12.5 **ENSEMBLE FORECASTING OF TROPICAL CYCLONE TRACKS**

Sim D. Aberson[1], Morris A. Bender[2], and Robert E. Tuleya[2]
[1] Hurricane Research Division, AOML/NOAA, Miami, FL
[2] Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ

1. INTRODUCTION

The atmosphere has long been recognized as a chaotic system (Lorenz 1963), e.g. very small perturbations to initial conditions result in increasingly large differences in the evolution of the atmosphere with time. Since the exact state of the atmosphere can never be measured, all analyses contain errors whose magnitudes can only be estimated. An indeterminate number of initial conditions consistent with the observational data can therefore be used in numerical weather prediction, and single model runs at any synoptic time only give one possible solution to the evolution of the atmosphere.

Many operational forecast centers around the world, therefore, now employ ensemble forecasting as a means of quantifying the uncertainty in the evolution of the atmospheric system. Small perturbations from a best "control" state are calculated and added to and subtracted from the control to allow for different integrations starting from theoretically equally likely initial states. These perturbations are designed so as to create the largest envelope of possibilities in the forecast. The method used to create these modes at the National Centers for Environmental Prediction (NCEP) are discussed in Toth and Kalnay (1993) and Tracton and Kalnay (1993).

Tropical cyclone track prediction is one of the simplest methods with which to test ensemble forecasting, since only one parameter is being forecast, the location. Simple models have already been utilized for ensemble forecasting of tropical cyclone tracks (Aberson et al. 1995). The Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model (Bender et al. 1993), a triply-nested movable mesh primitive equation model, has been run since 1992, and has proven to be the most accurate tropical cyclone track prediction model available in the North Atlantic basin (Aberson and DeMaria 1994). Ensemble forecasts from a two-mesh version of the GFDL model using the bred modes from the NCEP global model are presented for cases during the 1996 and 1997 hurricane seasons in the Atlantic and East Pacific basins. The model is initialized with the initial conditions corresponding to the various NCEP Medium-Range Forecast (MRF) ensemble members, with corresponding boundary conditions from the integrationof that particular member. Thirty-three cases during the 1996 hurricane seasons have been completed; almost 30 from the 1997 hurricane season have been run through the first week of September.

Corresponding author address: Sim D. Aberson, Hurricane Research Division, 4301 Rickenbacker Causeway, Miami, FL 33149.
e-mail: aberson@aoml.noaa.gov

2. FORECAST VERIFICATION

The "control" analysis is considered the best representation of the initial conditions. As such, the control forecast should provide the best forecasts of all the possible ensemble members. Of the remaining ensemble members, all are theoretically equally likely to perform best in each case, so that the errors, after a large number of cases is verified, should be approximately the same. Figure 1 shows the great-circle distance errors of each ensemble member compared to those from CLIPER, a simple statistical model based upon climatology and persistence (Neumann 1972); those members of the ensemble showing negative relative error are said to exhibit skill. The control forecast (GFCT) is either the best or the second best forecast at all times. The differences between the GFCT and the best ensemble member are not statistically significant at the 90\% level at any time based upon a paired t test with the null hypothesis that the mean of the differences is not significantly different from zero. However, the P4 ensemble member (GFP4) seems to perform better than the other members on average, and these differences are generally statistically significant between 12 and 72 h for a number of other ensemble members. This performance may be due to the relatively small number of cases in the sample, and does not

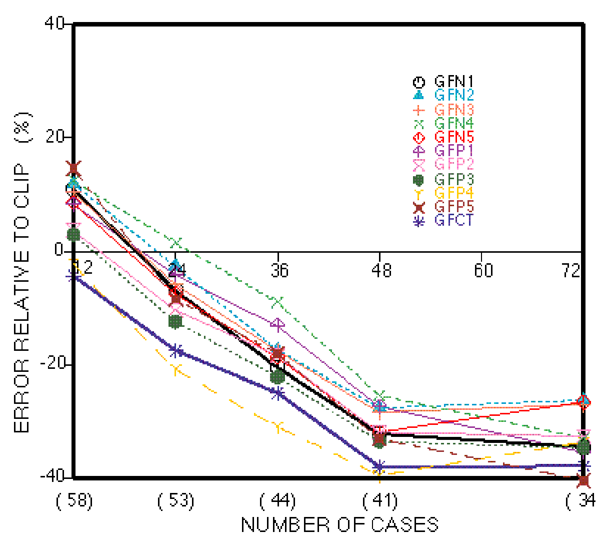

FIG 1. Relative errors of the 11 ensemble members versus CLIPER.

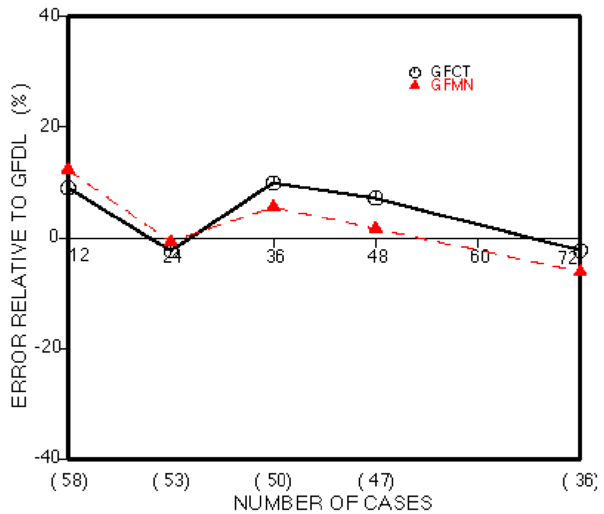

FIG 2. Comparison between operational GFDL model, the control ensemble (GFCT) and the ensemble mean (GFMN).

reflect upon the general value of the individual ensemble members.

The ensemble mean is expected to perform better than individual members. Figure 2 shows the ensemble mean (GFMN) compared to the control (GFCT) and the operational three mesh GFDL model. The quality of the forecasts from the high resolution GFDL model and the mean of the ensembles from the lower resolution GFMN is generally the same. The GFMN does improve upon the GFCT forecasts after 24 h, though the differences between the forecasts are statistically significant at the 90\% level only at 72 h.

3. RANKING OF ENSEMBLE FORECASTS

Given the perfect model assumption, any of the perturbed ensemble members are likely to be the best forecast at any particular time. Another way to test this is the rank distribution in which the best estimate of the atmospheric state at each forecast time is ranked between the individual ensemble forecasts ordered in some way. In the two-dimensional case of location, a line is fit to the forecast points based upon a least-square fit, and the forecast points are ordered along the line from west to east. The ranking occurs by finding the location of the true location of the tropical cyclone at the forecast time along the fitted line, and deciding between which forecasts positions this point exists. The results are shown in Fig. 3a for all 36 h cases. The results confirm a seeming westward bias of the GFDL forecast tracks, and the fact that the ensemble spread is unable to envelop the true forecast track in most cases. If the biases are removed from the forecasts (Fig. 3b), the rank distributions improve somewhat, but most of the forecasts remain in the wings of the distribution. This somewhat limits the usefulness of the ensemble and suggests that the perfect model assumption is not quite

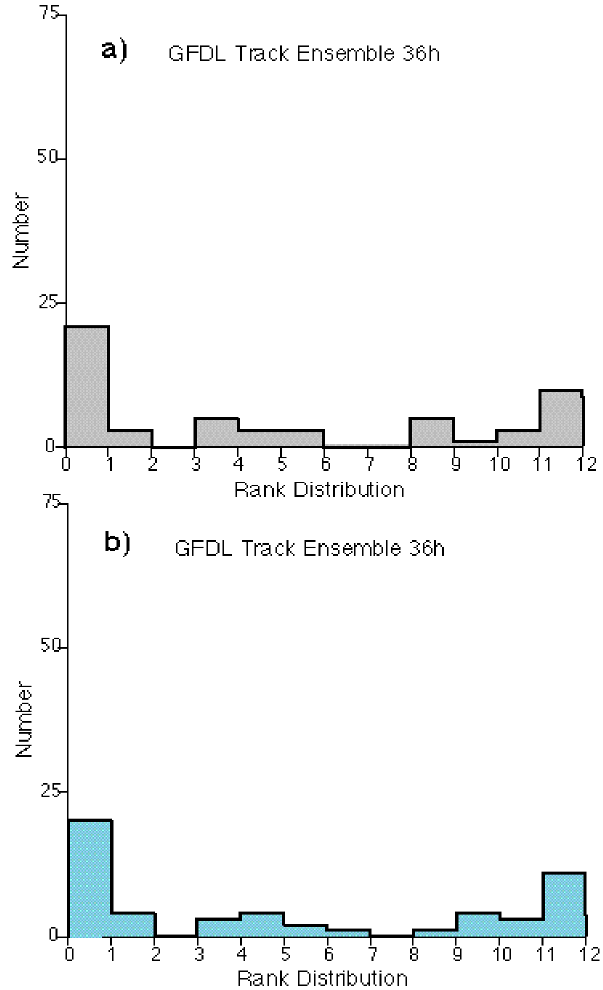

FIG 3. Rank distributions of the (a) GFDL ensemble track forecasts, and (b) GFDL ensemble track forecasts with the error bias removed.

applicable in this case. Results from other forecast times are similar.

4. SPREAD VS. ERRORS}

Ensemble forecasts are able to quantify the likely predictability of individual forecast cases. If the spread within the ensemble is large, the uncertainty, and therefore the forecast errors, can potentially also be large. Figure 4 shows the relationship between the spread within each ensemble, measured as the greatest distance between ensemble members, and the actual error at 36 h for all cases. There is no clear correlation between spread and error. However, the cases in which the forecast error is large tend to also have a large spread, and there seems to be a lower limit to the spread as the forecast error increases. A number of cases with large spread still show small forecast error, which can be expected if the verification is within the envelope of possibilities exhibited by the ensembles. Results at other forecast times are similar.

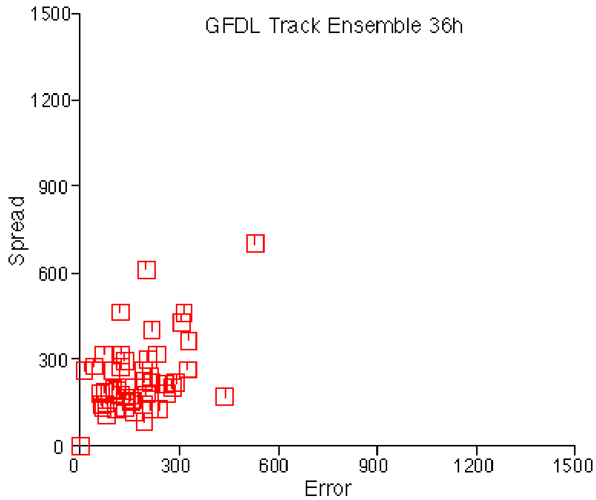

FIG 4. Relationship between forecast error (the great-circle distance between the forecast and verifying position) and the maximum spread among the ensemble members for the GFDL track ensembles at 36 h.

5. CONCLUSIONS

Ensemble forecasting with the GFDL hurricane model shows the general utility in ensemble forecasting. The ensemble mean is able to produce better forecasts than the control, and forecasts comparable to the high resolution GFDL model. The spread among the ensemble members gives an estimate of the potential forecast track errors in each case. However, the spread among the ensemble members is not yet large enough to encompass the actual evolution of the atmosphere in these cases, and model biases potentially limit the ability of the model to produce the most accurate forecasts. Planned improvements to the NCEP ensembles to increase the spread among the ensemble members in the tropics will certainly help to improve the track forecasts in the future.

6. REFERENCES

Aberson, S. D., and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.*, **122**, 2804-2815.

Aberson, S. D., S. J. Lord, M. DeMaria, and M. S. Tracton, 1995: Short-range ensemble forecasting of hurricane tracks. Preprints, *21st Conference on Hurricanes and Tropical Meteorology*, Miami, 494-496.

Bender, M. A., R. J. Ross, R. E. Tuleya, and Y. Kurihara, 1993: Improvements in tropical cyclone track and intensity forecasts using the GFDL initialization system. *Mon. Wea. Rev.*, **121**, 2046-2061.

Lorenz, E. N.., 1963: Deterministic non-periodic flow. *J. Atmos. Sci*., **20**, 130-141.

Neumann, C. J., 1972: An alternate to the HURRAN tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 22pp.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379-398.