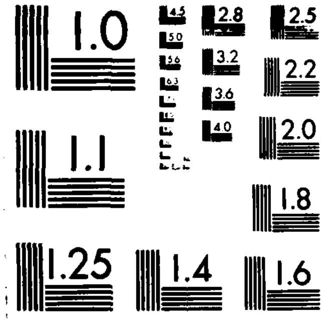
AD-A156 264 A STUDY OF MESOSCALE PROBABILITY FORECASTING 1/1
PERFORMANCE BASED ON AN ADVA..(U) NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION BOULDER CO EN..
UNCLASSIFIED G M WILLIAMS ET AL. 30 APR 84 F/G 4/2 NL

END
DATE
FILMED
8-85
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

AFGL-TR-84-0142

A STUDY OF MESOSCALE PROBABILITY FORECASTING PERFORMANCE BASED ON AN ADVANCED IMAGE DISPLAY SYSTEM

Gary M. Williams, Daniel Birkenheuer, and Duane A. Haugen

Program for Regional Observing and
Forecasting Services (PROFS)
NOAA/ERL/ESG, R/E23
325 Broadway
Boulder, CO 80303

April 30, 1984

Final Report
1 November 1981 - 30 September 1983

Approved for public release; distribution unlimited.

Air Force Geophysics Laboratory
Air Force Systems Command
United States Air Force
Hanscom AFB, Massachusetts 01731

"This technical report has been reviewed and is approved for publication"

Donald A. Chisholm
DONALD A. CHISHOLM
Contract Monitor

Donald A. Chisholm
DONALD A. CHISHOLM, Chief
Atmospheric Prediction Branch

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

FOR THE COMMANDER

Robert A. McClatchey
ROBERT A. McCLATCHEY, Director
Atmospheric Sciences Division

This report has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFGL/DAA, Hanscom AFB, MA 01731. This will assist us in maintaining a current mailing list.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFGL-TR-84-0142

6a. NAME OF PERFORMING ORGANIZATION
NOAA/ERL/ESG,R/E23
325 Broadway

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Air Force Geophysics Laboratory

6c. ADDRESS *(City, State and ZIP Code)*
Boulder, CO 80303

7b. ADDRESS *(City, State and ZIP Code)*
Hanscom AFB, Massachusetts 01731
Monitor/Donald Chisholm/LYP

8a. NAME OF FUNDING/SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
ESD3-0800

8c. ADDRESS *(City, State and ZIP Code)*

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
62101F	6670	10	BC

11. TITLE *(Include Security Classification)*
(over)

12. PERSONAL AUTHOR(S)
Gary M. Williams, Daniel Birkenheuer, and Duane A. Haugen

13a. TYPE OF REPORT
Final Report

13b. TIME COVERED
FROM 1 Nov 81 TO 30 Sep 83

14. DATE OF REPORT *(Yr., Mo., Day)*
1984 April 30

15. PAGE COUNT
54

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Forecast Display Systems, Forecasting and Observation Systems, Mesoscale Forecasting, Thunderstorm Forecasting, Mesoscale Observing Systems

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Forecast experiments were conducted at PROFS in the summers of 1981, 1982, and 1983 to determine the impact of an advanced technology workstation on forecaster performance. In each experiment, probability forecasts for wind, visibility and precipitation in northeast Colorado and severe weather warnings were issued and evaluated. The major efforts over these years have been to develop the sophisticated technology ~~that PROFS now uses~~ to conduct forecast exercises. Some of this development effort is outlined

(continued on back)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Donald A. Chisholm

22b. TELEPHONE NUMBER *(Include Area Code)*
(617) 861-4729

22c. OFFICE SYMBOL
AFGL/LYP

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

Block 11. A Study of Mesoscale Probability Forecasting Performance
Based on an Advanced Image Display System

Block 19, continued:

in this report. The results of the forecast exercises have been increasingly more convincing that the advanced ~~PROFS~~ forecaster workstations provide substantial improvement in forecast accuracy for short-term, site-specific forecasts. Various probability forecast exercises that have been incorporated in these experiments are presented in this report.

Probability forecasts of precipitation, a major part of the 1983 experiment, are the most recent set of results available. Brier score and skill score results are presented for the precipitation forecasts, comparing the forecasters to persistence, sample climatology and conditional climatology. The results show positive forecaster skill. Further improvements in skill are to be expected as future experimental plans are implemented. Keywords

back

TABLE OF CONTENTS

1. INTRODUCTION 1

2. THE 1981 FORECAST EXERCISE 4

3. THE 1982 FORECAST EXERCISE 8

4. THE 1983 FORECAST EXERCISE 15

4.1 Data and Products Available 16
4.2 Forecast Procedures 18
4.3 Forecast Evaluation 21
4.3.1 Wind forecast evaluations 23
4.3.2 Visibility forecast evaluations 26
4.3.3 Precipitation forecast evaluations 27

5. CONCLUSIONS 40

6. ACKNOWLEDGMENTS 41

7. REFERENCES 42

FIGURES

Page

Figure 1. Brier scores for the 20-kt wind speed category, 1981 exercise 16

Figure 2. Brier scores for the 20-kt wind speed category, initial condition winds < 20 kt, 1981 exercise 16

Figure 3. Brier scores for 20-kt wind speed category, initial condition winds > 20 kt, 1981 exercise 17

Figure 4. Schematic diagram of the PROFS Exploratory Development Facility (EDF) as configured for the 1983 exercise 22

Figure 5. Forecast area map for 1983 exercise 24

Figure 6. (a) Forecaster reliability chart for precipitation $\geq$ 0.01 inch. 42
(b) Forecast frequency plot for associated reliability plot. Forecast period: 0-2 hours 42

Figure 7. (a) Forecaster reliability chart for precipitation $\geq$ 0.01 inch. 43
(b) Forecast frequency plot for associated reliability plot. Forecast period: 2-4 hours 43

Figure 8. (a) Forecaster reliability chart for precipitation $\geq$ 0.01 inch. 43
(b) Forecast frequency plot for associated reliability plot. Forecast period: 4-6 hours 43

Figure 9. (a) Forecaster reliability chart for precipitation $\geq$ 0.01 inch. 44
(b) Forecast frequency plot for associated reliability plot. Forecast period: 6-8 hours 44

TABLES

Page

1. Data sets available in the 1981 forecast exercise.................. 7

2. Data sets available in the 1982 forecast exercise.................. 10

3. Data sets available in the 1983 forecast exercise.................. 20

4. Forecaster and persistence Brier scores vs. forecast period for winds $\geq$ 25 kt. Initial wind speeds < 25 kt........................ 29

5. Forecaster and persistence Brier scores vs. forecast period for winds $\geq$ 25 kt. Initial wind condition > 25 kt..................... 30

6. Forecaster skill scores relative to persistence vs. forecast period for wind speeds $\geq$ 25; light and strong wind initial conditions....... 30

7. Forecaster and sample climatology Brier scores and forecaster skill scores vs. forecast period for precipitation $\geq$ 0.01 in. All cases... 34

8. Forecaster and sample climatology Brier scores and forecaster skill scores relative to climatology vs. forecast period for precipitation $\geq$ 0.01 in. Rain-verified cases only.............................. 35

9. Forecaster, conditional climatology, sample climatology, and persistence Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. All verification cases............ 36

10. Forecaster, conditional climatology, and sample climatology Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. Rain-verified cases only................ 36

11. Forecaster, conditional climatology, sample climatology, and persistence Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. All verification cases............... 37

12. Forecaster, conditional climatology, and sample climatology Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. Rain-verified cases only.................. 38

13. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain condition. All verification cases.................................. 38

14. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain condition. Rain-verified cases only................................ 39

15. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain condition. All verification cases.................................. 39

16. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain condition. Rain-verified cases only................................ 39

1. INTRODUCTION

This is the final report of a contractual effort by the Program for Regional Observing and Forecasting Services (PROFS) for the Meteorology Division of the Air Force Geophysics Laboratory (AFGL). The contract was in effect from 1 November 1981 to 30 September 1983. However, we have delayed this report to include the results from a forecast exercise which ended mid-August 1983.

The purpose of the contract was to examine the impact of an advanced technology forecaster workstation on forecaster performance. The original plan was to accomplish this objective by evaluating documented forecaster appraisals of the system characteristics and by analyzing and evaluating forecaster skill scores obtained from forecast experiments. The realities of the evolutionary process PROFS has gone through during and after the period covered by this contract have modified this original plan considerably. Forecaster appraisals of the system have indeed led to significant engineering improvements in the system, but there has been neither time nor resources to devote to systematic documentation of these appraisals. Forecasting experiments have taken place, however, with continually increasing sophistication. This report will summarize the evolutionary growth of the PROFS system capabilities, present results of forecaster skill evaluations, and provide references to other reports and articles that contain pertinent material on the PROFS program.

PROFS was first organized in October 1979 as part of the Environmental Research Laboratories (ERL) of the National Oceanic and Atmospheric Administration (NOAA) to evaluate the improvement of local, short-term forecasts attributable to the use of advanced image and graphics display capabilities.

The period covered by this contract was a period of explosive growth for PROFS. In three years the staff grew from five to eighty people. The system capabilities increased from one data acquisition and recording computer to an extensive data acquisition, processing, and display capability based on a networked computer system of about 22 microprocessors, 10 minicomputers, and 5 midicomputers. This growth was essential to the central objective of the program. However, it has also limited the evaluation efforts to some extent, viz., it was necessary to design forecaster exercises in conditions of ever-changing system capabilities.

Three forecast exercises were conducted during the period of this contract: a displaced real-time experiment in late 1981 based on convective storm data collected during the summer of 1981; a real-time exercise in the summer of 1982; and a second real-time exercise in the summer of 1983. The evaluations based on these have provided quantitative and qualitative evidence for improvements in forecast accuracies. Moreover, considerable experience was gained in the design and operation of real-time image and graphics display systems, particularly in the selection of display products most useful to forecasters for local, short-term forecasts of severe weather.

Four types of data form the basis for the PROFS system: weather radar images; satellite visible and infrared images; National Weather Service surface observations, upper-air observations, and all the analyses and prognostic products prepared by the National Meteorological Center; and local mesoscale networks of surface observations. There has been considerable variation in the frequency and reliability of observations during the course of the three experiments. However, there are several system attributes that have been available from the onset and are important to understanding the expectation of improved forecast accuracies.

First, the system permits the integration and merging of information from a diverse set of data sources. For example, radar and satellite images can be combined; station plots can be overlaid on image data; and streamlines, isobars, and isotherms can be overlaid on image data. Second, the system provides rapid updates of information, an essential feature for observing and forecasting local, short-lived weather phenomena. Third, the system displays are in color, which permits a wide array of color enhancements of the information content in satellite and radar images. Fourth, the system can loop forward and backward through a series of images and graphics, permitting rapid assimilation of evolutionary processes by the forecaster. Fifth, zooming on selected features of the display and roaming across zoomed images are built-in features of the display system. Last, the workstation has become increasingly user-friendly over the last three years, an important attribute for forecaster use. These are the essential features of the advanced workstation. Some of them have been improved from 1981 to 1983, but the basic capabilities implied by these features have been available for each of the experiments.

It is useful to contrast these system attributes with the current technology presently available in many of the National Weather Service Forecast Offices (WSFO's) and planned for most USAF base weather stations (BWS's) in the late 1980's. The most prevalent NWS technology is the Automation of Field

Operations and Service (AFOS) system[1] while in the USAF it will be the Automated Weather Distribution Systems (AWDS). These systems provide the meteorologist with conventional surface and upper air observations and centralized products, analyses, and forecasts from the National Meteorological Center (NMC) and the Air Force Global Weather Central. The AFOS system is augmented by local radar data at all WSFO's responsible for severe weather warnings. Many WSFO's and BWS's display radar reflectivity data in the form of Plan Position Indicators (PPI's) on a color radar display system such as the Kavouras system.[2] Satellite data are available at WSFO's and BWS's on hard-copy satellite photographs from Unifax, which provides either visible or infrared images every 30 minutes and occasional water vapor images. Merging of images; overlaying of graphics and alphanumerics on images; and looping, zooming, and roaming are not possible with these sets of data. Each data set is independent of the other. It is thus natural to anticipate improved forecast accuracies with the advanced technology workstation.

2. THE 1981 FORECAST EXERCISE

During the summer of 1981, there was only one computer system available for data ingest and archiving. The basic data ingest computer was a VAX-11/750 with 2.5 megabytes of memory and a 256 megabyte hard-disk mass data storage device system. This system was capable of ingesting data in real-time and archiving

---

[1] Klein, W. H., 1976: The AFOS program and future forecast applications. Mon. Weather Rev., 104, 1494-1504.

[2] Design News, 1980: Color adds "pizzaz" to weather radar. 26, 13-14.

them on disk packs, but it could not support real-time processing and display of data. These functions were satisfied by writing software that could display data on a color display screen (Ramtek 9351) at a later time.

Product disk packs were prepared for a few selected days that could be accessed later by a forecaster workstation in a real-time playback mode. These data sets were used to design and conduct a displaced-real-time test in the late fall of 1981. Ten days were selected for study. Experienced forecasters from the private sector, the National Weather Service, and the Air Weather Service each spent two weeks in our forecast office going through real-time forecast exercises. Each forecaster was given three responsibilities in these exercises: prepare severe storm watches and warnings as deemed necessary; prepare probability point forecasts for selected stations in the PROFS computer-controlled mesoscale network of surface observations;[3] and prepare a subjective critique of the workstation capabilities. This critique addressed such points as the image and graphic products available, the coordinate systems in which these products were presented, the ease of operation of the workstation, and general appraisals of the strengths and weaknesses of the workstation system.

The test forecasts were prepared for eastern Colorado. Watches and warnings were defined in accordance with standard NWS practices. This included specifying type of forecast (watch or warning), type of event (severe thunderstorm, tornado, or flash flood), time of issuance, effective time and

---

[3] Pratte, J. F., and R. F. Clark, 1983: PROFS Mesonet-description and performance. Preprints, 5th Symp. on Meteorological Observations and Instrumentation. Am. Meteorol. Soc., Boston, Mass.

duration of the watch or warning, and the watch or warning area. Severe weather verification data were assembled for the purpose of evaluating these watches and warnings.

The point probability forecasts were made hourly during each test day for four forecast periods, each of 15 minutes duration ending 15, 30, 60, and 120 minutes after forecast time. The forecast elements were wind speed, visibility, and rainfall. For each period, the forecaster tabulated probability values for the following categories: wind speed $>$ 20 knots, wind speed $>$ 35 knots, wind speed $>$ 50 knots; visibility $<$ 3 miles, visibility $<$ 1 mile; rainfall $<$ 0.01 inches, 0.01 $\leq$ rainfall $<$ 0.04 inches; 0.04 $\leq$ rainfall $<$ 0.08 inches, rainfall $\geq$ 0.08 inches. The probabilities assigned to the four rainfall categories had to total 100 percent. Independent exceedance forecasts were made for each of the wind and visibility categories. The point probability forecasts were made for each of two stations in the mesoscale network which were specified for the forecaster at the beginning of each test day. Verification of the point forecasts was based on the surface network observations during the 15-minute periods immediately preceding the end of each forecast period.

The data sources and their update rates for this exercise are listed in Table 1.

Table 1. Data sets available in the 1981 forecast exercise.

Data Sets	Update Rate (min)
Digital GOES satellite imagery (Infrared and visible images)	30
Digital radar PPI's from NWS radars at Limon, CO and Cheyenne, WY	30
Surface mesonet observations of 5-min average values of wind, temperature, dewpoint, and visibility. Accumulated (5-min) values of precipitation.	5

This data set was of limited value for quantitative evaluation of forecast results. The update rates, particularly for the radar data, were simply too low to make warnings. Oftentimes, a storm appeared during the interval between data updates and had become severe by the time the forecaster first observed it on a radar image. Furthermore, there were many data reliability problems which arose from the difficulties of bringing a new automatic real-time data-handling system on-line for the first time. The visibility and precipitation verification data were not usable because of various calibration, siting, and equipment malfunction problems. The wind forecasts, however, did provide some useful information, but only for the category of greater than 20 knots. Wind gusts in excess of 35 knots occurred so infrequently that no useful forecast data were obtained for the higher two wind categories.

In addition to the point probability forecasts prepared by the forecasters, a climatological forecast algorithm furnished by AFGL was tested. The model was patterned after one developed by Gringorten[4] and was based on climatological

---

[4] Gringorten, I.I., 1972: Conditional probability for an exact non-categorized initial condition. Mon. Wea. Rev., 110, 796-798.

records from Buckley Air Force Base near Denver, Colorado, with terrain features similar to those mesonet stations used in the forecast exercise. Evaluation of this model over all the forecast periods showed substantial forecast skill for the greater-than-20 knots wind speed category. It is widely held that climatological models are valuable in making probability forecasts.[5] The limited results from this exercise subjectively confirmed that expectation. However, because of exercise scheduling problems the model was not furnished to the forecasters as guidance, and only a limited amount of forecast data were available to evaluate. It was therefore not possible to obtain quantitative evaluation results for the 1981 exercise.

The most valuable results to come from the 1981 exercise derive from the critiques of the workstation prepared by the forecasters. As one might expect, given the limited capabilities inferred from Table 1, there was widespread criticism of deficiencies in data type, data update rate, and in data reliability. Furthermore, the workstation was not considered user-friendly by most of the forecasters. Product display was controlled by keyboard entry of mnemonics similar to those in the AFOS system. The list of mnemonics was fairly short and detailed instruction sheets were readily available. Nevertheless, we concluded that future workstation development must be more user-friendly. These deficiencies were substantially corrected in the implementation of the system used in the 1982 exercise.

---

[5] Hughes, L. A., 1980: Probability forecasting - reasons, procedures, problems. NOAA Tech. Memo, NWS FCST 24.

In summary, the 1981 exercise was largely a learning experience. Quantitative forecast evaluation was simply not warranted, but subsequent system design and implementation efforts were substantially influenced by the expert opinions of the visiting forecasters.

3. THE 1982 FORECAST EXERCISE

By May 1982, system capabilities had been sufficiently expanded to warrant a real-time forecast exercise. Two separate workstations were installed in the PROFS forecaster office in Boulder. The advanced workstation incorporated color image and graphic display technology (Ramtek 9400); the other incorporated the AFOS system and Kavouras radar display monitor currently available in many NWS Forecast Offices (WSFO's). We also installed an identical advanced workstation at the Denver WSFO where it was merged into their severe weather warning station rather than treated as a separate workstation. The products displayed at the advanced workstation were generated by a large distributed-processing, networked computer system.[6] A detailed description of the advanced workstation and the products available in 1982 has been published by Reynolds.[7] Major components of the system included two VAX-11/780's (one for data ingest; one for product generation) and a PDP-11/24 to drive the displays at the forecaster workstation.

Data ingest capabilities were expanded to include pressure sensors in the surface network, an air-to-ground lightning detection system, and NWS

---

[6] Brown, R. C., 1983: Anatomy of a mesoscale instrumentation system. Preprints, Fifth Symposium on Meteorological Observations and Instrumentation, Am. Meteorol. Soc., Boston, 308-313.

[7] Reynolds, D. W., 1983: Prototype workstation for mesoscale forecasting. Bull. Am. Meteorol. Soc., 64, 264-273.

surface and upper-air observations that were obtained from the FAA 604-line, a national meteorological data communication system supported by the Federal Aviation Administration. Data sources and their update rates are given in Table 2. A most significant advance over the 1981 system was attaining the system capability to ingest the indicated data at the rates shown and to generate and display image and graphic products in real time. Further, the workstation was much easier to operate. Products were selected from a separate menu screen, activated by a light pen. This eliminated the operator frustrations of learning mnemonic codes to call up products as was done with the 1981 system.

Table 2. Data sets available in the 1982 forecast exercise.

Data Sets	Update rate (min)
Digital GOES satellite imagery (Infrared and visible images)	30
Digital radar PPI's from NWS radars at Limon, CO, and Cheyenne, WY	5
Surface mesonet observations of 5-min averaged values of wind, temperature, dew point, pressure, and visibility. Accumulated (5-min) values of precipitation.[5]	5
Air-to-ground lightning detection (location and frequency)[6]	5
NWS Surface Aviation Observations (SAO's)	60
NWS upper-air (rawinsonde) observations	7200

The second workstation included one alphanumeric and three graphic AFOS terminals which operated in a receive-only mode off the Denver WSFO distribution point. This system was augmented by a Kavouras color radar display system (dedicated line to the NWS radar at Limon, CO) and a Unifax satellite hard-copy device. The Kavouras system provided automatic updates of radar images every three minutes at four ranges: 60, 120, 180, and 240 nautical miles. The Unifax provided either visible or infrared satellite images every 30 minutes and occasional water vapor images.

In this report, the combined capabilities of the AFOS terminals, the Kavouras radar display, and the Unifax hard copy system will be referred to as the AFOS station. The advanced workstation will be referred to as the PROFS workstation. There were three workstations in the 1982 experiment: the PROFS workstation and the AFOS station in the PROFS forecast office in Boulder, and a merged AFOS and PROFS workstation in the Denver WSFO.

There were three goals of the 1982 exercise: (1) to provide an opportunity for senior research meteorologists (meteorologists with many years experience in research but limited experience in operational forecasting) to use and to constructively criticize the advanced workstation; (2) to train junior PROFS meteorologists (limited experience in either research or operational forecasting) in the use of the workstation and in the procedures and decision-making processes of preparing severe weather warnings and local forecasts; and (3) to evaluate the workstation and its capabilities as an aid in improving the timeliness and accuracy of short-term forecasts. Evaluation procedures were designed to compare like forecasts prepared at the various workstations.

Two types of forecast exercises were conducted, from 1 June through 15 August 1982. One was to prepare severe storm warnings as deemed necessary at each PROFS workstation. This permitted a comparison of warnings from the PROFS workstation in Boulder, the AFOS workstation in Boulder, and the merged AFOS/PROFS workstation in Denver where the official NWS severe weather warnings were issued.

The other was to prepare point probability forecasts of wind, visibility, and precipitation for two selected stations of the mesoscale network following the procedures already described for the 1981 exercise. (The same two stations were used throughout the 1982 exercise.) These forecasts were not made at the Denver WSFO, only at the Boulder workstations. As indicated in Table 2, the surface network data were available to the PROFS forecaster every five minutes. These data included such display products as station plots, streamlines, isotherms, and time series of selected parameters. The AFOS forecaster, on the other hand, was given only a tabulation of the hourly observations of wind, visibility, and precipitation for the two selected stations.

The results of this exercise are published in a NOAA Technical Report.[8] A brief summary of the major conclusions follows.

Quantitative evaluation results were obtained which suggest that improvements in severe weather warning accuracies can occur as a result of the advanced technology. Further, the relative forecaster experience levels of the participants in the exercise were detected and measurable. The junior

---

[8] Haugen, D. A., D. Birkenheuer, R. F. Bunting, M. Cairns McCoy, G. M. Williams, 1984: An evaluation of PROFS' 1982 real-time forecast experiment, NOAA Tech Report ERL 427-ESG1, ERL, Boulder, CO.

meteorologist, while helped by the advanced technology, was considerably less accurate than the highly experienced forecaster at the WSFO using the merged AFOS/PROFS workstation. However, it was not possible to distinguish between the relative value of forecaster experience and the advanced workstation capabilities. Furthermore, it was not possible to obtain results supported by tests of statistical significance due to various imperfections in the experiment design (discussed in Haugen et al.[8]) and to small sample sizes, which are an inherent characteristic of forecast experiments that focus on severe convective storms occurring infrequently in time and space.

The Brier score[9] was used for comparing point forecasts. It is computed from

$$BS = \frac{1}{N} \sum_{i=1}^{N} (F_i - O_i)^2 \qquad (1)$$

where F_i is the forecast probability; O_i is the observed probability, equal to one if the event occurred, zero if it did not; and N is the total number of forecasts. Brier scores range from zero for a set of perfect forecasts to one for a set of complete misses.

Only those cases for which the category value of the parameter for the forecast period differed from the value at the time of the forecast were

---

[9] Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. Mon. Weather Rev., 78, 1-3.

examined, i.e., persistence forecasts would not verify. This emphasizes those weather situations when the parameter values are changing, usually because of nearby convective events. These are expected to be the most difficult conditions to forecast and, hence, should have the greatest potential for discerning differences between the AFOS and PROFS workstations.

Change in wind speed category can occur from the influence of thunderstorms or more widespread mesoscale or synoptic-scale influences. In Colorado summers, a change in visibility or precipitation categories generally requires a rain cell over the station. Since the elements forecast in this experiment occurred rarely, useful statistics only for the category of wind speed greater than 20 knots could be obtained. No other categories occurred more than 12 times throughout the summer.

Analyses of the wind speed forecasts were based on three groupings of the forecasts: (1) all cases where the observed category for the forecast period differed from the observed category at the time the forecast was issued; (2) those cases where the wind speed at the forecast time was less than 20 knots but greater than 20 knots for the forecast period; and (3) those cases where the wind speed at the forecast time was greater than 20 knots but less than 20 knots for the forecast period. Note that the cases of groups (2) and (3) together constitute the cases of group (1).

The Brier score results for each of the three groupings for the wind speed forecasts are given in Figures 1, 2, and 3. The Brier scores for the PROFS workstation forecaster are consistently better than for the AFOS forecaster for all three groupings. This tends to support the hypothesis that the PROFS workstation capabilities provided improved short-range forecast accuracies. In particular, the data integration and display capabilities, especially the looping and overlay of mesonet data and the display of five-minute interval time series of simple station mesonet data, are probably the capabilities that led to the improved skill at the PROFS workstation. None of these capabilities were available to the AFOS forecaster. It should be recalled that the AFOS workstation had mesonet data only for the two stations being forecasted for and only the most recent data once an hour. Furthermore, the AFOS forecaster could not overlay these data on the Kavouras screen where radar images were being displayed. Nor was there a capability to loop or display radar images in sequence. This severely limited the AFOS forecaster relative to the PROFS workstation forecaster. The sample size, however, is too small to infer a statistically significant difference between the POWS and AFOS results.

It is seen that the skill was low, and worsened with length of forecast period for the cases where the wind speeds were initially less than 20 knots. The inverse is true for those cases where the wind speeds were initially greater than 20 knots. That is, forecasting the onset of strong winds one to two hours in advance is considerably more difficult than forecasting diminishing winds. This illustrates the difficulty of forecasting rare convective events. Indeed, none of the participants had previous experience in either probability forecasting or short-term forecasting of these particular elements. This led to the decision to use more experienced forecasters in the 1983 experiment.

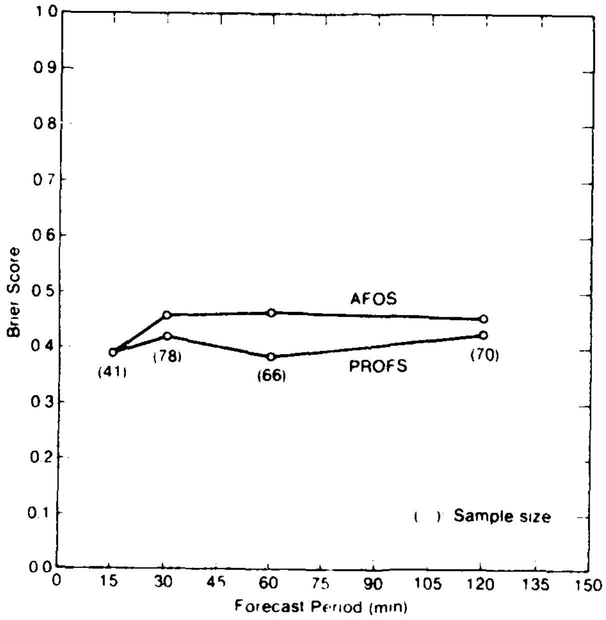


Figure 1. Brier scores for 20-kt wind speed category, 1981 exercise.

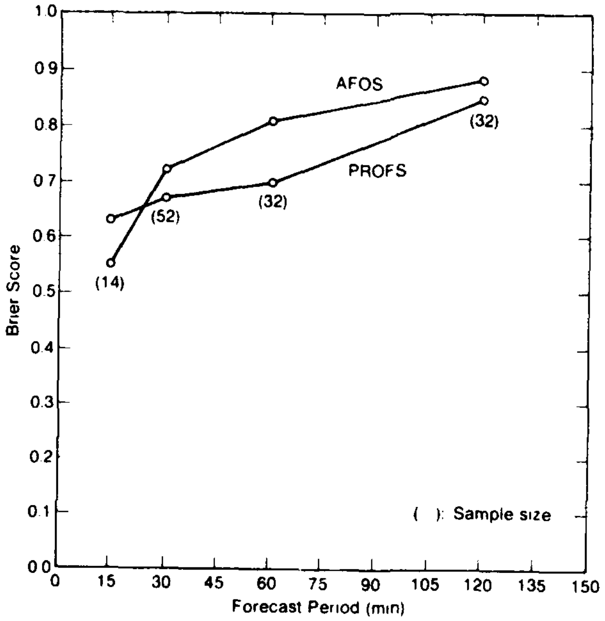


Figure 2. Brier scores for 20-kt wind speed category, initial condition winds < 20 kt, 1981 exercise.

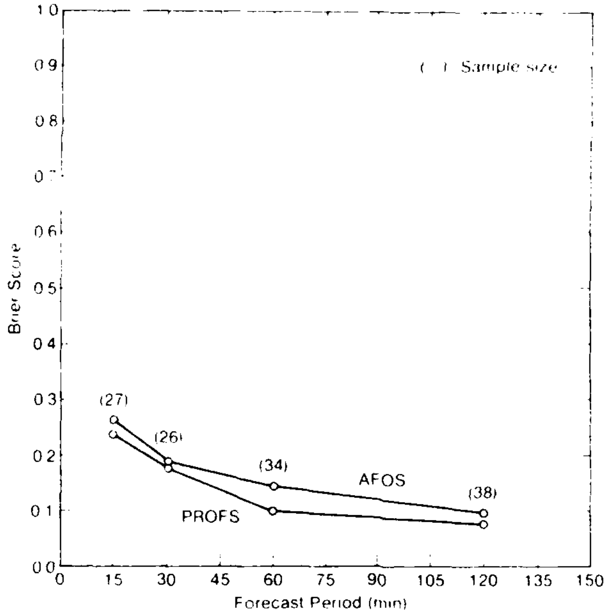


Figure 3. Brier scores for 20-kt wind speed category, initial condition winds > 20 kt, 1981 exercise.

4. *THE 1983 FORECAST EXERCISE*

Three major areas of effort were undertaken in preparing for the 1983 exercise: (1) adding new meteorological data sources to the data base; (2) improving system throughput and flexibility; and (3) sharpening a number of procedural matters in the execution of the forecast exercise. However, the ability to perform a comparative exercise was curtailed in that resources were not available to install an advanced workstation at the Denver WSFO. Nor was it possible to staff an AFOS forecast station in Boulder as had been done in 1982. Costs of replicating the advanced workstation for the Denver WSFO and staffing two forecast stations in Boulder were prohibitive.

4.1 Data and Products Available

Data sources and their update rates for the 1983 exercise are given in Table 3.

The major data sources added were real-time Doppler radar from the NCAR 10-cm Doppler radar (referred to as CP-2) and remotely-sensed profiles of wind, temperature, and humidity from the Wave Propagation Laboratory profiler system.[10] In addition, all the AFOS products from NMC and various NWS offices were now available on the advanced workstation. The data from the sources listed in Table 3 were processed to produce a variety of images on four different scales of motion: national, regional, eastern Colorado, and local. The national scale was a polar stereographic projection covering the contiguous United States and coastal waters, identical to that used on AFOS. The AFOS products produced by NMC were displayed on this scale. In addition, infrared satellite images were remapped and displayed on this scale. The remaining three scales were designed to provide increasingly detailed information on developing local weather situations with the so-called local scale being an area of 150 kilometers radius defined by the range of the CP-2 radar. All the products developed from data updated every five minutes were available on the local

---

[10] Hogg, D. C., M. T. Decker, F. O. Guiraud, K. B. Earnshaw, D. A. Merritt, K. P. Moran, W. B. Sweezy, R. G. Strauch, E. R. Westwater, and C. G. Little, 1983: An automatic Profiler of the temperature, wind, and humidity in the troposphere. J. Appl. Meteorol., 22, 807-831.

scale. This included satellite images during the last few days of the experiment when Rapid Scan Satellite Data were made available by NESDIS. Other products, updated less frequently, were available on the eastern Colorado and regional scales. Further discussion of products available on the 1983 workstation are given in Schlatter et al.[11] and the FY83 PROFS Annual Report.[12]

---

[11] Schlatter, T. W., P. Schultz, and J. M. Brown, 1984: Forecasting Convection with the PROFS system. Part I: Assessment of current capabilities and evaluation of the workstation. Submitted to Bull. Am. Meteorol. Soc.

[12] PROFS Annual Report for Fiscal Year 1983. NOAA/ERL/PROFS, 325 Broadway, Boulder, Colorado 80303.

Table 3. Data sets available in the 1983 exercise.

Data Sets	Update rate (min
Digital GOES satellite imagery (Infrared and visible images)	30 (5 min last few days of exercise)
Digital radar low-level PPI's from NCAR's 10-cm CP-2 Doppler radar	5
Digital Doppler velocity images from NCAR's 10-cm CP-2 Doppler radar	5
Mid-level (8 km) CAPPI from NCAR's 10-cm CP-2 Doppler radar (based on 11 tilt angle sequences from 0.5 to 14.5 degrees elevation)	15
Digital radar PPI's from NWS radars at Limon, CO, and Cheyenne, WY	5
Surface mesonet observations of 5-min averaged values of wind, temperature, dewpoint, pressure, and visibility. Accumulated (5-min) values of precipitation	5
Air-to-ground lightning detection (location and frequency)	5
Vertical profiles of wind to ≈ 15 km using ground-based Doppler wind profiling system of Wave Propagation Laboratory (WPL)	20
Vertical profiles of temperature and water vapor through stratosphere using ground-based passive microwave radiometer system of Wave Propagation Laboratory	20
NWS Surface Aviation Observations (SAO's)	60
NWS Upper-air (rawinsonde) observations	7200
All NMC products, analyses, prognoses, charts, graphics, text message, etc. available on AFOS	60 to 7200

The workstation was supported by a 32-bit architecture computer (VAX-11/750) capable of driving three workstations simultaneously. This vastly improved system throughput over the 1982 system which used 16-bit architecture computers (PDP-11/24's) for supporting only one workstation per computer. The products at the workstation were called from a separate menu screen as in the 1982 system. However, a touch screen rather than a light pen system was used for menu operation and the menu design was organized to provide rapid self-taught familiarity with all the products available on the system. A schematic diagram of the system as it had evolved by the 1983 exercise is given in Figure 4.

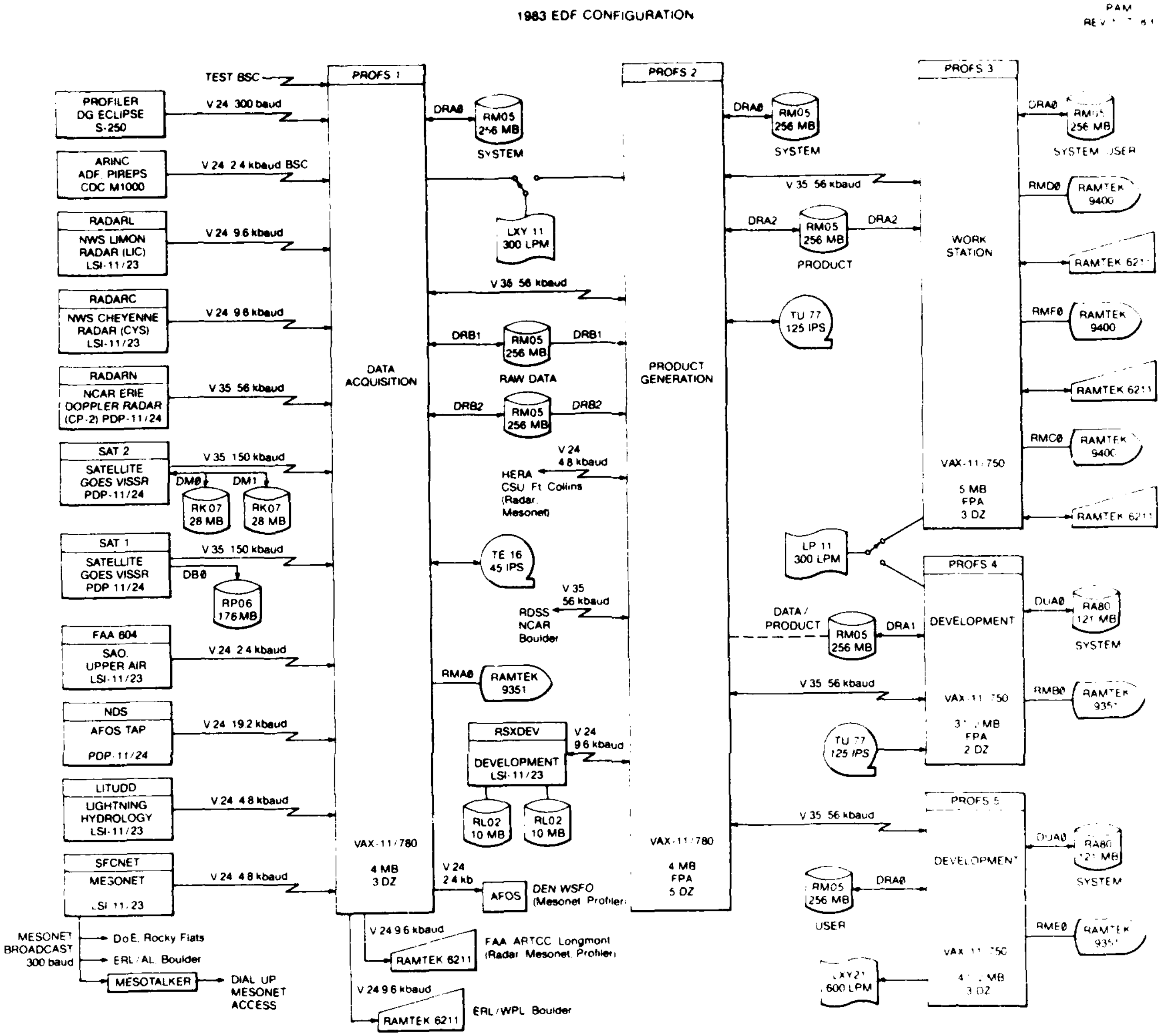


Figure 4. Schematic diagram of the PROFS Exploratory Development Facility (EDF) as configured for the 1983 exercise.

4.2 Forecast Procedures

There were three types of forecasts made during the 1983 exercise which extended from 1 June through 12 August. First, a convective outlook was prepared by noon each day. It consisted of probabilities of severe thunderstorms and tornadoes for specified regions within the forecast area and of a narrative describing the expected weather. Second, severe weather warnings for severe thunderstorms, tornadoes, or flash floods were prepared as deemed necessary during the forecast shift. Third, probability forecasts of wind, visibility, and precipitation were prepared for the forecast area at prescribed times during the forecast shift. The forecast shift was scheduled from 1600 to 2400 GMT (10:00 a.m. to 6:00 p.m. MDT) daily, and was extended three additional hours on days when convective weather was occurring near the end of the normal shift. The forecast area for all forecasts was the local area of a 150-kilometer radius from the CP-2 radar site, which was located roughly 30 kilometers east of Boulder, CO (Figure 5).

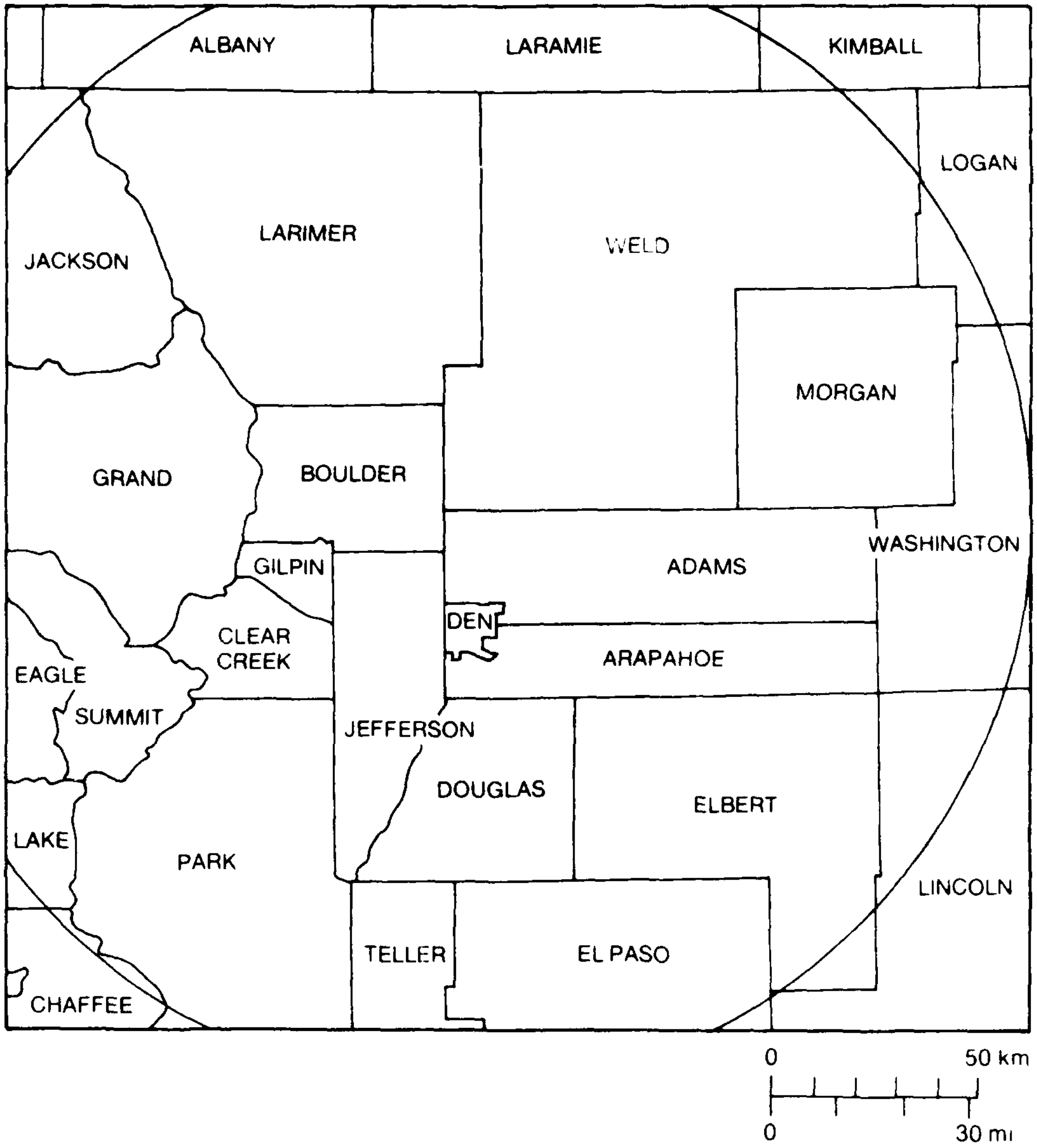


Figure 5. Forecast area map for 1983 exercise.

Probability forecast maps were prepared by drawing isopleths of probability levels of 0, 5, 10, 20, ... 100 percent on forecast maps for the parameter being forecast. The precipitation forecasts were made every two hours at 1845, 2045, and 2245 GMT for each of four successive two-hour periods starting at

1900, 2100, and 2300 GMT, respectively. Two separate precipitation categories were forecast: probability of precipitation $\geq$ 0.01 inch and probability of precipitation $\geq$ 0.5 inch. Visibility and wind forecasts were made every two hours at 1945, 2145, and 2345 GMT for each of four successive half-hour periods starting at 2000, 2200, and 2400 GMT, respectively. The wind speed category was speeds $\geq$ 25 knots; the visibility category was visibility less than three miles. Thus, eight different probability forecast maps were prepared every hour from 1945 GMT to 2345 GMT, and later if the shift was extended. The wind and visibility forecasts covered total periods of two hours, periods which expired just as a new set of forecasts was due. The precipitation forecasts covered total periods of eight hours, periods which were re-initiated every two hours with a new set of forecasts.

There were two advanced workstations in the 1983 exercise. One was in a forecast office; the other in a weather verification office. Each workstation was staffed by a team of two forecasters, one drawn from the National Weather Service, the other from the research community. These forecasters were a select staff with many years of operational forecasting or research experience with severe weather. In addition, an observer, either a PROFS staff member or a visitor from a university, the private sector, or some other government organization, assisted the forecast team, usually in a learning capacity.

The work schedule for each forecast team consisted of six days at the verification station, three days off, and six days at the forecast station. The verification team was responsible for obtaining severe weather verification data. This was done using radio-car chase teams, NWS weather spotters, contractor spotters, and after-the-fact survey operations. By the end of the six-day stint in the verification room, the forecast team was proficient in the use of the display system and, depending on the weather which had occurred, had developed a varying degree of knowledge of severe weather characteristics of the Colorado high plains.

The forecast station was maintained as an isolated environment. No visitors were permitted in the forecast office at any time. No discussion of the weather was allowed except among the forecaster team members on shift. No severe weather reports were permitted in the forecast office. Hence, all forecaster decisions were based solely on the information available from the workstation, i.e., the various display products generated from the data sources listed earlier.

4.3 Forecast Evaluation

The severe weather warning evaluations show substantial improvement in severe weather warning scores over scores obtained from the Denver WSFO warnings for the same time periods and forecast area.[13] It is inferred in these evaluations that the improvements are due to the enhanced information base available with the advanced workstation, the extensive experience of the select forecast staff, and the isolated, distraction-free environment of the experimental forecast office.

---

[13] McCoy, M. C., 1984: Severe storm warning scores from the PROFS 1983 forecast exercise. (Submitted to Bull. Am. Meteorol. Soc.)

The probability forecast evaluations will be summarized in the following sections. A more detailed discussion of these evaluations may be found in a paper by Williams.[14] As in the 1982 evaluation efforts, the Brier score, defined previously, has been used to analyze the probability forecasts. In addition, the reliability term of the so-called partitioned Brier score[15] has been evaluated. This term is defined as

$$\text{Reliability} = \frac{1}{N}\sum_{s=1}^{12} N^s \left(F^s - E^s\right)^2 \qquad (2)$$

where N^s is the number of forecasts for which $F = F^s$ (note that $\sum_s N^s = N$); E^s is the observed relative frequency of the event when F^s is the forecast probability. In this case, $s = 12$ and $F^s = 0, 5, 10, 20, \ldots.100$ percent. This term provides a quantitative measure of the reliability of the forecasts. It becomes zero only when the forecast probabilities and observed relative frequencies are equal for all S subsamples.

A skill score (SS) based on the Brier score is defined as

$$SS = [1 - (BS/BS_r)] \times 100, \qquad (3)$$

---

[14] Williams, G. M., 1984a: Probability forecast results of the PROFS 1983 Summer Exercise. (To be submitted to Mon. Wea. Rev.)

[15] Murphy, A. H., and R. Winkler, 1982: Subjective probabilistic tornado forecasts: some experimental results. Mon. Wea. Rev., 110, 1288-1297.

where BS_r denotes the Brier score for forecasts produced by some reference forecast, such as persistence or climatology. Such a skill score is positive (negative) when the forecasts being evaluated are more (less) accurate than the reference forecasts. Perfect forecasts ($BS = 0$) attain a skill score of 100, while forecasts equivalent in accuracy to the reference forecasts ($BS = BS_r$) receive a skill score of 0. Skill scores referenced to persistence, climatology, and conditional climatology are used herein. We discuss, in order, the wind forecast, the visibility forecast, and the precipitation forecast evaluations.

4.3.1 Wind forecast evaluations

There are two basic classes of weather situations that can cause high winds in the Colorado plains in the summer. One is phenomena generated by convective weather (e.g., downdrafts and outflow from strong to severe convective cells, gust fronts, tornadoes). The other is mesoscale to large-scale circulations, with or without topographical forcing (e.g., fronts, convergence zones, downslope winds). With wind information available from the mesoscale surface network and the Doppler radar, the forecaster had a rich data source for making the wind probability forecasts. The evaluation results indicate that these data sets (and possibly others as well) were used to advantage for forecasting a relatively rare event. Verification of the forecasts was accomplished using the observations from the mesoscale surface network.

Brier score results are presented in Tables 4 and 5 for two categories of initial conditions at the time of the forecast: initial wind speeds < 25 kt; initial wind speeds ≥ 25 kt. Note that over 4000 observations of light winds occurred as an initial condition and that only some 130 cases of strong winds occurred as an initial condition during the entire summer. That is, wind speeds ≥ 25 kt are relatively rare events and thus difficult to forecast accurately. The Brier score results for the light wind initial conditions are absolute scores of 0.02 to 0.03 for both persistence and forecaster Brier scores, a result which suggests that the forecasters well understood the rare event nature of strong winds. No deterioration of forecast accuracy with forecast lead time occurred; however, persistence forecasts consistently provided a better forecast in this large data sample.

The Brier score results for strong wind initial conditions are given in Table 5. The persistence scores, ranging from 0.48 to 0.73, and deteriorating with time, are consistent with the rare event aspect of the strong wind phenomenon. That is, the strong winds typically do not persist very long. The forecaster scores are substantially better than the persistence scores for this category, and they improve with time. However, the absolute values range from 0.2 to 0.33, with the poorer score (0.33) occurring for the 0-30 minute forecast period, not particularly impressive Brier scores. This suggests strongly that the forecasters overforecast the strong winds for the first forecast period particularly, with some improvement obtained for the longer lead forecast times.

Table 4. Forecaster and persistence Brier scores vs. forecast period for winds ≥ 25 kt. Initial wind speeds < 25 kt.

	Forecast Period (min)			
	0-30	30-60	60-90	90-120
Sample size (N)	4251	4258	4155	4106
Forecaster Brier score	0.030	0.032	0.037	0.029
Persistence Brier score	0.020	0.022	0.026	0.022

Table 5. Forecaster and persistence Brier scores vs. forecast period for winds $\geq$ 25 kt. Initial wind condition $\geq$ 25 kt.

	Forecast Period (min)			
	0-30	30-60	60-90	90-120
Sample size (N)	137	137	133	128
Forecaster Brier score	0.30	0.22	0.21	0.20
Persistence Brier score	0.48	0.66	0.69	0.73

Skill scores relative to persistence are shown in Table 6. For the initial light wind stratification, the skill scores are quite poor, ranging from -27 to -50. However, since both the persistence and forecaster scores were low for this category, these skill scores reflect to some extent the sensitivity of the score to small variations in small numbers rather than significant lack of skill on the part of the forecasters.

Table 6. Forecaster skill scores relative to persistence vs. forecast period for wind speeds $\geq$ 25; light and strong wind initial conditions.

	Forecast Period (min)			
	0-30	30-60	60-90	90-120
Skill score (initial winds < 25 kt)	-50	-45	-27	-32
Skill score (initial winds $\geq$ 25 kt)	37	67	70	73

The skill scores for the strong wind initial condition illustrate considerable skill relative to persistence, skill that improves with increasing forecast lead time, from 37 percent for the 0-30 minute forecast period to 73 percent for the 90-120 minute forecast period, although the absolute Brier scores show lack of absolute skill.

4.3.2 Visibility forecast evaluations

Low visibilities in Colorado during the convective storm season occur almost always only in the rain shaft of a heavy rainstorm or thundershower. Fog, blowing sand, or other obstructions to visibility occur so infrequently in the summer months they may be discounted. This is the third exercise in which visibility as a forecast parameter has been included - with the same result. The occurrence of the event is so rare, and the forecasters' confidence in forecasting non-occurrence (probabilities nearly always 0 to 30 percent) so well-justified that no insight can be gained by evaluating the forecasts. That is, forecasts of low visibility cannot be evaluated under these circumstances with the objective of determining improvements in forecast accuracy due to advanced workstation capabilities. There are therefore no quantitative analyses to present for the visibility forecasts. Furthermore, it is not likely that visibility as a forecast parameter will be included in future PROFS convective season forecast exercises.

4.3.3 Precipitation forecast evaluations

The precipitation forecasts from the 1983 exercise have offered the best opportunity for evaluation to date. It has been possible to use an expanded data base for forecast verification purposes and for developing a rainfall climatology for forecast comparison purposes. The National Weather Service manages a cooperative observer network in every state that provides observations of daily maximum and minimum temperatures and daily accumulated precipitation. Some of these cooperative stations employ rain gauges that provide hourly rainfall records. There are 19 such stations with usable records of 15 to 30 years in length in the local forecast area used in this exercise. Appropriate climatological summaries for this exercise were prepared under contract with the Colorado State climatologist.

Three sets of statistics were obtained based on the data for 1 June through 14 August, inclusively. One set was sample climatology, i.e., the probability of rainfall in two-hour periods identical with the forecast periods of the exercise. The second set was the probability of rainfall in four successive two-hour periods conditioned on the previous two-hour period having no rain. The third set was the probability of rainfall in four successive two-hour periods conditioned on the previous two-hour period having had rain. All three climatologies were generated for each of the two precipitation categories of our exercise, viz., rainfall $\geq$ 0.01 inch and rainfall $\geq$ 0.50 inch. The Brier score evaluations use these climatologies as well as persistence as reference or control forecasts in developing skill scores. None of these climatologies were available until the exercise was over, i.e., they were not available for forecast guidance during the exercise. Detailed discussion of these climatologies will be published as a NOAA Tech Memo.[16]

The probabilities based on the two-hour climatologies varied from .01 to .17, with highest values at the mountain and foothill stations. They also showed a marked diurnal variation, with the highest probabilities occurring

---

[16] Williams, G. M., 1984b: Climatology and conditional climatology of summer precipitation in Colorado. (To be submitted as a NOAA Tech Memo, ERL, Boulder, CO.)

between 2100 and 2300 GMT in the foothills and between 0100 and 0300 GMT at locations on the eastern plains, i.e., thunderstorms form over the Continental Divide and the eastern slopes of the Rockies in early to mid-afternoon and move over the eastern plains in the late afternoon and early evening.

The probability values based on conditional climatology varied considerably; further, the differences between values for precipitation as the initial condition and no precipitation as the initial condition were quite large. Probabilities for rain after no rain occurred ranged from .01 to .14, with an average of about .07. Probabilities for rain after rain occurred ranged from .03 to .79, and averaged about .40. As with climatology, some areas, especially the Palmer Divide south of Denver, were more favorable for precipitation, while values on the eastern plains were slightly lower than elsewhere. These probabilities also displayed a distinct diurnal variation, especially for rain-occurred values; for example, the probabilities for rain six to eight hours after 1900 were much higher than those for six to eight hours after 0100 GMT.

Our evaluations have been restricted to only the light to moderate rainfall category. Heavy rainfall, $\geq$ 0.50 inch in a two-hour period, occurred too infrequently to warrant analysis.

Two verification conditions have been defined for evaluation purposes: one independent of the weather during the verification period; the other, a subset of the first, chosen for only those cases (stations) when it rained ($\geq$ 0.01 inch) during the verification period. The latter condition highlights a very difficult situation to forecast, the probability of rainfall dependent on the expected development and movement of convective phenomena.

Forecaster and sample climatology Brier scores are presented in Table 7 for all cases obtained from the summer exercise. The absolute levels of the Brier scores for all cases range from 0.062 to 0.081 with little change in accuracy as a function of forecast period. There is also little to distinguish between forecaster and sample climatology scores.

Table 7. Forecaster and sample climatology Brier scores and forecaster skill scores vs. forecast period for precipitation ≥ 0.01 in. All cases.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	6964	6744	6487	6326
Forecaster	0.069	0.081	0.075	0.062
Sample Climatology	0.074	0.080	0.071	0.057
Forecaster skill score relative to climatology	7	-1	-6	-9

The Brier scores for the rain-verified cases only are presented in Table 8. It is seen that the absolute values of the Brier scores deteriorate substantially from those listed in Table 7. The forecaster scores range from 0.47 to 0.75; the sample climatology scores range from 0.84 to 0.88. The forecaster scores become worse as the forecast period increases; the sample climatology scores change only slightly as a function of forecast period. It is also seen that the forecaster skill scores are positive relative to sample climatology scores, although the skill decreases with increasing forecast period. It also seems noteworthy that rainfall occurred for only six to nine percent of all the cases evaluated, a statistic that emphasizes the rare event nature of the phenomenon being forecast.

Table 8. Forecaster, sample climatology Brier scores and forecaster skill scores relative to climatology vs. forecast period for precipitation ≥ 0.01 in. Rain-verified cases only.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	564	587	499	384
Forecaster Brier score	0.47	0.59	0.67	0.75
Sample Climatology Brier score	0.84	0.84	0.86	0.88
Forecaster Skill Score relative to climatology	44	30	22	15

Let us now examine forecaster, persistence, and conditional climatology scores for two sets of initial conditions: (1) no rain during the two-hour period preceding the first forecast period (no rain initially); (2) rain ≥ 0.01 inch during the two-hour period preceding the first forecast period (rain initially). For each of these two sets, two categories are examined: (1) all verification conditions (rain and no-rain); (2) rain cases only during the verification periods.

The results for the no-rain initial conditions, all verification conditions, are given in Table 9. Brier scores range from 0.051 to 0.077 with little to distinguish the forecaster, persistence, climatology, or conditional climatology scores. Nor does there appear to be any significant change in the level of these scores as a function of the forecast period.

Table 9. Forecaster, conditional climatology, sample climatology, and persistence Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. All verification cases.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	6564	6388	6151	6001
Forecaster	0.057	0.074	0.073	0.060
Conditional climatology	0.051	0.072	0.068	0.055
Sample Climatology	0.051	0.071	0.063	0.055
Persistence	0.054	0.077	0.074	0.058

The results for the rain-verified subset of these cases are given in Table 10. The forecaster absolute scores range from 0.52 to 0.76, deteriorating steadily with forecast period. The conditional climatology scores range from 0.86 to 0.89, with only slight variations with forecast period. The climatology scores range from 0.84 to 0.88 with only slight variations as well. By definition, the persistence scores for these cases are 1.0.

Table 10. Forecaster, conditional climatology, and sample climatology Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. Rain-verified cases only.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	351	493	452	350
Forecaster	0.52	0.61	0.67	0.76
Conditional climatology	0.89	0.86	0.87	0.89
Sample climatology	0.84	0.84	0.85	0.88

Forecaster, persistence, and conditional climatology Brier scores for the rain initial conditions, all verification cases are shown in Table 11. The forecaster scores range from 0.27 to 0.08, steadily improving with forecast period. They do not differ appreciably from the conditional climatology scores, in absolute magnitude or in behavior with forecast period, but the sample climatology scores are somewhat worse for the zero to two and two to four-hour forecast periods. The persistence scores, on the other hand, range from 0.47 to 0.9, steadily worsening with forecast period. Rain occurs increasingly rarely in the future as the forecast period increases when it is initially raining in a convective situation, as these persistence scores clearly show.

Table 11. Forecaster, conditional climatology, sample climatology, and persistence Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. All verification cases.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	400	356	336	325
Forecaster	0.27	0.20	0.11	0.08
Conditional climatology	0.25	0.19	0.12	0.09
Sample climatology	0.45	0.23	0.12	0.09
Persistence	0.47	0.74	0.86	0.90

The results for the rain-verified subset of these cases are given in Table 12. The forecaster scores range from 0.39 to 0.65, worsening with forecast period. The conditional climatology scores range from 0.25 to 0.73 and also steadily worsen with time. The sample climatology scores are particularly poor, ranging from 0.84 to 0.88. By definition, the persistence scores for these cases are 0.0.

Table 12. Forecaster, conditional climatology, and sample climatology Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. Rain-verified cases only.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Sample size (N)	213	94	47	34
Forecaster Brier score	0.39	0.53	0.58	0.65
Cond. climatology Brier score	0.25	0.53	0.68	0.73
Climatology Brier score	0.84	0.84	0.86	0.88

Forecaster skill scores referenced to conditional climatology, climatology, and persistence are presented in Tables 13 through 16 for the data categories previously defined.

It is seen from Table 13 that no forecaster skill is evident in forecasting precipitation given initial conditions of no rain, regardless of the referenced Brier score. The forecaster Brier scores for this data set were very low, however. This suggests that although the forecasters normally issued low probability forecasts in this situation and were rewarded with low Brier scores, their Brier scores were not low enough to show skill over climatology or persistence.

Table 13. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. All verification cases.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Forecaster/cond. climatology	-12	-3	-7	-9
Forecaster/climatology	-12	-4	-16	-9
Forecaster/persistence	-6	4	1	-3

Table 14. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial no-rain conditions. Rain-verified cases only.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Forecaster/cond. climatology	42	29	23	15
Forecaster/climatology	38	27	21	14
Forecaster/persistence	48	39	33	24

Table 15. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. All verification cases.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Forecaster/cond. climatology	-8	-5	8	11
Forecaster/climatology	40	13	8	11
Forecaster/persistence	43	73	87	91

Table 16. Forecaster skill scores referenced to indicated Brier scores vs. forecast period for precipitation $\geq$ 0.01 in. Initial rain conditions. Rain-verified cases only.

	Forecast Period (hours)			
	0-2	2-4	4-6	6-8
Forecaster/cond. climatology	-44	0	15	11
Forecaster/climatology	54	37	33	26

As seen in Table 14, the forecasters do show skill relative to all referenced scores for the rare event case: raining at a station some time in the future after an initial condition of no rain. As one might expect, this skill lessens to very low levels as the forecast period lengthens, suggesting that climatological guidance would probably have been very useful during the exercise, particularly for the later forecast periods.

For the rain initial conditions, all verification cases, shown in Table 15, little skill is shown by the forecaster except relative to persistence. That is, the forecasters generally were not swayed to issue high probabilities because of initial rain conditions and thus did quite well compared to persistence. Their score relative to climatology was quite good for the 0-2 hour forecast in this set as well. Furthermore, the skill scores are positive, although small, relative to climatology or conditional climatology for the later period forecasts.

The results for the case of rain-verified, initial rain conditions, Table 16, show that the forecasters do considerably better than climatology for all time periods. And as previously noted, the skill worsens with later forecast periods. A little skill is shown relative to conditional climatology for the later forecast periods, but no skill for the earlier periods.

Overall, it is encouraging to find suggestions of forecaster skill in these scores, albeit the skill is very low. This is particularly so given the difficult challenge this forecast exercise represents, increasingly so for the four to six-hour and six to eight-hour forecasts. Further, although nearly all the forecasters had severe weather warning experience, only a few of them were experienced with probability forecasting. Thus, to obtain positive skill scores that are reasonably well behaved is indeed encouraging.

Forecaster reliability plots are presented for the four different forecast periods in Figs. 6 through 9. In Figs. 6a through 9a the forecaster probabilities are plotted vs. the observed frequencies of occurrence of precipitation $\geq$ 0.01 inch. In Figs. 6b through 9b, the forecast probabilities are plotted vs. the relative frequency of forecasts made for each probability. The number at the right of each bar in this plot is the ratio of number of observed rainfall cases to the number of forecasts made in that forecast category.

It is seen that overforecasting prevailed for all forecast time periods. For example, in the zero to two-hour forecast period, 138 forecaster probabilities of 70 percent or greater were scored for stations at which precipitation occurred only 64 times. The forecasters were prone to issue somewhat lower probabilities for the later forecast periods, but were nevertheless inclined to overforecast considerably, particularly for the higher forecast probabilities. These plots graphically demonstrate possible inexperience in issuing probability forecasts. Presumably, seasoned forecasters used to this type of forecast would not be so prone to overforecasting.[17] Furthermore, had the climatological and conditional climatological probabilities been available for guidance during the exercise the tendency towards overforecasting might have been substantially reduced. In future forecast exercises, the forecast staff will be trained more thoroughly in probability forecasting and will have appropriate computer guidance available.

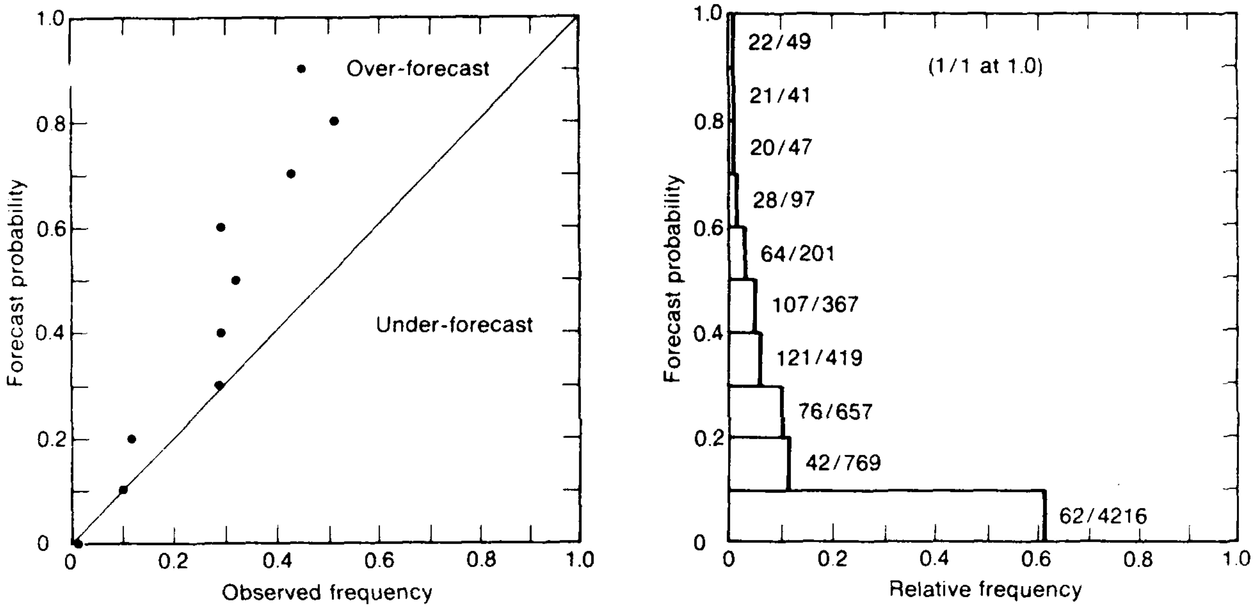


Figure 6. (a) Forecaster reliability chart for precipitation ≥ 0.01 inch.
(b) Forecast frequency plot for associated reliability plot.
Forecast period: 0-2 hours.

---

[17] Murphy, A. H., and H. Daan, 1984: Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee experiment. *Mon. Wea. Rev.*, *112*, 413-423.

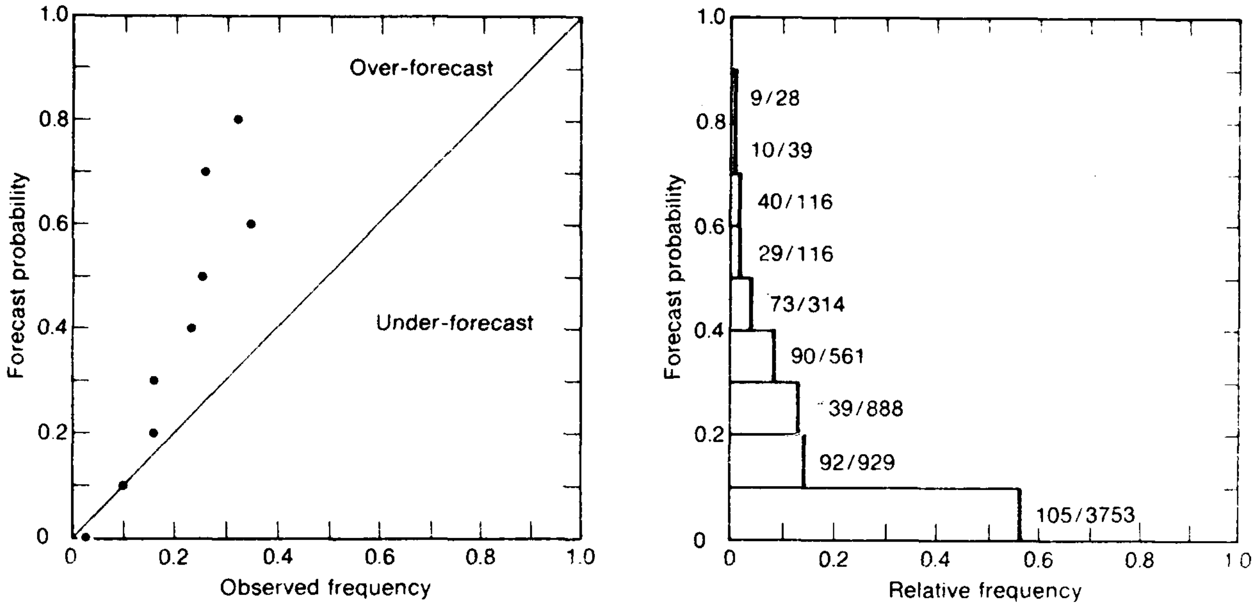


Figure 7. (a) Forecaster reliability chart for precipitation ≥ 0.01 inch.
(b) Forecast frequency plot for associated reliability plot.
Forecast period: 2-4 hours.

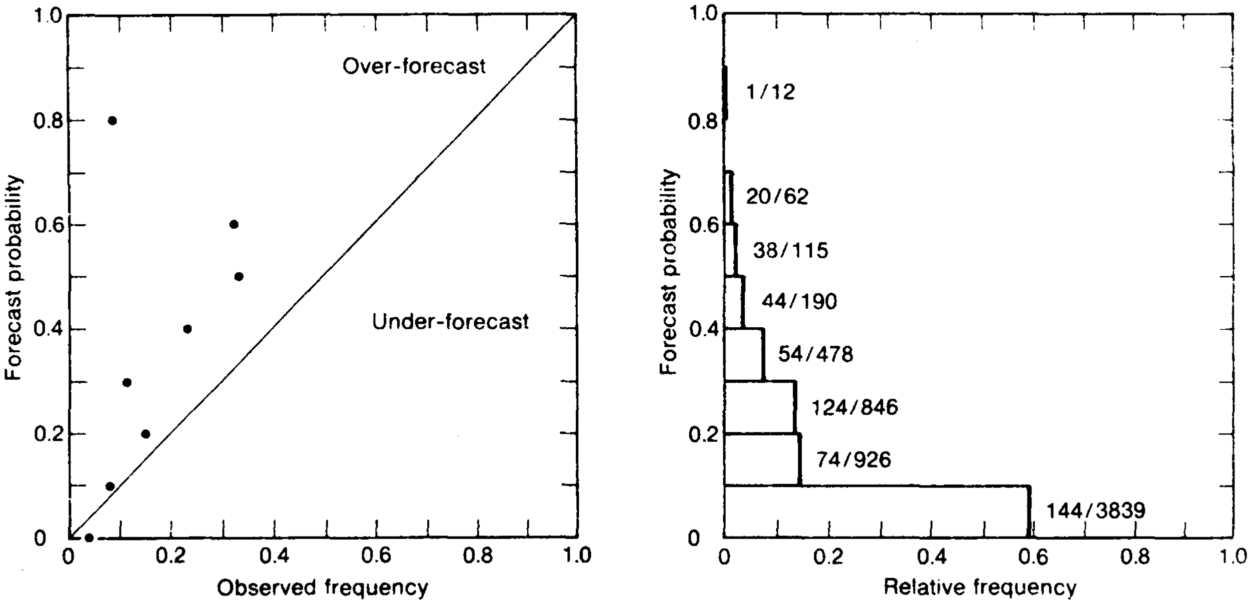


Figure 8. (a) Forecaster reliability chart for precipitation ≥ 0.01 inch.
(b) Forecast frequency plot for associated reliability plot.
Forecast period: 4-6 hours.

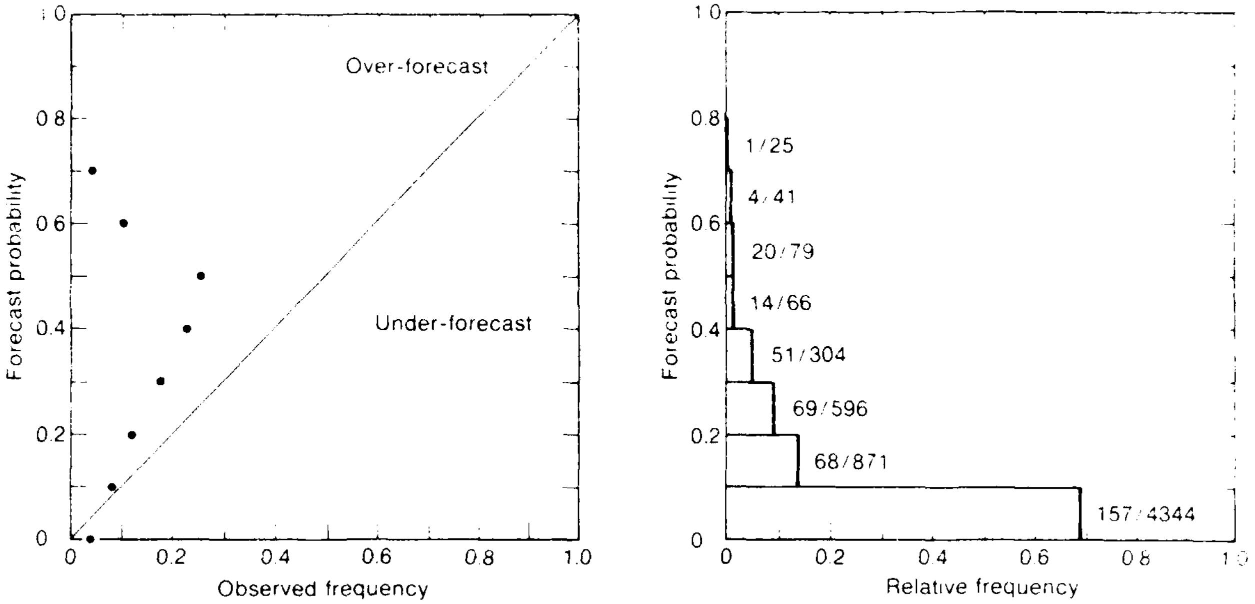


Figure 9. (a) Forecaster reliability chart for precipitation $\geq$ 0.01 inch.
(b) Forecast frequency plot for associated reliability plot.
Forecast period: 6-8 hours.

5. CONCLUSIONS

The results just presented do not reveal impressive skill scores. However, since positive skill scores did result for situations which are particularly difficult to forecast, it seems logical to expect that substantially improved results are possible in future experiments. Training forecasters to prepare probability forecasts is of paramount importance for future forecast exercises. Equally important is the use of conditional climatologies and sample climatologies as forecaster guidance tools.

This report describes a project development over the last four years that has been largely a building and learning experience. The primary emphasis in the PROFS program is now shifting to more sharply-focussed forecast experiments which should provide increasingly more solidly-based evaluations of forecast accuracies in the future.

The working hypothesis that the capabilities of an advanced workstation will lead to improved forecast accuracies by providing more useful information on rapidly changing weather situations in a timely manner to the forecaster has been given a basis of credibility to date by these forecast exercises. It is expected that this hypothesis will be strongly supported with future exercise results.

6. ACKNOWLEDGMENTS

We are particularly grateful for the support, advice, and encouragement given to us throughout this effort by Donald Chisholm, AFGL. Donald Beran and A. E. (Sandy) MacDonald, PROFS, have been a constant source of encouragement as well. Colin S. Ramage, University of Hawaii, and Allan H. Murphy, Oregon State University have each spent many hours with the authors discussing and contributing many stimulating ideas and concepts. It is our pleasure to acknowledge the help offered by all these people as well as the countless contributions by the PROFS staff over the years that have made these forecast exercises possible.

7. REFERENCES

1. Klein, W. H., 1976: The AFOS program and future forecast applications. Mon. Weather Rev., 104, 1494-1504.

2. Design News, 1980: Color adds "pizzaz" to weather radar. 26, 13-14.

3. Pratte, J. F., and R. F. Clark, 1983: PROFS Mesonet-description and performance. Preprints, 5th Symp. on Meteorological Observations and Instrumentation. Am. Meteorol. Soc., Boston, Mass.

4. Gringorten, I.I., 1972: Conditional probability for an exact non-categorized initial condition. Mon. Wea. Rev., 110, 796-798.

5. Hughes, L. A., 1980: Probability forecasting - reasons, procedures, problems. NOAA Tech. Memo, NWS FCST 24.

6. Brown, R. C., 1983: Anatomy of a mesoscale instrumentation system. Preprints, Fifth Symposium on Meteorological Observations and Instrumentation, Am. Meteorol. Soc., Boston, 306-313.

7. Reynolds, D. W., 1983: Prototype workstation for mesoscale forecasting. Bull. Am. Meteorol. Soc., 64, 264-273.

8. Haugen, D. A., D. Birkenheuer, R. F. Bunting, M. Cairns McCoy, G. M. Williams, 1984: An evaluation of PROFS' 1982 real-time forecast experiment. NOAA Tech Report , ERL, Boulder, CO, pp.

9. Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. Mon. Weather Rev., 78, 1-3.

10. Hogg, D. C., M. T. Decker, F. O. Guiraud, K. B. Earnshaw, D. A. Merritt, K. P. Moran, W. B. Sweezy, R. G. Strauch, E. R. Westwater, and C. G. Little, 1983: An automatic Profiler of the temperature, wind, and humidity in the troposphere. J. Appl. Meteorol., 22, 807-831.

11. Schlatter, T. W., P. Schultz, and J. M. Brown, 1984: Forecasting Convection with the PROFS system. Part I: Assessment of current capabilities and evaluation of the workstation. Submitted to Bull. Am. Meteorol. Soc.

12. PROFS Annual Report for Fiscal Year 1983. NOAA/ERL/PROFS, 325 Broadway, Boulder, Colorado 80303.

13. McCoy, M. C., 1984: Severe storm warning scores from the PROFS 1983 forecast exercise. (Submitted to Bull. Am. Meteorol. Soc.)

14. Williams, G. M., 1984a: Probability forecast results of the PROFS 1983 Summer Exercise. (To be submitted to Mon. Wea. Rev.)

15. Murphy, A. H., and R. Winkler, 1982: Subjective probabilistic tornado forecasts: some experimental results. Mon. Wea. Rev., 110, 1288-1297.

16. Williams, G. M., 1984b: Climatology and conditional climatology of summer precipitation in Colorado. (To be submitted as NOAA Tech Memo, ERL, Boulder, CO.)

17. Murphy, A. H., and H. Daan, 1984: Impacts of feedback and experience on the quality of subjective probability forecasts: comparison of results from the first and second years of the Zierikzee experiment. Mon. Wea. Rev., 112, 413-423.

ATE

LMED

-8