J13.5 COMPOSITE-BASED VERIFICATION OF PRECIPITATION FORECASTS FROM A MESOSCALE MODEL

Jason E. Nachamkin*, Sue Chen, and Jerome M. Schmidt

Naval Research Laboratory, Monterey, CA

1. INTRODUCTION

Mesoscale verification has become increasingly important in recent years, as computer power and numerical model resolution have increased. The term "mesoscale" is used rather loosely here, as verification methods are not necessarily associated with a particular scale. Higher grid resolutions produce increasingly complex and deterministic predictions where the details are explicitly provided. Thus, event-based standards such as those of Ebert and McBride (2000), Baldwin et. al (2001), and Nachamkin (2002, 2003) are important measures of model performance.

In this paper, composite methods introduced by Nachamkin (2002, 2003) are used to verify heavy precipitation events. Objective statistics are derived by collecting focused samples based on the existence of events in the forecasts and the observations. Though originally designed for event verification in areas with sparse observations, composite methods also work well when observations are more readily available.

Composite methods differ from the more direct object-oriented methods in that deterministic traits are not measured from each forecast-observation pair. Instead, the statistics are taken from the collective conditional distributions contingent on the existence of an event in either the observations or the forecasts. In this regard, verification becomes a data-mining problem. Organized collection strategies provide data structures that allow for complex diagnoses using relatively simple tools. For Instance, differences between the conditional biases from the observed and predicted events provide information regarding the contribution of false alarms and missed forecasts. The scale of the forecast error can also be estimated by varying the size of the sample. These statistics are demonstrated herein on operational forecasts from the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS™)[1] (Hodur, 1997) over the Continental US (CONUS) during the 2003 convective season.

2. MODEL AND OBSERVATIONAL DATA

The operational forecasts were generated at the Fleet Numerical Oceanography and Meteorology Center (FNMOC). Forecasts of 24-hr accumulated precipitation, valid at 24 and 48 hours, were collected from all forecasts initialized at 1200 UTC from 15 April through 7 September 2003. Each forecast was initialized with the multivariate optimum interpolation (MVOI) analysis, and boundary conditions were obtained from the Navy Operational Global Atmospheric Prediction System (NOGAPS). Two one-way nested grids were used in both areas with horizontal spacings of 81 and 27 km, respectively. The domain had 30 sigma levels in the vertical, with the lowest level at 10 m AGL. Subgrid-scale convective processes were parameterized using Kain and Fritsch (1993), while the explicit microphysics consisted of a modified Rutledge and Hobbs scheme (Schmidt, 2001). The forecasts were all run to 48 hours.

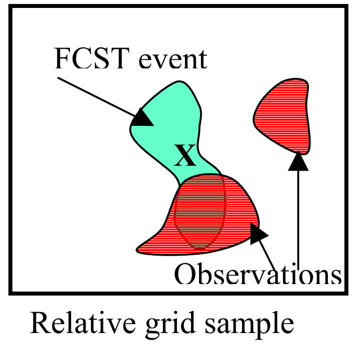

Fig. 1. Schematic depicting the relative grid centered on a predicted event (blue-green). The observations are red while the "X" marks the grid center.

The 24-hour accumulated precipitation analyses from the National Centers for Environmental Prediction (NCEP) River Forecast Center (RFC) were used as the observations. The analyses were valid at 1200 UTC with a grid resolution of 4 km. The RFC data were remapped to the COAMPS 27 km grid using an upscale discrete algorithm. The errors associated with the remapping were noted to be generally small by Chen et. al (2002).

3. VERIFICATION TECHNIQUE

Two separate distributions were collected, one based on the existence of an event in the forecasts and the other

* *Corresponding author address:* Jason E. Nachamkin
Naval Research Laboratory, Monterey, CA 93943-5502
email address: nachamkin@nrlmry.navy.mil

[1] COAMPS is a registered trademark of the Naval Research Laboratory.

contingent on an observed event. Heavy rain events were defined as contiguous precipitation areas anywhere over the CONUS containing 24-hr rain amounts of 25 mm (~ 1 inch) or greater. To ensure the events were well resolved by the model, only those events containing 50-500 grid points were selected. In practice, most events contained less than 300 grid points.

Once an event was identified in the forecasts (observations), all surrounding data were transferred to a 31X31 point relative grid with the same grid spacing as the model. The center of the event as defined by its area-weighted "center of mass" (Fig. 1) was positioned at the center of the relative grid. At that point, all available observational (forecast) data were also positioned on the relative grid. Model data were templated by the available observations, such that all forecasts outside of the contiguous coverage of the RFC analysis were cut from the set.

Ebert and McBride (2000) noted that the existence of boundaries within the data lead to errors in determining event-related quantities. The conditional distributions contingent on the forecast events are unaffected by this because the relative grid position is determined by the forecasts alone. However, the observation-based conditional distributions will be affected since partially observed events lead to errors in the location of the event center. These errors effectively increase the observational variance because the observed events are not all directly superimposed. In the data-mining sense, the data structures are less coherent and features like systematic phase errors are less apparent. Ebert and McBride (2000) noted that the phase error, and thus event position, was among the least sensitive parameters to the data boundary errors. In a Monte-Carlo experiment, they noted standard displacement errors on the order of two grid points or less for most events.

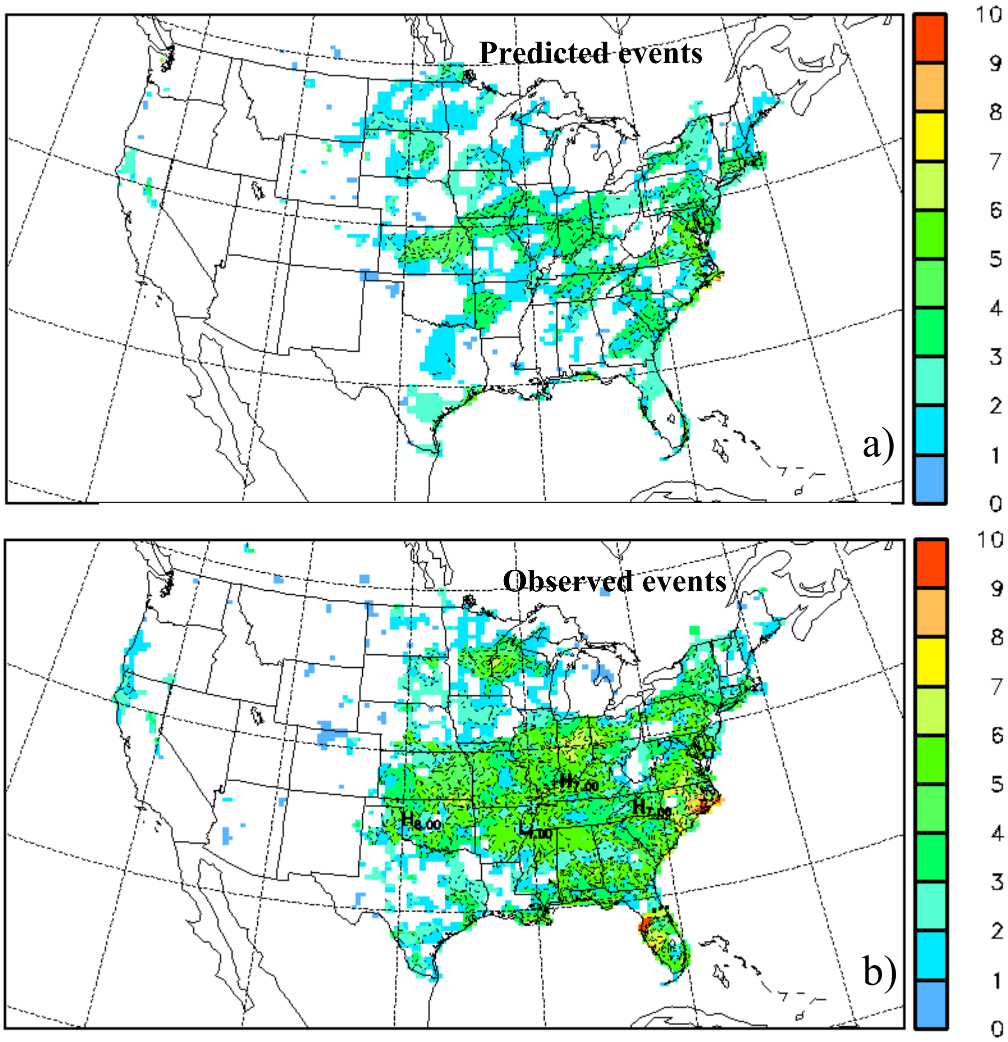

Fig. 2. Number frequency of 24-hour rainfall accumulations greater than or equal to 25 mm in a) the 24-hour forecasts and b) the RFC precipitation analyses. Statistics were collected from 15 April through 7 September 2003. Coverage of the predicted events was templated against the observation coverage.

4. RESULTS

The 24-hour event occurrence frequencies on the native 27 km model grid, templated by the available observations, are depicted in Figure 2. These represent the number of days with 24-hour accumulations of 25 mm or greater in a given place over the entire convective season. Most of the observed events occurred east of the Rocky Mountains with local maxima in North Carolina and Florida. The number of forecast events (Fig. 2a) was lower in almost all areas, most notably Florida where almost no events were predicted. Low biases at high precipitation amounts are a common problem in most mesoscale models. The convective forcing mechanisms, such as the sea breeze, are not well resolved by the 27 km model grid.

The precipitation distributions on the relative grid (Fig. 3) indicate the model possessed some skill, especially when an event was predicted. The conditional distributions of the observations and forecasts, contingent on an event in the forecasts (Fig. 3a), were similar in structure. Both the observations and the forecasts displayed consolidated maxima of 40 and 50 mm, respectively. The forecast precipitation fields were systematically phase shifted by about 3 grid points to the north and west of the observations. For the cases contingent on an observed event (Fig. 3b), the bias was strongly negative, indicating many events in which not enough precipitation was predicted. The forecast distribution was diluted by the missed forecasts. Despite this, some evidence of the phase shift was apparent as the maximum average forecast rainfall amounts (~ 18 mm) was shifted about 2 grid points north and west of the observed maximum.

The event occurrences can be further parsed by investigating the number distributions of the event-scale precipitation statistics (Fig. 4). Here the forecast-observation pairs in each conditional distribution were sorted by the ratio of the grid-average forecast precipitation to that of the observations (F/O) taken over the relative grid. These statistics indicate that majority of the forecasts in the forecast-based contingency sample (blue line in Fig. 4) were within 25% of the observations on the area of the

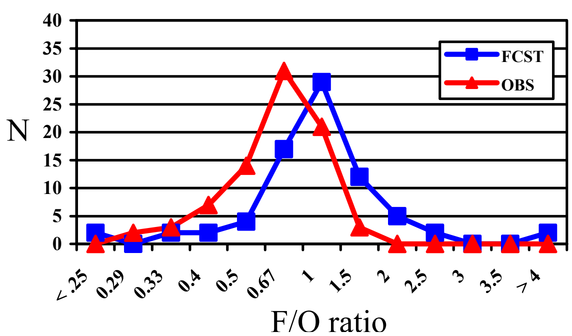

Fig. 4. Number frequencies of the ratio of the predicted rainfall to the observed rainfall on the relative grid. Forecast-based contingencies are blue while observation-based contingencies are red.

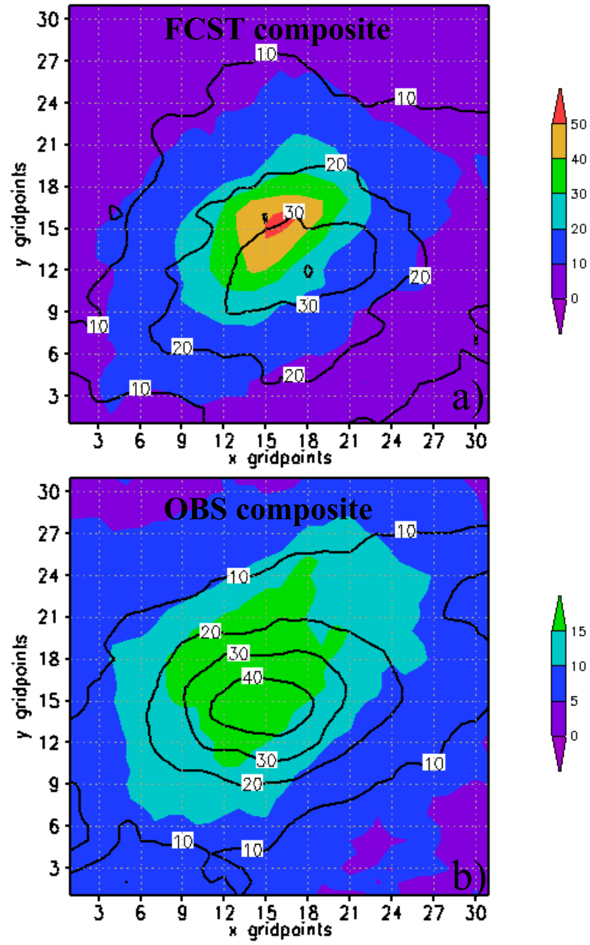

Fig. 3. Average precipitation amounts (mm) on the relative grid. Model precipitation from the 27 km grid is shaded while observed values are contoured. The distribution contingent on the existence of an event in the forecasts is displayed in a), and the distribution contingent on the observed events is displayed in b).

relative grid. However the tails of the distribution did contain false alarms and even some missed forecasts. The observation-based contingency samples (red line) contained almost no false alarms, and the peak was shifted towards the lower ratios. However, many of these forecasts did contain some precipitation. Isolating systematic traits of this distribution of forecasts would be of considerable aid to diagnosing the under-prediction problem. The composite method is well posed for this because any field can be collected and averaged on the relative grid. This reiterates the connection between verification and data mining.

5. ERROR QUANTIFICATION

Objectively quantifying the effects of phase errors is perhaps the most elusive aspect of mesoscale verification. As the forecasts become more detailed, small differences in position or timing result in major penalties in the objective statistics. Much of this problem arises from the way the statistics are collected. The equitable threat score (ETS) is based on dichotomous point-wise samples. No consideration is given to the surrounding points, and thus no consideration is given to the scale of hits and misses. The result is an extremely sensitive parameter where the error is rapidly saturated by even the smallest deviations.

Alternately, the bias presents the opposite problem. Biases are often calculated over the entire native model grid, and all information regarding scale or location is lost in the averaging process. However, this problem can be alleviated by sorting the samples before any of the statistics are calculated. Focusing the samples limits the degrees of freedom within the data, paving the way for consistent, objective statistics. The conditional composites are useful in this regard in that an event of a given type and scale is known to exist near grid center. Statistics taken at varying scales on the relative grid reveal information regarding the nature and scale of the errors.

These concepts are demonstrated in Figure 5a, which shows 24-hour relative grid bias for both conditional composites calculated at successive cocentric squares centered at grid center. The bias scores for the smallest areas are quite large, in part because of small area. They are taken at the center of the events where precipitation rates and their associated gradients are large. Small displacement errors lead to large differences in the statistics at this scale. The errors are more systematic in the observation-based composite, but statistics taken at this point alone are oblivious to this. As the sample scale increases the biases trend closer to zero, with the forecast-based biases actually reaching zero as the box size reaches 15X15 grid points. The conditional distributions in Figure 3a indicate that an area of this size encompasses the average phase error. This means that when the model predicts a rain event at 24-hours, forecasts for that amount of rain spread over a 15X15 grid point area will on average be well calibrated.

Of course this is only half of the story because the distribution of the forecasts given the observations has yet to be considered. The biases there never cross the zero line due to the under-prediction problem. The decreasing magnitude of the bias with increasing scale graphically illustrates the decreasing sensitivity of the bias at larger scales. Figure 3b shows that forecasts issued on the full 31X31 point relative grid will not be much better than those issued at smaller scales. The bias simply decreases away from the event center as the magnitude of the average rainfall decreases. This behavior becomes apparent in the grid integrated rainfall errors (Fig. 5b). Here, the forecast-based errors still cross the zero line at the 15X15 point scale, but the observation-based errors steadily accumulate. This indicates that both the bias and the accumulated error should be examined when evaluating event displacement on the relative grid. These measures taken together represent useful tools for objectively characterizing the error.

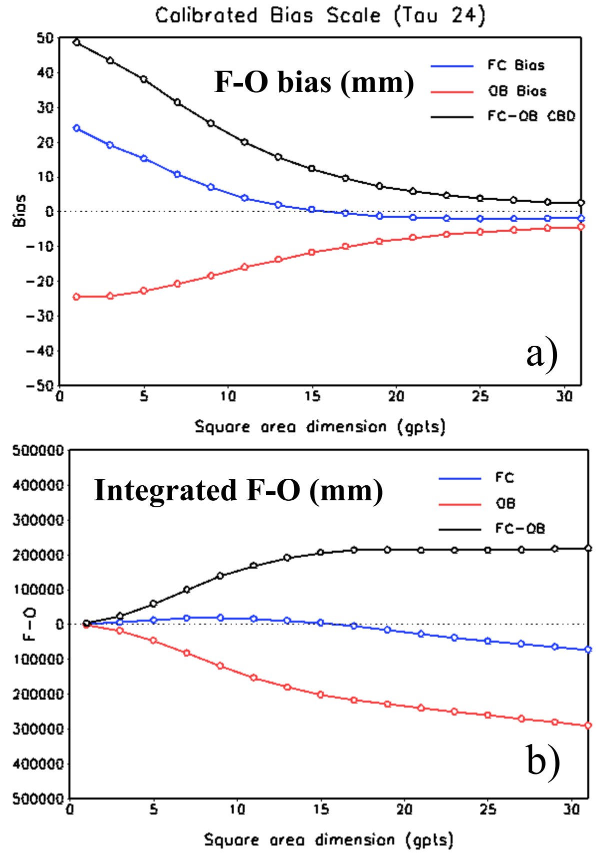

Fig. 5. Grid-total statistics are plotted as a function of relative grid size for the 24-hr forecasts. Statistics from the forecast-contingent distributions are blue while the observation-contingent statistics are red. The black lines represent the difference between the two statistics. In a) relative grid biases are plotted while in b) the integrated F-O differences are plotted (both in mm).

6. CONCLUSIONS

Taken together, these statistics presented here can be used to objectively and subjectively describe the prediction of discrete events by a numerical model. The end result is a set of verification statistics that is consistent with the kinds of forecasts that are currently issued in many circumstances. Although the model output is becoming increasingly detailed, operational forecasts will likely remain probabilistic. To be useful, verification should communicate the optimal conditions where event-oriented forecasts have a high probability of being correct. At the same time a verification package should allow for the simple investigation of the forecasts that went bad. Systematic traits of the poor forecasts can be collected with the goal of informing the user of impending error and ultimately improving the forecasts themselves. In that regard verification and statistical post-processing are intertwined.

7. REFERENCES

Baldwin, M. E., S. Lakshmivarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. *Preprints, 9th Conference on Mesoscale Processes,* 30 July-2 August, Ft. Lauderdale, FL, American Meteorological society, 255-258.

Chen, S, J. E. Nachamkin, J. M. Schmidt, and C-S Liou 2002: Quantitative precipitation forecast for the coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Preprints 19th Conference on Weather Analysis and Forecasting,* Amer. Meteor. Soc., San Antonio, TX 202-205.

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

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.* **125,** 1414-1430.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. *The Rerepresentation of Cumulus Convection in Numerical Models, Meteor. Monogr.*. No. 46. Amer. Meteor. Soc., 165-170.

Nachamkin J. E. 2002: Forecast verification using meteorological event composites. *Preprints, 19th Conference on Weather Analysis and Forecasting*, 12-16 August, San Antonio, TX 206-209.

Nachamkin J. E. 2003: Mesoscale verification using meteorological composites. *Mon. Wea. Rev* Accepted for publication, copies available from author.

Schmidt, J. M., 2001: Moist physics development for the Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). BACIMO, Fort Collins, CO.

8. ACKNOWLEDGEMENTS

This research is supported by the Office of Naval Research (ONR) through program element 0602435N, and the Space and Naval Warfare Systems Command (SPAWAR) through program element 603207N. Special thanks go to Fleet Numerical Meteorology and Oceanography Center (FNMOC) for providing computer time and support.