Ninth Conference on Aviation, Range, and Aerospace Meteorology,
11-15 September, Orlando, FL

5.19 CONVECTIVE INTERCOMPARISON EXERCISE: BASELINE STATISTICAL RESULTS

Jennifer Luppens Mahoney[1], Barbara G. Brown[2], Cynthia Mueller[2], and Joan E. Hart[3]

1. INTRODUCTION

Various convective forecasts, which are often used by air-traffic decision-makers, were verified in near real-time from 1 June - 31 August 1999 using the Real-Time Verification System (RTVS; Mahoney *et al.* 1997) to develop a statistical baseline for the quality of convective forecasts and to illustrate the underlying differences in the forecasts. The forecasts used in this convective forecast intercomparison fell in two categories. One category includes forecasts that are issued frequently and produce short 1 - 2 h forecasts, such as the National Convective Weather Forecast (NCWF) and C-SIGMETs (Convective Significant Meteorological Advisories). The other forecasts extend from 2 - 6 h and are issued less frequently, such as the Collaborative Convective Forecast Product (CCFP) and the C-SIGMET Outlooks (Outlooks).

Basic statistical results of the evaluation, and the underlying differences in the forecast products are presented in this paper, with the data described in Section 2 and the methodology summarized in Section 3. The results are presented in Section 4, with future plans described in Section 5.

2. DATA

The convective forecast products and verifying observations are briefly described in this section.

2.1 Forecast Products

Collaborative Convective Forecast Product (CCFP) - This experimental forecast was generated from input provided by participating airline, Center Weather Service Units (CWSU) and Aviation Weather Center (AWC) meteorologists, and staff at the FAA Air Traffic Control System Command Center. The CCFP product was generated as a graphic depicting forecasts of convective activity valid at specific times. The forecast product was ultimately used by decision-makers for routing traffic around convective areas (Phaneuf and Nestoros 1999).

First Guess Forecast (FG) - The FG forecast (Phaneuf and Nestoros 1999) was generated by AWC meteorologists as a precursor to the CCFP. The forecast was made available to the CCFP participants, who evaluated the forecast and provided feedback that was ultimately incorporated into the CCFP.

Convective SIGMET (C-SIGMET) - This product, generated by AWC forecasters, is an operational text forecast of convective activity. The forecast is issued hourly and is valid for up to 2 h (NWS 1991). More recently, however, these forecasts are issued as a convective detection field (personnel communication AWC), but were not evaluated as such in this exercise. The forecasts are issued to capture severe or embedded thunderstorms and their hazards (*e.g.,* hail, high winds) that are either occurring or forecasted to occur for more than 30 minutes of the valid period. C-SIGMETs are also issued for thunderstorm lines and areas of active thunderstorms affecting at least 3,000 square miles. For this evaluation, the C-SIGMETs are treated in three different ways; as a forecast of 1 h duration, valid at the *end* of the period; a forecast of 2 h duration, valid at the *end* of the period; and 2 h, valid *throughout* the entire 2-h period.

Convective SIGMET Outlook (Outlook) - The Outlook is an operational text forecast of convective activity, generated by AWC meteorologists, issued hourly, and valid from 2-6 h after the issuance time of the Outlook (NWS 1991). The forecast area encompasses moving and changing weather over the 4-h period and emphasizes thunderstorms that form lines or active clusters, organized severe thunderstorms, or areas of embedded thunderstorms. For this evaluation, the Outlooks are considered in two ways: as a forecast of 6 h duration, valid at the *end* of the period, and as a forecast of 4 h duration, valid *throughout* the 2 to 6-h period after issuance.

National Convective Weather Forecast (NCWF)- The NCWF, designed and implemented by the National Center for Atmospheric Research (NCAR; Mueller *et al*. 1999), provides current convective hazards and 1 hour extrapolation forecasts of thunderstorm hazard locations. The hazard field and forecasts update every 5 minutes. The NCWF development is sponsored by the Federal Aviation Adminstration (FAA) Aviation Weather Research Program (AWRP; Sankey *et al.* 1997). The NCWF targets airline dispatchers, general aviation, and FAA Traffic Management Units (TMU).

[1] **Corresponding author address**: Jennifer Mahoney, R/FS5 325 Broadway, Boulder, CO 80303.
Email: Mahoney@fsl.noaa.gov
[2] Research Applications Program, National Center for Atmospheric Research
[3] Cooperative Institute for Research in the Environmental Sciences, University of Colorado/Forecast Systems Laboratory, Boulder, CO.

2.2 Verifying Observations

The observations used by RTVS to verify the forecasts include lightning reports, radar data, and the National Convective Weather Detection Product (NCWD; Mueller *et al.* 1999). The results presented in this paper rely on the NCWD only.

The NCWD combines a two-dimensional mosaic of radar reflectivity with radar-derived cloud top data and a grid of lightning detections from the National Lightning Data Network (NLDN; Orville 1991). The cloud top data are primarily used to remove anomalous propagation and ground clutter. The lightning data help to keep the NCWD current, since lightning data have a lower latency than radar data. The NCWD fields were made available on a 4-km grid, with convective storms delineated by a threshold of 40 dBZ, or more than 3 lightning strokes in 10 minutes.

3. METHODOLOGY

3.1 Matching Methods

Before the forecasts were matched to the observations, a 20-km grid (*i.e*., 20 x 20 km) was overlaid on the observation field. Each box on the overlay grid was assigned a *Yes* or *No* value depending on whether a positive observation (i.e., one 4-km NCWD observation with reflectivity greater than 40 dBZ) fell within the 20-km box. The same procedures were applied to the forecasts, with a 20-km box labeled as *Yes* when any part of the forecast polygon intersected that box. If a forecast polygon did not intersect the 20-km box, then a *No* forecast was assigned to the box.

Once the matching process was completed, each 20-km box on the observation grid was matched to each 20-km box on the forecast grid. This technique produced the forecast/observation pairs used to generate the verification statistics. For example, a *Yes* forecast box overlapping a *Yes* observation box produced a *Yes-Yes* pair. Similarly, a *Yes* forecast and *No* observation produced a *Yes-No* pair, and so on, filling the two-by-two contingency table shown in Brown *et al.* (1997).

Observations that fell within a 10-minute time window prior to the forecast valid time were mapped to the 20-km grid and used for verification. To ensure consistency among the results, all forecast products, excluding the 2-6 h Outlook and the 0-2 h C-SIGMET were subjected to this criterion. Different criteria were applied to the 2-6 h Outlook and the 0-2 h C-SIGMET; for these forecasts, all observations within the 4-h and 2-h period, respectively, were mapped to the grid and used to verify the forecasts.

3.2 Statistical Measures

The verification methods used in this study are based on standard verification concepts that take into account the underlying statistical basis for verification, as well as the associated high dimensionality of the verification problem (*e.g*., Murphy and Winkler 1987; Brown *et al.* 1997). The primary verification statistics used in this analysis include the following:

- PODy and PODn are estimates of the proportion of *Yes* and *No* observations, respectively, that were correctly forecast (*e.g.,* Brown *et al.* 1999; Brown *et al.* 1997).
- FAR is the proportion of *Yes* forecasts that were incorrect.
- The Bias represents the ratio of the number of *Yes* forecasts to the number of *Yes* observations and is a measure of over and underforecasting.
- The Critical Success Index (CSI; Schaefer 1990), also known as the Threat Score, is the proportion of hits that were either forecast or observed.
- % Area is the percentage of area of the forecast domain where convection is forecast to occur (*e.g*., Brown *et al*. 1997).

4. RESULTS

Throughout the convective exercise, statistical results were generated by the RTVS and presented on the Web-based interface in near real-time (http://www-ad.fsl.noaa.gov/afra/rtvs/RTVS-project_des.html). A limited selection of those results is presented here.

4.1. Overall Results

Overall results for the exercise are shown in Table 1. The statistics were computed for each forecast type by combining the forecast/observation pairs for all issue times for the 92 days of the exercise. Further discussion of the results for the C-SIGMETs, Outlooks, CCFP and the NCWF can be found in Mahoney *et al.* (2000b), Mahoney *et al.* (2000a), and Brown and Mahoney (2000).

The results in Table 1 suggest that large differences in PODy and % Area are associated with the 2-6 h forecasts. In particular, the Outlooks cover over 14% of the country. These areas are large to allow for developing and moving convective activity over a 4-h period. As a result, the large areas produced by the Outlooks capture a significant amount of convective activity, which is evident in the large PODy values. In contrast, the % Area is smaller for the CCFP forecasts than for the Outlooks, and the CCFP has a correspondingly small FAR. The collaborative manner in which the CCFP forecasts are produced, and the

Table 1. Overall verification results, by forecast product and forecast length. Statistics were computed by combining counts for all days and hours using the NCWD.

Product	Fcst. length	Filter	PODy	PODn	FAR	CSI	Bias	%Area
NCWF	1	None	0.13	1.00	0.32	0.12	0.2	0.51
CCFP	1	None	0.27	0.98	0.80	0.13	1.4	2.25
C-SIGMETs	1	None	0.26	0.98	0.74	0.15	1.0	2.30
NCWF	2	None	0.07	1.00	0.50	0.06	0.1	0.31
C-SIGMETs	2	None	0.18	0.98	0.82	0.10	1.0	2.27
C-SIGMETs	0-2	None	0.21	0.99	0.45	0.18	0.4	2.30
CCFP	3	None	0.31	0.95	0.84	0.11	2.1	5.73
Outlooks	2-6	None	0.44	0.89	0.73	0.20	1.6	14.19
Outlooks	6	None	0.39	0.87	0.93	0.06	6.0	14.11
CCFP	5	None	0.28	0.94	0.86	0.10	2.1	6.20
CCFP	7	None	0.24	0.96	0.89	0.08	2.3	4.81

specific time at which they are valid, both may contribute to the smaller areas. As a consequence, however, the PODy values are smaller for the CCFP than for the Outlooks.

A large variation in FAR and Bias are evident in Table 1 comparing the NCWF and C-SIGMETs. In particular, the NCWF statistics have a very small overall Bias, which indicates that the NCWF consistently underforecasts the convective regions. In contrast, the 1-h C-SIGMETs have a Bias of about 1, indicating that the forecast captures approximately the correct amount of convection. Correspondingly, the FAR values are 50% larger for the C-SIGMET point forecasts than for the NCWF. Due to the small areas covered by the NCWF, and the low Bias values associated with those forecasts, the PODy values for the NCWF forecast are relatively small.

4.2. *Scaling Influences*

The statistical results presented in Section 4.1, are influenced, in part, by both the scale at which the forecasts are produced and the grid sized used to verify them. A comparison between the 4-km NCWD and the NCWD mapped to the 20-km grid is shown in Fig. 1a. The mapping of the NCWD to the 20-km grid clearly expands the domain of observed convective activity. This mapping also influences the size of the forecast area (not shown), however, possibly to a lesser degree.

The mapping of the NCWD to the 20-km grid has at least some (possibly significant) effect on the statistical results computed for the various forecasts, which makes it difficult to directly compare these forecasts. For instance, the NCWF appears to greatly underforecast the convection (c.f. Fig. 1b and the overall low Bias listed in Table 1). However, statistics computed from 18 May - 11 June 2000 for the NCWF using the 4-km grid with no relaxation to the 20-km grid, indicated a bias of 0.8, PODy of 0.30, FAR of 0.63, and a CSI of 0.20 (Megenhardt et al. 2000). The 20-km grid seems to be the appropriate scale for the C-SIGMETs where the observation area is equal to the forecast area (Fig. 1c). For the CCFP, the convection is under-represented by using the 20-km grid as shown in Fig. 1d. These results suggest that the appropriate grid size used to map the observations should be similar to the scale at which the forecasts are issued and intended to be used.

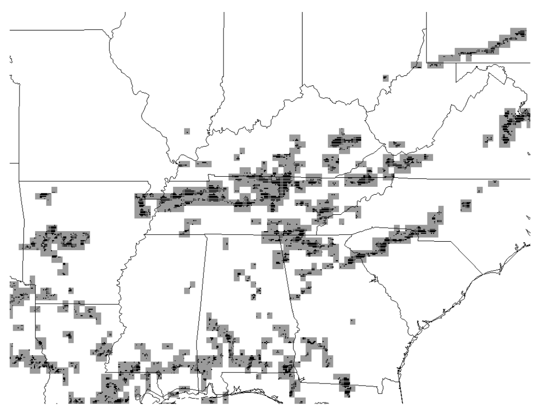

Fig. 1a. Maps of the 4-km (black) and the 20-km (gray) NCWD.

4.3. *Forecast Comparisons*

The basic differences between the algorithms are highlighted in Table 2. As shown in this table, the various forecasts differ from each other in terms of lead-time, valid

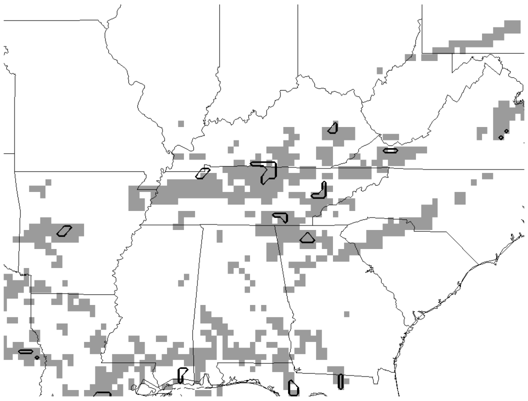

Fig. 1b. Maps of 20-km NCWD (gray) and NCWF (solid black lines).

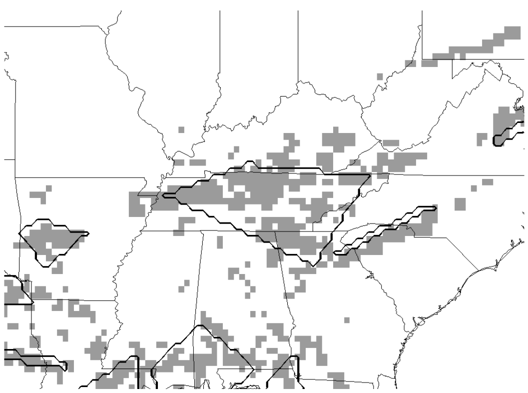

Fig. 1c. Maps of 20-km NCWD (gray) and C-SIGMETs (solid black lines).

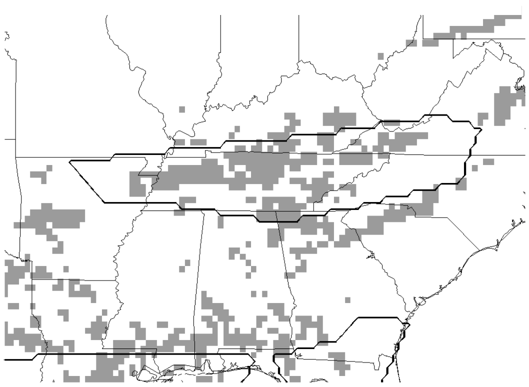

Fig. 1d. Maps of 20-km NCWD (gray) and CCFP (solid black lines).

period/time, and average areal coverage. The NCWF is the only automated algorithm; the others are human-generated. The NCWF and the CCFP are instantaneous snapshots, while the C-SIGMETs and the Outlooks are valid over 2 and 4-h time periods. Moreover, the NCWF focuses on active convection that is expected to persist, and is often associated with long-lived multi-cellular storms; in addition, the NCWF does not focus on isolated convective cells, which are included in the C-SIGMETs. The NCWF also does not attempt to forecast the development of new convective regions, rather it focuses on frequent updates of the movement of existing convective areas. In contrast, the development and movement are considered by the C-SIGMETs and Outlooks. Moreover, the CCFP provides a forecast for convective activity that is expected to directly impact aviation operations. These areas include hazards that are directly associated with the convective activity, but not necessarily covered by a convective cell. One example is the "down wind denied area" which is airspace effected by a cloud's anvil resulting from the convective activity (Foss, personal communication).

The NCWF is an automated extrapolation algorithm that is proficient in moving mature convective cells, but has difficulty with extrapolating cells in the initiation and dissipation stages (Megenhardt *et al.* 2000). This forecast uses observations at 1-km resolution. The C-SIGMETs, CCFP, and Outlook, on the other hand, include forecaster interpretation and analysis and benefit from model-based forecasts at a 40-km resolution.

4.4. Summary of Results

Some characteristics of the quality of the NCWF, 1 and 2 h C-SIGMETs, 6-h Outlooks, and the CCFP are summarized in Fig. 2. Figs. 2a-d are based on concepts from Browning *et al.* (1980), Doswell (1986), Austin *et al.* (1987), and Wilson *et al* (1998) (henceforth BDAW) in an attempt to quantify their assessment of convective forecast accuracy. Fig. 2 was created for each forecast by combining the counts using the unfiltered NCWD over 92 days for all issue times. In some cases, the counts include 24 issue times per day (*e.g.* NCWF). Others include only 1 issue time per day (*e.g.* 1-h CCFP).

The results shown in Fig. 2 are consistent with those presented by BDAW. The NCWF, an automated forecast based entirely on storm extrapolation, rapidly decreases to relatively low values of PODy (Fig. 2a), Bias (Fig. 2c), and CSI (Fig. 2d), with an increase in the FAR (Fig. 2b) after 1 h. The C-SIGMETs and CCFP produced by an expert system (*e.g.,* human forecaster) outperform the extrapolation forecast in terms of PODy, Bias, CSI, and FAR at the shorter time scales. The expert system, however, has the advantage of using numerical model output, the automated

Table 2. Characteristics of the Convective Forecasts

Product	Issue Time (UTC)	Forecast Length	Valid Period	Human / Automated	Avg % Area Covered over all forecast lengths
CCFP Final	1500 and 1900	1-, 3-,5-h and 3-, 5-,7-h	Valid at end of forecast period	Human	5.17 %
C-SIGMET	Hourly	1-, 2-, and 0-2 h	Valid at end of period and throughout 0-2 h period	Human	2.3 %
Outlook	Hourly	2-6 h and 6-h	Valid at end of period and throughout 2-6 h period	Human	14.91 %
NCWF	5 min.	1- and 2-h	Valid at end of period	Automated	0.51 %

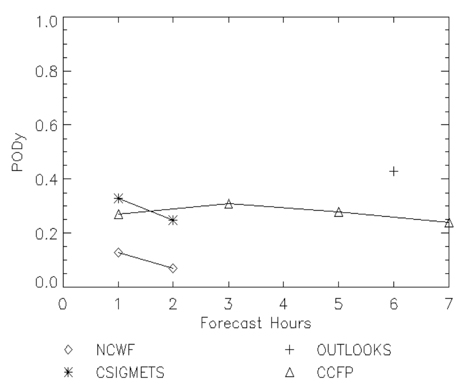

Fig. 2a. PODy for NCWF (diamond); 1 and 2 h CSIGMETs ('*'); 6 h Outlooks (+); and the CCFP (triangle).

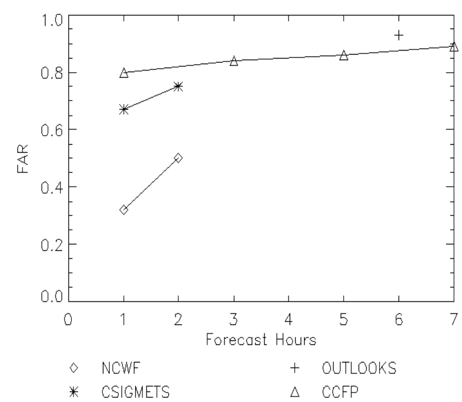

Fig. 2b Same as Fig. 1a, except for FAR.

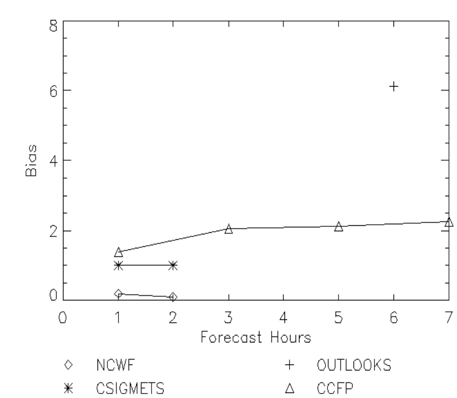

Fig. 2c. Same as Fig. 1a, except for Bias.

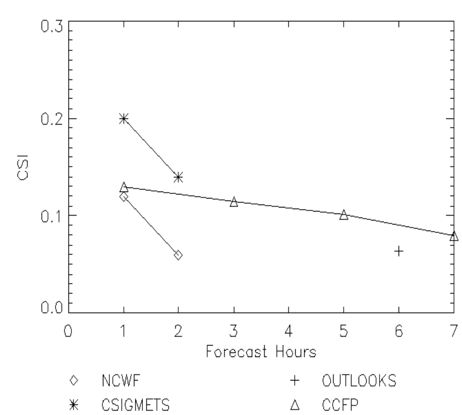

Fig. 2d. Same as Fig. 1a, except for CSI.

extrapolation forecast, and fuzzy logic techniques (Wilson *et al.* 1998) to produce its forecasts. For longer forecast lengths, the expert systems, represented by the CCFP and the Outlooks, maintain relatively consistent values of PODy and have slowly decreasing values of CSI. However, the FAR and Bias are somewhat larger than the NCWF values at the 2 h time period. These forecasts typically contain larger areas (as shown in Table1) than the 1 - 2 h forecasts. As shown earlier, their sizes are large to include convective initiation, propagation, and dissipation, and the hazards associated with the convective activity (e.g., hail, tornados, high winds).

5. FUTURE PLANS

The verification exercise will continue during the summer of 2000. The results will be available through the RTVS Web-based interface (http://www-ad.fsl.noaa.gov/afra/rtvs/RTVS-project_des.html). The verification methods used in the 2000 exercise will be modified in order to address: 1) grid size used to map the observations, 2) time window used to collect the observations, and 3) statistical boundaries used to identify "good" forecasts. Further evaluations will be undertaken to sort out scale issues from the analyses.

Acknowledgments

This research is in response to requirement and funding by the Federal Aviation Administration Aviation Weather Research Program. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

The authors would like to Nancy Rehak (NCAR) for providing the NCWD and the NCWF, Don Frank (AWC) and Clinton Wallace (AWC) for providing the CCFP and FG forecasts and valuable feedback, Judy Henderson (FSL) for her work on RTVS, the AWC forecasters who developed the CCFP and FG, and Ed Tollerud and Nita Fullerton for their helpful reviews.

References

Austin. G.L., A. Bellon, P. Dionne, and M. Roch, 1987: On the interaction between radar and satellite image nowcasting systems and mesoscale numerical models. *Proc. Mesoscale Analysis and Forecasting*, Vancover, BC, Canada, European Space Agency, 225-228.

Brown, B.G. and J.L. Mahoney, 2000: Quality Assessment of the National Convective Weather Forecast Product. (Available from Barbara Brown, P.O. Box 3000, Boulder, CO 80303).

Brown, B.G., T.L. Kane, R. Bullock, and M.K. Politovich, 1999: Evidence of improvements in the quality of in-flight icing algorithms. *Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, 10-15 Jan., AMS, 48-52.

Brown, B.G., G. Thompson, R.T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. and Forec*., 12, 890-914.

Browning, K.A., 1980: Local weather forecasting. *Proc. Roy. Soc. London, Ser. A*, **371**, 179-211.

Doswell, C.A., III, 1986: Short-range forecasting. *Mesoscale Meteorology and Forecasting*. P. Ray, Ed., Amer. Meteor. Soc., 689-719.

Mahoney, J.L., B.G. Brown, and J. Hart. 2000a: Statistitical verification for the Collaborative Convective Forecast Product. NOAA Technical Report OAR 457-FSL 6, U.S. Dept. of Commerce, 30pp.

Mahoney, J.L., B.G. Brown, and J. Hart, 2000b: 1999 Convective SIGMETs and convective SIGMET outlooks: Verification methods and statistical results. (Available from Jennifer Mahoney, R/FS5 325 Broadway, Boulder,CO 80303).

Mahoney, J.L., J.K. Henderson, and P.A. Miller 1997: A description of the Forecast Systems Laboratory's Real-Time Verification System (RTVS). *Preprints, 7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, CA, AMS, J26-J31.

Megenhardt, D., C.K. Mueller, N. Rehak, G. Cunning, 2000: Evaluation of the National Convective Forecast Product. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Amer. Meteor. Soc. (this publication)

Mueller, C.K., C.B. Fidalego, D.W. McCann, D. Meganhart, N. Rehak, and T. Carty, 1999: National Convective Weather Forecast Product. *Preprints, 8th Conference on Aviation, Range, and Aerospace Meteorology*, Amer. Meteor. Soc, 230-234.

Murphy, A.H. and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

NWS, 1991: National Weather Service Operations Manual, D-22. National Weather Service. (Available at Website http://www.nws.noaa.gov).

Orville, R.E., 1991: Lightning ground flash density in the contiguous United States-1989. *Mon. Wea. Rev*., **119**, 573-577.

Phaneuf, M. W. and D. Nestoros, 1999: Collaborative convective forecast product: Evaluation for 1999. (Available from the author at CygnaCom Solution, Inc.)

Sankey, D., K.M. Leonard, W. Fellner, D.J. Pace, and K.L. Van Sickle, 1997: Strategy and direction of the Federal Aviation Administration's Aviation Weather Research Program. *Preprints, 7th Conference on Aviation Range, and Aerospace Meteorology*, Long Beach, Amer. Meteor. Soc, 7-10.

Schaefer, J.T., 1990: The Critical Success Index as an indicator of warning skill. *Wea. and Forec*., **5**, 570-575.

Wilson, J.W., N. A. Crook, C.K. Mueller, J. Sun, and M. Dixon, 1998. Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc*., **79**, 2079-2099.