Quality Assessment Report

National Convective Weather Forecast (NCWF-2)

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

[1]Forecast Systems Laboratory, NOAA Research, Boulder, Colorado
[2]Joint collaboration with the Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado
[3]National Center for Atmospheric Research, Boulder, Colorado

29 February 2004

Contacts: Jennifer Mahoney (NOAA/FSL 325 Broadway, R/FS5, Boulder, CO 80305; Jennifer.mahoney@noaa.gov) or Barbara Brown (NCAR, P.O. Box 3000, Boulder, CO; bgb@ucar.edu).

Quality Assessment Report

National Convective Weather Forecast (NCWF-2)

Jennifer Luppens Mahoney, Partice Kucera, Joan E. Hart, and Barbara G. Brown

Quality Assessment Product Development Team

26 February 2004

SUMMARY

This report summarizes a statistical evaluation of the quality of the second generation National Convective Weather Forecast (NCWF-2) product that was developed by the Convective Weather Product Development Team under the Federal Aviation Administration's Aviation Weather Research Program. The document provides an assessment of the NCWF-2 forecasting capability, including quantitative verification of 1- and 2-hour probabilistic forecasts. The results of this evaluation will be provided to the FAA and National Weather Service (NWS) Aviation Weather Technology Transfer (AWTT) Board for its consideration of whether NCWF-2 is ready for transition to an experimental product.

The components of the NCWF-2 considered in this evaluation include the 2-hour forecast and the probabilistic capability. The NCWF-2 is evaluated using the operational National Convective Weather Detection product (NCWD-VIL) as well as a reflectivity-based version of the NCWD (referred to as NCWD-ref). For a measure of comparison, the quality of the NCWF-2 forecasts is compared to the quality of the operational 1-hour NCWF and the 2-hour Collaborative Convective Forecast Product (CCFP).

The study includes a variety of analyses. Overall statistics for the NCWF-2 were computed from 15 June – 31 August 2003. Analyses were run for several summer 2002 cases to determine the impact on the statistical results of using NCWD-ref vs. NCWD-VIL to initialize and evaluate the forecasts. In addition, a subset of 11 days from active convective cases during 2003 was chosen for the NCWF-2 and CCFP comparison.

The statistical methodology is consistent with the approach used in previous evaluations of the NCWF (e.g., Brown and Mahoney 2000).

The results indicate that:

- The NCWF-2 forecasts are provided in probabilistic terms which will allow better decision-making procedures.
- One-hour NCWF-2 forecasts are as skillful as the operational NCWF when NCWD-VIL is used to generate and verify the algorithm.
- An elliptical filter of 60 km seems to be better suited than the 30-km filter for capturing the correct amount of convective activity.
- Two-hour NCWF-2 forecasts are nearly as skillful as 2-hour CCFP forecasts.

1. INTRODUCTION

New developments by the FAA Aviation Weather Research Program's Convective Weather Product Development Team (FAA/AWRP/CW PDT) have led to the creation of a 1- and 2-hour probabilistic convective forecast which is considered as the next generation National Convective Weather Forecast (NCWF-2). An early version of the NCWF-2 was evaluated by the Quality Assessment PDT during the summer of 2003. The evaluation results presented in this report are provided to the FAA and National Weather Service (NWS) Aviation Weather Technology Transfer (AWTT) Board for use in their process of determining whether to transition the NCWF-2 from "test" to "experimental" status. Note that the AWTT Board has defined experimental products as those products that "show promise" to become useful operational products in the future.

Performance of the NCWF-2 was evaluated from 15 June – 31 August 2003 and compared to the performance of the operational NCWF. Three additional evaluation periods for a subset of cases were investigated to help address issues between observation datasets and comparisons with the CCFP.

The report is organized as follows. The approach is presented in Section 2. Section 3 briefly describes data (e.g., algorithms, forecasts, and observations) included in the evaluation. The verification methods are summarized in Section 4, and the results of the evaluation are presented in Section 5. Finally, Section 6 includes the conclusions.

2. APPROACH

During the evaluation period, a probabilistic forecast was produced by the NCWF-2 every 5 minutes. NCWF-2 forecasts issued on the hour were evaluated using the Real-Time Verification System (RTVS; Mahoney et al. 2002) and included in the verification study.

The verification approach applied in this study is identical to the approach undertaken in previous studies associated with the operational version of the NCWF (Brown and Mahoney 2000). In particular, the algorithm forecasts were verified using the operational NCWD-VIL (a convective observation product based on radar-derived VIL combined with lightning observations) and the NCWD-ref (a convective observation product based on radar reflectivity observations). The algorithm probabilities were transformed into Yes/No convective fields by determining if the probability at a grid point exceeded or was less than a pre-specified threshold; a variety of different thresholds was utilized to examine the full range of performance of the NCWF-2.

As a supplement to the 2003 comparison:

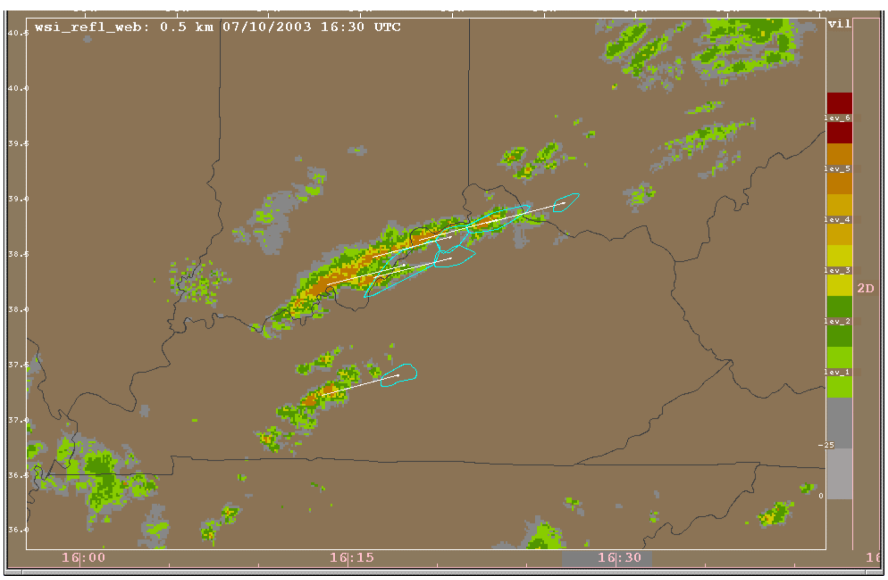
- Comparisons between the NCWF-2 and the CCFP were performed for 11 days in 2003 where the two forecasts were appropriately scaled to the match the other.
- Comparisons between the observations (NCWD-VIL vs. NCWD-ref) were performed for the 11 days in 2003 to directly determine the differences between the two observations.
- Comparisons between the NCWF-2 and the NCWF generated using the NCWD-VIL observations were performed on a subset of 8 days in 2002 to investigate the impacts on the verification results associated with (a) the observations (NCWD-VIL vs. NCWD-ref) and (b) the size of the elliptical filter used to generate the NCWF-2.

In evaluating the NCWF-2, it is important to compare the quality of the forecast to the quality of one or more standards of reference. Therefore, the quality of the NCWF-2 is compared to the quality of the operational NCWF and the 2-hour CCFP. However, it is important to emphasize that the NCWF-2 and CCFP are very different types of forecasts, with different objectives. The comparisons are designed to be as fair as possible to both types of forecasts, as described in Section 4, while still obtaining the information needed. Nevertheless, users of these statistics should keep these assumptions in mind when evaluating the strengths and weaknesses of each type of forecast.

3. DATA

3.1 National Convective Weather Forecast-2 (NCWF-2)

The operational version of the NCWF (i.e., "NCWF") is an automated system that predicts the location of convective storm areas (Mueller et al. 1999). Every five minutes, the NCWF generates a 1-hour convective extrapolation forecast that is valid at a specific time (e.g., Fig. 1a). The operational NCWF provides a binary forecast of storm location with a one hour lead time (Megenhardt et al. 2000). In contrast, the new NCWF-2 provides a probabilistic forecast (i.e., the probability that convection will occur at a specific grid point at a specific time) at 1 and 2 hours in the future (e.g., Fig. 1b); the NCWF-2 forecasts are also updated every 5 minutes. The version of NCWF-2 that was run during summer 2003 at NCAR (e.g., Fig. 1b) was based primarily on 2-km reflectivity observations and the RUC forecast fields.



a)

b)

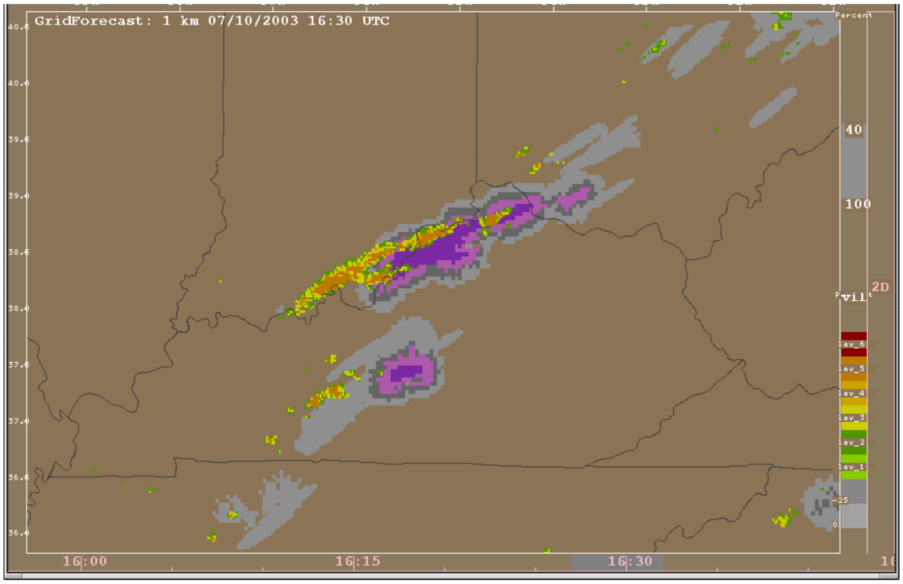


Figure 1. Examples of the 1-hour forecast produced by NCWF (Fig. 1a) and NCWF-2 (Fig. 1b). In Fig. 1a, NCWD-ref is the background field and the cyan polygons indicate the one hour NCWF forecast position. Figure 1b shows the NCWF-2 probabilities where the NCWD-ref values greater than 40 dBz are overlaid.

The probabilistic forecasts produced by NCWF-2 are based on work by Germann and Zawadski (2004). The probabilities are calculated by determining the area coverage of convection within an elliptical filter region. The elliptical filter is rotated at 10-degree intervals to determine the orientation with the maximum area coverage. The maximum area coverage calculated within the elliptical filter is mapped as the probability level (purple shades in the Fig. 2). An elliptical filter size of 30 km was used as the basis for generating the probabilities in the 2003 version of NCWF-2 (Fig. 2a). Based on investigations by the NCWF-2 developers, the elliptical filter size will increase to 60 km for summer 2004, which will slightly increase the area of convective activity identified (Fig. 2b). Once the coverage values are mapped, the probabilities are then advected using storm motion vectors. In addition to extrapolating the convection, the NCWF-2 also attempts to capture regions of growth using output from the Rapid Update Cycle numerical weather prediction model (RUC; Benjamin et al. 1998) along with radar trending and diurnal considerations. Mueller et al. (2004) describes NCWF-2 in detail.

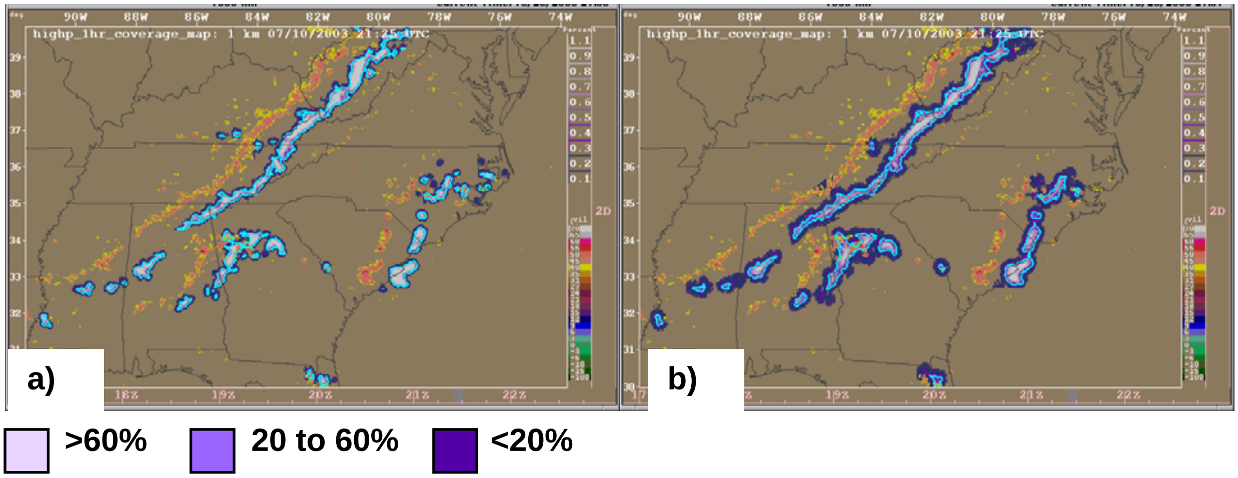


Figure 2. Probability fields based on an elliptical filter with a scale of (a) 30 km and (b) 60 km. The 40% contour is shown as the cyan overlay. The yellow and red overlay is the NCWD-ref.

3.2 Collaborative Convective Forecast Product (CCFP)

The CCFP is prepared through a multiple-set collaborative process (Weather Applications Workgroup 2003; Hudson and Foss 2002; Phaneuf and Nestoros 1999) that begins with AWC forecasters, but includes participation from airline meteorologists and dispatchers, as well as meteorologists from the Center Weather Service Units (CWSUs) at the Air Route Traffic Control Centers (ARTCCs). The CCFP is used as a strategic decision aid by the decision-makers at the airlines and the Air Traffic Control System Command Center (ATCSCC) for rerouting air traffic around convective weather. The issue and lead times produced during 2003 for the CCFP are summarized in Table 1. The times listed in the table occurred after the change to daylight savings. Prior to daylight savings, the CCFP is issued 1 hour earlier.

3.3 Observations

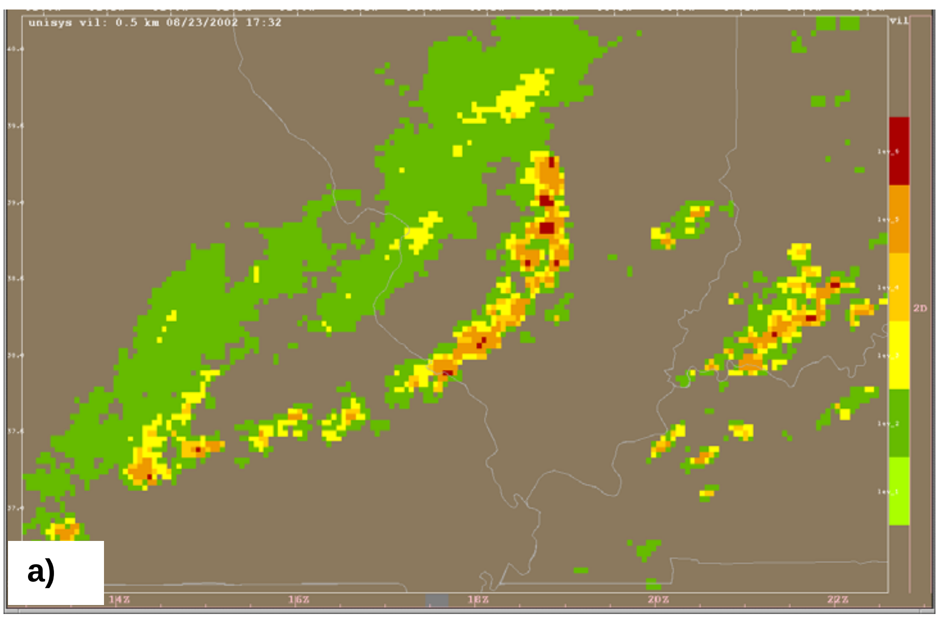
Two observation datasets are used to evaluate the quality of the NCWF-2: (a) the operational NCWD-VIL which combines lightning and VIL-based radar data to build a convective field and (b) NCWD-ref which is strictly based on reflectivity observations.

The NCWD-VIL is a convective hazard field depicting areas of convective weather that may be hazardous to aviation. The hazard field is based on WSR-88D National Radar Mosaics and National Lightning Detection Network cloud-to-ground lightning data (Orville 1991). The echo tops data are used to threshold the radar-derived VIL observations. The VIL data are provided in the NIDS WSR-88D product stream and are mapped to a national mosaic by UNISYS. The VIL field is calculated by using an empirical formula to derive liquid water content from radar reflectivity at each elevation. The data are then integrated with height to obtain VIL. The VIL observations provide information about the intensity of a storm throughout its vertical extent, and are a proxy for vertical development. VIL values are translated to a VIP scale. The VIP values 3 and greater are used to evaluate the NCWF-2.

In addition to the NCWD-VIL, 2-km reflectivity observations, which were mapped to a 4-km grid for this evaluation (i.e., NCWD-ref), are also used to evaluate the NCWF-2. The reflectivity observations are provided to NCAR through the WSI WSR-88D product stream. Reflectivity values of 40 dBZ and greater are used to evaluate the NCWF-2. Lightning data were not considered in this version of the NCWD-ref.

Because VIL data were not available at NCAR, the version of NCWF-2 that was run during summer 2003 was based on NCWD-ref. However, the operational NCWF and the version of NCWF-2 that will be implemented at AWC will be based on NCWD-VIL.

The fundamental differences between the two observational datasets proved to have a large impact on the statistical results. In particular, NCWF-2 skill varied depending on whether the forecasts were evaluated against the NCWD-VIL or the NCWD-ref. Therefore, it is critical that the differences in the two datasets be noted. For example, Fig. 3a shows the NCWD-VIL and Fig. 3b shows the NCWD-ref for a particular time period. The convective areas identified by the NCWD-VIL data often appear as smoother more intense regions (as can be seen by the red values in Fig. 3a as compared to lighter green areas in Fig. 3b). In addition, smaller isolated areas of convective activity are often eliminated from the NCWD-VIL convective field when compared to the NCWD-ref as illustrated in the figures. Differences in the observations are also considered in Section 5.2.



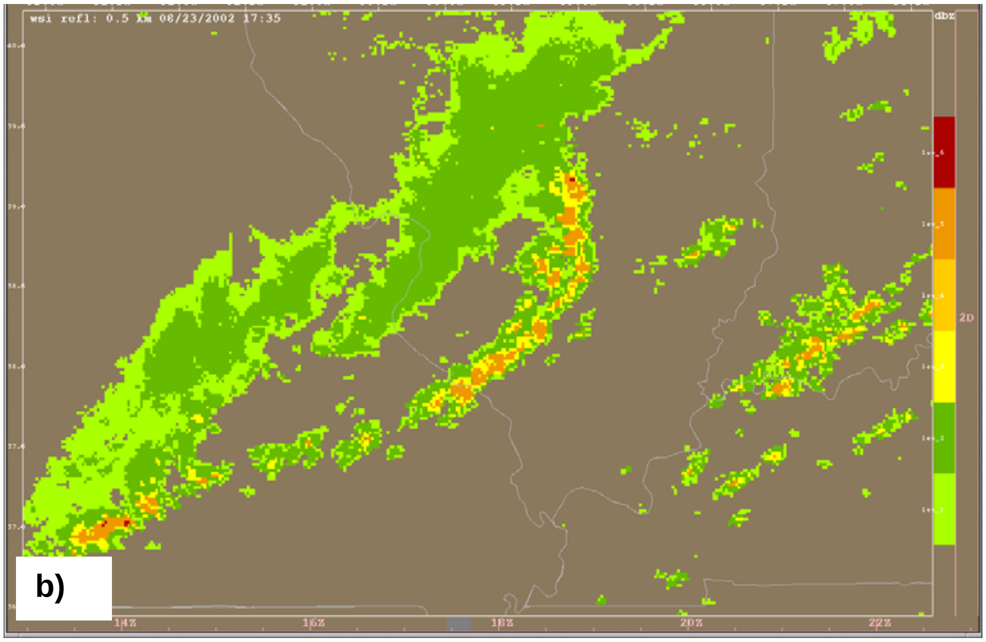


Figure 3. Example radar data sets: a) NWS 4-km VIL used to create the NCWD-VIL and b) WSR-88D reflectivity used to create the NCWD-ref.

4. METHODS

This section summarizes the methods used to match the forecasts and observations, and it describes the various statistical measures that are used to evaluate the forecasts.

4.1 Spatial Matching of Forecasts to Observations

The matching procedures described in this Section are the same as have been used in previous evaluations of NCWF (e.g., Brown and Mahoney 2000).

When comparing the forecasts and the observations, the first procedure includes moving the forecasts and observations to a common resolution or scale. For evaluation of the NCWF-2 with the NCWD-VIL or NCWD-ref, once the 2-km NCWD-ref observations were interpolated to a 4-km grid, no further scaling was necessary. However, when comparing the NCWF-2 with the CCFP forecast, the NCWF-2 and the NCWD-VIL are up-scaled to a 40-km grid to match the resolution of the CCFP. Conversely, the CCFP is down-scaled to a 4-km grid to match the resolution of the NCWF-2.

The scaling technique is as follows: A grid with a specified resolution (4 km or 40 km) is overlaid on the observation field. Each box on the overlay grid is assigned a *Yes* or *No* value depending on whether a positive observation fell within the grid box. For each grid box, the criteria for positive observations defined in Section 3 (e.g., a VIL value of 3.5 kg/m^3 or greater or a lightning rate of at least 3 to 5 strokes in 10 minutes) is used to determine if the grid box would be assigned a *Yes* or *No* observation; specifically, if at least one 4-km NCWD box in the grid area met the criteria, the entire box is assigned a *Yes* observation; otherwise it is assigned a *No* observation. The same procedure is applied to forecasts that are based on polygons (e.g., NCWF, CCFP), with a grid box labeled with a *Yes* forecast if any part of the forecast polygon intersected that box. If a forecast polygon does not intersect the grid box, then a *No* forecast is assigned to that box. Forecasts based on a grid (e.g., NCWF-2) are treated in the same manner as the gridded observations.

The NCWF-2 was verified with the NCWD-VIL or the NCWD-ref file nearest to the forecast valid time. However, the CCFP was verified using a ± 10 minute window surrounding the valid time. Therefore, in some cases, up to four additional 5-minute NCWD files were used to evaluate the CCFP.

4.2 Statistical Verification Methods

By overlapping the Yes/No forecast and observation grids for a particular forecast, each grid point was assigned a forecast/observation designation (i.e., YY, YN, NY, NN). The *Yes/No* forecast/observation pairs were then counted, to fill in a 2x2 contingency table like the one shown in Table 1. That is, for a given forecast, all of the grid boxes with a *Yes* forecast and a *Yes* observation were counted to obtain the YY count; all of the grid boxes with a *Yes* forecast and a *No* observation were counted to obtain the YN count; and so on. Individual forecast contingency tables were accumulated to obtain tables representing particular days, months, or other periods (including the entire forecast period). Overall results are presented in this report. For further information visit the RTVS web site at www-ad.fsl.noaa.gov/fvb/rtvs/; link convection.

Table 1. Basic contingency table for evaluation of dichotomous (e.g., Yes/No) forecasts. Elements in the cells are the counts of forecast-observation pairs.

Forecast	***Observation***		***Total***
	Yes	***No***	
Yes	YY	YN	YNYY+YN
No	NY	NN	NNNY+NN
Total	YY+NY	YN+NN	YN+NNYY+YN+NY+NN

Table 2 lists the verification statistics that were included in the evaluation, with PODy, PODn, and FAR representing the basic verification statistics. General descriptions of these statistics include the following:

- PODy and PODn are estimates of the proportions of *Yes* and *No* observations, respectively, that were correctly forecast (e.g., Brown et al. 1997).
- FAR is the proportion of *Yes* forecasts that were incorrect.
- Bias is the ratio of the number of *Yes* forecasts to the number of *Yes* observations, and is a measure of over- or under-forecasting.
- The Critical Success Index (CSI), also known as the Threat Score, is the proportion of correct *Yes* forecasts that were either forecast or observed.
- The True Skill Statistic (TSS) (e.g., Doswell et al. 1990) is a measure of the ability of the forecast to discriminate between *Yes* and *No* observations; TSS also is known as the Hanssen-Kuipers discrimination statistic (Wilks 1995).
- The Heidke Skill Score (HSS) is the percent correct, corrected for the number expected to be correct by chance.

- The Gilbert Skill Score (GSS) (Schaefer 1990), also known as the Equitable Threat Score, is the CSI corrected for the number of correct *Yes* forecasts expected by chance.
- The % Area is the percent of the total possible area that had a *Yes* forecast (Brown et al. 1997).

Table 2. Verification statistics used in this study.

Statistic	***Definition***	***Description***
PODy	YY/(YY+NY)	Probability of Detection of "Yes" observations
PODn	NN/(YN+NN)	Probability of Detection of "No" observations
FAR	YN/(YY+YN)	False Alarm Ratio
CSI	YY/(YY+NY+YN)	Critical Success Index
Bias	(YY+YN)/(YY+NY)	Forecast Bias
TSS	PODy + PODn – 1	True Skill Statistic
HSS	[(YY+NN)-C1]/(N-C1), where N=YY+YN+NY+NN C1=[(YY+YN)(YY+NY) + (NY+NN)(YN+NN)] / N	Heidke Skill Score
GSS	(YY-C2)/[(YY-C2)+YN+NY], where C2=(YY+YN)(YY+NY)/N	Gilbert Skill Score
% Area	(Forecast Area) / (Total Area) x 100	% of the area of the continental U.S. where convection is forecast to occur

5. RESULTS

The results of the NCWF-2 evaluations are described in this section. A variety of time periods were included in the analysis:

- Overall statistics for the period 15 June – 31 August 2003 where the 30-km elliptical filter was applied to NCWF-2.
- Comparisons of NCWF-2 and the CCFP performed for 6 active convective cases (11 days) in 2003.
- Comparisons between the observations (NCWD-VIL vs. NCWD-ref) performed for the 11 days in 2003.
- Case study comparisons between the NCWF-2 regenerated using the NCWD-VIL observations and 60-km filter and the NCWF performed on a subset of 4 cases (8 days) in 2002.

It is important to note that during the summer 2003 period, the version of NCWF-2 that was evaluated was derived using the 30-km elliptical filter and the NCWD-ref. On the other hand, the operational NCWF was derived from the NCWD-VIL dataset. This inconsistency in the fundamental make-up of the two algorithms led to a misinterpretation of the NCWF-2 skill as compared to the NCWF. Although the results are presented for the NCWF-2 and the NCWF with this inconsistency, additional analyses are provided that allow a direct comparison between the NCWF and the NCWF-2 derived using NCWD-VIL.

5.1 Overall Statistics

Overall verification results for the 1-hour NCWF-2 forecasts for 15 June – 31 August 2003 are shown in Fig. 4. The statistics (i.e., PODy, CSI, and FAR) are shown as a function of Bias, where the curves represent the quality of the NCWF-2 at a variety of thresholds in 0.1 increments ranging from 0.0 at the top of the curve to 1.0 at the bottom of the curve. The single points represent the statistics for the 1-hour NCWF forecast. The results presented in Fig. 4 are convoluted by the observational inconsistency between the NCWF-2 and the NCWF.

Figure 4 indicates that, overall, the NCWF-2 is skillful at capturing convective activity at 1 and 2 hours. Positive skill is noted in all scores. To determine the relative skill of NCWF-2, the quality of NCWF-2 is compared to the 1-hour NCWF and the 2-hour CCFP.

Figure 4 suggests that the NCWF is slightly more skillful than the NCWF-2 at a threshold of 0.3 since the PODy and CSI values for NCWF are larger and the FAR value is smaller than the corresponding values for the NCWF-2. However, after closely investigating these differences, it became apparent that the observation field (i.e., NCWD-VIL vs. NCWD-ref) had an important impact on the relative skill of the NCWF-2. Additional analyses of the NCWF and NCWF-2 based entirely on VIL-based radar observations are summarized in Section 5.3.

Figures 5a and b show verification statistics for the 2-hour NCWF-2 and the CCFP. Figure 5a shows the verification measures for the 2-hour NCWF-2 up-scaled to 40 km as a function of Bias, in comparison to the verification measures for the 40-km CCFP, for 6 cases (based on 11 days) in 2003. The Bias of the CCFP remains close to 1.0 while the Bias for the NCWF-2 ranges from 0.27 (for a threshold of 0.4) to 1.3 (for a threshold of 0.2). The CSI for NCWF-2 is larger for all thresholds than the CSI for the CCFP. However, the PODy and FAR values for the two forecasts are nearly identical.

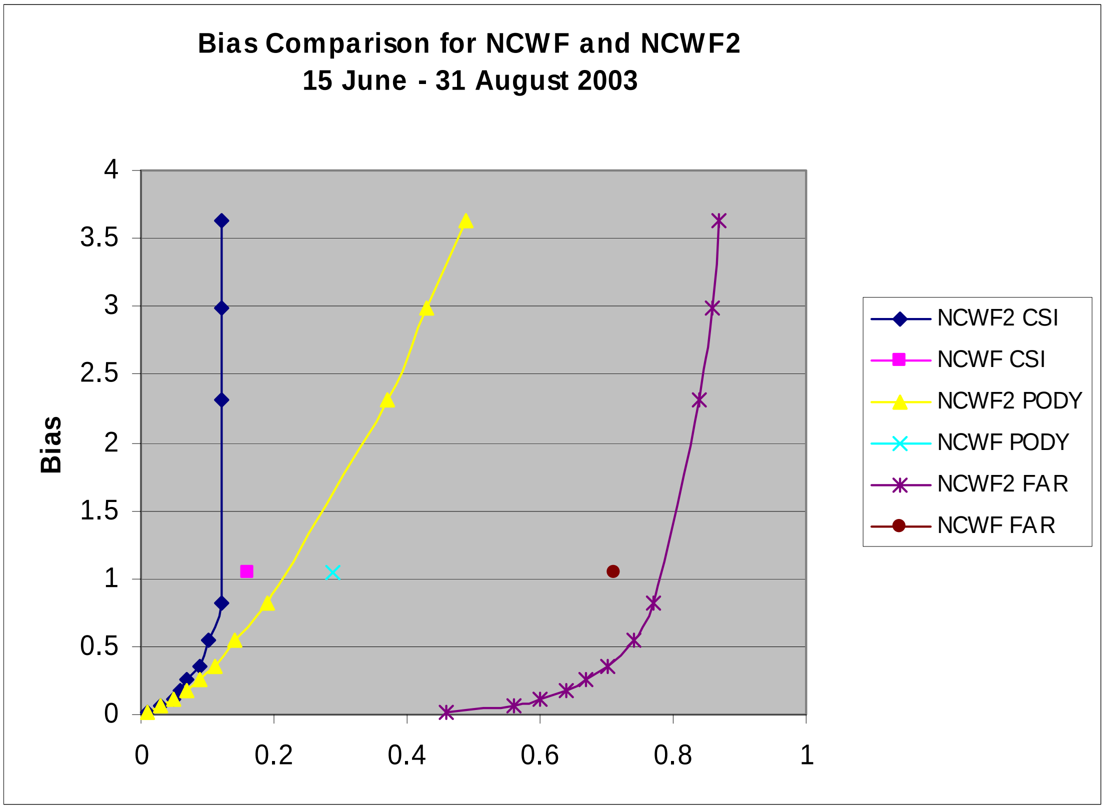
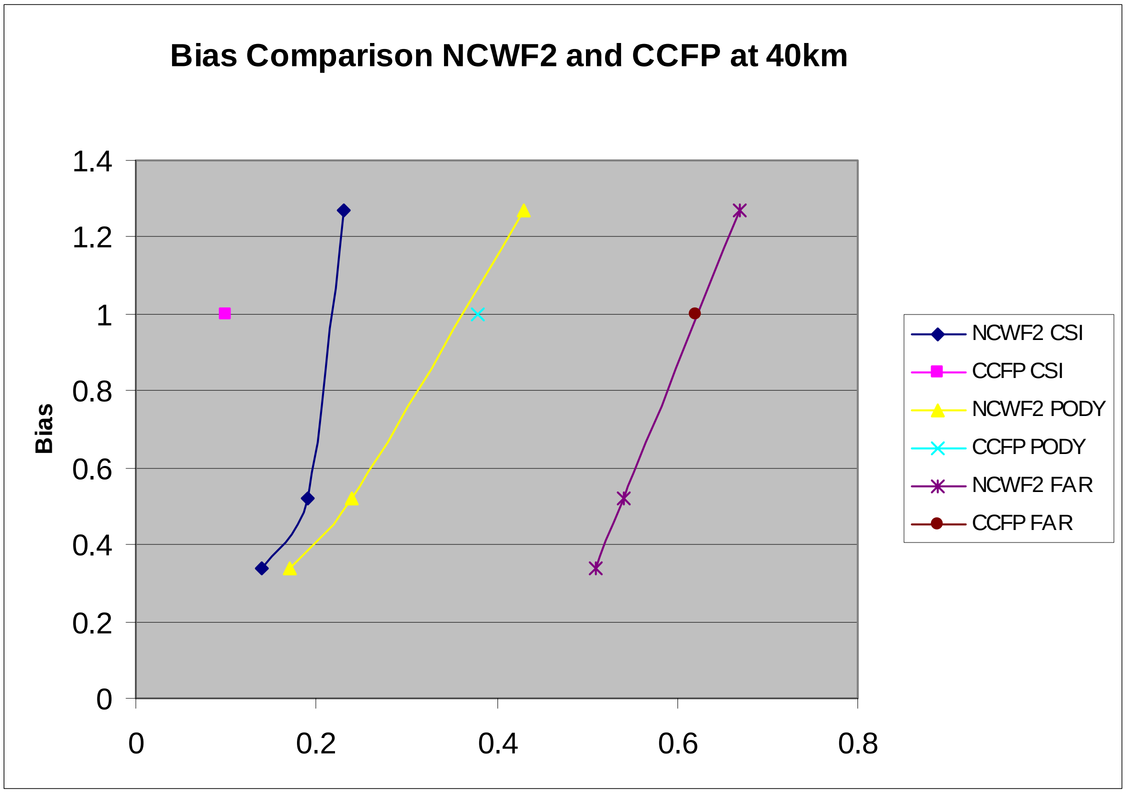


Figure 4. Statistical results from 15 June – 31 August 2003 as a function of Bias for NCWF-2 (CSI; diamond, PODy; triangle, FAR; '*') and NCWF (CSI; square, PODy; 'x', FAR; closed circle). Thresholds for NCWF-2 are 0.0 (top of plot) to 1.0 (bottom of plot) in 0.1 increments.

Figure 5b shows the forecast verification statistics as a function of Bias for the 4-km NCWF-2 compared to the verification statistics for the CCFP down-scaled to 4-km, for the same 6 cases in 2003. The Bias for CCFP increases to nearly 5.0 on the 4-km scale (Fig. 5b), suggesting more overforecasting of the convective activity. However, the other verification statistics for the NCWF-2 are not quite as good as the verification statistics for the CCFP on a 4-km scale, with the CSI and PODy values for NCWF-2 (with a threshold of 0.1) smaller than the corresponding values for CCFP, and larger values of FAR for NCWF-2 as compared to CCFP.



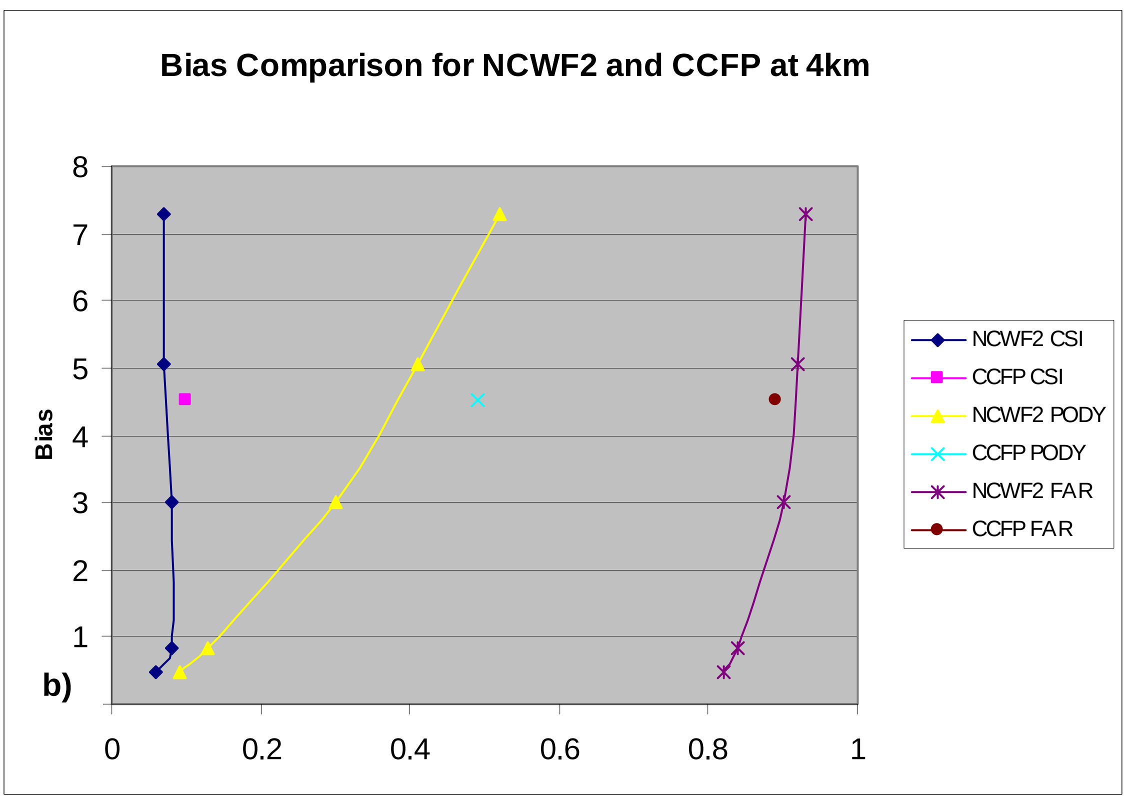


Figure 5. Statistical results from 4 cases in 2003 for 2-hour NCWF-2 (CSI; diamond, PODy; triangle, FAR; '*') and CCFP (CSI; square, PODy; 'x', FAR; circle). Thresholds for NCWF-2 in Fig. 5a are 0.2 (top), 0.3, and 0.4 (bottom); for Fig. 5b 0.0 (top), 0.1, 0.2, 0.3, 0.4 (bottom).

5.2 Summer 2003 results based on NCWD-VIL vs. NCWD-ref

During the summer 2003 verification period, verification statistics were generated for the NCWF-2 using both the NCWD-VIL and the NCWD-ref. Comparisons of the results based on these two observation sets are summarized in Table 3. These results are examined in more detail by selecting appropriate thresholds for NCWF-2 and comparing individual statistics. Table 3 shows a variety of statistics associated with three thresholds: 0.2, 0.3, and 0.4.

Overall, the performance of NCWF-2 verified by the NCWD-VIL is slightly better than the NCWF-2 performance verified by the NCWD-ref. All scores for 1-hour forecasts are slightly better for NCWF-2/NCWD-VIL than for NCWF-2/NCWD-ref. The Bias is slightly smaller for the comparison to NCWD-VIL than for the comparison to NCWD-ref. These results suggest that the smaller isolated convection often present in the reflectivity data impacts the skill of the forecasts. At 2 hours, the differences in skill are quite small, although a very slight, general reduction in skill for the NCWD-ref comparison is evident.

Table 3. Overall statistics for 1- and 2-hour NCWF-2 forecasts verified using NCWD-VIL and NCWD-ref observations, for 15 June – 30 August 2003.

Product	Fcst Length (h)	PODy	PODn	FAR	CSI	TSS	HSS	GSS	Bias	% Area
					NCWD-VIL					
NCWF2 (0.2)	1	0.37	0.99	0.84	0.12	0.36	0.21	0.12	2.32	0.87
NCWF2 (0.3)	1	0.19	1.00	0.77	0.12	0.19	0.21	0.12	0.82	0.31
NCWF2 (0.4)	1	0.14	1.00	0.74	0.10	0.14	0.18	0.10	0.54	0.20
NCWF2 (0.2)	2	0.23	0.99	0.91	0.07	0.22	0.12	0.06	2.60	0.96
NCWF2 (0.3)	2	0.10	0.99	0.87	0.06	0.10	0.11	0.06	0.73	0.27
NCWF2 (0.4)	2	0.06	0.99	0.84	0.05	0.06	0.09	0.05	0.40	0.15
					NCWD-ref					
NCWF2 (0.2)	1	0.35	0.99	0.86	0.11	0.34	0.19	0.11	2.52	0.88
NCWF2 (0.3)	1	0.18	1.00	0.80	0.10	0.18	0.18	0.10	0.89	0.31
NCWF2 (0.4)	1	0.13	1.00	0.78	0.09	0.13	0.16	0.09	0.58	0.20
NCWF2 (0.2)	2	0.22	0.99	0.92	0.06	0.21	0.11	0.06	2.73	0.96
NCWF2 (0.3)	2	0.09	0.99	0.88	0.06	0.09	0.10	0.05	0.80	0.27
NCWF2 (0.4)	2	0.06	1.00	0.86	0.04	0.06	0.08	0.04	0.42	0.15

The differences in the verification statistics for NCWF-2 shown in Table 3 are simply a result of differences in the observation fields. To investigate these differences further, a grid-by-grid comparison of NCWD-VIL and NCWD-ref was undertaken using data for the subset of 11 days (6 cases) in summer 2003. This comparison involved a simple overlay of the Yes/No fields representing the two types of observations and creating contingency tables like Table 1 for each forecast. Statistics were computed for each forecast using the definitions in Table 2, with NCWD-VIL treated as the "forecast" and NCWD-ref treated as the "observation." The results for two representative statistics (Bias and HSS) for all the forecasts on the 11 days are shown in Fig. 6. These diagrams show histograms of the two measures for all of the days and hours.

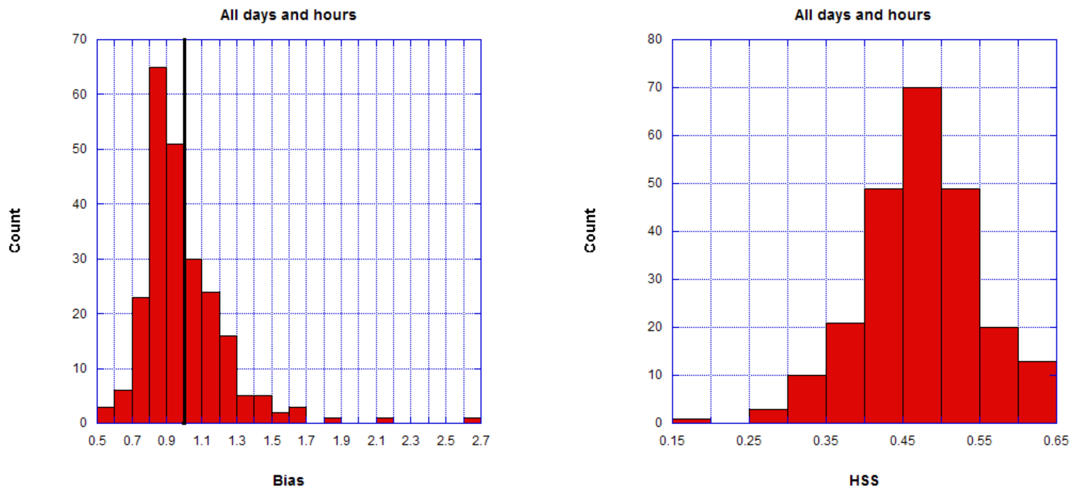


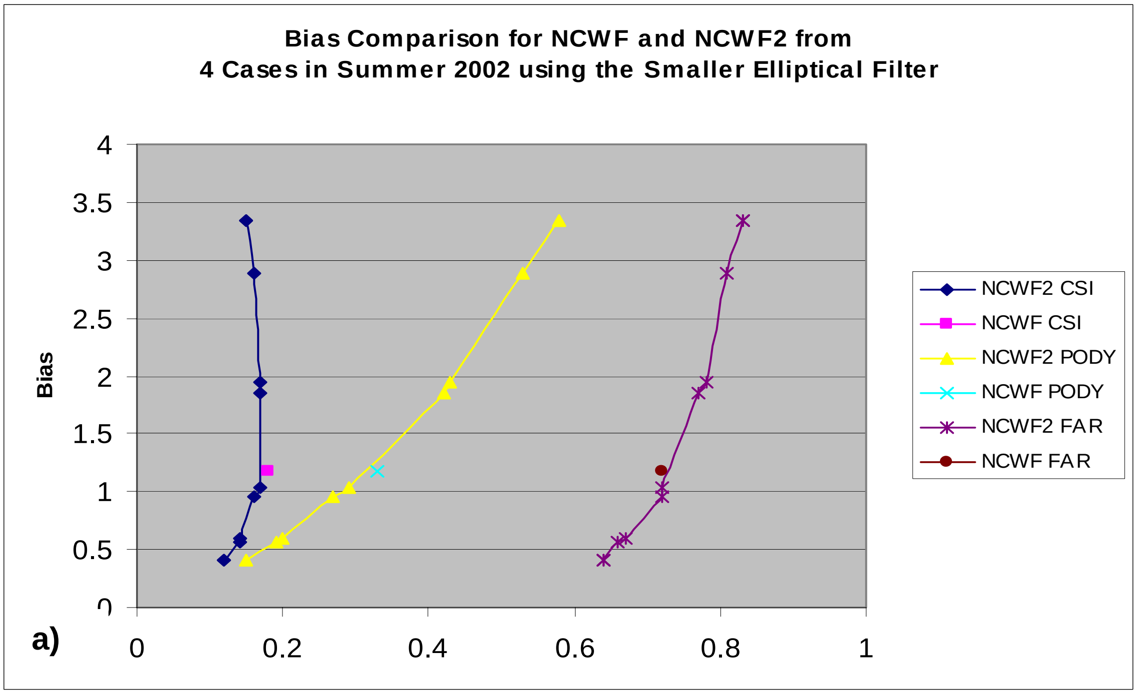
Figure 6. Histograms showing distributions of statistics comparing NCWD-VIL to NCWD-ref, for 11 days in summer 2003: (a) Bias; and (b) HSS. NCWD-VIL was treated as the “forecast” to compute the statistics; NCWD-ref was treated as the “observation.”

The results in Fig. 6 indicate that the two observation fields frequently are very different from each other, with Bias values ranging from 0.5 to 2.7. The mode (most common value) of the Bias is somewhat less than one, although very large values of Bias also occur on occasion. The overall Bias value, for all forecast hours combined, is 0.90, indicating that NCWD-VIL, overall, covers less area than NCWD-ref. This result supports the idea that NCWD-ref often shows small areas of convection that don’t appear in the NCWD-VIL field. The mode of the HSS values is around 0.5, which indicates that the two fields often disagree. Although not shown here, the Bias values vary by time of day, with Bias values greater than one common in the morning hours and Bias values less than one more frequent during afternoon and evening hours. These differences in the observation fields suggest that it is not surprising that verification statistics based on these two types of observations also are not in agreement.

5.3 Summer 2002 Comparisons between the NCWD-VIL derived NCWF and NCWF-2

The statistical results presented in the previous sections indicated a strong sensitivity to the type of observation used to generate and verify the NCWF-2. To evaluate the impact of the observations on the forecast quality, the NCWF-2 was re-computed for 4 cases (8 days) in 2002 when NCWD-VIL data were available. The results are compared to the results generated for NCWF for the same 4 cases in 2002. In addition, because the NCWF-2 developers are planning to change the elliptical filter size in 2004, results were examined for forecasts based on the old (30-km) and new (60-km) filter sizes.

Figure 7 shows the forecast verification statistics for the NCWF-2 as a function of Bias for 4 cases (8 days) in the summer of 2002 with forecasts based on the 30-km and 60-km elliptical filters (Figs. 7a and 7b, respectively).



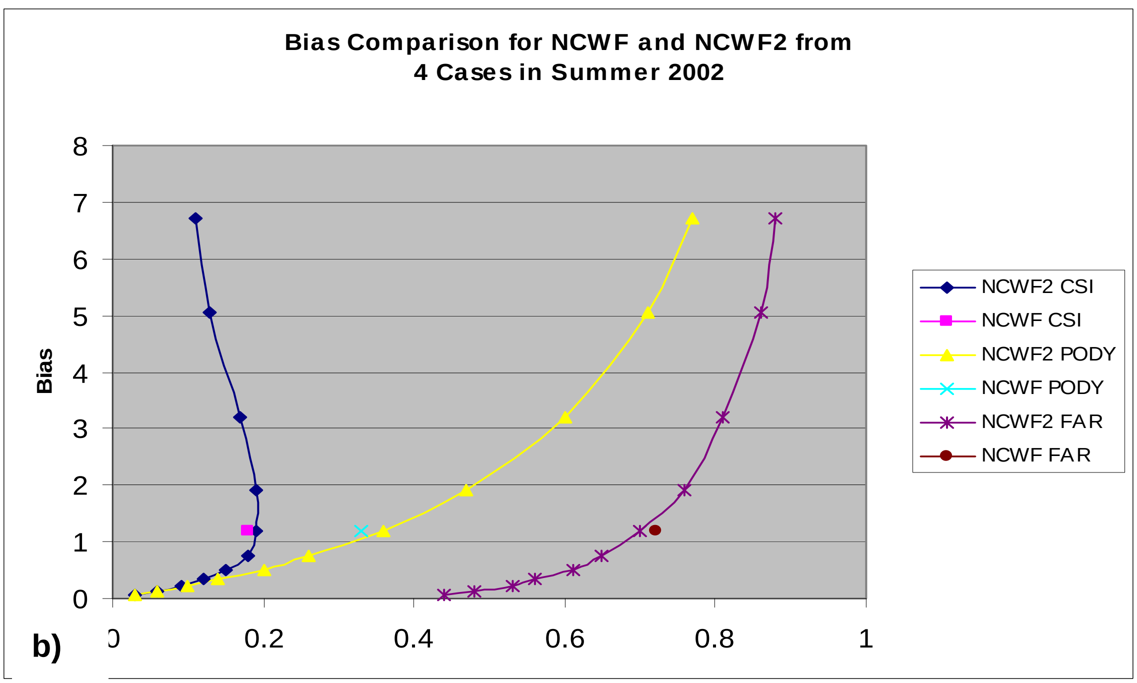


Figure 7. Same as Fig. 4, except NCWF-2 was regenerated using NCWF-VIL observations for 8 days in 2002; Fig. 7a; small filter (30-km), Fig. 7b; large filter (60-km).

The results presented in Fig. 7a suggest that the 1-hour NCWF-2 with the 30-km elliptical filter is nearly as skillful as the operational NCWF. The CSI and PODy values for NCWF-2 are slightly smaller than the values for NCWF, and the FAR values are slightly larger. However, when the elliptical filter size is increased to 60-km (Fig. 7b), the skill of the NCWF-2 increases over that of the NCWF. In particular, the CSI and PODy values for the NCWF-2 are slightly larger than the values for NCWF. The FAR is somewhat smaller for NCWF-2 as compared to NCWF. Although these results are based on a small subset of cases, they suggest that the skill of the NCWF-2 is comparable to the skill of the operational NCWF when a larger elliptical filter is used and the algorithm is derived and verified using the NCWD-VIL dataset.

6. CONCLUSIONS

This report has summarized the evaluation of the NCWF-2, with a particular focus on the new probabilistic capability and the 2-hour forecast. The evaluation has followed techniques used in previous evaluations of NCWF and other convective forecasts. The results described in this report suggest that NCWF-2 is a skillful convective product. In particular:

- The NCWF-2 forecasts are provided in probabilistic terms allowing for better decision-making procedures.
- One-hour NCWF-2 forecasts are as skillful as the operational NCWF when NCWD-VIL is used to derive and verify the algorithm.
- An elliptical filter of 60-km seems to be better suited for capturing the correct amount of convective activity.
- Currently, 2-hour NCWF-2 forecasts are nearly as skillful as 2-hour forecasts produced by the CCFP.

Many of the results presented in this report were performed on a subset of cases. Additional analyses of the NCWF-2 algorithm will be performed during the summer of 2004, allowing further investigation into the role of the NCWD-VIL and the elliptical filter. In addition, it will be important to consider the reliability of the probabilistic forecasts.

ACKNOWLEDGEMENTS

This research is in response to requirements and funding provided by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

We would like to thank the members of the Convective Weather Product Development Team for their support of the independent verification effort over the last several years. We also would like to thank the verification groups at FSL and NCAR for providing outstanding support and many contributions that made this study and report possible. Finally, we would like to thank the AWRP Leadership Team for their support of the independent verification activities undertaken by the Quality Assessment Product Development Team.

REFERENCES

Benjamin, S.G, J.M. Brown, K.J. Brundage, B.E. Schwartz, T.G Smirnova, T.L. Smith, 1998: The operational RUC-2. *Preprints, 16th Conference on Weather Analysis and Forecasting*, Amer. Meteor. Soc., 249-252.

Brown, B.G. and J.L. Mahoney, 2000: Quality assessment of the National Convective Forecast Product. Quality Assessment Report provided to the Technical Review Panel of the Aviation Weather Technology Transfer Board.

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. *Weather and Forecasting*, **12**, 890-914.

Doswell, C.A., R. Davies-Jones, and D.L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, **5**, 576-585.

Germann and Zawadzki, 2004: Scale dependence of the predictability of precipitation from continental radar images. Part II: Probability Forecasts. *Journal of Applied Meteorology*, **43**, 74-89.

Hudson, H. R. and F. P. Foss, 2002: The Collaborative Convective Forecast Product from the Aviation Weather Center's Perspective. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Oregon, WA, Amer. Meteor. Soc., 73-76.

Mahoney, J.L, J. K. Henderson, B.G. Brown, J.E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002: The real-time verification system (RTVS) and its application to aviation weather forecasts. *Preprints,* 10th *Conference on Aviation, Range, and Aerospace Meteorology*, OR, Amer. Meteor. Soc. (Boston), 323-326.

Megenhardt, Dan, C.K. Mueller, N. Rehak, and G. Cunning, 2000: *Preprints 9th Conference on Aviation, Range, and Aerospace Meteorology*, Amer. Met. Society, Orlando, Fl.

Mueller et al 2004. NCWF-2. Report to the AWTT Board.

Orville, R.E., 1991: Lightning ground flash density in the contiguous United States – 1989. *Mon. Wea. Rev*., **119**, 413-429.

Phaneuf, M. W. and D. Nestoros, 1999: Collaborative convective forecast product: Evaluation for 1999 (available from the author at AvMet Applications, Inc.).

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

Weather Applications Workgroup, 2003: Statement of User Needs CCFP/2003. FAA, CDM, CR-Workgroups. February 2003. 26 pp.

Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, San

Diego, 467 pp.