QC
807.7
.U66 c.2
no. 427

NOAA Technical Report ERL 427–ESG 1

An Evaluation of the PROFS 1982 Real-Time Forecast Experiment

Duane A. Haugen
Daniel Birkenheuer
Robert F. Bunting
Mary Cairns McCoy
Gary M. Williams

June 1984

LIBRARY
FEB 27 1985
N.O.A.A.
U. S. Dept. of Commerce

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Environmental Research Laboratories

QC
807.5
.U66
no.427
c.2

NOAA Technical Report ERL 427–ESG 1

An Evaluation of the PROFS 1982 Real-Time Forecast Experiment

Duane A. Haugen
Daniel Birkenheuer
Robert F. Bunting
Mary Cairns McCoy
Gary M. Williams

Program for Regional Observing and Forecasting Services
Boulder, Colorado

June 1984

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Oceanic and Atmospheric Administration
John Byrne, Administrator

Environmental Research Laboratories
Vernon Derr, Director
Boulder, Colorado

NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA Environmental Research Laboratories. NOAA does not authorize any use of information from this publication for advertising or publicity purposes.

CONTENTS

Page

ABSTRACT 1

1. INTRODUCTION 1

2. THE 1982 REAL-TIME FORECAST EXPERIMENT 2
   2.1 Forecaster Workstations 2
   2.2 Forecaster Staffing 4
   2.3 Forecast Procedures 5

3. SEVERE WEATHER VERIFICATION PROGRAM 6
   3.1 Characteristics of Verification Data 6
   3.2 Denver WSFO Verification Data 7
   3.3 Telephone Call Program 7
   3.4 Chase Program 7
   3.5 Spotters 8
   3.6 Verification Data File 8
   3.7 Discussion 8

4. WARNING EVALUATION RESULTS 10
   4.1 Objective 10
   4.2 Definition of Data Sets 10
   4.3 Scoring Procedures 11
   4.4 Discussion of Results 11

5. PROBABILITY FORECAST EXPERIMENT 16
   5.1 Experimental Procedures 16
   5.2 Evaluation Procedures 16
   5.3 Results 17

6. SUMMARY 18

7. ACKNOWLEDGMENTS 19

8. REFERENCES 19

ACRONYMS 21

FIGURES

Page

1. Severe weather reports ranked by source..... 9

2. Non-severe weather reports ranked by source..... 9

3. Brier scores for 20-kn wind speed category..... 17

4. Brier scores for 20-kn wind speed category, initial condition winds < 20 knots..... 17

5. Brier scores for 20-kn wind speed category, initial condition winds > 20 knots..... 18

TABLES

1. Data sets available on the 1982 advanced workstation..... 3

2. Total number of warnings issued at each station by type of severe weather event..... 10

3. Severe thunderstorm warning scores, hits, false alarms, and misses for each workstation..... 12

4. Tornado warning scores, hits, false alarms, and misses for each workstation..... 13

5. Comparative severe thunderstorm warning scores for Category I..... 14

6. Comparative severe thunderstorm warning scores for Category II..... 14

7. Comparative severe thunderstorm warning scores for Category III..... 15

An Evaluation of the PROFS 1982 Real-Time Forecast Experiment

Duane A. Haugen, Daniel Birkenheuer,*
Robert F. Bunting, Mary Cairns McCoy, Gary M. Williams

ABSTRACT. From 1 June to 15 August 1982 the Program for Regional Observing and Forecasting Services (PROFS) conducted a forecast exercise in the Front Range area of Colorado. One purpose of the exercise was to evaluate the PROFS real-time operational workstation as an aid to improving the timeliness and accuracy of short-term forecasts. The authors describe the workstation, the forecasts issued (severe storm warnings, and short-term point forecasts of wind, visibility, and precipitation), and the verification program. Forecasts issued from an AFOS (Automation of Field Operations and Services) workstation are compared with forecasts prepared on the PROFS workstation. Probability of detection scores, false alarm rates, and critical success indices are given for each workstation. Small sample sizes and problems in exercise design preclude definitive answers. However, the exercise results suggest that improved forecasting performance is possible with the advent of advanced technology and training of forecasters in preparing short-term forecasts as well as in using advanced technology capabilities.

1. INTRODUCTION

This report presents the results of a forecast experiment conducted by the Program for Regional Observing and Forecast Services (PROFS) from 1 June to 15 August 1982. A central objective of the PROFS program is to determine if improvements in short-term, local forecasts and warnings of severe weather are possible with the use of advanced computer and display technology at a forecaster workstation. The 1982 exercise was our first real-time forecast experiment, and, as such, was largely a learning experience. There were actually three goals of the 1982 exercise: (1) to provide an opportunity for senior research meteorologists to use and to constructively criticize the PROFS advanced workstation; (2) to train junior PROFS meteorologists in the use of the workstation and in the procedures and decision-making of preparing severe weather warnings and local forecasts; (3) to evaluate the workstation

*Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado.

and its capabilities as an aid to improving the timeliness and accuracy of short-term forecasts. It is the last goal which is the subject of this report.

Our working hypothesis is that forecast accuracies are improved by use of the advanced workstation. We envision our evaluation efforts as a two-stage process. The first stage is to determine whether improvements indeed occur. If they do, the second stage is to evaluate quantitative measures of the improvements. These two stages might be likened, respectively, to exploratory and confirmatory experiments in terms frequently used to describe experimental design criteria (Flueck, 1982). The seasoned experimentalist will certainly appreciate that it is difficult at best to test and evaluate our working hypothesis. And our task was made substantially more difficult by the fact that the 1982 exercise had several goals. However, even though compromises were made in the design of the experiment, the results of our evaluation provide some qualitative evidence that supports our hypothesis.

2. THE 1982 REAL-TIME FORECAST EXPERIMENT

2.1 Forecaster Workstations

Two separate workstations were installed in the PROFS forecast office in Boulder. The advanced workstation incorporated color image and display technology; the other incorporated the vector graphic and limited image display technology currently available in many National Weather Service (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 advanced workstation included a high-resolution color image and graphics system for displaying and manipulating products generated by a large distributed-processing, networked computer system (Brown, 1983). The system routinely acquires, processes, and displays meteorological data from four major sources: weather satellites, weather radars, local mesoscale networks, and NWS surface and upper-air observation networks. A detailed description of the advanced workstation and the products available in 1982 has been published by Reynolds (1983). Data sources and their update rates are given in Table 1.

There are several attributes of the advanced workstation that deserve mention. First, the system permitted the integration and merging of information from diverse data sources. For example, we could combine radar and satellite images; we could overlay station plots on image data; we could overlay streamlines, isobars, and isotherms on image data. Second, the system provided rapid updates of information, an essential feature for observing and forecasting local, short-lived weather phenomena. Third, the system displays were in color, permitting a wide array of color enhancements of the information content in satellite and radar images. Fourth, the system had the capability to loop forward and backward through a series of four frames containing 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 were built-in features of the display system. Last, the workstation was user-friendly. Products were easily selected from a

Table 1.--Data sets available on the 1982 advanced workstation

Data sets	Update rate (min)
Digital GOES satellite imagery (IR and visible images)	30
Digital radar data from Limon, Colo. and Cheyenne, Wy.	5
Surface mesonet observations of 5-min average values of wind, temperature, dew point, pressure, and visibility. Accumulated (5-min) values of precipitation*	5
Air-to-ground lightning detection (location and frequency)†	5
NWS Surface Aviation Observations (SAO's)	60
NWS upper-air (rawinsonde) observations	720

* See Pratte and Clark (1983).
† See Brown (1983).

separate menu screen with a light pen. It is the consideration of these attributes that leads to the hypothesis that the capabilities of the advanced workstation should provide improvements in local, short-term forecasts.

The technology available in many WSFO's includes a set of alphanumeric and graphic terminals on the NWS Automation of Field Operations and Services (AFOS) system. The AFOS system (Klein, 1976) provides the meteorologist with conventional surface and upper-air observations and centralized products, analyses, and weather forecasts prepared by the National Meteorological Center (NMC) and the National Severe Storms Forecast Center (NSSFC). (The AFOS terminals in the PROFS forecast office are used as a data receiver system, operating as a node off the Denver WSFO distribution point.) The AFOS system is augmented by local radar data at all NWS WSFO's responsible for severe weather warnings. Many WSFO's display radar reflectivity data in the form of Plan Position Indicators (PPI's) on a Kavouras color radar display system (Design News, 1980). Reflectivities are displayed in six levels of intensity, each assigned a different color. Data are updated automatically every 3 minutes at each of four selectable ranges: 60, 120, 180, and 240 nautical miles. Satellite data are available at WSFO's on hard-copy satellite photographs as either visible or infrared images every 30 minutes and as occasional water vapor images. The combination of the AFOS system, Kavouras radar display, and Unifax satellite images was available at the PROFS forecast office and at the Denver WSFO. Merging of images; overlaying of alphanumerics and graphics on images; and looping, zooming, and roaming of images are not possible on the WSFO system. It is thus a natural thought process to pursue comparative experiments in order to evaluate an advanced workstation.

In this report, we refer to the combined capabilities of the AFOS terminals, the Kavouras radar display, and the satellite hard-copy system as the AFOS station. We refer to the advanced workstation as the PROFS workstation. In summary, there were three workstations in the 1982 experiment: separate PROFS and AFOS stations in the PROFS forecast office in Boulder, and a merged AFOS and PROFS workstation in the Denver WSFO.

Within the overall experiment, two types of forecast exercises were conducted. One was to prepare severe storm warnings as deemed necessary at each workstation. This permitted a comparative evaluation effort based on three sets of warnings: those 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.

It is customary NWS practice to issue severe storm warnings either in response to observations of severe weather or in response to decisions based on current (and recent) radar and satellite data. Observations come from a variety of sources, including NWS-trained weather spotters, public law enforcement officials, private meteorologists, and the general public. During this exercise, there were also observations available from chase teams supporting a National Center for Atmospheric Research (NCAR) project. All these severe weather observations were available in common to all three workstations.

The other forecast exercise was to prepare short-term point forecasts of wind, visibility, and precipitation for two stations selected from the PROFS surface mesoscale meteorological network (Pratte and Clark, 1983). These forecasts were made only at the two workstations in Boulder. As indicated in Table 1, the surface network data were available to the PROFS workstation forecaster every 5 minutes. These data included such products as station plots, streamlines, isotherms, and time series of selected parameters. All the workstation capabilities of looping, overlaying, and merging of images and graphics, etc., were also available. The AFOS forecaster, on the other hand, was given only the hourly observations of wind, visibility, and precipitation for the two selected stations. Further details of this part of the exercise are given in Section 5.

2.2 Forecaster Staffing

The PROFS workstations were staffed with forecasters drawn from PROFS, WRP, NESDIS/RAMM, and NCAR.* There were two forecaster categories: lead forecaster and assistant forecaster. The lead forecaster was responsible for the forecasts prepared at the assigned workstation and for training the assistant forecaster. The selection of lead forecasters was based on the amount of forecast experience or expertise in a particular specialty such as radar or satellite meteorology. Lead forecasters averaged about 2 years forecasting experience, mostly in meteorological field experiments or making forecasts for personal use. In particular, the lead forecasters were not experienced in

*WRP is the Weather Research Program of the Environmental Research Laboratories (ERL) of the National Oceanic and Atmospheric Administration (NOAA). RAMM is the Regional and Mesoscale Meteorology branch of the National Environmental Satellite, Data, and Information Service (NESDIS) of NOAA.

making daily operational forecasts as are the NWS forecasters. The assistant forecasters had little or no forecast experience.

Every meteorologist had one tutorial day on both of the PROFS workstations before beginning the forecast shift. New forecasters were rotated through 8-day cycles of daily forecast shifts from noon to 8:00 p.m. local time. Each assistant forecaster worked 4 days on each workstation in turn. There were two assistant forecasters on duty at all times. The lead forecaster, on the other hand, alternated between the workstations (1 day on the AFOS workstation, the next day on the PROFS workstation) for a total of 8 days. Thus, only one of the workstations had a lead forecaster on duty any particular day. The assistant forecaster was solely responsible for the forecasts at his workstation on the days he worked alone.

The above work schedules were followed without exception for the assistant forecasters. There were occasions, however, during the summer when no lead forecaster was present at either workstation. This occurred because many of the lead forecasters were from organizations other than PROFS; they occasionally were absent from their assigned shifts because of high priority requirements from their parent organizations.

The PROFS workstation at the Denver WSFO was staffed daily from noon to 8:00 p.m. by an experienced meteorologist, usually a PROFS staff member. The major responsibility of this person was to aid the WSFO severe weather forecasters in using the PROFS workstation. This person often discussed the current weather situation with the duty forecaster and provided guidance in the interpretation of the displays. The decision to issue official severe storm warnings remained a sole responsibility of the WSFO forecaster, however.

2.3 Forecast Procedures

Severe storm watches and warnings were prepared as deemed appropriate by the forecasters at each workstation. There were two differences between the severe weather forecast procedures followed in the PROFS forecast office and those followed in the Denver WSFO, however. The PROFS forecasters were permitted to issue watches as well as warnings for their area of responsibility (all of Colorado east of the Continental Divide). The WSFO forecasters issue only warnings for severe thunderstorms and tornadoes, and watches and/or warnings for flash floods. The intent was to compare watches with those issued by the National Severe Storms Forecast Center (NSSFC) and warnings with those issued by the Denver WSFO. As it turned out, insufficient watches were issued by the PROFS forecasters to warrant this comparison.

A more important difference was the procedure for designating the warning areas. At the PROFS workstations, warning areas were specified by a three- or four-sided polygon drawn on warning forms. The warning areas specified by the WSFO forecasters followed standard NWS procedures which specify use of geopolitical boundaries, viz., counties or portions of counties. These areas were drawn as four-sided polygons on warning forms by the PROFS meteorologist on duty at the WSFO.

Point probability forecasts of wind, visibility, and precipitation threshold values were prepared hourly for two stations: Aurora (Stapleton International Airport) and Fort Collins. They were made for the 15-min periods

ending 15, 30, 60, and 120 minutes after forecast time.

The PROFS workstation forecasters had access to the AFOS data base at any time during the forecast shift, but AFOS forecasters were prohibited from viewing the PROFS workstation. As a general rule, the AFOS and PROFS workstation participants briefed themselves at the beginning of each shift by reviewing the synoptic-scale data and prognoses available on AFOS. The WSFO forecasters were not required to use the PROFS workstation in the preparation of their warnings, but the PROFS staff member on duty between noon and 8:00 p.m. assisted the WSFO forecasters in its use if they so desired. It was agreed that since the NWS forecaster had official warning responsibilities for the general public, his use of the PROFS workstation should be voluntary, not mandatory. No records were kept of WSFO forecaster usage of the advanced workstation.

The intent of the PROFS forecast office procedures was to separate the AFOS and PROFS workstations. As the summer went on, however, the separation in the PROFS forecast office was frequently compromised. Daily, the attraction and stimulation of the PROFS workstation drew many interested visitors who looked over the shoulders of the forecasters. They occasionally made comments and even offered advice, usually to the PROFS workstation forecaster, despite requests for silence. In addition, the AFOS and the PROFS workstation were located in the same room, separated by a 6-ft partition. However, all visitors' comments and questions could be heard at both forecaster stations. These conditions may have influenced some of the comparison statistics developed from this exercise.

3. SEVERE WEATHER VERIFICATION PROGRAM

3.1 Characteristics of Verification Data

Evaluation of forecasts requires comprehensive verification of the events. Verification of severe weather occurrence, however, is particularly difficult. Severe thunderstorms and tornadoes are extremely rare events. A thunderstorm must produce hail greater than 0.75 inches in diameter and/or wind gusts of greater than 50 knots to be classified as severe according to NWS procedures. Severe thunderstorms, tornadoes, and flash floods can be verified only by visual observations. The quality and the quantity of severe storm verification data therefore depend heavily on the number of observers, their degree of training in scientific observation procedures and specific severe storm identification, and the fortuitous occurrence of storm tracks in regions where observers are located. This last factor is particularly pertinent to Colorado severe weather in that there are large unpopulated areas with few roads. Does the lack of a severe weather observation in a particular convective situation automatically mean that severe weather did not exist? Does a severe weather report automatically provide usable verification data? Unfortunately the answer to these questions is no. In short, severe weather verification data in any severe storm study depend at least partly on human judgment and luck. Thus, in obtaining the verification data set we used for evaluation purposes, we found it necessary to reject a small number of reports, which in our judgment lacked credibility. Fortunately, however, these shortcomings are considerably less troublesome for comparison evaluations than for evaluating absolute levels of skill.

The sources of verification data for this exercise were severe weather reports collected by the National Weather Service offices in Denver, Colorado Springs, and Pueblo; reports sent in by or solicited from public officials; a written log and accompanying photographs from chase teams; and reports from private spotter networks and a variety of other data sources. Occurrences of convection, both severe and non-severe, were verified in real time and after the fact.

3.2 Denver WSFO Verification Data

The PROFS staff member at the Denver WSFO recorded information on convective activity received from the public, NWS spotters, law enforcement and other public officials, and a ham operator network that is activated when severe weather occurs close to Denver. The PROFS monitor immediately relayed all severe weather reports to the lead forecaster in Boulder who, in turn, informed the PROFS workstation and the AFOS assistant forecasters.

After-the-fact information, including clippings from newspapers around the state, was exchanged between the severe weather focal point at the Denver WSFO and PROFS to help complete the set of verification data. Additional severe weather reports were obtained semi-monthly from the Weather Service Offices (WSO's) in Colorado Springs and Pueblo.

3.3 Telephone Call Program

Additional after-the-fact information on storm occurrence and severity was obtained from county civil defense and emergency preparedness authorities, county sheriffs, agricultural extension service agents, police and fire departments. We telephoned them the day after convective activity occurred in their county to ask about type of storm, time, duration, estimated damage, etc. We normally obtained accurate information; occasionally we discounted a report because storm location or time of occurrence was vague. We soon learned that this part of the verification exercise depended on our being able to contact interested, committed officials.

Sometimes telephone calls could not be completed. We then sent a letter requesting verification of convective episodes. Twenty letters were sent to various sources in the course of the summer; 14 replies were received.

3.4 Chase Program

Another source of information was a group of meteorologists from ERL and NCAR who agreed to gather information on storms for PROFS and for the Joint Airport Weather Studies (JAWS) Project. The objective of JAWS, involving NCAR and the University of Chicago (McCarthy et al., 1982), is the study of the convective microburst event.

Early in the day, as convective weather began, chase team volunteers met at PROFS and decided when and where to gather the best information on the convective activity. When teams went out, any weather report of interest to JAWS or PROFS was radioed in to the JAWS radar site. Radio communications between

JAWS and chase car personnel were monitored by the AFOS and the PROFS workstation during most of the summer. JAWS radar operators directed the car when it was within radio range; otherwise, the chase team members depended on infrequent contact with PROFS by telephone or on their own knowledge. The radar operators relayed reports to PROFS and the Denver WSFO.

The next day a detailed report including a map of routes driven, weather observations, and any photos taken (complete with photo log), was turned in to the verification staff. Over a dozen tornadoes and funnel clouds were filmed by the chase teams during the summer. The Denver WSFO staff also used the reports as part of their own real-time and after-the-fact severe weather verification efforts.

3.5 Spotters

Spotters are essential for verification of convective activity. Most spotters are volunteers enlisted by the Denver WSFO and trained to recognize severe weather, but there are many other people in the Front Range area who can assist in reporting severe weather, including private meteorologists (some with their own spotter networks), research meteorologists at ERL and NCAR, and meteorology students, primarily at Fort Collins, Boulder, and Denver. To reach as many of these as possible, we presented several widely advertised seminars at Front Range locations and informed the attendees of the summer experiment, stressing our need for accurate verification data.

At each seminar we distributed franked post cards for persons to mail in with information on type of severe weather, time of occurrence, location, etc. We also published an in-state WATS (Wide Area Telephone Service) number that people could call, 24 hours a day, leaving name, phone number, and verification information. A PROFS staff member telephoned any caller the next day to get detailed information on severe weather.

The WATS line was used mainly by persons who lived outside the Denver metropolitan area, or by chase team members in the field. The phone was staffed by PROFS forecasters during the real-time test and placed on automatic record after hours.

3.6 Verification Data File

The sum of these activities was a daily tabulation of specific weather events, including type, time of occurrence, location, severity, name of observer, and a text message. All information, current or after-the-fact, was then entered and stored in the PROFS computer data base. Nearly 900 convective reports were collected during the experiment.

3.7 Discussion

The distribution of verification reports according to source is presented in Fig. 1 for severe weather and in Fig. 2 for non-severe weather. It is noteworthy that normal Denver WSFO verification efforts contributed 62 percent of the severe weather observations and 53 percent of the non-severe weather observations.

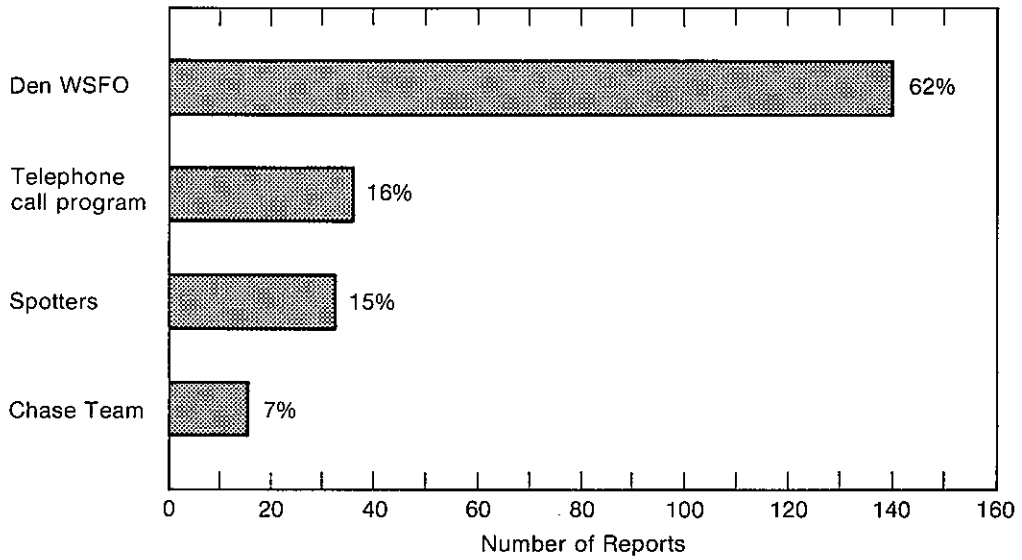

Figure 1.--Severe weather reports ranked by source.

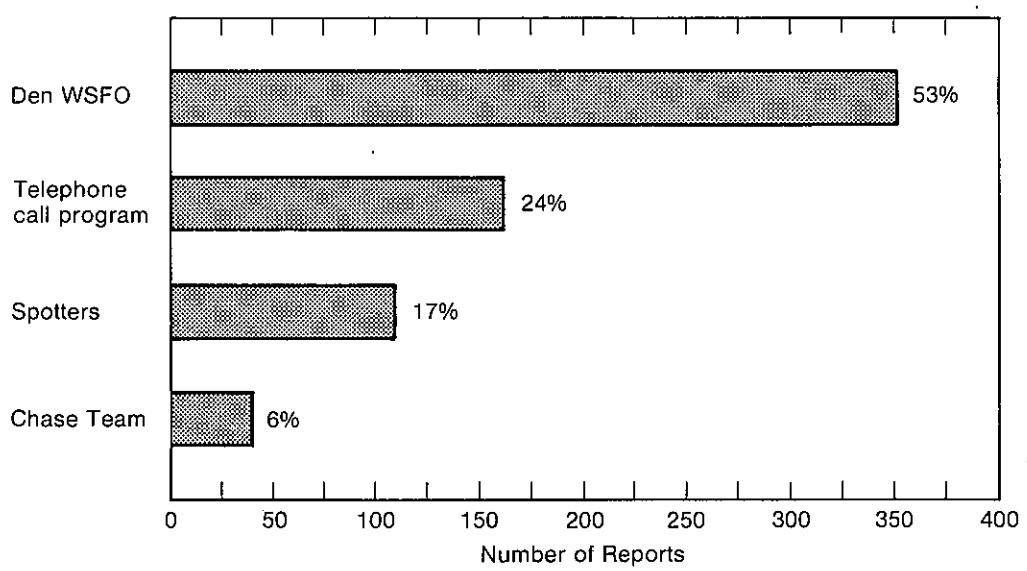

Figure 2.--Non-severe weather reports ranked by source.

An additional 31 percent of severe weather reports and 41 percent of non-severe weather reports were obtained by the supplemental follow-up programs implemented by the PROFS staff. Only 7 percent of the severe weather reports came from the chase team operations. This suggests that normal WSFO verification operations could be improved significantly with relatively little additional effort or expense by implementation of a telephone call program and augmentation of spotter networks.

Chase teams are costly and difficult to justify on a regular basis. Nevertheless, they provide essential information, and we plan to increase

their value by scheduling experienced chasers rather than relying on volunteers and by improving the area coverage of radio communications.

The verification data files for this exercise are recorded on hard disk. They are available to any interested person. Inquiries should be sent to Director, PROFS, 325 Broadway, Boulder, CO 80303.

4. WARNING EVALUATION RESULTS

4.1 Objective

The objective of the evaluation is to examine the hypothesis that the POWS capabilities helped improve severe weather warnings. The initial approach was to compare skill scores computed for the three workstations, but that led to questions about the impact of forecaster experience on the results, the statistical stability of skill scores based on the occurrence of rare events, and the impact of warning area sizes and warning duration times on skill scores.

4.2 Definition of Data Sets

Forecasters at each workstation in Boulder manually entered watches or warnings on a form. The forecaster specified type of issuance, type of severe event, issue time, effective time, duration time, and area. The forecaster specified areas by drawing three- or four-sided polygons on the map section of the form. Identical information was recorded at the WSFO, although the warning areas were inferred from the text messages of the official warnings, which customarily specify area boundaries (e.g., county lines).

After the summer exercise, we realized that only severe storm warnings could be studied, because there were insufficient watches prepared by the PROFS forecasters to warrant analysis. The warnings selected for analyses were issued between noon and 8:00 p.m. local time for only those days when both AFOS and the PROFS workstation were operational in Boulder. The numbers of warnings issued for each type of severe weather event by each workstation for the selected days are given in Table 2. There are very few tornado warnings and even fewer flash flood warnings in these data sets. Accordingly, we

Table 2.--Total number of warnings issued at each station by type of severe weather event

Event	AFOS	PROFS	WSFO
Severe thunderstorm	88	106	109
Tornado	17	16	22
Flash flood	7	9	6
Total	112	131	137

have chosen to study the severe thunderstorm warnings in some detail, the tornado warnings to a limited extent, and the flash flood warnings not at all.

4.3 Scoring Procedures

The following scoring indices (Donaldson et al., 1975) were calculated for each workstation:

probability of detection (POD) = X/(X+Z),

false alarm rate (FAR) = Y/(X+Y),

critical success index (CSI) = X/(X+Y+Z),

where X is the number of forecast severe events that occur, Y is the number of forecast severe events that do not occur, and Z is the number of non-forecast severe events. The verification data were used to tabulate values of X, Y, and Z. A hit (X) was recorded whenever one or more severe events occurred within the warning area during the valid warning time. No attempt was made to record hit (X), false alarm (Y), or miss (Z) values for each county covered by a warning area as is customarily done by NWS in official scoring of NWS warnings (Pearson and David, 1979).

Severe thunderstorms and tornadoes were treated as separate events. That is, if large hail and a tornado were reported from the same convective storm, and no severe weather warning was issued, two misses were recorded, one for severe thunderstorms and one for tornadoes. Further, a severe thunderstorm warning verified if a tornado was reported within a thunderstorm warning area. But if a concurrent tornado warning was not in effect, the tornado observation was recorded as a miss for a tornado warning evaluation.

No distinction was made between warnings based on decisions made from workstation information alone and those issued in response to direct storm observations. Warnings issued from any of the three workstations and based on immediate response to spotter, chase team, or public official reports were automatically defined as hits. Since these spotter reports were available at each workstation, this provides a common element in the probabilities of detection computed for the workstations, but not so for the false alarm rates and the critical success indices.

4.4 Discussion of Results

The scores for the severe thunderstorm warnings are listed in Table 3 along with the tabulations of hits, false alarms, and misses for each workstation.

We note first that the AFOS and the PROFS workstation scores differ only in the probability of detection. The false alarm rates and the critical success indices are identical. The WSFO results, on the other hand, are consistently better for all three scores. These results are a direct consequence of the number of hits, false alarms, and misses recorded for each workstation. The misses are nearly the same, but the number of hits increases steadily from the AFOS to the PROFS to the WSFO workstation. Further, although the false

Table 3.--Severe thunderstorm warning scores, hits, false alarms, and misses for each workstation

	AFOS	PROFS	WSFO
POD	0.36	0.41	0.45
FAR	0.80	0.80	0.77
CSI	0.15	0.15	0.18
X	18	21	25
Y	70	85	84
Z	32	30	31

alarms at the AFOS station are fewer than those recorded for the PROFS and WSFO stations, the FAR and CSI skill scores are not better because of the fewer recorded hits.

These results, taken at face value, do not support our hypothesis that the use of advanced technology will improve the accuracy of severe weather forecasts. There are at least three possible reasons for this. First, the separation between the two workstations in Boulder was inadequate because of the lack of visitor control. The very entrance of visitors into the forecast office engendered a heightened feeling of awareness at each workstation. The lack of isolation further tended to blur possible differences between the two workstations.

Second, the WSFO results seem to suggest strongly the effect of forecaster experience. The WSFO staff were experienced operational forecasters (7 to 20 years experience, with 5 to 12 years experience in Colorado). This experience level alone could have been sufficient to produce the above results. In addition to their experience, the WSFO forecaster had the additional benefit of an advanced workstation and a PROFS meteorologist in attendance providing assistance in the use of the station's advanced capabilities. However, it is not possible to separate out the relative value of experience and advanced workstation capabilities in these results.

Third, the range in the absolute values of the scores is quite small. Slight changes in the recorded values of hits, false alarms, or misses at any of the three workstations could have produced different results. Stated differently, only dramatic large differences between the workstations would support credible inferences with respect to our initial hypothesis. Since large differences did not result, we can state only that experience levels are surely important. Here again, it is not possible to differentiate between the effects of experience and of advanced workstation capabilities.

The scores for the tornado warnings are listed in Table 4 along with the tabulations of hits, false alarms, and misses for each workstation.

Table 4.--Tornado warning scores, hits, false alarms, and misses for each workstation

	AFOS	PROFS	WSFO
POD	0.11	0.16	0.33
FAR	0.82	0.75	0.59
CSI	0.07	0.11	0.23
X	3	4	9
Y	14	12	13
Z	24	21	18

In contrast to the situation for the severe thunderstorms, there is a consistent progression in all skill scores from the AFOS to the PROFS to the WSFO workstation for the tornado warnings. We note, comparing AFOS and PROFS, that these results are consistent with our hypothesis. These two stations had overall comparable levels of experience so that the differences might well be inferred to result from the advantages of the advanced workstation capabilities. But we also note that the sample size is surely too small to lend credibility to such an inference. That is, the differences between 3 hits, 24 misses (AFOS) and 4 hits, 21 misses (PROFS) are small differences that well could have occurred by chance. Indeed, it is also possible that the differences are quite small because of the lack of isolation between the two workstations.

The WSFO results, on the other hand, appear to us to be much more impressive. There is substantial improvement in the WSFO scores over the scores for either AFOS or the PROFS workstation. Unfortunately, as already noted, we cannot separate the relative effects of forecaster experience and advanced workstation capabilities. It remains for future experiments to focus on that issue more sharply.

We now turn to a more detailed analysis of the severe thunderstorm warning scores listed in Table 3. As noted in Section 2.2, the staffing schedule provided days when the assistant forecaster worked alone on either the AFOS or PROFS workstation. Furthermore, there were times when the scheduled lead forecaster was not present at either station. This state of affairs provides an opportunity to examine further the relative effect of forecaster experience vs. advanced workstation capabilities. We define three subsets of days as follows:

Category I: Those days when the assistant forecasters were each alone at the AFOS and the PROFS workstation.

Category II: Those days when a lead and an assistant forecaster (team of two) were at the AFOS workstation; an assistant forecaster was alone on the PROFS workstation.

Table 5.--Comparative severe thunderstorm warning scores for Category I: assistant forecaster alone at both AFOS and the PROFS workstation*

	AFOS	PROFS	WSFO
POD	0.21	0.38	0.24
FAR	0.77	0.76	0.76
CSI	0.13	0.17	0.13
X	3	6	4
Y	10	19	13
Z	11	10	13

*Number of days: 7

Table 6.--Comparative severe thunderstorm warning scores for Category II: lead forecaster and assistant forecaster (team of two) at AFOS, assistant forecaster alone at the PROFS workstation*

	AFOS	PROFS	WSFO
POD	0.44	0.40	0.54
FAR	0.77	0.81	0.73
CSI	0.18	0.14	0.22
X	11	10	14
Y	36	44	38
Z	14	15	12

*Number of days: 11

Category III: Those days when a lead and an assistant forecaster (team of two) were at the PROFS workstation; an assistant forecaster was alone at the AFOS workstation.

As noted previously, the assistant forecasters had little or no prior forecast experience. They were all reasonably proficient with the operation of AFOS and the PROFS workstation, but had no developed forecast skills. The lead forecasters were, by comparison with the assistant forecasters, considerably experienced in forecast skills and were also proficient in the operation of both workstations. And, as already discussed, the WSFO forecasters were the most experienced, with well-developed forecast skills and with long-term experience in forecasting Colorado severe weather. The skill scores for each category are tabulated in Tables 5, 6, and 7.

Table 7.--Comparative severe thunderstorm warning scores for Category III: lead forecaster and assistant forecaster (team of two) at the PROFS workstation, assistant forecaster alone at AFOS*

	AFOS	PROFS	WSFO
POD	0.36	0.50	0.54
FAR	0.86	0.81	0.83
CSI	0.11	0.16	0.15
X	4	5	7
Y	24	22	33
Z	7	5	6

*Number of days: 11

We note first, from Table 6, that the experienced lead forecaster on AFOS outperformed the assistant forecaster on the PROFS workstation. From Table 7, we note that the experienced lead forecaster on the PROFS workstation outperformed the assistant forecaster on AFOS. In both of these categories, the experienced forecasters at the WSFO performed best except that they had a disproportionately high number of false alarms in Category III. This aside, these skill scores imply that forecaster experience is indeed a strong factor in improved forecaster accuracy.

The only comparison between workstations with forecasters of comparable skill is between the assistant forecasters working alone at AFOS and the PROFS workstation, as seen in Table 5. The PROFS scores are better than the AFOS scores for this category and are even slightly better than the WSFO scores. We do not infer from this that the advanced workstation capabilities led to the improved performance of the assistant forecaster over his counterpart at the WSFO. Rather, this is more likely the result of the natural variability one expects in small sample statistics.

We conclude that the results shown in Tables 5-7 imply that forecast experience is measurable and that it is an important factor in performance. However, these conclusions are tentative at best. Our sample sizes are small. There are some deficiencies in the experimental procedures that tend to compromise the statistics for comparative purposes. Although small sample sizes can always be a factor in experiments of this sort, procedural deficiencies need not be. They will be corrected in future PROFS experiments.

In the course of our analyses, we noted that warnings issued by the PROFS forecasters were often for smaller areas and shorter durations than those issued by the WSFO forecasters, although the assistant forecaster often issued warnings for larger areas and longer durations than his more experienced counterpart. However, the median values for each workstation were nearly identical, a warning area of 2000 square kilometers and a warning duration time of 60 minutes. In future experiments we plan to investigate a scoring procedure

that rewards a forecaster for specifying small warning areas and short warning times because the advanced workstation lends itself easily to frequently updated and site-specific warnings. This in turn suggests the possibility of developing an improved nowcasting service that effectively communicates site-specific warnings and reasonably short warning times to the public. This concept needs further testing, particularly in nowcasting dissemination techniques. It will be the subject of future PROFS experiments.

5. PROBABILITY FORECAST EXPERIMENT

5.1 Experimental Procedures

Probability forecasts were made on the hour for four periods, each 15 minutes long, ending 15, 30, 60, and 120 minutes after the issuance time for Aurora and Fort Collins, two stations in the PROFS mesonet. For each period, the forecaster estimated the probabilities of occurrence for three wind speed categories: $\geq$ 20, $\geq$ 35, and $\geq$ 50 knots; for visibility $\leq$ 1 and $\leq$ 3 miles; and for four mutually exclusive rainfall categories: trace or less, 0.01-0.03 inches, 0.04-0.08 inches, and greater than 0.08 inches. The probabilities for the precipitation forecasts had to add to 100 percent.

These forecasts were prepared only at AFOS and the PROFS workstation. Radar and satellite data were available at each workstation (see Tables 1 and 2). The mesonet data were available at the PROFS workstation every 5 minutes in the form of station plots, streamlines, and time series of selectable parameters. Once an hour the AFOS forecaster was given only the Aurora and Fort Collins mesonet observations of wind, visibility, and precipitation. The full data integration and display capability of the PROFS workstation encouraged overlays of mesonet data, primarily on radar images that were updated every 5 minutes. The comparison of the forecast results thus measures the value of data from mesoscale meteorological networks combined with radar and satellite data for short-range probability forecasts, but not the value of mesoscale data alone since AFOS and the PROFS workstation did not have the same set of mesoscale data.

5.2 Evaluation Procedures

The Brier score (Brier, 1950) is the statistic we use here for comparing the point forecasts. It is computed from

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

where F_i is the forecast probability; O_i is the observed probability, equal to 1 if the event occurred, 0 if it did not; and N is the total number of forecasts. Brier scores range from 0 for a set of perfect forecasts to 1 for a set of complete misses.

We have elected to examine only those cases for which the value of the parameter for the forecast period differed from the value at the time of the forecast, i.e., those cases for which a persistence forecast would not verify.

This emphasizes those weather situations when the parameter values are changing, usually because of nearby convective events.

Change in wind speed category can occur from the influence of thunderstorms or more widespread mesoscale or synoptic-scale influences. A change in visibility or precipitation category requires a rain cell over the station. Since the elements forecast in this experiment occurred rarely, we were able to obtain statistics only for the category of wind speed greater than 20 knots. 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 for which the observed category for the forecast period differed from the observed category at the time the forecast was issued; (2) those cases for which the wind speed at the forecast time was less than 20 knots but greater than 20 knots for the forecast period; (3) those cases for which 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).

5.3 Results

The Brier score results for each of the three groupings for the wind speed forecasts are given in Figs. 3, 4, and 5. The Brier scores for the PROFS workstation forecaster are consistently better than those for the AFOS forecaster for all three groupings. This tends to support the hypothesis that the advanced workstation capabilities provide improved short-range forecast accuracies. In particular, the data integration and display capabilities,

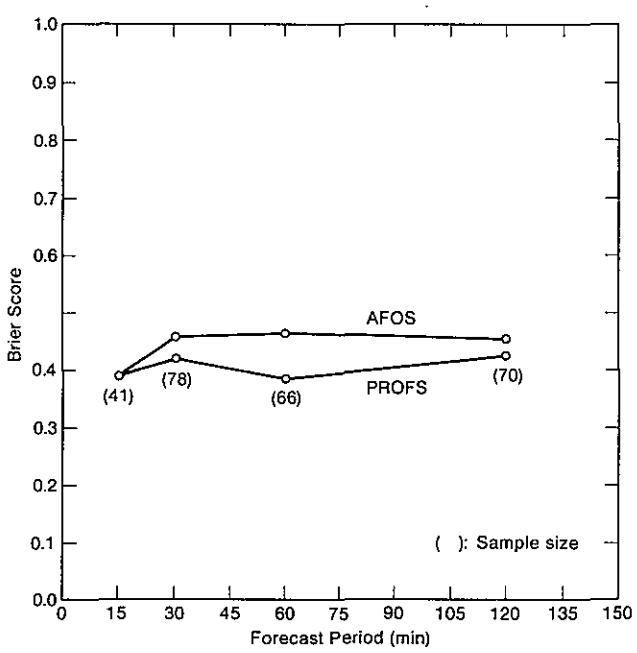

Figure 3.--Brier scores for 20-kn wind speed category.

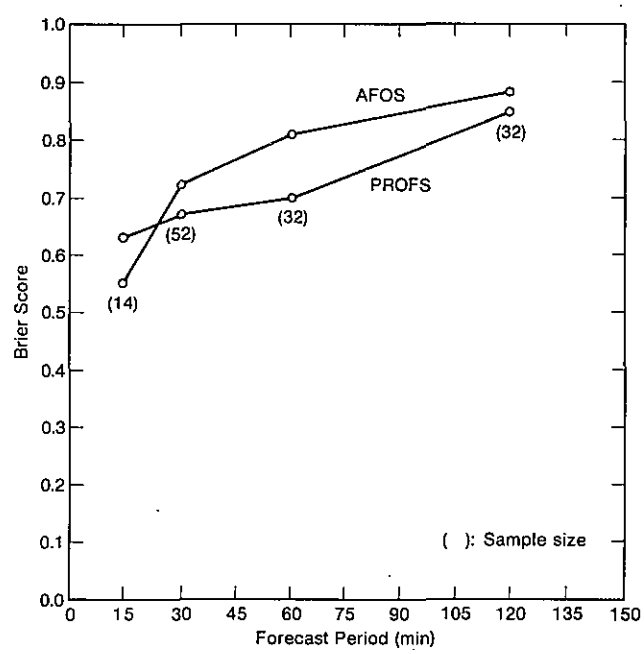

Figure 4.--Brier scores for 20-kn wind speed category, initial condition winds < 20 knots.

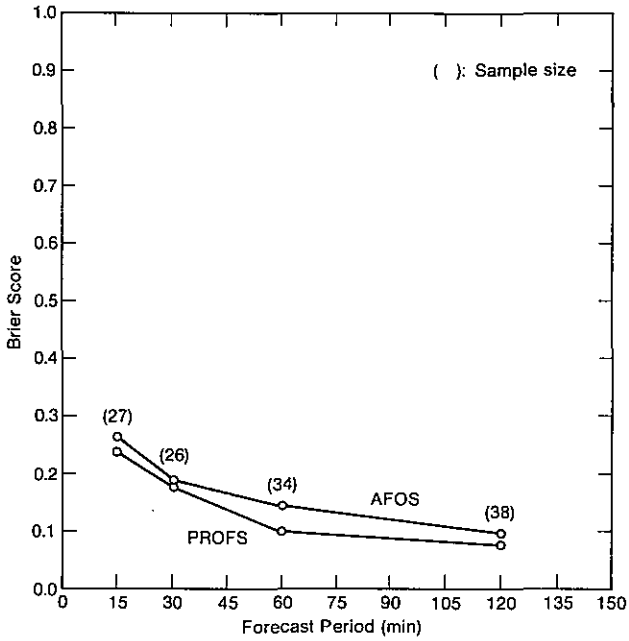

Figure 5.--Brier scores for 20-kn wind speed category, initial condition winds > 20 knots.

especially the looping and overlay of mesonet data and the display of 5-min interval time series of single-station mesonet data, could be capabilities that contributed to the improved skill. None of these capabilities were available to the AFOS forecaster. The sample size, however, is too small to infer a statistically significant difference between the results.

We note also that the skill is low, and worsens with length of forecast period for the cases for which the wind speeds were initially less than 20 knots. The inverse is true for those cases for which the wind speeds were initially greater than 20 knots. That is, forecasting the onset of strong winds 1 to 2 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 short-term probability forecasts of these particular elements.

6. SUMMARY

These analyses are the result of a forecast comparison experiment that, at best, must be considered preliminary in nature. The matters of design, execution, and analysis of experiments for which conclusions must depend on statistical inferences have not been explicitly discussed in this report. These matters, however, are extremely important, and procedures for designing and conducting such experiments are well established (Flueck, 1982).

Our results are not supported by tests of statistical significance because of various imperfections in the experiment design (discussed in Sections 2 and 4), and the small samples inherent in the severe storm problem. However, our analyses suggest that improved warning scores will result by (1) the operational use of the data integration and display capabilities of advanced workstations, (2) intensified and improved training of severe weather forecasters, and (3) improved severe weather verification. The second and third

suggestions could be implemented even before advanced technologies become operational.

We look forward to future forecast experiments with improved designs and yet further advanced data sets incorporated into the PROFS workstation. In particular, Doppler radar velocities and rapid-scan satellite image data should be important additions to the PROFS workstation in future experiments.

7. ACKNOWLEDGMENTS

Prof. Colin S. Ramage, University of Hawaii, has made many contributions to this report. His critical reviews and stimulating, provocative discussions have done much to influence the work that was performed. Dr. Thomas W. Schlatter, PROFS, and Dr. John M. Brown, WRP, contributed many valuable suggestions in a critical review of the drafts leading to this manuscript. Dr. A. E. (Sandy) MacDonald, PROFS Director, has provided constant support and encouragement throughout the course of this work. Our special thanks to these people for their help and support.

8. REFERENCES

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

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

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

Donaldson, R. J., Jr., R. M. Dyer, and M. J. Kraus, 1975. An objective evaluator of techniques for predicting severe weather events. Preprints, 9th Conference on Severe Local Storms, Am. Meteorol. Soc., Boston, 321-326.

Flueck, J. A., 1982. Comparative experimentation: some principles and prescriptions. Teaching of Statistics and Statistical Consulting, Academic Press, New York, 443-463.

Kelly, D. L., and J. T. Schaefer, 1982. Implications of severe local storm warning verification. Preprints, 12th Conference on Severe Local Storms, Am. Meteorol. Soc., Boston, 459-462.

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

McCarthy, J. M., J. W. Wilson, and T. T. Fujita, 1982. The joint airport weather studies project. Bull. Am. Meteorol. Soc. 63:15-22.

Pearson, A., and C. L. David, 1979. Tornado and severe thunderstorm warning verification. Preprints, 11th Conference on Severe Local Storms, Am. Meteorol. Soc., Boston, 567-568.

Pratte, J. F., and R. J. Clark, 1983. PROFS mesonet: description and performance. Preprints, Fifth Symposium on Meteorological Observations and Instrumentation, Am. Meteorol. Soc., Boston, 303-307.

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

ACRONYMS

AFOS	Automation of Field Operations and Services
CSI	Critical Success Index
ERL	Environmental Research Laboratories
FAR	False Alarm Rate
IR	Infrared
JAWS	Joint Airport Weather Studies
NCAR	National Center for Atmospheric Research
NESDIS	National Environmental Satellite, Data, and Information Service
NMC	National Meteorological Center
NSSFC	National Severe Storms Forecast Center
NWS	National Weather Service
POD	Probability of Detection
PPI	Plan Position Indicator
PROFS	Program for Regional Observing and Forecasting Services
RAMM	Regional and Mesoscale Meteorology (branch of NESDIS)
SAO	Surface Aviation Observations
WATS	Wide Area Telephone Service
WRP	Weather Research Program
WSFO	Weather Service Forecast Office
WSO	Weather Service Office