Comments on ''Using Ensembles for Short-Range Forecasting''

LANCE M. LESLIE

School of Mathematics, University of New South Wales, Sydney, Australia

MILTON S. SPEER

Bureau of Meteorology, Sydney, Australia

8 December 1999

In their paper, Stensrud et al. (1999, henceforth referred to as S99) have reported on findings from a ''pilot'' study in short-range ensemble forecasting (SREF), using ensembles from the National Centers for Environmental Prediction (NCEP). The ensembles used in S99 were derived from two NCEP models, the 80-km Eta Model and the regional spectral model, which has approximately the same resolution as the 80-km Eta Model. We are interested in two main findings from the S99 study. Our primary interest is in the first of these, in which S99 showed that the accuracy of the ensemble mean of a medium-sized selection of 80-km coarse-resolution ensemble forecasts was comparable with that obtained from single forecasts using the much higher resolution, 29-km fine-resolution Meso Eta Model. This result holds true in their study for two very different numerical experiments and for a wide range of forecast parameters. It is a very significant finding and has profound consequences if it is unconditionally true. A second result of interest to us is that they found almost no correlation between the spread of the ensemble members and the accuracy of the ensemble mean in the prediction of cyclone locations. The authors noted the challenge presented to SREF by this poor correlation. One of the driving forces for SREF has been the potential of forecast spread as a major input parameter (and often the sole one) in producing a priori estimates of forecast uncertainty. This second finding therefore is also of utmost importance.

We wish to state at the outset that we regard S99 as a very valuable and thought-provoking study. It provides an excellent review of the background to and motivation for SREF work, presents some very interesting SREF results, and raises a number of important questions. Like the authors, we also are motivated primarily by the need for accurate and timely predictions of a whole range of short-range regional forecasts of primary and derived meteorological variables, together with statistical information about the forecasts. We also wish to provide, along with the predictions, quantitative measures of confidence in the predictions, by obtaining knowledge of the evolution of the initial error fields over the forecast period. The viability of the SREF approach is central to this goal. Our SREF approach was summarized in a recent study (Leslie and Speer 1998, hereafter referred to as LS98). It presented the results of a case study of explosive Australian east coast cyclogenesis. LS98 found that the application of an SREF approach yielded valuable information that, if followed, would have prevented an underforecasting of the wind strength over land. The official forecast was for strong winds whereas the ensemble procedure suggested that a land gale was more probable than strong winds. This indeed turned out to be correct. The work of LS98 was commented on formally by Hamill (1998) and also informally in a fruitful exchange of electronic mail between a range of researchers and practitioners. Here we mention only one of Hamill's comments, which was echoed by the informal comments, namely, the need for a continuing, careful, discussion of the cost–benefit analysis of the resources required for an ensemble of coarse-resolution forecasts when compared with those for a single high-resolution prediction.

Turning to specifics, S99 carried out an experiment in which ensembles of ten 80-km Eta Model initial states were used to generate forecasts from 20 different archived analyses. Our interest centers on two findings by S99. Both relate to comparisons of the mean average errors (MAEs) between the ensemble mean of the 80-km Eta Model and the single Meso Eta Model predictions. First, they showed that the ensemble mean of the 80-km resolution forecasts ''compares favorably with the Meso Eta Model for all parameters and pressure levels'' (p. 436). Second, they compared the locations

Corresponding author address: Lance M. Leslie, School of Mathematics, University of New South Wales, Sydney 2052, Australia.
E-mail: L.Leslie@unsw.edu.au

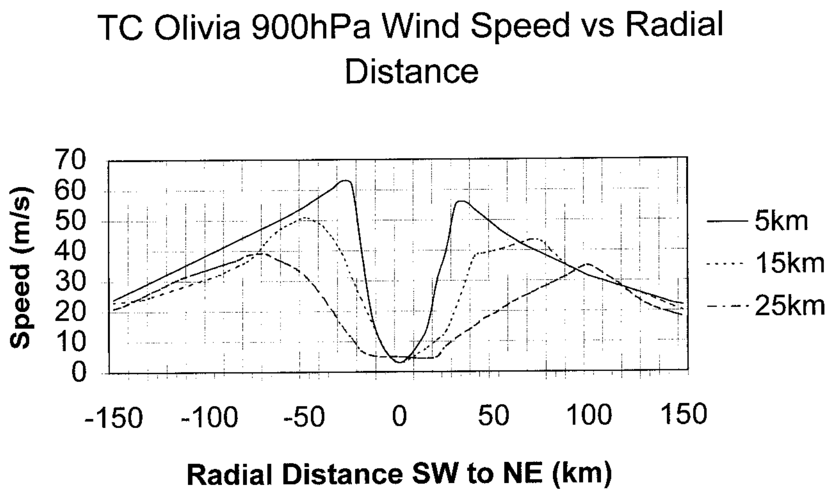

FIG. 1. A radial SW–NE cross section through Tropical Cyclone Olivia at 900 hPa. Olivia occurred off the NW coast of Western Australia and the cross section was valid at 0000 UTC 9 Apr 1996. Note the dramatic improvement in wind strength and structure as the resolution is increased from 25 to 5 km.

of cyclones over North America again from the ensemble mean of the 80-km and the Meso Eta Models. In this experiment a total of 33 cyclone cases was used and, once again, S99 found that ''results indicate that the MAEs of the ensemble mean and Meso Eta Model cyclone locations are virtually identical'' (p. 438).

Our difference in interpretation of the results obtained by S99 can be stated very simply. They imply, seemingly without a caveat, that an ensemble of 80-km forecasts is as accurate as a single high-resolution 29-km resolution forecast. This simply is not true, as a blanket statement. There are many meteorological systems in which the model resolution must be below a threshold value to enable the intensity, location, path, and other parameters such as rainfall totals and patterns to be predicted with an acceptable degree of accuracy. Such systems are readily represented by tropical cyclones and strong to explosive extratropical cyclones. There are numerous other meteorological systems for which this statement holds but those will suffice here. In the case of tropical cyclones it is well established that to obtain hurricane force winds and realistic structure requires horizontal resolutions that are able to resolve down to scales of the order of the eye of the storm. This threshold resolution is approximately 10 km or less. Figure 1 is an adaptation of the study of tropical cyclogenesis by LeMarshall and Leslie (1999) and clearly illustrates the futility of inadequate resolution in forecasting tropical cyclone intensity and structure. An ensemble of 25-km resolution forecasts, no matter how many and how well chosen, simply will not reproduce the central pressures and wind strengths of a single 10- or 5-km model forecast. As such, the ensemble mean is of little value. Simple scaling analysis readily confirms what the NWP models produce. In the case of explosive marine cyclones, such as Australian east coast lows or the North American Atlantic and Pacific Ocean ''bombs,'' the threshold values depend upon the strength of the storms but 80-km resolution predictions again are incapable of acceptable simulations of such storms. Horizontal resolutions of at least 25 km or less appear to be required (Leslie et al. 1987), with resolutions of 10 km or less necessary for the accurate depiction of the most severe events (Buckley and Leslie 1999, manuscript submitted to *Wea. Forecasting*). To quantify our assertion we have performed a study that compared the MAEs from 10-member ensembles at 75-km resolution, with single forecasts at 25-km resolution on a total of 30 east coast lows. The east coast lows were of varying strengths from relatively weak to intense, explosive east coast cyclones. We looked at errors in only four variables at 24 h: location, central pressure, wind speed, and maximum rainfall amounts. The results are summarized in Table 1. They bear little resemblance to the findings of S99, in which the differences were not statistically significant.

TABLE 1. Mean average forecast error in east coast low position from 30 cases comparing a 10-member 75-km ensemble version of The University of New South Wales high-resolution NWP model with single forecasts at 25-km resolution. The four meteorological variables are cyclone location, central pressure, maximum wind speed, and area-averaged rainfall amount (above 5 mm).

	Mean average error (MAE)	
	Resolution	
Variable	75-km ensemble	25-km single forecast
Location	114 km	87 km
Central pressure	7 hPa	3 hPa
Max wind	7 m s^{-1}	4 m s^{-1}
Mean areal rainfall amount	58 mm	22 mm

We do not believe that S99 think that it is a universal truth that the ensemble mean of 80-km resolution forecasts can produce the same wind strengths and central pressures as a single 29-km forecast. However, a major motivation for this comment is to encourage studies like S99 to be as clear as possible when discussing SREF studies and their implications. The importance of SREF requires, we think, the minimizing of possible misinterpretation. New ideas, strategies, and applications are needed to employ SREF fruitfully and to this end we wished to point out that there are some overgeneralizations in S99.

Finally, we turn to the second aspect of S99 that we wish to comment on, namely the low level of correlation they found between ensemble mean cyclone location forecasts and ensemble spread. We also carried out a similar series of 10-member ensemble forecasts of east coast lows, at 25-km resolution, to calculate the correlation between ensemble mean and ensemble spread, in the manner of S99. S99 found a correlation coefficient of $r = 0.36$ (p. 439) between the Meso Eta Model cyclone location errors and the ensemble spread. This was a disappointing result for predicting forecast skill, as they state. Unfortunately, our 25-km resolution east coast low forecasts, described above, fared only a little better. For our 30 cases we found a correlation coefficient of $r = 0.45$, again a value that is too small for confidence in predicting the skill of forecast cyclone locations. We are continuing to work on this aspect, testing a variety of approaches that will be reported on later, but at this stage our results are very much in accordance with the findings of S99. As a totally independent study, our results add weight, even if not particularly encouraging, to the findings of S99. More work is needed in this crucial area of research.

Acknowledgments. The first author (LML) was supported in this work by an Australian Research Council Large Grant and the U.S. Office of Naval Research under Grant N-00014-94-1-0556.

REFERENCES

Hamill, T. M., 1998: Comments on ''Short-range ensemble forecasting of explosive Australian east coast cyclogenesis.'' *Wea. Forecasting,* **13,** 1205–1207.

LeMarshall, J. F., and L. M. Leslie, 1999: Modelling tropical cyclone intensity. *Aust. Meteor. Mag.,* **48,** 147–152.

Leslie, L. M., and M. S. Speer, 1998: Short-range ensemble forecasting of explosive Australian east coast cyclogenesis. *Wea. Forecasting,* **13,** 822–832.

——, G. J. Holland, and A. H. Lynch, 1987: Australian east coast cyclones. Part II: Numerical modeling experiments. *Mon. Wea. Rev.,* **115,** 3037–3053.

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