

9.4 NEW VERIFICATION APPROACHES FOR CONVECTIVE WEATHER FORECASTS

Barbara G. Brown*, Randy R. Bullock, Christopher A. Davis, John Halley Gotway,
Michael B. Chapman, Agnes Takacs, Eric Gilleland, Kevin Manning
National Center for Atmospheric Research, Boulder, Colorado

Jennifer L. Mahoney

NOAA Research-Forecast Systems Laboratory, Boulder, Colorado

1. INTRODUCTION

In recent years, the need for alternative verification approaches to evaluate convective forecasts has become apparent. In particular, the aviation weather community has a strong need for verification measures that are more clearly tied to the operational usefulness of the forecasts and are able to diagnose the sources of the errors in the forecasts (e.g., Brown et al. 2002). The need for these alternative approaches also extends to forecasts of more standard weather elements (e.g., winds, precipitation) produced by mesoscale numerical weather prediction (NWP) models, where standard verification approaches penalize the models' abilities to capture finer scale features. These penalty effects have been documented, for example, by Baldwin (2003), Rife et al. (2004), and others (e.g., Mass et al. 2002). Developing verification methods that satisfy the needs for operationally relevant and diagnostic approaches will be beneficial for managers who need to understand the benefits and usefulness of particular types of forecasts; for operational users of the forecasts who need to understand how to make use of the forecasts; and for forecast developers who need to know how their forecasts can/should be improved.

In response to these needs, several efforts are in progress to develop verification approaches that are more diagnostic. Development of these approaches will eventually lead to verification measures that are operationally meaningful in the context of specific forecast users, such as air traffic or water managers. Some developments in this area are described in this paper, along with examples of applications to several types of forecasts that have relevance for aviation forecast developers and decision makers. The focus here is on quantitative precipitation forecasts (QPFs) and forecasts of convection.

The methodology considered here is based on an "object-oriented" approach to verification. With the object-oriented approach under development, forecast and observed precipitation/convection areas are reduced to regions of interest that can be compared to one another in a meaningful way. In the case of most human-generated forecasts, the forecast objects are pre-defined by the forecaster(s). In contrast, gridded forecasts, which are the norm for forecasts produced by NWP models, must be converted to objects.

Following a short discussion of the verification problem in Section 2, a general description of the verification approach is presented in Section 3. Several examples of applications of the approach are presented in Section 4, and future work on this topic is considered in Section 5. Some conclusions are presented in the final section.

2. MOTIVATION

Standard approaches for verification of spatial QPFs and convective forecasts are based on simple grid overlays in which the forecast grid is matched to an observation grid or set of observation points. For Yes/No forecasts of convective weather, the forecast/observation (Yes/No) pairs are counted, to complete the standard 2x2 verification contingency table. The counts in this table can be used to compute a variety of verification measures and skill scores, such as the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Equitable Threat Score (ETS) (e.g., Doswell et al. 1990; Wilks 1995). For continuous variables (e.g., precipitation amount), the forecasts and observations are used to estimate the mean squared error (MSE), mean absolute error (MAE), and other standard verification measures for continuous variables (e.g., Wilks 1995; Jolliffe and Stephenson 2003).

An important concern is that this approach does not provide information needed to diagnose particular problems with the forecasts, to reveal the steps that can be taken to improve the forecasts or to give meaningful guidance to forecast users.

*Corresponding author address: Barbara G. Brown,
NCAR, PO Box 3000, Boulder CO 80307-3000; e-mail:
bgb@ucar.edu

Uncertainty and scaling issues associated with the observations also are of importance (e.g., Casati et al. 2004), but are beyond the scope of this paper; characteristics of convective observations are treated more fully by Mahoney et al. (2004a,b).

Figure 1 illustrates some of the difficulties associated with diagnosing forecast problems using standard verification statistics. This figure shows five examples of forecast/observation pairs, with the forecasts and observations represented as areas. For a forecast user, cases a-d clearly demonstrate four different types or levels of "goodness": (a) appears to be a fairly good forecast, just offset somewhat to the right; (b) is a poorer forecast since the location error is much larger than for (a); (c) is a case where the forecast area is much too large and is offset to the right; (d) shows a situation where the forecast is both offset and has the wrong shape. Of the four examples, it appears that case (a) is the "best." Given the perceived differences in performance, it is dismaying to note that all of the first four examples have identical basic verification statistics (POD=0, FAR=1, CSI=0) indicating no skill. Thus, the verification approach is insensitive to differences in location and shape errors. Similar insensitivity could be shown to be associated with timing errors. Moreover, example (e) - which could be considered a very poor forecast from a variety of points of view - actually has some skill (POD, CSI >0), suggesting it is a *better* forecast than the one depicted in example (a).

These examples demonstrate that it is difficult to diagnose some of the sources of error in a forecast using traditional verification statistics alone. While these measures may be useful for providing an aggregate measure of overall performance or for tracking performance over time, they tell us little about the sources of errors or their operational impacts. For example, each type of error illustrated in Fig. 1 would lead to different aircraft route impacts, which cannot be distinguished by examining the overall verification statistics. The goal of the work described here is to develop a verification approach that will clearly distinguish the types of errors associated with a forecast, including forecasts like those illustrated in Fig. 1.

3. ALTERNATIVE APPROACHES

3.1 Background

The general goal of the object-oriented verification approach is to reduce the forecast and observed convection/precipitation areas to regions

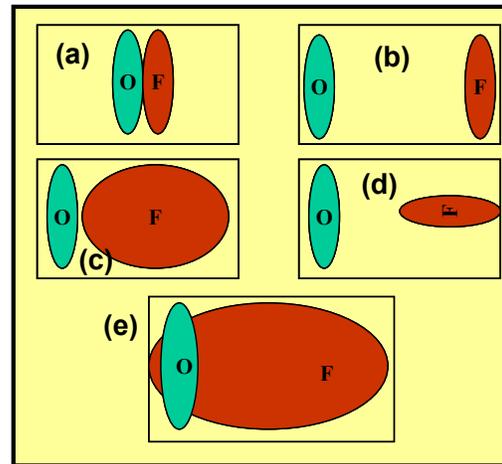


FIGURE 1. Schematic example of various forecast (F) and observation (O) combinations.

of interest that can be compared to one another in a meaningful way. Ebert and McBride (2000) had a similar goal in development of their entity-based verification approach, but they approached the problem from a somewhat different point of view. The main goals of the Ebert-McBride approach are to identify displacement vectors and to decompose the error statistics into their sources (e.g., displacement errors, pattern errors). An approach developed by Hoffman et al. (1995) also focused on decomposing the forecast error according to various attributes.

In general, the object-oriented approach requires several steps:

- Objects are defined in the observation and forecast fields;
- Adjacent objects or objects in close proximity that naturally "belong" together are merged into composite objects;
- Forecast and observed objects are optimally matched to each other;
- Differences between matched forecast and observed objects are computed for all relevant attributes; other verification measures (e.g., the size of the intersection area) are also computed, as well as measures that are meaningful for forecast developers and operational forecast users;
- Verification measures are aggregated across a variety of forecasts to reveal any systematic errors and to understand the overall performance of the forecasts.

Specific details of the application of the object-oriented approach depend on the characteristics of

the forecasts and observations. For example, in the case of forecasts that are already defined as shapes (e.g., human-generated forecasts), the forecast objects essentially are pre-defined and it is only necessary to define the observed objects. In some cases the observed objects may also be pre-defined.

More details about the application of the object-oriented approach are presented in the following sections. Specific information for particular types of forecasts and observations is also provided.

3.2 Gridded forecasts and observations

As suggested in the previous section, the more complex application of the technique involves situations where both the forecasts and observations are represented by a grid. In this case, objects must be defined for both the forecast grid and the observation grid. However, it is important to remember that all spatial shapes on a map can be represented as a gridded field. Thus, the “gridded forecasts and observations” situation is the most generic application of the approach.

3.2.1 Characterizing areas of precipitation/convection

The approach for defining the objects, for both the forecasts and the observations, involves application of a convolution function (essentially a smoothing function) to the spatial grid, followed by use of a threshold to eliminate areas that are not of interest. Our experience with precipitation fields has indicated that a circular convolution function provides good results in most cases. The convolution filter essentially assigns the average value in the surrounding circular region to the point in the center of the circle. Thresholding the convolved field allows object boundaries to be detected. The combination of these two steps results in boundaries that are similar to those that a human would draw. It might seem more straightforward to simply apply a threshold to the raw grid. However, this approach would lead to fields that are very spiky and discontinuous and not at all like the regions that a human would identify subjectively. Finally, although the field is thresholded, the underlying values (i.e., precipitation intensity or convection occurrence/non-occurrence) are retained for further analysis.

In summary, only two parameters are needed for this process: the convolution radius and

the masking threshold. Figure 2 shows an example of the process applied to precipitation forecast fields from the Weather Research and Forecasting (WRF) NWP system on a 22-km grid scale across the CONUS. Further information about these steps in the process is presented by Bullock et al. (2004) and Davis et al. (2004).

3.2.2 Merging and matching objects

The purpose of the merging step is to combine objects that naturally go together (e.g., adjacent objects that may appear to be part of the same weather system), in either the observation or forecast field. Similarly, the purpose of the matching step is to optimally match forecast and observed objects that naturally go together. By nature, these two steps are interrelated; in fact, in the current version of the approach, the merging of objects follows directly from the process of matching objects. However, future enhancements to the system are likely to include additional criteria for both merging and matching of objects. Merged objects are referred to as composite objects.

The process of matching objects is based on an application of fuzzy logic (e.g., Yager et al. 1987). First, a set of attributes or characteristics is measured for each forecast object and each observed object. These attributes include the distribution of intensity values within the object (represented by various percentile values); the object area; the centroid location; the orientation angle; and the object curvature. The attributes for each pair of forecast and observed objects are then compared by computing the difference or ratio of the attribute values, and other attributes that apply to that pair (e.g., the area of overlap) are also computed. Each of the attribute comparison values for a pair of objects is then mapped to a predefined “interest” curve to determine the interest value that is assigned to that pair of objects for that particular attribute.

An example of one of the interest maps is shown in Fig. 3. For this attribute comparison - the ratio of the area of the forecast object to the area of the observed object - the interest value is low when the ratio is small (i.e., when the forecast shape is much smaller than the observed shape); increases to one when the ratio is near one (i.e., the forecast object and observed object have about the same size); and decreases toward zero as the ratio becomes large (i.e., the forecast object is much bigger than the observed object).

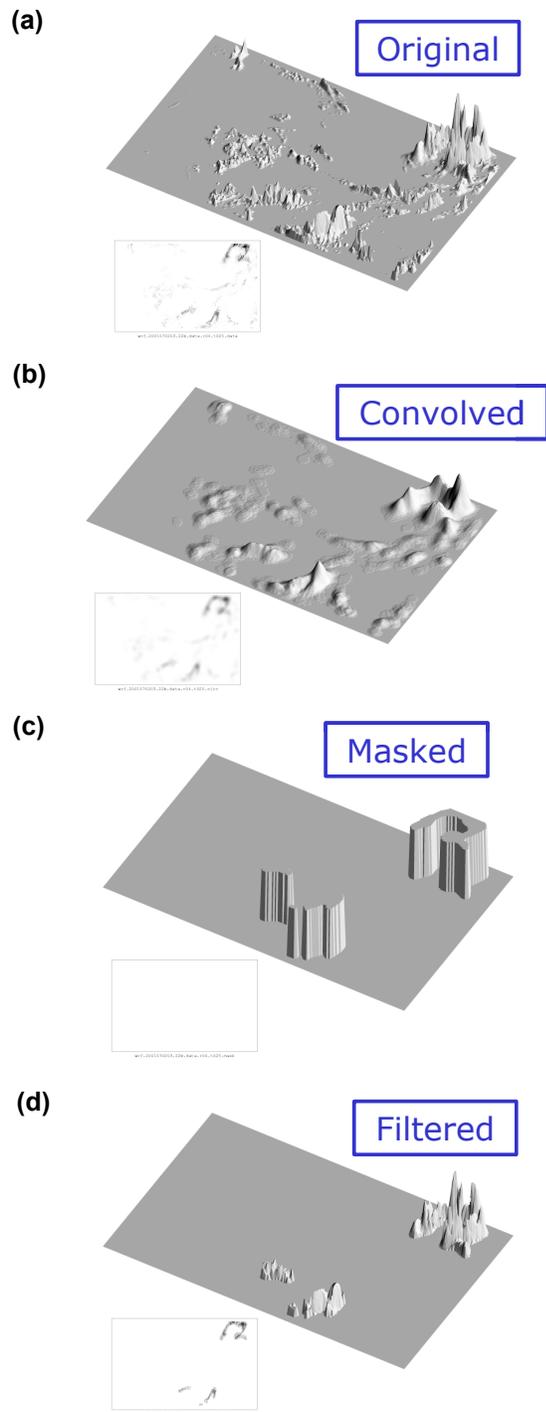


FIGURE 2. Example (in 3 dimensions) of a precipitation forecast field from the WRF model and the steps in the process of identifying objects in the field: (a) original precipitation field; (b) convolved (smoothed) field; (c) “masked” field after a threshold has been applied, leaving the areas of interest; (d) original precipitation values underlying the masked objects.

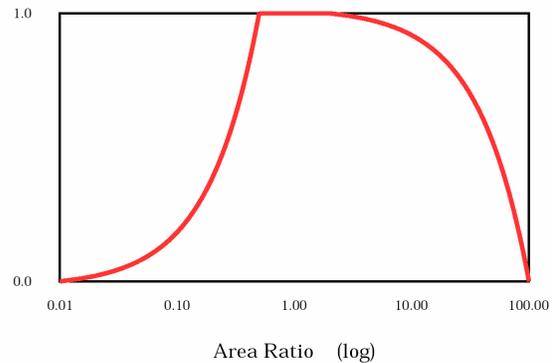


FIGURE 3. Example of an interest map, in this case for the ratio of the area of the forecast object to the area of the observed object.

Finally, the interest values for all of the attributes are combined as a weighted sum to determine an overall interest value that can be assigned to the pair of objects. These overall interest values are used to objectively define pairs of matched forecast and observed objects, by comparing the overall interest values to a threshold. Two (several) objects are merged if they both (all) are matched to the same objects in the other field. That is, if two forecast objects are matched to the same observed object, the two forecast objects are combined into a composite object and matched to the observed object, and *vice versa*.

Note that only object pairs with an overall interest value that exceeds the threshold are matched to one another. In some cases an object may not have a matching object in the other field, and thus it is an “unmatched” object. Such cases are not unanticipated; essentially these objects are false alarms if they are in the forecast field and they are missed objects if they are in the observed field.

3.2.3 Verification measures

Many different measures can be computed based on the results of the object matching process. In particular, characteristics of each of the attributes defined in the previous section can be compared between the forecast and observed objects. For example, the location of the centroid (i.e., “center of mass”) of the forecast object relative to the centroid of the observed object may be of interest to answer questions regarding whether the forecast is consistently located too far to the east or west. Characteristics and differences can be summarized using various statistical measures

(e.g., mean, median, extreme percentiles) or distribution representations (e.g., histograms, box plots, scatter plots), as appropriate.

3.3 *Forecast shapes and gridded observations*

The object definition and matching process can be simplified if the forecasts are already defined as shapes, such as the forecasts and advisories that are commonly produced by human forecasters. Examples of shape forecasts are the Convective Significant Meteorological Advisory (C-SIGMET; NWS 1991) and the Collaborative Convective Forecast Product (CCFP). Both of these forecasts are issued by forecasters at the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) Aviation Weather Center (AWC). The C-SIGMETs are advisories and 1- and 2-h forecasts of significant convective activity over regions covering an area at least 3,000 mi² in size, with a minimum areal coverage of 40%. Similarly, the CCFP forecasts are 2-, 4-, or 6-h outlooks for convective activity with at least 25% coverage over a minimum area of 3,000 mi². The CCFP is produced by AWC forecasters in collaboration with meteorologists at airline meteorology departments, Center Weather Service Units, and other facilities.

Essentially two different approaches can be taken with regard to these types of forecasts: (i) Optimize the “placement” of the forecasts relative to the gridded observations; or (ii) Define objects in the observation field and match them to the forecast shapes using the fuzzy logic approach. Applications thus far have involved the first approach. The disadvantage of this approach is that it requires an immense amount of processing time because it is necessary to test a large variety of locations and orientations for the forecast shapes. The advantage of this more straightforward optimization approach is that the matching process may be somewhat less arbitrary; the goal simply is to find the best possible score given a particular set of forecast shapes, regardless of whether they clearly match the corresponding shapes in the observations. The main advantage of the fuzzy logic shape matching approach in this context is that it requires much less processing. Moreover, the object characteristics are clearly defined by the matching process. Both approaches leave open the possibility of future enhancements to the verification procedure in which additional forecast shapes might be

added or forecast shapes might be altered to make the forecasts more optimally fit the observations.

4. EXAMPLES

4.1 *Gridded forecasts and observations*

The main application in this context so far has been to spatial QPFs produced by the WRF model. The forecasts for this example are on a nominal 22-km horizontal grid, and are matched to precipitation observations from the NCEP Stage IV precipitation analysis (see <http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/QandA/#STAGEX>). These observations combine radar and rain gauge measurements and are produced hourly on a 4-km grid; for this study the observations were re-mapped to the 22-km WRF grid.

Figure 4 shows an example of a set of forecast and observed objects for a particular case: the 12-h WRF precipitation forecast valid at 0000 UTC on 2 July 2001. Although the geographic map is not shown, the scale of the forecast essentially includes the whole CONUS. The figure shows a case in which 10 forecast objects were identified, along with 11 observed objects (Fig. 4a and c). However, the merging process led to four composite forecast objects, four composite observed objects, and two individual observed objects (Fig. 4b). Each composite forecast object was matched to one of the composite observed objects, but the two observed objects that were not merged into a composite object were not matched to a forecast object. The final panel in Fig. 4 (Fig. 4d) shows the overlap of the observed shapes on the forecast objects and the overlap of the forecast objects on the observed shapes.

Some basic verification statistics for the example composite shapes are shown in Tables 1 and 2. Table 1 shows the intersection area (IA) and symmetric difference (SD) area values, which are two basic indicators of the correspondence between the forecast and observed objects. In particular, IA measures the area of overlap between the two objects. The SD is the total area covered by the forecast and observed objects, less the intersection area. As indicated in Table 1, SD is generally much larger than IA, which indicates the differences between the forecast and observed objects are quite large. In one case (composite objects A) the forecast and observed objects do not intersect.

Table 2 shows the values of some basic attributes of the forecast and observed objects.

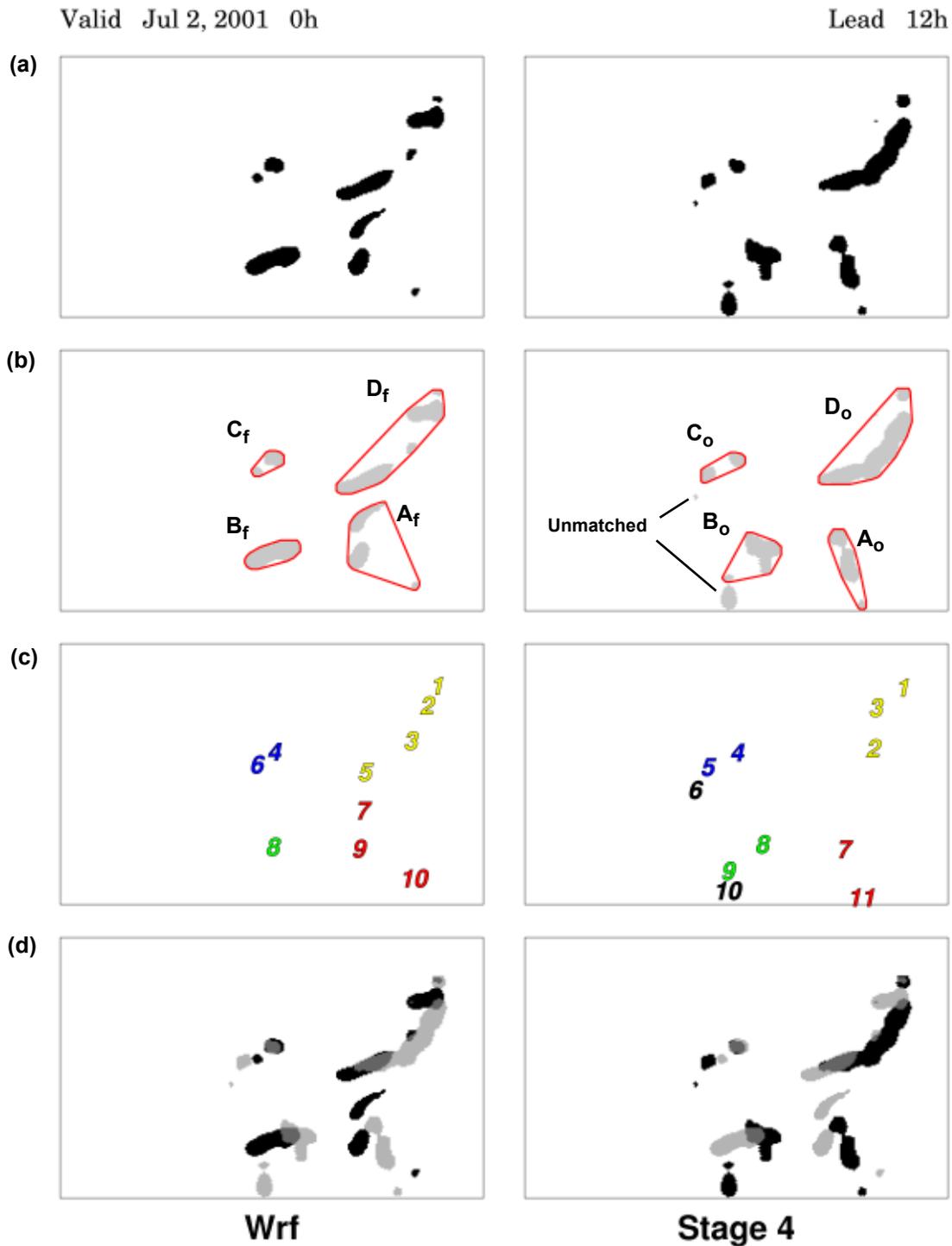


FIGURE 4. Example WRF and Stage 4 case for 12-h WRF precipitation forecasts valid on 2 July 2001 at 0000 UTC: (a) precipitation objects defined by convolution-threshold approach; (b) Merged and matched areas; (c) individual object identifiers, where colors indicate objects that are matched and merged (a black number indicates the object was not matched to any objects in the other field); and (d) overlap areas between forecast and observed objects.

TABLE 1. Verification statistics (intersection area and symmetric difference area) for example presented in Fig. 4. Intersection area is the area of overlap between the forecast and observed object; symmetric difference is the total area covered by the forecast and observed object, excluding the intersection area. Both values are in units of gridboxes.

Composite object	Intersection Area (IA)	Symmetric Difference (SD) Area
A	0	638
B	100	488
C	66	134
D	259	905

TABLE 2. Verification measures/attributes for objects shown in Fig. 4. Area values are in units of grid boxes; intensity is in mm; and centroid location is measured in grid squares.

Attribute	WRF	Stage IV	Difference
Composite Objects "A"			
Centroid X	187	197	-10
Centroid Y	44	31	13
Intensity (0.50)	4.7	2.5	2.2
Intensity (0.90)	8.5	13.9	-5.4
Area	319	319	0
Composite Objects "B"			
Centroid X	130	144	-14
Centroid Y	36	36	0
Intensity (0.50)	4.7	2.0	2.7
Intensity (0.90)	8.7	9.7	-1.0
Area	355	333	22
Composite Objects "C"			
Centroid X	128	121	7
Centroid Y	93	90	3
Intensity (0.50)	4.0	2.4	1.6
Intensity (0.90)	8.5	11.3	-2.8
Area	126	140	-14
Composite Objects "D"			
Centroid X	205	215	-10
Centroid Y	102	100	2
Intensity (0.50)	3.8	3.8	0
Intensity (0.90)	7.4	13.8	-6.4
Area	585	838	-253

These attributes measure such characteristics as the location of the objects (centroid location in the x and y directions), precipitation intensity 0.50th and (0.90th percentiles of the intensity distributions) and overall object size (area). For this simple case, some basic characteristics of the forecasts can be inferred. For example, for all composite objects except one (C), the forecast object is located too far to the west; in all cases except for one (B), the forecast object is located at least somewhat too far to the north; generally the median forecast intensity is somewhat too large, whereas the 0.90th percentile of forecast intensities is too small (implying the forecast did not correctly capture extreme precipitation occurrences); and forecast areas C and D are too small, while forecast area B is somewhat too large. Finally, it is important to note that two relatively small observed objects were not matched to any of the forecast objects; the areas of these observed objects were 7 and 121 grid boxes.

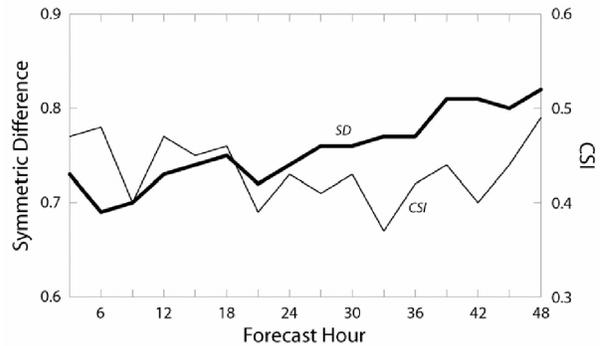


FIGURE 5. Symmetric difference (SD) and critical success index (CSI) for matched objects as a function of lead time (h).

Although these results are based on a single case, they are indicative of the types of information that can be obtained from these types of comparisons. Figure 5 is an example of one way these statistics can be summarized across a larger set of cases. For this analysis, WRF precipitation forecasts for July through August 2001 were analyzed; only objects in the area of the CONUS east of the Rocky Mountains were considered, to make the results more homogeneous. The lines in Fig. 5 show the CSI value for matching objects and the average SD value for matched objects as a function of lead time. Because the CSI line is relatively constant for all lead times, it can be inferred from this figure that the ability to create matched forecast and observed objects does not diminish with increasing lead time. In contrast, the average SD value increases with lead time, which indicates that

the errors in the forecasts grow with lead time. Additional summary plots are presented in Davis et al. (2004).

4.2 Forecast shapes and gridded observations

Examples of two types of human-generated forecasts are considered in this application of the general object-matching approach: the CCFP and the C-SIGMETs. Figures 6 and 7 show examples of each type of forecast, along with the resulting object optimization. In the case of the CCFP, the forecast locations and orientations were optimized directly by searching the grid, rather than by matching shapes. For the C-SIGMETs, objects in the observation field were identified and used to provide an initial guess regarding the optimal location of a C-SIGMET. Because no additional forecast areas were included in the forecasts (i.e., the bias and FAR values could not be changed by moving the forecast shapes), the optimizations were based on the value of POD_y alone. For these forecasts, the observations used for verification represent the “convective constrained area” (CCA) estimated from radar and lightning measurements. The CCA is intended to represent the region around convection where air traffic is impacted by the weather, and is expressed as a percent coverage value across a 3,000 mi² area; for more information about computation of the CCA, see Mahoney et al. (2004a,b).

In the case of the CCFP (Fig. 6), the forecasts are optimized by relatively small movements of the forecast shapes, as shown in Table 3. For example, shape number 2 is optimized by a small rotation as well as a small change in location in the north-south direction. Although these optimization strategies do not result in large increases in POD_y for the optimal forecasts, they clearly diagnose the characteristics of the errors in the forecasts.

As noted earlier, the C-SIGMET optimization process is somewhat more complex than the process applied to the CCFP. For the initial step, the centroid of each C-SIGMET was matched to the centroid of the best-fit observed CCA object. Following this matching, the location and orientation of the best-fit forecast shape were optimized by testing a variety of possible choices. The green shapes in Fig. 7 show the actual C-SIGMET location and orientation; the gold shapes show the first guess at an optimal location and orientation; and the blue shapes show the final optimization.

TABLE 3. CCFP optimization statistics showing displacement and orientation errors for the sample forecasts in Fig. 6.

Shape number	Rotation (deg)	x-movement (km)	y-movement (km)
1	180	40	-40
2	-5	0	-120
3	-95	-160	0
4	15	-120	80
5	0	-240	-80

The case presented in Fig. 7 shows C-SIGMETs that were initiated at 0200 UTC on 5 July 2003. The C-SIGMETs were advected using the motion vector included in the C-SIGMET message, so all of the C-SIGMETs shown in Fig. 7 were valid at 0400 UTC. Several small line shapes in Fig. 7 were not changed through the optimization process; these shapes could not be matched to a region with observed convection.

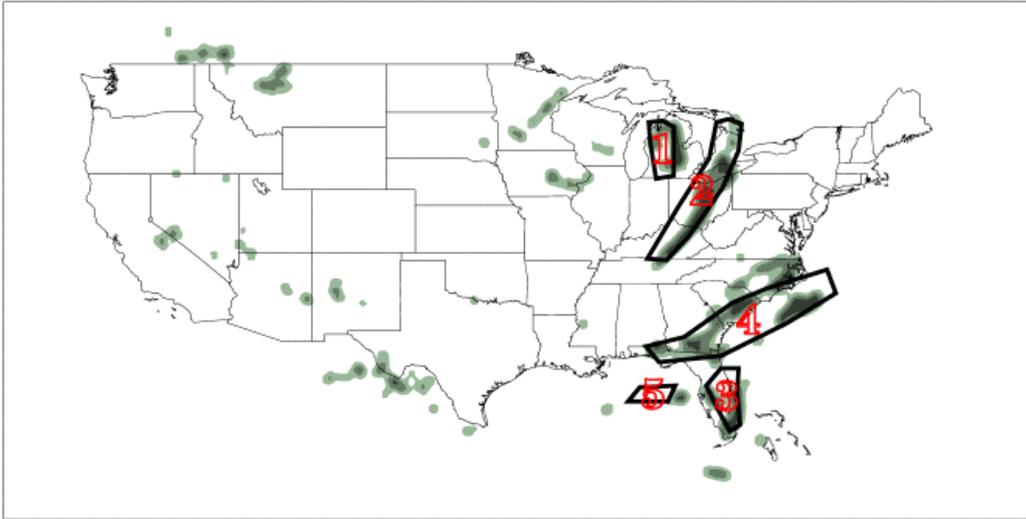
As was the case for the CCFP example, the changes in the verification scores achieved by the optimal forecasts was not large; nevertheless, it is clear that at least in some cases (e.g., observed shape 12 in Fig. 7) the actual location of the forecast was not optimal; it appears that the forecast could have been improved by not advecting the C-SIGMET area as quickly over the 2-h time period. This type of diagnostic information on its own could provide useful feedback to the C-SIGMET forecasters. It also could be summarized across a set of cases to characterize overall performance of the forecasts.

4.3 Probabilistic forecasts

The issues that have been considered here essentially apply to all types of spatial forecasts. It is important to note that although probabilistic forecasts provide numerous advantages over the more common deterministic types of spatial forecasts that have traditionally been produced and used, they do not eliminate the need to apply approaches that are more diagnostic and that are able to indicate the sources of the errors in the forecasts. For example, traditional verification statistics for probabilistic forecasts are very sensitive to displacement errors. This effect is illustrated here with a simple example. In particular, the cartoons in Fig. 8 represent a spatial probability forecast in which five different probability values can be forecast, levels 0-4, where level 0 represents the area outside the blue contours).

Threshold: 0.25

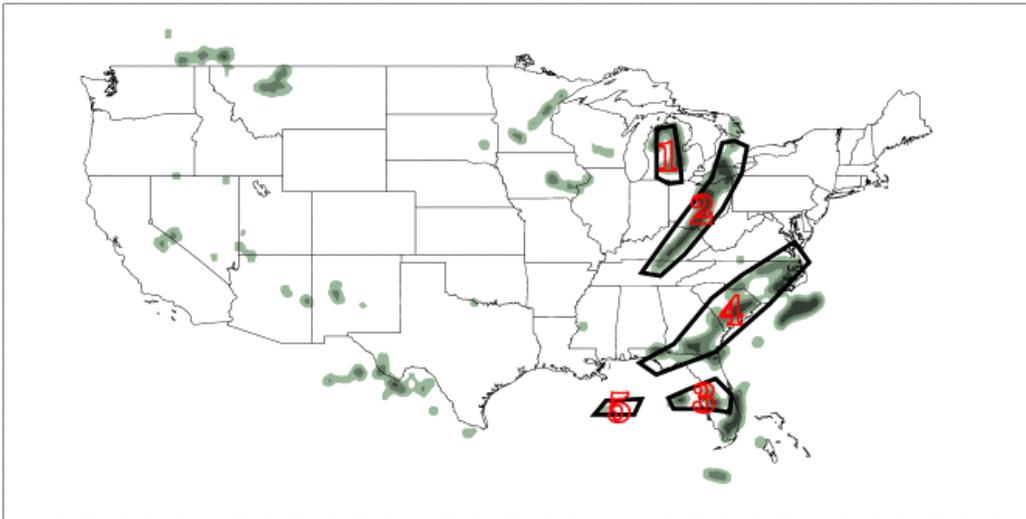
(a)



Original Configuration

PODY = 0.1859

(b)



Optimized Configuration

PODY = 0.2183

FIGURE 6. Example of CCFP optimization for 2-h CCFP forecasts valid at 2100 UTC on 8 June 2003: (a) original forecasts; and (b) forecasts with optimized location and orientation.

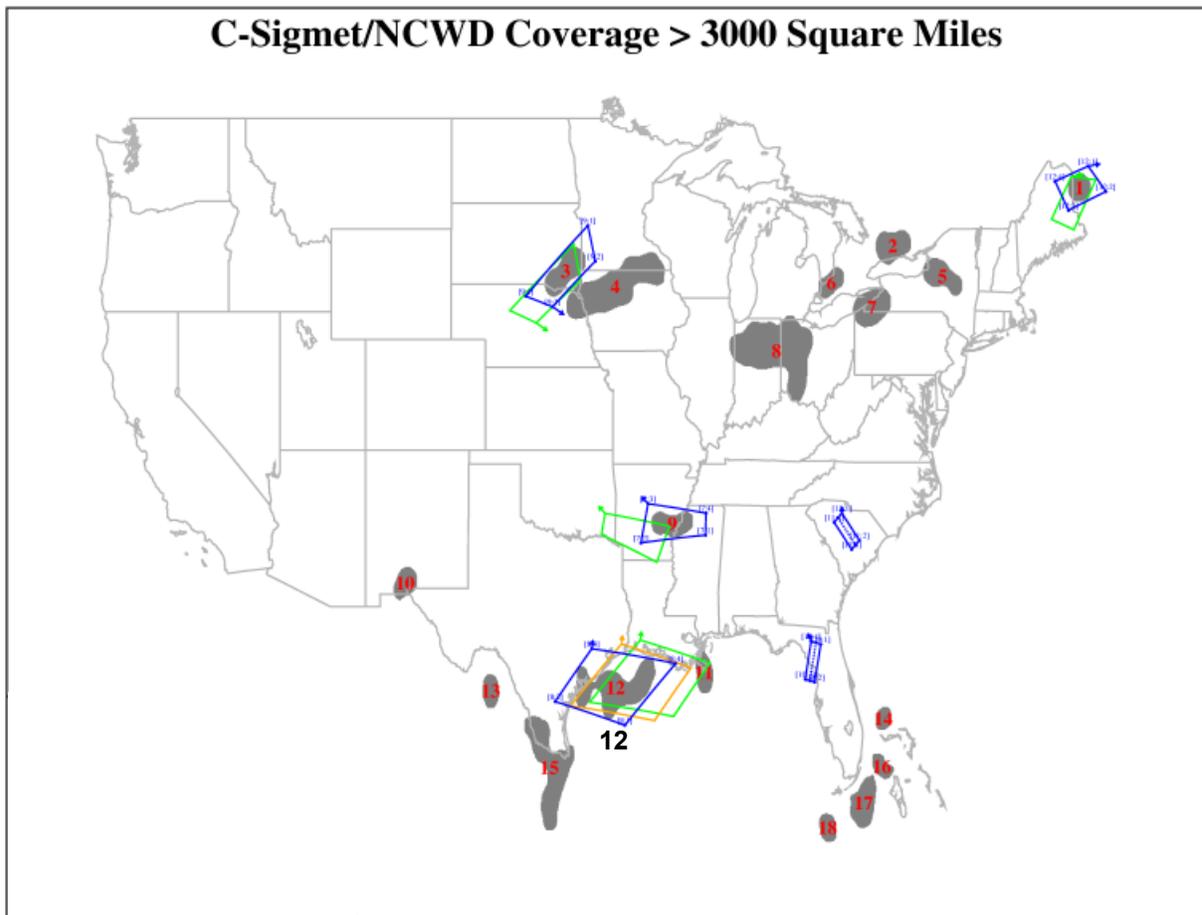


FIGURE 7. C-SIGMET optimization example for C-SIGMETs valid at 0400 UTC on 5 July 2003. Original C-SIGMET locations are shown in blue, initial optimal locations in gold, and final optimal locations are shown in light green. Observed objects are identified using red numbers; black number 12 for that object is presented for clarity only.

Figure 9 shows the frequency of use of each of the forecast probability levels. Note that Forecasts A and B differ **only** in the location of the forecast regions relative to the observed region of convection: the areas representing the forecast probability levels are the same size for both forecast examples. In particular, the forecasts are the same size and shape and have the same values. However, the ellipses in Forecast A completely overlap the convection area, whereas the forecast ellipses in Forecast B are displaced to the south-east.

An overall evaluation of the quality of the probability forecasts is provided by reliability diagrams and relative operating characteristic (ROC) plots. Together these diagrams measure two important attributes of the capabilities of the forecasts. In particular, the reliability diagrams show

how consistent the relative frequency of occurrence of the event is with the forecast probability value. The ROC measures the forecast's ability to discriminate between situations when the event (i.e., convection) occurs and when it does not occur; it is a plot of the hit rate vs. the false alarm rate for various probability thresholds.

The reliability diagrams for the two forecasts in Fig. 8 are shown in Fig. 10. The lines in this figure indicate that Forecast A is much more reliable than Forecast B (or at least has the potential to be more reliable, depending on the calibration of the forecast values). In particular, for Forecast B the frequency of convection varies only slightly with the probability level, whereas a great deal of differentiation is demonstrated by Forecast A.

The ROC diagrams presented in Figure 11 also show that Forecast A has much more skill

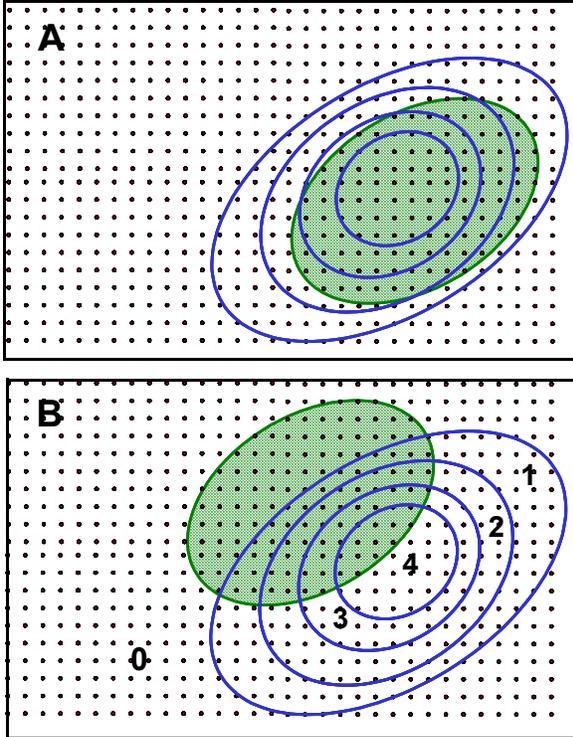


FIGURE 8. Schematic spatial probability forecasts. Blue contours represent different probability forecast levels (0-4). Green ellipse represents the observed Yes field.

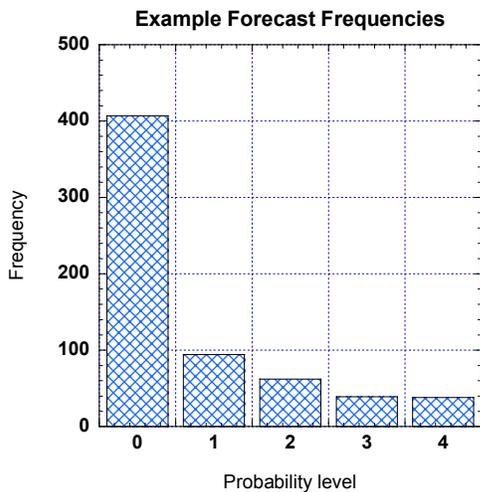


FIGURE 9. Relative frequency of use of forecast probability levels in Fig. 8.

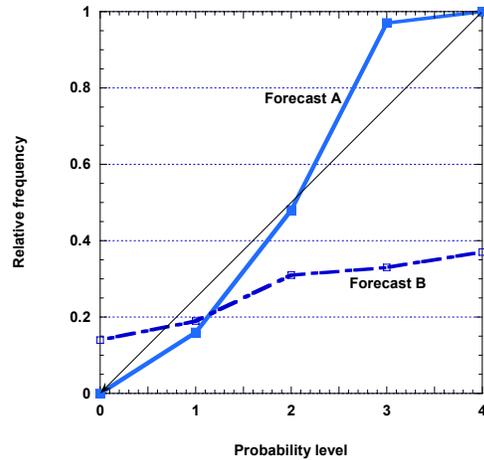


FIGURE 10. Reliability diagrams for forecasts shown in Fig. 8. X-axis is the probability level; Y-axis shows the frequency of occurrence of the "event" for each of the probability levels.

than Forecast B. In the ROC diagrams, greater skill is demonstrated by lines that lie closer to the upper left corner of the diagram, and it is clear that the line for Forecast A is much further to the upper left than the line for Forecast B. Another way to consider this measure of skill is the area under the ROC curve; clearly, the area under the curve for Forecast A is much larger than the area under the curve for Forecast B. These results indicate that Forecast A is much more skillful than Forecast B in terms of its ability to correctly classify Yes and No observations of convection.

These results suggest that the verification results for spatial probabilistic forecasts can be strongly impacted by displacement errors. Application of object-oriented verification techniques to these types of forecasts would make it possible to decompose the errors according to their various sources (including displacement and shape errors). This approach would also make it possible to diagnose the impacts of various improvements in placement, orientation, size, and so on.

5. FUTURE WORK

A number of tests and enhancements are planned for the future. For example, the verification technique will be applied to a wide variety of cases and to a variety of types of forecasts. The new cases will allow more complete testing of the methodology. Although it is difficult to objectively test the approach and unambiguously measure its specific capabilities, examination of many different cases

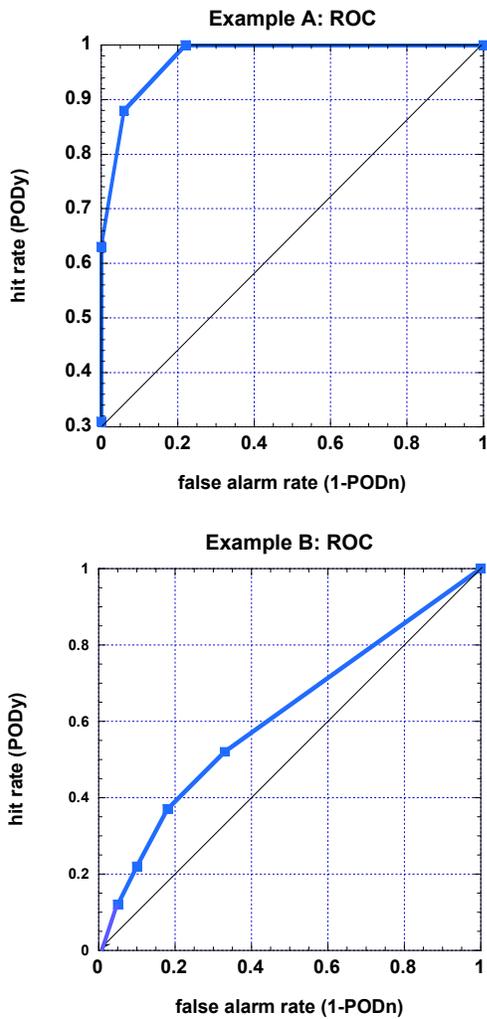


FIGURE 11. ROC plots for two forecasts shown in Fig. 8.

will allow us to evaluate its consistency and reasonableness. Sensitivity tests will clarify the robustness of the technique relative to the various settable parameters. New types of forecasts to be investigated include the National Convective Weather Forecast (NCWF) and higher resolution WRF precipitation forecasts. The new version of the NCWF (NCWF-2) is in the form of a probability forecast, which will allow us to test the methodology on actual probabilistic forecasts.

A number of additional steps are needed to make the approach fully functional as a verification tool. For example, we will identify a set of general verification measures that summarize the basic attributes of the forecasts and their quality in the

context of the object-oriented approach. Some of the likely measures have been described here, but others will be defined as we examine additional datasets. In addition to the basic statistics, it will be necessary and valuable to define specific measures that are meaningful to particular users and provide information that is operationally relevant for particular applications.

Several enhancements to the method have already been identified. For example, it has become apparent that the methodology can be improved by including the time dimension in the object identification and matching process (much as a human would do). We also may investigate the possibility of evaluating objects in three dimensions. Alternative approaches for matching objects are also being investigated. For example, we are testing a binary image matching technique developed by Baddeley (1992) and comparing the results of this approach to the results of the fuzzy logic approach. In addition, as the verification technique matures, local/regional characteristics (e.g., orography) may be incorporated into the matching process.

Finally, the approach described here clearly is highly dependent on the scale of the forecasts and observations. The approach naturally takes these effects into account through the definition of the radius and threshold parameters. Thus, scale effects could simply be handled by applying a variety of different parameter combinations and examining the variations in the verification statistics as a function of the parameters. In addition, we plan to investigate more direct approaches such as the wavelet (intensity-scale) decomposition approach recently developed by Casati et al. (2004).

6. CONCLUSIONS

The object-oriented verification approach for spatial forecasts has the potential for providing new useful verification information that cannot be obtained using standard approaches. The method described here is complementary to the approach developed by Ebert and McBride (2000) as well as other approaches that are under development (e.g., Baldwin 2003; Casati et al. 2004). Future work on this topic will lead to verification information that is meaningful in a variety of operational settings and provides feedback to forecast developers to implement meaningful improvements to forecasting systems.

ACKNOWLEDGMENTS

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

This work was partially supported by the National Science Foundation and the U.S. Weather Research Program. NCAR is sponsored by the National Science Foundation.

REFERENCES

- Baddeley, A.J., 1992: An error metric for binary images. In *Robust Computer Vision: Quality of Vision Algorithms*, W. Forstner and S. Ruwiedel, Editors, Wichmann Verlag, 59-78.
- Baldwin, M.E., 2003: Development of an Events-Oriented Verification System Using data mining and Image Processing Algorithms. 3rd Conference on Artificial Intelligence Applications to the Environmental Science. *Preprints, 83rd AMS Annual Meeting*, 9-13 February, Long Beach, CA, American Meteorological Society (Boston), available at <http://ams.confex.com/ams/pdfview.cgi?user-name=57821>.
- Brown, B.G., J.L. Mahoney, C.A. Davis, R. Bullock, and C.K. Mueller, 2002: Improved approaches for measuring the quality of convective weather forecasts. *Preprints, 16th Conference on Probability and Statistics in the Atmospheric Sciences*. Orlando, 13-17 May, American Meteorological Society (Boston), 20-25.
- Bullock, R., B.G. Brown, C.A. Davis, M. Chapman, K.W. Manning, and R. Morss, 2004: An Object-Oriented Approach to the Verification of Quantitative Precipitation Forecasts: Part I—Methodology. 17th Conference on Probability and Statistics in the Atmospheric Sciences, Preprints, 84th Annual Meeting, American Meteorological Society, <http://www.ametsoc.org/meet/index.html>.
- Casati, B., G. Ross, and D.B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, **11**, 141-154.
- Davis, C., B. Brown, and R. Bullock, 2004: Object-based verification of precipitation forecasts. Submitted to *Monthly Weather Review*.
- Ebert, E., and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, **239**, 179-202.
- 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.
- Hoffman, R.N., Z. Liu, J.-F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Monthly Weather Review*, **123**, 2758-2770.
- Jolliffe, I.T., and D.B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Wiley and Sons Ltd., 240 pp.
- Mahoney, J.L., J.E. Hart, and B.G. Brown, 2004a: Defining observation fields for verification of spatial forecasts of convection. *Preprints, 17th Conference on Probability and Statistics in the Atmospheric Sciences*, 12-16 January, Seattle, WA, American Meteorological Society (Boston). Available from www.ametsoc.org.
- Mahoney, J.L., S. Seseske, J.E. Hart, M. Kay, and B.G. Brown, 2004b: Defining observation fields for verification of spatial forecasts of convection: Part 2. *Preprints, 11th Conference on Aviation, Range, and Aerospace Meteorology*, 4-8 October, Hyannis, MA, American Meteorological Society (Boston). Available from www.ametsoc.org.
- Mass, C.F., D. Ovens, K. Westrick, and B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, **83**, 407-430.
- NWS, 1991: National Weather Service Operations Manual, D-22. National Weather Service. (Available at <http://www.nws.noaa.gov>).
- Rife, D. L., C. A. Davis, Y. Liu, and T. T. Warner, 2004: Predictability of low-level winds by mesoscale meteorological models. *Monthly Weather Review*, in press.
- Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Yager R.R., S. Ovchinnikov, R.M. Tong, and H.T. Hguyen, 1987: *Fuzzy sets and Applications, Selected Papers by L.A. Zadeh*. John Wiley and Sons, 685 pp.