18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Verification of the time evolution of precipitation systems in numerical weather forecasts

Baldwin, M.E. [1]

[1] *Department of Earth and Atmospheric Sciences, School of Science,*
Purdue University, West Lafayette, Indiana, USA
Email: baldwin@purdue.edu

Abstract: Precipitating weather systems have life cycles involving formation, evolution, and decay. These systems can translate across spatial regions spanning 1000's of km. Climate scientists, numerical modelers, and operational hydro-meteorologists are especially interested in understanding how the characteristics of precipitation systems change over space and time. For example, a topic of current interest is whether global climate change can be observed in the changing characteristics of precipitating weather systems.

In addition, there are many outstanding questions regarding the ability of numerical weather prediction systems to predict realistic weather events that translate and evolve over time and space that could be addressed with appropriate forecast evaluation methods. For example, do state-of-the-art numerical weather prediction models accurately predict temporal changes in the morphological characteristics of convective precipitation systems? In order to answer these types of questions, automated techniques of identifying and tracking individual weather systems must be developed and applied to forecast and observed data, and methods of comparing forecasts and observations of time-varying features must be developed.

This paper will focus on the problem of developing a framework for the comparison of meteorological features that evolve with time. Since weather systems typically evolve and translate in space and time, explicit analysis of their characteristics requires a feature-based or "object-oriented" approach. Such an approach involves identification of the precipitating weather system of interest and measurement of appropriate characteristics of each system. Individual systems within the forecast and observed data must be tracked over time, which involves issues of determining the beginning and ending of specific events. Issues related to comparing predicted and observed weather events must also be addressed, as well as the problem of comparing forecast and observed events with unequal lifetimes.

Keywords: *Numerical weather prediction, forecast verification, image processing, quantitative precipitation forecasting*

1. INTRODUCTION

The development of meaningful, objective methods for verifying numerical predictions of spatial fields that contain realistic detail which also satisfy the needs of a diverse user community continues to be a difficult issue. Realistic spatial fields of meteorological variables such as precipitation and clouds contain a great deal of structure and variability across a wide range of spatial and temporal scales. Traditional measures of accuracy which directly compare predicted and observed values of these highly-variable fields tend to produce very large errors given even slightly imperfect forecasts. Since displacement errors in weather forecasts are to be expected, a forecast containing realistic structure will likely produce the so-called "double penalty" effect on traditional scores (e.g., heavy rain predicted where no rain was observed and no rain predicted where heavy rain was observed). Such scores typically do not provide very useful information regarding the performance and value of these kinds of forecasts (e.g., Davis et al. 2006; Keil and Craig 2007).

Recent research in this area has primarily focused on methods of evaluating the "realism" of forecasts, generally following suggestions made by Anthes (1983). A recent review of this type of work was presented by Casati et al. (2008). The newer methods can be generally classified into four categories: optical flow, neighborhood-based, scale-decomposition, and feature-based approaches. Optical flow methods (e.g., Keil and Craig 2007) involve distorting the spatial forecast field to closely match the observed field and determining errors in displacement, amplitude, and shape/structure as a result. Neighborhood-based methods (e.g., Atger 2001; Roberts and Lean 2008) relax the requirement of matching the exact spatial and/or temporal location of forecasts and observations, giving credit for forecasts that are "close" or nearby in space and time. Scale-decomposition methods (e.g., Briggs and Levine 1997; Zepeda-Arce et al. 2000; Casati et al. 2004) examine forecast errors as a function of spatial scale. Feature-based methods compare predicted and observed characteristics of specific meteorological phenomena, and are often called "object-oriented" approaches (e.g., Ebert and McBride 2000; Nachamkin 2004; Davis et al. 2006). For purposes of forecast verification, the terms *feature*, *event*, *object*, and *entity* have been used interchangeably. "Objects" are considered here in much the same way as they are in cluster analysis, which is the systematic approach of combining individual objects into groups based on their similarity. It is the general category of "object-oriented" methods that will be the focus of this paper.

One common characteristic that these recent methods generally share is that predicted and observed fields valid at the same single snapshot in time are compared in order to evaluate the performance of forecast systems. Feature-based verification methods identify "objects" in spatial fields (forecast and observed separately) typically by locating contiguous regions of variable values greater than a specified threshold (such as 1h precipitation greater than 5mm). Attributes for each object that describe relevant characteristics of each object, such as their location, intensity, and size are then determined. Where forecast and observed objects are simultaneously found close enough together to be considered "matching", attributes of forecast objects are compared to those from objects observed at the same snapshot in time. This allows for the analysis of the distribution of errors in specific aspects of the spatial forecasts, such as location errors. However, what is typically not considered by these methods is the time evolution of the predicted and observed objects.

There are many interesting questions related to the performance of weather prediction systems that contain realistic meteorological features that move and evolve over time that could be addressed with suitable verification techniques. For example, does a numerical prediction system have a bias in forecasts of the lifetime of convective systems? Do numerical models predict the average speed of certain classes of precipitating weather systems better than other classes? In order to answer these kinds of questions, automated techniques of time tracking of weather systems must be applied to forecast and observed fields, and methods of verifying forecasts of time-varying features must be developed.

There are several methods available to track features in meteorological data over time. These have been developed for short-term forecasting of convective storms ("nowcasting") based on radar data (e.g., Dixon and Wiener 1993; Johnson et al. 1998; Wilson et al. 1998) as well as tracking of cyclones in climate analyses and numerical weather prediction output (e.g., Hodges 1995). While the development of robust methods of tracking features in meteorological data over time is a topic of great interest, it is also important to develop a general verification framework that can utilize information obtained from any feature tracking algorithm. This paper will focus on the problem of developing a framework for the comparison of feature attributes that evolve over time.

2. VERIFICATION FRAMEWORK

It is upon the well-developed framework for verification of Murphy and Winkler (1987) that any new verification technique should be built. This general framework involves analysis of the joint distribution of forecasts and observations. While Murphy and Winkler (1987) did not explicitly consider the verification of meteorological features that evolve with time, their framework does provide the foundation for the verification of time-varying features. A more specific framework for feature-tracking verification involves four basic steps: feature identification, characterization, tracking, and comparison. In the first step of this process, specific meteorological features must be located and identified using multivariate data. Criteria for object identification will vary depending upon the phenomena of interest. Routines for identifying objects should not be a function of both the observed *and* predicted data, for example, the definition of an observed object should be independent of the forecast data. Otherwise, the same set of observed data will result in different observed objects being defined for different forecast systems, making comparative verification infeasible. In this work, the meteorological phenomena of interest are precipitating weather systems. The automated procedure for identifying such precipitation systems is based on the contiguous rain area method developed by Ebert and McBride (2000), where connected regions of precipitation greater than some threshold are identified as separate features (Baldwin et al. 2005).

After the features have been identified within the forecast and observed fields, the characteristics of those objects must be extracted in order to provide a useful description of each object. A set of attributes that can describe the most important and discriminating aspects of an object at a particular instant in time must be collected. For example, the i^{th} object would be described by a vector of dimension *m* that contains attributes describing that object. These attributes could be associated with the spatial location (perhaps latitude and longitude) and other meteorological and morphological characteristics (size, amplitude, orientation, structure, continuity) relevant to the phenomena of interest. Figure 1 shows an example of the results of simple object identification and characterization procedures using a numerical prediction of radar reflectivity. The object identification procedure of Baldwin et al. (2005) is used with a 40 dBZ radar reflectivity threshold to produce the numbered objects in the right-hand panel of Figure 1. Contiguous regions of reflectivity greater than 40 dBZ are labeled. Attributes related to the location, size, and intensity of these precipitation objects are also collected. In this example, object #63 is located at 36.02° N, 263.61°E, is 612 km^2 in size, a mean value of reflectivity of 48 dBZ and a maximum value of 56 dBZ. Figure 2 shows the results from a similar analysis using the observed radar reflectivity data valid at the same time (+/- 2 minutes). In the observed data, object #8 is located at 35.75°N, 263.66°E, is 702 km^2 in size, a mean value of reflectivity of 47 dBZ, and a max reflectivity value of 58 dBZ. In this example, the forecast generally produces a spatial field that contains realistic-looking structure, except with some displacement error. This could also be a displacement in the temporal evolution of the precipitation systems, since the forecast generally initiated precipitation 1-2h later than what was observed in this particular example.

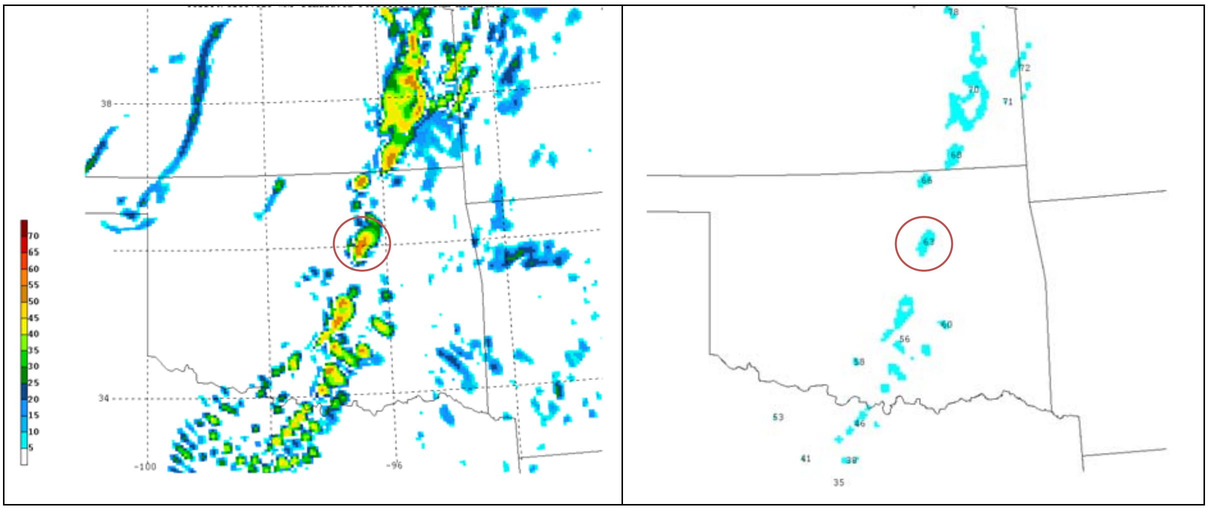

Figure 1. Example of object identification from a single snapshot in time. Left panel, 25h forecast of simulated radar reflecitivity (1km above ground level) from a 4.25km WRF model forecast valid 0100 UTC 06 Nov 2008 using the Purdue WRF configuration. Right panel, objects identified using modified Baldwin et al. (2005) algorithm and 40 dBZ threshold. Object #63 is circled in both panels. (NCEP NAM initial and boundary conditions, 45 vertical levels, Purdue Lin microphysics, YSU turbulence, RRTM/Dudhia radiation, NOAH land-surface model 25s timestep, see http://wxp.eas.purdue.edu/wrfdata/)

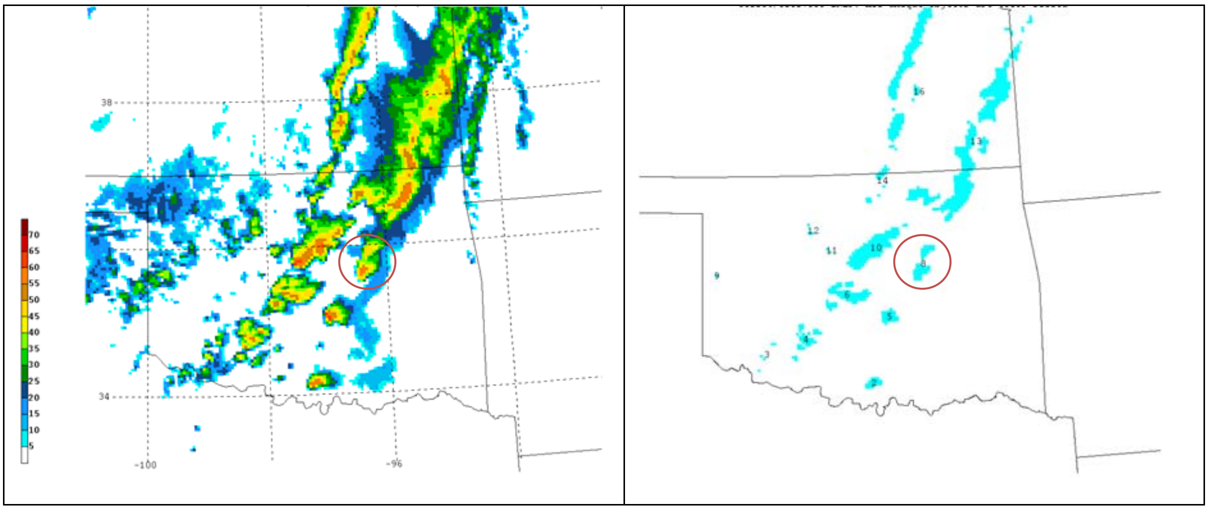

Figure 2. Example of object identification from observed radar data. Left panel, radar reflectivity (lowest elevation angle) remapped to the identical 4.25km grid as in the WRF forecast data (valid 0101 UTC 06 Nov 2008). Right panel, objects identified using modified Baldwin et al. (2005) algorithm and 40 dBZ threshold. Object #8 is circled in both panels.

3. TIME EVOLUTION

A feature tracking algorithm typically takes objects that have been identified and characterized separately at discrete snapshots in time and "connects" similar objects from one time to the next. This is commonly referred to as the *data association* problem. In general, this procedure is similar to that followed in cluster analysis algorithms, where individual objects are connected based upon their similarity. One can choose from a wide variety of measures of similarity/dissimilarity in order to perform this task (e.g., Romesburg 1984). Measures of similarity between objects establish the degree of "closeness" between two objects. At each snapshot in time, each object is described by a vector of size *m* containing the attribute values corresponding to that object. Conceptually, one can measure the degree of difference between two objects via some measure of the size (or *norm*) of the vector that is created by subtracting the two attribute vectors associated with those two objects. For example, the so-called *generalized Euclidean distance* (Carroll and Wish 1974)

$$d_{Aij} = [(obj_i - obj_j)^T \mathbf{A} (obj_i - obj_j)]^{1/2} \qquad (1)$$

is the square root of the weighted inner product of the difference between attribute vectors associated with two objects (obj_i and obj_j). For this distance measure to meet the qualifications of a distance metric, the weight matrix **A** must be a positive-definite symmetric matrix. The use of the weight matrix allows one to weigh certain attributes more heavily than others, or to account for differences in the units between attributes. When the identity matrix is used for the weight matrix, the familiar Euclidean distance is obtained (which is used in this work).

There are numerous feature-tracking procedures currently in use for tracking convective storms in radar and satellite data (e.g., Lakshmanan et al. 2003; Morel and Senesi 2002; Storlie et al. 2009) as well as tropical and extra-tropical cyclones in global weather prediction and climate analysis products (e.g., Bengtsson et al. 2006; Froude et al. 2007). The choice of tracking algorithm will depend upon the phenomena of interest, and different algorithms will produce different results. For example, Joe et al. (2004) compared several nowcasting algorithms that were used as part of the Sydney 2000 Forecast Demonstration Project (FDP) to evaluate the utility of these routines. The various algorithms produced vastly different numbers of storm tracks depending on the threshold used to identify convective cells in the radar data as well as radar scan times. Ebert et al. (2004) verified the performance of the short-term forecasts from these FDP nowcast algorithms and decided to evaluate each algorithm individually, not comparing results from different algorithms due to the numerous differences in their operation. Regardless of which time-tracking algorithm is selected, the results of that algorithm should provide summary information about relevant attributes for each feature across the lifetime of that feature. We can assume that the i^{th} feature would be described by a vector of dimension $m \times n$ that contains *m* attributes describing that object over *n* separate snapshots in time. The first index of the time attribute indicates the initial time of the feature, and the n^{th} index of the time attribute indicates the last time that the feature could be detected in the data. Therefore, each of the *m* attributes will contain *n* specific values indicating how those attributes changed during the lifetime of the feature, whether those attributes are related to the spatial location or other meteorological and morphological characteristics relevant to the phenomena of interest.

As an example, Figure 3 displays the results of a simple object tracking algorithm (Baldwin and Carley 2009). With this procedure, the attributes of each object identified at time = t are compared to every object identified at t = t+ Δt (where Δt = 10 min in the forecast data and ~5 min in the observed data) using a Euclidean distance metric. The attribute vector contains information regarding location, size, and intensity characteristics. Since these attributes posses different units and ranges of values, they must be standardized to allow vector calculations such as a Euclidean distance measure. In this example, the standardization was a simple normalization where the minimum value of each attribute was subtracted and then the result was divided by the range (determined by a small sample data set consisting of a single 36h period), such that each attribute varies between 0.0 and 1.0. The location attributes were weighted such that a 125km distance would relate to a 1.0 difference in normalized attribute space. The resulting "distance" between objects considers not only their spatial separation, but also differences between intensity and size characteristics. Therefore, objects are considered similar if all of their attributes are similar. The feature tracking procedure is often complicated by the fact that meteorological features (such as convective storms) can appear or disappear at any particular instant in time ("birth" and "death"). An object at time = t is connected or associated with an object at time = t + Δt if the Euclidean distance between them was less than a specified threshold. In this example, the threshold was determined by analyzing the cumulative distribution of between-object distances computed for simultaneous images from a single sample dataset (36h period). This threshold is one where greater than 99% of simultaneous objects would be considered separate. In this example, the example forecast object #63 from Figure 1 could be tracked for ~2h, and had a track of length = 150 km. The example observed object #8 from Figure 2 could be tracked for 96 min and had a track length of 90 km. The relevant attributes for each feature can be tracked and recorded over time.

In this example, comparing the distribution of forecast and observed tracks over the 36h period of this forecast shows that the forecast produced a reasonably realistic prediction of the overall time evolution of many of the individual convective features that were observed in the radar data. The forecast missed several of the features in the west-central portion of Oklahoma, mainly due to the fact that it initiated storms later in the day than what was observed for this particular case.

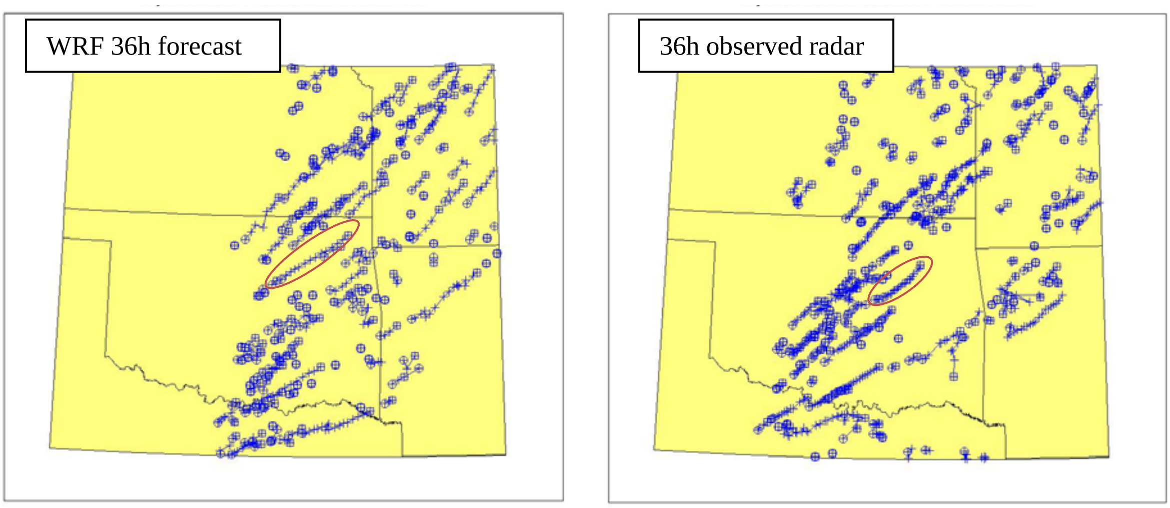

Figure 3. Results of a simple object tracking algorithm (Baldwin and Carley 2009). The Euclidean distance between objects at successive time intervals are computed using normalized location, size, and intensity attributes. Objects between successive output times (forecast = 10 min, observed ~ 5 min) that are closer than a given threshold are connected. Circles indicate the initial location of a feature and squares indicate the final location of a feature. The time-track for the example object #63 from Figure 1 is highlighted in the left hand panel, and the track for the example object #8 from Figure 2 is highlighted in the right hand panel.

4. VERIFICATION OF FEATURES THAT EVOLVE OVER TIME

To this point, the steps in feature tracking verification have involved separate analysis of the numerical forecast information and observational data. Those wishing to follow the general verification framework of Murphy and Winkler (1987) should analyze the distributions of forecast and observed attributes (e.g., Davis et al. 2006) along with the joint distribution of forecast and observed attributes of features that can be paired in time and space. In order to analyze the joint distribution of forecast and observed attributes, one must pair up "matching" features in the predicted and observed datasets. This comparison step shares many common issues with the data association/time tracking step discussed previously. For example, we may want to measure the multi-variate Euclidean distance between the i^{th} forecast attribute vector (f_i) and the j^{th} observed attribute vector (o_j). Euclidean distance between these two features would be defined as $d_{ij} = [(f_i - o_j)^T (f_i - o_j)]^{1/2}$. Once the similarity measure has been chosen, overall summary verification scores or accuracy measures could then be obtained. The Euclidean distance between a forecast feature and all observed features can be calculated, those forecast/observed feature that are closer in multi-variate distance than some specified threshold can be paired. Attributes of paired sets of forecast/observed features can be directly compared using the joint distribution approach or by computing a series of "measures-oriented" statistics. Forecast features that remain unpaired with any observed feature can then be considered "false alarms" and observed features that remain unpaired with any forecast feature can be considered "missed events".

This approach relies on the forecast and observed feature attribute vectors being of identical size. This is not a problem if one is comparing forecast and observed features obtained from snapshots (not considering time evolution, attribute vector length = m). However, if one is comparing attribute vectors that vary with time, the length of attribute vectors will be $m \times n$, where n will be equal to the age of the feature in units of time-steps. Forecast and observed attribute vectors cannot be compared if they are of different sizes. The challenge is to develop attribute vectors of a *fixed* length that contain enough information to describe the complete time evolution of objects of interest. For comparison purposes, one could compute a time-evolution attribute vector that contains a set of summary statistics describing how the location, intensity, morphological, and meteorological attributes vary with time. For example, the attribute vector that is used in this preliminary work consists of the time, size, location, and intensity-related attributes at the beginning and ending of the track of the system, the average size, location, and intensity, and the maximum size/intensity attributes.

5. CONCLUSIONS

There are mainly outstanding issues that must be resolved before a robust verification system can be implemented that provides useful information regarding the time evolution of weather systems. For instance, the object identification procedure is quite sensitive to the choice of threshold, and has a great impact on the resulting system size. One may be interested in measuring the performance of a forecast system in predicting fine-scale meteorological features, in this case, a relatively large reflectivity threshold (such as the one used here) will be appropriate. However, in many cases, it will likely prove very difficult for numerical models to match similar features in the observed data for small-scale features. A lower threshold will cluster the small-scale features that are located within a larger-scale precipitation system into a "composite" system, which may be easier to track and match in the forecast data. The dimensionality of the verification problem increases tremendously with the addition of time-evolving information. Regarding the Euclidean distance calculations, ideally one would prefer to use a long-term climatology to normalize the object attributes. This will certainly require extensive analysis in order to obtain such a feature-based climatology. This ongoing work is in its preliminary stage, and examples of this numerical forecast verification approach using high-resolution numerical forecasts and observed radar data will be presented at the conference.

ACKNOWLEDGMENTS

Jacob Carley assisted in the generation of figures that were used in this paper. Two anonymous reviewers provided very helpful comments and suggestions that resulted in substantial improvements to the paper.

REFERENCES

Anthes, R.A. (1983), Regional models of the atmosphere in middle latitudes. *Monthly Weather Review*, 111, 1306–1335.

Atger, F. (2001), Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes in Geophysics*, 8, 401–417.

Baldwin, M.E., Kain, J.S., and Lakshmivarahan, S. (2005), Development of an Automated Classification Procedure for Rainfall Systems. Monthly Weather Review, 133, 844–862.

Baldwin, M. E., and Carley, J. (2009), Tracking precipitating weather systems in forecast and observed data. *Presentation for 7th Conference on Artificial Intelligence and its Application to the Environmental Sciences*, AMS Annual Meeting, Phoenix, AZ, January 11-15, 2009, J7.3.

Bengtsson, L., Hodges, K.I., and Roeckner, E. (2006), Storm tracks and climate change. *Journal of Climate,* 19, 3518–3543.

Briggs, W.M., Levine, R.A. (1997), Wavelets and field forecast verification. *Monthly Weather Review*, 125: 1329–1341.

Casati, B., Ross, G., Stephenson, D.B. (2004), A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications,* 11, 141–154.

Casati, B., Wilson, L.J., Stephenson, D.B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E.E., Brown, B.G., and Mason, S. (2008), Forecast verification: current status and future directions. *Meteorological Applications*, 15, 3-18.

Carroll, J. D. and Wish, M. (1974), Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (eds.), Contemporary Developments in Mathematical Psychology. (Vol. II.: Measurement, Psychophysics and Neural Information Processing). W. H. Freeman, 57-105.

Davis, C., Brown, B., and Bullock, R. (2006), Object-based verification of precipitation forecasts. Part I: methodology and application to Mesoscale Rain Areas. *Monthly Weather Review*, 134, 1772–1784.

Dixon, M., and Wiener, G. (1993), TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol*., 10, 785–797.

Ebert, E.E., McBride, J.L. (2000), Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, 239, 179–202.

Ebert, E.E., Wilson. L.J., Brown, B.G., Nurmi, P., Brooks, H.E., Bally, J., and Jaeneke, M. (2004), Verification of nowcasts from the WWRP Sydney 2000 Forecast Demonstration Project. *Weather and Forecasting*, 19, 73-96.

Froude, L.S.R., Bengtsson, L., and Hodges, K.I. (2007), The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Monthly Weather Review*, 135, 315-333.

Joe, P., Burgess, D., Potts, R., Keenan, T., Stumpf, G., and Treloar, A. (2004), The S2K Severe Weather Detection Algorithms and Their Performance. *Weather and Forecasting*, 19, 43–63.

Johnson, J. T., MacKeen, P.M., Witt, A., Mitchell, E.D., Stumpf, G.J., Eilts, M.D., and Thomas, K.W. (1998), The Storm Cell Identification and Tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting,* 13, 263–276.

Hodges, K. I., (1995), Feature tracking on the unit sphere. *Monthly Weather Review*, 127, 3458–3465.

Keil, C., and Craig, C.G. (2007), A Displacement-Based Error Measure Applied in a Regional Ensemble Forecasting System. *Monthly Weather Review*, 135, 3248–3259.

Lakshmanan, V., Rabin, R., and DeBrunner, V. (2003), Multiscale storm identification and forecast. *Journal of Atmospheric Research*, 67–68, 367–380.

Morel C. and Senesi, S. (2002), A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology. *Quarterly Journal of the Royal Meteorological Society*, 128, 1953-1971.

Murphy, A.H. and Winkler, R.L. (1987), A general framework for forecast verification. *Monthly Weather Review*, 115, 1330-1338.

Nachamkin, J.E. (2004), Mesoscale verification using meteorological composites. *Monthly Weather Review,* 132, 941–955.

Roberts, N.M. and Lean, H.W. (2007), Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* 136, 78–97.

Romesburg, C. H., 1984: Cluster Analysis for Researchers. Life Time Learning, 334 pp.

Skamarock, W.C. and Klemp, J.B. (2008), A Time-Split Nonhydrostatic Atmospheric Model for Weather Research and Forecasting Applications. *J. Computational Physics*, 227, 3465-3485.

Storlie, C.B., Lee, T.C.M., Hannig J., and Nychka, D. (2009): Tracking of multiple merging and splitting targets: A statistical perspective, *Statistica Sinica*, 19, 1-52.

Wilson, J. W., Crook, N.A., Mueller, C.K., Sun, J., and Dixon, M. (1998), Nowcasting thunderstorms: A status report. *Bulletin of the American Meteorological Society*, 79, 2079–2099.

Zepeda-Arce, J., Foufoula-Georgiou, E., Droegemeier, K.K. (2000), Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *Journal of Geophysical Research*, 105(D8), 10129–10146.