10A.2 **APPLICATION OF THE MODE OBJECT-BASED VERIFICATION TOOL FOR THE EVALUATION OF MODEL PRECIPITATION FIELDS**

Barbara G. Brown[1], Randy Bullock, John Halley Gotway,
David Ahijevych, Christopher Davis, Eric Gilleland, and Lacey Holland

National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

In recent years, the verification and numerical modeling communities have increasingly focused attention on development of new tools for evaluation of spatial forecasts of elements such as precipitation, convection, and clouds. Because these elements exhibit significant spatial variability, their structure and specific location can be difficult to forecast; traditional verification measures [e.g., Critical Success Index (CSI), Root-mean-squared error (RMSE)] penalize the performance of forecasts with these types of errors without identifying the cause of the poor performance. In fact, for many users or decision making situations, small location errors may not be important; yet traditional verification measures indicate a forecast with these errors has little or no skill.

Spatial forecast verification issues are particularly relevant for high resolution forecasts (e.g., from fine scale or mesoscale models). Many of the new verification developments have focused on methods that provide more diagnostic information about forecast performance and can separate location errors from magnitude and other errors. Several different types of approaches have been developed, including scale-separation techniques (e.g., Casati et al. 2004); entity-based approaches that involve optimal matching of forecast and observed precipitation regions (Ebert and McBride 2000); and object-based approaches, in which forecast and observed areas of precipitation (or other element of interest) are represented and compared as objects, characterized by attributes such as location, size, and intensity (e.g., Baldwin and Lakshmivarahan 2003; Marzban and Sandgathe 2006, Davis et al. 2006).

This paper describes a new object-based verification tool, the Method for Object-based Diagnostic Evaluation (MODE) and its application to precipitation forecasts from the Weather Research and Forecasting (WRF) model. The MODE tool is included in the first version of the Model Evaluation Tools (MET), a set of model verification tools that will soon be available to the WRF community.

The MODE tool is described more completely in Section 2, and an example of the application of MODE to WRF precipitation forecasts is presented in Section 3. Because results obtained from the application of spatial verification methods vary significantly as a function of the spatial scale of the forecast and observation fields, variations in performance as a function of scale are considered in Section 4. Future development and applications of MODE are discussed in Section 5.

[1]*Corresponding author address*: Barbara G. Brown, National Center for Atmospheric Research (NCAR), PO Box 3000, Boulder CO 80307-3000; e-mail: bgb@ucar.edu

2. THE MODE TOOL

The initial motivation for development of MODE was to provide a tool that would be able to mimic a human analyst's evaluation of forecast performance. For example, a skilled analyst is able to examine graphics showing forecast and observed patterns and infer the quality of a forecast - for example, whether the forecast area is too far north, too large, not intense enough. Our goal was to develop a tool that could objectively make these comparisons. The tool would then be able to provide meaningful diagnostic information regarding the good and bad attributes of large sets of forecasts.

MODE is based on a multi-step automated process, which includes the following stages:

(a) Object identification;
(b) Measurement of object attributes;
(c) Merging of objects in the same field;
(d) Matching of objects from the forecast and observed fields;
(e) Comparison of forecast and observed object attributes;
(f) Summarization and comparison across many cases.

These steps are summarized in Figure 1 and are described more specifically in the following subsections.

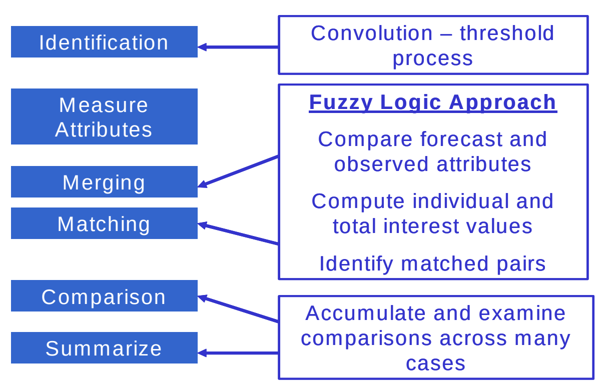

FIGURE 1. Summary of steps involved in application of the MODE approach.

2.1 *Object identification*

MODE is designed with the assumption that the forecasts and observations are on 2-dimensional scalar grids, with the same coordinates (i.e., matching grid points). The first step in the MODE process is to convert gridded precipitation values into precipitation objects. Although there are many possible ways to accomplish this step, the approach that was selected for MODE involves two processes: (a) convolution and (b) thresholding. The objects identified by this process are called "simple" objects.

The convolution process smooths the raw data (using a convolution filter); then a threshold is applied to create objects that look similar to what a human might draw After thresholding, the original data values are restored to object interiors, and the rest of the field is zeroed out. This process is illustrated with the example in Figure 2.

The convolution filter applied in Figure 2 is a simple circular function. Other types of functions (e.g., Gaussian, elliptical) could be applied instead of the circular function; however our experience has indicated that the circular function works well in most cases. The circular filter can be defined using a single parameter, R, the radius of the averaging region. In Figure 2, the radius applied is four gridpoints.

The second parameter required for object identification is the threshold, T. The choice of this parameter essentially depends on whether we are interested in more or less intense precipitation areas. Use of a large T will eliminate areas with low intensity precipitation and will result in smaller objects; use of a small T will result in broader areas of precipitation, and (in many cases) fewer objects. The value of T used in Figure 2 is 2.5 mm.

Together R and T can be used to represent and select the scale of precipitation of interest. This capability is discussed further in Section 4.

2.2 *Object attributes*

Attributes of the precipitation objects are used for three purposes: (a) to merge objects within a single forecast or observed field; (b) to match forecast and observed objects; and (c) to summarize the performance of forecasts by comparing attributes between matched forecast and observed objects. Many of the attributes are defined geometrically, including object characteristics such as location, size, aspect ratio, and complexity. All of these attributes are defined mathematically in Bullock et al. (2007). Other attributes are based on the precipitation values inside the objects; currently the MODE computes several quantiles of precipitation values (e.g., the 0.25th, 0.50th, 0.75th, and 0.90th quantiles) within each object. These quantiles provide a representation of the distribution of precipitation values, including the "average" value and the extremes, within an object.

In addition to the attributes mentioned here (and other typical attributes), specialized attributes related to a particular forecast application can also be defined. These attributes might only be used for comparing forecast and observed objects (i.e., they might not be used in the matching process). For example, for aviation applications it is of interest to know the density of precipitating systems. An attribute related to the density of objects could easily be devised and compared between forecast and observed objects.

Using these single object attributes, attributes for object pairs can now be defined. Typically one of the two objects constituting the pair will be a forecast object, and the other will be an observed object, though this need not always be true. These "pair" attributes are used in

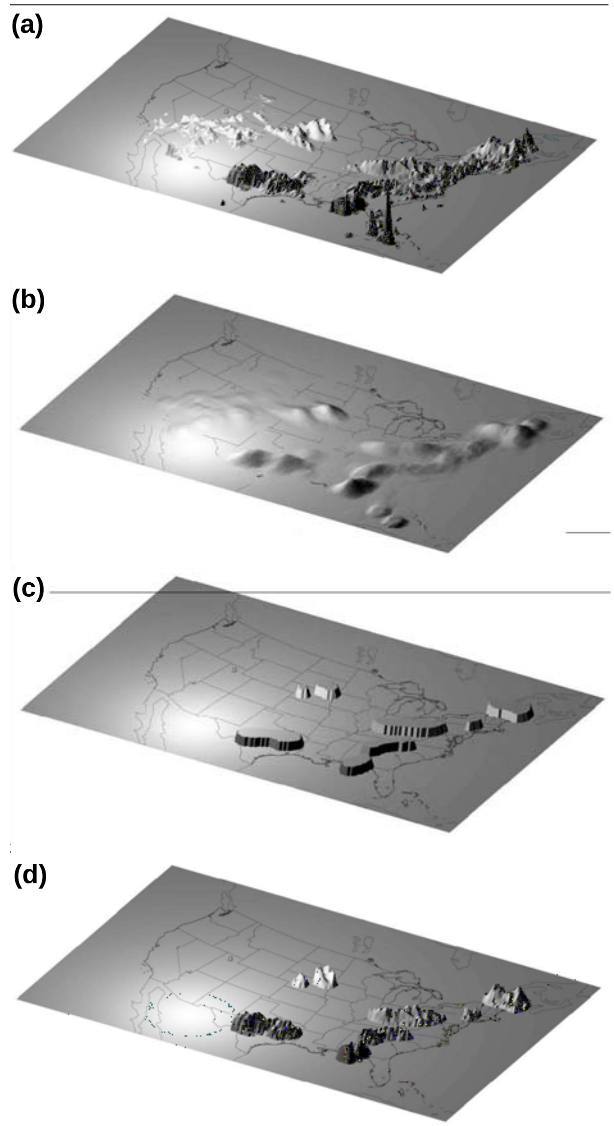

FIGURE 2. Example of a forecast of 3-h precipitation accumulation from the WRF model (a 12-h forecast valid at 0000 UTC on July 2, 2001) on a 22-km grid. Raw precipitation values are shown in (a), and the convolved values are shown in (b). The masked regions shown in (c) result from application of the threshold. The final field of objects is shown in (d), after the raw values are restored to each of the object grid points. For this example, the convolution radius, R, is 4 gridpoints, and the threshold, T, is 2.5 mm.

the process of object matching and merging (discussed in the next section). In particular, a pair attribute measures the similarity of the values of a particular attribute for the two objects in the pair. Typical pair attributes include the differences or ratios of the geometric and intensity attributes. Examples include centroid distance (i.e., the distance between two centroids); area ratio (i.e., the ratio of the areas of the two objects); intersection area (i.e., the amount of overlap between the two objects); angle difference (i.e., the difference in orientation between the two objects); and median intensity ratio (i.e., the ratio of the median intensity values for the two objects).

2.3 *Merging and matching*

Object matching is the process of associating objects in one field with objects in the other field. Object merging is the process of associating one or more objects in the same field; following merging the individual simple objects are considered to be parts of a larger composite object. When merging is done, attributes can be recalculated (if desired) for the composite object in the same way they were calculated for the component simple objects.

Currently MODE uses a fuzzy logic process (e.g., Yager et al. 1987) for both merging and matching objects. Although a variety of alternative approaches, including binary image matching (Gilleland et al. 2007) and cluster analysis (Marzban and Sandgathe 2006) could also be used, the fuzzy logic approach has the advantage of objectively applying human judgments about how objects should be matched, and allows matching and merging to be based on a large variety of attributes.

Suppose one has a collection of attributes $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ that are believed to be relevant to the merging and matching problem. The first step is to define interest maps $I_i(\alpha_i)$ to indicate which values of the attribute i should "count" heavily and which should not. Interest maps typically take values either in the range [0, 1] or [-1, 1], with 0 indicating no interest, and 1 indicating strong interest. Simple examples include (i) I_i is an increasing function (high values of α_i are more interesting), and (ii) I_i is a decreasing function (low values of α_i are more interesting). In general though, attributes often require slightly more complicated interest maps.

An example of an interest map used by MODE is shown in Figure 3. In this mapping, median intensity ratios between 0.7 and 1.5 are assigned an interest value of 1. Values less or greater than 0.7 are assigned smaller interest values, decreasing to a value of 0 for ratios of 0 and 4.

Next, confidence maps $C_i(\alpha)$ are chosen. These confidence maps reflect, roughly, how much we "believe" the (measured or calculated) value of the attribute i.

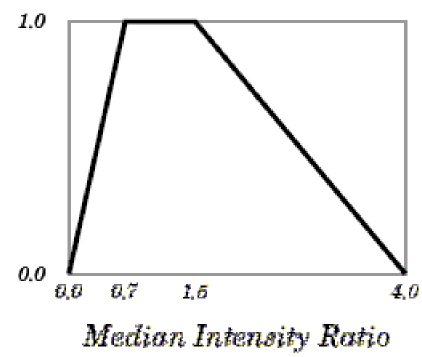

FIGURE 3. An example of an interest value map, for median intensity ratio.

Confidence maps typically take values in the range [0, 1]. Note that while each interest map I_i is a function of the single attribute i only, each confidence map C_i is a function of the entire attribute vector α.

As a simple example, consider the situation where the local wind is measured with a combined anemometer/windvane. If α_1 is measured wind speed, and α_2 is measured wind direction, then low values of measured wind speed will typically result in a poorly resolved wind direction. In this case the confidence map C_2 for the wind direction should be chosen so that low values of wind speed α_1 give a low value of C_2. For MODE, a confidence map is used for the angle difference paired attribute. In particular, for objects with an aspect ratio (ratio of the lengths of the two major axes) near to one, the orientation angle is not a meaningful measure. The confidence for the angle difference between two orientation angles is the geometric mean of the confidence values for the individual orientation angles.

Finally, scalar weights w_i are chosen for each attribute to reflect their perceived relative importance. Typically the weights are nonnegative. Assigning a value of zero to some w_i effectively "turns off" the contribution of α_i to the final decision. In our case, weights and interest and confidence maps were chosen through discussions with researchers in the field of convective precipitation systems. In general, the participation of such "experts" in the design of fuzzy logic systems is highly desirable.

All of these ingredients are combined to form the total interest function $T(\alpha)$:

$$T(\alpha) = \frac{\sum_i w_i C_i(\alpha) I_i(\alpha_i)}{\sum_i w_i} . \qquad \textbf{(1)}$$

A threshold is applied to the total interest function to determine whether a good match has been made. A typical total interest threshold applied in MODE is 0.7.

The fuzzy logic engine is used both to merge objects in individual fields and to match objects between forecast and observed fields. These steps can either take place in sequence or in parallel. In the sequential approach, the simple objects in the forecast and observed fields are individually merged into composite objects that are then matched. In the parallel approach, matching and merging occur at the same time. For example, the merging of observed simple objects may depend on whether particular sets of observed objects are well matched to the same forecast object.

Several approaches for the merging/matching steps have been investigated, and each has its own set of advantages and disadvantages. An approach that seems to perform quite well - in terms of identifying and matching reasonable composite objects - involves both the parallel approach and the application of a second threshold to the observed field to combine neighboring objects. This approach was used for the example shown in the following section.

3. APPLICATION TO WRF PRECIPITATION FORECASTS

The MODE has been applied to many cases from the WRF model. The example presented in this section is based on WRF output from the Storm Prediction Center's (SPC's) 2005 Spring Program, and in particular precipitation fields from the Advanced Research WRF (ARW) run at a 2-km horizontal resolution but provided on a 4-km grid. The observations are based on the NCEP Stage II precipitation analysis (Lin and Mitchell 2005).

Figure 4 shows the "raw" WRF and Stage II grids for a case valid at 0000 UTC on 1 June 2005. The values shown are 1-h precipitation accumulations; the WRF output is based on a 24-h forecast initialized at 0000 UTC. Figure 5 shows the precipitation objects created using a convolution radius of 15 grid squares and a threshold of 0.05 in. Color coding indicates which objects were merged and matched. Dark blue objects were unmatched.

For this case, five simple objects were identified in the WRF field and six simple objects were identified in the Stage II analysis. Two small WRF objects were unmatched and could be categorized as false alarm areas. The interest values for the matched simple objects were quite large, ranging from 0.85 to 0.97. A total of three composite objects were matched between the forecast and observed fields.

Some attributes of the composite objects shown in Figure 5 are listed in Table 1. These attributes include area, distance, and intensity measures, as well as the object intersection and union areas. Both individual and pair attributes are shown. As shown in Table 1, the three composite forecast objects were somewhat too large, compared to the analysis objects, indicating some overforecasting. This difference is relatively small for Object 3. Centroids of the forecast and observed composites were located 20-30 gridpoints apart; however, the convex hulls for all three objects overlapped, indicating the objects were in relatively close proximity. Finally, the WRF tended to overforecast precipitation intensity, for both the median and 0.90th quantile values of precipitation.

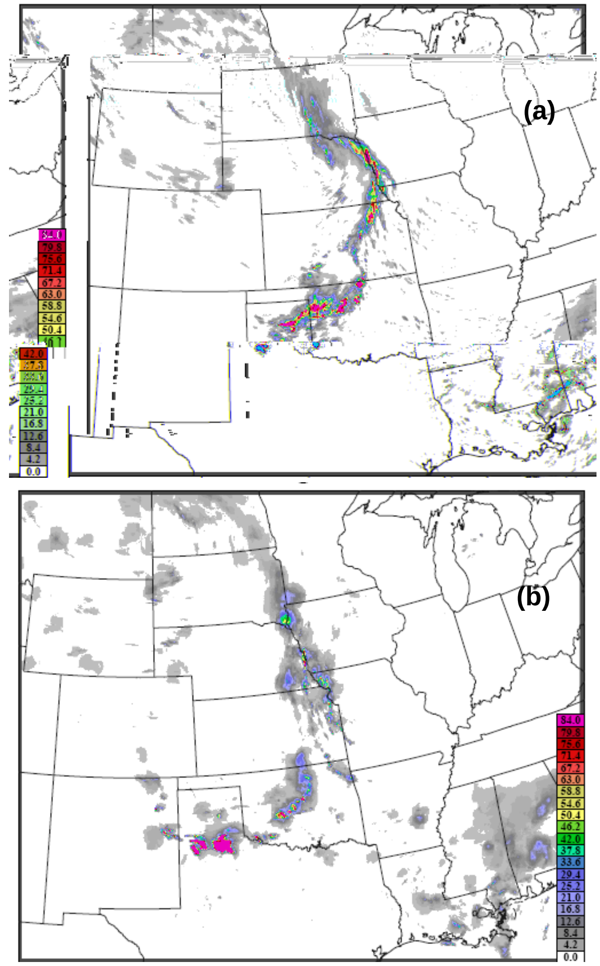

FIGURE 4. (a) WRF ARW-2 and (b) Stage II precipitation values for 0000 UTC 1 June 2005.

In contrast to these relatively simple but extensive attribute measures, traditional verification of this forecast would indicate POD=0.40, FAR=0.56, and CSI=0.27.

Ideally results like those computed for the 1 June case would be combined with results for a large number of cases to create a more complete picture of forecasting performance under varying conditions. In addition, as noted earlier, attributes could be evaluated that would be meaningful for specific forecast users. These steps would complete the set of steps involved in using MODE for forecast verification (Figure 1).

4. SUMMARIZING PERFORMANCE VARIATIONS WITH SCALE

The convolution radius, R, and precipitation threshold, T, are closely tied to the scale of the precipita-

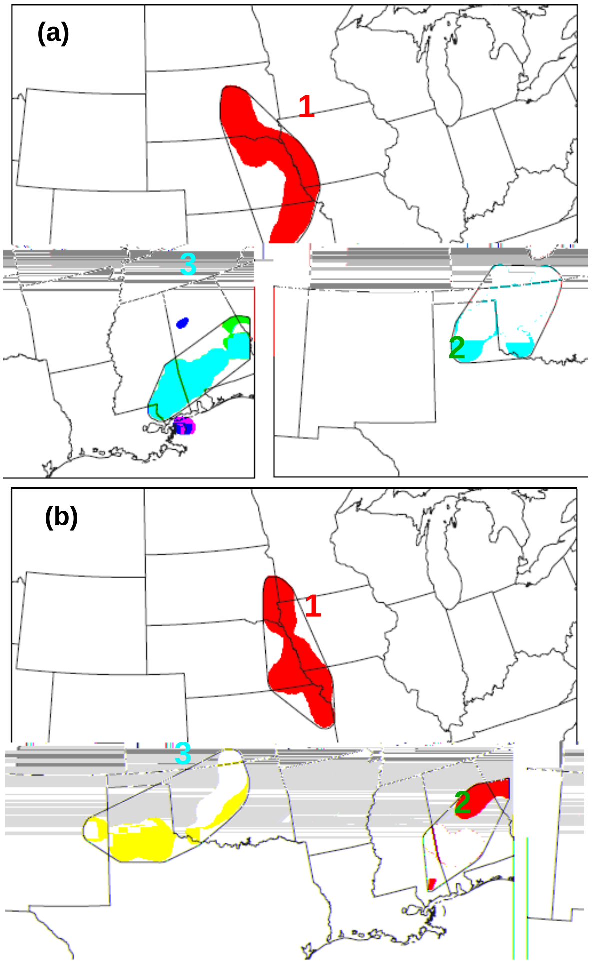

FIGURE 5. Single and composite objects identified for the case shown in Fig. 4 for (a) WRF and (b) Stage II analysis. Colors identify objects that are matched and merged. Dark blue objects are not matched.

tion areas that are considered by MODE. Thus, it is possible to examine forecast performance across scales simply by examining particular performance statistics as a function of R and T. The “quilt” plot presented in Figure 6 shows an example of a statistic that measures how well forecasts and observation objects are matched, as a function of R and T. The statistic shown is the median of the maximum interest values for each of the forecast simple objects. Larger values of this statistic indicate that the forecast and observed objects have a greater chance of matching (i.e., having an interest value that exceeds the threshold of 0.7 in this case). The values in Figure 6 are based on nine cases from the 2005 SPC spring program. Examination of this figure indicates that the (R,T) combination used to define the objects in Figure 5 are associacted with a region of the diagram with relatively large total interest values. In addition to examining performance (as defined by any attribute or attribute function of interest) as a function of scale, figures like this can help in the selection of reasonable values of R and T for defining objects.

TABLE 1. Example attributes and attribute comparisons for 1 June 2005 case shown in Figures 4 and 5.

Attribute or comparison	Composite object number		
	1	**2**	**3**
WRF area	7,830	4,632	6,933
Stage II area	5,973	3,746	6,585
Area diff	1,867	886	348
Area ratio	1.31	1.24	1.05
Intersection area	2,753	1,534	2,074
Union area	10,467	6,335	10.913
Int/Union	0.26	0.24	0.19
Centroid distance	255	288	226
Convex hull distance	0	0	0
WRF Median intensity	10.00	5.01	15.00
St II Median intensity	8.00	7.00	8.00
Median intensity ratio	1.25	0.72	1.88
WRF 0.90th intensity	47.01	43.01	67.00
St II 0.90th intensity	26.01	15.00	63.00
0.90th intensity ratio	1.81	2.87	1.06

5. SUMMARY AND CONCLUSIONS

This paper has described one of several emerging tools for the evaluation of spatial forecast fields. The MODE tool is able to provide a large variety of diagnostic and meaningful information about the performance of forecasts of elements such as precipitation or convection.

Although generally designed for the evaluation of forecasts from mesoscale models, the MODE could easily be adapted to evaluate information from regional climate or chemistry models. New capabilities that are in progress include methods for evaluating ensemble forecasts and incorporation of the time dimension (e.g., through creation and evaluation of 3-dimensional objects).

ACKNOWLEDGMENTS

NCAR is sponsored by the National Science Foundation. This work was sponsored in part by NCAR’s project on Short-Term Explicit Prediction (STEP), which is sponsored by NSF; and by the Developmental Testbed Center, which is sponsored by NOAA and the Air Force Weather Agency. We thank Dr. Mike Baldwin for provid-

ing the SPC spring program WRF model output and Stage II precipitation observations.

REFERENCES

Baldwin, M. E., and S. Lakshmivarahan, 2003: Development of an events-oriented verification system using data mining and image processing algorithms. Preprints, Third Conf. on Artificial Intelligence Applications to Environmental Science, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, 4.6.

Bullock, R., B. Brown, C. Davis, and E. Gilleland, 2007: Object-based verification of precipitation forecasts. In preparation (available from R. Bullock at bullock@ucar.edu).

Casati B., G. Ross, and D.B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.,* **11**, 141-154.

Davis, C.A., B.G. Brown, and R.G. Bullock, 2006: Object-based verification of precipitation forecasts, Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772-1784Ebert, E.E., and J.L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrology*, **239**,179--202.

Gilleland E, T.C.M. Lee, J. Halley Gotway, R.G. Bullock, and B.G. Brown, 2007: Quantifying differences between defined binary image objects of forecast and analysis of quantitative precipitation for forecast verification. Submitted to *Mon. Wea. Rev.*

Lin, Y. and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. *Preprints, 19th Conf. on Hydrology*, American Meteorological Society, San Diego, CA, 9-13 January 2005, Paper 1.2.

Marzban, C., and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, 824-838.

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..

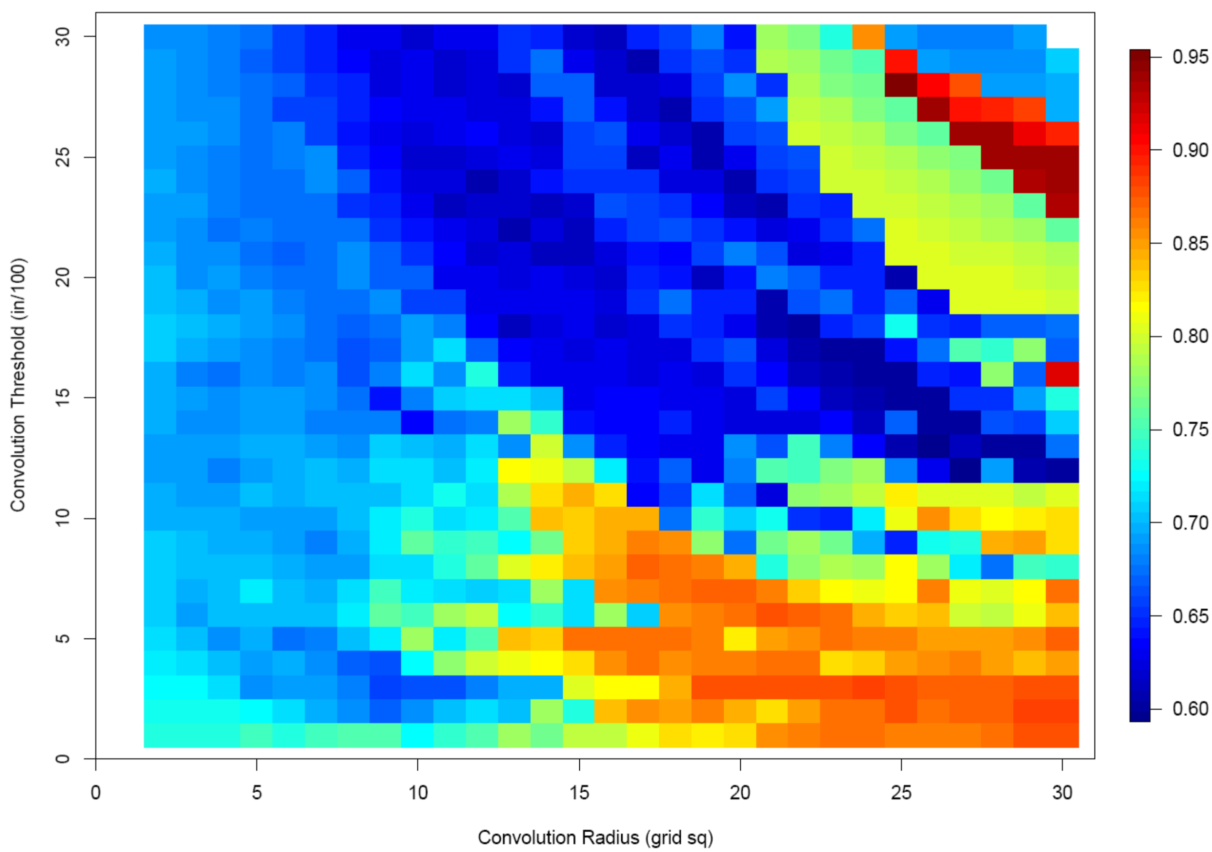

FIGURE 6. Example “quilt” plot showing a statistic that characterizes the overall matching strength of a forecast at different scales, as a function of the convolution radius (R) and the threshold (T). The results in the figure are based on nine WRF-ARW cases from the SPC spring program in 2005.