Reply

DAVID J. STENSRUD AND HAROLD E. BROOKS

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

JUN DU, M. STEVEN TRACTON, AND ERIC ROGERS

NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland

20 January 2000

_____

We welcome the opportunity provided by the comments of Leslie and Speer (2000, hereafter referred to as LS00) to further discuss and clarify the conclusions reached in our recent publication on short-range ensemble forecasting (SREF; Stensrud et al. 1999). In that paper we calculate the mean absolute errors of basic model variables for both the ensemble mean and a single higher-resolution forecast model, using rawinsonde data as the verifying observations for twenty 36-h forecast periods. We conclude that the mean values from a 10-member ensemble of 80-km Eta model forecasts of these basic atmospheric variables (temperature, relative humidity, and wind speed and direction) are as accurate as the 29-km Meso Eta Model values at standard mandatory pressure levels. LS00 suggest that this conclusion implies, seemingly without a caveat, that an ensemble of 80-km forecasts is as accurate as a single high-resolution 29-km resolution forecast and they disagree with this conclusion. In particular, they point out that for some meteorological features a model must have sufficiently high resolution to resolve the feature, and an ensemble of model forecasts at a lower resolution would yield little forecast value since the feature could not be predicted to a sufficient degree of accuracy.

First, we want to emphasize that our conclusions are correct, since whenever one discusses the accuracy of a model forecast it must always be measured using a particular metric. The metric used in our analysis is the mean absolute error of the model forecasts as verified against rawinsonde data, which for the North American region have a data spacing of approximately 400 km. Thus, the verification is calculated with respect to large-scale features. Indeed, for most meteorological fields, the data available for verification capture at most the large-scale features of the atmosphere. Therefore, even

when a higher-resolution model is used, it is difficult to assess the ability of any of the models to reproduce features with small-scale structures, unless special data are used. One of the only exceptions is a national precipitation analysis based upon radar data (e.g., Baldwin and Mitchell 1996) that has higher resolution than the present operational models and can capture small-scale features. This makes the comparison of models with differing horizontal and vertical resolutions a challenging task, and more research is needed in this area of verification.

Yet the point raised by LS00 is an interesting one, and has been heard many times during the presentation of our results of SREF experiments in regard to topographically induced flow patterns in the western United States. We fully agree with LS00 that a 25-km resolution forecast cannot adequately predict tropical cyclone intensity and structure, as indicated clearly in their Fig. 1. A 25-km model also cannot capture many of the local flow patterns in regions of complex terrain. However, the goal of ensemble forecasting is to maximize the total utility of the forecasts instead of the model skill (cf. Tracton and Kalnay 1993). By focusing only upon the intensity and structure of tropical cyclone forecasts, LS00 seem to be emphasizing the goal of model skill over the goal of total model utility, in which multiple realizations of the forecast outcome are produced and the forecast is inherently probabilistic.

Owing to the discrete nature of numerical weather prediction models, there will always be features that are small enough that the model cannot reproduce them. If one such feature has operational forecast significance, does this imply that an ensemble of lower-resolution model forecasts has little or no value? No, because we know that the evolution of the large-scale flow patterns often influences the development and evolution of smaller-scale atmospheric features. One example of this effect is a sea-breeze circulation. The sea breeze typically extends farther inland when the low-level, large-

scale flow is also directed in the inland direction (Atkinson 1981). Thus, the evolution of a relatively small-scale feature depends intimately upon the evolution of the large-scale flow. It then becomes very important to provide information about the uncertainty of the larger-scale flow, which is a goal of ensemble forecasting.

This appears to imply a trade-off between model resolution and probabilistic information content, since computer time is finite. While it is well known that the atmosphere exhibits a sensitive dependence to initial conditions (Lorenz 1963), there is a strong desire to obtain as much information from a model forecast as possible. In the case of tropical cyclones, information on changes in intensity and structure would be very helpful in an operational setting. However, the question then becomes whether or not the forecasts from a model with sufficient resolution to capture tropical cyclones are more useful than an ensemble of lower-resolution model forecasts. What happens when the single, high-resolution model forecast goes awry, leaving the forecaster with little or no guidance? Is this a better use of our computer resources than an ensemble of lower-resolution models that provide information on a variety of possible outcomes, even if these forecasts do not provide explicit guidance on tropical cyclone intensity changes?

We believe, and think LS00 would agree, that the apparent competition between single, very high resolution model forecasts and an ensemble of model forecasts is a harmful viewpoint. In reality, both are needed to provide the most useful forecast guidance. One way to view a merger of high-resolution models and lower-resolution ensemble forecasts is to use the ensembles to provide information on the uncertainty in the larger scales. Then the one or two most likely large-scale forecasts, as determined using the ensemble output, could be used to provide the initial and boundary conditions for a very high resolution model forecast. This high-resolution forecast(s) then is used to provide the details missing from the lower-resolution ensemble forecasts, possibly including information on tropical cyclone intensity and structure changes. Yet it would be based upon the most likely forecast scenario, and even if the high-resolution forecast goes awry the ensemble data are still available to provide information on alternative forecast scenarios. In this way, probabilistic information is used to help determine the forecast produced from the high-resolution model.

The probabilistic information provided by the ensemble also can be postprocessed to provide more accurate information than is available from the raw model output, such as is available using the model output statistics (Jacks et al. 1990). Vislocky and Fritsch (1995) present compelling evidence for why we should be combining all the available postprocessed forecast products instead of relying upon the single most superior postprocessed forecast product. In addition, this postprocessing can often produce details not seen in the actual model output. Hall et al. (1999) show results from a neural network that uses large-scale model data and an observed sounding as input to produce very detailed precipitation forecasts for one location. Leung et al. (1996) obtain good results using output from an atmospheric model with 90-km resolution to drive a high-resolution watershed model with 180-m resolution over the Pacific Northwest, illustrating yet another form of model postprocessing. These results suggest that it may be possible for a model with 25-km grid spacing to predict tropical cyclone intensity and structure changes when the model output is combined with a good postprocessing scheme.

The second point that LS00 raise regards the low level of correlation between the ensemble mean cyclone location forecasts and the ensemble spread. Unlike LS00, Stensrud et al. (1999) only compared the cyclone locations at the 36-h forecast time and did not examine central pressure, wind speed, and maximum rainfall amounts as in Table 1 of LS00. However, the differences in mean absolute errors in east coast low position reported in LS00 from the ensemble mean forecast and the single higher-resolution model forecast are similar to those found in our study, if the ensemble mean cyclone location is defined using the mean sea level pressure (MSLP) of all the ensemble members. If one instead uses the mean position of the individual cyclones (MPIC) of all the ensemble members to determine the ensemble mean cyclone location, then we find that the accuracy of the ensemble mean and the high-resolution cyclone location are indistinguishable. Since LS00 do not indicate which method is used to determine the ensemble mean cyclone location, it is difficult to compare the results, although we suspect that they used the MSLP approach since they also report the errors in central pressure. In addition, it would be helpful if a significance level could be calculated for the differences in LS00 to help the reader determine if the differences are large enough to be important.

Finally, LS00 find a correlation coefficient of $r = 0.45$ between cyclone location errors and the ensemble spread, very similar to our results that produce a correlation coefficient of $r = 0.36$. While we are always glad to see corroborating evidence, this lack of correlation between skill and spread is troublesome. It may be that spread is only a useful predictor of skill when the spread values are extreme, as suggested in the study of Whitaker and Loughe (1998), and that we should not expect large correlation coefficients between spread and skill. This indicates that we must begin to use metrics that verify the entire probability distribution function for various forecast aspects. Regardless, we fully agree with LS00 that more work is needed in trying to predict forecast skill, and we will continue to explore the richness of ensemble forecasts.

doctoral fellowship sponsored by the NWS Office of Meteorology.

REFERENCES

Atkinson, B. W., 1981: *Meso-scale Atmospheric Circulations.* Academic Press, 495 pp.

Baldwin, M. E., and K. E. Mitchell, 1996: The NCEP hourly multisensor U.S. precipitation analysis. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., J95–J96.

Hall, T., H. E. Brooks, and C. A. Doswell III, 1999: Precipitation forecasting using a neural network. *Wea. Forecasting,* **14,** 338–345.

Jacks, E., J. B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson, and S. C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting,* **5,** 128–138.

Leslie, L. M., and M. S. Speer, 2000: Comments on "Using ensembles for short-range forecasting." *Mon. Wea. Rev.,* **128,** 3018–3020.

Leung, L. R., M. S. Wigmosta, S. J. Ghan, D. J. Epstein, and L. W. Vail, 1996: Application of a subgrid orographic precipitation/surface hydrology scheme to a mountain watershed. *J. Geophys. Res.,* **101,** 12 803–12 817.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Stensrud, D. J., H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433–446.

Tracton, S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational implementation. *Wea. Forecasting,* **8,** 379–398.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved model output statistics forecasts through model concensus. *Bull. Amer. Meteor. Soc.,* **76,** 1157–1164.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.,* **126,** 3292–3302.