Ionospheric F-region storms: unsolved problems

Gerd W. Prölss
Argelander Institut für Astronomie, Universität Bonn,
Auf dem Hügel 71, 53121, Bonn, Germany

Abstract: Befitting the venue of this meeting, we first investigate the polar ionosphere. The most pressing problem here is that no reliable description of this region is available, not even for undisturbed conditions. To improve on this situation we have investigated the properties of some of the more prominent anomalies observed in this region, including the heating effect below the magnetospheric cusp, the subauroral electron temperature enhancement and the main ionospheric trough. Using DE-2 satellite data, we find, for example, that all these features move towards lower latitudes with increasing geomagnetic activity in a systematic way. We also show that the subauroral electron temperature enhancement and the main ionospheric trough are co-located, on average. With regard to the mid-latitude region, the positive phase of ionospheric storms remains the most challenging problem. Different mechanisms have been proposed to explain this phenomenon, including neutral gas composition changes, equatorward directed winds, and east- and northward directed electric fields. Up to now, the measurements available are not sufficient to single out the correct explanation(s). Ionospheric holes are one of the most spectacular disturbance effects observed at equatorial latitudes. These holes are marked by a steep drop in the electron density to very low values. Also their bottom is rather flat and almost without any structure. Different explanations of this phenomenon have been offered, none of which is generally accepted.

1 Introduction

The term ionospheric storm is used here to designate a global perturbation of the ionosphere caused by the strongly enhanced dissipation of solar wind energy in the near-Earth space environment. Evidently, such disturbances form an important link in the complex chain of solar-terrestrial relations. They are also of great practical interest since sub- and transionospheric radio communications may be severely degraded or even disrupted during such events.

Ionospheric storms were discovered more than 75 years ago (Hafstad and Tuve, 1929). Prior to that, radio engineers found that communication links involving wave propagation through the ionosphere were materially affected
**Title:** Ionospheric F-region storms: unsolved problems

**Performing Organization:** Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121, Bonn, Germany
Figure 1: Latitudinal variations of the electron temperature and density in the polar region during moderately disturbed conditions (AE= 227). The data were obtained by the DE-2 satellite in the noon-midnight meridian at about 21 UT on September 5, 1982. Prominent anomalies are the electron temperature increase beneath the magnetospheric cusp; the subauroral electron temperature enhancement (SETE); and the main ionospheric trough.

during magnetic storms (Espenschied et al., 1925). Since then more than 600 papers have been published on this subject. In spite of this tremendous effort, ionospheric storms remain a very challenging topic even today. This is because many aspects of this fascinating phenomenon remain incompletely documented and understood. In what follows only a few of these unsolved problems will be discussed. In our selection we proceed from polar to equatorial latitudes. First, some anomalies of the polar and subpolar ionosphere and their dependence on geomagnetic activity are investigated (Section 2). Next, positive ionospheric storms at middle latitudes and their possible origins are discussed (Section 3). Finally, ionospheric holes at equatorial latitudes will be described (Section 4).
2 Ionospheric storm effects at higher latitudes

Several features combine to make the polar ionosphere a rather complicated region. It is here where most of the solar wind energy is dissipated and where this dissipation often dominates the upper atmospheric energy budget. Accordingly, changes induced by this energy source are not of a transient nature. Rather, there is a continuous transition between more and less disturbed conditions. Moreover, satellite measurements show that the ionization density and temperature in this region are highly structured, particularly in the absence of sunlight. Typically, a satellite will observe a number of peaks and troughs in these parameters as it traverses the polar region. This is illustrated in Figs. 1 and 2 for moderately and more strongly disturbed conditions, respectively. The dynamics of these features in response to the continually varying level of magnetospheric activity is considerable. It renders the determination of a quiet-time pattern which could serve as a reference in storm studies rather difficult. In this situation, it seems more profitable to concentrate on the storm behavior of specific morphological features identified in the satellite data. These include the cusp heating effect, the subauroral electron temperature enhancement (SETE), and the main ionospheric trough; see Figs. 1 and 2. In what follows, their origins and their dependences on geomagnetic activity will be briefly discussed.

2.1 Cusp heating effect

The cusp represents an important topological feature of the terrestrial magnetosphere. Here magnetic field lines separate, extending to different parts of the magnetosphere; see Fig. 3. Because of this special topology, plasma from the magnetosheath and the magnetospheric boundary layers have direct access to the dayside polar atmosphere. Accordingly, large numbers of low-energy electrons are observed to precipitate into this region (e.g. Newell et al., 2004; and references therein). This soft electron precipitation is responsible for a significant heating of the electron gas in the upper ionosphere, leading to the prominent peak in the electron temperature documented in Figs. 1 and 2. In what follows we are interested in how the location of this heating effect depends on the level of geomagnetic activity.

A comparison of Figs. 1 and 2 indicates that during disturbed conditions the cusp-related temperature peak is located at lower latitudes. In order to describe this displacement effect in a systematic way, the latitude of the temperature peak and the associated auroral electrojet index AE were determined for a large number of satellite passes. Subsequently, these latitudes were sorted into 100
Figure 2: Latitudinal variations of the electron temperature and density in the polar region during more strongly disturbed conditions (AE = 702). The data were obtained by the DE-2 satellite in the noon-midnight meridian at about 23 UT on March 1, 1982. As in Fig. 1 the temperature increase beneath the magnetospheric cusp, the subauroral electron temperature enhancement (SETE), and the main ionospheric trough are indicated.

nT wide intervals of the AE index. For each of the intervals, the median and the upper and lower quartiles were determined. These are indicated by the dots and bars in Fig. 4. Finally, a regression line was fitted to the medians. As is evident, an excellent linear correlation exists between the AE index and the position of the temperature peak. For completely quiet geomagnetic conditions (AE = 0), the regression line predicts the temperature peak to be located near 79° invariant latitude. For each increase in the AE index by 100 nT, it moves equatorward by about 1 degree. During strongly disturbed conditions, it may be observed at a latitude as low as 61 degrees.

There is also a tendency for the magnitude of the temperature enhancement
Figure 3: Topology of the magnetosphere in the cusp region and associated plasma populations. The Tsyganenko model T96-1 was used to calculate the field lines for winter solstice conditions in the northern hemisphere. The footpoints of the field lines are located at 70°, 71.5°, 73.5° and 75° magnetic invariant latitude. NP stands for (geographic) north pole, LLBL for low-latitude boundary layer. (From Prölls, 2006a)

to increase with increasing magnetic activity. This correlation, however, is relatively weak, and the scatter of data points is large. In fact, the magnitude of the heating effect varies much more strongly with altitude (Prölls, 2006a).

2.2 Subauroral electron temperature enhancement (SETE)

Another prominent peak in the electron temperature is observed in the nighttime ionosphere at subauroral latitudes; see again Figs. 1 and 2. One possible explanation of this heating effect is illustrated in Fig. 5.

During magnetically disturbed conditions, large numbers of energetic particles are injected into the inner magnetosphere. These ring current particles are a major heat source for the thermal plasma in this region. Whereas the primary energy transfer channel from the hot ring current ions to the cool plasma-particles is through Coulomb collisions (e.g. Kozyra et al., 1997 and references therein), wave particle interactions may also be important (e.g. Mishin and Burke, 2005 and references therein). In both cases, the heating
Figure 4: Position of the cusp-related temperature peak as a function of geomagnetic activity. The ordinate indicates the invariant latitude of the temperature peak, $latm$, and the abscissa the associated $AE$ index. For the $AE$ index, linearly-interpolated, hourly-averaged values are used. Both parameters were determined for a total of 1409 electron temperature profiles. Only data obtained in the 9 to 15 h magnetic local time sector were considered. For each 100 nT-wide interval of the $AE$ index, the median and the upper and lower quartiles of the latitudes of the temperature peak were determined; they are shown by the dots and bars in the figure. A regression line has been fitted to the medians, and its ordinate intersection $a$ (in K) and slope $b$ (in K/nT) are given in the lower left-hand corner. (From Prölls, 2006a)

occurs preferentially near the equatorial plane of the ring current. However, due to the high thermal conductivity of the electrons, part of this energy is carried along magnetic field lines down into the topside ionosphere; see Fig. 5. Accordingly, any satellite passing through the footpoint region of field lines threading the equatorial heating region should observe a sudden increase in the electron temperature, and this is indeed the case.

Now the ring current population turns out to be highly dynamic. During disturbed conditions its density may increase by a factor of ten or more within
Figure 5: Topology of magnetic field lines threading the inner-magnetospheric heating region. A projection onto the x - z plane of a geocentric solar-magnetospheric (GSM) coordinate system is shown. The field lines were calculated for 5.3 UT on September 5, 1982 using the Tsyganenko model T 96-1. The footpoints are located at 53.5, 56.6 and 58.4 invariant latitude, respectively. RC stands for ring current and SETE for subauroral electron temperature enhancement. (From Prölls, 2006b)

a few hours. At the same time, the centroid of the particle distribution shifts towards Earth. Accordingly, the magnitude of the heating effect should also increase during disturbed conditions, and its location should move towards lower latitudes. Both effects are observed.

In Fig. 6 we have plotted the magnitude of the temperature peak as a function of the Dst index. The general format of data presentation is similar to that of Fig. 4 except that this time we use the Dst index, which is a good indicator of the ring current intensity. As is evident, a fairly good linear correlation exists between this index and the magnitude of the heating effect. During quiet conditions, the temperature at the location of the peak is of the order of 2100
K, on average. For each decrease of the Dst index by 10 nT, the temperature increases by 100 K. For a large geomagnetic storm of -300 nT, this would lead to a peak temperature of about 5000 K, and such large temperatures are observed.

There is also a very good correlation between the Dst index and the location of the temperature peak (Prölls, 2006b). During quiet conditions, the temperature peak is located slightly poleward of 60 degrees invariant latitude. For each decrease in the Dst index by 10 nT, the temperature peak moves towards lower latitudes by about 0.8 degrees. For a large geomagnetic storm of -300 nT, this would lead to a total shift of 24 degrees, in agreement with observations.

2.3 Spatial correlation between the subauroral electron temperature enhancement and the subauroral ionization trough

The scenario depicted in Fig. 5 offers just one possible explanation for the subauroral electron temperature enhancement. An alternate explanation is based on the observation that in almost all cases SETE effects are located in the immediate neighborhood of the ionospheric trough. In fact, on average, both phenomena are exactly co-located. This is documented in Fig. 7. Plotted is the mean latitudinal profile of the subauroral electron temperature enhancement, together with the associated changes in the electron density. The excellent spatial correlation of both phenomena is obvious. Given the decrease of the electron density in the trough, even a uniform distribution of the magnetospheric heat source should lead to a localized increase in the electron temperature. This is because each electron in the trough region will get a larger share of the heat addition. Moreover, the heat loss will be reduced because of the lower ion density. A more detailed discussion of this topic can be found in Wang et al. (2006).

3 Positive ionospheric storms at middle latitudes

In contrast to the situation in the polar region, the morphology of ionospheric storms at middle latitudes is well documented. Basically, the ionization density may either increase or decrease during disturbed conditions, and these changes are traditionally designated as positive and negative ionospheric storms, respectively. Using ground-based ionosonde measurements, both disturbance effects are illustrated in Fig. 8. Here the maximum ionization density of the F2-layer serves as a convenient parameter to specify the ionospheric behaviour. Similar results are obtained if the maximum ionization density is replaced by the column density or total electron content (TEC) of the ionosphere, as is
Figure 6: Magnitude of the subauroral electron temperature enhancement as a function of geomagnetic activity. The Dst index is indicated on the abscissa, the associated peak temperature Tem on the ordinate. Both parameters were determined for a total of 1159 electron temperature profiles. Only data obtained in the 21 to 03 h magnetic local time sector were considered. Also, all temperature values were adjusted to a common altitude of 600 km. For each 10 nT wide interval of the Dst index, the median and upper and lower quartiles of the magnitudes of the temperature peak were determined; these are shown by the dots and bars in the figure. Also shown is the regression line fitted to the medians. The associated line parameters are given in the lower left-hand corner (ordinate interception \( a \) in K, slope \( b \) in K/nT). (From Pröß, 2006b)

frequently the case today.

Whereas the origin of negative ionospheric storms is well understood (it is attributed to changes in the neutral gas composition of the upper atmosphere; see, for example, Pröß and Werner, 2002), this is not the case for positive ionospheric storms. In fact, people have been trying to understand this latter phenomenon for more than 70 years without coming up with a generally accepted explanation. Some of the most frequently discussed mechanisms are listed in Fig. 9.

It has been repeatedly suggested that positive ionospheric storms are also
caused by changes in the neutral gas composition (e.g. Rishbeth, 1991). Thus far, however, no experimental evidence has been presented to support this supposition for more severely disturbed conditions. The moderate increase in the $O/N_2$ density ratio frequently observed at middle or lower latitudes is certainly not sufficient to explain the rather large positive ionospheric storms sometimes observed in these regions.

In contrast, the idea that positive ionospheric storms are caused by upward transport of ionization is well supported by observations. Consider, for example, the data presented in Fig. 10. In response to a sudden energy injection at polar latitudes (as indicated by the AE index), a well-defined positive ionospheric storm develops at middle latitudes. The immediate cause of this density increase is a sudden uplifting of the F layer, as demonstrated in the middle part of this figure.

That an increase in layer height will lead to positive storm effects is easily understood if the height dependences of the ionization production and loss
Figure 8: Magnetic storm-induced changes in the ionospheric electron density. F2-layer maximum electron densities ($N_{\text{max}}$, heavy dotted lines) measured at two mid-latitude stations during the magnetic disturbance event of February 21 and 22 of 1973 (upper panel) illustrate positive and negative ionospheric storm effects (middle and lower panels). The monthly medians serve as quiet-time references (thin lines). Brisbane is located at 27°S, 152°E, and Pt Arguello at 36°N, 239°E. (From Prölls, 1980)

According to basic theory, the loss rate is proportional to the molecular nitrogen and molecular oxygen densities (e.g. Prölls, 2004). Therefore, it decreases much faster with height than the production rate, which is proportional to the atomic oxygen density. An upward displacement of the
The big and as yet unanswered question is what causes this upward motion of the ionosphere? Here the scientific community is divided into two factions. One advocates winds, the other one electric fields as the primary cause. Both mechanisms are illustrated in Fig. 11. In the case of equatorward directed winds, we expect a drag-induced drift of the ionospheric plasma along the magnetic field lines which is directed upward and equatorward. In the case of an eastward directed electric field we expect an \( \vec{E} \times \vec{B} \) drift of the ionospheric plasma which is directed upward and poleward. Therefore measuring the local drift vector should tell us which mechanism is at work. Unfortunately, unambiguous measurements of this kind are scarce. Another way to single out the correct explanation would be to measure the zonal electric fields and the meridional winds directly. Such measurements during positive ionospheric storm conditions appear to be even rarer.

Even if winds are identified as the primary cause of positive ionospheric storms, there remains the question as to the origin of these winds. Currently, two different types of wind perturbations are discussed in the literature. First, changes in the large-scale wind circulation, and second, so-called travelling atmospheric disturbances (TADs). In the first case, storm-induced heating of the high-latitude upper atmosphere drives a large scale Hadley-type wind circulation which extends from high to low latitudes. Such winds may be responsible for large-scale, long-duration positive ionospheric storms like that illustrated in

Figure 9: Mechanisms contributing to the positive phase of ionospheric storms at middle latitudes.
Figure 10: Short-duration positive ionospheric storm in the local afternoon sector. In response to an isolated burst of substorm activity on January 23, 1973 (hourly-averaged AE index, upper panel), the ionosonde at Slough (50° invariant magnetic latitude) first observes an impulse-like uplifting of the F2-layer (dhm, middle panel), and subsequently an impulse-like increase in the ionization density (dNm, middle panel, and Nm, lower panel). Data recorded on January 22 serve as a quiet-time reference (dotted lines). (From Prölss, 1993)

Fig. 8.

The second case is illustrated in Fig. 12. During a magnetic substorm
Figure 11: Uplifting of the ionospheric F-layer by (a) an equatorward directed wind, and (b) by an eastward directed electric field. The situation illustrated applies to the northern hemisphere. $B$ is the Earth’s magnetic field, and $u_z$ the upward directed component of the drift velocity.

(indicated by an increase in the AE index) energy is injected into the polar upper atmosphere. This sudden energy addition launches a so-called travelling atmospheric disturbance (TAD). By this we mean a pulse-like atmospheric perturbation formed by a superposition of gravity waves which propagates with high velocity (here, 600 m/s) towards the equator. An essential feature of such a TAD is that it carries along equatorward-directed winds of moderate magnitude (here, 150 m/s). At middle latitudes these meridional winds cause an uplifting of the F2 layer which in turn leads to an increase in the ionization density. This mechanism is better suited to explain short-duration positive ionospheric storms like that illustrated in Fig. 10.

On the other hand, if electric fields turn out to be the primary cause of positive ionospheric storms, we are still faced with the question as to the origin of these fields. Are they externally imposed on the ionosphere or are they generated by neutral winds in the dynamo region of the lower ionosphere? And if they are externally imposed, do they propagate from high to low latitudes or are they produced by polarization effects in the inner magnetosphere? On the
other hand, if dynamo effects predominate, are they caused by changes in the large-scale wind circulation or by travelling atmospheric disturbances?

Horizontal transport of the ionospheric plasma may also be important. Foster and co-workers, for example, promote the idea that a special kind of positive storm effect, called either dusk effect (e.g. Mendillo, 2006) or storm-enhanced density (SED; e.g. Foster et al., 2004), is produced by horizontal advection of high density plasma from lower to higher latitudes. They attribute this plasma transport to a combination of poleward and eastward directed electric fields.

Evidently, an evaluation of the many different possibilities listed in Fig. 9 requires simultaneous electric field and wind measurements of sufficient temporal and spatial resolution. Presently, such measurements are not available.

4 Ionospheric holes at equatorial latitudes

Ionospheric ‘holes’ certainly belong to the most spectacular storm phenomena at low latitudes. An example of this kind of perturbation is presented in Fig. 13. The upper part shows a subsatellite track of the DMSP-F15 satellite during the storm event of October 29/30, 2003. This was one of the largest solar-terrestrial storms ever observed, with an X17/4B flare on October 28, a solar wind velocity of nearly 2000 km/s and a minimum Dst index of less than -350
Figure 13: Storm-induced ionospheric hole in the equatorial ionosphere. The upper part of this figure shows a subsatellite track of the DMSP-F15 satellite during the large geomagnetic storm of October 30, 2003. In the lower part, the ion density measured along the associated satellite trajectory is plotted as a function of geographic latitude. Note that this density is on a logarithmic scale. The solar local time of observation is approximately 21:30 h, the observation height 840 km. An arrow indicates the location of the magnetic equator. (After Kil et al., 2006b)

nT. In the lower part of this figure, the ion density measured along the satellite trajectory is plotted on a logarithmic scale. The densities refer to an altitude of 840 km and have been plotted on a logarithmic scale.
The most striking feature of this data set is the large drop in the ion density near the magnetic equator. This ionospheric hole is characterized by rather steep walls on both sides of the density depletion, by a flat and almost structureless bottom, and by densities which are less than 1/100 of the values recorded outside the hole. Up to today, similar density dropouts have been observed during at least six different geomagnetic storms, all of which were unusually large (e.g. Greenspan et al., 1991; Basu et al., 2001; Lin et al., 2001; Lee et al., 2002; Su et al., 2002; Vlasov et al., 2003; Lin and Yeh, 2005; Kil et al., 2006a,b). Also, all these observations were made after sunset, indicating that we are dealing here with a nighttime phenomenon.

As to the origin of this spectacular disturbance phenomenon, different mechanisms have been proposed. One group of authors suggests that an equatorial super-fountain removes the ionization from the equatorial region (e.g. Greenspan et al., 1991; Batista et al., 1991; Basu et al., 2001). Others assume that the ionospheric layer rises above the observation heights of the satellites (Su et al., 2002). Some support for this latter explanation comes from ionosonde measurements which show that the layer height may indeed increase to high altitudes (Su et al., op. cit.; Sobral et al., 1997; Abdu et al., 2005). Sunward convection of the nighttime plasma represents another possibility (Vlasov et al., 2003). Finally, Kil et al. (2006a, b) propose that ionospheric holes are formed when so-called ionospheric bubbles merge to form a big bubble. Evidently, simultaneous measurements of all the key parameters involved are needed to single out the correct explanation.

As we have seen, this situation is not uncommon in ionospheric storm research. Some of the storm effects have been known for decades and are well documented. Also various mechanisms have been proposed to explain the perturbations. What is missing are suitable measurements which allow us to single out the correct explanation(s). Therefore progress in this field will critically depend on better and more comprehensive measurements. In this sense, ionospheric storms represent a real challenge for experimentalists.

Acknowledgments. The DE-2 data used in this study were kindly provided by the NASA National Space Science Data Center. The ionosonde data were obtained from World Data Center A. I am grateful to all the experimenters who contributed to these data sets. I am also indebted to M. Hanussek and K. Schrüfer for their help in preparing this manuscript.
References


