolution of these observations, however, has not been adequate to see the detailed structure on the height scale ($\ll$ 1 km) that is inferred from the electron density profiles. It is possible that a specially designed experiment might be able to show the existence of the fine structure.

Comment by R.A.Goldberg, US

In response to W.I.Axford's question (a) regarding the high abundance of Si^+ ions above 100 km shown in the paper, Si^+ appears as a strong metallic ionic constituent in layers above 100 km. Below this level Si^+ chemistry with O_2 causes it to deplete rapidly so that it is no longer observed. Within the layers above 100 km there is no confusion with N_2^+ since the gaseous molecular species are depleted by recombination with the highly abundant electrons. Other than in the sporadic-E layers the 28^+ mass is likely to be N_2^+.

THE SEASONAL AND GEOGRAPHICAL VARIATION OF EQUATORIAL SPREAD-F IRREGULARITIES INFLUENCED BY ATMOSPHERIC GRAVITY WAVES AND ELECTRIC FIELDS DUE TO THUNDERSTORMS

by J.Röttger

H.Volland, Ge

The model used by Hays and Roble has a moderate spatial resolution and considers only average thunderstorm activity. For an estimate of local thunderstorm fields at ionospheric heights, the model of Park seems to be more appropriate.

Author's Reply

This would yield an improvement if it were possible to carry out case studies for a localized area, but this does not seem to be practicable from the experimental viewpoint. In my opinion, however, the global effect of internal gravity waves appears to be more important that the electric field control, and I would question whether future efforts to carry out case studies of the influence of thunderstorm-generated electric fields on ionospheric irregularities are worthwhile.

J.Aarons, US

(a) It should be emphasized that only range spread-F data should be used in analyses of this type.

(b) The summer/winter differentiation is not valid for field-aligned equatorial observations.

(c) Hard evidence is needed to support the very valid suggestion of trophospheric initiation of F-layer irregularities.

Author's Reply

(a) This has effectively been done in all the investigations of the author, e.g. by restricting to large scale wavelike structures of equatorial spread-F (see appropriate references listed in my paper).

(b) It is assumed that the question refers to observations that irregularity occurrence follows a fieldline, viz. correlation of occurrence at both "terminals" of the fieldline. This has to be considered and may account for some of the discrepancies occurring in seasonal variations. The summer/winter picture, however, is more complicated than that implied in the question. As pointed out, it is not so much "summer/winter" differences but the position of the region of interest with respect to the intertropical convergence zone which has to be taken into account in the author's model, and this zone differs considerably from the geographic equator which defines summer/winter seasons. Furthermore, field-alignment depends on magnetic coordinates, making the definition of summer/winter in the conventional sense even less meaningful.

(c) There is a reasonable and accepted theory on the spatial resonance mechanism: steepening TIDs which seed large-scale equatorial spread-F bubbles (see references in my paper). There is also reasonable evidence that internal gravity waves, being responsible for TIDs, are connected with dynamical processes in the troposphere. No other theory or interpretation of experimental results is known which can explain wavelike structures in the equatorial spread-F appearance and their general occurrence maxima over large continents.

It should be stressed that it never was claimed that tropospheric influences can explain equatorial spread-F other than some of its numerous features.

J.Forbes, US

Don't you think that use of annual average rainfall to deduce geographical distributions of convective activity can be misleading? For instance, in India the intense convective activity that occurs during the monsoon season is washed out in an annual average, as shown by your figure.

Author's Reply

The model takes into account the mean position of the intertropical convergence zone during northern winter when the scintillation data were obtained. During this time only secondary convergence areas are south of India, so that the average rainfall rate with slight maxima over the Indian subcontinent does not show up in the model assumption. On the other hand, it is accepted that thunderstorms during the monsoon season are of a different nature from the high-reaching cumulonimbus convection over large tropical land masses which evidently cause penetrative convection resulting in gravity wave generation.

INSTABILITES DE PLASMA DANS LES ELECTROJETS
PLASMA INSTABILITIES IN THE ELECTROJETS

par M.Crochet

J.Forbes, US

You mentioned that your fit to the data assuming linear theory for a 7 km scale size was improved by taking into account neutral wind effects. How were the winds obtained to make this correction, and would this correction be relatively more or less important in consideration of the auroral electrojet?

Author's Reply

The neutral wind profile was determined using the 29 MHz radar measurements at different ranges (ref: HANUISE and CROCHET, 6th Int. Symposium on Equatorial Aeronomy, Porto Rico, 1980). The neutral wind effect will not be so important in the auroral zone if marginal stable waves ("type 1") are absent.

Comment by J.Röttger, Ge

I wish to comment that this work is another example of experiments carried out using transmissions with considerable bandwidth on frequencies allocated (exclusively) to the amateur radio service. It should be noted that this could be misleading to potential (military) users of the scientific results who may erroneously assume that any kind of operation is authorised in these bands.

DYNAMICS OF THE DAYSIDE CUSP AURORA

by A.Egeland, P.E.Sandholt, K.Henriksen, C.S.Deehr and G.G.Sivjee

G.P.Gregori, It

Can you state whether the activity in the noon sector starts slightly earlier or later than the activity in the midnight sector? In any case, what is your present time resolution?

Author's Reply

Our data are based on standard instrumentation and thus the time resolution is better than ± 2 min. It is not possible to conclude where the activity starts first along the oval, but probably the auroral disturbances start in the night sector slightly before they do so in the day side cusp.

J.S.Nisbet, US

Have you been relating your observations in the dayside cusp region to the interplanetary magnetic field?

Author's Reply

We have not yet received interplanetary magnetic field data (IMF) for the actual recording periods, and have thus not compared our dayside cusp results with IMF-data. I would like to add that because of the intimate connection between the night-side substorm activity and the dayside cusp aurora, we are very uncertain about the relation to IMF during disturbed conditions.

LARGE-SCALE WAVES IN THE IONOSPHERE OBSERVED BY THE AE SATELLITES
by S.H.Gross, C.A.Reber and F.Huang

J.Röttger, Ge

Assuming that the phase shift between N_2 and n_e is significant, what is a possible physical reason for it?

Author's Reply

Phase shift arises in the relative responses as a result of the different masses of the various gases and the ionization, as well as different temperatures. In the thermosphere one gas slips relative to another due to diffusion. Thus, helium is usually phase shifted through a large angle relative to nitrogen.

THE GENERATION AND PROPAGATION OF ATMOSPHERIC GRAVITY WAVES WITH SPECIAL REFERENCE TO RADIO PROPAGATION
by T.B.Jones

J.Aarons, US

Earlier data on TIDs originating in the auroral oval seem to me sparse and selective. What are your views relative to the auroral oval origin of TIDs?

Author's Reply

Small and medium scale TIDs do not necessarily originate in the auroral zone and other sources e.g. tropospheric storms, have been identified. Large scale TIDs are normally thought to have auroral zone sources. However, even for these events non-auroral sources have been identified. I would stress that the neutral wind can act as a TID filter and the TID direction observed is the combination of the source location and azimuthal filter imposed by the neutral wind.

K.Rawer, Ge

The upwards attenuation of gravity waves seems still to be rather uncertain. At the time when Hines proposed gravity waves to explain travelling ionospheric disturbances theoretical estimates indicated such strong attenuation that the waves should not be able to reach the F2-peak. The results of Francis you have shown give about 800 km as the limiting level. I remember, however, a French investigation of topside ionograms from Alouette I where TIDs were seen up to 1000 km, the height of the satellite. Thus it seems that the attenuation of gravity waves is still not well understood. Perhaps neutral winds could also have some influence in this context.

Author's Reply

The attenuation of the gravity wave will depend on factors such as viscosity, thermal conductivity and ion drag. Whether the wave will propagate to a particular height will also depend on the atmospheric parameters such as the neutral wind vector and on the atmospheric parameters such as the temperature profile which determines whether ducting, etc., can occur.

It should be noted that a TID is the consequence in the ionosphere of the passage of a gravity wave. Since the electron mobility is much larger along the geomagnetic field lines than in any other direction, the electron density profile change is not necessarily identical to that of the gravity wave which produces the TID.

W.I.Axford, Ge

(a) In reply to Prof. Rawer's comment I would like to emphasize that TIDs are in themselves not gravity waves but are only associated with gravity waves. Thus the presence of a TID at very high altitudes does not imply that there is a gravity wave at those altitudes – probably the wave is at low (F-region) altitudes and its effects on the plasma are communicated via the geomagnetic field to high altitudes.

(b) What is the role of surface waves (i.e. waves which propagate around the surface of the Earth analogous to Lamb waves in seismology) in the atmosphere? Presumably disturbances at larger distances from the source must contain such a component.

Author's Reply

(a) I agree. The gravity wave is essentially a perturbation of the neutral atmosphere and the electon density change is a result of this. The electron displacement will not necessarily be the same as that of the neutrals since, for example, the electron mobility is larger along the geomagnetic field lines than across them.

(b) The Lamb wave will propagate along the surface of the Earth with little attenuation. Indeed it is possible for waves of this kind to travel two or three times around the Earth's circumference, as reported in the case of the Krakatoa eruption. However, waves of this kind are attenuated strongly as they move away from the surface and their influence at thermospheric heights is quite small.

PREDICTION TECHNIQUES FOR FORTHCOMING SOLAR MAXIMA

G. M. Brown
Department of Physics, University College of Wales, Aberystwyth, U. K.

ABSTRACT

This paper reviews the various techniques which have been used to predict the magnitude of a forthcoming sunspot maximum. Details of individual methods are not considered, but full bibliographical references are given. Most of the techniques are discussed in relation to their applications to the current solar cycle number 21. During the last decade there has been substantial improvement in prediction abilities arising from the realization that a solar cycle is being built up for several years prior to its conventional beginning at sunspot minimum. This means that information about the nature and size of an up-coming cycle must be contained in both solar and terrestrial data obtained during the declining phase of an 'old' cycle and around the minimum. Various prediction applications of this principle are discussed which show considerable promise. These methods break new ground in having a degree of physical backing rather than being purely empirical or dependent on the existence of periodicities in the sunspot number time series. Significant improvement in the future may confidently be expected.

1. INTRODUCTION

There are many scientific, commercial, and economic reasons why it is desirable to be able to predict the magnitude of a forthcoming maximum of solar activity in the eleven-year cycle. This poses quite a problem since measurements over the last two centuries show that the size of the maximum has varied from cycle to cycle in an apparently random manner. Nevertheless over recent cycles many attempts have been made to predict at least the likely size of the next maximum, using a variety of techniques. Unfortunately most of the earlier attempts were largely empirical and lacked physical understanding of the underlying mechanisms, with the result that success rates have been unspectacular.

In recent years there has been considerable progress in understanding how a solar cycle is built up, and in parallel with this a new concept has come into prediction applications which shows promise. In this paper we briefly review the various solar forecasting techniques so far employed, with particular emphasis on the more recent developments. It should be noted that we shall be entirely concerned with the problem of predicting the magnitude of a subsequent solar maximum and not with the time of its occurrence. There are several empirical (inverse) relationships between the risetime of a cycle from its commencement at solar minimum and the peak value of sunspot number achieved in that cycle, so to a limited extent one of these parameters can always be derived from the other. We shall discuss solar activity in terms of the conventional Zurich sunspot number R, but, in passing, it must be pointed out that the cycle-to-cycle variability in the size of peak R is not necessarily mirrored in other indices of activity, for example solar faculae.

2. SUMMARY OF PREDICTION TECHNIQUES

The wide range of techniques employed and results obtained in attempting to forecast the magnitude of a forthcoming sunspot maximum are well illustrated by considering the prognostications for the current cycle number 21. This cycle conventionally began at the minimum in sunspot number which occurred in June 1976, and after a relatively slow start activity increased rapidly. The maximum has probably now passed and the peak 13-month (R_m-max) running mean value of R will be high, probably equivalent to, or greater than, that of cycle 18 (R_m-max = 152). The plethora of predictions for cycle 21 is illustrated in Figure 1: no reasonable value for the magnitude of sunspot maximum has not been proposed. It should be noted that in compiling this histogram no distinction has been made between values of peak annual mean (R_a-max) and peak smoothed monthly mean (R_m-max): some authors quote one, some the other; some do not state which. In practice, R_m-max is always greater than R_a-max, with an average difference over the last thirteen cycles of 4%, although for cycle 12 the difference was as large as 17%. Where authors quote a prediction range, the middle value has been plotted in Figure 1. Brief consideration of these predictions for cycle 21 will usefully serve to illustrate most of the techniques employed. It will also appear that there is, in fact, some structure to this histogram.

2.1. Statistical Treatment of Past Solar Cycles

Of all the methods used, those which depend on statistical analyses of the past time series of sunspot numbers are by far the most numerous. They depend on the existence of periodicities in the data and many different techniques have been employed. In recent years spectral analysis has been a favoured approach.

Projections a short time ahead (up to one year) can be fairly accurate (e.g. the regression analysis technique used in the McNish-Lincoln (1949) method), but uncertainties accumulate rapidly for longer time scales where the predictions are based on extrapolations of already predicted data. The methods usually give an estimate of the forthcoming trend and are not designed specifically to forecast the magnitude of the peak of a cycle. Some techniques, however, are geared to the prediction of this parameter, such as those based on the statistics of the distribution of the sizes of the peaks and on the existence of long period cycles in the magnitude of the maxima. On Figure 1 the following references make use of statistical analyses in one form or another: Bonov 1970, Cohen and Lintz 1974, Cole 1973, Fougere 1979, Gleissberg 1971, Henkel 1971, Hill 1977, Hunter 1979, Peking Observatory 1979, Reddy et al 1979, Sakurai 1977, Schove 1955, Suda 1979, Vasilyev and Kandaurova 1971 and 1979, Vasilyev et al 1975, Vitinski 1962, Zhen-Tao et al 1979. Other relevant references include: Chadwick 1959, Herrinck 1959, Kalinkov 1968, Minnis 1960 and 1961.

2.2. Initial Progress of a Cycle

Another commonly adopted prediction technique seeks to establish quantitative relationships between the rate of rise of sunspot number over the first 1½ to 2 years following sunspot minimum and the magnitude of

the peak subsequently attained. In general, these parameters are directly related to each other, but no universally valid equation has been deduced. The accuracy obviously improves as the lead-time to the maximum decreases and the actual cycle progression becomes more clearly established. Methods listed on Figure 1 which are based on this approach include Cole 1973, Jose 1965, King-Hele 1966, Reddy et al 1979, Waldmeier 1964, Xanthakis 1967, Xanthakis and Poulakos 1978 and 1979, Zhen-Tao et al 1979.

A related technique has recently been developed by Brown and Evans (1980). This depends on a study of the development of sunspots with respect to photospheric faculae during the early stages of a cycle, rather than the temporal variation of sunspots alone over this time. An apparently close and direct relationship between the rate of this development for a given cycle and the magnitude of the peak sunspot number reached has been found. This is physically reasonable in that the greater is the intensity of a cycle (measured by R_m-max) the greater is the area of sunspots on the disc relative to the area of the faculae within which the spots form, at any time including the early stages of a cycle. The reference Brown 1979 (SVF:spots versus faculae) in Figure 1 refers to this technique.

2.3. Planetary Positions

The concept that tidal action of the planets, related to their alignment, plays a role in the development of a solar cycle has been proposed many times but has never been either convincingly established or convincingly refuted. Accordingly, predictions based on the supposed relationships (Jose 1965, Smith 1976 in Figure 1; also Anderson 1954, Bagby 1975, Bigg 1967, Okal and Anderson 1975, Wood 1965 and 1972) must be treated with reserve.

2.4. Phenomena during the Preceding Cycle and near Sunspot Minimum

Within the last decade a new approach has appeared which as well as showing promise has the added strength that it has some definite, though as yet rather ill-defined, physical basis. In fact, the realization of the principle employed has modified our whole concept of the length of a real solar cycle. This will be discussed in the next section, followed by a summary of the various applications to prediction so far reported.

3. WHEN DOES A SOLAR CYCLE BEGIN?

Traditionally, the first appearance of the new-cycle sunspots in middle solar latitudes, overlapping the old-cycle spots around the solar equator near the end of its cycle, has been taken as the signature of the commencement of a new cycle. On this picture a solar cycle begins at sunspot minimum. However, this is an over-simplification. Sunspots are visual manifestations of the build-up of sub-surface solar magnetic fields of a critical intensity and the build-up process will have been proceeding for several years prior to the 'break-out' of the first spots.

One of the earliest indications of the beginning of a solar cycle currently known to solar astronomers is a reversal of the polar magnetic field of the Sun which takes place separately in each hemisphere (and at different times) within a year or two of the preceding sunspot maximum, i.e. anytime from 2 to 5 years before minimum (Howard 1977). According to the dynamo theory (Babcock 1961, Leighton 1969) it is out of these polar fields that the toroidal magnetic flux is wound by the Sun's differential rotation, eventually being amplified to an intensity sufficient to generate the next cycle of sunspots. Thus, the next cycle is probably being built during most of the declining phase of the preceding cycle. During this phase also there are several other interesting solar happenings: e.g. the equatorial rotation accelerates and there appears to be a reorganization of the large-scale solar structures with the development of large unipolar features underlying long-lived coronal holes. These coronal holes are the source of solar wind streams which are faster and wider during the declining phase than in any other part of the solar cycle.

Figure 2 illustrates a terrestrial ramification of the importance of the declining cycle in terms of geomagnetic disturbance for the average of solar cycles 12 to 19 (1879-1965) synchronizing the minima in R. The centre panel shows that the minimum of the magnetic activity index 'aa' coincides with sunspot minimum, but that there are two peaks of activity, one before sunspot maximum and one on the declining cycle. The first peak is commonly attributed to the influence of flare-associated, sudden-commencement type magnetic storms, and the second peak to coronal hole (M-region)-associated recurrent storms. The lower panel in Figure 2 shows the corresponding variation of the auto-correlation of 'aa' with a 27-day lag, and confirms that 27d recurrent magnetic activity is most evident on the declining phase of the cycle, towards sunspot minimum.

Thus, physically, there is sound evidence to believe that by the time of sunspot minimum, when a cycle conventionally begins, the type and size of cycle ahead will be already largely determined. It is therefore logical that there should be precursors to solar activity in a new cycle several years before sunspot minimum. This is the basis of the growing number of solar predictions which employ both solar and terrestrial data obtained during the declining phase and at the minimum of an 'old' cycle. The eleven predictions indicated with stippling on Figure 1 are all based, in one form or another, on this concept. It is interesting to note that all these methods have predicted a fairly large maximum for cycle 21. In the next section we summarize the techniques so far employed.

4. PRECURSORS TO SOLAR ACTIVITY

4.1. Methods based on Magnetic Activity

It is evident from the results in Figure 2 that magnetic activity, and in particular recurrent-type magnetic activity, during the declining phase of a solar cycle may be a terrestrial teller of the build-up of the next cycle. Pioneer work in this area was published by Ohl (1971) in which he showed that the difference between a magnetic activity index based on Kp and sunspot number over the whole declining cycle was highly correlated with the peak sunspot number of the following cycle. Improved results were subsequently obtained

(Ohl 1976) by restricting the data to the three years preceding sunspot minimum; Ohl (1979) obtained similar results using a "recurrent aurora" index over the same period.

Ohl's work was extended by Sargent (1977,1978) using the magnetic index aa and developed it into a useful prediction technique. Valtchuk and Feldstein (1978) correlated aa (over one year and over three years before sunspot minimum) with peak succeeding sunspot number by plotting the parameters logarithmically. Simon (1979) noted an equivalence between maximum magnetic activity (aa) on the declining phase and the first maximum of magnetic activity near sunspot maximum (cf. Figure 2) of the following cycle, which itself is well correlated with the maximum sunspot number of that cycle.

The original work by Ohl (1966) concerned a relation between the minimum average annual value of a magnetic index (based on Kp) and the magnitude of R-max following. A more sophisticated version (using aa) along the same lines was developed by Kane (1978).

4.2. Use of "Abnormal Quiet Days"

The principle behind this method was described by Brown and Williams (1969) and Brown (1974); its use as a prediction technique is outlined by Brown (1979). The basic result employed is that the number of international quiet days (magnetic) when the solar diurnal variation of the horizontal magnetic field at a middle latitude station has an abnormal phase at sunspot minimum correlates closely with the magnitude of R-max of the following cycle. This rather curious relationship has been studied further by Brown and Butcher (1981) in the light of their improved understanding of the nature and cause of AQDs (Butcher and Brown 1980 and 1981), and evidence is presented which suggests that the main factor which determines the connection may be quantitatively related to the magnitude of southward swing of the component B_z of the interplanetary magnetic field normal to the ecliptic plane.

4.3. Methods based on Sunspot Number

It was shown by Brown (1976) that the correlation coefficient between secularly-smoothed (over five cycles) values of R-max and preceding R-min is significantly higher than that between R-max and any other R-min, in particular the following R-min. This result, however, is not directly amenable to prediction applications as it stands because of a varying sensitivity of the two parameters over the 80-year cycle and also because the secular smoothing ensures that data are always two and a half cycles in arrears. Nevertheless, the same phenomenon was used by Wood (1979) for prediction purposes, employing annual mean sunspot numbers.

Ramaswamy (1977) developed a relationship between the relative skewness of a sunspot number cycle and the ratio of the values of R-max in the next and the current cycles and discussed prediction applications.

4.4. Methods based on Solar Phenomena

The earliest mention of a precursor to solar activity of the type under discussion was given by Waldmeier (1939), although it could not be used for prediction purposes. A smoothed analysis for seven solar cycles showed that the average smoothed heliographic latitude at which sunspots of a new cycle develop is correlated with the subsequent maximum activity reached. In fact such a relation is inherent in the Babcock-Leighton theory of sunspot formation. On the basis of the same theory, Schatten et al (1978) argue that the magnitude of the solar polar field near sunspot minimum should anticipate the size of the next cycle which is generated from it, and by considering four separate methods of estimating this field they propose a linear relationship to the succeeding value of R-max in each case. They use the method to predict the size of cycle 21.

In addition to the recent suggestion by Brown and Butcher (1981) mentioned in section 4.2 that the interplanetary magnetic field (IMF) may be a useful parameter for prediction purposes, Brown (1979) reported a likely relation between the occurrence of long-lived away-sectors of the IMF near sunspot minimum and the magnitude of R-max following.

5. CONCLUSION

Few solar activity prediction techniques claim intrinsic accuracies of better than 20 per cent. It is therefore impracticable to pick a "best method" in the light of subsequent performance (cf. the narrow differences between varying methods applied to cycle 21 in Figure 1). It is probably significant, though, that the high predictions for cycle 21 are almost the exclusive prerogative of applications of the principle that the magnitude of a cycle is predetermined at or before the preceding solar minimum, and the signs are that cycle 21 will vindicate the general principle. As discussed, there is also substantial physical backing for the concept and it seems certain that refinements and improvements to these techniques over the next few years will lead to significant advances in our prediction abilities.

REFERENCES

ANDERSON, C. N., 1954, Notes on the sunspot cycle. J. Geophys. Res. 59, 455-461.

BABCOCK, H. W., 1961, The topology of the sun's magnetic field and the 22-year cycle. Astrophys. J. 133, 572-587.

BAGBY, J. P., 1975, Sunspot cycle periodicities. Nature 253, 482.

BEZRUKOVA, A. Y., 1959, Soln. Dann. No. 11, 72-74.

BIGG, E. K., 1967, Influence of the planet Mercury on sunspots. Astron. J. 72, 463-466.

BONOV, A. D., 1970, Soln. Dann. No. 7, 111-113.

BROWN, G. M., 1974, A new solar-terrestrial relationship. Nature 251, 592-594.

BROWN, G. M., 1976, What determines sunspot maximum? Mon. Not. Roy. Astron. Soc. 174, 185-189.

BROWN, G. M., 1979, New methods for predicting the magnitude of sunspot maximum. Solar-Terr. Predictions Proc., Boulder, Vol. 2, 264-279.

BROWN, G. M. & BUTCHER, E. C., 1981, The use of abnormal quiet days in Sq(H) for predicting the magnitude of sunspot maximum at the time of preceding sunspot minimum. Planet Space Sci. (in press).

BROWN, G. M. & EVANS, D. R., 1980, The use of solar faculae in studies of the sunspot cycle. Solar Phys. 66, 233-243.

BROWN, G. M. & WILLIAMS, W. R., 1969, Some properties of the day-to-day variability of Sq(H). Planet. Space Sci., 17, 455-470.

BUTCHER, E. C. & BROWN, G. M., 1980, Abnormal quiet days and the effect of the interplanetary magnetic field on the apparent position of the Sq focus. Geophys. J. Roy. Astron. Soc. 63 (Dec.)

BUTCHER, E. C. & BROWN, G. M., 1981, On the nature of abnormal quiet days in Sq(H). Geophys. J. Roy. Astron. Soc., 64 (Feb).

CHADWICK, W. B., 1959, Prediction of sunspot numbers for cycle 20. Nature 184, 1787.

COHEN, T. J. & LINTZ, P. R., 1974, Long term periodicities in the sunspot cycle. Nature 250, 398-400.

COLE, T. W., 1973, Periodicities in solar activity. Solar Phys. 30, 103-110.

FOUGERE, P. F., 1979, Sunspots: power spectra and a forecast. Solar-Terr. Predictions Proc., Boulder, Vol. 3, A1-A9.

GLEISSBERG, W., 1971, The probable behaviour of sunspot cycle 21. Solar Phys. 21, 240-245.

HENKEL, H. R., 1971, A feature of the secularly smoothed maxima of sunspot frequency. Solar Phys. 20, 345-347.

HERRINCK, P., 1959, Prediction of sunspot numbers until the end of the present cycle. Nature 184, 51-52.

HILL, J. R., 1977, Long term solar activity forecasting using high-resolution time spectral analysis. Nature 266, 151-153.

HOWARD, R., 1977, Large-scale solar magnetic fields. Ann. Rev. Astron. Astrophys. 15, 153-174.

HUNTER, H. E., 1979, Application of an eigenvector technique to prediction of sunspot activity. Solar-Terr. Predictions Proc., Boulder, Vol. 3., A22-A36.

JOSE, P. D., 1965, Sun's motion and sunspots. Astron. J. 70, 193-200, and HENKEL 1971 (q.v.).

KALINKOV, M., 1968, A possible method of predicting solar activity. C. R. Acad. Bulg. Sci. 21, 311-314.

KANE, R. P., 1978, Predicted intensity of the solar maximum. Nature 274, 139-140.

KING-HELE, D. G., 1966, Prediction of the dates and intensities of the next two sunspot maxima. Nature 209, 285-286.

LEIGHTON, R. B., 1969, A magneto-kinematic model of the solar cycle. Astrophys. J. 156, 1-26.

McNISH, A. G. & LINCOLN, J. V., 1949, Prediction of sunspot numbers. Trans. Amer. Geophys. Union 30, 673-685.

MINNIS, C. M., 1960, An estimate of the peak sunspot number in 1968. Nature 186, 462.

MINNIS, C. M., 1961, An estimate of the peak sunspot number in 1968. J. Atmos. Terr. Phys. 20, 94-99.

OHL, A. I., 1966, Wolf's number prediction for the maximum of the cycle 20 (in Russian). Soln. Dann. 12, 84.

OHL, A. I., 1971, Physics of the 11-year variation of magnetic disturbances. Geomag. Aeron. 11, 549-551.

OHL, A. I., 1976, A preliminary forecast of some parameters of cycle no. 21 of the solar activity (in Russian). Soln. Dann. No. 9, 73-75.

OHL, A. I. & OHL, G. I., 1979, A new method of very long-term prediction of solar activity. Solar-Terr. Predictions Proc., Boulder, Vol. 2, 258-263.

OKAL, E & ANDERSON, D. L., 1975, On the planetary theory of sunspots. Nature 253, 511-513.

PEKING OBSERVATORY, 1979, Solar activity predictions at Peking Observatory. Solar-Terr. Predictions Proc., Boulder, Vol. 1, 154-162.

RAMASWAMY, G., 1977, Sunspot cycles and solar activity forecasting. Nature 265, 713-715.

REDDY, B. M., AGGARWAL, S., LAKSHMI, D. R., SHASTRI, S. & MITRA, A. P., 1979, Long term solar activity and ionospheric prediction services rendered by the National Physical Laboratory, New Delhi. Solar-Terr. Predictions Proc., Boulder, Vol. 1, 118-133.

SAKURAI, K., 1977, Equatorial solar rotation and its relation to climatic changes. Nature 269, 401-402.

SARGENT, H. H., 1977, A prediction of the next sunspot maximum. Trans. Amer. Geophys. Union. 58, 1220.

SARGENT, H. H., 1978, A prediction for the next sunspot cycle. Proc. 28th IEEE Vehicular Technology Conf. (Denver, Colo.), 490-496.

SCHATTEN, K. H., SHERRER, P. H., SVALGAARD, L. & WILCOX, J. M., 1978, Using dynamo theory to predict the sunspot number during solar cycle 21. Geophys. Res. Letters 5, 411-414.

SCHOVE, D. J., 1955, The sunspot cycle, 649 B.C. to A.D. 2000. J. Geophys. Res. 60, 127 -146.

SIMON, P. A., 1979, Polar coronal holes and solar cycles. Solar Phys. 63, 399-410.

SMITH, F. M., 1976, Some new insights into the mechanism of the sunspot cycle. Radio Comm. 52, 494-500.

SUDA, T., 1979, An attempt at prediction of solar-activity indices which are closely related to climate. Solar-Terr. Predictions Proc., Boulder, Vol. 3., A96-A110.

VALTCHUK, T. E. & FELDSTEIN, Y. I., 1978, Soln. Dann. No. 12.

VASILYEV, O. B. & KANDAUROVA, K. A., 1971, Soln. Dann. No. 11, 109.

VASILYEV, O. B. & KANDAUROVA, K. A., 1979, A statistical model of solar periodicity and its prediction. Solar-Terr. Predictions Proc., Boulder, Vol. 3, A10-A21.
VASILYEV, O. B., VITINSKI, Y. I. & KANDAUROVA, K. A., 1975, Forecast for solar cycle 21. Soln. Dann. No. 10, 55-59.
VITINSKI, Y. I., 1962, Solar-Activity Forecasting. Acad. Nauk. SSSR, Leningrad.
WALDMEIER, M., 1939, Astr. Mitt. Zurich 14, 439 and 470.
WALDMEIER, M., 1964, in Xanthakis & Poulakos 1978 (q.v.).
WOOD, C. A., 1979, Solar activity prediction for cycle 21. Solar-Terr. Predictions Proc., Boulder, Vol. 3, A37-A40.
WOOD, K. D., 1972, Sunspots and planets. Nature 240, 91-93.
WOOD, R. M. & WOOD, K. D., 1965, Solar motion and sunspot comparison. Nature 208, 129-131.
XANTHAKIS, J. N., 1967, Probable values of the time of rise for the forthcoming sunspot cycles. Nature 215, 1046-1048.
XANTHAKIS, J. N. & POULAKOS, C., 1978, A forecast of solar activity for the 21st solar cycle. Solar Phys. 56, 467-469.
XANTHAKIS, J. N. & POULAKOS, C., 1979, Prediction of the radio emission indices of the sun in the frequency range $1000 \leqslant f \leqslant 3750$ MHz. Solar-Terr. Predictions Proc., Boulder, Vol. 3, A77-A95.
ZHEN-TAO, X., AI-DI, Z., YAN-LIN, M. & QUAN-SHI, G., 1979, Long-term forecasting of solar activity. Solar-Terr. Predictions Proc., Boulder, Vol. 1, 163-167.

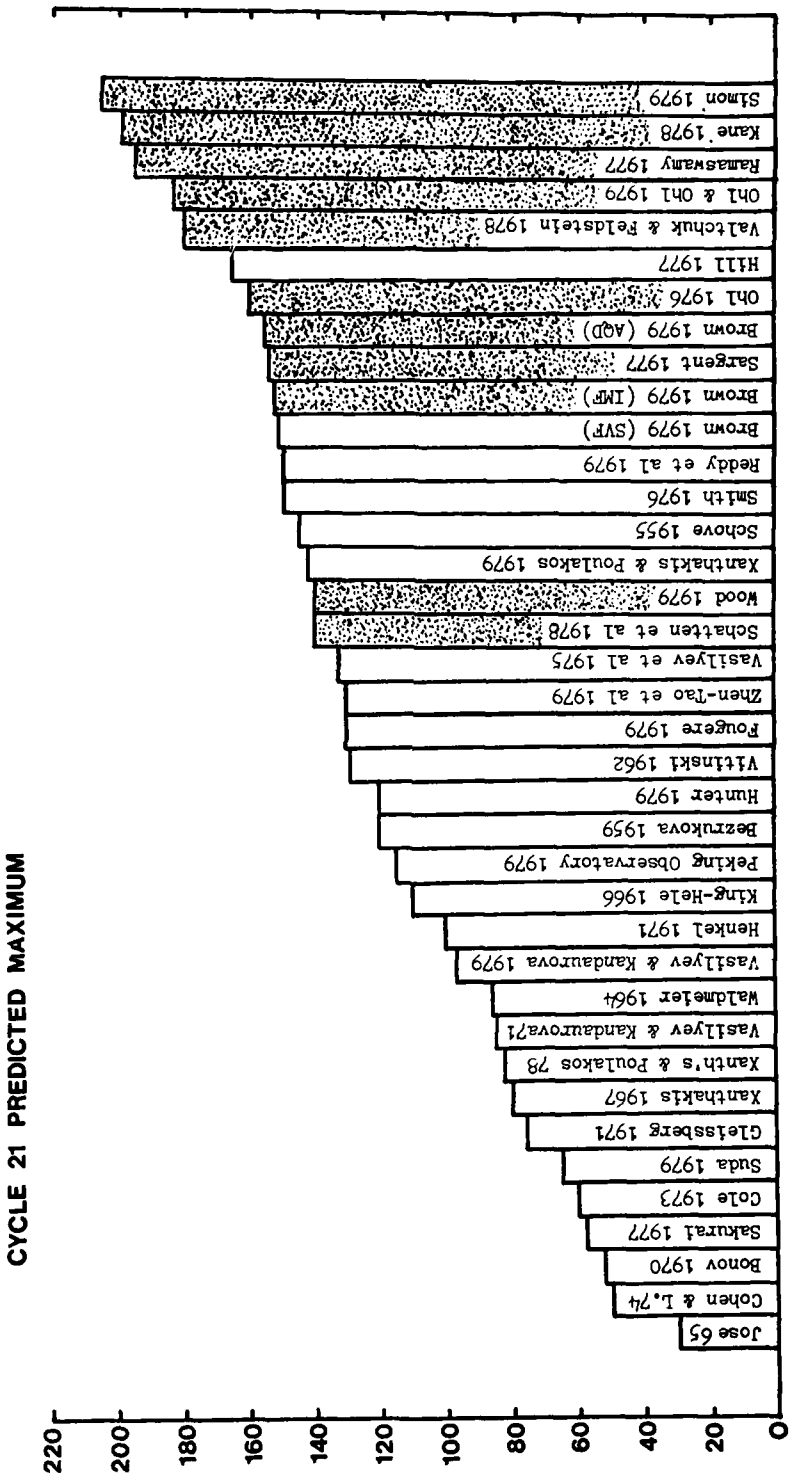


FIGURE 1. Thirty-eight published predictions for the magnitude of peak mean sunspot number for cycle number 21. Stippled predictions employ methods based on precursors near sunspot minimum or during the preceding cycle.

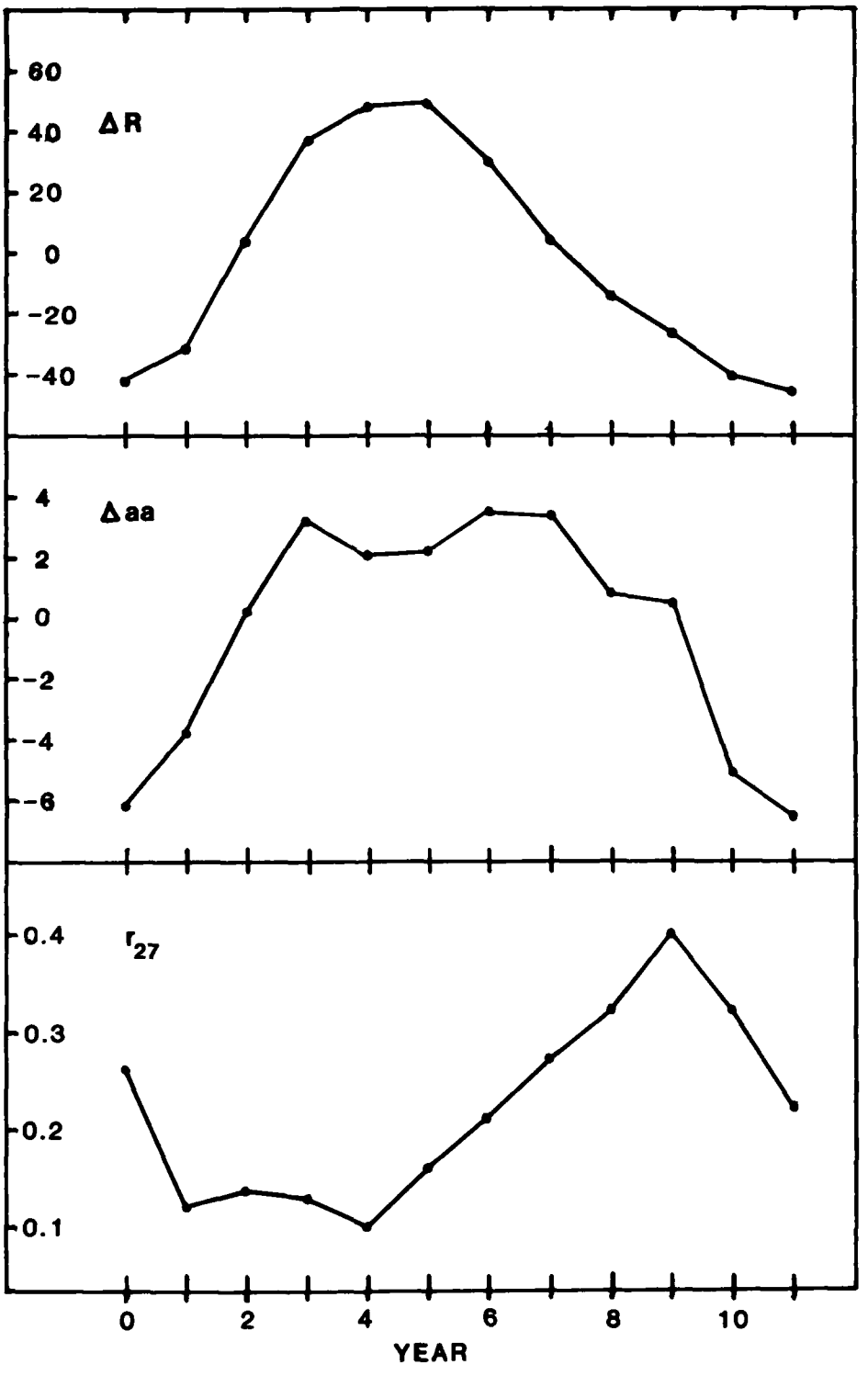


FIGURE 2. Average solar cycle variations for cycles 12-19 (1879-1965) of sunspot number R (upper panel), magnetic activity aa (centre panel), 27-day lag auto-correlation of aa (lower panel). In the two upper curves trends in R and aa have been removed by taking departures from a smoothed curve.

IONOSPHERIC PREDICTIONS FOR HF RADIO SYSTEMS : THE FUTURE

P. A. Bradley and M. Lockwood

S.R.C., Rutherford and Appleton Laboratories, Ditton Park, Slough SL3 9JX, U.K.

SUMMARY

Current prediction procedures are reviewed briefly and the uses for which they are needed are considered. Examples are presented of requirements for long-term, short-term and storm propagation predictions. Desirable and likely future improvements in knowledge of ionospheric morphology leading to more accurate predictions are discussed.

Specific suggestions are made for extended prediction procedures to aid optimisation of communication systems. These involve estimates of additional parameters and improved representations of other factors. In particular, predictions of background interference, signal dispersion and error rates are proposed.

1. INTRODUCTION

Since the earliest days of radio, the objective of developing accurate predictions of received HF sky-wave signal characteristics has been pursued vigorously in many countries. Prediction procedures were devised to give estimates of median values of the maximum usable frequency (MUF), received signal strength, background noise and lowest usable frequency (LUF), and to indicate their diurnal, seasonal and solar-cycle variations. The techniques adopted have usually involved the following stages : (i) determination of a representative model of the electron concentration over the propagation path, taken as being along the great circle between transmitter and receiver, (ii) some kind of ray assessment leading to an estimate of the modes present, (iii) calculation of the received signal intensity in terms of the various separate transmission-loss factors judged to be significant, (iv) estimation of the intensities of atmospheric noise and background man-made noise arising from unintended emissions, and (v) choice of some reference required signal/noise ratio to yield an acceptable grade of service.

Despite the significant amount of calculation involved, in many cases prediction accuracy has been poor. This has led to the pursuit of refinements and recourse to computer evaluation techniques. So there has been a shift in the last 20 years to the development of a prediction capability at only a few centres. Prediction services have been established in a number of countries (STPP, 1979). These meet most requirements more effectively than if users were to undertake their own calculations. At the same time, efforts have been in progress under the auspices of the International Radio Consultative Committee (CCIR) of the International Telecommunication Union (ITU) to develop agreed procedures for international use (CCIR, 1970; 1978a,b). An example of the computer printout from the first CCIR interim prediction procedure is given in Fig. 1.

In recent years there has been a growing awareness that predictions are still often considerably in error, that methods of prediction are too complicated for the accuracy they achieve, that requirements exist for simplified predictions where some loss of accuracy can be tolerated in the interests of speed of evaluation, that there is a need also for short-term predictions for periods of from hours to a day or more ahead and that estimates of additional parameters would be of value. This paper addresses itself to some of these topics, paying particular attention to operational factors and to areas where improvements in prediction capability are possible in the foreseeable future. Section 2 reviews current prediction requirements and in Section 3 developments related to a better understanding of ionospheric morphology are considered. Examples of possible extended predictions and predictions which have been improved in other respects are examined in Section 4.

2. CURRENT PREDICTION REQUIREMENTS

It is generally assumed by radio propagation experts that the need exists for long-term predictions in support of system design, based on a median or statistical representation of the ionosphere, and for short-term predictions taking into account likely ionospheric changes over periods of up to about a day for use in the frequency management aspects of system operation. Besides these two types of predictions for reliability considerations, there is also a need for compatability predictions to assess possible interference between individual circuits affecting channel sharing, and for the development of service-planning procedures based on optimum channel usage. These cases are considered separately below, with contrasting factors highlighted in Table I.

2.1 System Design

There is no major restriction on the permissible amount of calculation or the speed with which the results are needed. Accuracy is the prime consideration and most radio users are willing either to apply accepted procedures or to have these carried out for them by some agency. Computing costs are low compared with installation costs. In principle, calculations for a range of conditions representative of the different hours, seasons and solar epochs that operation is planned serve to identify required transmitter size, number of frequencies for which assignments are wanted and antenna design details. It is difficult to obtain statistics on how much use is made in practice of predictions for system design and how much reliance is placed on past experience. It seems probable though, where comparable services are already in operation under conditions similar to those planned, that their performance data may be extrapolated with adequate results. Under other conditions and where a contractor needs to convince his customer or sponsor that proposals being made are valid, recourse to predictions is essential.

Some circuit-design parameters, such as maximum permitted transmitter power, channel centre frequencies, types of modulation and bandwidth are fixed under the ITU Radio Regulations (ITU, 1976) and by Conventions of other international agencies. For example, HF radiotelephone links of the aeronautical

mobile service are governed by Annex 10 to the Convention of the International Civil Aviation Organisation (ICAO, 1972). Other parameters, notably the choice of antennas and the approximate frequencies for which assignments are wanted, must be determined by the operator. Cases where predictions are potentially of particular value are those where the transmitter location can be varied for optimisation and where networks of transmitters are required to provide area coverage, say for broadcasting or distress protection. Long-term predictions similarly serve an important role in the siting and planning of surveillance systems such as HF over-the-horizon radars and passive monitoring stations.

2.2 Frequency Management

Frequency management may be defined as the selection of the frequency to use on a particular occasion from those assigned and available. The circuit operator's ideal is to have a large number of frequencies assigned and to be able to choose the best; alternatively to employ frequency diversity. However, in the interests of efficient spectrum utilisation the objectives must be to keep the number of frequency assignments and simultaneous transmissions to a minimum, although it is to be noted that multiple-frequency transmissions are used in some applications. Given a realistic set of assigned frequencies, frequency management in principle could be aided by some form of short-term prediction procedure. Possible approaches to the development of such procedures are discussed in Section 3. It is evident that any short-term method adopted needs to be capable of rapid evaluation and requires on-line data links to a main-frame computer, or local use of a minicomputer or pocket calculator. Such an approach must be seen in perspective in comparison with alternative techniques and in the light of existing operating practices which differ appreciably for the separate radio services.

It is instructive to consider specific examples. Major use is made of HF for sound broadcasting and there are nine general frequency bands within which this is permitted together with two additional bands for use in the Tropical Zone (see Table II). At present broadcasting schedules, consisting of frequency assignments to each operating agency, are determined four times a year. Broadcasters need to publish their frequencies well in advance and listeners prefer them to be fixed. When preparing a schedule, frequencies are bartered among agencies, but there are operational and administrative advantages in retaining as far as possible the same frequencies for successive schedules, even if these are then used at different times of day or for different coverage areas. Transmissions are planned to optimise site resources, so that a choice of frequency for one coverage area can influence what frequencies are used for another. Simultaneous three-frequency band working is not uncommon, particularly where the same programme is being radiated to different coverage areas. In principle, day-to-day changes in propagation conditions can be overcome by the listener himself selecting the best frequency received. The position may change following the ITU World Administrative Radio Conference (WARC) on HF Broadcasting with a first session in 1983, which has as an objective the establishment of a frequency-assignment plan, but this seems unlikely. Short-term predictions appear to have little application, though storm warnings could be useful in advising listeners how to retune.

For point-to-point communications in the fixed service, duplex working is commonplace. Commercial applications require running costs to be minimised and so frequency diversity is avoided as far as possible. Where 24-hour operation is required, three different frequencies are normally used : one for daytime, one for night and one for the dawn/dusk transitional periods. A convenient and effective frequency-management procedure is illustrated in Fig. 2. With frequencies already assigned, the times when these ought to be most effective are determined on the basis of long-term predictions of the MUF and LUF. Then a schedule is prepared involving dual-frequency transmissions over periods of 1-2 hours centred on the times when the predictions indicate that a frequency change is needed, thereby providing means of allowing for the day-to-day fluctuations in propagation conditions. The signals at the new frequency are monitored at the receiver, and when their quality is judged to be better, that channel is used and a short coded message sent to cease transmissions of the original frequency. The approach is a form of real-time channel sounding aided by long-term predictions and again it seems to preclude the need for short-term predictions on a regular basis. However, storm predictions could be of particular value when conditions differ appreciably from those normally experienced.

The maritime mobile service at HF provides long-range radiotelegraph, radiotelex, radiotelephone and radio facsimile links between ships and shore stations. Slightly different frequencies are used for each of these within the bands indicated in Table II, but in every case the shore stations operate in a broadcast role, transmitting the call signs simultaneously on defined groups of frequencies in the 4, 6, 8, 12, 17 and 22 MHz bands of all ships for which there is traffic. The ships then answer on a prespecified paired frequency to that for which reception is found to be best and the messages are passed. Since the requirement is to serve ships over a wide area, the use of so many frequencies by the shore station is justifiable. The system works well and there is no need for prediction aids. A somewhat similar situation applies for the aeronautical mobile (route) service mainly using HF for long-distance radio telephone links between commercial aircraft and air-traffic control centres. Ground stations transmit using groups of frequencies and aircraft respond in a time-sharing mode on the one of these same frequencies found to be best. The channel noise background can be monitored in the aircraft and the operator has the advantage in his frequency selection of also being able to assess how well communication is achieved between ground and other aircraft in his vicinity. Again then the arrangement provides its own real-time channel sounding.

From the foregoing it may be concluded that there is no case for routine short-term predictions in support of civilian communications systems, but storm predictions would be of particular value to the sound broadcasting and point-to-point services. There are, however, other examples of where routine short-term predictions would be useful, such as in the frequency management of over-the-horizon radars. A further situation is that of an off-route aircraft with a small crew making short flights where the objective is for limited but reliable communications to one of a number of ground stations which are not necessarily transmitting signals. Other military applications for short-term predictions can be envisaged, although the possibility of ionospheric modification affecting their accuracy cannot be overlooked entirely.

2.3 Channel Sharing

The large numbers of separate circuits in each of the HF service bands mean that channel sharing is essential. Sharing agreements are concluded at the time of frequency assignment, in principle having regard for the notified hours of operation of the circuits, but in practice relying almost entirely on geographical separation for interference avoidance. Without actually setting up the circuits, the likelihood of interference must be determined using a long-term prediction procedure. All frequency assignments are recorded in a Master Register (ITU, 1979) collated and maintained by the secretariat of the International Frequency Registration Board (IFRB) of the ITU. Each proposed new assignment is given a favourable or unfavourable notification depending on whether or not it is judged to lead to unacceptable interference to existing co-channel and adjacent-channel circuits. The IFRB secretariat sometimes receives several thousand HF proposals within a week and so their requirements are for a modest prediction procedure that can be evaluated rapidly on a main-frame computer; some loss of accuracy compared with the requirements for system design has to be tolerated in the interests of expediency.

If an unfavourable notification is received the Administration responsible for the new circuit has the option of making a revised submission or trying to secure the agreement of its counterpart responsible for the circuit potentially interfered with that sharing is acceptable. This agreement may be secured following a period of test transmissions or as part of package proposals. When a new service plan is being introduced, as during a WARC, there is also a need for an agreed simple prediction procedure of limited accuracy that can be implemented using a pocket calculator. This provides an aid to negotiations aimed at formulating proposals that can be given a more rigorous evaluation by the IFRB Secretariat.

2.4 Service Planning

To date, relatively little effort has been applied on a worldwide basis to the optimisation of the different radio services. Most of these have grown in a haphazard fashion and their spectrum utilisation is far from efficient. The ideal arrangements of optimum modulation techniques, channel bandwidths and spacings are understood, but the inertia to change is overwhelming. Arrangements which assume continuous use of a frequency, when this is employed for only a fraction of the time, are a luxury that ultimately cannot be accepted, yet the administrative machinery to supervise any alternative procedure seems insurmountable. As social and economic patterns change, requirements for different types of radio systems vary. At the present, with the greater availability of satellite and cable links, there is a reduction in the number of HF point-to-point circuits, particularly for long-haul routes, but a growing demand for new broadcasting stations. The 1980 WARC has re-allocated certain frequency bands to the sound broadcasting service to meet this trend, but not without opposition from some quarters. It is almost certain that no investigations have been carried out on the real spectrum requirements of the different services. Furthermore, frequency sharing between the services is a useful means of optimising spectrum utilisation but any changes to current practice need very careful review before being introduced. There is plenty of scope for further studies based on long-term predictions to determine the ideal service-planning strategies.

3. IMPROVEMENTS IN IONOSPHERIC KNOWLEDGE

3.1 Introduction

The accuracy of existing HF propagation predictions is governed primarily by the ability to model the ionosphere. Current models have various major deficiencies that could be overcome with improved ionospheric knowledge; the nature of these deficiencies depends on latitude and their importance is a function of the application of the prediction. There are other limitations which can be directly attributed to the modelling approach, which for most parameters is in terms of a small number of factors.

A key variable in any model must be some form of activity index. The most commonly used index is smoothed sunspot number, but other indices exist and more could be developed to quantify ionospheric morphological features : all are measures of some observable effect which either shares a common cause with the ionospheric feature or is a link in the solar-terrestrial interaction on which it depends. The usefulness of any given index to propagation predictions is a function of :

(i) the degree of correlation with the ionospheric feature;

(ii) the sensitivity of the propagation to the ionospheric feature;

(iii) the ease of observation of the index, and

(iv) the accuracy of prediction of the index.

The relevance of the index is determined by (i) and (ii), but practical constraints are set by (iii) and (iv), depending on how far into the future the prediction is needed.

Curent prediction procedures are of greatest accuracy at middle and low latitudes where the configuration of the geomagnetic field leads to partial shielding from magnetospheric effects. Magnetospheric interactions produce variable and complex ionospheric features at higher latitudes outside the plasmasphere, where therefore median long-term predictions using an activity index based only on sunspot number are of limited accuracy. Correlations between certain high-latitude ionospheric features and other indices are becoming established, but it is not yet possible to determine which indices will prove most useful, and over what time periods. As a consequence, high-latitude features are discussed here in general terms without separation into long and short-term time scales as in the lower-latitude treatment. The difference between a long-term propagation prediction and a short-term propagation prediction depends essentially on the form of ionospheric modelling adopted.

3.2 Modelling the Ionosphere at High Latitudes

Consider first the case at latitudes within the plasmasphere (less than about 55° geomagnetic

latitude) where there is thermal plasma exchange between the ionospheric F-region and the magnetosphere. Both co-rotate with the Earth, and so this exchange exhibits a systematic diurnal cycle, as discussed for example by Murphy et al. (1976). The interaction is subject to modification by storm-time depletion of the protonosphere, especially at middle latitudes, where complete recovery may be achieved only rarely. Other factors influencing the ionosphere at low and middle latitudes likewise have a regular diurnal variation except when storm perturbed : these include production rate (because of the dominance of photo-ionisation), loss rate (because thermospheric winds and ionospheric composition are controlled mainly by solar EUV heating) and electric fields (because of the shielding from magnetospheric fields). Hence the concept of characterising a median quiet ionosphere for a given location in terms of solar-zenith angle and sunspot number, regarded as an index quantifying the incident solar-energy flux, has a good physical basis.

By contrast, poleward of the plasmapause the F-region no longer co-rotates with the Earth, but moves under the influence of the so-called magnetospheric-convection electric field as given by a model of Heppner (1977), reproduced in Fig. 3. Evidence is growing that this motion is a very important factor in the formation of ionospheric features like the main trough (Spiro et al., 1978) and the high-latitude hole (Brinton et al., 1978). The thermospheric winds and their effects are also more complex at high latitudes because of the plasma convection and energy deposition during auroral activity.

Although maps of high-latitude ionospheric features are best expressed in a local time-invariant latitude coordinate system, it should be appreciated from the above remarks that local time does not have the same significance as at lower latitudes. Geographical smoothing between measured F-region parameters is most accurate when performed along the lines of flow. The convection pattern varies with the direction of the interplanetary magnetic field (IMF) and with the Kp index, both of which can be predicted accurately only in the short term. But this does not necessarily mean that long-term median predictions of propagation conditions at high latitudes would not benefit from inclusion of an allowance for the influence of these factors. For example, accepting that the convection pattern depends critically on whether there is a southward or northward component of the IMF, it may prove valuable to attempt to produce ionospheric models for each case separately. These give more information on extreme propagation conditions than is obtained using a combined model as at present. Satellite observations of the IMF are used as input data to an existing short-term ionospheric prediction procedure developed by Zavakina and Lavrova (1979).

At the cleft, particles with energies less than 1 keV have access to the ionosphere without prior magnetospheric acceleration. These produce enhanced ionisation in both the topside and bottomside F-region, depending on the convection pattern (e.g. Foster et al., 1980). Plasma which passes under the cleft is moved rapidly anti-sunward, and hence the regions for which it can act as an ionisation source are determined by the flow pattern. In addition, there have been suggestions of a throat restricting flow into the polar cap (Spiro et al., 1978). This would cause the flow under the cleft to be faster, and hence reduce the magnitude of the enhancement whilst increasing its spatial extent. It follows therefore that the relationship between the cleft and the dayside convection pattern is of great importance to predictions. The cleft is known to move equatorward during southward turnings of the IMF, and this motion relative to the convection pattern may explain why the plasma concentration poleward of the cleft is subsequently sometimes decreased (Foster et al., 1980). Other suggested possible mechanisms which may be important, like enhancements of loss rates due to the electric field (Schunk et al., 1975), need to be better understood before they are likely to be used effectively to introduce improvements in existing models of the polar-cap F-region.

The extreme temporal and spatial variability of the high-latitude ionosphere leads to observational data requirements that are not currently met by existing sounders. This situation should improve a little with the introduction of new sounding stations at Spitsbergen, Cape Parry and Sondrestrømfjord. However, there are other problems in the interpretation of conventional ionograms due to off-vertical reflections in the presence of layer tilts, such as those which arise at the edges of the trough, auroral oval, hole and cleft. Examples by Nygren (1977) of ray-tracing simulations for a tilted ionosphere are shown in Fig. 4. To interpret ionograms unambiguously under these conditions requires either successive observations with the ionosonde mounted on a moving platform such as an aircraft, or that a direction-finding capability be introduced. Other useful experimental data likely to lead to an improved understanding of the high-latitude ionosphere are becoming available from specialised facilities such as the advanced vertical sounder and the incoherent-scatter radar.

Ionospheric models have been developed recently incorporating high-latitude characteristic features (e.g. Vondrak et al., 1977); Besprozvannaya et al., 1979). The question arises whether inclusion of a specific feature leads to better propagation predictions than with heavily-averaged maps in which the feature becomes obscured. Consider for example Fig. 5 derived by Lockwood (1980) from experimental observations on an HF link along the length of a mid-latitude trough. This gives estimates of the difference ΔG in received field strength from that to be expected in the absence of ionospheric curvature. Defocusing is seen to arise ($\Delta G < 0$ dB), particularly for the higher levels of magnetic activity. Comparisons have been made with ray-tracing results based on ionospheric models including a trough at different positions relative to the path. These confirm that the trough is responsible for the defocusing. The observed larger scatter at the lower Kp can only be explained in terms of the fluctuations in trough parameters. Mendillo et al. (1978) have considered the modelling of the trough in terms of Kp and shown that there is considerable scatter about the best-fit regression lines. This serves to illustrate that sometimes incorporation of correct, yet incompletely specified relationships, in this case into a short-term propagation prediction procedure, can do more harm than good.

The growth in knowledge of the morphology of particle precipitation in different energy ranges should help the development of improved models of the high-latitude D, E and Es regions. D-region data are needed particularly for the modelling of auroral absorption. Two auroral absorption models derived mainly from riometer data have been proposed by Foppiano. One of these (Foppiano, 1975) averages over all Kp values and is for use in long-term predictions; the other method (Foppiano, 1977), together with an extension of the first method by Vondrak et al., (1977) and a method produced in the USSR are dependent upon Kp, and so are intended for short-term prediction use. All these procedures are based on a statistical concept of an oval within which auroral activity occurs. At any one time the precipitation is over a narrower band of latitudes which will not always lie within the statistical oval (Kamide and Winningham, 1977). It seems

possible that overestimation of the absorption on trans-auroral paths may arise from the use of the statistical auroral oval concept; ideally short-term predictions should take account of the position of the instantaneous aurora. Similarly, the morphology of the auroral E-region enhancement closely resembles that of the particle precipitation and is likely to be modelled with too wide a latitudinal extent.

At present, auroral substorms are largely unpredictable, other than in a general warning of the probability of substorm occurrence during the incidence of a geomagnetic storm. Models of substorm dynamics in terms of the poleward motion after onset offer promise of improved warnings for the higher latitudes.

3.3 Modelling the Ionosphere at Lower Latitudes

At plasmaspheric latitudes it is meaningful to develop predictions for use on a range of different timescales. Although the mechanisms responsible for day-to-day ionospheric variations at these latitudes are largely understood, accurate representations cannot yet be formulated. The extremes in the variation may be attributed usually to geomagnetic storms. No complete model encompassing the diversity of ionospheric storm responses has been produced; such models as exist can be misleading if implemented injudiciously.

3.3.1 Long-term Predictions

The most widely used models of F-layer characteristics are those presented by the CCIR (1978c). They include maps of foF2 based on data collected during only three separate years and are now over a decade old. As such, they may well not be representative of all epochs, particularly since no account is taken of any changes in different solar cycles, or of the effects of the secular variation of the Earth's magnetic field. The models are derived from vertical soundings. Satellite-probe measurements and results from thermospheric wind theory have not been considered in the geographical smoothing adopted, despite inconsistencies which these sources of information reveal. Nevertheless the enormity of the task of producing new maps from an extended data base should not be underestimated; ultimately the justification for an updated model must arise from a requirement for propagation predictions of increased accuracy. In the meantime, existing map limitations should always be borne in mind when contemplating other changes in propagation prediction techniques, which may not then be found worthwhile.

Additional parameters are mapped in some more recent models. For example, Anufneva et al., (1979) give the maximum gradient of the vertical plasma concentration together with the concentration at the height where it occurs. These parameters are use in calculations of ducted propagation modes. There are other models such as the International Reference Ionosphere and that of Tacione et al., (1979) giving the complete height distribution of electron concentration. It remains to be established if these are sufficiently representative to be of significant value to general propagation predictions, or whether they include too detailed a fine structure to be preferable to the simpler idealised models consisting of parabolic and linear segments.

3.3.2 Short-term Routine Predictions

There are various approaches presently being considered. Rush et al., (1974) and King and Slater (1973) have concluded that mid-latitude F-mode propagation variability is caused mainly by changes in foF2 rather than in hmF2. Therefore, greatest improvements from a long-term prediction could be achieved by means of updated foF2 estimates. One proposal is to use measured data, such as yesterday's value or the avera of the last several days. A case can be made for taking figures gathered earlier the same day from a location to the east where the sun has risen earlier. However, the density of required sounding stations (Rush, 1976) tends to preclude this scheme in most areas of the world.

Other ideas being pursued, which hold promise, involve establishing correlations between foF2 and activity indices which can be determined in the short term. These indices may be solar, magnetic, ionospheric or a combination of the three.

3.3.3 Short-term Storm Predictions

The deposition of energy at auroral latitudes arising from particle precipitation and Joule heating by dissipative electric currents, leads to the production of thermospheric wind followed by composition changes which perturb the mid-latitude ionosphere (Volland, 1979). The energy deposited from the magnetosphere depends not only on the size of the perturbation of the solar wind, but also on the reservoir content of the plasma sheet. Electric fields introduce a local-time variation into the ionospheric response to storms. The enhanced convection electric field forces plasma down into the nightside ionosphere, but on the dayside plasma is extruded from the plasmasphere. Hence on the nightside there are initial downward fluxes followed by upward replenishment fluxes; on the dayside a.' fluxes are upward. The severity and development of mid-latitude storms depends on the time of storm onset.

A fuller understanding of all these mechanisms is required before an effective short-term prediction capability can be developed. Any adopted procedure is likely to require continuous monitoring of the interplanetary, magnetospheric and ionospheric media.

4. REQUIREMENTS FOR IMPROVED AND EXTENDED PREDICTIONS

The techniques of signal and noise estimation have been mentioned in the introduction and likely improvements of relevant ionospheric knowledge in the foreseeable future have been considered in Section 3. Here the application of these improvements and of other data to existing procedures is examined and the prediction of further system-performance parameters is discussed.

Present methods of raypath assessment give only a limited treatment of mixed modes involving successive reflections from different ionospheric layers, and the effects of layer tilts cannot be included meaningfully with the scale size of current mapping representations (Bradley, 1979). It seems possible,

however, that the regular layer tilts associated with both high- and low-latitude F-region features could be mapped separately and introduced into the predictions. At the same time it would then be possible to allow for off-great-circle propagation in the presence of ionisation gradients. An empirical procedure based on experimental data to predict signal amplitudes for off-great-circle paths involving ground side-scatter has been developed by Kanaya and Wakai (1976), but this has not received the attention it deserves : it should be applied and fully tested.

Another important factor affecting raypaths is the presence of spread-F irregularities. These can occur at a wide range of heights and so undoubtedly both cause scattering losses and provide a reflection mechanism. Important advances have been made in recent years in the modelling of sporadic-E reflection and obscuration losses (Sinno et al., 1976; CCIR, 1978d); it seems possible that a similar approach could be applied to modelling spread-F effects. This is particularly needed before there can be any accurate transequatorial propagation prediction procedure.

In recent years there has been a growing awareness that it is not meaningful to assume in a long-term prediction that propagation ceases abruptly at a limiting frequency given as the maximum frequency that can be reflected from the monthly median model ionosphere. So the term availability (F.DAYS) has been defined as the probability of mode support, allowing for day-to-day ionospheric variability. This can be estimated for all frequencies of interest in terms of the monthly median MUF and its decile range (Bradley and Bedford, 1976). At frequencies above the monthly median MUF an extra loss term is incorporated into the calculations; this becomes larger the greater the frequency (Wheeler, 1966). The rationale is claimed to be that with a spatially inhomogenous ionosphere there is no unique MUF, but a reducing number of patches of ionisation capable of supporting reflection as the frequency is increased.

More work is needed in developing techniques for estimating the combined effect when different propagation modes are present. If a 1F2 mode has a median field strength of 42 dB > 1 μV/m and an availability of 73%, and the 2F2 mode has a median field strength of 46 dB > 1 μV/m and an availability of 27%, what is their resultant? - The answer requires more information than is currently available on ionospheric spatial variability, but it is by no means certain on the basis of existing knowledge that present techniques cover this case as well as they might. Useful information on the accuracy of signal prediction procedures comes from comparisons with measured data. The CCIR has assembled a Data Bank of measurements for this purpose (CCIR, 1978e), but it remains to establish how to compare a prediction with a measurement when the availabilities in the two cases differ.

Background noise intensities are currently estimated using CCIR models for atmospherics arising from lightning discharges (CCIR, 1963) and for wideband unintended man-made emissions from sources of industrial and domestic activity (CCIR, 1978f). Yet it is known that in many cases the background to satisfactory reception arises from co-channel signals. Although there have been several spectrum-monitoring campaigns aimed at quantifying the interference environment, these have been limited by lack of a coherent analysis strategy for processing the vast amounts of non-stationary data collected. Here it is suggested that a statistical approach may be possible based on the probability of finding clear channels, and that it would be more meaningful to introduce the results of this into estimates of the likely background, rather than giving best-case figures when a clear channel can be found.

The proposal is to record data in a large number of narrowband channels, say of bandwidth 1 kHz, at different times of day and for different locations. Some convenient reference threshold intensity would be selected and by combining results within each frequency band of each service at a given hour on the different days of a season, the fraction of the number of channels sampled with an interfering signal in excess of the threshold would be determined and used as a measure of spectrum occupancy. Grouping of results in accordance with service allocation is important, because for example there is experimental evidence that occupancy is considerably greater in the sound-broadcasting bands than those of other services; this is to be expected in view of the larger transmitter powers involved. Figure 6 is an idealisation of the type of results to be expected. This shows greatest occupancy in the 5-15 MHz range. The frequency of maximum occupancy will vary with time-of-day and season. Reductions in occupancy at lower frequencies can be attributed to the effects of increased ionospheric absorption, and those at the higher frequencies both to an absence of transmitters and to the closer of those that are operative lying within the skip range. It may prove possible to map occupancy at say 10 MHz as a function of geographical position for different times-of-day and seasons, and then to have families of curves showing the frequency variations in a similar fashion to that employed in the modelling of atmospheric noise (CCIR, 1963). Other mean relationships for the dependence of occupancy on threshold and channel bandwidth would also need to be derived.

Assuming the median signal/noise ratio is given and estimates of the day-to-day variabilities of signal and noise intensities are known, it is possible to calculate the probability (F.S/N) of achieving some desired threshold signal/noise ratio when a propagation path exists. With circuit reliability (REL) defined as the probability of achieving the desired threshold signal/noise ratio, it has been usual to evaluate this as

$$REL = (F.DAYS).(F.S/N)$$

on the assumption that the two probabilities are uncorrelated. However, recent work by Thrane and Bradley (1981) based on analyses of measured communication-circuit performance data have confirmed that this approach underestimates the reliability at the higher frequencies. An alternative technique is therefore required to be developed.

With a trend in recent years towards increased introduction of digital communication systems, the requirement now exists for predictions to be extended to include estimates of error rates. Figure 7 shows that these depend not only on signal/noise ratio, but also on channel dispersion. As well as attenuation, both time and frequency dispersion arise during ionospheric propagation. These differ for the separate modes. Time dispersion results from variations in group path and mean values can be estimated readily. On the other hand frequency dispersion is due to ionospheric movements which are not currently modelled. An objective should be to try to map Doppler shifts and Doppler spreads, perhaps in a similar way to the mapping of satellite scintillations. Much work has been undertaken in assessing error rates for different data

speeds, modulation systems and orders of diversity as a function of signal/noise ratio and channel dispersion (e.g. Akima et al., 1969). Results from separate sources need collation, international recognition under the auspices of the CCIR and integration into circuit-performance prediction procedures. There are additional requirements involving a knowledge of fading characteristics and phase coherence across a channel bandwidth; a further objective should be to map also parameters descriptive of these factors.

5. CONCLUSIONS

Despite other possible approaches, such as real-time channel sounding and reliance on past measured data, a need exists for a range of propagation prediction procedures with different degrees of complexity and accuracy. These include long-term, short-term and storm predictions, and are wanted to test for compatability with other circuits and services as well as for reliability considerations. Both accurate and simple procedures are required, adapted to main-frame computer, minicomputer and pocket-calculator applications. The estimation of additional parameters would be of particular value, especially to the growing numbers of radio circuits now employing digital techniques.

Prediction accuracy depends critically on the ability to model the ionosphere. A review of causal features reveals that there are several areas where recent knowledge holds promise for improvements in ionospheric modelling. Understanding of the influence of auroral precipitation and convective motions on ionisation morphology now permits the development of improved long-term models of the high-latitude ionosphere, but care needs to be exercised to ensure that specific features are not misrepresented with the form of averaging adopted. For all latitudes long-term models could be improved using the larger data base of conventional ionospheric-sounding measurements currently available, and also by incorporating results from the range of other techniques that exist.

6. REFERENCES

AKIMA, H., AX, G. G. and BERRY, W. M., 1969, "Required signal-to-noise ratios for HF communication systems", ESSA Tech. Rept. ERL 131-ITS 92, US Govt. Printing Office, Washington, DC 20402.

ANUFNEVA, T. A., GULYAEVA, T. L., KADUKHIR, G. K., 1979, "Prediction of parameters of the maximum of the vertical electron density gradient", S.T.P.P., 4, C-57.

BESPROZVANNAYA, A. S., SHIROCHKOV, A. V. and SHCHUKA, T. I., 1979, "On the approach to forecasting polar ionospheric conditions", S.T.P.P., 2, 528-542.

BRADLEY, P. A. and BEDFORD, C., 1976, "Prediction of circuit availability", Electronics Letters, 12, 32-33.

BRADLEY, P. A., 1979, "Propagation at medium and high frequencies, 2: Long and short-term models" in "Aerospace propagation media modelling and prediction schemes for modern communications, navigation and surveillance systems", AGARD Lecture Series 99, Technical Editing and Reproduction Ltd., London.

BRINTON, H. C., GREBOWSKY, J. M. and BRACE, L. H., 1978, "The high-latitude winter F region at 300 km: thermal plasma observations from AE-C", J. Geophys. Res., 83, 4767-4776.

CCIR, 1963, "World distribution and characteristics of atmospheric radio noise", Report 322, International Telecommunication Union, Geneva.

CCIR, 1970, "CCIR interim method for estimating sky-wave field strength and transmission loss at frequencies between the approximate limits of 2 and 30 MHz", Report 252-2, International Telecommunication Union, Geneva.

CCIR, 1978a, "Sky-wave field strength and transmission loss at frequencies above 1.6 MHz", Decision 6-2, International Telecommunication Union, Geneva.

CCIR, 1978b, "Second CCIR computer-based interim method for estimating sky-wave field strength and transmission loss at frequencies between 2 and 30 MHz", Supplement to Report 252-2, International Telecommunication Union, Geneva.

CCIR, 1978c, "CCIR atlas of ionospheric characteristics", Report 340-3, International Telecommunication Union, Geneva.

CCIR, 1978d, "Method for calculating sporadic-E signal strength", Recommendation 534, International Telecommunication Union, Geneva.

CCIR, 1978e, "Measurement of sky-wave signal intensities at frequencies above 1.6 MHz", Report 253-3, International Telecommunication Union, Geneva.

CCIR, 1978f, "Man-made radio noise", Report 258-3, International Telecommunication Union, Geneva.

FOPPIANO, A. J., 1975, "A new method for predicting the auroral absorption of HF sky waves", CCIR IWP 6/1 Docs. 3 and 10, International Telecommunication Union, Geneva.

FOPPIANO, A. J., 1977, "Prediction of auroral absorption of high and medium frequency waves at oblique incidence", Ph. D. Thesis, University of London.

FOSTER, J. C., STILES, G. S. and DOUPNIK, J. R., 1980, "Radar observations of cleft dynamics", J. Geophys. Res., 85, 3453-3460.

HALCROW, B. W. and NISBET, J. S., 1977, "Model of F2 peak electron densities in the main trough region of the ionosphere", Radio Sci., 12, 815-820.

HEPPNER, J. P., 1977, "Empirical models of high latitude electric fields", J. Geophys. Res., 82, 1115-1127.

HOLZWORTH, R. H. and MENG, C. I., 1975, "Mathematical representation of the auroral oval", Geophys. Res. Lett., 2, 377-380.

ICAO, 1972, "Aeronautical Telecommunications", Annex 10 to the Convention on International Civil Aviation, International Civil Aviation Organisation, Montreal.

ITU, 1976, "Radio Regulations", International Telecommunication Union, Geneva.

ITU, 1979, "International Frequency List", International Telecommunication Union, Geneva.

KANAYA, S. and WAKAI, N., 1976, "Method for estimating the signal strength of ground-scatter modes at frequencies above the classical MUF for transmitter-receiver separations beyond 4000 km", CCIR IWP 6/1 Doc. 39, International Telecommunication Union, Geneva.

KAMIDE, Y. and WINNINGHAM, J. D., 1977, "A statistical study of the 'instantaneous' nightside auroral oval - the equatorial boundary of electron precipitation as observed by the Isis 1 and 2 satellites", J. Geophys. Res., 82, 3573-3588.

KING, J. W. and SLATER, A. J., 1973, "Errors in predicted values of foF2 and hmF2 compared with the observed day-to-day variability", Telecommunication Journal, 40, 766-770.

LOCKWOOD, M., 1980, "The bottomside mid-latitude ionospheric trough", J. Atmos. Terr. Phys., 42, 605-615.

MENDILLO, M., CHACKO, C. C., LYNCH, F. and WILDMAN, P. L. J., 1978, "Attempts to predict trough/plasmapause boundaries in real time", AFGL-TR-78-0080, Air Force Geophysics Labs., Hanscom A.F.B., Mass.

MURPHY, J. A., BAILEY, G. J. and MOFFETT, R. A., 1976, "Calculated daily variations of O^+ and H^+ at mid-latitudes - 1 protonospheric replenishment and F region behaviour at sunspot minimum", J. Atmos. Terr. Phys., 38, 351-364.

NYGREN, T., 1977, "Simulation of vertical incidence ionograms by ray tracing method in the presence of replacement layer and ionospheric trough", J. Atmos. Terr. Phys., 39, 733-739.

RUSH, C. M., 1976, "An ionospheric observation network for use in short-term propagation predictions", Telecommunication Journal, 43, 544-549.

RUSH, C. M., MILLER, D. and GIBBS, J., 1974, "The relative daily variability of foF2 and hmF2 and their implications for HF radio propagation", Radio Sci., 9, 749-757.

SCHUNK, R. W., RAITT, W. J. and BANKS, P. M., 1975, "Effect of electric fields on the daytime high latitude E and F regions", J. Geophys. Res., 80, 3121-3130.

SINNO, K., KAN, M. and HIRUKAWA, Y., 1976, "On the reflection and transmission losses for ionospheric radio wave propagation via sporadic E", J. Radio Res. Labs. Japan, 23 (110), 65-84.

SPIRO, R. W., HEELIS, R. A. and HANSON, W. B., 1978, "Ion convection and the formation of the mid-latitude F region ionization trough", J. Geophys. Res., 83, 4255-4264.

STPP, 1979, Solar-Terrestrial Predictions Proceedings, Ed. R. F. Donnelly, Volume I - containing a series of papers describing different national prediction services, National Oceanic Atmospheric Administration, Boulder, Colorado, US Govt. Printing Office Stock No. 005-117-00479-1.

TACIONE, T. F., FLATTERY, T. W., PATTERSON, V. G., SECAN, J. A. and TAYLOR, J. W., 1979, "Ionospheric modelling at Air Force Global Weather Central", S.T.P.P., 1, 367-377.

THRANE, E. V. and BRADLEY, P. A., 1981, "High frequency sky-wave prediction methods and observational data for high latitude communication circuits" - to be presented at IEE Second International Antennas and Propagation Conference, York.

VOLLAND, H., 1979, "Magnetospheric electric fields and currents and their influence on large scale thermospheric circulation and composition", J. Atmos. Terr. Phys. 41, 853-866.

VONDRAK, R. R., SMITH, G., HATFIELD, V., TSUNODA, R., FRANK, V. and PERREAULT, P., 1977, "Chatanika model of the high latitude ionosphere for application to HF propagation prediction", RADC-TR-78-7, Rome Air Development Centre, Hanscom A.F.B., Mass.

WHEELER, J. L., 1966, "Transmission loss for ionospheric propagation above the standard MUF", Radio Science, 1, (11), 1303-1308.

ZAVAKINA, R. A. and LAVROVA, E. V., 1979, "On the possibility to predict variations in the F2-region parameters as a function of the IMF direction", S.T.P.P., 4, C - 27.

TABLE I Requirements for propagation-prediction procedures

Requirement	Reliability considerations		Compatability considerations	
	System design	Frequency management	Channel sharing	Service planning
Ionospheric model	long-term	short-term	long-term	long-term
Permissible amount of calculation	considerable	considerable	little	considerable
Speed of calculation	routine	fast	fast	routine
Available means of calculation	large computer	large computer minicomputer or pocket calculator	large computer and pocket calculator	large computer
Data transfer to user	routine	near real time	routine	routine
Alternative approach	past experience and current convention	past experience real-time sounding multiple transmission	none	none

TABLE II HF allocated bands for the sound broadcasting, maritime mobile and aeronautical mobile services in Europe and Africa (ITU Region 1) - frequencies in kilohertz.

Sound broadcasting	Maritime mobile	Aeronautical mobile
3200 - 3400T	3155 - 3400	2850 - 3230
3950 - 4000	3500 - 3800	3400 - 3500
4750 - 4995T	4063 - 4650	3800 - 3950
5005 - 5060	6200 - 6525	4438 - 4850
5950 - 6200	8195 - 8815	5430 - 5730
9500 - 9775	12330 - 13200	6525 - 6765
11700 - 11975	16460 - 17360	8815 - 9040
15100 - 15450	22000 - 22720	10005 - 10100
17700 - 17900	25010 - 25600	11175 - 11400
21450 - 21750	26100 - 27500	13200 - 13360
25600 - 26100		15010 - 15100
		17900 - 18030
		21870 - 22000
		23200 - 23350

T - frequency bands allocated for use in the Tropical Zone only (ITU, 1976)

MARCH 0 SUNSPOT NUMBER 125.0

SYDNEY	TO ADELAIDE	AZIMUTHS	MILES	KM.
33.92S - 151.17E	34.92S - 138.58E	260.95 88.09	720.5	1159.4

MINIMUM ANGLE 0.0 DEGREES POWER: 10.00KW REQ.SIG.-110.0 DBW

FREQUENCIES IN MHZ

UT	MUF	2.0	3.0	5.0	7.5	10.0	12.5	15.0	17.5	20.0	25.0	30.0	
04	16.7												
	1F	1E	1E	1E	2F	1F	1F	1E	1F	-	-	-	MODE
	30.9	6.2	6.2	6.3	41.7	22.9	23.0	9.0	30.9	-	-	-	ANGLE
	4.8	3.9	3.9	3.9	5.4	4.4	4.4	4.0	4.8	-	-	-	DELAY
	392	90	90	91	258	264	268	120	392	-	-	-	VIRT HT
	.50	.99	.99	.99	.99	.99	.99	.99	.28	-	--	-	F. DAYS
	133	258	223	182	154	136	134	133	133	-	-	-	LOSS DB
	39	-105	-67	-21	11	31	35	37	39	-	-	-	DBU
	-93	-218	-183	-142	-114	-96	-94	-93	-93	-	-	-	SIG.DBW
	.96	.0	.0	.0	.23	.93	.95	.96	.96	-	-	-	F. SIG
08	16.3												
	1F	1E	1E	1F	1F	1F	1F	1F	1F	-	-	-	MODE
	30.0	6.2	6.3	22.9	21.7	21.7	22.8	25.1	30.0	-	-	-	ANGLE
	4.7	3.9	3.9	4.4	4.3	4.3	4.4	4.5	4.7	-	-	-	DELAY
	380	91	92	246	250	260	277	310	380	-	-	-	VIRT HT
	.50	.99	.99	.99	.99	.99	.98	.79	.20	-	-	-	F. DAYS
	130	135	132	123	125	126	127	129	130	-	-	-	LOSS DB
	42	18	25	38	40	41	42	42	42	-	-	-	DBU
	-90	-95	-92	-83	-85	-86	-87	-89	-90	-	-	-	SIG.DBW
	.95	.91	.94	.98	.97	.97	.96	.95	.95	-	-	-	F. SIG
12	11.6												
	1F	1F	1F	1F	1F	1F	1F	-	-	-	-	-	MODE
	32.9	22.9	22.4	22.7	23.8	26.4	32.9	-	-	-	-	-	ANGLE
	4.9	4.4	4.4	4.4	4.4	4.5	4.9	-	-	-	-	-	DELAY
	428	269	271	278	294	329	428	-	-	-	-	-	VIRT HT
	.50	.99	.99	.99	.99	.89	.24	-	-	-	-	-	F. DAYS
	127	117	119	121	123	125	128	-	-	-	-	-	LOSS DB
	42	36	38	40	41	42	42	-	-	-	-	-	DBU
	-87	-77	-79	-81	-83	-85	-88	-	-	-	-	-	SIG.DBW
	.97	.99	.99	.99	.98	.98	.97	-	-	-	-	-	F. SIG
16	10.1												
	1F	1F	1F	1F	1F	1F	1F	-	-	-	-	-	MODE
	34.0	23.7	23.4	24.0	25.7	31.8	34.0	-	-	-	-	-	ANGLE
	5.0	4.4	4.4	4.4	4.5	4.8	5.0	-	-	-	-	-	DELAY
	446	283	286	296	320	410	446	-	-	-	-	-	VIRT HT
	.50	.99	.99	.99	.97	.52	.11	-	-	-	-	-	F. DAYS
	126	117	119	121	123	126	127	-	-	-	-	-	LOSS DB
	41	36	38	40	42	42	42	-	-	-	-	-	DBU
	-86	-77	-79	-81	-83	-86	-87	-	-	-	-	-	SIG.DBW
	.98	.99	.99	99	.98	.98	.97	-	-	-	-	-	F. SIG

FIGURE 1 Computer prediction of HF signal strength by the first CCIR interim procedure

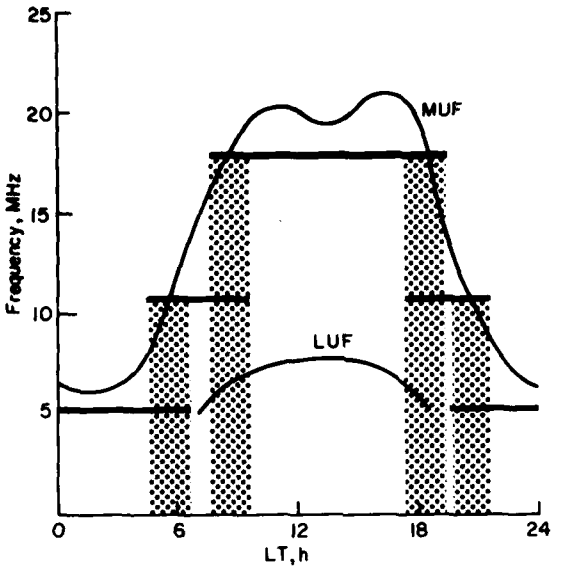


FIGURE 2 Use of propagation predictions in point-to-point frequency management

assigned frequencies

periods of dual-frequency transmission

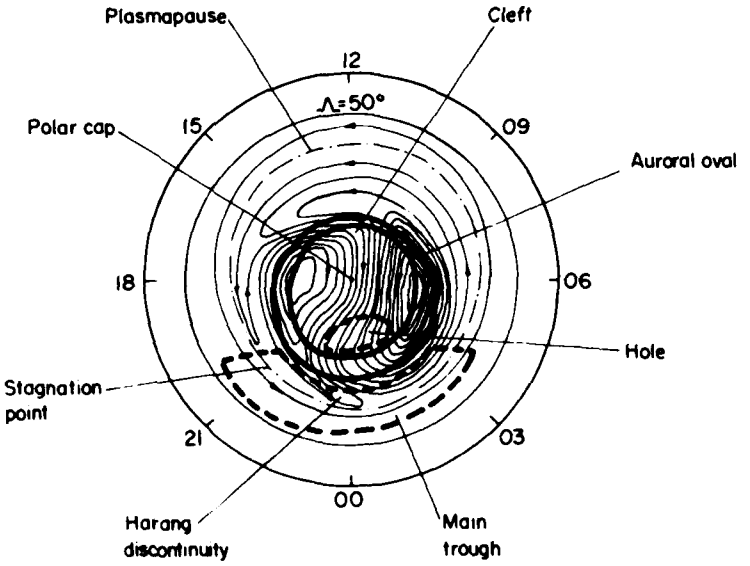


FIGURE 3 Polar convection model for the interplanetary magnetic field directed away from the sun and Kp ~ 3 (from Heppner, 1977). The positions of the auroral oval and the mid-latitude trough for equivalent conditions are also shown from Holzworth and Meng (1975) and Halcrow and Nisbet (1977), respectively.

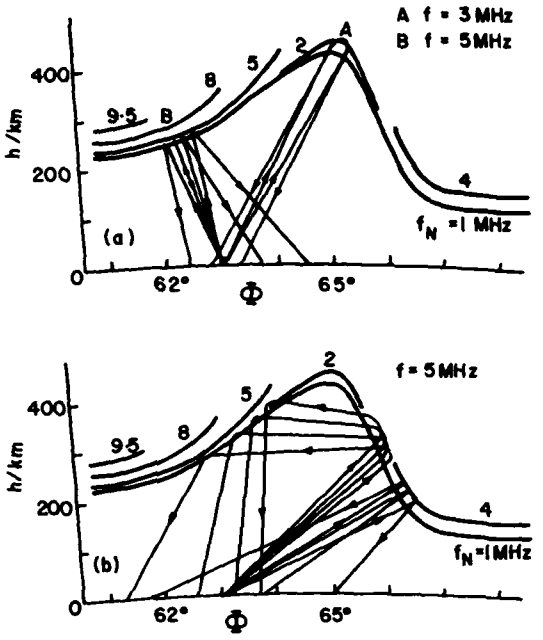


FIGURE 4 Off-vertical reflections in the high-latitude ionosphere simulated by ray tracing (from Nygren, 1977)

(a) sample model ionosphere given by plasma frequency f_N and selected raypaths for frequencies of f = 3 and 5 MHz from a transmitter at geomagnetic latitude $\Phi = 63^{\circ}$

(b) additional raypaths for 5 MHz

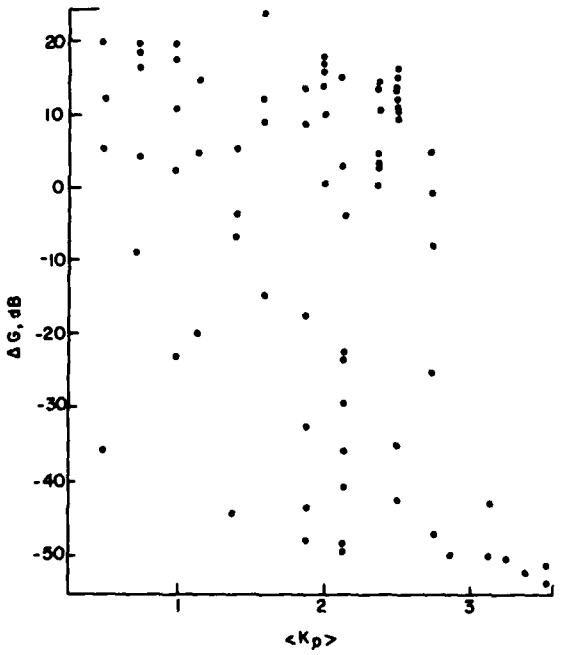


FIGURE 5 Signal focusing ΔG as a function of averaged magnetic activity index <Kp> estimated from measured data collected over a 5000 km path traversing the high-latitude trough (from Lockwood, 1980)

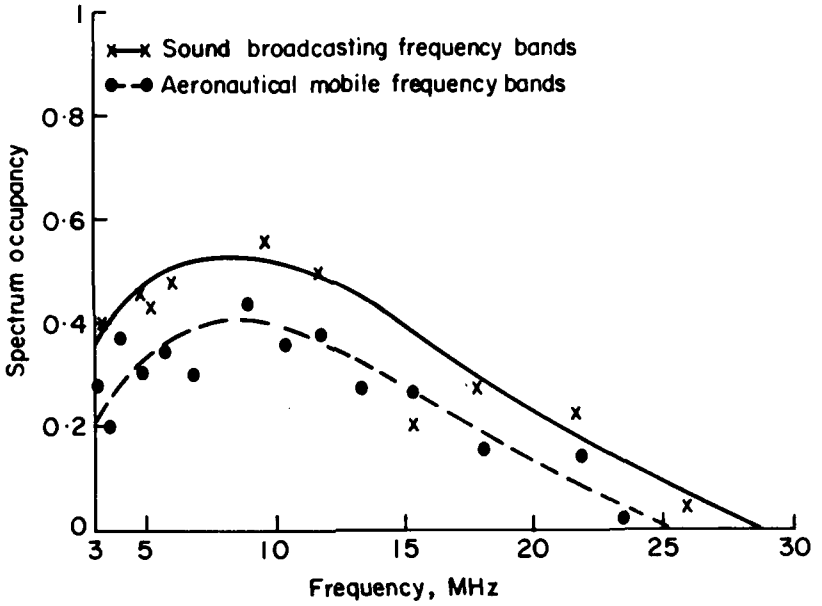


FIGURE 6 Idealised representation of spectrum occupancy above a fixed threshold for the sound broadcasting and aeronautical mobile frequency bands

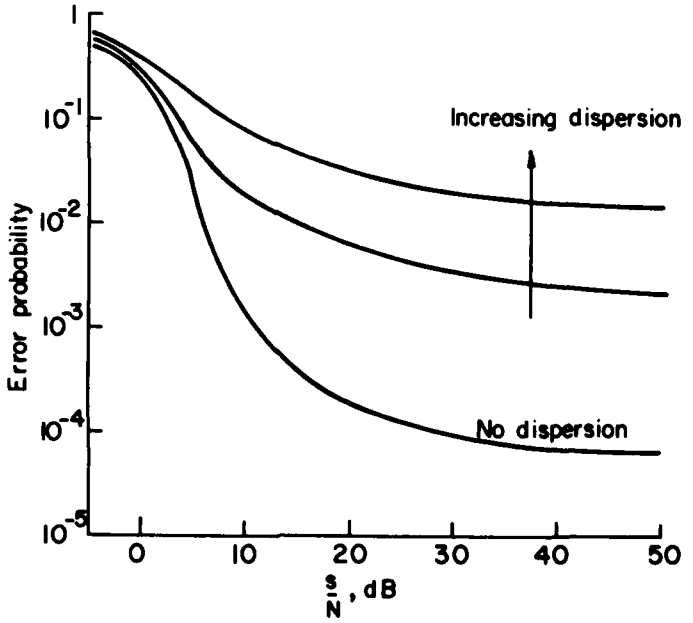


FIGURE 7 Error probability as a function of signal/noise ratio and dispersion

Ultimate Limits to Error Probabilities for Ionospheric Models Based on Solar Geophysical Indices and How These Compare with the State of the Art

J. S. Nisbet and C. G. Stehle

Ionosphere Research Laboratory
The Pennsylvania State University
University Park, Pennsylvania 16802

ABSTRACT

An "ideal" model based on a given set of geophysical indices is defined as a model that will provide a least squares fit to the data set as a function of the indices considered. Satellite measurements of electron content for three stations at different magnetic latitudes have been used to provide such data sets which were each fitted to the geophysical indices. The magnitude of the difference between the measured value and the derived equation for the data set has been used to estimate the probability of making an error greater than a given magnitude for such an ideal model. Typical values range from a 50% probability of an error greater than $5 \times 10^{16} m^{-2}$ to a 20% probability of an error greater than $10^{17} m^{-2}$. The calculations were repeated for the same data sets using the current version of the Penn State MkI theoretical ionospheric model. The error probabilities were found to be comparable, particularly for higher confidence estimates. Atmospheric Explorer C data is used to examine the causes of the fluctuations and suggestions are made about how real improvements can be made in ionospheric forecasting ability. Joule heating inputs in the auroral electrojets are related to the AL and AU magnetic indices. New magnetic indices based on the time integral of the energy deposited in the electrojets are proposed for modeling processes affected by auroral zone heating.

1. INTRODUCTION

In recent years a large amount of new information has become available on the parameters basic to ionospheric models. A new generation of atmospheric models such as those of Hedin et al. (1979), Laux and Von Zahn (1979), Barlier et al. (1978) and Stehle (1980) provide accurate information on the average behavior of not only the total density but also of the relative densities of the constituents which control both the ion production and loss processes. Accurate detailed EUV spectra have been obtained for an extensive period around the last sunspot minimum and the increasing portion of the recent solar cycle that can be used for direct aeronomical calculations, Hinteregger (1980). The Atmospheric Explorer series of satellites has provided a detailed opportunity to compare the predictions of the atmospheric ion chemistry with actual neutral ion and electron data in a series of papers such as those of Torr and Torr (1979). It thus appears a suitable time to take a new look at ionospheric models and to see in what direction progress can be made in developing better theoretical models for prediction purposes.

2. Comparison of an Existing Model with an Ideal Model Based on the Same Geophysical Indices

Nisbet et al. (1980) have examined the probability of errors in electron content estimates for three stations, Huancayo 12°S on the magnetic equator, Kharkov 50°N and Murmansk 68.5°N. Figure 1 shows the fraction of the time the magnitude of the differences between electron contents measured using the Transit satellites and those estimated using the Penn State MkI ionospheric model, Nisbet (1971) were greater than a given value at closest approach of the satellite to the observing station. All local times were included in these observations. For Huancayo, 10% of the time the errors were in excess of $1.45 \times 10^{17} m^{-2}$; for Kharkov this 10% limit was $6 \times 10^{16} m^{-2}$ and for Murmansk $7.2 \times 10^{16} m^{-2}$.

To see to what extent these errors could be reduced by an improved model, it was decided to compare these tests of the Penn State MkI model with an "ideal" model based on the same geophysical indices. An ideal model in this context is assumed to be that model which would provide a least squares fit to each actual data set. In this way, because actual measurements of electron content at that location have been fitted, all assumptions arising out of the relations between the photochemistry and the modeled parameters and about the effects of geographical variations have been taken care of.

The equation used for the least squares fitting is

$$N_t = a_o + k_1 F_{10.7} + k_2 \bar{F}_{10.7} + k_3 Kp(t - \tau) + \sum_{n=1}^{2} a_n \sin\left(\frac{2\pi n D}{365.5} - \theta_n\right) + \sum_{n=1}^{3} b_n \sin\left(\frac{2\pi n H}{24} - \phi_n\right)$$

where $F_{10.7}$ and $\bar{F}_{10.7}$ are the daily and monthly averaged 2800 MHz solar flux intensities
$Kp(t - \tau)$ is the Kp index for τ hours before the observation
D is the day number
and H is the local time in hours.

Figure 2 shows the probability of errors for these ideal models for Huancayo, Kharkov and Murmansk data for comparison with Figure 1. The errors were in excess of $7.5 \times 10^{16} m^{-2}$ at Huancayo, $5 \times 10^{16} m^{-2}$ at Kharkov and $3.5 \times 10^{16} m^{-2}$ at Murmansk, for 10% of the observations. The reduction in errors obtained by using a

least squares fit to the actual data thus represents approximately a factor of two improvement in accuracy for electron content measurements. While a factor of two improvement is not negligible, it is probably safe to say that it would not prove of overriding importance in most engineering applications. To obtain such an improvement on a worldwide basis would require considerable effort and a much larger data base than is presently available. This factor of two is probably an upper limit on the improvement that would be obtained in practice using these magnetic and solar indices because of the long term fluctuations that did not show up in these observations. It would thus appear that any order of magnitude improvement in ionospheric forecasting accuracy must come from other approaches. In particular the substitution of indices more directly related to the actual energy inputs than those used in equation 1 appears important to the understanding of the mechanisms responsible for the day to day variation.

3. High Latitude Energy Inputs

The Atmospheric Explorer series of satellites provided a unique opportunity to measure the important parameters of the high latitude inputs. Griffis et al. (1980) have used electron and ion spectra from the Low Energy Electron Experiment of Hoffman et al. (1973) as input to a computer program which calculates energy deposition and electron production due to energetic electrons and protons as a function of altitude below the satellite. This program is derived from the program TANGLE of Voudrak et al. (1978). Electron production and solar EUV measurements were used in a revised model based on the Penn State MkI model of Nisbet (1971) to calculate electron and ion densities and Hall and Pedersen conductivity profiles along the satellite orbit. Electric field measurements were derived from the ion drift measurements made by the retarding potential analyzer system of Hanson et al. (1973). Figure 3 shows the power deposition integrated over altitude for the particles and for Joule heating for orbit #4708 through the cusp and electrojet regions. The region of the cusp is denoted by a C and the evening electrojet by an E. It is apparent that in both the cusp and electrojet regions, the Joule heating exceeds that of the particle inputs by more than an order of magnitude. Figure 4 shows the altitude profiles of the heating for this orbit. It is apparent that the Joule heating maximizes in the region of 120 km. Figure 5 shows the density of O and Ar reduced to 120 km and He at 300 km calculated from the satellite observations assuming diffusive equilibrium. These measurements show very large perturbations. The increases for Ar and decreases for He and O of the densities in the region are sensitive indicators of vertical transport. The electrojet heating is important because it drives the wind system that causes the density perturbations in the auroral zone and affects the low latitude composition. These in turn influence F region densities. Transient changes are important because they launch gravity waves. Figure 6 taken from Nisbet and Glenar (1977) shows the correlation of the south component of the atomic oxygen flux at 70° magnetic latitude, height integrated above 120 km from rocket measurements, and derived atomic oxygen densities at that latitude for $0<Kp<1.2$. The equatorward fluxes in the early morning sector are most effective in reducing the atomic oxygen densities in this sector. Because it is in this sector that the equatorward thermospheric winds maximize, it is in this sector that the greatest effects of high latitude energy deposition on the low latitude thermosphere are seen.

4. Quantizing the Energy Input

It has been customary to relate ionospheric and thermospheric models to the planetary ap of Kp indices. These are based mainly on low latitude data and so are not directly related to the cause of the thermospheric perturbations, the energy deposited in the auroral zone. To illustrate the difficulties of using the ap index, figure 7 shows the variation in $\log_{10} n(O)$ at 120 km at 6 hours MLT just equatorward of the electrojet region, compared with the 3 hour ap index for 16 March 1969, taken from Stehle (1980). It is apparent that the transient time behavior of the two is very different.

It would appear that a magnetic index should have the following properties:

a. It should be directly related to the energy input using geophysically measurable parameters.

b. It should represent the integral of the energy input over a period comparable with the time constant for the energy storage and transport processes.

c. It should have sufficient time resolution to model rapid changes in energy input.

The auroral electrojet indices AL and AU give the magnetic perturbation in the north-south direction in the region of the auroral electrojet and are related to current in the morning and evening electrojets. They are available at 2½ minute intervals.

It can be shown that the power dissipated in the electrojets is given by

$$P = \frac{\int \sigma_P \, dz}{\left(\int \sigma_H \, dz\right)^2} \left(\frac{A}{400 \pi f}\right)^2 \quad Wm^{-1}$$

where $\int \sigma_H \, dz$ and $\int \sigma_P \, dz$ are the vertical integrals of the Hall and Pedersen conductivities.

A is the AL index in the morning sector or the AU index in the evening sector in nT.

$$f = \left|\frac{\Delta H}{K_y}\right|$$

where ΔH is the perturbation in the north-south magnetic field strength and K_y the electrojet current intensity, both in amps/meter.

Li (1979) has examined the available incoherent scatter and rocket data on densities in the region of the electrojet and determined that in the range $A>100$, $\int \sigma_P dz$ and $\int \sigma_H dz$ are approximately related to A. Kamide and Brekke (1975) have investigated the factor f using the Chatanika incoherent scatter radar and get average values of .2 and .34 for positive and negative bays respectively. The AL or AU indices may thus be

presumed to be relatively linear measures of the power inputs to the morning and evening electrojets.

We have been investigating the correlations with indices ML and ME based on the integrated AL and AE indices

$$ML(\tau, t_1)\ (t_o) = \int_{-\infty}^{t_o - t_1} AL(t)\ \exp\left(\frac{t + t_1 - t_o}{\tau}\right) \frac{dt}{\tau}$$

where τ is a time constant dependent on the time necessary to establish the wind system and t_1 is a propagation time delay. For the region just equatorward of the morning electrojet, it may be presumed that t_1 will be small. Figure 8 shows the same density data plotted on figure 7 with the ML (τ = 24 hrs, t_1 = 0) index. It is apparent that there is a much better correlation between the transient behavior of this index and the atomic oxygen densities than there is with the ap index.

5. CONCLUSIONS

While considerably better models of the mean ionosphere can be developed using the new information on the thermospheric densities, solar fluxes and ion chemistry, these changes are likely to make at best a reduction by a factor of two in the errors between predictions and actual values on given days. To improve estimates by an order of magnitude will require new methods of compensating for short term changes. A major source of such variations appears to be winds and gravity waves associated with Joule heating in the auroral electrojets. Understanding the energy balance of this region and magnetic indices better related to fundamental energy inputs would appear to be critical to future progress in this area.

6. ACKNOWLEDGEMENTS

Some of this work was done when one of the authors, J. S. Nisbet, was a Fulbright Fellow at Kharkov State University. The authors wish to thank W. Hanson, R. Heelis and R. Power for the ion drift data, R. Hoffman for the electron spectra, A. Nier for fifteen seconds averages of the neutral densities and the entire AE team for many helpful discussions. This research was supported by NASA grants NSG-5212 and NGL-39-009-003.

7. REFERENCES

Barlier, F., C. Berger, J. L. Falin, G. Kockarts and G. Thuillier, 1978, "A Thermospheric Model Based on Satellite Drag Data", Ann. Geophys., 34, 9-24.

Griffis, M., J. S. Nisbet and E. Bleuler, 1980, "Particle and Joule Heating of the Neutral Polar Thermosphere in the Cusp Region Using Atmospheric Explorer - C Satellite Measurements", COSPAR Twenty-third Plenary Meeting, Budapest, Hungary, 9-14 June 1980. To be published in Space Science.

Hanson, W. B., D. R. Zuccaro, C. R. Lippincott and S. Santani, 1973, "The Retarding-Potential Analyzer on Atmosphere Explorer", Radio Science, 8, 333-339.

Hedin, A. E., C. A. Reber, N. W. Spencer, H. C. Brinton and D. C. Kayser, 1979, "Global Model of Longitude/UT Variations in Thermospheric Composition and Temperature Based on Mass Spectrometer Data", J. Geophys. Res., 84, 1-9.

Hinteregger, H. E., 1980, "Various Representations of Solar EUV Variability for Aeronomical Applications", COSPAR Twenty-third Plenary Meeting, Budapest, Hungary, 9-14 June 1980.

Hoffman, R. A., J. L. Burch, R. W. Janetzke, V. F. McChesney, S. H. Way and D. S. Evans, 1973, "Low Energy Electron Experiment for Atmosphere Explorer-C and -D", Radio Science, 8, 393-400.

Kamide K. and A. Brekke, 1973, "Auroral Electrojet Current Density Deduced from the Chatanika Radar and from the Alaska Meridian Chain of Magnetic Observations", J. Geophys. Res., 80, 587-594.

Laux, U. and U. von Zahn, 1979, "Longitudinal Variations in Thermospheric Composition under Geomagnetically Quiet Conditions", J. Geophys. Res., 84, 1942-1946.

Li, Chi-Hsi, 1979, "Influence of Cusp Current Variations and Magnetic Activity on the Polar Ionospheric Current System", M.S. Thesis, Electrical Engineering, The Pennsylvania State University, University Park, PA 16802.

Nisbet, J. S., 1971, "On the Construction and Use of a Simple Ionospheric Model", Radio Science, 6, 437-464.

Nisbet, J. S. and D. A. Glenar, 1977, "Thermospheric Meridional Winds and Oxygen Depletion at High Latitudes", J. Geophys. Res., 82, 4685-4693.

Nisbet, J. S., O. F. Tyrnov, G. N. Zintchenko and W. J. Ross, 1980, "Limits on the Accuracy of Correction of Trans-Ionospheric Propagation Errors Using Ionospheric Models Based on Solar and Magnetic Indices and Using Local Measurements", Accepted for publication in Radio Science.

Stehle, C. G., 1980, "A Global Model of the Neutral Thermosphere in Magnetic Coordinates Based on AE-C Data", M.S. Thesis, Electrical Engineering, The Pennsylvania State University, University Park, PA 16802.

Torr, O. G. and M. R. Torr, 1979, "Chemistry of the Thermosphere and Ionosphere", J. Atmos. Terr. Phys., 41, 797-839.

Vondrak, R., O. de la Beaujardiere and R. Hoffman, 1978, "Simultaneous AE-C and Chatanika Radar Measurements of Auroral Particle Fluxes and Ionospheric Electron Density Profiles", presented at Atmosphere Explorer Symposium II, 1978.

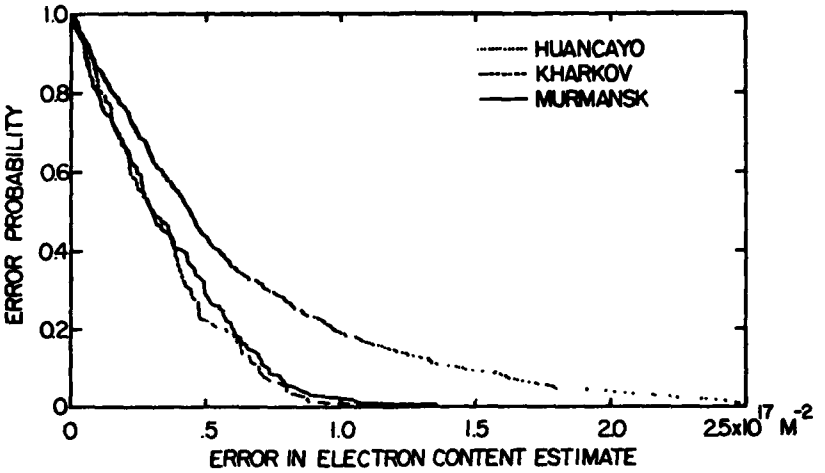


Fig. 1 Probability of making an error greater than a given magnitude when using estimates from Penn State MkI Model for Huancayo, Kharkov, and Murmansk for all local times. (Nisbet et al., 1980)

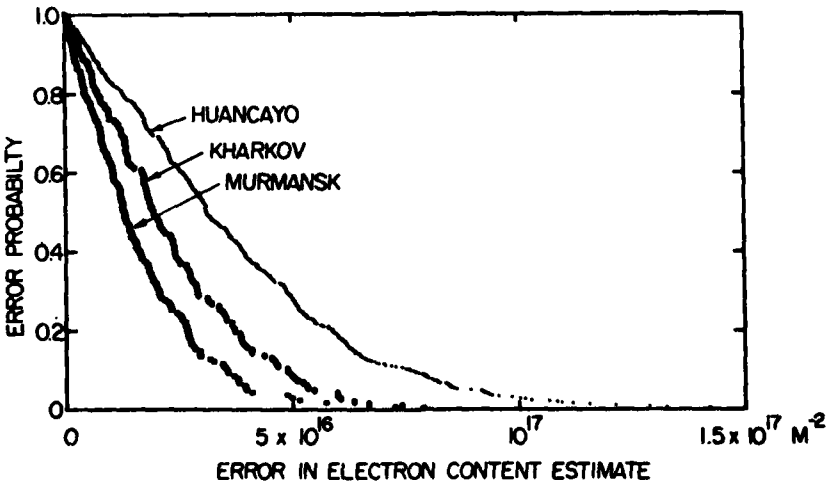


Fig. 2 Probability of making an error greater than a given magnitude when using estimates derived from least squares fit to the data set for Huancayo, Kharkov and Murmansk for all local times. (Nisbet et al., 1980)

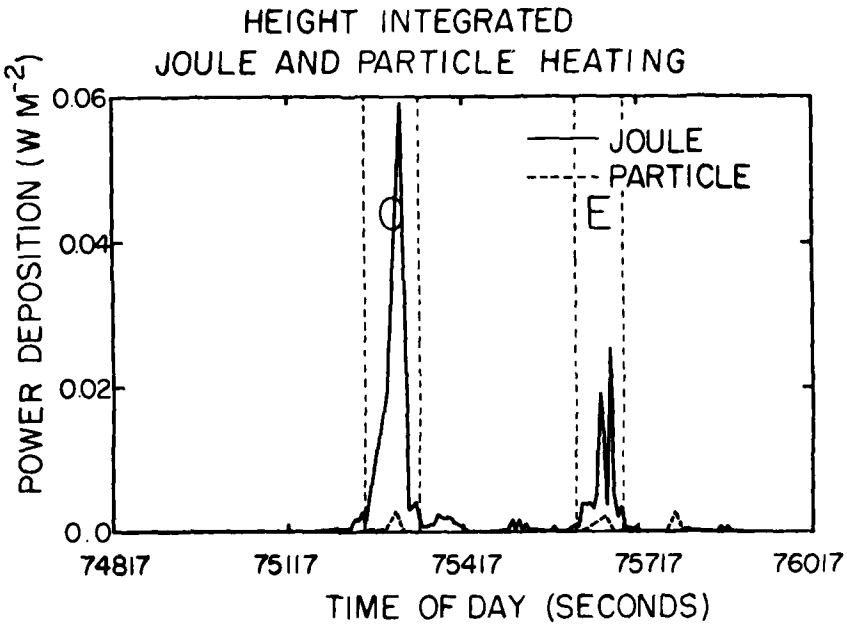


Fig. 3 Height integrated Joule heating and particle energy flux along orbit 4708 of AE-C. At 75190 sec, the orbit passes the magnetic noon meridian at magnetic latitude 71.5°, then skirts the cusp region (C), passes the 1800 hr meridian (70.6° latitude) at 75590 sec, and traverses the evening electrojet region (E). (from Griffis et al., 1980)

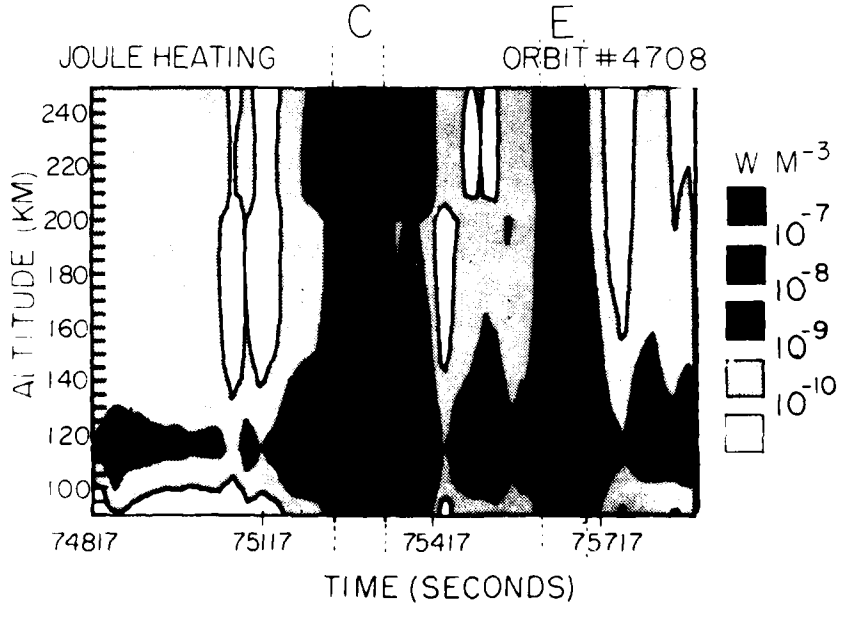


Fig. 4 Joule heating contours along orbit 4708 of AE-C. (Griffis et al., 1980)

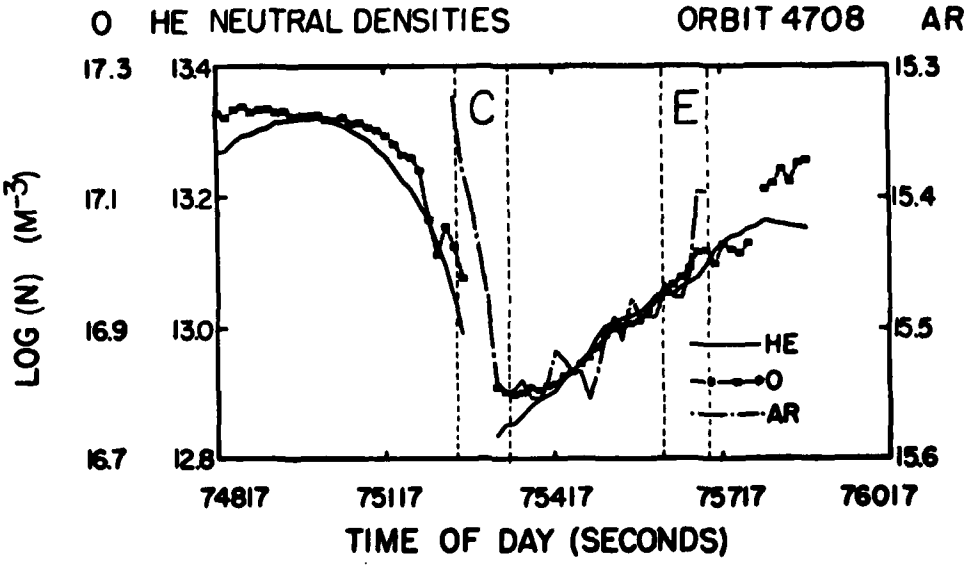


Fig. 5 Variation of neutral densities along orbit 4708 of AE-C. Helium densities are reduced to a height of 300 km, O and Ar densities to 120 km using N_2 densities and assuming diffusive equilibrium. (Griffis et al., 1980)

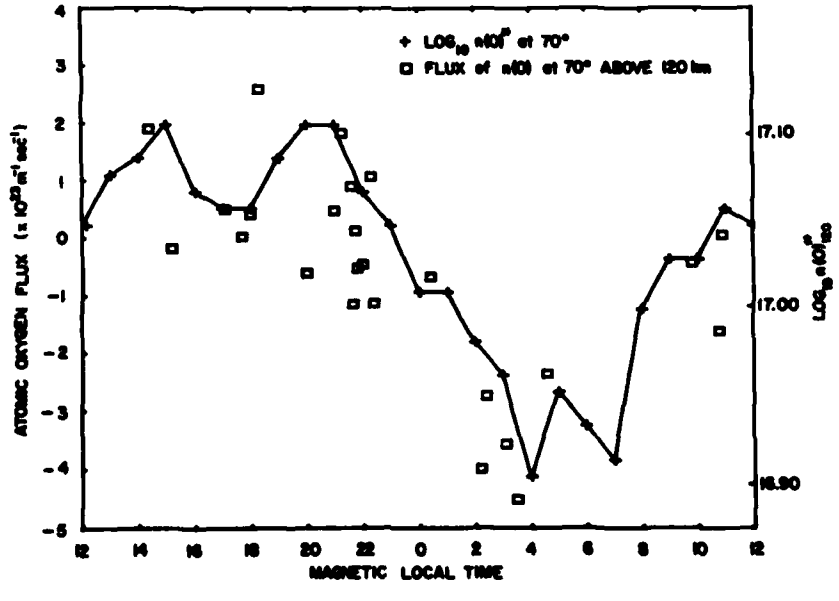


Fig. 6 Poleward component of atomic oxygen flux at 70°N geomagnetic latitude, height integrated above 120 km, and derived oxygen densities at 70° geomagnetic latitude for equinox and Kp = 0 to 1.2. (Nisbet and Glenar, 1977)

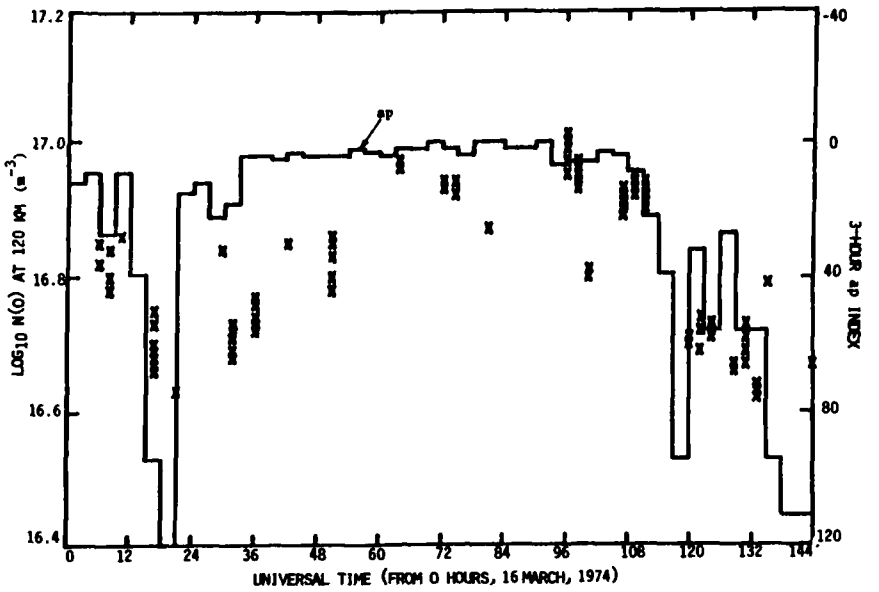


Fig. 7 $Log_{10}n(O)$ at 120 km and 3-hour ap index as functions of universal time from 16 March 1974. (Stehle, 1980)

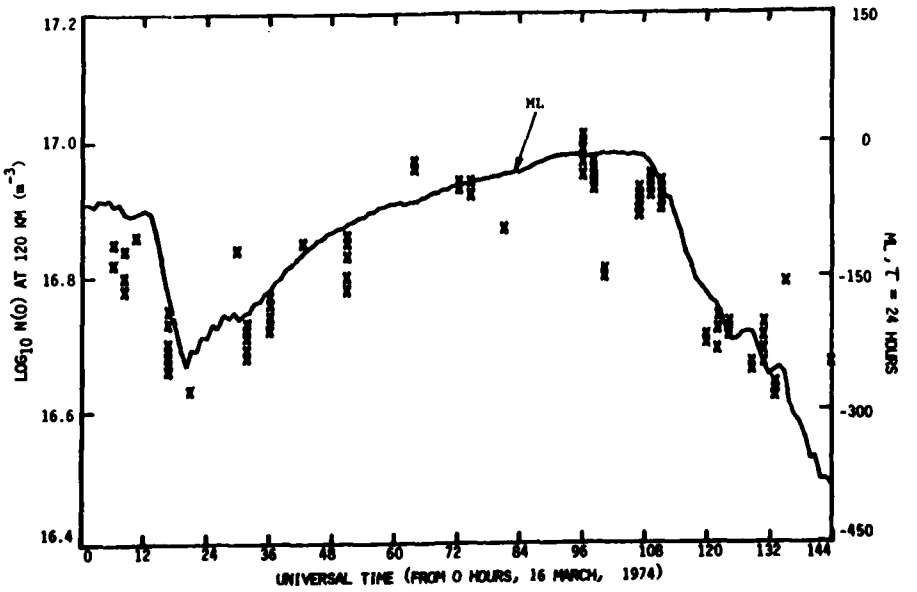


Fig. 8 $Log_{10}n(O)$ at 120 km and the ML index (τ = 24 hours, t_1 = 0) as functions of universal time from 16 March 1974. (Stehle, 1980)

THE DAY-BY-DAY VARIABILITY OF THE IONOSPHERIC PEAK DENSITY

K. Rawer / E. Harnischmacher / R. Eyfrig

University of Freiburg
D-7800 Freiburg

Dedicated to the memory of Georg G o u b a u
who died 17 Oct. 80 at Eatontown, N.J., U.S.A.

SUMMARY

In order to extend the present, monthly ionospheric prediction schedule to, finally, daily predictions, different causes of day-by-day variability are discussed. The effect of lunar tides is shown to be predictable. It is particularly important at certain hours.

1. INTRODUCTION

When the first ionospheric predictions were made in USA, the main predicted parameter was the monthly average critical frequency, foF2. In Germany, we first followed this schedule but our users were not in a position to choose their working frequencies liberally. Therefore, we soon introduced a "statistical prediction" by drawing lines of 90, 50 and even 10 % probability for establishing the desired communication (see RAWER, K., 1975). To this end curves corresponding to these probabilities were empirically determined in mass-plots with axes foF2 vs. hour, obtained for that month in the past which was used as basic input. The user, when applying the prediction to a certain day, had to bear the uncertainty which is inherent to statistics. Only in cases of stronger magnetic disturbances a warning was given with the advice to use lower frequencies whenever possible.

During the last years, one of us (E.H.) has taken up the problem trying to find some regular behaviour in the uncertainty range. He could show that an important part of this range, at least at certain hours, is synchronous with the Moon (see Section 4 below). At least, for this influence we feel that a step towards individual predictions for a given day is now feasible.

2. SEASONAL VARIATION

It is important to note that tidal influences of solar origin cannot cause a day-by-day variability, except for smaller changes which are caused by seasonal variations. Since the diurnal variation is empirically determined, all Sun-synchronous effects including tidal motions due to gravity as well as atmospheric heating by solar radiation and, of course, photo-ionization are not distinguished but taken into account as far as they are linked with the solar hour.

Seasonal variations are not quite negligible since, though systematic, they contribute to the apparent dispersion of hourly measurements. In particular, the variation of the sunrise and sunset hours provokes systematic shifts visible on the steeper parts of the diurnal curve by a broadening of the mass plot along the time axis. Also, since the timing of measurements is usually made in mean time, not in true solar hour, there is a small but visible effect of the variable equation of time. These are regular effects which could easily be taken into account when producing daily predictions. However, their importance remains negligible, except during the Sun-rise hours for which the monthly averaging process decreases the steepness of the morning rise. Individual days, in fact, often show a more drastic change than does the monthly average curve.

3. SOLAR ACTIVITY

Since the solar activity measures, e.g. the Zürich Sunspot number or the COVINGTON index (solar 10.7 cm radio emission) are often largely variable from one day to another, there was a strong feeling that variations of the solar radiation might explain most of the ionospheric variability. Such an influence is certainly existing but, unfortunately, not straightforward. Since long time the solar X-ray intensities are continuously measured aboard the SOLRAD satellites of U.S. NRL. This wavelength range is, however, absorbed in lower regions of the ionosphere, not in the F-region. Regular and continuously calibrated satellite measurements of the EUV-spectrum which is decisive for the aeronomy of this latter region were first made since the end of 1972 aboard the German-US satellite AEROS-A (SCHMIDTKE, G., et al., 1974). These authors claim to give absolute energy fluxes. When the data so measured were used as input into a modern, theoretical aeronomic model computation, it was found that the fluctuation so computed was almost one order smaller than that of observed electron densities (ROBLE, R.G., and SCHMIDTKE, G., 1979). An invariable neutral atmosphere model was used in these computations. Meanwhile G. SCHMIDTKE (see LÄMMERZAHL, P., et al., 1979) repeated the computations for the same test days admitting different neutral atmosphere models according to the daily solar and magnetic activity indices as used in the MSIS-model (HEDIN, A.E., et al., 1977). Despite the fact that his EUV intensities showed some 27 d recurrence tendency (an effect

of solar rotation) SCHMIDTKE could not detect a significantly larger variability of the production term than with the average MSIS-model.

We draw two conclusions from these findings. First, the production term in the ionization balance equation cannot be the main source of the observed day-by-day variability so that we remain with the loss and transport terms as possible causes of larger variations. Second, the description of the neutral atmosphere as given by the MSIS-model in function of magnetic and solar activity indices is insufficient for detailed aeronomic applications. We suggest that the behaviour of minor neutral constituents and their transport be further investigated and appropriate indices be established for depicting the neutral atmosphere more appropriately in view of the aeronomy of individual days.

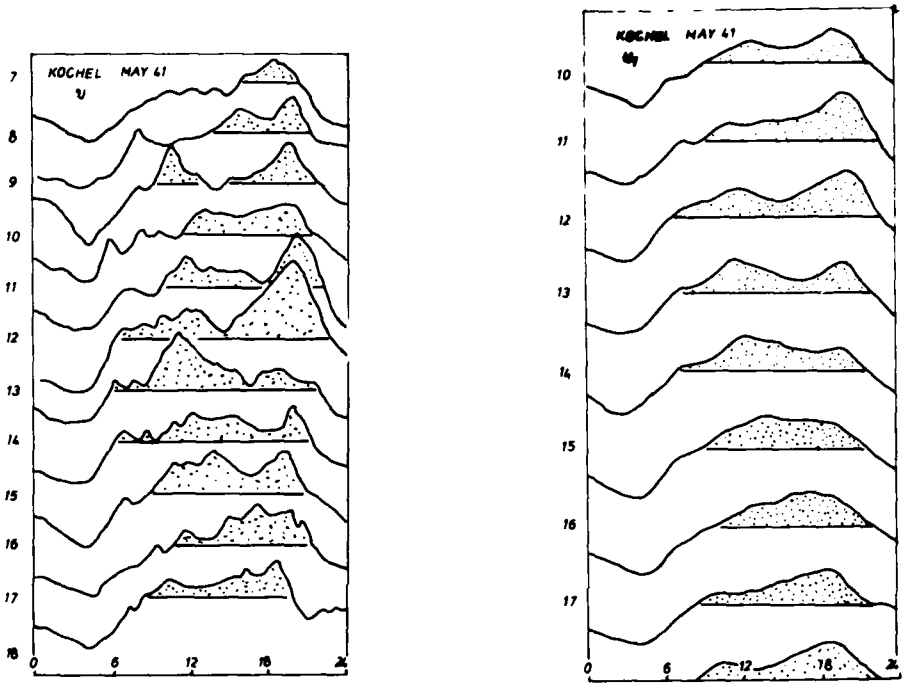


Fig.1 Measured daily curves of U = foF2 (Fig.1a, left hand) and "oblique sliding average" U_1 for same data (Fig.1b, right hand). Kochel, 1941, May 7 (1o) to 18 (17); dates left hand side.

4. LUNAR TIDES

In his (1967) survey paper S MATSUSHITA reported on different investigations which had been made with the classical method of tidal analysis, i.e. considering all data from a few years and assuming that the lunar effect should be independent of the solar hour. The tides so found are quite small, particularly at midlatitudes where the amplitude of the semi-diurnal lunar tide was found to be inferior to 0.1 MHz. Somewhat higher values are reported from the equatorial belt but 0.15 MHz is no-where exceeded. However, as early as 195o J.BARTELS had shown that by restricting the time range of the analysis to southern summer months and the hours 10 to 14 h, an extremely large amplitude of 1.1MHz was obtained for Huancayo. One year later, O. BURKARD (1951) found that the lunar effect at the same station depends strongly on the solar hour, large effects appearing at certain solar hours only. These results show that the lunar influence cannot be considered as an independent contribution of fixed phase through all hours and seasons; with other words: linear perturbation theory is apparently not adequate.

Having this in mind, two of us (HARNISCHMACHER, E., and K. RAWER, 1978/9, 1979) recently reconsidered the problem at temperate latitudes. We could show that slightly averaged diurnal curves of the critical frequency foF2 are astonoshingly similar when selecting days from years of comparable solar activity, same month and - most important - same lunar phase. This was proven for summer months at temperate latitudes when the sunrise/sunset effect is less important than in winter. We felt it might be interesting to extend the time range of these investigations further into the past and consider also the effects of geomagnetic disturbances at temperate latitudes. Therefore, we want back to the ionosonde measurements obtained during the three years 1941..43 at the former ionospheric station Kochel in Bavaria - besides one of the worlds oldest ionospheric sounding stations (founded by J. ZENNECK and G. GOUBAU).

These years of low solar activity were compared with the measurements obtained at Freiburg (same latitude) during the (solar maximum) years 1954 and 1965. Thus the present investigation is mainly based upon midsummer measurements obtained during three minima of the solar activity cycle.

4.1 Data Preprocessing

In order to eliminate irregular, short term perturbations (i.e. features lasting less than an hour, see Section 5) we preprocessed the daily/hourly measured values of foF2 (here called U) in a weak, two step smoothing procedure after the following rules:

For day D and (solar) hour H the actually measured value U (D,H) was replaced by

$$U_1 (D,H) = (U (D-1, H-1) + 2U (D,H) + U (D+1, H+1)) / 4 \quad . \tag{1}$$

Since the moon has an apparent diurnal period of 24.82 h, our averaging process means smoothing over ± 1d (with weights 1:2:1) at almost fixed lunar hour. It can thus be considered as a weak filtering process with preference in favour of the lunar tides while the solar ones are slightly decreased (for the half day tide in the ratio 0.7:1). The main effect is, however, a strong reduction of irregular components which are to a large part due to the so-called "travelling ionospheric disturbances" (see Section 5). Their influence is largely filtered away as seen by comparing Figs. 1a (for U) and 1b (for U_1).

Let us now consider the lunar phase $\mu(D)$: this is the difference (in h) between solar and lunar noon and was tabulated by J. BARTELS and G. FANSELAU (1937). We proceed to an averaging over the lunar phase by replacing $U_1(D,H)$ obtained for a day D with $\mu(D) = \mu_1$ and an hour H by

$$U_2 (D,H) = (U_1 (D_{-12}, H) + 2U_1 (D,H) + U_1 (D_{+12}, H)) / 4 \tag{2}$$

where day number D_{-12} is selected by the condition that its μ-value is $(\mu_1 - 12h)$, and D_{+12} similarly for $(\mu_1 + 12h)$. The procedure means averaging over equal phases of the main (12 h) soli-lunar tide which comes again to the original phase after about half a month. In this step phenomena with periods around half a month are fully conserved while phenomena of shorter and longer periods are decreased in importance.

Since for these two smoothing processes we need an uninterrupted set of hourly measurements, missing data were interpolated by hand-drawing. When doing so, the over-all similarity of diurnal curves was taken into account; this is easier in a man-controlled procedure than by automatic interpolation methods e.g. splining. Anyway, except for June 1941, the basic measurements were almost uninterrupted so that the uncertainty due to these interpolations remains negligible.

4.2 Assessment of the Lunar Effect

Smoothed (U_2) diurnal curves show a few typical features which depend clearly on the lunar phase. As first noted by O. BURKARD (1951) the lunar effect is more pronounced at certain solar hours, in particular around the evening maximum which uses to take its minimum value around half Moon. We find that even the sign of the effect is variable with solar hour and lunar phase. (This explains why classical tidal analysis ended up with extremely small amplitudes.)

Since one expects the largest difference between opposed lunar phases the lunar tidal effect is better seen when the U_2 curve of a given day with lunar phase μ_1 is compared with that of the day at $\mu_2 = \mu_1 - 6h$. Both curves were drawn in each of the diagrams shown in Fig.2 obtained 1954 at Freiburg during midsummer (17 May ... 15 July). The difference range between both curves is dotted where the effect was positive (i.e. U_2 at μ_1 was greater than at $\mu_1 - 6h$) and open in the reverse condition. There appears a clear lunar modulation with a 12h periodicity in μ, the phase of which, however, depends on the solar hour. In this midsummer period the largest (positive) lunar effect appears around 20 h with maximum at $\mu = 0$ and 12 h respectively. A second maximum is seen around 8 h, with slightly smaller amplitude and similar phase. Maxima with reversed sign appear, however, at hours between these values. Thus the soli-lunar 12 h effect may be described as difference against the average, at least roughly, by an expresssion of the kind

$$M (\mu,H) = C_1 + C_2 \cos (\omega_1(\mu+c_3)) + D_1 \cos (\omega_1(t+d_2)) + E_1 \cos (\omega_2(t+e_2)) \cdot \cos (\omega_2(\mu+\phi)) + F_1 \cos (\omega_3(t+f_2)) \cdot \cos (\omega_2(\mu+\psi)) \quad ,$$

$$\phi = e_3 + e_4 \cos (\omega_2(\mu+e_5)) \quad , \quad \psi = f_3 + f_4 \cos (\omega_2(\mu+f_5)) \tag{3}$$

where t is solar hour of day, $\omega_1, \omega_2, \omega_3$ are the pulsations of the first three solar tides of 24, 12 and 8 h, resp.:

$\omega_1 = 0.26180/ h$; $\omega_2 = 0.5236 / h$; $\omega_3 = 0.7854 / h$.

(We felt, it is sufficient to use the lunar phase μ as day-by-day variable instead of giving a continuous description with the lunar pulsation of 0.21277 / d, corresponding to the lunar synodal period of 29.5306 d).

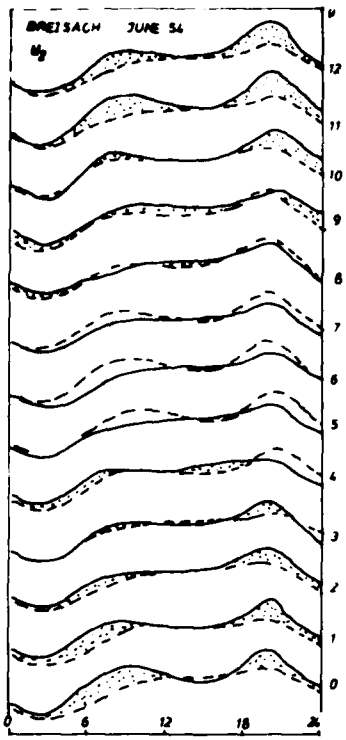


Fig.2 Comparison of daily curves U_2 (after second smoothing) for a given day (full curve) with a later day of opposed phase of semi-diurnal lunar tide (broken curve). Lunar phase μ at right hand. Breisach, 1954, June.

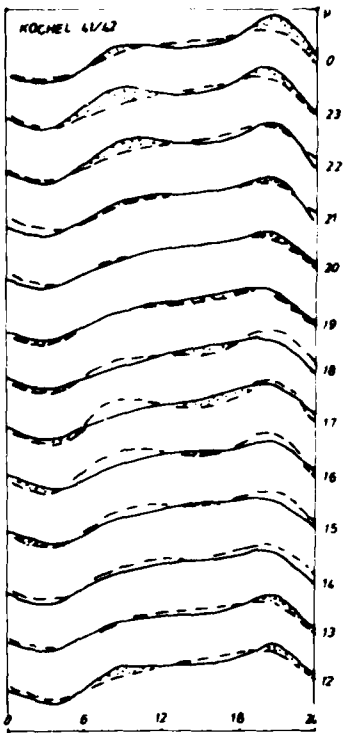


Fig.3a Same as Fig.2 for Kochel, but averaged over summer 1941 and 1942.

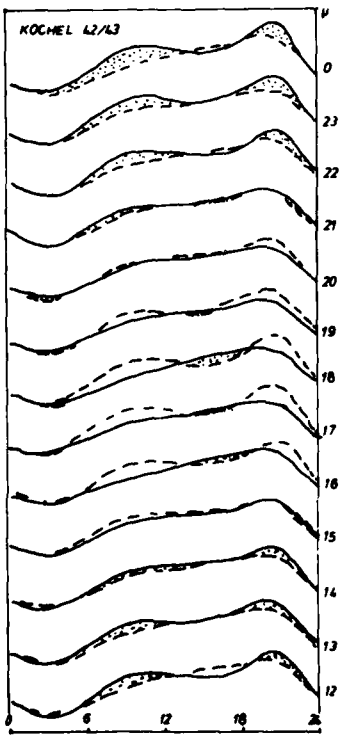


Fig.3b Same as Fig.2 for Kochel, but averaged over summer 1942 and 1943.

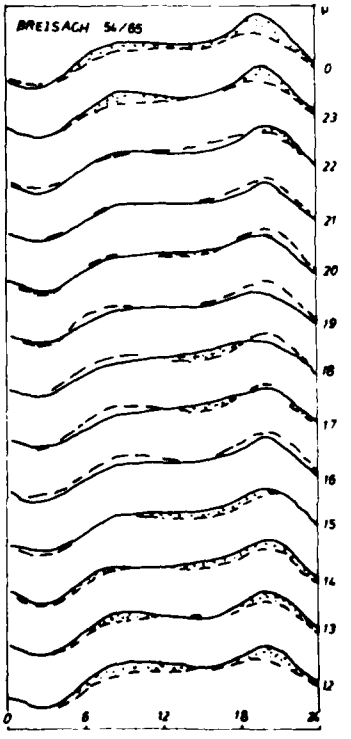


Fig.3c Same as Fig.2 for Breisach, but averaged over summer 1954 and 1965 (solar minimum years).

Fig.3 shows that the amplitude of the main, semi-diurnal effect, E_1 in Eq.(3), is of the order of [illegible] MHz. This is not at all negligible for practical applications. The first three terms of Eq.(3) are needed when numerical optimization is applied for determining the 15 parameters appearing there; however, they are without interest for the purpose of prediction since they describe features of the monthly pattern which do not occur systematically, e.g. influences of magnetic disturbances.

These terms are much smaller, and the lunar effects appear clearer but somewhat smoothed when data from more than just two lunar months are combined. Therefore, we finally considered the whole midsummer period and two years of measurements, namely 1941/42 and 1942/43, for Kochel, and similarly with the two solar minimum years 1954 and 1965 for Freiburg. Results are shown in Figs.3. It is seen from these figures that the lunar tidal pattern is quite similar for these three combinations and has rather similar amplitudes, depending on the solar hour.

4.3 Discussion

Such description is, of course, a tool for predicting the lunar effect upon foF2. To this end one has to use the last two terms in Eq.(3) with amplitudes E_1 and F_1. For most hours the correction remains inferior to [illegible] MHz but at certain hours it can even be larger. Our results, when expressed numerically, could therefore be used for a short term prediction, about one week in advance, starting with measured diurnal curves of foF2. Except in cases where a magnetic disturbance intervenes (see Section 6 below) this procedure should be quite helpful.

Another use would be with predictions over longer terms, e.g. one year in advance, or even over a sunspot cycle, by specifying a lunar modulation, day by day, around the usually predicted monthly average curve. This long term procedure, mostly quite successful, occasionally breaks down, when the planetary constellation in one of both periods is 'disturbed' i.e. when gravitational effects of Jupiter (and sometimes Venus) are not negligible ([illegible], and k. [illegible], 1[illegible]). Short term prediction, however, does not fall under such a criticism.

In conclusion, we feel that up to [illegible] of the so-called day-to-day fluctuation of foF2 [illegible] Regular lunar tidal phenomena and therefore predictable. We could, at least with [illegible] terms to this effect in the foregoing.

We note again that effects of solar time, by definition, are depending on the solar hour and thus taken into account by current prediction methods anyway.

5. [illegible]

The large influence of plasma transport induced by neutral winds has been demonstrated by [illegible] and [illegible] (1[illegible], 1[illegible]). These authors assumed a neutral wind system deduced from the world map of average exospheric temperature as given by [illegible] and [illegible] (1[illegible]). Even if seasonal variations are now admitted there can be no doubt that actual neutral winds are more variable than deduced with this oversimplified model. Day-to-day changes due to lunar tides should be the cause of the phenomena discussed in Section 4. However, there are also short term phenomena of more local character which are known as 'travelling ionospheric disturbances'. These plasma motions are very probably caused by neutral motions propagating as internal gravity waves in the atmosphere ([illegible], [illegible], and [illegible]). Locally, the passing of such phenomena lasts less than half an hour, but they appear, by day, several times an hour. In hourly measurements they just cause a fluctuation of foF2, the amplitude of which is up to [illegible] (Bibl, k., and k. [illegible], 1[illegible]), only very rarely more than [illegible] MHz. While this is not negligible it is, unfortunately, not predictable either. Therefore, unpredictable dynamic phenomena in the atmosphere explain a part of the observed variability but only a minor one.

6. [illegible]

These effects are [illegible] even more complex. They are extremely strong in the auroral zone and at least quite important in the subauroral zones. In Section 4 they were largely [illegible] by carefully averaging such that tidal effects appeared clearly in the averaged data. For individual days (1), and even after the first, short term smoothing procedure (2), their influence is sometimes quite important. We have checked in particular the '4' data and found all stronger magnetic perturbations appearing in the F_1 data but more [illegible]. Unfortunately, it is not possible to give, rather than "warnings", precise predictions which could be used for a numerical application. Quite generally magnetic disturbances constitute a major source of uncertainty when ionospheric predictions shall be utilized for an individual day.

The Kochel measurements of '4' are particularly instructive in this context. In Figs. 4 and 5 we present examples of strongly disturbed and magnetically quiet periods using the [illegible] of the planetary geomagnetic indices ki as an indicator. Fig. 4a shows daily [illegible] curves for one disturbed period (with maximum of [illegible] on May [illegible]) continuing with values above [illegible] until May [illegible], see right hand disturbance value [illegible]. Fig. 4b refers to a much stronger case (with values above [illegible] during three days, maximum [illegible] on July [illegible]). In both cases the interesting [illegible] clearly stand out in the last [illegible]. Apart from this, the shape of the diurnal curves is strongly [illegible] the forenoon maximum disappears and that in the afternoon is strongly increased. In the stronger July disturbance a noon maximum all are [illegible] the only one during the two most disturbed days.

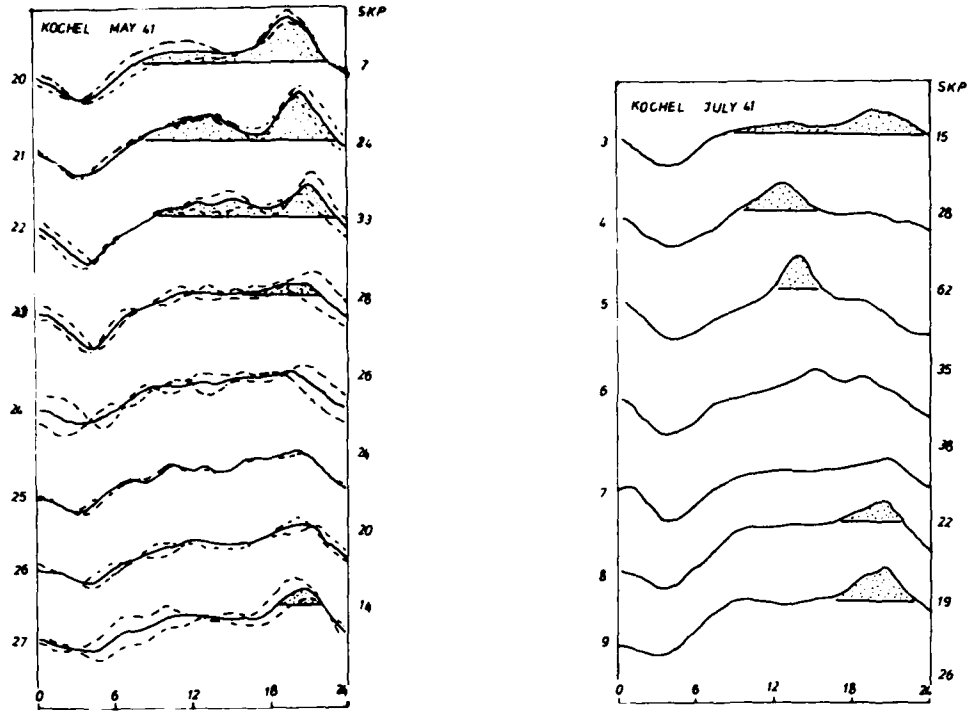


Fig.4 Full curves: U_1 for Kochel 1941 during disturbed periods in May (left) and July (right). Left hand: date; right hand: daily sum of three-hourly indices Kp. (Broken curves in Fig.4a: shifted fore- and after-day U-curves as used for establishing U_1).

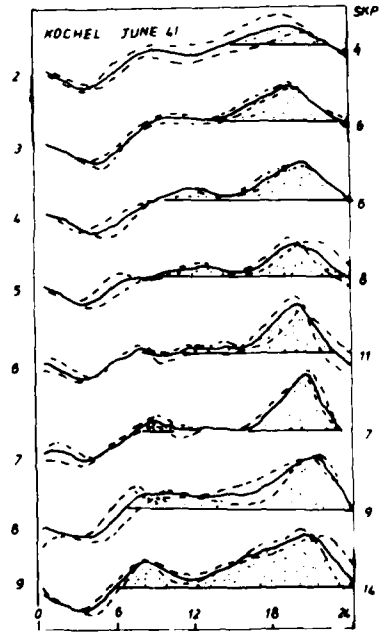


Fig.5 Same as Fig.4a for a quiet period in June 1941 (Kochel).

Quite different is Fig. 5 for a quiet period in June showing the forenoon and, in particularly, the afternoon maximum clearly present.

These examples prove that strong magnetic disturbances apparently destroy the usual neutral wind system which redistributes the plasma by transport along the magnetic field lines (KOHL, H., and J.W. KING, 1965, 1967). The overwhelming influence of heat deposit in the auroral oval makes these conditions so particular that they should be described by a special set of diurnal curves as we have done in the past for prediction purposes (RAWER, K., 1947). Fortunately, such cases are the exception, at least at temperate latitudes.

7. CONCLUSIONS

Apart from the still obscure short-term effects of variable solar activity (Sect. 3) and of dynamic phenomena in the atmosphere (Sect. 5) there are two kinds of systematic variations which might become helpful for future day-by-day predictions . With present time knowledge the great changes occurring during magnetic disturbances can not yet be predicted safely.

7.1 References

BARTELS, J., 1950, "Ebbe und Flut in der Ionosphäre", Ber. dtsch. Wetterdienstes US-Zone, Nr.12, 30-33.

BARTELS, J., and G. FANSELAU, 1937, ZS f. Geophysik 13, 311-328, (reprint as "Geophysikalische Mondtafeln 1850-1975", Berlin: Springer, 1938).

BIBL, K., and K. RAWER, 1959, "Travelling disturbances originating in the outer ionosphere", J. Geophys. Res. 64, 2232-2238.

BURKARD, O., 1951, "Studie zum ionosphärischen Gezeiteneffekt", J. Atmosph. Terr. Phys. 1, 349-352.

HARNISCHMACHER, E., and K. RAWER, 1978/79, "Lunar tides in foF2", Riv. Ital. die Geofisica e Scienze affini V, 105-108.

HARNISCHMACHER, E., and K. RAWER, 1979, "Basic findings helpful for daily ionospheric predictions", in: AGARD Conference Proceedings No. 263, on "Special Topics in Propagation", 16-1...16-10.

HEDIN, A.E., C.A. REBER, G.P. NEWTON, N.W. SPENCER, H.C. BRINTON, H.G. MAYR and W.E. POTTER, 1977, "A global thermospheric model based on mass spectrometer and incoherent scatter data - MSIS 2. Composition", J. Geophys. Res. 82, 2148.

HINES, C.O., 1968, "Applications of gravity wave theory to upper atmospheric studies", 364-370, in: RAWER, K., (ed.), "Winds and Turbulence in Stratosphere, Mesosphere and Ionosphere", Amsterdam: North-Holland Publ. Comp., and Newyork: Academic Press.

JACCHIA, L.G., and J. SLOWEY, 1963, "Accurate drag determinations for eight artificial satellites: atmospheric densities and temperatures", Smithson. Contr. Astrophys. 8, 1.

KOHL, H., and J.W. KING, 1965, "Atmospheric winds between 100 and 700 km and their effects on the ionosphere", Nature 206, 699-701.

KOHL, H., and J.W. KING, 1967, "Atmospheric winds between 100 and 700 km and their effects on the ionosphere", J. Atmosph. Terr. Phys. 29, 1045.

LÄMMERZAHL, P., K. RAWER and M. ROEMER, 1979, "Ergebnisse des AEROS-Satellitenprogramms", Max-Planck-Inst. f. Kernphysik, Heidelberg (W.-Germany).

MASUSHITA, S., 1967, "Lunar Tides in the Ionosphere", Handbuch der Physik (Encyclopedia of Physics, ed. J. BARTELS, Berlin: Springer Verlag), vol. 49/2, 547-602.

RAWER, K., 1947, "Les perturbations ionosphériques dans la zone des aurores polaires", Revue Scientifique 85, 287.

RAWER, K., 1975, "The historical development of forecasting methods for ionospheric propagation of HF waves", Radio Sci. 10, 669-679.

ROBLE, R.G., and G. SCHMIDTKE, 1979, "Calculated ionospheric variation due to changes in the solar EUV flux measured by the AEROS spacraft", J. Atmosph. Terr. Phys. 41, 153-160.

SCHMIDTKE, G., W. SCHWEIZER and M. KNOTHE, 1974, "The AEROS EUV-spectrometer", J. Geophys. 40, 577-584.

EXPERIMENTAL VALIDATION OF THE ONSOD OMEGA PREDICTION METHOD

T.B. Jones and K. Mowforth
Physics Department
University of Leicester
Leicester, U.K.

SUMMARY

The diurnal and seasonal changes in the D-region electron density influence the phase and amplitude of very low frequency (VLF) radio waves propagating in the waveguide formed by the earth as one wall and the ionosphere as the other. In the Omega navigation system, position is determined by comparing the phase of the signals received from two pairs of VLF transmitters, each pair providing a line of position (LOP). Thus the changes in the D-region will influence the measured position and the extent of the resulting error will depend on the deviation of the phase velocity from its nominal assumed value.

The Omega Navigation System Operations Detail (ONSOD) issue predictions of the expected LOP errors on a world-wide basis for 15 day intervals. These corrections are based on both physical and empirical modelling of the D-region. From these models, the phase velocity of the three Omega frequencies (10.2, 11.3, 13.6 kHz) are computed and the expected LOP errors derived. The predictions are available in a simple tabular form which allows the navigator to correct his measured position given his approximate distance from the transmitters, and the time of day and season.

In the present investigations the position of three fixed receiving sites has been determined over a two-year period using the following Omega transmitters: A - Aldra, Norway, B - Liberia and D - North Dakota, U.S.A. From these observations and the known location of the receivers, the actual system errors have been determined. These have been compared with predicted errors of the ONSOD program.

The day to day variability of the ionosphere is found to restrict the usefulness of any prediction method based on mean models. It appears that the addition of some form of real time updating might lead to an improved error correction technique.

1. INTRODUCTION

Omega is a radio navigation system employing Very Low Frequency (VLF) signals which provide positional information to users on a world-wide basis. Position is determined by a phase technique and the stability of VLF propagation to great distances via the ionosphere is therefore an essential requirement. When fully implemented, eight transmitting stations will be deployed to provide uniform coverage over the globe. The primary navigational signal is transmitted at a frequency of 10.2 kHz at a power of 10 kW and additional frequencies of 13.6 kHz and 11.3 kHz are also provided.

The transmissions from each station are synchronized, (by means of Caesium standards), unmodulated, continuous wave signals which are time-shared with 0.9 seconds on and 9.1 seconds off intervals. Outside the "near-field" zone, each transmitter radiates a stable signal pattern which is repeated in a radial direction at approximately 30 km segments (for 10.2 kHz transmissions) from the station. These repeated segments (or wavelengths) thus provide a measure of great-circle distance.

The system is customarily used in a 'hyperbolic mode'. In this mode signals from a pair of transmitters are compared with an internal oscillator not synchronized to the transmitter and a phase difference measured. The locus of geographic positions whichgives rise to the same phase difference with respect to two transmitters is a closed curve on the earth's surface. Clearly two such curves, referred to as lines of position (LOPs) obtained from two transmitter pairs (i.e. three or four transmitters), will establish a position fix. Since the calculation of these LOPs is somewhat involved, charts of their location have been prepared for large areas of the world. These charts indicate the LOP lattices corresponding to several transmitter pairs and are plotted in sufficient detail to allow accurate interpolation by the user.

2. THE ONSOD PREDICTION PROGRAM

The Omega navigation charts are constructed on the assumption that VLF radio waves travel with the same phase velocity (V_p) in all directions, specifically:-

$$V_p = \frac{C}{0.9974} \qquad (1)$$

As a result of non-uniformities in the earth-ionosphere waveguide, the phase velocity, V_p is not constant. Its value will depend on factors such as ionospheric and ground conductivities, and the direction of propagation relative to the geomagnetic field. Thus V_p will vary with time of day, season and location. A user's position, as determined by received phase information and charts alone, will not necessarily coincide with his true position and the magnitude of the error is related to the difference between the actual V_p and its assumed constant value. A method for predicting the magnitude of these errors has been developed by the Omega Navigation System Operations Detail (ONSOD).

The ONSOD prediction program (Morris & Milton, 1974) calculates the phase of the VLF signal received from an Omega transmitter, at a point anywhere on the earth's surface for any time of day or season. The program models the effects of changes in various geophysical quantities on the phase velocity of the Omega signal. This approach differs from previous analytical methods, such as the Waveguide Mode Theory, in that it does not seek to relate analytically the observed radio wave characteristics (e.g., phase or amplitude), to a given ionospheric electron density model. Instead the relevant values of attenuation, phase velocity and mode excitation are assumed to be specified in terms of readily defined characteristics of the path such as its orientation, latitude, ground conductivity, diurnal and seasonal period, etc. The terms of the equation in the ONSOD program are parameterised by three space dependent quantities:

(1) ground conductivity
(2) geomagnetic latitude (dip angle)
(3) magnetic path bearing

2.1. Path Segmentation

The propagation path is divided into segments, each of 0.01 radians arc length. The segments are then classified into three groups (see figure 1) according to their proximity to the transmitter and receiver as follows:

(a) Excitation Zone and (b) De-Excitation Zone

The excitation and de-excitation zones are regions similar to the Fresnel zones of antenna theory, and it is assumed that propagation in these regions is only weakly affected by the presence of the ionosphere. However, the characteristics of the signal are dependent upon the amount of energy transferred to and from the dominant propagating mode at the transmitter (excitation) and receiver (de-excitation) antennas. In the Waveguide Mode Theory this effect is described by the modal excitation factor.

(b) Mid-path Region

In this region propagation is strongly affected by the ionosphere. The signals are assumed to consist essentially of a single TM mode propagating in the earth-ionosphere waveguide.

2.2. The Diurnal Functions

For a given segment of an Omega propagation path the VLF phase velocity is strongly controlled by the *solar zenith angle* (χ). To account for this variation, the ONSOD program employs a 'diurnal function' which is derived (Swanson & Bradford, 1971) from empirical and theoretical considerations and relates the local phase velocity to Cos χ. As experimental observations have demonstrated, anomalous propagation features are sometimes observed during the sunrise period. A further weighting function is introduced into the prediction algorithm to account for this phenomenon.

2.3. Arrangement of Propagation Correction Tables

The results of the ONSOD prediction program are compiled into a set of correction tables which provide the navigator with a simple means of improving the accuracy of the measured position fix (figure 2). The corrections are tabulated for a series of locations covering the whole globe, each location being four to six degrees of latitude or longitude from its neighbours. The tables contain a set of 24 hourly corrections each set being valid for a 15 day (half monthly) period. The corrections for individual transmitters are expressed in ceticycles (cecs), or hundredths of a wavelength. When converted to LOP's the phase differences are expressed in centilanes (cels), the geographic size of the cel is dependent on the navigator's position.

3. EXPERIMENTAL ARRANGEMENT

An experimental investigation of the performance of the Omega navigation system in the United Kingdom has been carried out by the Leicester Group in conjunction with the Admiralty Compass Observatory (ACO). The accuracy of the predicted corrections was determined from observations of LOP's at three known receiver locations. A detailed study of the propagation mechanisms for the propagation paths involved was undertaken in order to determine the likely sources of the prediction errors.

Monitoring of a number of Omega transmitters has been carried out by ACO at two fixed locations in the United Kingdom; RAE, Farnborough and the Butt of Lewis (figure 3). In addition, a fully automated Omega receiving facility has been developed at Leicester University and this has recorded data continuously over the past $2\frac{1}{2}$ years. At Farnborough and the Butt of Lewis hourly LOP values in digital form are obtained from standard Omega navigation receivers. At Leicester, LOP data are collected by an Omega navigation receiver at five minute intervals. In addition, facilities are available for recording the phase values of individual Omega transmitters using phase tracking receivers.

The United K ngdom experiments have been mainly concerned with the following Omega transmitter pairs:

Norway (A) and Liberia (B) - LOP AB

Norway (A) and North Dakota (D) - LOP AD

Liberia (B) and North Dakota (D) - LOP BD

The data has been collected for both the primary navigation frequency of 10.2 kHz and the alternate frequency of 13.6 kHz. Table 1 lists the Omega lane widths in kilometers for LOP's AB, AD and BD at the

three monitor sites for 10.2 kHz and 13.6 kHz.

4. COMPARISON OF PREDICTED AND EXPERIMENTAL LOP DATA

The validity of the prediction program has been assessed by comparing the predicted LOP error with the experimentally determined values. The position of the LOP is measured from the received phase information and its error determined from the difference between this and the known position of the receiving location. The variation of the LOP error and its comparison with the predicted LOP are discussed below.

4.1. The Farnborough Monitor

4.1.1. 10.2 kHz Results

Figure 4(a) illustrates the typical LOP error data collected at Farnborough for transmitter pair AB measured at 10.2 kHz. A regular diurnal variation is clearly evident and is of the order of 40 cels. This demonstrates the regular solar control of the lower ionosphere and the resulting effect on the VLF propagation. The pattern is fairly repeatable from day to day. The magnitude of the LOP error is at its smallest around noon, demonstrating that at this time the VLF phase velocity deviates little from the constant value of equation (1). VLF waves propagating along the low latitude path from Liberia, (transmitter B), to Farnborough, are influenced strongly by the occurrence of solar flares. These cause a phase advance and the resulting effect on the LOP AB can be seen in the data set, e.g., 10:00 UT, 23rd September.

LOP error data are also recorded for the transmitter pairs AD and BD as indicated in figures 4(b) and 4(c). Diurnal patterns of a different shape to those of LOP AB are recorded and there is more variability from day to day. The lower signal levels encountered on the high latitude path from North Dakota, (transmitter D), tend to increase the phase noise on this path. A noteworthy feature of the data reproduced in figures 4(a), 4(b) and 4(c) is the disturbance due to the large geomagnetic storm of 23rd September. Only a minor perturbation is apparent on the AB LOP, but major effects are evident on both the AD and BD LOPs. This suggests that large phase anomalies are present in the signal received from North Dakota at Farnborough during this period since transmitter D is common to both LOP AD and BD. The other two propagation paths are less severely affected.

The measured and predicted LOP errors are compared in figures 5(a), 5(b) and 5(c) for LOPs AB, AD and BD respectively. The predicted values are represented by the broken line and the full lines indicate the diurnal error changes for each individual day within the 15 day period for which the prediction applies. Typical 15 day periods representing summer and winter conditions are reproduced in the figures.

The varying degree of agreement between predicted and measured values can be ascertained by inspection of figure 5. This also illustrates the extent of the day to day and seasonal variability in the experimental results. Note, for example, the marked reduction in the variability during the summer months. The important features of this comparison can be summarized as follows:

(a) LOP AB

For this transmitter pair the predicted and measured values agree well during the day time but a systematic error of approximately 20 to 30 cels is present at night. Transient changes due to solar flares are evident in the experimental data during the period 1st-15th July, 1978. The day to day variability is greater in winter than in summer.

(b) LOP AD

There is good agreement between the measured and predicted errors from 00:00 to 12:00 UT during the winter. From 12:00 to midnight the errors are underestimated by the prediction program, sometimes by as much as 20 cels. The day to day variability is however very large at times and in particular during geomagnetic activity. In the summer months, the predicted errors are shifted by approximately 10 cels during the whole day with respect to the measured values, however, the day to day variability is less than that during winter.

(c) LOP BD

Large differences between predicted and measured LOP errors in the early morning period, 01:00 to 09:00 UT, when the prediction program overestimates the dawn diurnal phase change in magnitude and slope.

4.1.2. 13.6 kHz Results

The LOP error data measured at Farnborough for the 13.6 kHz frequency have been examined in a similar manner to those of the 10.2 kHz transmissions already described. The measured LOP errors are reproduced in figures 6(a), 6(b) and 6(c) for transmitter pairs AB, AD and BD respectively.

The diurnal patterns for LOP AB (figure 6(a)) at 13.6 kHz are similar to those at 10.2 kHz. There is however a systematic shift in the whole data set of approximately -40 cels with respect to the 10.2 kHz results. Similar observations can be made for LOP AD and LOP BD. A systematic shift of about -60 cels with respect to the 10.2 kHz results is present in the LOP AD data and the corresponding value for LOP BD is approximately -20 cels.

The comparison of the predicted and measured errors is represented by the typical examples of winter and summer conditions reproduced in figure 7. The important features of this comparison can be summarized as follows:

(a) LOP AB

The predicted LOP error at 13.6 kHz for LOP AB (figure 7(a)) differs from that at 10.2 kHz, principally in its position with respect to the measured LOP data. At 13.6 kHz large systematic errors (10 to 20 cels) occur between predicted and measured values during the day, the converse of the situation at 10.2 kHz. During the night period, differences are reduced to approximately 10 cels. The effects of solar flare activity produce similar results at 13.6 kHz as at 10.2 kHz for this LOP.

(b) LOP AD

Differences between measured and predicted errors for LOP AD at 13.6 kHz are presented in figure 7(b). Similar comparisons can be made between the predicted curves and measured data as were made previously for LOP AD at 10.2 kHz.

(c) LOP BD

Large differences between predicted and measured values occur during the dawn period for this LOP at 13.6 kHz (figure 7(c)) as for the 10.2 kHz transmission. However, at 13.6 kHz differences remain large (10 to 20 cels) during the period from dawn to 21:00 UT in the summer months.

4.2. The Leicester Monitor

The Leicester monitor differs from those at Farnborough and the Butt of Lewis in that LOP information is recorded at five minute intervals rather than hourly. The Leicester data is therefore particularly suitable for studies of short duration changes in LOP. The LOP errors recorded at Leicester exhibit a similar behaviour to those for Farnborough. The differences between the predicted and measured LOP errors are also very similar for the two stations. In order to illustrate the short period changes in LOP error, typical examples of Leicester data are presented in figures 8, 9 and 10.

Further discussion of the Leicester data and its comparison with the predicted errors is not presented as the conclusions drawn from the Farnborough results apply equally to the Leicester site.

4.3. The Butt of Lewis Monitor

Reliable data from the Butt of Lewis monitor has been obtained for the summer months of 1979 only. The LOP errors for transmitter pairs AB, AD and BD are compared with the predicted values for this receiver location in figures 11 and 12. Excellent agreement between the predicted and measured errors are obtained for LOP AB (figure 11(a)), at 10.2 kHz, and there is little day to day variability in the measured values. Very good agreement is also obtained for LOP AD (figure 11(b)) and LOP BD (figure 11(c)) at 10.2 kHz. However, appreciable day to day variability is sometimes present in the data.

The comparisons at 13.6 kHz are illustrated in figures 12(a), 12(b) and 12(c). The agreement between predicted and measured errors is not so close at this frequency as for the 10.2 kHz data. The conclusions derived from the Farnborough 13.6 kHz data could equally well apply at the Butt of Lewis at this frequency.

An important feature of the Butt of Lewis results is the very close agreement, at 10.2 kHz, between the errors predicted by the ONSCD program and those measured experimentally.

5. COMPARISON OF ERRORS MEASURED AT THE VARIOUS RECEIVING STATIONS

The comparisons of predicted and experimental LOP data presented in section 4 suggest that the performance of the Omega system differs even within the confines of the United Kingdom. An attempt was therefore made to compare the LOP errors measured at the various monitor sites. The object of this study was to try to quantify the similarity (or difference) in the LOP errors measured at the various sites. This proved to be rather more difficult than expected since the LOP can move to either side of its true position. The possible cases are illustrated in Table 2.

Six different possible situations can give rise to the same numerical value for the difference in errors, although some can be distinguished from sign changes. Care must therefore be exercised in interpreting the LOP error differences for the same LOP measured at the two monitoring sites, $(X-Y)$, where X and Y are the errors measured at each of the two sites respectively.

The difference in errors for LOP AB measured at Farnborough and Leicester are shown in figure 13(a) There is a small systematic bias of about 5 cels and there is also some evidence of a very small diurnal variation. Otherwise the difference is random. The error differences for LOP AD is shown in figure 13(b). Again there is a systematic difference of about 5 cels, but in the opposite sense to the LOP AB shift. There is little evidence of a diurnal pattern and the changes appear to be quite random. The BD comparison presented in figure 13(c) indicates no systematic shift and the differences in errors for the LOP at the two monitoring stations are quite random.

These comparisons indicate that small constant differences are present in LOP AB and AD measured at Farnborough and Leicester. The differences in LOP BD error are quite random. There is some evidence of a small diurnal variation in the LOP errors measured at the two locations.

A comparison of the Farnborough and Butt of Lewis LOP errors has also been undertaken. The comparison of the LOP AB errors at the two locations is reproduced in figure 14(a). The strong diurnal variation in the error differences is immediately apparent. The night time values exhibit the greatest differences, (~ 25 cels) with much smaller differences, (~ 10 cels), present during the day time. The LOP AD error comparison, (see figure 14(b)), also exhibits a strong diurnal influence. The errors at the two locations are rather simila. at night as indicated by the near zero error difference values. During day time, large error differences

occur and the errors observed in this LOP at the two receiving locations are quite different.

Very large diurnal changes are evident in the comparison of LOP BD presented in figure 14(c). The night time difference is approximately -20 cels and the day time value approximately +10 cels. This represents a diurnal change in difference of about 30 cels and quantifies the very marked differences in the LOP errors recorded at Farnborough and the Butt of Lewis.

The comparisons presented serve to illustrate the very real differences in Omega performance at the observing sites considered, all of which are within the United Kingdom. They also serve to highlight the difficulties in applying existing prediction techniques which are not capable of resolving these differences in performance. If adequate predictions are to be made for systems such as Differential Omega, then these differences, which exist over small spatial distances, must be incorporated into the prediction scheme.

6. CONCLUSION

A detailed analysis of the experimentally determined performance of the Omega system at three locations within the United Kingdom has been presented. The predicted performance of the system has been obtained using the ONSOD program. The validity of these LOP corrections for the three locations has been assessed and their diurnal and seasonal variations quantified. The corrections are more accurate for the Butt of Lewis data than for the results obtained at the other two receiving sites.

The variation in performance of Omega at the three receiver sites has been studied by examining the difference in the measured error of the same LOP at the three locations. The performance at Farnborough and Leicester is very similar but there is a marked difference between the Butt of Lewis errors and those which occur at the two southerly receivers. In particular, there is a diurnal variation in the error difference. This feature has an important implication for the use of Differential Omega in the United Kingdom.

This study has shown that VLF propagation conditions can be accurately predicted by the ONSOD method for a location such as the Butt of Lewis. At other locations its accuracy is reduced by an amount that varies with time of day and season. This variation in accuracy over short distances of a few hundred kilometres suggests that other factors, such as ground conductivity changes, may influence the system performance. These anomalies will be the subject of further investigation as will be the variation in position fixes determined from the LOP information.

ACKNOWLEDGEMENT

The support of the United Kingdom Admiralty Compass Observatory is greatfully acknowledged.

REFERENCES

Morris, P.B. and Milton, Y.C., OMEGA Propagation Corrections: Background and Computational Algorithm, ONSOD Report No. ONSOD-01-74, December 1974.

Swanson, E.R. and Bradford, W.R., Diurnal Phase Variation at 10.2 kHz, NELC Technical Report 1781, 1971.

TABLE I

Omega Lane Widths at the UK Monitoring Sites

Monitor Site	LOP	Lane Width (km)	
		10.2kHz	13.6kHz
Leicester	AB	14.7	11.0
	AD	24.0	18.0
	BD	17.6	13.2
Farnborough	AB	14.7	11.0
	AD	25.0	18.6
	BD	17.6	13.2
Butt of Lewis	AB	15.4	11.6
	AD	19.2	14.4
	BD	17.5	13.2

TABLE II

The Possible Variations of LOP Errors and the Resulting Differences

LOP Error Location X	LOP Error Location Y	Difference in errors (X-Y)
lop, +5, true position X	+1 Y	+4
+1	+5	-4
-5	-1	-4
-1	-5	+4
+3	-1	+4
-1	+3	-4

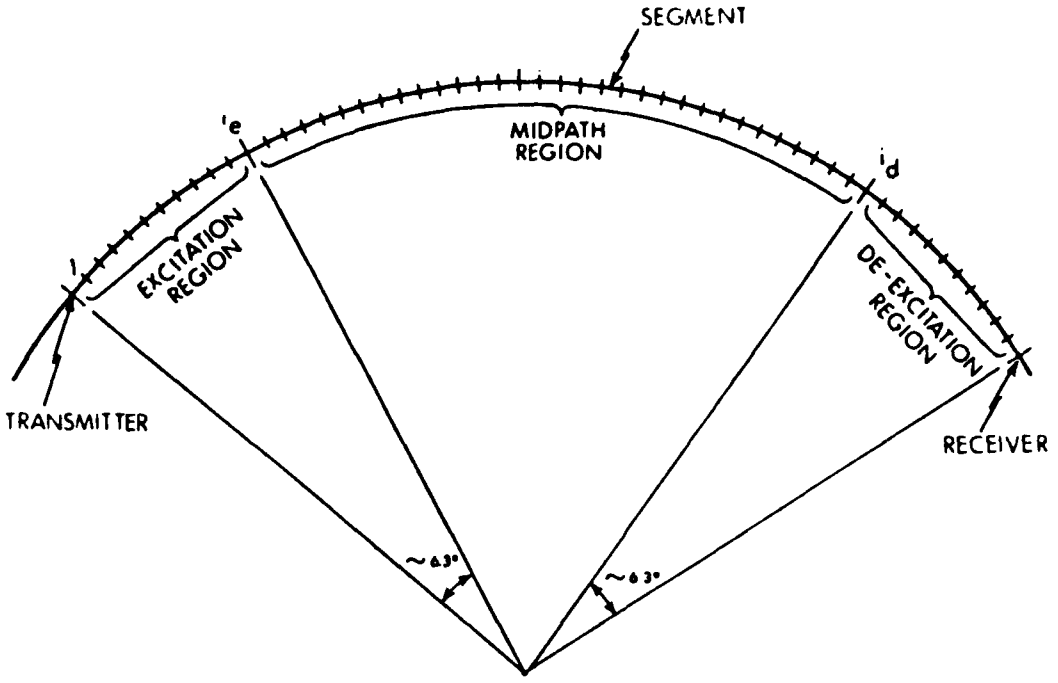


Fig.1 Schematic of the Omega signal path segmentation (Morris & Milton, 1974)

OMEGA PROPAGATION CORRECTIONS FOR 10.2 KHZ

LOCATION 16.0 N 40.0 W
STATION A NORWAY

DATE	GMT																							
	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1–15 JAN	−71	−71	−71	−71	−71	−71	−71	−71	−71	−14	−5	−2	−1	0	1	0	−3	−12	−24	−40	−61	−71	−71	−71	−71
16–31 JAN	−71	−71	−71	−71	−71	−71	−71	−71	−68	−11	−6	−3	−1	0	1	0	−1	−8	−20	−36	−57	−71	−71	−71	−71
1–14 FEB	−71	−71	−71	−71	−71	−71	−71	−71	−59	−11	−8	−4	−2	0	0	0	−2	−5	−16	−31	−52	−71	−71	−71	−71
15–28 FEB	−71	−71	−71	−71	−71	−71	−71	−67	−44	−10	−6	−3	0	1	2	1	0	−3	−9	−24	−45	−71	−71	−71	−71
1–15 MAR	−71	−71	−71	−71	−71	−71	−70	−59	−32	−8	−4	−1	1	3	3	3	1	−2	−5	−17	−39	−70	−71	−71	−71
16–31 MAR	−71	−71	−71	−71	−71	−71	−64	−48	−18	−6	−3	1	3	5	5	4	2	0	−4	−9	−29	−67	−71	−71	−71
1–15 APR	−71	−71	−71	−71	−71	−66	−55	−37	−10	−5	−1	2	5	6	6	6	4	1	−3	−7	−16	−60	−71	−71	−71
16–30 APR	−71	−71	−71	−71	−68	−59	−47	−28	−7	−3	1	4	6	7	8	7	5	2	−2	−6	−10	−41	−71	−71	−71
1–15 MAY	−71	−71	−71	−69	−62	−53	−41	−22	−6	−2	2	5	7	8	8	8	6	3	−1	−5	−9	−16	−70	−71	−71
16–31 MAY	−71	−71	−69	−65	−58	−49	−37	−18	−5	−1	3	6	8	9	9	8	6	4	0	−4	−8	−12	−62	−70	−71
1–15 JUN	−69	−69	−66	−62	−56	−47	−34	−16	−4	−1	3	6	8	9	9	9	7	4	1	−3	−7	−11	−50	−67	−69
16–30 JUN	−68	−68	−66	−61	−55	−47	−34	−16	−5	−1	3	6	8	9	9	9	7	4	1	−3	−7	−10	−45	−65	−68
1–15 JUL	−69	−69	−66	−61	−55	−45	−31	−12	2	4	6	8	9	10	10	10	9	7	5	3	0	−2	−43	−66	−69
16–31 JUL	−71	−71	−69	−65	−58	−48	−35	−15	1	4	6	8	9	10	10	10	9	7	5	3	0	−2	−57	−70	−71
1–15 AUG	−71	−71	−71	−69	−63	−53	−40	−20	1	3	5	7	9	9	10	9	8	6	4	2	0	−5	−70	−71	−71
16–31 AUG	−71	−71	−71	−71	−68	−59	−45	−25	0	2	5	7	8	9	9	9	7	6	4	1	−1	−34	−71	−71	−71
1–15 SEP	−71	−71	−71	−71	−71	−65	−52	−32	−1	2	4	6	7	8	8	8	7	5	3	0	−13	−65	−71	−71	−71
16–30 SEP	−71	−71	−71	−71	−71	−70	−59	−39	−3	1	3	5	7	7	7	7	6	4	1	−6	−33	−71	−71	−71	−71
1–15 OCT	−71	−71	−71	−71	−71	−71	−67	−48	−7	0	3	4	6	6	6	6	4	3	−2	−19	−48	−71	−71	−71	−71
16–31 OCT	−71	−71	−71	−71	−71	−71	−71	−59	−15	0	2	4	5	5	5	5	3	1	−11	−30	−59	−71	−71	−71	−71
1–15 NOV	−71	−71	−71	−71	−71	−71	−71	−69	−33	−1	1	3	4	4	4	4	2	−5	−19	−38	−65	−71	−71	−71	−71
16–30 NOV	−71	−71	−71	−71	−71	−71	−71	−71	−57	−2	0	2	3	4	4	3	0	−10	−24	−42	−67	−71	−71	−71	−71
1–15 DEC	−71	−71	−71	−71	−71	−71	−71	−71	−69	−3	−1	1	2	3	3	2	−2	−12	−25	−43	−67	−71	−71	−71	−71
16–31 DEC	−71	−71	−71	−71	−71	−71	−71	−71	−71	−7	−2	−1	1	2	2	1	−3	−13	−26	−43	−65	−71	−71	−71	−71

Fig.2 A typical Omega propagation correction table

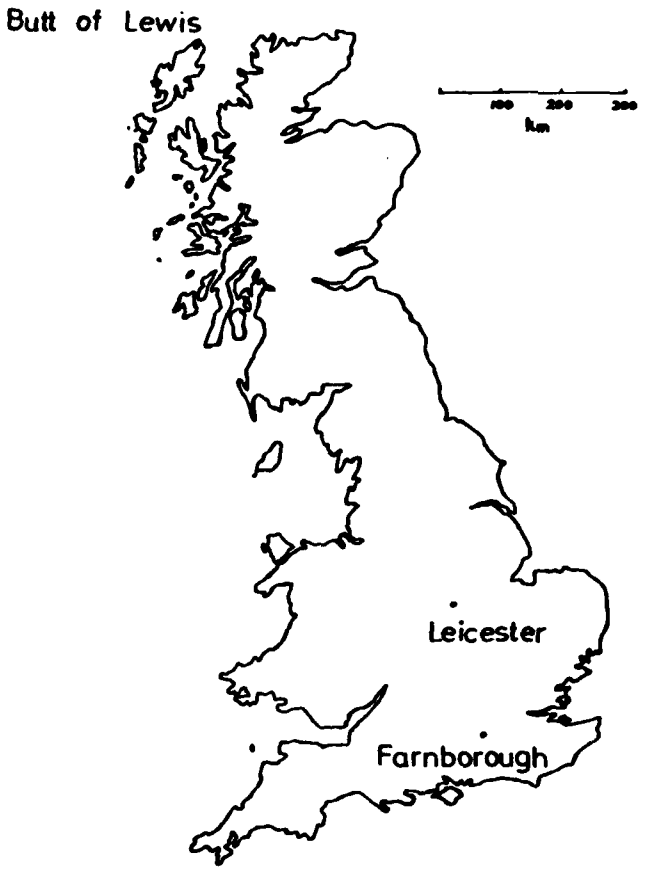


Fig.3 The location of the three UK monitoring sites

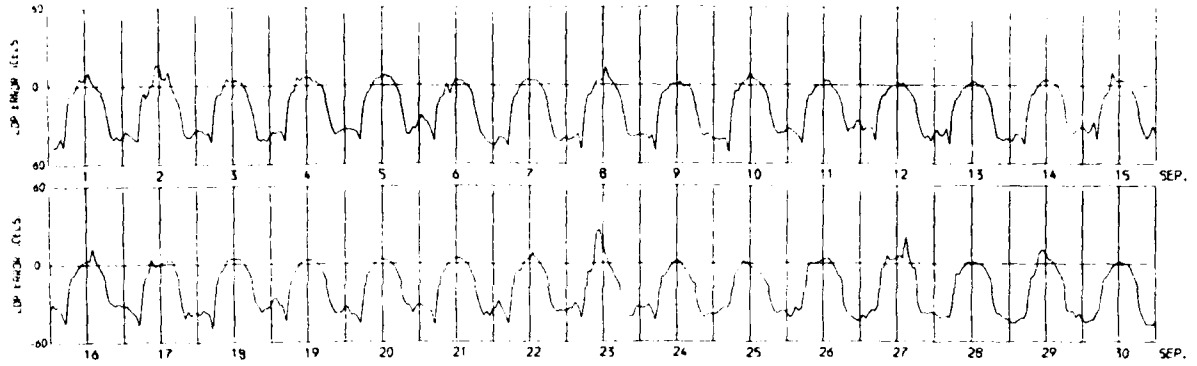


Fig.4(a) Farnborough LOP error data for LOP AB at 10.2 kHz, recorded during September 1978

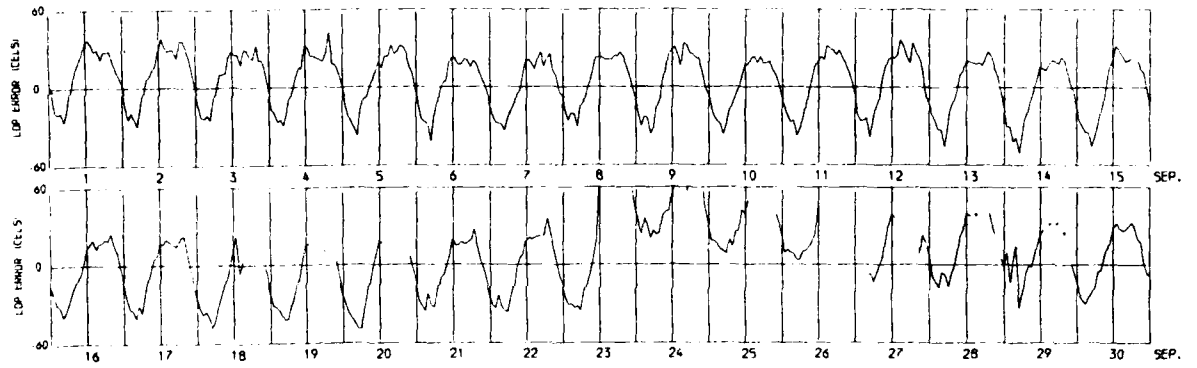


Fig.4(b) Farnborough LOP error data for LOP AD at 10.2 kHz, recorded during September 1978

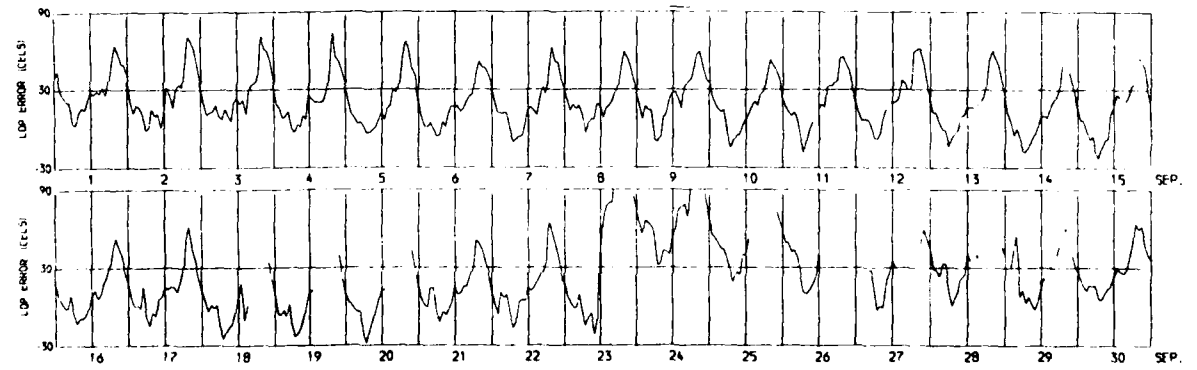


Fig.4(c) Farnborough LOP error data for LOP BD at 10.2 kHz, recorded during September 1978

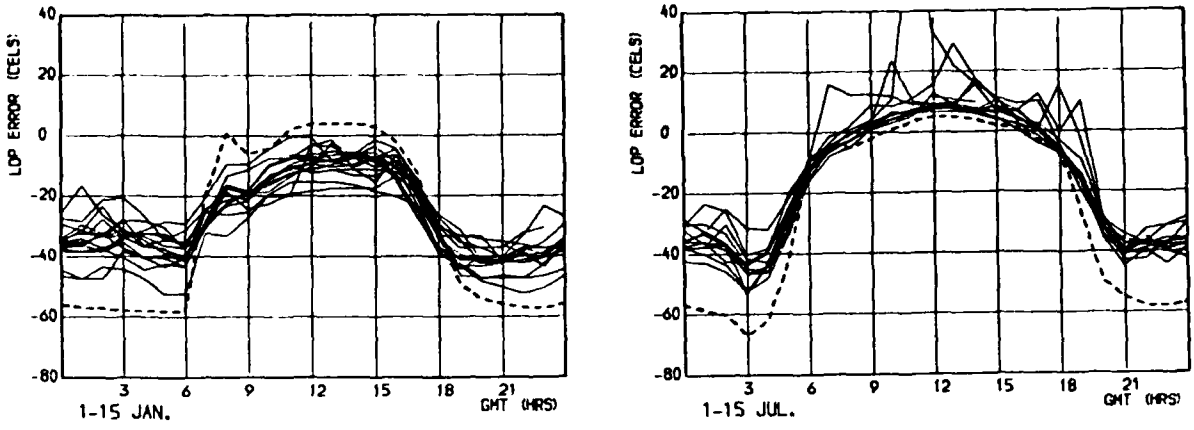


Fig.5(a) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP AB, 10.2 kHz, winter and summer 1978

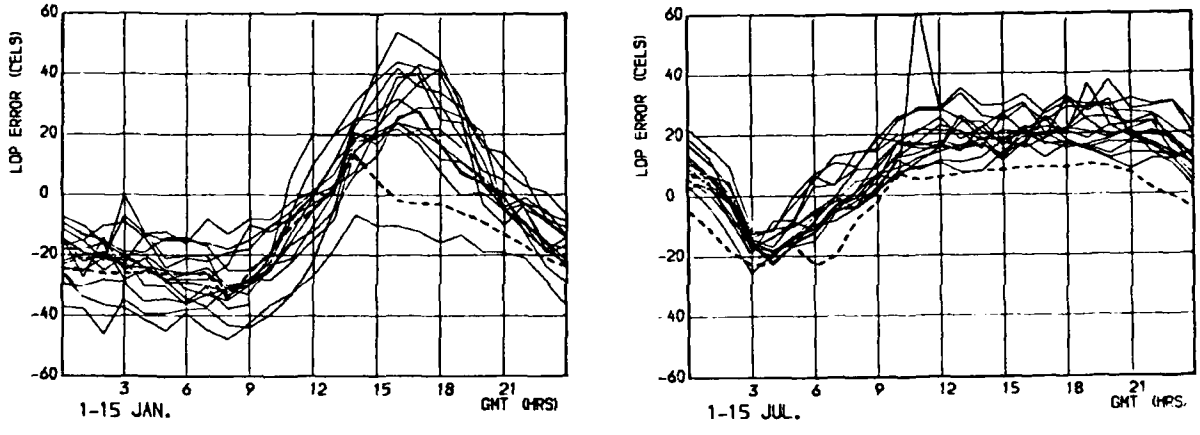


Fig.5(b) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP AD, 10.2 kHz, winter and summer 1978

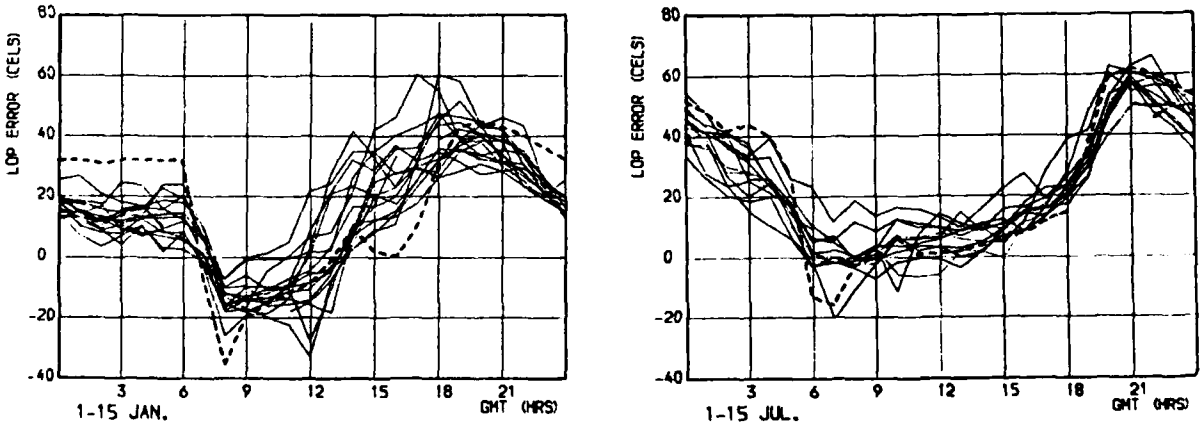


Fig.5(c) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP BD, 10.2 kHz, winter 1979 and summer 1978

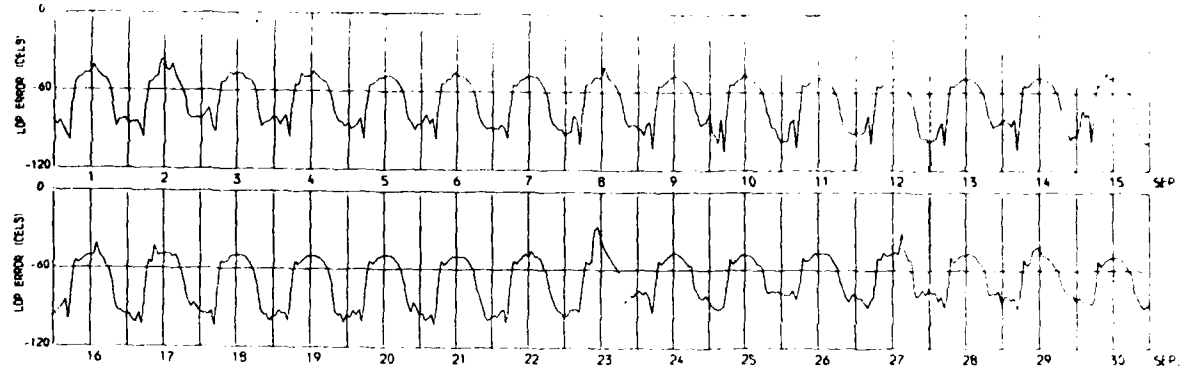


Fig.6(a) Farnborough LOP error data for LOP AB at 13.6 kHz, recorded during September 1978

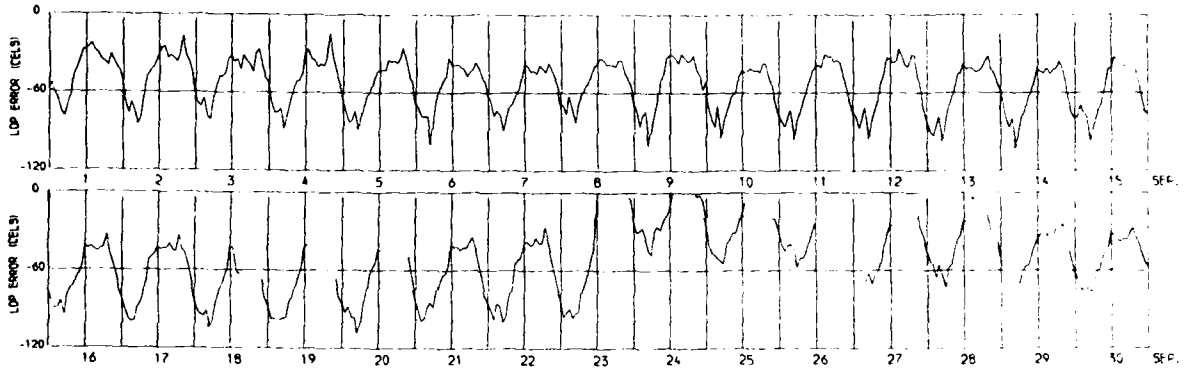


Fig.6(b) Farnborough LOP error data for LOP AD at 13.6 kHz, recorded during September 1978

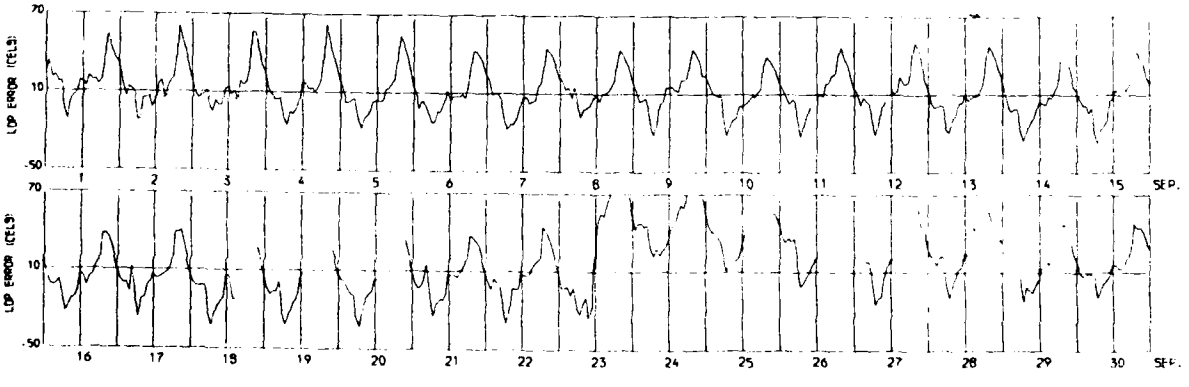


Fig.6(c) Farnborough LOP error data for LOP BD at 13.6 kHz, recorded during September 1978

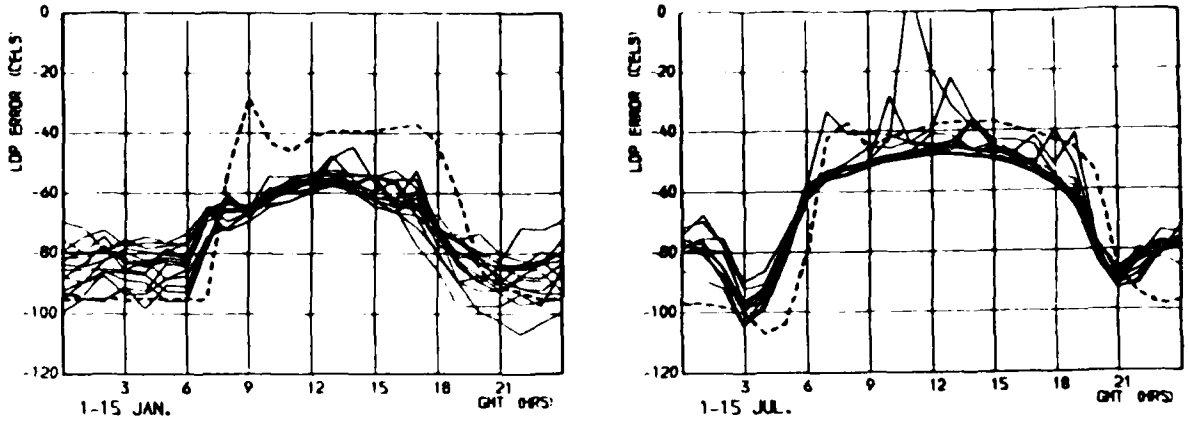


Fig.7(a) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP AB, 13.6 kHz, winter and summer 1978

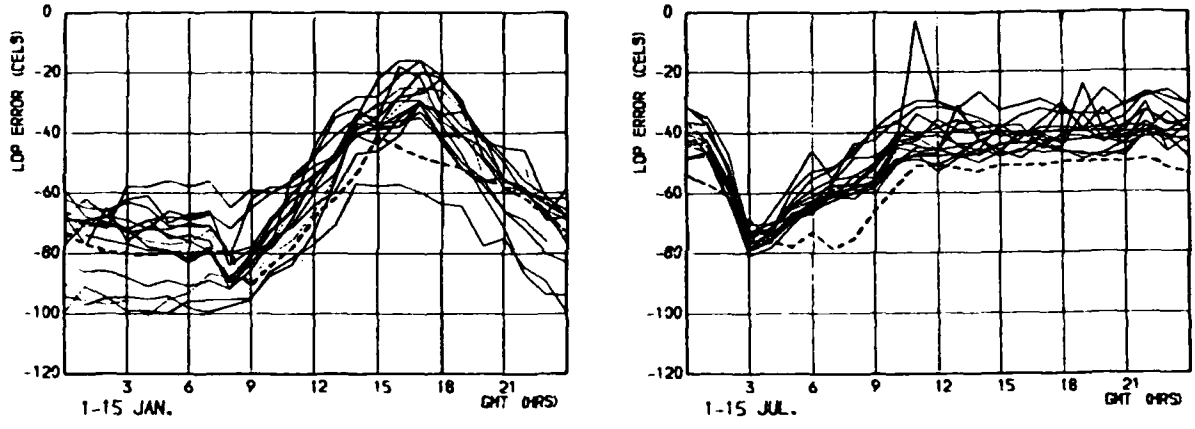


Fig.7(b) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP AD, 13.6 kHz, winter and summer 1978

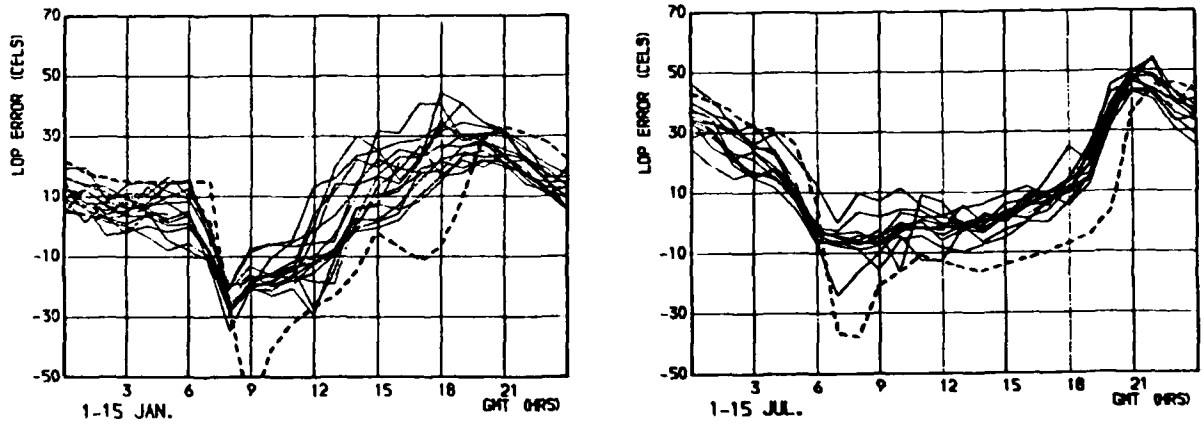


Fig.7(c) Comparisons of half monthly LOP errors measured at Farnborough (solid lines) with the ONSOD prediction (broken line). Data for LOP BD, 13.6 kHz, winter 1979 and summer 1978

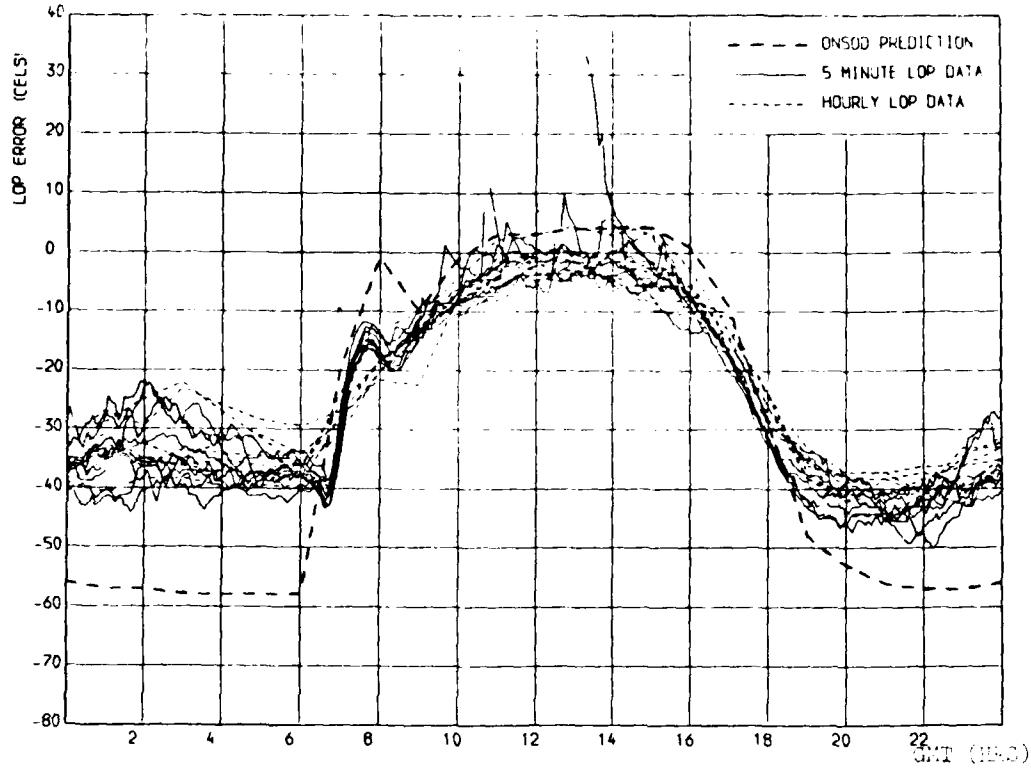


Fig.8(a) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP AB, 10.2 kHz, 16- 31 January 1979

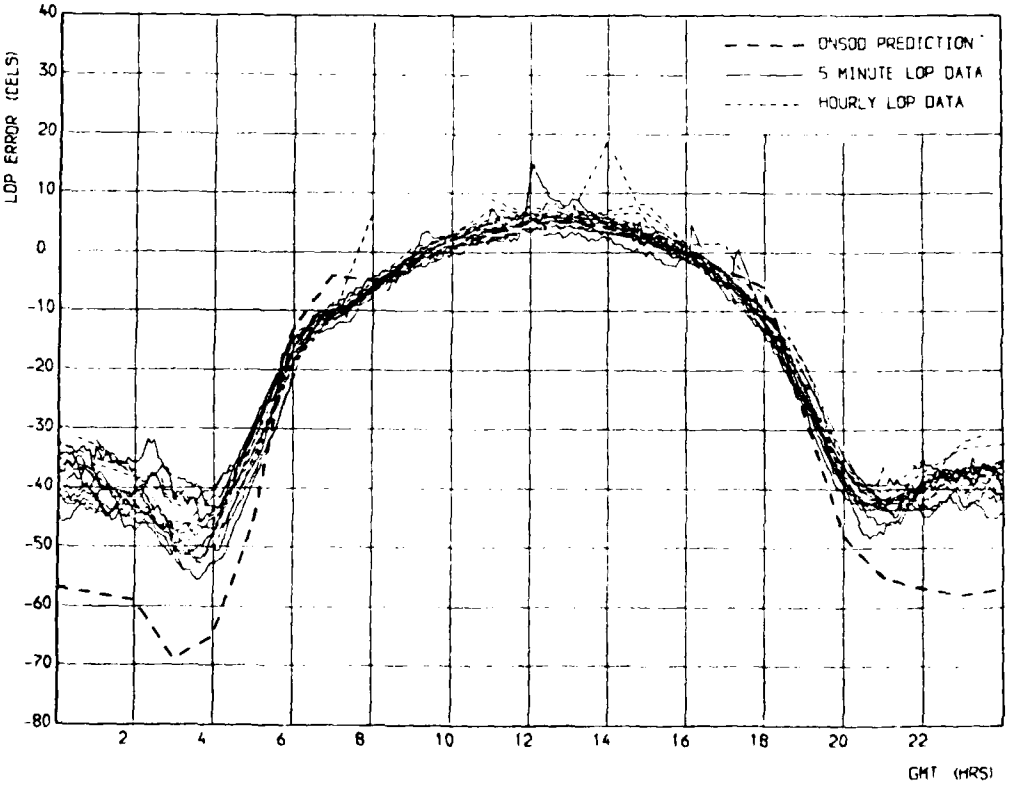


Fig.8(b) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP AB, 10.2 kHz, 16 31 July 1979

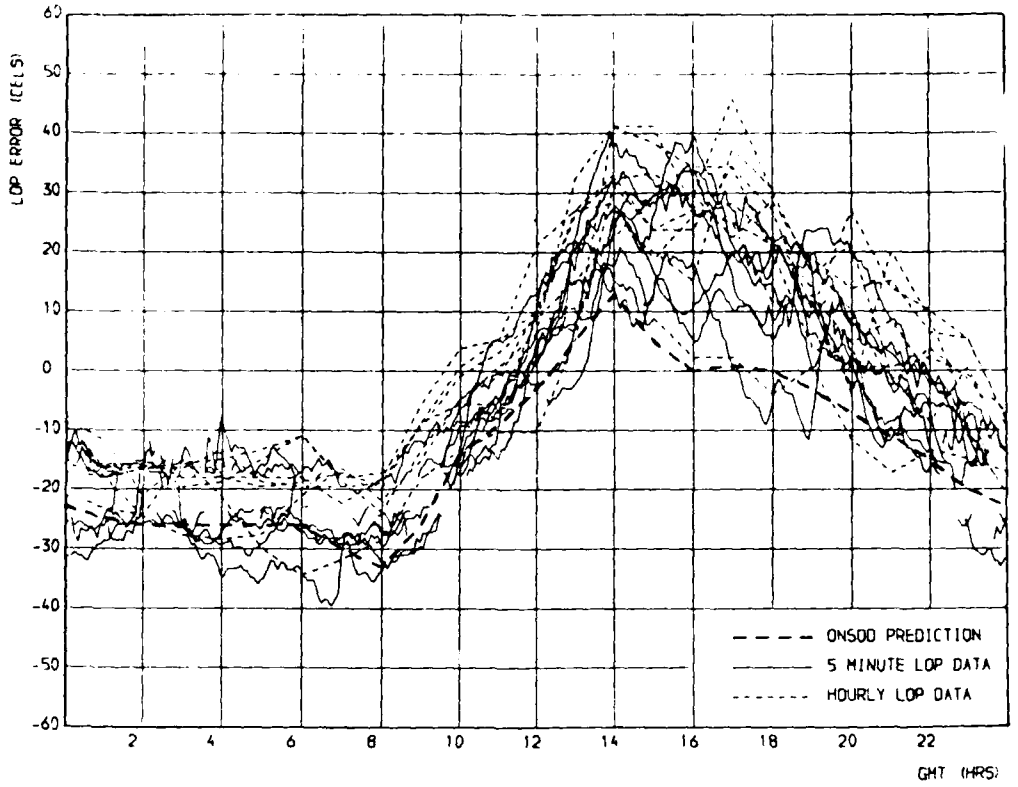


Fig.9(a) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP AD, 10.2 kHz, 16–31 January 1979

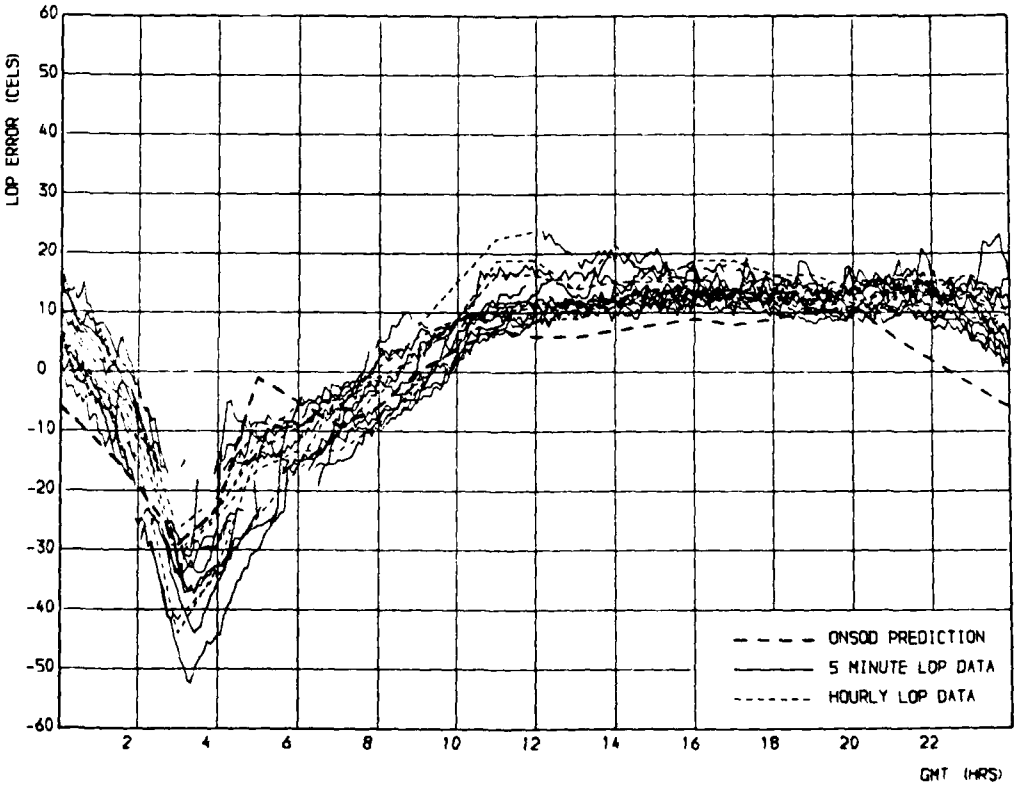


Fig.9(b) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP AD, 10.2 kHz, 16–31 July 1979

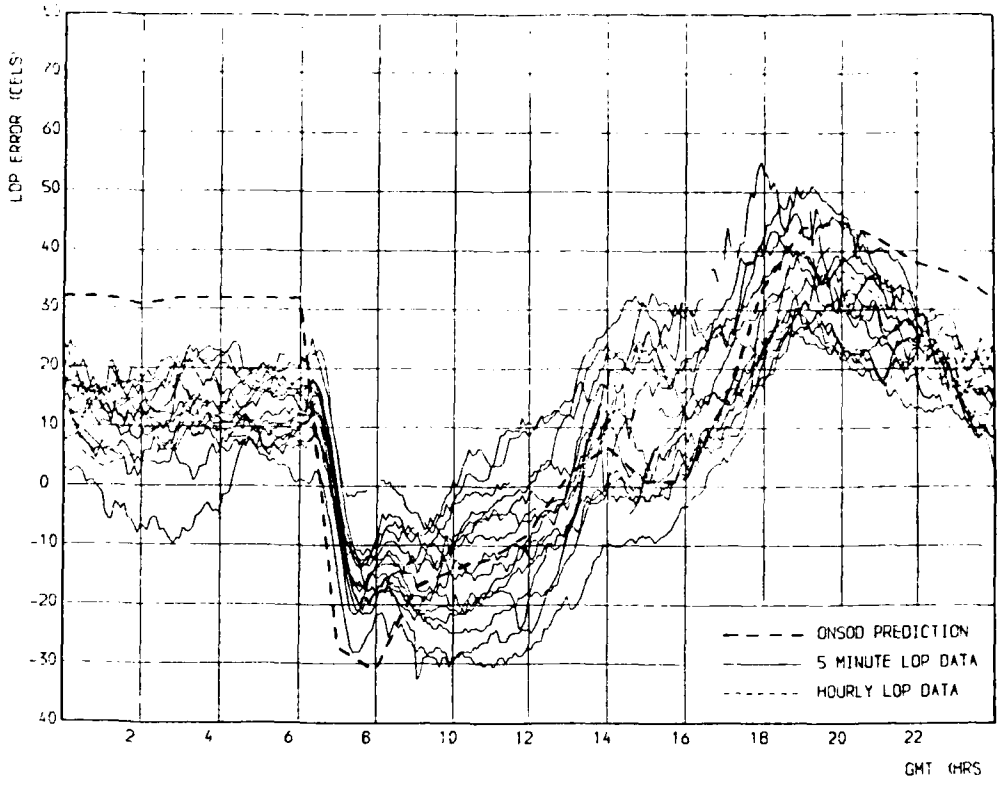


Fig.10(a) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP BD, 10.2 kHz, 1 15 December 1979

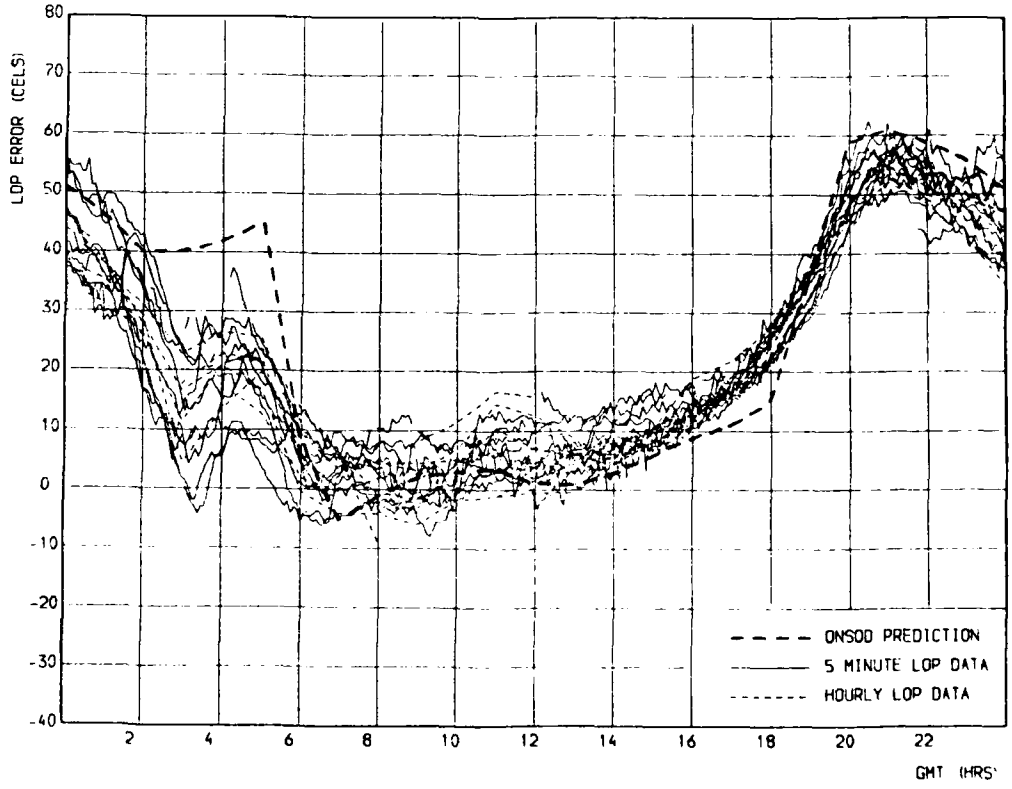


Fig.10(b) Comparison of half monthly LOP errors measured at Leicester with the ONSOD prediction. Data for LOP BD, 10.2 kHz, 16 31 July 1979

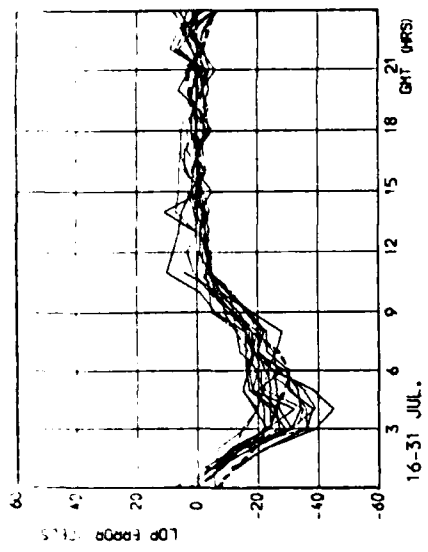


Fig.11(b) LOP AD

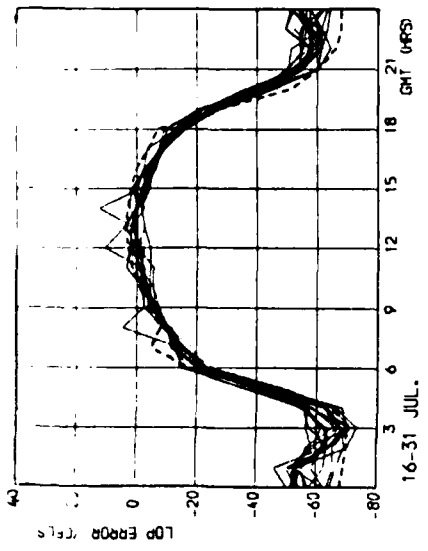


Fig.11(a) LOP AB

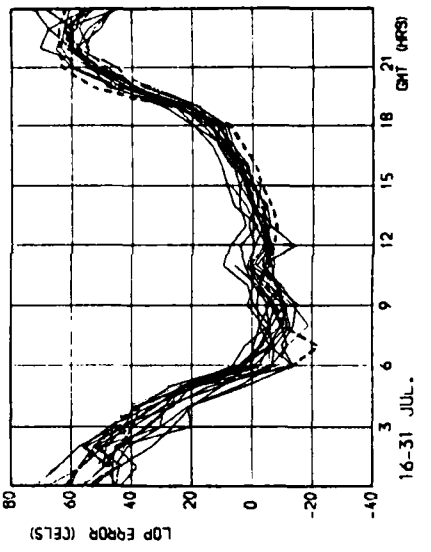


Fig.11(c) LOP BD

Fig.11 Comparisons of half monthly LOP errors measured at the Butt of Lewis (solid lines) with the ONSOD prediction (broken line). 10.2 kHz data

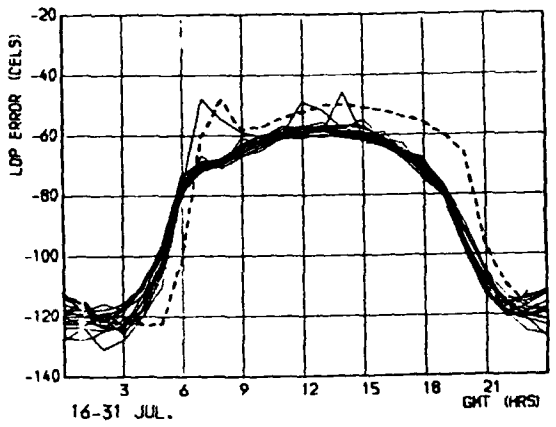


Fig.12(a) LOP AB

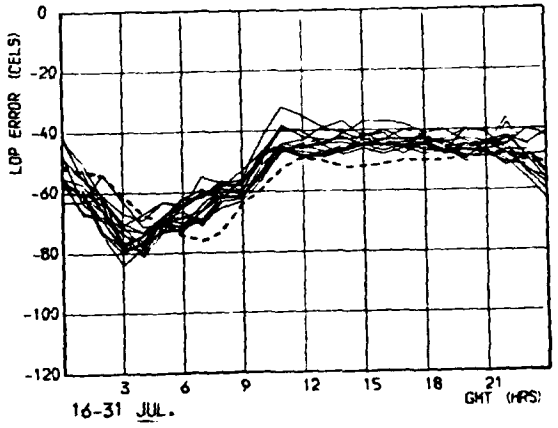


Fig.12(b) LOP AD

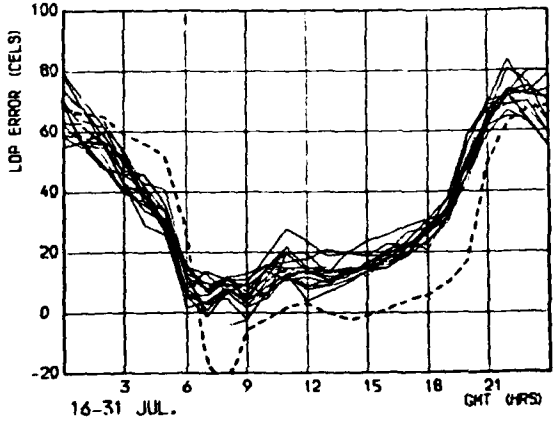


Fig.12(c) LOP BD

Fig.12 Comparisons of half monthly LOP errors measured at the Butt of Lewis (solid lines) with the ONSOD prediction (broken line). 13.6 kHz data

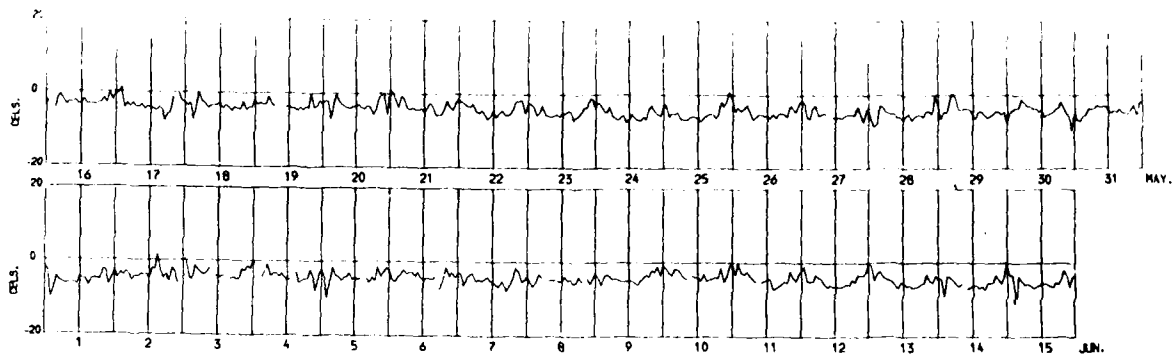


Fig.13(a) Differences in LOP error between Leicester and Farnborough. Data for LOP AB at 10.2 kHz recorded from 16th May to 15th June 1979

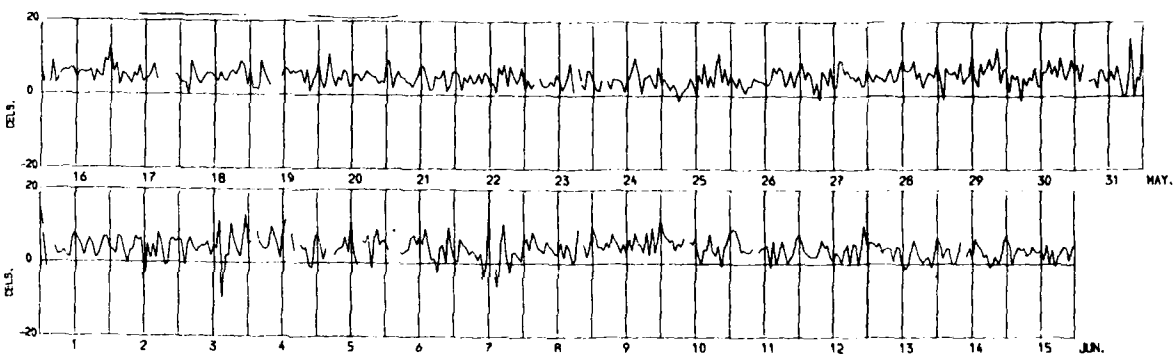


Fig.13(b) Differences in LOP error between Leicester and Farnborough. Data for LOP AD at 10.2 kHz recorded from 16th May to 15th June 1979

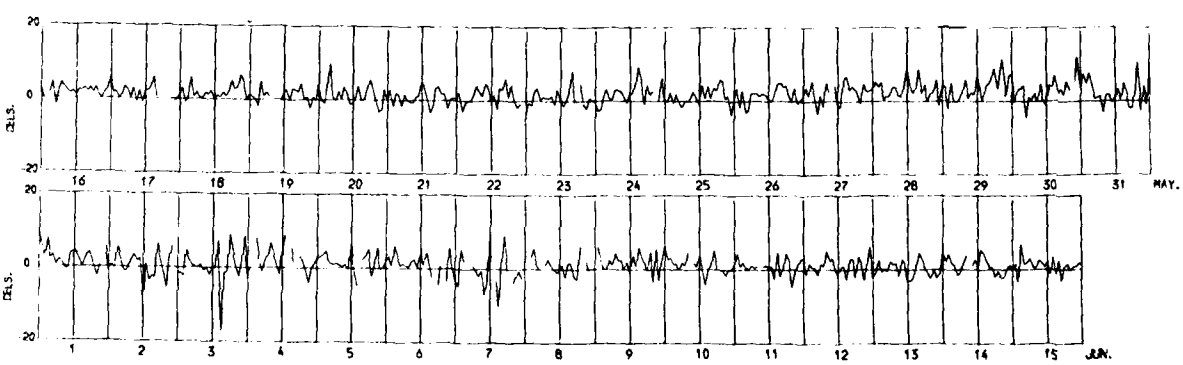


Fig.13(c) Differences in LOP error between Leicester and Farnborough. Data for LOP BD at 10.2 kHz recorded from 16th May to 15th June 1979

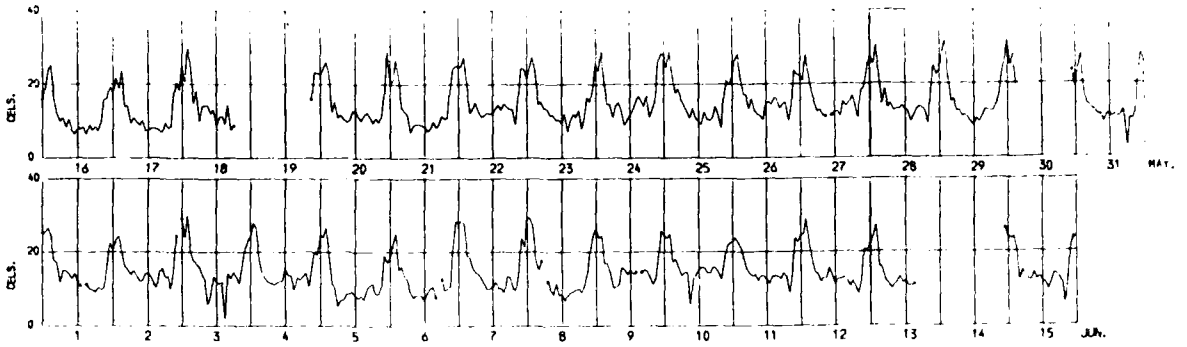


Fig.14(a) Differences in LOP error between Farnborough and the Butt of Lewis.
Data for LOP AB at 10.2 kHz recorded from 16th May to 15th June 1979

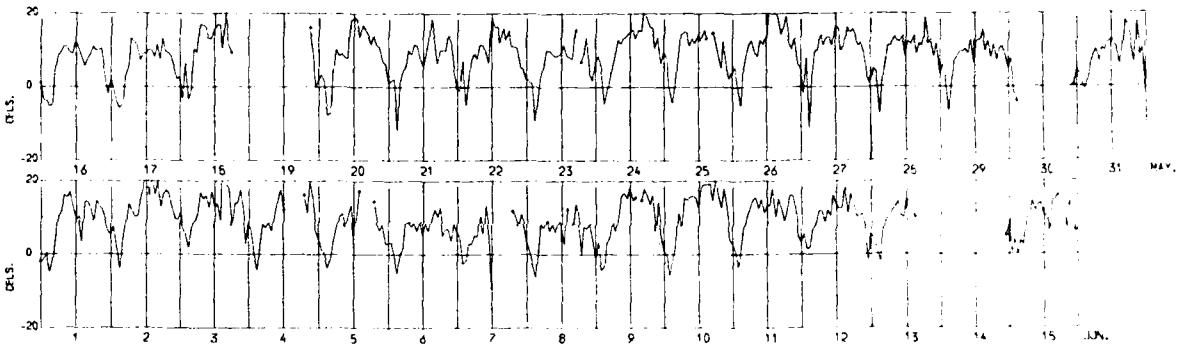


Fig.14(b) Differences in LOP error between Farnborough and the Butt of Lewis.
Data for LOP AD at 10.2 kHz recorded from 16th May to 15th June 1979

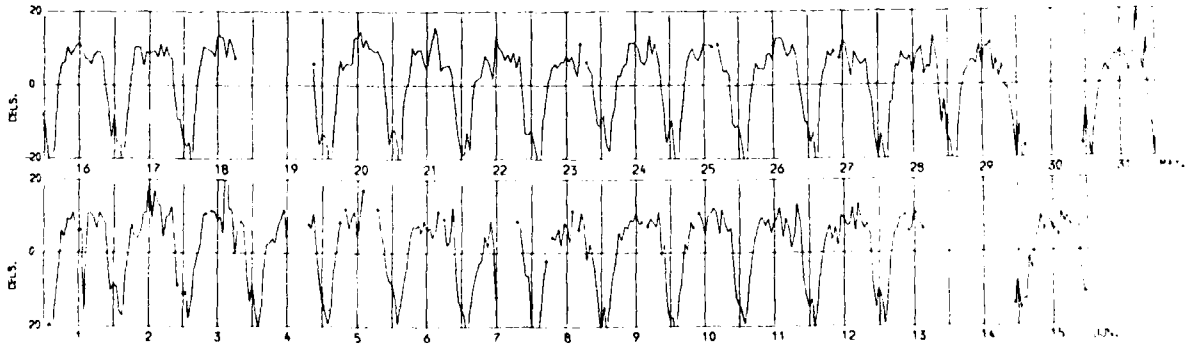


Fig.14(c) Differences in LOP error between Farnborough and the Butt of Lewis.
Data for LOP BD at 10.2 kHz recorded from 16th May to 15th June 1979

SOME CONSIDERATIONS UPON MODELS FOR IONOSPHERIC RADIO COMMUNICATIONS FORECASTING

P. DOMINICI and B. ZOLESI
Ionospheric Department
Istituto Nazionale di Geofisica
Rome, Italy

SUMMARY

The ionospheric radio communications exhibit, along with consistent advantages, the well known disadvantage that the essential parameters show large variations, which must be forecast. By appropriate models, some of these variations are reconductible to known behaviors of solar and geophysical quantities and are therefore foreseeable with satisfactory statistical accuracy; on the contrary, other variations are completely casual. There is a point of view, that of geophysicists, according to which the problem of the ionospheric modelling for the forecasts is a normal problem of scientific knowledge, with the assumption that the last aim, that is to know completely the ionospheric phenomena, is certainly reachable; the actual trend is to elaborate mathematical models of the ionosphere, which are used for forecasting in near real time, by inserting suitable recent data. Other points of view, belonging to people concretely interested in ionospheric radio communications, are based upon the concept of a forecast suitable for the management of the links; according to what is a "suitable" forecast, there is to distinguish the point of view of people planning the links from the point of view of people operating the links. It is a common opinion among planners that the actual forecasts are sufficiently accurate in order to the allowance of the radio frequencies. On the contrary, it is a common opinion among operators that the actual forecasts are intrinsically unable to make a link sure with "critical" conditions of some parameters and above all in presence of casual variations and/or perturbations. A comparison between measured and calculated data is made, which seems to justify these contrasting opinions. A brief discussion of the above points of view is made, and the conclusions are: (i) as to the ionospheric physics, as well as to the planning of ionospheric radio communications, the studies upon ionospheric adaptive models physically based (for instance, the recent models based on 3 overlapped Chapman layers) are to be encouraged; (ii) as to the management of the links, it seems to be convenient to prefer techniques of ionospheric monitoring in real time instead of ionospheric forecasting in near real time.

1. PREFACE

As in all the geophysical sciences, ionospheric physics (including its applications) uses models to support and visualize its theories.

It is well known that the development of ionospheric physics has its roots in the development of long distance radio communications. The way in which an ionospheric physicist considers ionospheric models is different from the way in which a person involved in ionospheric radio communications considers them.

From a radio engineer's point of view, ionospheric models are useful in schematizing the properties of the transmission medium between the antennas of an ionospheric radio link. As to the ability to represent temporal behaviors of such properties, it has to be noted the necessity one has of forecasting the ionospheric characteristcs of reflectivity and absorptivity influencing the operative parameters of HF radio links (optimum working frequency, power, types of antenna, etc.); these previsions are for future forecasting, even at long time, for the radio links planning and for radiofrequencies allowance, whereas they are for the immediate future (the closest possible to the present) for the management of the radio links.

In the geophysical field, the aim of the models (which assume heuristic and symplifying functions) is the understanding of ionospheric phenomena, which is assumed fully attainable and not conditioned by questions relating to immediacy or not.

In the development of ionospheric models the applicative view has up to now had the greatest relevance, with consequential accentuation of the statistical and empirical aspects above physical ones. This can be seen in the current definition of the ionosphere which, referring normally to the propagation of radio waves, makes the assumption of the "ionosphere" in a limited way, as an "electronosphere", and the assumption of the electron density as its representative quantity. Moreover, because the synoptic observations on a planetary scale of the electron density in the atmosphere are done exclusively through vertical inci-

dence soundings with HF radio waves, the experimental base on which the ionospheric models are built is restricted to what is observable by such radio soundings.

From the examination of ionograms taken over a period of almost 50 years by a net of stations covering the greatest part of the Earth's surface, a standard model was developed and used for studies of both ionospheric physics and radio propagation. In table I we can see the principal features of this reference ionosphere (standard ionosphere from ground radio sounding).

As it is well known, the model extends itself (Figure 1) to the absolute maximum of the electron density, that is, it includes only the so-called bottomside ionosphere. To this absolute maximum and to lower relative maximums a corresponding number of ionospheric zones are normally associated: the region F (between 200 and 400 km altitude) and the region E (approximately 100 km altitude) are directly observable on the ionograms, while the region D (approx. 70 km altitude) is observable indirectly through its absorption property; during daylight time in summer months the F region frequently appears divided into two layers, F1 and F2 (the highest). Besides these regular regions there is occasionally present at the E region levels, a thin and strongly ionized layer, called "sporadic E layer". Let us also remember that the maximum electron density of an ionospheric region is derived in a very simple manner from the so-called critical frequency of the region, which is the maximum frequency of the radio waves reflected under vertical incidence from the region, relative to the ordinary ray between the two rays to the birefringence induced by the terrestrial magnetic field.

The behavior of the components of such a model is individuated by the temporal and geographical behaviors of some numerical data taken from the ionograms according to detailed conventions and rules (standard ionospheric characteristics). Schematically, for the regular regions E and F (this last one eventually split in the F1 and F2 layers) and for the sporadic E layer, it is a matter of the critical frequencies of the ordinary ray (symbols: foE, foF1, foF2, foEs), depending on respective maximum electron densities, and of the virtual heights of reflection (h'E, h'F, h'F2, h'Es), deduced from the echo delay and referring to the respective altitudes of maximum electron density; for the region D it is a matter of the minimum frequency (fmin) of the radio waves reflected from either the E or F regions, that changes according to the variations of the maximum electron density of the region D itself.

Each of these characteristics is a function of the site, as to say, of the geographic latitude φ and longitude λ, and of the time t. In a given site, as a sounding station, only the dependence on time remains: this lead us to distinguish between local models, or unidimensional (the only variable is t), and global models, or 3-dimensional (in the 3 variables φ, λ, t).

2. LOCAL MODELS

These models were developed by analyzing historical series of data taken in every sounding station, with reference to the natural intervals of the variable t, that is in connection to the hours of the day, to the days of the year (grouped into the 12 Calendar months), and to the passing of the years. The discovery being quickly made of clear correlations, partly positive and partly negative, between the ionospheric characteristics and the position of the Sun, brought to place side by side the statistical model derived from the previously mentioned analysis, and an analytical model established from the theory of the atmospheric photoionization developed by S. Chapman (1931), explaining a large part of the observed statistical peculiarities.

According to such a theory, for instance, the analytical model for the ordinary critical frequency in quasi-stationary conditions consists of the equation:

$$1 \qquad fo = a(R) \cos^{0.25}\chi \qquad (\chi \leqslant 90°),$$

where χ is the zenith angle of the Sun. This last one implies the diurnal and annual variations by a well known astronomical formula in which appear, besides to the latitude of the site, the local hour angle (that is the local time) and the declination (that is the epoch of the (year) of the Sun. The quantity a(R) depends on the electron production rate, that is on the intensity of the ionizing solar photonic radiation, and therefore depends, as indicated, on the sunspot number R, assumed as an index of solar activity, which varies with a pseudo-period of about 11 years.

As an example of a local statistical model, we shall refer to the one deduced from the ionograms taken in Rome in the period 1948-1970. Limiting ourselves, for reasons of simplicity, to the critical frequencies, for the normal E region and for the F1 layer, the following formulae were obtained (90°):

$$2a \qquad \log foE = 0.520 + 0.0006\,R + 0.28 \log \cos\chi,$$

$$2b \qquad \log foF1 = 0.645 + 0.0008\,R + 0.20 \log \cos\chi.$$

As to the median values of the critical frequency of region F, night-time, and of the F2 layer, daytime, the statistical model is expressed graphically from the istograms of Figure 2, image of the correlation formula:

$$3 \qquad \overline{foF2} = A_H + B_H \bar{R} \ ,$$

relative to the single local hours H, where A_H, B_H are tabulated coefficients and $\bar{R}$ is the monthly mean value of the sunspot number R. The fluctuations of foF2 are much larger than those of foE (Figure 3) and foF1. As for the sporadic E layer, because in many cases a true critical frequency corresponding to a maximum echo delay doesn't exist, it wouldn't be appropriate to look for a correlation between foEs and other ionospheric or astrophysic quantities; so, the monthly median of the fraction of time that foEs exceeds some reference value (Figure 4) is enough significative.

A simple comparison between 2 , 3 and 1 shows that the E normal region and the F1 layer are, at least in first (but quite good) approximation, "Chapman layers" (that is verifying the previously mentioned theory of Chapman), which cannot be said for the whole region F.

3. GLOBAL MODELS

Local models are useful for theoretical studies and for radio purposes only in an area of a few hundred km; for larger areas, it becomes necessary to use global models.

It could seem that, being the latitude included in the Sun zenith angle χ (see 1), a kind of time shift in a local model could be enough for the transport in longitude; really, it is not so: the exponent of $\cos\chi$ in 1 varies with the longitude. A similar, but much more larger "longitude effect" can be found through the numerical coefficients of type 3 formulae if one tries to transport in longitude local statistical models for foF2. Therefore, the global models are preferably made linking opportunely together a sufficient number of statistical local models.

Until a few years ago, the global models were given by world maps of isolines of some ionospheric characteristics, in particular foF2 and MUF(3000)F2, the last one being the maximum usable frequency for a 1--hop link between two terminals at 3000 km, with ionospheric reflection on the F2 layer. These maps manually traced by some ingenious methods of interpolation and extrapolation between and from local models; let us remind the method by K. Rawer (1960), among the good qualities of which is the introduction of "magnetographic" coordinates in place of the geographic ones, in order to consider the longitude effect, due to the anisotropy of the terrestrial magnetic field (RAWER, 1963).

Presently, world maps of the most important ionospheric characteristics are traced by electronic computers. The method (numerical mapping) consists in the assumption that an ionospheric characteristic can be expressed through terms of a Fourier series in time domain, whose coefficients are expressed through terms of a serie expansion of the product of a function depending on the coordinates but not on the time, and therefore tabulated permanently, by a function of the coordinates and time, tabulated for actual ionospheric conditions (JONES and GALLET, 1962; JONES and others, 1966). As it is known, this procedure has been adopted for some years now, by various institutions interested in ionospheric radio communications, among which are the CCIR and the Istitute for Telecommunications Sciences (ITS) of the Environmental Science Services Adiministration (ESSA), Dept. of Commerce, USA.

As it was pointed out earlier, the numerical values that play in the models are generally monthly median values, that is probable at 50%. The passage to more significative values, for example to lower deciles, that is with a 90% probability, it is done by multiplying the medians by numerical coefficients, variable by site, hour, season and solar activity taken from the statistical distribution of the characteristics themselves; for example, from the appropriated tables (BARGHAUSEN and others, 1969; pg. 34), it results that for foF2 the lower decile is on average 85% of the median, varying between 70% and 90% of that median.

To indicate the ionospheric model situation at the beginning of the 1970's, we can observe the principal characteristics in table II of the ionospheric model adopted in the ITS-78 program of the ESSA for the operative parameters of HF ionospheric radio links still largely diffused today, both in its original form and in derived forms (BARGHAUSEN and others, 1969). It is an empirical model whose principal assumption -- which is typical of the simplicistic mentality that dominates models used exclusively for radio-technical purposes - is a two parabolic regions scheme, as indicated in Figure 5 A: region E, with constant maximum altitude and semi-thickness, and region F, with maximum altitude deduced from M(3000)F2 and semithickness derived from the minimum virtual height h'F of the region, corrected by the delay due to the E region. The foF2 and M(3000)F2 characteristics are derived from "numerical maps" issued 3 months in advance; the other characteristics are derived "permanent" world maps of isolines. Region D is considered as an absorbing layer of which the produced attenuation L_D is given.

4. RELIABILITY AND USEFULNESS OF THE MODELS

In Figure 6 monthly median values of the critical frequency foF2 at 00 and 12 local hours observed in recent years at Rome are compared with corresponding values, calculated by the local statistical model of Figure 2: there is an excellent accord. We are referring to the foF2 characteristic which is, as we earlier, the most indicative one: in terms of geophysical interest because it concerns the absolute maximum electron density of the whole ionosphere, in terms of radio propagation because it determines the upper frequency limit of radio waves reflected vertically by the ionosphere, and finally from a comparative point of view because it shows very large statistical fluctuations.

As to the global models, many comparison studies between observed and calculated values have been made, that generally also demonstrate an excellent accord. In Figure 7 we can see a comparison of this kind relative to the aforementioned ITS-78 Program.

The situation is very much different if from monthly median values one passes to instantaneous values: as to a local situation, we present some data in Figures 8 and 9, which are self-explanatory. Data derived from global models show a similar situation.

The inadequacy of all the models, when from a median case one goes to an instantaneous one is not surprising, considering the way in which the models are constructed. It should be noted that the way to give more or less importance to the differences between observed and calculated values, and therefore the way of use, or not use some methods to avoid such differences are notably different in the geophysical and radio fields.

5. GEOPHYSICAL CONSIDERATIONS: IMPORTANCE OF ANALYTICAL MODELS

From a geophysical point of view, the fact that an ionospheric model, normally dependable regarding median values, is unable to explain some variations, is not by itself a negative fact: the difference between actual and theoretical values can itself be indicative of a phenomenon. Consider for example, the fact that all the studies on ionospheric perturbations (SWF, SID, PCA, eclipse effects, etc.) are based on the divergence between instantaneous values and the so-called normal values, that is, in respect of median, mean or even instantaneous values deduced from an opportune local statistical model. So, models of this last type, i.e. with median values, that, as we will shortly discover, are not fully utilizable for radio-technical purposes, can find ample and advantageous uses in ionospheric physics. Neverthless, more important are the analytical models.

It is known that the only theoretical model that can be at present practically used is the Chapman model, in which the density profile is satisfactorily approximable as a parabolic layer; one of the consequences is the relation 1 between the critical frequency and the zenith angle of the Sun, that is adequately satisfied for the critical frequency foE_4 of the E region. Now, if one looks for the functional dependence of monthly medians of hour values $(foE)^4$ from corresponding values of $\cos\chi$ the 15th day of every month for a number of years in a certain site, one discovers that the quantity $C_o = a^4(O)$, that is $(foE)^4/\cos^p\chi$ calculated for R=O by the regression method, shows a surprising and evident winter maximum (Figure 10): it is the so-called Appleton anomaly of the E region. This means that the equation of electron equilibrium in the E region, that in the Chapman model is assumed in the form $\delta N/\delta t = q - \alpha N^2 = O$ in quasi-stationary conditions (q rate of electron production, α effective coefficient of ionic ricombination), must be corrected or considering an opportune variability of q and α or introducing (as it seems more acceptable) an opportune transport term.

A second example on the physical significance of the differences from an analytical model is, even for the E region, the previously mentioned longitude effect consisting in the fact that the exponent of $\cos\chi$ in the 1 varies around the theoretical value 0.25 in dependency on the geographic longitude λ; it is aknowledge that the localities in which the differences are the largest, correspond to what the geomagnetists call the foci of the ionospheric Sq currents, that determine the diurnal variation of the magnetic elements in magnetically calm days.

Another important example is the so-called winter anomaly in the ionospheric absorption determined by region D. If the absorption is considered as produced by a Chapman layer, it could be larger in summer months rather than, as usually happens, in winter months (that justifies the corrective winter factor W inserted in the absorption formula of the above cited ITS-78 Program: see table II).

The results obtained for both regions E and D from the comparison between observed data and data taken from Chapman-like models, and between these last data and data from purely statistical models, have for some time brought to the attention the opportunity of introducing convenient analytical models even for re

gion F and for sporadic E layer.

As to the F region, the last years show a situation of stall. There is infact no consistent reason for which a model, like that of Chapman, in the first approximation, which is well accepted for the E region, cannot be equally accepted for a slightly higher region; and yet, the anomalies, and especially the seasonal anomaly of foF2, of F region still resist to explanations based on Chapman's theory, even if one modifies the theory by introducing consistent terms of transport. A further difficulty consists, for all the "2 regions models", in the fact that the profile of electron density should present a clear "valley" between the two regions, while instead, as it is welle known, either measurements with rockets or ionograms in true heights give profiles with barely perceptible relative maximums in a monotonically increasing course, till the absolute maximum corresponding to the plasma frequency foF2.

An attempt to obviate this last disadvantage through a purely empiric way is given by a model (BRADLEY and DUDNEY, 1973) in which the above mentioned "valley" is filled in the way indicated in the Figure 5 B, that is linking linearly the maximum of the parabolic profile of the E region with the point of the F region parabolic profile corrisponding to the frequency 1.7 foE. Even if such a model seems to have given encouraging results, the evident empiricism of its construction brings to serious doubts about its heuristic capacities.

Really, the problem of developing a not completely empiric ionospheric model that therefore can bring to discover and/or well understand the phenomena from which ions and electrons are created and variously grouped in those formations we call ionospheric regions and layers is a problem of great difficulty. The ionosphere is infact a system of complex and not completely known constitution, where complicated chemical--physical (ionization, capture, etc.) and dynamic (gravitational tides, wave movements, winds, etc.) processes take place, with which the ionospheric particles (electrons, ions, neutral atomic groups) interact among themselves and the entire ionospheric system reacts with the lower "meteorological" atmosphere and with the upper magnetosphere; such a sistem is immersed, as on the other hand the whole atmpsphere, in terrestrial force and radiative fields (gravitational, magnetic, electric, thermal), and cosmic (planetary gravitational field, solar radiation, solar wind, cosmic radiation). The phenomena of interaction with the "outside", and that being, from one side with the "solid" Earth, and from the other principally with the Sun, some have a largely deterministic character, for ex. all those related to the heliocentric movements of the Earth, and therefore to the zenith angle of the Sun; others, on the contrary, have an undoubtely casual character, which, for instance, are those linked to sudden variations in solar activity.

In a similar situation, an ionospheric model cannot be satisfactory if, at least for its formulation, it is not an interactive model with remaining atmospheric systems. Even if we are not lacking in some attempts of formulating general atmospheric models of interactive kind, the fact remains that it seems reasonable to go along this way by small steps.

The first step , of course, is to consider once more what experience gives us , that is the traditional scheme based on the ionograms, with the innovation, however, of giving a particular attention to external interactions. The simpler "interactive" assumption is that of Chapman: the normal E region and F1 and F2 layers (as well as the D region, even tough observable only indirectly), are produced by solar photons and the profile of the ionosphere results in a linear (or a more complicated) superimposition of parabolic profiles relative to the E, F1, F2 formations (Figure 5 C).

The first of these models, gained attention in 1972 (DEDKER, 1972), for the presence of a parabolic layer, respective to the previous models with two parabolic layers; it regards, however, a model requiring complex algorithms and being relatively rigid. Much more interesting appears to be a successive "phenomenological model" (CHING and CHIU, 1973; CHIU, 1975), which is simpler regarding the calculations, and has the advantage of easily adapting itself to successive improvements, deriving form experience. The physical model consists in assuming the electron density at a generic altitude as a sum of the electron density of three Chapman-like layers, each of which density is given through a combination of functions of six variables: year time (order number of the day), day time (local hour), geomagnetic longitude, geographical latitude, altitude, sunspot number; these functions, that are derived substantially from Chapman's theory and in which appear therefore also the effective coefficients of ionic recombination, the atmospheric scale height, and the maximum height of ionization, are structured in such a way to include additive or multiplicative "phenomenological" terms, so that they can give coherent results with statistically observed values. A recent study (FLEURY and GOURVEZ, 1979) has compared monthly median values of foF2 calculated by this model, and by two models of the type used in the ITS-70 Program, and it has put in relief that the dependability is about the same for the three models; the phenomenological model of Ching and Chiu would seem to represent therefore a progress not so much in terms of precision, but rather in terms of physical formulation.

All these models concern monthly median values. Values of this type can be accepted in this phase,

which is still the first systematization of ionospheric models on physical base; in successive phases, when the analyses will be of second approximation, it will probably be convenient to refer to time intervals that are less conventional than from the calendar months; for instance, at periods more or less of equal variation of the solar declination, possibly in a simple relation with the solar rotation period.

As to instantaneous values, and even better as to their differences from the "normal" values defined by the median values of the model, much has been already made for the suitable construction of "perturbative models": think for example, of the rich literature available on ionospheric storms, on the ionospheric effects of solar eclipses, on radio waves fade-outs. It is this indeed the sector of ionospheric physics appeared immediately as the richest of geophysical contents and whose development was therefore relatively free from applicative conditionings. A noticeable improvement that must be undertaken in this sector is the formulation of norms for the reduction of the ionograms whose aim is to give, more effectively than at present, informations relative to ionospheric perturbations, that at the moment are accessible only to those who have the possibility of seeing the ionograms. A significant percentage of ionograms display very interesting formations for perturbative models and some, for example the layers very often present between the E region and the F region, interesting also for "normal" or "median" models.

6. THE PREVISION OF DATA FOR THE MANAGEMENT OF RADIO LINKS

From a radio-technical point of view, in order to express an opinion about the utility of an ionospheric model, it is necessary to distinguish between the exigences of the designers of radio links via the ionosphere, and those of the operators of such radio links.

As to the planning of an ionospheric radio link, it is necessary a reliable estimate of the values of some ionospheric characteristics at some places in a future time frame, which generally is enough distant; this last circumstance implies that the values in question can be median values. Well, the reliability of almost all the present models giving monthly median values is, as we already said, excellent, and in our opinion, sufficient generally speaking for the needs of a circuit planner and frequency allocator. Possible improvements are auspicable when their interest concerns not too much large areas; in such cases it would seem more convenient to use "zonal" models, rather than global, in such a way as to bring into the calculations some important particularities, and expecially the local characteristics of the sporadic E layer, that in global models are totally ignored, or else represented in a too large mean.

The situation facing the operator of ionospheric radio links is quite different, because a fundamental requirement is the continuity of the link. Let us make an example. Imagine an operator that at 1000 UT every day during the month of June 1980 to operate a link between England and Cyprus; the distance is about 3000 km and the ionospheric reflection, for a 1-hop path via F region, takes place on central Italy, such that the maximum usable frequency via F2 layer is approximately the MUF(3000)F2 measured in the sounding station at Rome. The planner of the link would have given, from an available zonal model of previsions (DOMINICI, 1975), a 90% MUF of 22.0 MHz, which corresponds to a median of 25.8 MHz, for the MUF(3000)F2; according to this prevision, the operator would have used a working frequency equal to 22.0 MHz. Well, in June 1980 at Rome the median MUF(3000)F2 at 1000 UT resulted to be 25.6 MHz, in good accord with the forecast, but the instantaneous MUF resulted in less than 22 MHz in 7 days among the 30 of the month; so, the link via F would have resulted sure only for 77% of the time, instead of 90% of the time. The situation described by this example is consistent with that described by previous Figures 6, 8, 9.

The question to adapt to the exigences of the operators the "median" models has been and is still the main problem of the ionospheric radio communications, and the great difficulties in solving it are one of the reasons of the relative decline these radio communications have had. Working on the same statistical base that has brought to the excellent median models now used, it is possible to develope some perturbative models based on the fact that the greatest discrepacies of the effective foF2 values in contrast with median values are related to solar flares and subsequent magnetic activity. Briefly, the photonic radiation suddenly arising from such solar events produce a contemporary sudden increase of the ionization of the D region, and therefore strong fadings of the signals (so-called Short Wave Fade-out), until the interruption of the links; the successively arriving solar corpuscolar radiation (protons and fast electrons) strongly perturbe the ionospheric equilibrium, producing large variations of the electron density in the F region (that is of foF2) and, in a much smaller way, in the lower regions, with a very accentuated geographical anisotropy due to the terrestrial magnetic field: they are the phenomena called "ionospheric storms" and "polar cap absorption".

The statistical models now used for ionospheric radio communications have all an input for data playing as perturbation indexes. For instance, in the many times mentioned ITS-78 Program, for short time forecasting it is used to this aim the K index of magnetic activity, given by geomagnetic observatories (BARGHAUSEN and others, 1969; pg. 93). In this order of ideas, however, it seems to be more rational to use

the ionospheric characteristics themselves as an index of the more or less perturbed state of the ionosphere, also because the communications between the agencies presiding short time ionospheric forecastings and those doing ionospheric measurements are easier than with the agencies doing other geophysical or astrophysical measurements. At this regard, an interesting example is the so-called "hybrid model" (FLATTERY and RAMSEY, 1975): there are two steps, the first of which is the simple application of the ITS-78 Program (included the correction by the magnetic K index), while the second is based on the comparison between the foF2 and M(3000)F2 median so obtained and mean values of the same characteristics in the last five days at a certain number of sounding stations. Adaptive models of this kind seems to be able to include, besides perturbative fluctuations, also the fluctuations present in quiet days, that are not correlated with solar or geomagnetic events but perhaps with tide effects and therefore with planetary configurations (HARNISHMACHER and RAWER, 1979).

The ionospheric modelling for radio communications purposes seems to go toward this direction, that is toward the interaction between a reliable statistical model and some kind of significant actual ionospheric data.

At this point we have to consider the following circumstancies:

(a) because from the point of view of the operators of ionospheric radio links the auspicable situation is the one in which the informations from forecasting agencies are suitable to assure the continuity of the link, it is necessary to have a model whose output is formed by operational parameters with 100% of probability: this aim does not seems reachable.

(b) a gain in reliability of the present forecasts implies a rapidly increasing complexity of the models and their associated activities; surely, the models would be unmanageable already well far the 100% probability level.

(c) because the efficacy of the introduction of ionospheric control data into a short time forecast model is so great as those data are near the present time, the limit of this process is to use data in real time; then it is much simpler to renounce the model with its intrinsical inadequacies, and to measure directly the operative parameters.

The (a) and (b) considerations bring to accept realistically the incovenience to look for increasing the data reliability from present models more than a reasonable limit, perhaps already very close, if not already reached. To avoid the operators unsatisfaction, the way is given by (c), that is to use adaptive systems of radio management.

As it is known, there are already systems of this kind. As to point-to-point links, the best of these systems (see, for instance, the commercially available swept spectrum system described in FENWICK and WOODHOUSE, 1979) consist in a transmitter and a receiver, each being at one of the terminals, making between them an oblique ionospheric sounding; at the receiving side, all the elements for the choice of the optimum working frequency are displayed, that are the MUFs relative to all the effective paths (including ground waves path), with their own signal intensity, the OWF being univocally determined as the one to which only one path and greatest signal intensity correspond; the occupacy of radio spectrum is also displayed. Such informations exceed the possibilities of our present forecast models and programs.

REFERENCES

BARGHAUSEN A.F., FINNEY J.W., PROCTOR L.L., SCHULTZ L.D., 1969, "Predicting long-term operational parameters of high-frequency sky-wave telecommunication systems", ESSA Tecn. Rpt. ERL110-ITS 78

BRADLEY P.A., DUDENEY T.R., 1973, "A single model of the vertical distribution of electron concentration in the ionosphere", Jour.Atm.Terr.Phys., Vol. 35, pp.2131-2146

CHING B.K., CHIU Y.T., 1973, "A phenomenological model of global ionospheric density in the E-, F1-, F2-regions", Jour.Atm.Terr.Phys., Vol. 35, pp.1615-1630

CHIU Y.T., 1975, "An improved phenomenological model of ionospheric density", Jour.Atm.Terr.Phys., Vol. 37, pp.1563-1570

DEDKER P.P., 1972, "Techniques for syntetizing median true height profiles from propagation parameters", Jour.Atm.Terr.Phys., Vol.34, pp.451-464

DOMINICI P., 1975, "Manuale di dati di previsione per radiocollegamenti ionosferici nell'area italiana", Servizio Ionosferico Nazionale CTSD-ING, Rome

FENWICK R.B., WOODHOUSE T.J., 1979, "Real-time adaptive HF frequency management", AGARD Conference "Special topics in HF propagation", Lisbon May 28-June 1, 1979

FLATTERY T.W., RAMSEY A.C., 1975, "Derivation of total electron content for real-time global applications", in "Effects of the ionosphere to space systems and communications", edit. J.M. Goodman

FLEURY R., GOURVEZ P., 1979, "Influence des models d'ionosphére sur les calculs de propagation des ondes décametriques", AGARD Conference "Special topics in HF propagation", Lisbon May 28-June 1, 1979

HARNISCHMACHER E., RAWER K., 1979, "Basic findings helpful for ionospheric predictions", AGARD Conference "Special topics in HF propagation", Lisbon May 28-June 1, 1979

JONES W.B., GALLET R.M., 1962, "The representation of diurnal and geographic variations of ionospheric data by numerical methods", Jour.Res. NBS, Sect.D; Vol.66D, pp.419-438

JONES W.B., GRAHAM R.P., LEFTIN M., 1969, "Advances in ionospheric mapping by numerical methods", ESSA Techn. Rpt. ERL 107-ITS 75

RAWER K., 1963, "Propagation of decametric waves", in "Meteorological and astronomical influences on radio wave propagation", edit. B. Landmark, pp.221-250, Pergamon, New York

TABLE I

Standard Ionosphere

Height [km]	50	100	200	300	500	1000
Zones	D Region	E Region	F1 Layer	F2 Layer	Heliosphere	Protonosphere
Prevalent ions	O_2^+ , NO^+	O_2^+, NO^+, N_2^+	O_2^+ , NO^+	O^+	He^+	H^+
Mean electron density [m^{-3}]						
day	10^9	10^{11}	10^{11}	10^{12}		
night	10^8	10^9	-	10^{11}	decreasing →	
Normal ionizing agents	← solar photons L_α 1261 Å		10 to 1000 Å →		← 100 to 1000 Å	
Geographic and time behavior	← simple			→ ← complex		→
Principal radio methods	← - - ← ground ionosondes →					
	radar scattering →					

TABLE II

Program ITS-78 (Numerical values are monthly medians)

F Region

foF2 — numerical maps

M(3000)F2 — numerical maps

h'F — maps

true height of the maximum ionization

$$hMF = \frac{1490}{M(3000)F2} - 176 \text{ [km]}$$

true height of the minimum ionization

$$\begin{cases} hmF = h'F - 20\,(z \ln\frac{z+1}{z-1} - 2) \text{ [km]} \\ z = 0.834\,\frac{foF2}{foE} \end{cases}$$

semi-thickness

$$ymF = hMF - hmF$$

E Region

foE — maps

hME — 110 km

semi-thickness 20 km

D Region

hMD — 63 km

absorption (non deviative)

$$\begin{cases} L = \frac{286\,(1 + 0.087|\varphi|)}{10 + (f+f_H)^2}\,FNW\,(\sec i)(1+0.005\,R_{13}) \text{ [dB]} \\ \begin{cases} F = \left[(\cos\chi_{12})^{(2.25-0.032|\varphi|)}\right]\frac{\cos(0.893\chi)}{\cos(0.893\chi_{12})} \\ 0 \le \chi \le 90^\circ \end{cases} \\ \begin{cases} F = 0.01 \\ \chi > 90^\circ \end{cases} \end{cases}$$

φ geographical latitude, degrees

W winter anomaly factor, world maps

N number of hops

i angle of incidence at 100 km

f_H gyrofrequency at absorbing height, MHz (world maps)

χ sun zenith distance, at noon

R_{13}^{12} smoothed 13-month sunspot number

Sporadic E

ftEs — maps

hMEs — 110 km

non-parabolic layer (thickness about 2 km)

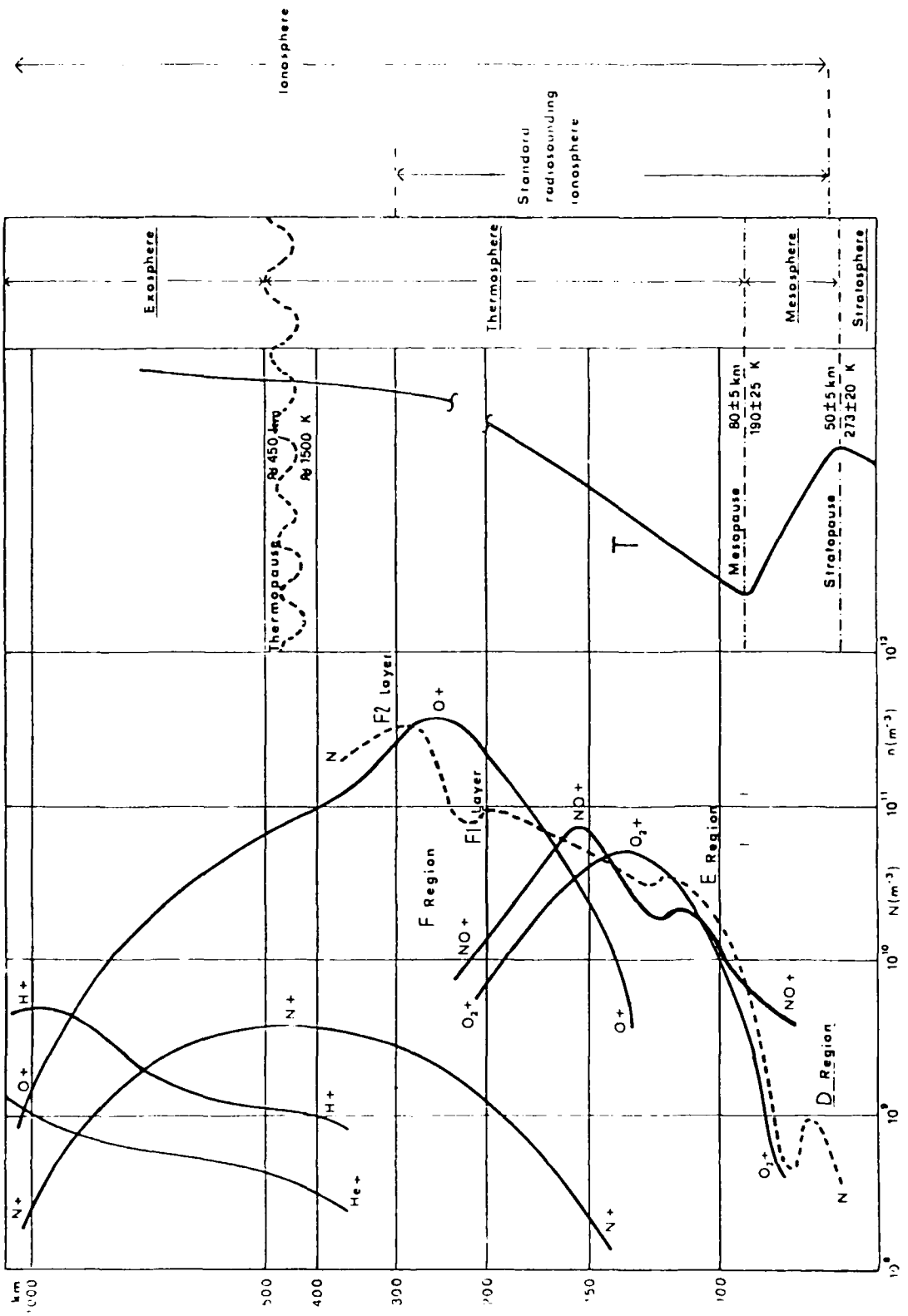


Fig. 1 Schematic behavior of some ionic densities n and electron density N, and of the temperature T of the terrestrial atmosphere: at right, atmospheric zones

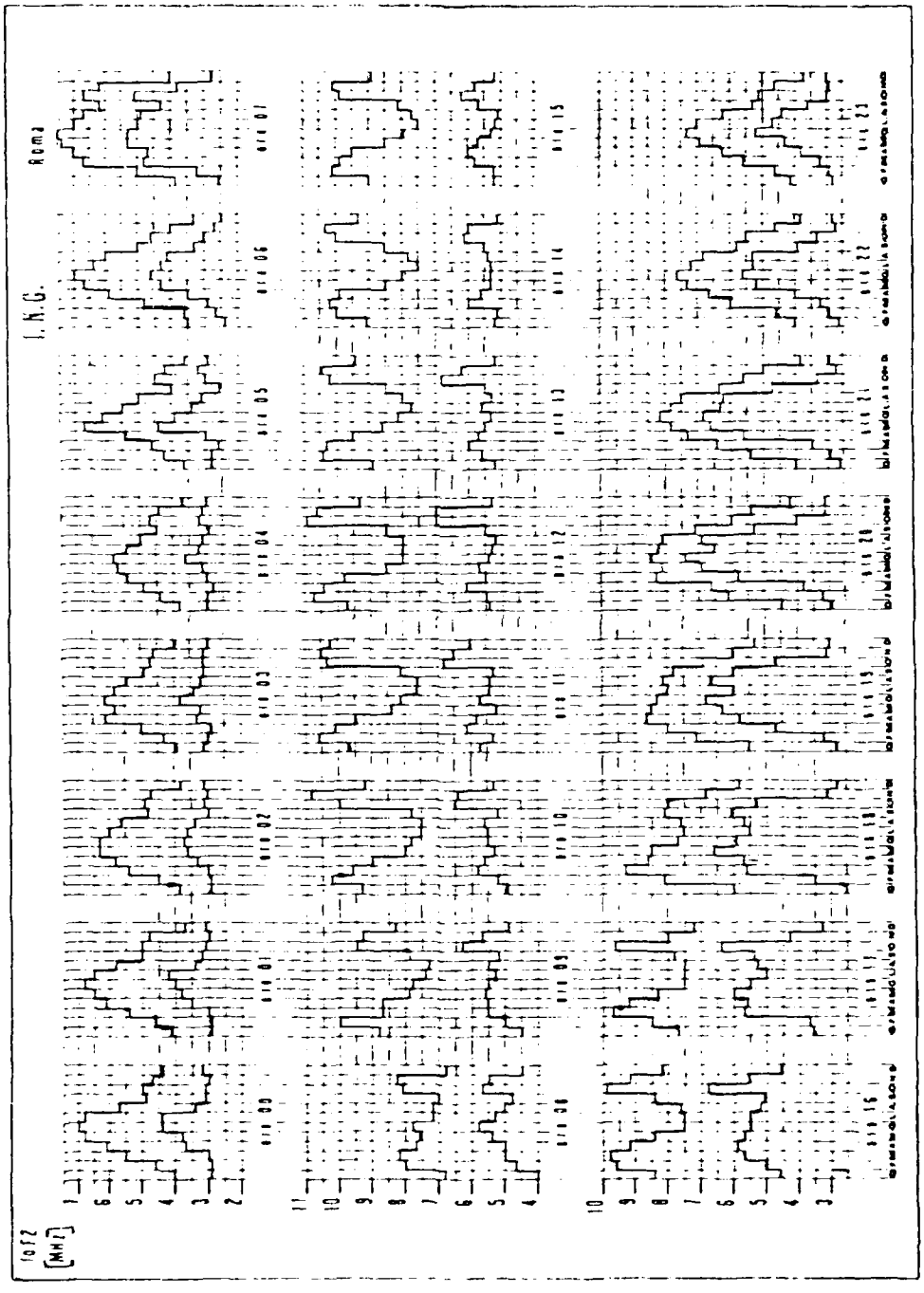


Fig.2 Annual behaviors of the monthly median values of the critical frequency foF2 at Rome (Italy) for R = 0 (lower curves) and for R = 100 (upper curves)

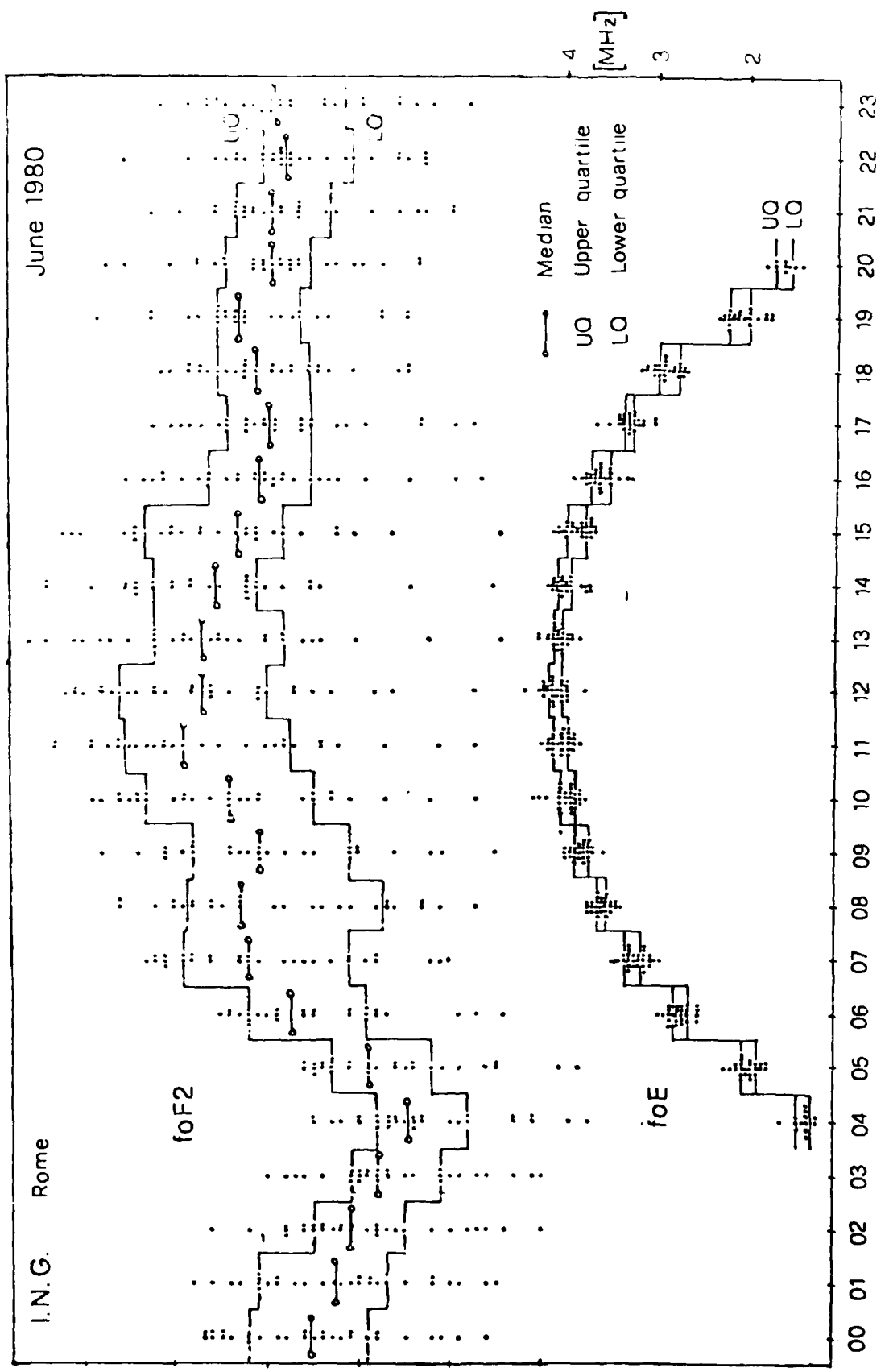


Fig.3 Monthly distribution of the critical frequencies foF2 and foE at Rome (Italy)

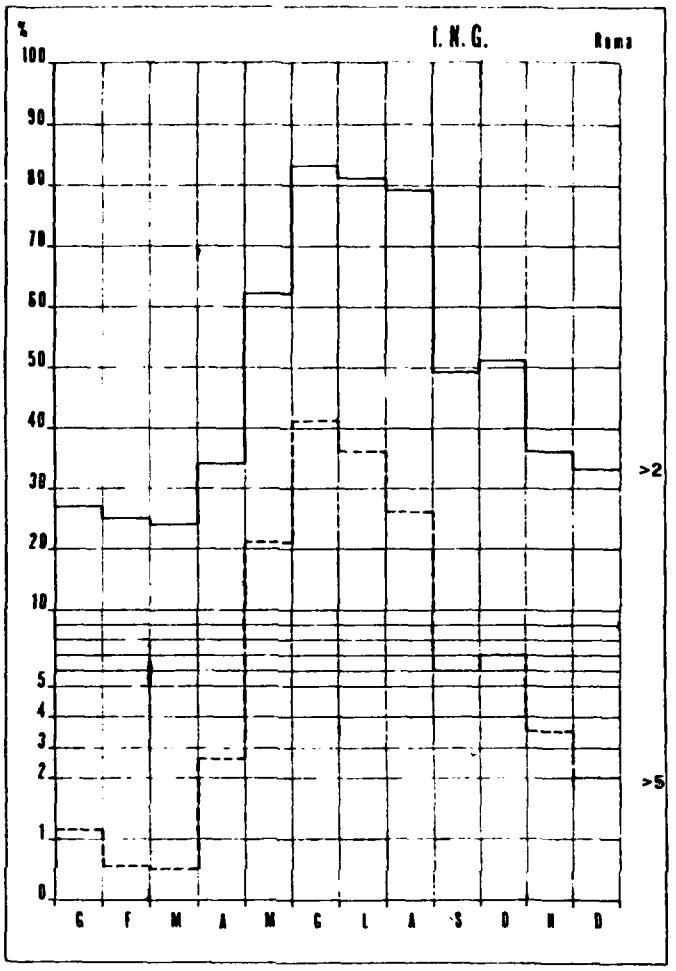


Fig.4 Monthly median appearance percentage of sporadic E at Rome (Italy), for frequencies exceeding 2 and 5 MHz

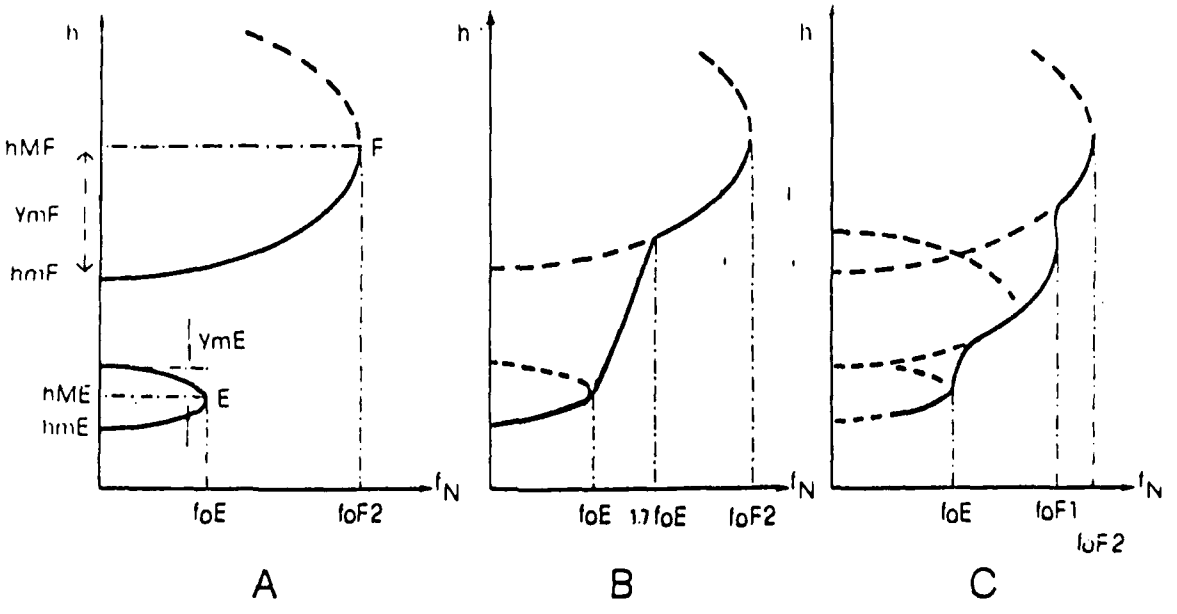


Fig.5 Ionospheric models with two (A, B) and three (C) parabolic layers

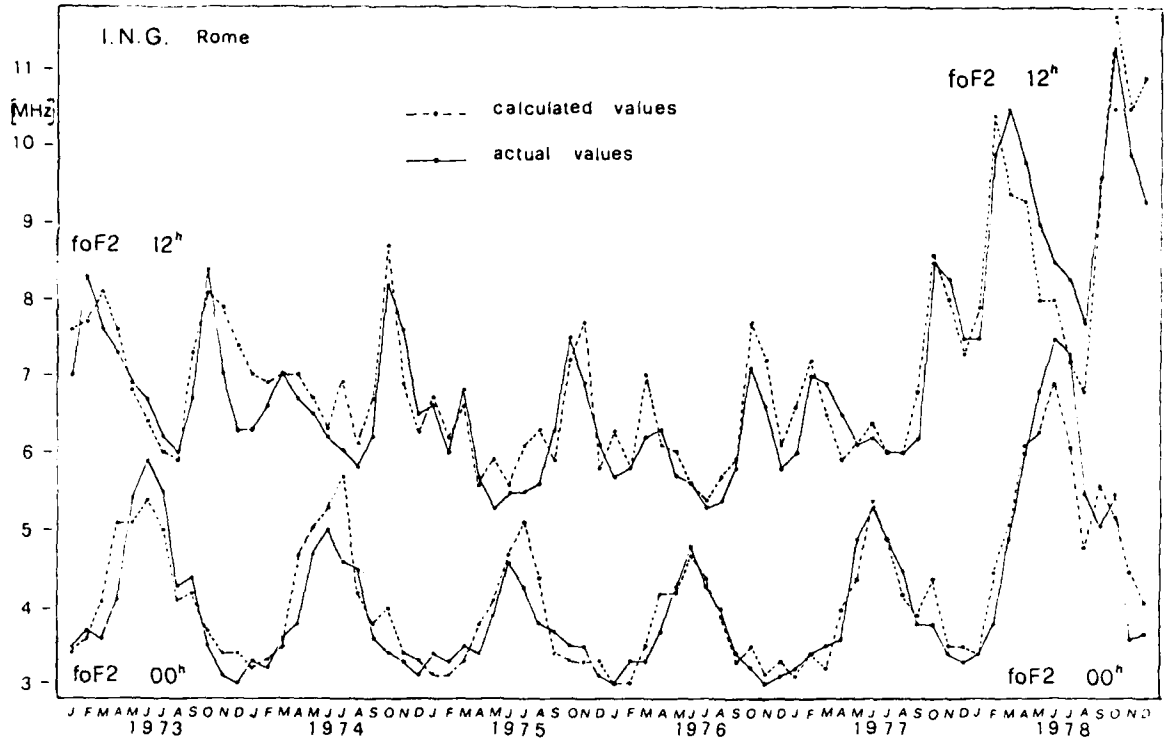


Fig.6 Comparison between values calculated by a local model and actual values of the critical frequency foF2 at local midnight and noon, at Rome (Italy)

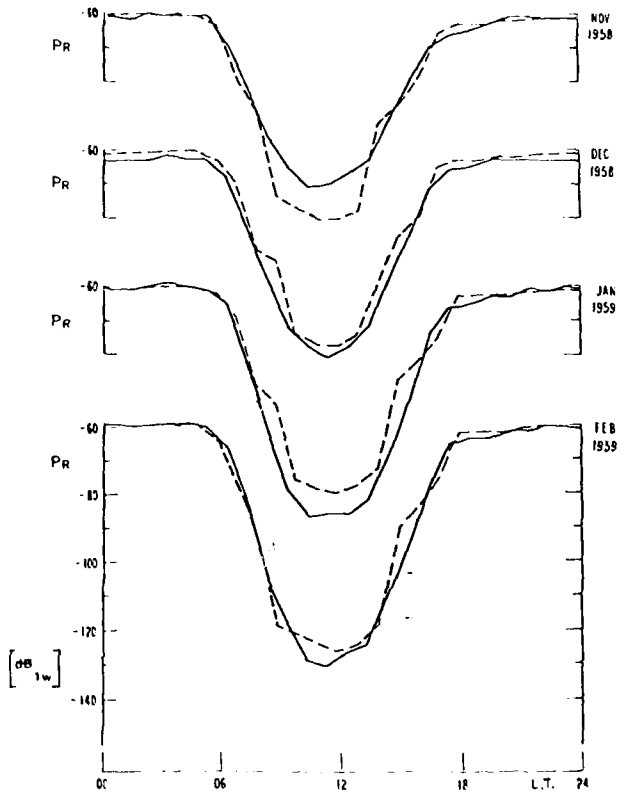


Fig.7 Median values observed (continuous curves) and calculated by the program ITS-78 (broken curves) of the received power P_R for a radio link on 5 MHz between Long Branch, Ill. and Boulder, Co., 1292 km long (*Barghausen et al.*, 1969)

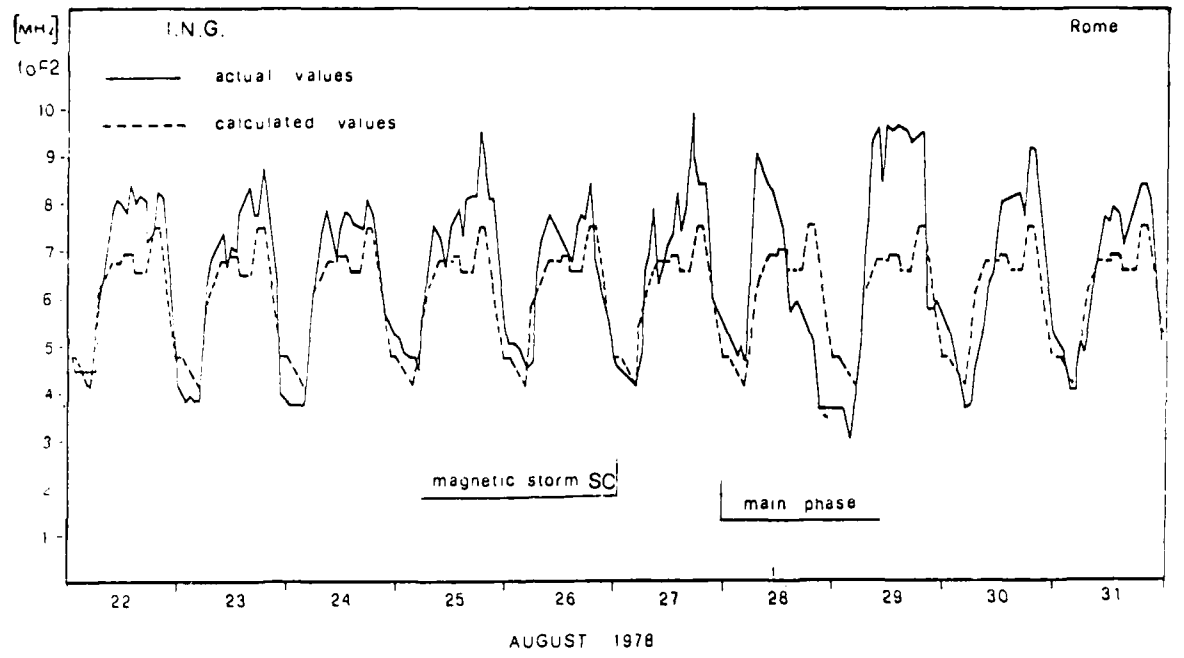


Fig.8 Diurnal behaviors of the critical frequency foF2 at Rome (Italy)

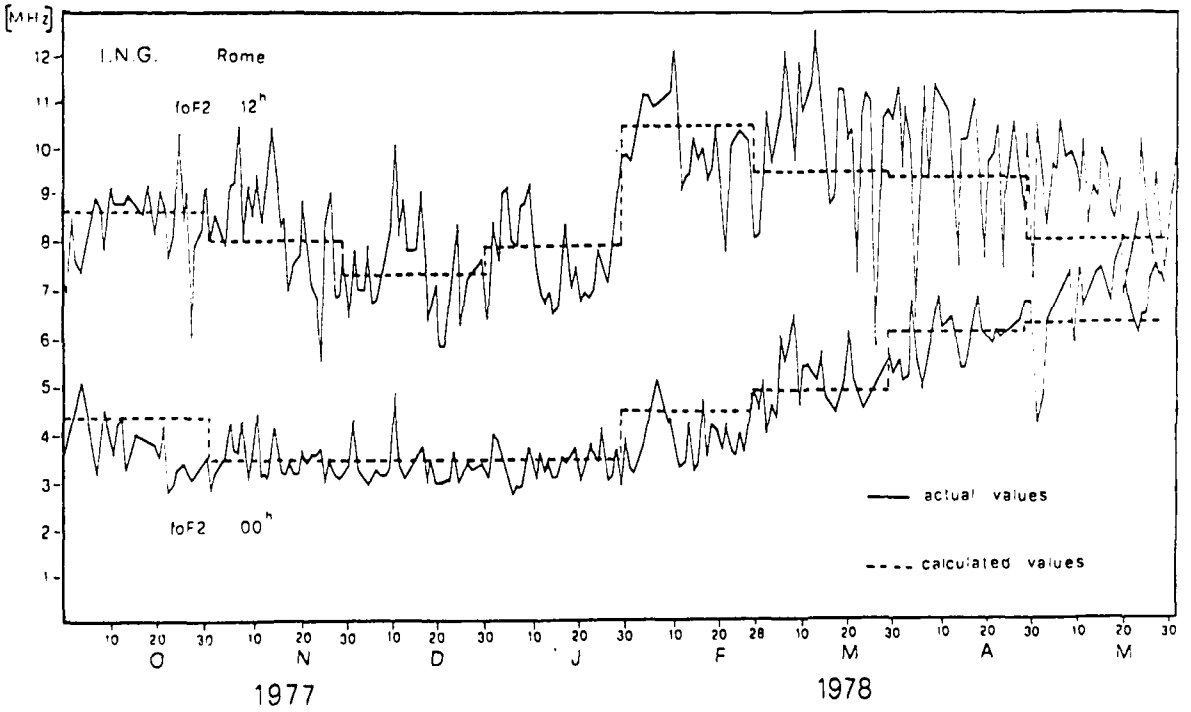


Fig.9 Comparison between day-to-day behaviors and monthly median values calculated by a local model at local midnight and noon, at Rome (Italy)

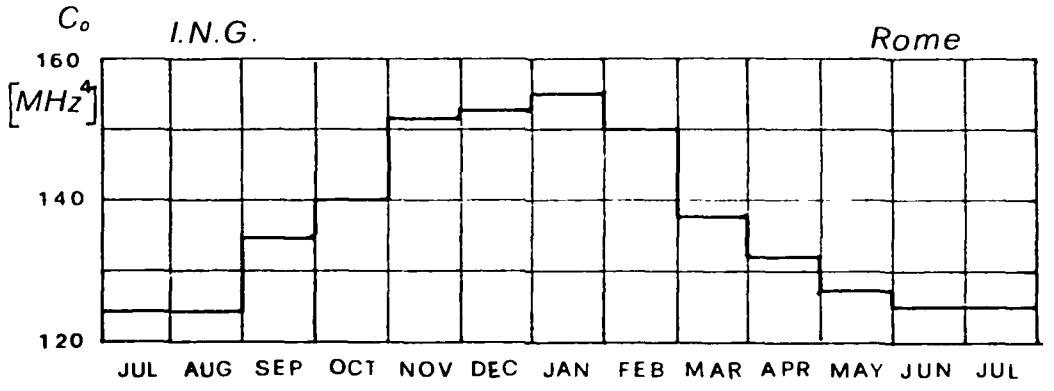


Fig.10 Appleton anomaly of the critical frequency foE at Rome (Italy)

SESSION DISCUSSION

SESSION VI – SOLAR AND IONOSPHERIC PREDICTIONS

Chairman and Editor – Prof. Dr H.Volland
Radioastronomical Institute
University of Bonn
Auf dem Huegel 71
53 Bonn 1, West Germany

PREDICTION TECHNIQUES FOR FORTHCOMING SOLAR MAXIMA
by G.M.Brown

T.Damboldt, Ge

May I remark that the mentioned "successful" predictions can predict the sunspot maximum only after the cycle has already begun, whereas the statistical methods predict the maximum sunspot activity of a cycle at any time even before the cycle has begun. I noticed however that the better predictions (your Figure 1) were made not only after the beginning of a cycle but in most cases only shortly before the maximum. As a user who needs forecasts of sunspot numbers for HF propagation predictions, I would like to know the chances for reliable predictions being made for the height of the maximum several years in advance.

Author's Reply

It depends on what you mean by "after (or before) the cycle has begun". If you consider that a new cycle starts with the reversal of the solar polar field, then this gives a lead-time of several years on the first appearance of the new cycle sunspots at minimum epoch, as I stress in the paper. If you refer to the conventional beginning of a cycle at sunspot minimum, I do not agree with your statement. Several of the precursor-type of predictions summarized in the paper are capable, in principle, of yielding their results around the time of sunspot minimum or even a little before this.

With regard to your second point it should perhaps be stressed that, although it is true that the majority of the more recent predictions for cycle 21 have been the ones indicating a substantial maximum, it should not therefore be concluded that they are cases of being "wise after the event". It is just the historical development of the subject and the relatively recent growth in the variety of approaches. Providing equivalent data will be available, there is no reason why any or all of the methods used rather late for cycle 21 should not be applied to the appropriate time to cycle 22. This should indeed be a salutory exercise!

C.M.Minnis, UK

Do the 3 "Brown 1979" predictions, which all agree and which will all be about right represent independent predictions?

Author's Reply

Yes, they are totally independent predictions based on three quite separate analyses. The letters in parentheses after the 1979 date in the histogram were inserted to indicate this point, although they are unexplained in the text. The methods are based on (i) considerations of the growth of 'Sunspots versus faculae' (SVF) during the early stages of a cycle, (ii) the occurrence of long-lived "away" sectors of the interplanetary magnetic field (IMF) at the preceding sunspot minimum, and (iii) the occurrence of 'abnormal quiet days' (AQD) of unusual Sq-phase at the preceding sunpot minimum.

H.E.Hinteregger, Ge

Besides a curiosity about the magnitude of the *maximum number of sunspots* to be expected for a new solar cycle, perhaps even greater interest would be related to estimating the corresponding *maximum EUV flux*, if one wishes to produce some sort of guess about the ionospheric-thermospheric implications expected for the maximum of a new sunspot cycle. Actual EUV irradiance observations of long-term variations unfortunately exist in good quality only for a part of cycle 20 and the present cycle 21. The finding of a "chromospheric EUV minimum" about 14 months before the June/July 1976 minimum of monthly mean R_z illustrates the point you made regarding the certainly long extent of the birthperiod of a new solar cycle. The occurrences of (1) an unexpected rise in the

quiet-disk chromospheric EUV and (2) unexpected higher ratios of relative variations of EUV versus $F_{10.7}$ observed already early in cycle 21, by June 1977, may have been important clues for expecting a rather impressive cycle 21 maximum in the EUV as I mentioned at the 1977 IAGA meeting in Seattle.

Author's Reply

This is an important point which I meant to have emphasised. I have restricted my review to considerations of the prediction of the conventional sunspot number, and it is right to stress that this is only one index of solar activity which certainly does not meet the total demands of those concerned with predictions of ionospheric electron densities or other solar-terrestrial relations. I mention in the paper that the cycle-to-cycle variability in the magnitude of peak sunspot number is not mirrored, for instance, in the corresponding variability in peak faculae areas, despite the fact that these are two intimately connected aspects of photospheric activity. This is just one example of the complexity of what we call a solar cycle. The different phasing of different aspects of activity which you mention may well, in some cases, be found to provide further useful precursors to the magnitude of a cycle.

IONOSPHERIC PREDICTIONS FOR HF RADIO SYSTEMS: THE FUTURE
by P.A.Bradley and M.Lockwood

K.Rawer, Ge

As for the mid-latitude trough it is certainly not reproduced in the CCIR numerical maps. (This is due to the *reduction schedule for ionograms used in the past on the one hand, and on the other to the mathematical* representation used in the model.) As bad as this is from a geophysical viewpoint, it may not be so disturbing for propagation predictions because the radio waves are not so stupid as our simplified ray-tracing is, and they avoid these zones of depletion, being reflected at places of higher electron density.

Author's Reply

This is indeed true for higher frequencies. If, however, the frequency is such that penetration does not occur in the great circle phase, then focusing and defocusing effects can occur and the trough modulates the received signal strength. (Lockwood, 1980 reference as in preprint).

T.B.Jones, UK

The value of predictions for long term planning is quite clear. However, for the user, the value must be limited since he is interested in the ionosphere as it exists at the time he wishes to communicate. Moreover, he generally has no access to a large computer on which to make a prediction even if he knew what ionospheric model to pull into it! In my view some form of real-time update of the prediction will be essential for the future.

Author's Reply

Such an update would have to be generated at a remote prediction centre and then communicated to the user. This would have to be based on a model of ionospheric storm effects or upon extrapolation of ionosonde soundings into the immediate future. Would the ionosonde used there be close enough to the path in question to give useful updates, bearing in mind the spatial variability of the ionosphere?

T.Damboldt, Ge

The influence of day-to-day variations of the ionosphere on HF radio propagation is so large that a further increase in accuracy of long-term predictions does not seem to be very sensible. Although long-term predictions are needed for systems design and planning purposes, don't you think that the future emphasis for HF predictions should be (besides the improvement of the prediction of noise and frequency occupancy) on real-time prediction systems?

of noise and frequency occupancy) on real-time prediction systems?

Author's Reply

Real-time methods are useful in frequency management and as we have seen many communication systems have developed real-time channel evaluation methods of some kind. Such systems remove the need for any type of predictions for frequency management once the design is decided.

C.M.Minnis, UK

In view of the high cost of establishing a short-term ionospheric disturbance service, has any consideration been given to measuring the "value" of the service to the actual user? An important factor is what action the user would like to take to maintain communications in the absence of a forecast.

Reference: Minnis, C.M.: Ionospheric stormwarning services, *Wireless Engineer*, *30*, 103, 1953.
Minnis, C.M.: Ionospheric disturbance forecasts: assessment of their economic value, *Telecommunication Journal*, *44*, 328, 1977.

Author's Reply

In the paper we wished to distinguish between routine short term predictions and storm predictions, and we feel the storm predictions would be of value, except for fully adaptive systems. The "over-reaction" of an operator to the onset of storm conditions (tending apparatus, trying frequencies at random) would be removed if he had been warned of the impending event.

ULTIMATE LIMITS TO ERROR PROBABILITIES FOR IONOSPHERIC MODELS BASED ON SOLAR GEOPHYSICAL INDICES AND HOW THESE COMPARE WITH THE STATE OF THE ART

by J.S.Nisbet and C.G.Stehle

M.Nicolet, Be

I would like to know if there is a possibility of dividing the atmosphere (ionosphere) into several regions such as the tropical region, the mean latitudes and the polar regions, which would have different behaviours?

Author's Reply

The heat input due to Joule heating in the auroral electrojets causes winds that cause density perturbations to propagate in a few hours to the equator. I really feel the only hope is to have global neutral models, but they must use better estimates of the energy inputs.

EXPERIMENTAL VALIDATION OF THE ONSOD OMEGA PREDICTION METHOD

by T.B.Jones and K.Mowforth

U.B.Mitchell, UK

Could you higher error rate on the receivers inland in the UK be due to ambient radio noise?

To quote an example, there is a very considerable amount of power from domestic T.V. receivers very close to 10 kHz. This is generated by the line time bases (not the r.f. sections).

Author's Reply

I do not think so since the bandwidth of the receivers is very narrow – they are in fact phase tracking receivers. There could, however, be a field site effect. Two of the receivers are inland whereas the other is on the sea shore.

Note also, similar effects are seen at 13.6 kHz.

H.Volland, Ge

Is it possible to eliminate the remaining errors in the forecasting model?

Author's Reply

Not at the moment. We are working at present on the problem.

J.S.Nisbet, US

I am very pleased to see these statistics on the variations from the models and variations from the mean. For the design of systems the accuracy often depends on these statistics, and it is important to get them published - not just the mean monthly models.

Author's Reply

I agree. This approach was made to help the system designers.

SOME CONSIDERATIONS UPON MODELS FOR IONOSPHERIC RADIO COMMUNICATIONS FORECASTING

by P.Dominici and B.Zolesi

K.Rawer, Ge

I would like to get some information on how the model you used for comparison with your measured data was obtained. Was it purely empirical or have certain geophysical reasonings been used?

Author's Reply

For the F2 region we use a statistical model based on the correlation formula $\overline{f_o F2_H} = A_H + B_H R$ relative to the single local hours H, where A_H, B_H are tabulated coefficients and R is the monthly mean value of the sunspot number R.

LIST OF ATTENDEES

AARONS, J. Dr	Senior Scientist, Air Force Geophysics Lab. L.G. Hanscom Field, Bedford, MA 01731, USA
ALBRECHT, H.J. Dr-Ing.	FGAN, Königstrasse 2, 5307, Wbg-Werthoven, W. Germany
AXFORD, W.I. Dr	Max-Planck Institute fur Aeronomie, Postfach 20, D-3411, Katlenburg-Lindau 3, W. Germany
BALCIOGLU, O. Miss	Istanbul University, Verbilimleri, Fakultesi Jeofizik Muh Bolumu, Vezneciler-Istanbul, Turkey
BARBATI, S. Dr	Contraves Italiana, Sp.A Via Tiburtina 965, 00156 Roma, Italy
BAUER, P. Dr	Matre de Recherche, CNET/CRPE, 3 Av de la Republique, 92131 Issy-les-Moulineaux, France
BELROSE, J.S. Dr	Communications Research Centre, Dept. of Communications, P.O. Box 11490 Stn. H, Ottawa K2H8S2, Canada
BLANC, M. Dr	CNET/RPE, 38–40 rue du General Leclerc, 92131 Issy-les-Moulineaux, France
BLYTHE, J.H. Dr	Marconi Research Labs, Great Baddow, CM2 8HN, Essex, UK
BOITHIAS, L. Mr	Ing en Chef du CNET, 38–40 rue du General Leclerc, 92131 Issy-les-Moulineaux, France
BOSSY, L. Prof.	Institut Royal Meteorologie, 3 Ave Circulaire, B-1180 Bruxelles, Belgium
BOTTARI, E. Ing.	Accademia Aeronautica, Pozzuoli, Italy
BREKKE, A. Prof. Dr	The Auroral Observatory, P.O. Box 953, N9001 Tromso, Norway
BROWN, G.M. Dr	Dept. of Physics, University College of Wales, Aberystwyth, Dyfed, UK
BULAT, T. Assoc. Prof. Dr	I.U. Yerbilimeri Fakultesi, Jeofizik Muh Bolumu, Univ. of Istanbul, Vezneiler Istanbul, Turkey
COWLEY, S.W.M. Dr	Blackett Laboratory, Imperial College, London, SW7 28Z, UK
COYNE, V. Mr	Asst. Chf Strategic Surveillance Br, RADC. OCS, Griffiss AFB, N.Y. 13441, USA
CROCHET, M. Prof.	L.S.E.E.T., La Giponne, Bd des Armaris, 83100 Toulon, France
CUTOLO, M. Prof.	Universita di Napoli, Istituto di Fisica, Via Monteoliveto 3, 80134 Napoli, Italy
DAMBOLDT, T. Dr	Forschunginstitut der DBP, Postfach 5000, D-6100 Darmstadt, W. Germany
DI COCCO, R. Mr	c/o S.I.A., via Canova n.25, 10126 Torino, Italy
EGELAND, A.	University of Oslo, Norway
ELLINGSEN, E. Mr	ELAB, Norwegian Institute of Tech., N-7304 Trondheim-NTH, Norway
FABI, G.A.Umberto Generale Isp.	Aeronautica Militaire, Ufficio del Delegato Nazionale all'A, Piazzale K Adenauer, 3 00144 ROMA/EUR, Italy
FORBES, J. Prof	Dept. of Astronomy, Boston University, Boston, Mass., USA
GOLDBERG, R.A. Mr	NASA/Goddard Space Flight Center, Laboratory for Planetary Atmospheres, Greenbelt, MD 20771, USA
GREGORI, G.P. Prof.	Istituto di Fisica d'Atmosphere, Rome, Italy
GROSS, S.H. Prof.	Polytechnic Institute of New York, ROUTE 110, Farmingdale, N.Y. 11735, USA
HANBABA, Dr	MER/MPE, CNET 22301 Lannion, France
HINTERREGER, H.E. Dr	Air Force Geophysics Lab., Hanscom AFB, MA 01731, USA
IASELLI, P. Col.	Viale dell'Universita 4, 00185 Roma, Italy
IRVINE, W. Mr	Communications Research Center, Dept. of Communications, P.O. Box 11490, Station H, Ottawa, K2H 8S2, Canada

JARVIS, M.J. Dr	HSO (Ionospherics), British Antartic Survey, Madingley Rd., Cambridge, UK
JONES, T.B. Prof.	Physics Dept., The University, Leicester, UK
KOHL, H. Prof. Dr	Max-Planck Institut fur Aeronomie, Postfach 20, D-3411 Lindau 3, W. Germany
LAMPERT, E. Dr Ing.	Siemens AG, Postfach 70 00 70, 8000 Munchen 70, W. Germany
LANGE-HESSE, G. Dr	Max-Planck Institut fur Aeronomie, Postfach 20, D-3411 Lindau/Harz, W. W. Germany
LOCKWOOD, M. Dr	Rutherford & Appleton Labs, Ditton Park, Slough, SL3 9JX, UK
MERKEL, M.A. Mr	US Army Communications Electronics, CC-EMED, Engineering Installation, Ft. Huachuca, AR 85613, USA
MINNIS, C.M. Dr	URSI Secreatariat, Ave. A. Lancaster 32, B-1180 Brussels, Belgium
MITCHELL, V.B. Prof.	University of Exeter, Physics Dept., Stocker Rd., Exeter, EX4 4QL, UK
NESKE, E. Mr	IPM, Heidenhoffstr. 8, 78 Freiburg, W. Germany
NICOLET, M. Prof.	External Geophysics, Brussels University, 30 Ave. Den Doorn, 1180 Brussels, Belgium
NISBET, J.S. Prof.	CODE 960, Goddard Space Flight Center, Greenbelt MD, 20771, USA
POEVERLEIN, H. Prof. Dr	Technische Hochschule, 6100 Darmstadt, W. Germany
PONTICELLO, Capt.	Academmia Aeronautica, Pozzuoli, Italy
PROLSS, G. Dr	Inst. fur Astrophysik, Auf dem Hugel 71, D5300 Bonn, W. Germany
RAITT, W.J. Prof.	Center for Atmospheric & Space Sci., UMC 41, Utah State University, Logan, Utah 84322, USA
RAWER, K. Prof.	D-7801 March 2, Herrenstrasse 43, W. Germany
RICHMOND, A.D. Dr	NOAA/ERL/SEL R43, 325 Broadway, Boulder, CO 80303, USA
ROTTGER, J. Dr	Max-Planck Institut fur Aeronomie, P.O. Box 20, D-3411 Katlenberg-Lindau 3, W. Germany
SCHEGGI, A.M. Prof.	IROE, Via Panciatichi 64, 50127 Firenze, Italy
SCHMERLING, E.R. Dr	Chief, Plasma Physics (Code ST5), Office of Space Science, NASA Hqs., Washington D.C., 20546, USA
SCHMIDTKE, G. Dr	Institut fur Physikal Messtechnik, Heidenhofstr. 8, D-7800 Freiburg, W. Germany
SHARP, R. Dr	Lockheed, Palo Alto Research Lab., Dept. 52 -12 Bldg., 205 Palo-Alto, CA 94304 USA
SMITH, L.G. Dr	Dept. of Electrical Eng., University of Illinois, Urbana, Ill 61801, USA
SOICHER, H. Dr	c/o USDAO, US Embassy, Tel Aviv, APO New York 09672, USA
SPALLA, P. Dr	Istituto di Ricerca sulle Onde, Elletromagnetiche del C.N.R., via Panciatchi n.64 50127 Firenze, Italy
SPITZ, E. Dr	Thomson-CSF, Corbeville, 91401 Orsay, France
SPRENKELS, C. Col. D'Aviation Ir.	Commandant 22 Log W, Quartier Roi Albert I, rue de la Fusee 70, B-1130 Bruxelles, Belgium
STARK, A. Dipl. Ing.	Fa. Rhode u. Schwarz, Unterbereich Funkerfassung, Funkortung Postfach 80 14 69, D-8000 Munchen 80, W. Germany
STUBBE, P. Mr	Max-Planck Institut für Aeronomie, 3411 Katlenburg-Lindau 3, W. Germany
TAAGHOLT, J. Mr	Ionosphere Laboratory, Technical University of Denmark, Bldg. 349, DK-2800 Lyngby, Denmark
THIEMANN, H. Dr	Rotlaubstrasse 20, D-78 Frieburg, W. Germany
THRANE, E. Mr	N.D.R.E., 2007 Kjeller, Norway
TORR, M.R. Prof.	Center for Atmospheric and Space Science, Utah State University, Logan, Utah 84322, USA
TORR, D.G. Prof.	Center for Atmsopheric and Space Science, Utah State Univeristy, Logan, Utah 84322, USA
TSOUKIAS, T. L/Col.	Technology Research Center (KETA), Delta Fallirou, Palaion Faliron, Athens, Greece

VAGNARELLI, F. L/Col. Dr	Aeronautica Militaire Italiana, Ufficio del Delegato Nazionale all'A, 3 Ple Adenauer, Roma/EUR, Italy
VOLLAND, H. Prof. Dr	Radioastronomical Institute, University of Bonn, Auf dem Huegel 71, 53 Bonn 1, W. Germany
WILSON, Q. Mr	Mitre Corp., USA
ZOLESI, B. Dr	Istituto Nazionale di Geofisica, via Ruggero Bonghi 11b 00184 Roma, Italy

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-295	ISBN 92-835-0284-1	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM

7. Presented at
the 27th Meeting of the Electromagnetic Wave Propagation Panel held in Pozzuoli, Italy 28–31 October, 1980

8. Author(s)/Editor(s)
Various

9. Date
February 1981

10. Author's/Editor's Address
Various

11. Pages
410

12. Distribution Statement
This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

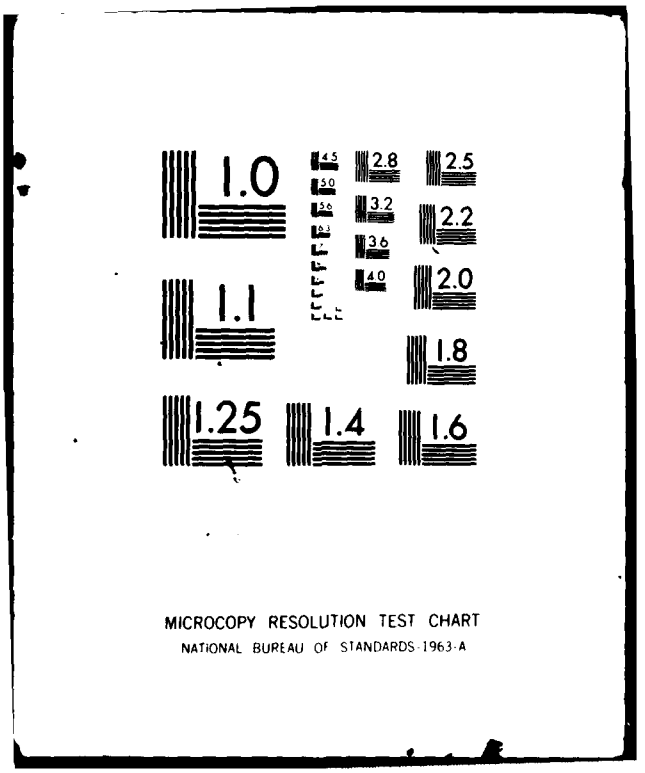
13. Keywords/Descriptors

Electromagnetic wave transmission
Ionospheric propagation
Ionospheric disturbances
Sun
Solar wind
Magnetosphere

14. Abstract

These Proceedings consist of the 36 papers and the discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel in Pozzuoli. There are 4 papers on the sun, solar wind, and magnetosphere; 6 on solar wind, magnetosphere ionosphere coupling; 7 on dynamical coupling; 7 on ionospheric prediction and morphology; 6 on irregularities and waves; and 6 on solar and ionospheric predictions.

AD-A098 119 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 4/1
THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL S--ETC(U)
FEB 81
UNCLASSIFIED AGARD-CP-295 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AGARD Conference Proceedings No.295 Advisory Group for Aerospace Research and Development, NATO THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM Published February 1981 410 pages These Proceedings consist of the 36 papers and the discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel in Pozzuoli, Italy on 28–31 October, 1980. There are 4 papers on the sun, solar wind, and magnetosphere; 6 on solar wind, magnetosphere ionosphere coupling; 7 on dynamical coupling; 7 on ionospheric prediction and morphology; 6 on irregularities and waves; and 6 on solar and ionospheric predictions. ISBN 92-835-0284-1	AGARD-CP-295 Electromagnetic wave transmission Ionospheric propagation Ionospheric disturbances Sun Solar wind Magnetosphere	AGARD Conference Proceedings No.295 Advisory Group for Aerospace Research and Development, NATO THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM Published February 1981 410 pages These Proceedings consist of the 36 papers and the discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel in Pozzuoli, Italy on 28–31 October, 1980. There are 4 papers on the sun, solar wind, and magnetosphere; 6 on solar wind, magnetosphere ionosphere coupling; 7 on dynamical coupling; 7 on ionospheric prediction and morphology; 6 on irregularities and waves; and 6 on solar and ionospheric predictions. ISBN 92-835-0284-1	AGARD-CP-295 Electromagnetic wave transmission Ionospheric propagation Ionospheric disturbances Sun Solar wind Magnetosphere
AGARD Conference Proceedings No.295 Advisory Group for Aerospace Research and Development, NATO THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM Published February 1981 410 pages These Proceedings consist of the 36 papers and the discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel in Pozzuoli, Italy on 28–31 October, 1980. There are 4 papers on the sun, solar wind, and magnetosphere; 6 on solar wind, magnetosphere ionosphere coupling; 7 on dynamical coupling; 7 on ionospheric prediction and morphology; 6 on irregularities and waves; and 6 on solar and ionospheric predictions. ISBN 92-835-0284-1	AGARD-CP-295 Electromagnetic wave transmission Ionospheric propagation Ionospheric disturbances Sun Solar wind Magnetosphere	AGARD Conference Proceedings No.295 Advisory Group for Aerospace Research and Development, NATO THE PHYSICAL BASIS OF THE IONOSPHERE IN THE SOLAR-TERRESTRIAL SYSTEM Published February 1981 410 pages These Proceedings consist of the 36 papers and the discussions presented at the 27th Meeting of the Electromagnetic Wave Propagation Panel in Pozzuoli, Italy on 28–31 October, 1980. There are 4 papers on the sun, solar wind, and magnetosphere; 6 on solar wind, magnetosphere ionosphere coupling; 7 on dynamical coupling; 7 on ionospheric prediction and morphology; 6 on irregularities and waves; and 6 on solar and ionospheric predictions. ISBN 92-835-0284-1	AGARD-CP-295 Electromagnetic wave transmission Ionospheric propagation Ionospheric disturbances Sun Solar wind Magnetosphere