8.3 VERIFICATION OF MESOSCALE FEATURES IN NWP MODELS

Michael E. Baldwin*[1,2], S. Lakshmivarahan[3], John S. Kain[2]

[1] NOAA/NWS/SPC Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK

[2] NOAA/NSSL Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK

[3] School of Computer Science, University of Oklahoma, Norman, OK

1. INTRODUCTION

The primary goal of forecast verification is to answer the question: "how good is this forecast?" or "how good is this set of forecasts?" As is the case for most deceptively simple questions, a simple answer is usually not sufficient. In particular, some of the practical difficulties associated with verifying forecasts from meso- or smaller-scale models have not yet been satisfactorily resolved. As computing power increases, operational weather forecasting centers obtain the capability to run numerical models that contain increasingly higher resolution. Since the amplitude of forecast features (e.g., precipitation maxima) from these models tends to increase as the horizontal grid spacing decreases, relatively small errors in space can cause very large differences between forecast and observed values at a specific location. As a result, statistical measures of performance obtained by traditional verification approaches will look poor when forecast and observed fields containing small-scale, high-amplitude features are compared. This is perhaps best illustrated by an example.

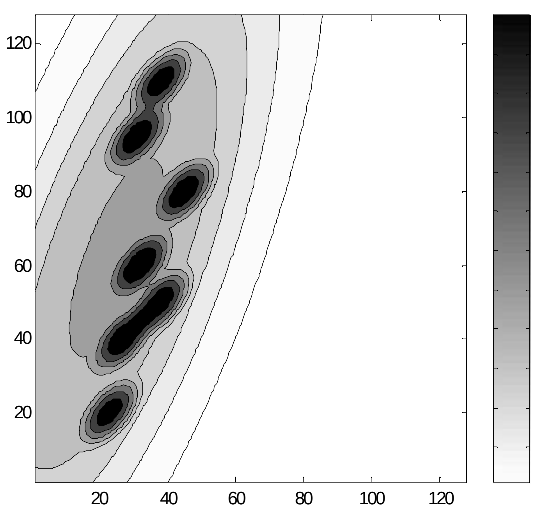

Figure 1: Simulated observed precipitation field.

Corresponding author address: Michael E. Baldwin, Storm Prediction Center, 1313 Halley Cir, Norman, OK, 73069 email: <Mike.Baldwin@noaa.gov>

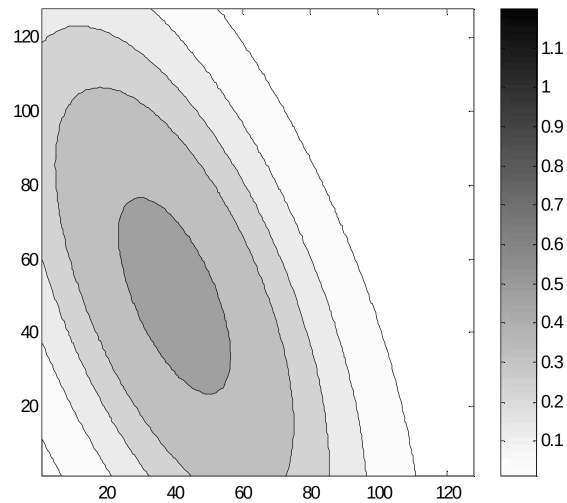

Figure 2: Simulated precipitation forecast #1.

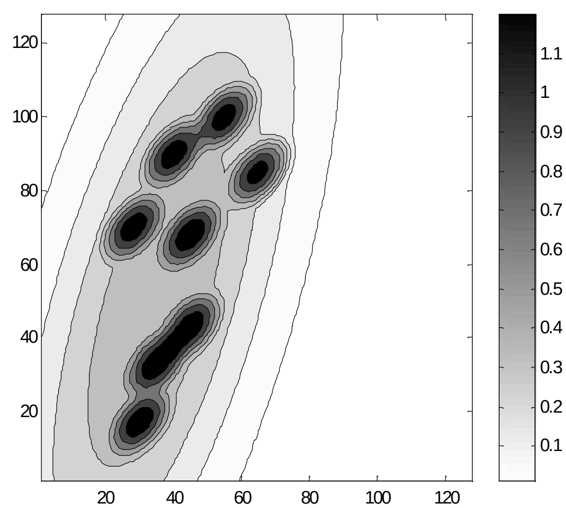

Figure 3: Simulated precipitation forecast #2.

2. THEORETICAL EXAMPLE

In this example, *simulated* precipitation fields were generated using an elliptical shape function (Williamson, 1981). In the "observed" field (Fig. 1), a relatively large scale ellipse is found containing several smaller-scale, higher-amplitude ellipses embedded within it. The domain consists of 128 x 128 grid points. For the sake of providing dimensions to the problem, if we assume the grid spacing is 5km, the large-scale ellipse is approximately 1000km long and 300km wide while the smaller-scale ellipses are approximately 100km long and 50km wide. Simulated forecast #1 (Fig. 2) consists of a single larger-scale ellipse, whose center is displaced compared to the observed larger-scale ellipse but with a similar amplitude. The orientation of the ellipse is also in error, and the forecast ellipse is also wider than the observed. Simulated forecast #2 (Fig. 3) contains features that are shaped similarly (both larger and smaller-scale) to the observed field. The entire area is displaced to the "southeast" compared to the observed field and the amplitude of the larger-scale ellipse is slightly less than the observed. The randomly configured smaller-scale ellipses are positioned differently relative to the center of mass of the larger-scale feature than in the observed field. Subjective visual inspection of these two forecasts indicate that forecast #2 is more *realistic* than forecast #1. This may translate into more *valuable* forecast information to a end-user or decision maker who is sensitive, for example, to the occurance/non-occurance of the smaller-scale, higher-amplitude features found in the observed field. On the other hand, other users of the forecast information may find that the smoother forecast is of greater value for their particular situation. We will now attempt to determine whether or not various methods of verifying these forecasts will uncover the value of this particular forecast situation.

2.1 Measures-oriented results

Table 1 displays the results of several traditional verification measures applied to these two forecast fields. The two forecasts have the same bias (ratio of average forecast to average observation) which indicates that neither forecast is suffering from a large bias error. In terms of the mean absolute error and root mean square (RMS) error, forecast #1 produced lower scores, which are preferred. For the threat score and equitable threat score (Mesinger, 1996), using the 0.45 threshold, forecast #1 produces higher scores, which are preferred. In this case, the various traditional verification meausres shown here indicate that the more detailed forecast #2 is of *lesser* quality than the smoother forecast #1. It seems clear that answering the question of "how good is this forecast?" with only a few scalar measures is unsatisfactory.

Verification measure	Forecast #1	Forecast #2
Mean absolute error	0.157	0.159
RMS error	0.254	0.309
Bias	0.98	0.98
Threat score	0.214	0.161
Equitable threat score	0.170	0.102

Table 1: Results of traditional verification measures for simulated precipitation fields

2.2 Distributions-oriented results

Conclusions regarding the absolute or relative performance of forecast systems are often made based upon a few measures of performance (e.g., Mesinger, 1996). This type of scalar analysis of verification information is defined by Brooks and Doswell (1996) as a *measures-oriented* approach to verification. An alternate and more complete approach to verification involves the analysis of the joint distribution of forecast and observations (Murphy and Winkler, 1987), dubbed by Brooks and Doswell (1996) as the *distributions-oriented* approach. Although analyzing and explaining the results of a more complex verification requires a great deal of effort, the time and effort required to perform a distributions-oriented verification should be considered more than just a luxury. Previous research (e.g., Brooks and Doswell, 1996) highlights the danger of ignoring the complexity and dimensionality of verification and touts the advantages of a more complete analysis of the relationship between forecasts and observations. Here we provide a brief analysis of the previous example using the distributions-oriented approach through examination of scatter plots (Figures 4 and 5).

In Figure 4 the joint distribution of observed and and forecast values for forecast #1 is plotted in a scatter diagram while Figure 5 shows the same information for the more detailed forecast #2. Within the distributions-oriented approach to verification, Murphy (1993) defines several specific aspects of forecast quality, such as bias, accuracy, association, etc. A complete analysis of the forecast verification information provided by the distributions-oriented approach will involve examination of all of the various aspects. We have already found that the bias of each forecast, defined as the ratio of the average forecast to the average observation, is equal to 0.98. Forecast accuracy, defined as the average degree of correspondence between individual forecasts and observations, is typically measured by scores such as the mean absolute error or root mean square error (Table 1). According to the measures found in Table 1, forecast #1 is more accurate than forecast #2. Association, defined as the overall strength of the linear relationship between forecasts and observations, is typically measured by the correlation coefficient. In this case, forecast #1 has a correlation coefficient of 0.486 while forecast #2 has a correlation of 0.429. Therefore, a brief examination of the distributions-oriented verification information shows that for some (but not all) aspects of forecast quality, forecast #1 is preferred over the more detailed forecast #2.

3. VERIFICATION OF FORECAST REALISM

In previous hypothetical example, both traditional measures-oriented and distributions-oriented approaches to verification represented a forecast system that contained *realistic* small-scale, high-amplitude features as lesser quality when compared to one that did not. Despite the potential for large errors at particular points, predicted fields that contain realistic spatial structures, scales, and amplitudes may be of considerable value to certain users (e.g., mesoscale forecasters). Consequently, from the point of view of these users, the value of the more detailed forecast has not be properly determined via these methods of analyzing forecast quality. Anthes (1983) recognized this problem early on and suggested an alternate type of verification for the evaluation of meso- and smaller-scale prediction models, that is, the determination of the "realism" of a forecast.

One technique suggested by Anthes (1983) to evaluate the realism of a forecast involves the examination of characteristics of significant phenomena, such as the central pressure of cyclones, or maximum wind speeds of thunderstorms. Along these lines, Williamson (1981) presented a method of pattern recognition to objectively identify geopotential height systems in a constant pressure-surface field. An empirical function was fit to the field that represents a high or low center, elliptically shaped with amplitude, position, and shape parameters. The parameters defining the function were determined by minimization, and good first guesses were required. A major issue regarding this technique is that it relies on an empirical function to fit shapes in the forecast and observed fields. This has the advantage of being able to explicitly define the attributes of the phenomena of interest. However, this also has the disadvantage of trying to fit possibly complex natural patterns by an empirical shape function. Ebert and McBride (2000) also present methods of verifying characteristics of "phenomena" (contiguous rainfall areas). When there is some overlap between observed and forecast precipitation areas, Ebert and McBride (2000) decompose the forecast error into components due to displacement, amplitude, and "shape" errors. While information on the differences between forecast and observed spatial structure is certainly useful, it is unrelated to the types of meteorological phenomena associated with the forecast and observed areas of rainfall, whether these were in agreement, etc. Again, returning to our simulated example, both forecast #2 and observed rainfall fields contained the same scales of elliptically-shaped rainfall, but distributed differently in space. In this case, the diagnosed "shape" errors would be large, even though the size and shape of the smaller-scale rainfall predicted features were nearly identical to those observed.

Anthes (1983) also suggested comparing the

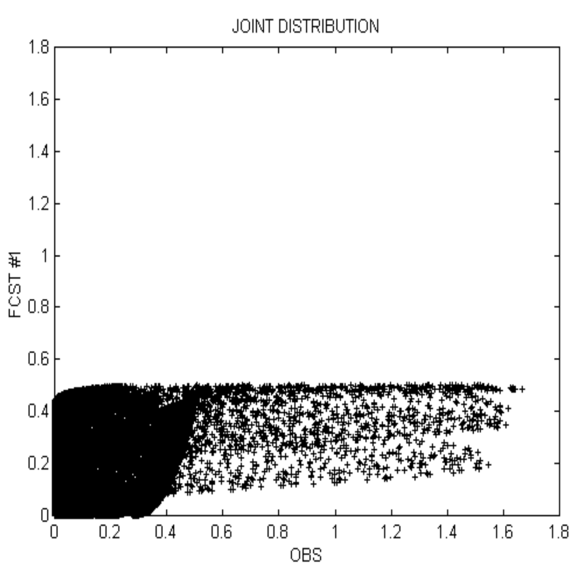

Figure 4: Scatter diagram of the joint distribution of observed and forecast values for forecast #1

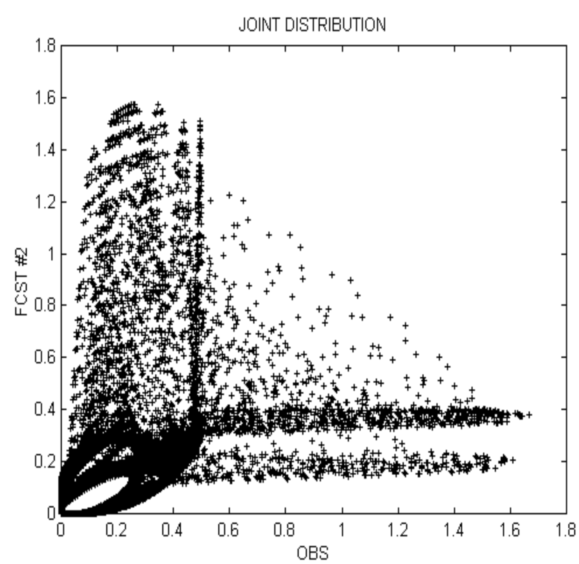

Figure 5: As in Figure 4 except for forecast #2

observed and forecast spectra of certain fields. Zepeda-Arce et al (2000) provide a recent example, using wavelet transforms to compute the spatial variation of the rainfall field as a function of horizontal scale. Examination of how the variance of the spatial fluctuations change as a function of scale for the observed and forecast fields shows how well the forecast is capturing the spatial structure of the field. This technique provides information on the "climatology" of a forecast system, but no information on the forecast accuracy. Information on displacement and phase errors will not be provided by this type of technique.

Anthes (1983) also recommended the use of a correlation matrix scoring method (Tarbell et al, 1981). Although this may be able to provide some information on phase or displacement errors as well as the spatial structure of the fields, this method cannot objectively determine whether the maximum correlation is the result of the same (or similar) meteorological phenomena. In addition, the presence of small-scale, high-amplitude features may cause substantial uncertainty in the determination of the phase or displacement error. Returning to our example, one finds several different local maxima through the auto-correlation technique.

4. DEVELOPMENT OF AN "EVENTS-ORIENTED" APPROACH TO VERIFICATION

The previously discussed research on verifying forecast realism only provided a portion of the information that one could possibly obtain by a more complete analysis. It is possible that more useful verification information could be obtained if one were to classify or categorize the forecast and observed fields prior to verification. For example, the forecast and observed fields could be decomposed into subsets of small domains of a predetermined size. The predominant meteorological phenomena, or event contained within each subset could be classified. Once identified within the observed and forecast fields, the joint probability of the forecasts and observations of particular events could also be examined. There appears to be an opportunity to extend the general framework of the distributions-oriented verification approach to verifying the realism of forecasts.

Therefore, we plan to develop a method to obtain verification information on spatial patterns found in forecast and observed fields. By applying pattern recognition techniques, fields can be decomposed into sets of different *events*. Rather than trying to fit empirical shape functions to the fields (Williamson, 1981), the pattern recognition techniques will allow attributes associated with naturally occurring patterns to define the events. Ideally, the events will represent or at least correspond to significant meteorological phenomena. The approach of analyzing the joint distribution of the set of forecast events compared to the set of observed events could be described as an *events-oriented* approach to verification.

The bulk of this work (in progress) will involve developing an objective method of identifying and classifying events. This research will follow the process and use the well-established techniques in the field of *knowledge discovery in databases* (KDD, Fayyad et al 1996) and *data mining*, associated with the tools of discovery of patterns within large and complex sets of data. A large historical database, richly populated with a variety of interesting and important phenomena that span a large portion of the entire range of possible events, will be analyzed. Several methods of data reduction will be tested, including: analysis of statistical distributions, cluster analysis, principal component analysis, and spectral/wavelet analysis.

When thoroughly and carfefully analyzed, proper verification information allows for optimal use of forecast information by knowledgeable decision makers. However, regarding the issue of verifying forecast and observed fields that contain high-amplitude, small-scale features, current approaches can often be misleading and are not providing optimal information on the true quality of those forecasts. For the forecaster who is concerned with the problem of predicting mesoscale phenomena, such as forecasters at the National Weather Service's Storm Prediction Center, (SPC) or perhaps a model developer attempting to improve numerical guidance for this purpose, various methods of determining the value of forecasts containing small-scale, high-amplitude features need to be developed.

References

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

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.

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

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth, 1996: From data mining to knowledge discovery: An overview. *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. MIT Press. 1-34.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.*, **77**, 2637–2649.

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

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.

Tarbell, T. C., T. T. Warner, and R. A. Anthes, 1981: An example of the initialization of the divergent wind component in a mesoscale numerical weather prediction model. *Mon. Wea. Rev.*, **109**, 77–95.

Williamson, D. L., 1981: Storm track representation and verification. *Tellus*, **33**, 513-530.

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**, 10129-10146.