11B.4 OBJECTIVE VERIFICATION OF HIGH-RESOLUTION WRF FORECASTS DURING 2005 NSSL/SPC SPRING PROGRAM

Michael E. Baldwin* and Kimberly L. Elmore
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK.
Also affiliated with NOAA/NSSL

1. INTRODUCTION

The development of meaningful, objective methods for verifying gridded, high-resolution forecasts containing realistic detail that satisfy the needs of a diverse user community continues to be an elusive problem. Recent research along these lines has focussed primarily on ways of evaluating the "realism" of forecasts, following suggestions made by Anthes (1983). One such general method involves the comparison of measures related to the structure of detailed fields, such as Fourier power spectra (e.g., Skamarock 2004; Harris et al. 2001; Zepeda-Arce et al. 2000). Another potentially useful suggestion involves the comparison of characteristics of specific meteorological phenomena, often called the "object-oriented" approach (e.g. Ebert and McBride 2000; Nachamkin 2004; Case et al. 2004). Examination of the spatial distribution of errors along with their significance has also been considered (Elmore et al. 2005).

In this work, we examine the utility of verification methods that provide information on the quality of high-resolution numerical forecasts. Forecast output from several versions of the WRF model were evaluated during the 2005 NSSL/SPC Spring Program. These include a ~4km grid-spacing version of the WRF-NMM run at NCEP, a ~4km version of WRF-ARW run at NCAR, and a ~2km version of WRF-ARW run by CAPS at the Pittsburgh Supercomputing Center. These forecasts were evaluated subjectively in real-time by teams of operational forecasters and researchers (Weiss et al. 2005). Results from the subjective evaluation will be reported by Kain et al. (2005) at this conference. In this work, several automated/objective verification techniques have been applied to these forecasts, in particular, predicted and observed radar reflectivity fields are compared. Object-oriented techniques have been used to compare the forecast and observed characteristics of a variety of rainfall systems. Other statistical techniques are used to measure biases in the forecast fields, structure, etc. This paper will report on ongoing research related to meaningful, objective verification of forecasts that contain realistic, high-resolution detail.

**Corresponding author address:* Michael E. Baldwin, CIMMS/OU, 1313 Halley Cir, Norman, OK, 73069
Email: mbaldwin@ou.edu

2. OBJECT-ORIENTED VERIFICATION FRAMEWORK

In general, the framework for object-oriented verification consists of three basic steps: object identification, characterization, and comparison. To complete the first step in this process, specific meteorological phenomena must be located and identified using weather-related information. The object-identification process could be performed manually (e.g., Smith and Mullen 1993), although such a process would usually involve considerable time and labor. Automated procedures for identifying meteorological objects are necessary in order to perform long-term verification studies and obtain comprehensive information on forecast performance. Criteria for object identification must be established and documented so that results can be duplicated by other researchers. Such criteria will vary depending upon the phenomena of interest. Results will also be sensitive to the spatial and temporal scales that the meteorological data can resolve, data analysis techniques, etc. Routines for identifying objects should not be a function of both the observed and predicted fields, otherwise different objects will be defined for different forecast systems, making comparative verification infeasible. Examples of automated object-identification procedures that have been established in previous work include agglomerative cluster analysis methods (Lakshmanan et al. 2003; Peak and Tag 1994), as well as thresholding-type methods of identifying sea breeze fronts (Case et al 2004), Mistral wind storms (Nachamkin 2004), and contiguous rain areas (Ebert and McBride 2000). In this work, the meteorological phenomena of interest are precipitation-producing systems. The automated procedure for identifying such precipitation systems was developed by Baldwin and Lakshmivarahan (2003). This automated procedure identifies rainfall systems as connected regions of precipitation through the use of image processing routines. The definition of connected regions is relaxed to allow systems that are situated very close together to be grouped as a single precipitation system. This procedure was used as the basis of an automated rainfall system classification procedure (Baldwin et al. 2005).

Once objects have been identified within the forecast and observed meteorological data, the characteristics of those objects must be extracted in order to provide a useful description of each object. Meteorological phenomena can be described by statistical character-

istics, properties, or *attributes*. Ideally, one would select a set of attributes that can describe the most important and discriminating aspects of an event in a concise fashion. For example, the i^{th} forecast event could be described by an *attribute vector* of *m* dimension $\mathbf{f_i} = (\alpha_i,\beta_i, \ldots, x_i, y_i)^T$ where x_i, y_i are the attributes associated with the spatial location of this event (perhaps latitude and longitude), and α_i, β_i,..., are attributes that could be associated with the size, intensity, orientation, continuity, intermittancy, etc., of the event. Of course, observed events must be described with the same set of attributes, for example, the vector describing the j^{th} observed event would contain $\mathbf{o_j} = (\alpha_j,\beta_j, \ldots, x_j, y_j)^T$.

In order to measure the accuracy of the forecast and quantify the agreement between forecast and observed events, the similarity between these vectors can be measured. There are numerous possible choices of similarity/dissimilarity measures, for example, the correlation coefficient between $\mathbf{f_i}$ and $\mathbf{o_j}$ is an example of a *similarity* measure, since the higher the correlation coefficient is, the more similar $\mathbf{f_i}$ and $\mathbf{o_j}$ are. Another possible candidate would be the generalized Euclidean distance, defined as $d_{ij} = (\mathbf{f_i} - \mathbf{o_j})^T \mathbf{A} (\mathbf{f_i} - \mathbf{o_j})$, a measure of dissimilarity. Here **A** is a weight matrix that could allow certain attributes to have greater weight than others, due to differences in units, relative importance, etc. Once the similarity measure has been chosen, overall summary verification scores or accuracy measures could then be obtained. This approach to verifying events would be analogous to the "measures-oriented" approach to verification (Brooks and Doswell 1996). A more comprehensive analysis of the verification information could also be obtained by examination of the joint distribution of forecast and observed events, dubbed the "distributions-oriented" approach by Brooks and Doswell (1996). This could be considered an extension to the verification framework outlined by Murphy and Winkler (1987). Considerable categorization or classification will likely be required to make this tractable, in order to limit the complexity and dimensionality of the joint distribution.

3. DATA

In order to prepare forecast and observed data for verification, a considerable amount of work is typically required. In this case, an archive was produced of forecasts from three versions of the WRF model, each run at grid-spacings less than 5km for purposes of evaluation during the 2005 NSSL/SPC Spring Program. This program attempted to assess the utility of near cloud-resolving numerical models to provide detailed and useful guidance to forecasters on environmental conditions and development and evolution of convective systems. The 2005 Spring Program lasted a period of seven weeks (18 Apr - 3 Jun 2005) where forecast guidance was evaluated by teams of forecasters, researchers, and visiting scientists at the Storm Prediction Center (SPC). High-resolution WRF model runs were provided to SPC by NCEP/EMC, NCAR, and OU/CAPS. Model configurations used for these runs are briefly described in Table 1. Each forecast was initiated at 00 UTC and extended at least 30h. In this work, the 18-30h forecasts were selected for verification.

	WRF-NMM4	WRF-ARW4	WRF-ARW2
Grid spacing	4.5km	4.0km	2.0km
Vertical levels	35	35	51
PBL.	MYJ	YSU	YSU
Microphysics	Ferrier	WSM6	WSM6
Rad. (SW/LW)	GFDL/GFDL	Dudhia/RRTM	Dudhia/RRTM
Init Cond.	NAM	NAM	NAM

Table 1: Model configurations used for the daily 0000 UTC WRF runs, evaluated during the 2005 NSSL/SPC Spring Program. (adapted from Weiss et al. 2005)

Direct output from the WRF model requires further post-processing in order to calculate various diagnostic variables (such as pressure level data, CAPE, reflectivity) and provide output in a format (GRIB) that is compatible with SPC visualization software. Modifications were made to the NCEP WRFPOST to calculate equivalent reflectivity from the predicted hydrometeor fields, and interpolate those reflectivity values vertically to constant height levels above ground. The reflectivity at 1km above ground was selected as the forecast field for verification purposes in this work.

Observed reflectivity images were archived from a 2km national mosaic of NEXRAD base reflectivity data (also known as "NIDS" data). These image files were converted to GRIB format to ensure compatibility with the object identification and characterization routines. The base reflectivity data was originally unfiltered, therefore ground clutter and other non-precipitation related signals were included in the data. In order to roughly filter out such signals, a threshold of 30 dBZ was applied to the fields prior to the object identification/analysis steps. To be consistent, the same 30 dBZ threshold was also applied to the forecast 1km reflectivity fields. Therefore, the precipitation systems in this work will primarily be convective in nature, with minimum precipitation rates of approximately 0.1 in/hr (assuming standard Z-R relationship). In order to apply the same object identification and characterization routines to all fields, all forecast and observed reflectivity fields were remapped to a common 4km grid (same grid as used in NCEP Stage IV precipitation analysis) using an area-average interpolation algorithm.

4. OBJECT IDENTIFICATION & CHARACTERIZATION

The object identification and characterization routines are provided by Baldwin et al. (2005). Since the previous work used hourly rainfall data instead of reflectivity, some modifications were required. In addition, other changes were made in order to speed up the processing. The following will briefly outline the object analysis process. In general, rainfall objects are defined as contiguous regions of precipitation, as indicated by the reflectivity fields. A simple threshold (30 dBZ) is applied to each field to convert each image into a binary image. A connected component labeling algorithm (Klette and Zamperoni 1996) is then applied to this binary image to provide a separate label for each individual contiguous region of reflectivity. The size of each contiguous region is then increased by 15%, which is intended to include a surrounding region of "trace" rainfall for each object. Since it is not unusual for find small gaps between nearby regions of reflectivity, those regions that are within ~20km of each other are connected together. Therefore, the definition of connected points within an object is relaxed so that points within ~20km are considered connected.

An example of this process is shown in figure 1. The original base reflectivity field is shown in Fig. 1a, while the same field after the application of the 30 dBZ threshold is shown next in Fig 1b. These figures show a fairly long line of convection located from West Virginia southward through northern Georgia. A broad region of lighter precipitation is located across the lower Great Lakes. The final result of the object identification procedure is displayed in figure 1c. Only those objects of size greater than 150 pixels (~3000 km^2) are included in this plot. The procedure has identified 5 objects, the largest of which represents the main convective line. The next largest object represents the broad region of lighter rain over Ohio and Michigan. A smaller object is found over Alabama which is separate and to the southwest of the large line. Other small objects are found in northern Illinois and near St. Louis, Missouri.

This example illustrates several issues related to object identification. First of all, some of the isolated, intense convective cells indicated by the radar (such as those found to the south of the main line in Georgia) are not included in any of the identified objects. These cells were identified as small objects, each less than 150 pixels in size. The 150 pixel threshold used here is fairly arbitrary, and could be relaxed to allow more objects to be identified and verified. For now, the verification will focus on the meso-β and larger sized objects identified by this procedure. There is no perfect object identification routine that will satisfy every user. For example, a human analyst may decide to split the main north-south oriented line into two objects. Another analyst may decide to include the convection in Alabama as part of the main line. Ideally, an object identification procedure could be tuned in order to produce a set of objects consistent with the scales and phenomena of interest to each user. Inclusion of additional meteorological variables in the identification procedure, such as CAPE, surface temperature and winds, etc., will likely improve the identification scheme.

Once objects have been identified, various attributes or characteristics are assigned to each of them. These include the date, time, location of the object's centroid, the mean, variance, and maximum value of the reflectivity within the object, and attributes related to the shape and spatial continuity of the reflectivity field. These shape-related attributes are a function of the autocorrelation, specifically, the difference between the maximum and minimum correlation value at specific lags (50, 100, 150 km) every 10 deg are computed. This provides similar information to the correlogram contour ellipticity that was used by Baldwin et al. (2005) to help discriminate between linear and cellular classes of convective systems, while requiring much less computation cost. Orientation angles (angle of the maximum autocorrelation at 50, 100, 150 km lags) are also used to characterize each object.

For example, the attributes assigned to the largest object (object id #37)in figure 1 will be described. The centroid location for this object is 35.7° N 83.15° W. This object is 4923 grid boxes in size, or approximately 100000 km^2. The mean reflectivity value is 32 dBZ, the variance of the reflectivity is 112 $(dBZ)^2$, and the maximum value for this object is 60 dBZ. The differences between the maximum and minimum correlation at 50, 100, and 150 km (respectively) are 0.23, 0.24, and 0.55. Typical values of this attribute for linear-type convective systems are ~0.4 and larger, therefore the largest correlation lag indicates that this object would fall into the linear classification. The orientation angles (angle of maximum auto-correlation at 50, 100, and 150 km lags) are nearly 60° for each lag. This angle is relative to E-W at the centroid location (positive counterclockwise) therefore 60° indicates a object oriented SSW-NNE.

5. OBJECT VERIFICATION

In order to compare predicted and observed objects, one must first establish a set of matching rules to associate predicted objects with corresponding observed objects. Figure 2 illustrates one way to accomplish this task. The forecast and observed objects identified at the same time are collected. First, dissimilarity/similarity measures between all observed and forecast objects are computed, the degree of dissimilarity between forecast and observed objects is indicated in figure 2 by the length of the lines connecting each pair of objects. An example of a dissimilarity measure is the generalized Euclidean distance, defined in section 2. Such a measure is similar to computing the distance between objects in multi-dimensional space, where the

various dimensions are the attribute variables associated with each object. By using this type of measure, differences between objects in any particular attribute will result in large dissimilarity, even if those objects are close together in physical space (x,y). This way, errors associated with all attributes (location, orientation, size, shape, intensity) can be combined and summarized with a single measure. Links are established between objects that are considered "close enough" together, as determined by some threshold measure of similarity (d_T). In this example, both the 02-F2 and O2-F3 pairs are close enough together to be considered matching. If a particular forecast object is found to not be within this threshold "distance" of any observed object, it would be considered a "false alarm". In this example, forecast object F1 is considered a false alarm. Similarly, if a particular observed object was found to not be within this threshold distance of any forecast object, it would be considered a "missed event". In this example, observed object O1 is considered a missed event.

There are many complicating factors involved in this process. The attribute variables have a wide range of values and different units. The distribution of values within a particular attribute may also be important. For example, a 5 mm difference in average precipitation between objects could be considered large, for example, if the observed object averaged 6.5 mm and the forecast object averaged 1.5 mm. On the other hand, a 5 mm difference could also be considered relatively small if the difference was taken between objects averaging 65 and 60 mm. Also, it is not clear how to weigh the different attributes in order to obtain meaningful dissimilarity measures. Results will likely be sensitive to the choice of similarity/dissimilarity measure. An added complication would be to consider time as a separate variable, rather than comparing forecast and observed objects occurring at the same time. This would allow forecast objects to match observed objects occurring at a later/earlier time. The goal is to produce a set of matching forecast and observed objects in a manner consistent with a subjective analyst. It is expected that a considerable amount of tuning will be required in order to achieve this goal. Another complicating factor is the likelihood of obtaining multiple matching objects for a given forecast or observed object. Analysis of the joint distribution of forecast and observed object attributes is fairly straightforward if there are simple matching pairs of observed and forecast objects. It is not clear how one should perform such an analysis if more than one matching object is found.

The current status and results from this ongoing research will be presented at the conference.

References

Anthes, R. A., 1983: Regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.*, **111**, 1306–1335.

Baldwin, M. E., and S. Lakshmivarahan, 2003: Development of an events-oriented verification system using data mining and image processing algorithms. Preprints, *3rd Conf. on Artificial Intelligence,* Long Beach, CA, Amer. Meteor. Soc., paper 4.6.

Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an Automated Classification Procedure for Rainfall Systems *Mon. Wea. Rev.*, **133**, 844-862.

Brooks, H.E. and C.A. Doswell III, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. Wea. Forecasting, **11**, 288-303.

Case, J. L., J. Manobianco, J. E. Lane, C. D. Immer, F. J. Merceret, 2004: An objective technique for verifying sea breezes in high-resolution numerical weather prediction models. *Wea. Forecasting* **19,** 690-705.

Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Elmore, K. L., M. E. Baldwin, and D. M. Schultz, 2005: Field significance revisited: Spatial bias errors in forecasts as applied to the Eta Model. *Mon. Wea. Rev.*, in press.

Harris, D., E. Foufoula-Georgiou, K.K. Droegemeier and J.J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydromet.*, **2**, 406-418.

Kain, J. S., S. J. Weiss, M. E. Baldwin, D. Bright, and J. J. Levit, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL Spring Program. *Preprints, 17th NWP Conf.*, AMS, Washington D.C., Aug 1-5 2005, paper 2A.5.

Klette R. and P. Zamperoni, 1996: *Handbook of image processing operators.* John Wiley and Sons, 397pp.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atm. Res.*, **67-68**, 367-380

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

Nachamkin, J. E., 2004: Mesoscale verification using meteorological composites. *Mon. Wea. Rev.,* **132**: 941-955.

Peak, J. and P. Tag: 1994, Segmentation of satellite weather imagery using hierarchical thresholding and neural networks. *J. Appl. Meteor.*, **33**, 605–616.

Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Mon. Wea. Rev.* **132**, 3019-3032.

Smith, B. B. and S. L. Mullen, 1993: An Evaluation of Sea Level Cyclone Forecasts Produced by NMC's Nested-Grid Model and Global Spectral Model. *Wea. Forecasting,* **8**, 37–56.

Weiss, S. J. Kain, J. Levit, M. Baldwin, D. Bright, G. Carbin, and J. Hart, 2005: NOAA Hazardous Weather Testbed. SPC/NSSL Spring Program 2005. Program overview and operations plan. available at http://www.spc.noaa.gov/exper/Spring_2005/2005_ops_plan.pdf

Zepeda-Arce, J., E. Foufoula-Georgiou, and K.K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res.*, **105** (D8), 10,129-10,146.

.

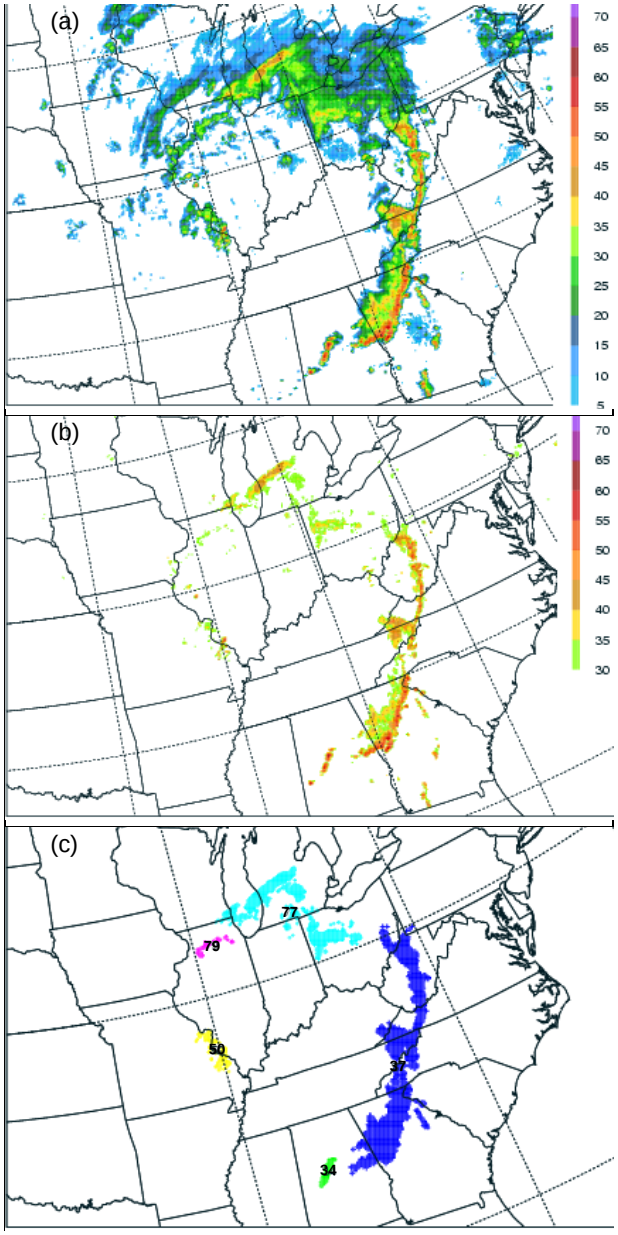

Figure 1: Illustration of the object identification process. Top panel (a) NEXRAD national mosaic base reflectivity (dBZ) valid 1800 UTC 22 Apr 2005. Middle panel (b) as in (a) except for only reflectivity greater than 30 dBZ. Bottom panel (c) is the result of the object identification process. Only objects with size greater than 150 pixels are shown. Each object is plotted with a different color, and the object ID number is plotted at each object's centroid location.

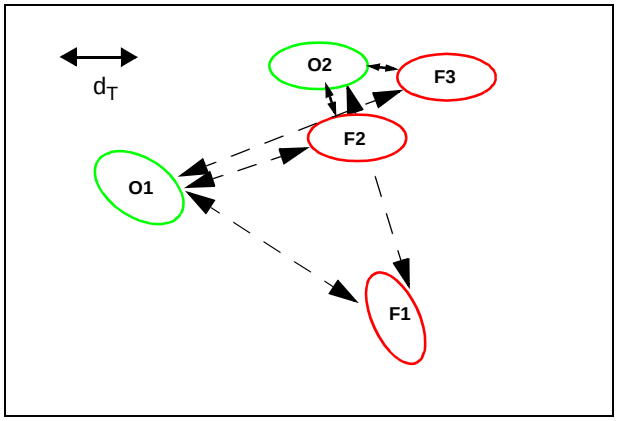

Figure 2: Hypothetical example illustrating the forecast/observed object matching process. Forecast objects are in red, observed objects are in green. The dissimilarity between each pair of observed and forecast objects is indicated by a line between the objects. The dissimilarity measure could be considered a distance in a multiple dimension space, where the dimensions are the attribute variables associated with each object. Dashed lines indicate the object pair is beyond the dissimilarity threshold d_T.