P1.31 ENSEMBLE FORECASTING OF TROPICAL CYCLONE INTENSITY

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 the initial conditions result in increasingly large changes 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 given the observations.

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 deviations 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 deviations 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 intensity prediction remains one of the most difficult problems in weather forecasting. Only recently have intensity forecasts shown any skill. Estimates of the uncertainty and the possible variability of such forecasts can be highly valuable in allowing forecasters to improve their predictions. 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). However, intensity forecasts from the GFDL model have not regularly shown skill. 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 GFDL model is initialized with the initial conditions corresponding to the various NCEP Medium-Range Forecast (MRF) ensemble members, with corresponding boundary conditions from the integration of that ensemble member. Thirty-three cases during the 1996 hurricane seasons have been completed; almost 30 from the 1997 hurricane season have been completed through the first week of September.

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

2. FORECAST VERIFICATION

Because the two mesh version of the GFDL model has lower resolution than the three mesh operational version, the GFDL ensembles do not represent the inner core structure of tropical cyclones well. Therefore, the model is unlikely to accurately represent the current and forecast intensity of the tropical cyclones. As a result, a statistical correction is made to the intensity forecasts from the two mesh version. The operational three-mesh GFDL and the two mesh control runs are compared, since they both are run from the best (control) member of the MRF ensemble. The average ratio at each forecast time between the GFCT (GFDL ensemble control) and the operational GFDL is calculated, and this ratio is used to adjust the intensity of the ensembles. The ratio varies by forecast time (0.87, 0.81, 0.75, 0.84, 0.87 at 12, 24, 36, 48, and 72 h, respectively), but not by intensity. The ratios do not vary much from case to case.

The "control" analysis is considered the best representation of the initial conditions. As such, the control forecast should provide the best forecasts among the 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 relative errors of each ensemble member compared to those from SHIFOR, a simple statistical model based upon climatology and persistence (Jarvinen and Neumann 1979); those members of the ensemble showing negative relative error are said to exhibit skill.

The GFCT ranks among the top three forecasts on average, but the differences between none of the ensemble members with smaller average errors than the GFCT, and the GFCT, are statistically significant at the 90\% level based upon a paired t test with the null hypothesis that the mean of the differences is not significantly different from zero. The N5 ensemble member (GFN5) seems to perform better than the other members on average. This may be due to the relatively small number of cases in the sample, and does not reflect upon the relative value of the individual ensemble members. The lack of skill before 36 h is reminiscent of

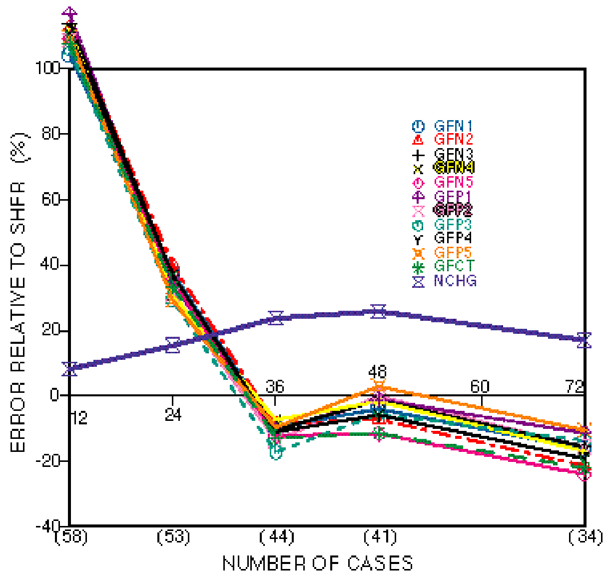

FIG 1. Relative errors of the adjusted GFDL ensemble intensity forecasts. NCHG represents a persistence (no change) forecast.

early tropical cyclone track forecasting with primitive equation models (DeMaria et al. 1990).

The ensemble mean is expected to perform better than individual members. Figure 2 shows the ensemble mean (GFMN) as compared to the control (GFCT) and the operational three mesh GFDL model. Neither the GFCT nor the GFMN performs as well as the GFDL in this sample except at 36 h. The GFCT is superior to the GFMN at all times except 48 h, and only at this time are the differences statistically significant at the 95\% level.

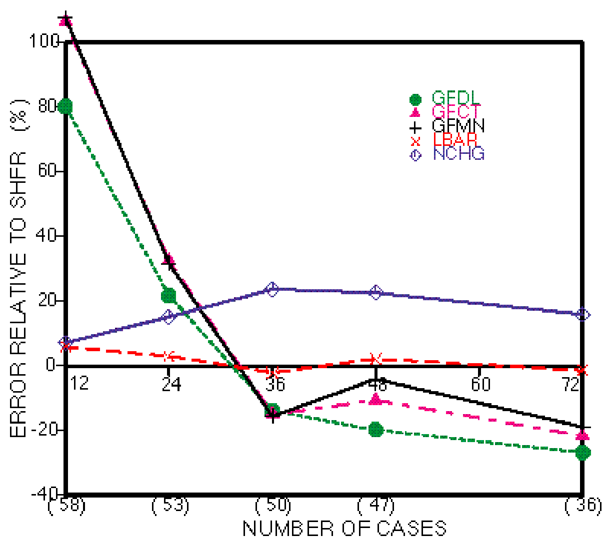

FIG 2. Comparison between the different operational tropical cyclone intensity models (Shifor [SHFR], Ships [LBAR], and GFDL), and the ensemble control (GFCT) and mean (GFMN).

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 by 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. Figure 3a shows the rank distribution from lower to higher forecast intensity for all 36 h cases. The figure shows a bias toward under-forecasting tropical cyclone intensity in the ensemble, and only a few cases are within the envelope of the possibilities presented. If the biases are removed (Fig. 3b), most forecasts remain in the wings of the distribution, though the distribution itself is less skewed toward forecasts of weaker tropical cyclones. This result shows that the usefulness of the ensemble is somewhat limited and that the perfect model assumption is not quite applicable. Results from other forecast times are similar.

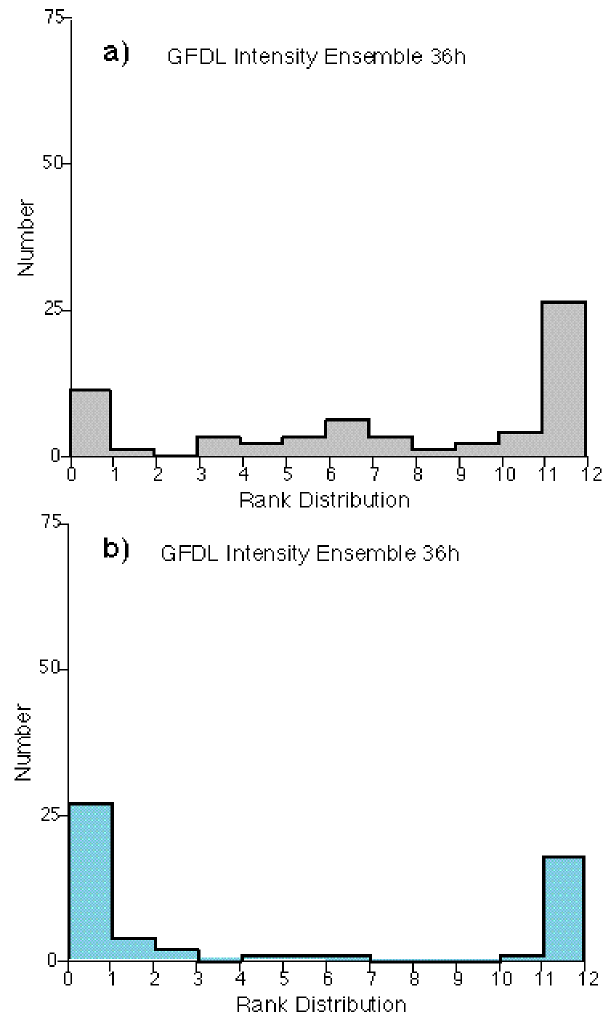

FIG 3. Rank distribution of the (a) GFDL ensemble intensity forecasts, and (b) GFDL ensemble intensity forecasts with the error bias removed.

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 is also large, 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 difference between ensemble members, and the actual error, at 36 h. There is little 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 especially if the verification is within the envelope of possibilities exhibited by the ensembles. Results at 48 and 72 h are similar. The relationship is not as strong at earlier time periods where the ensemble members show little overall skill.

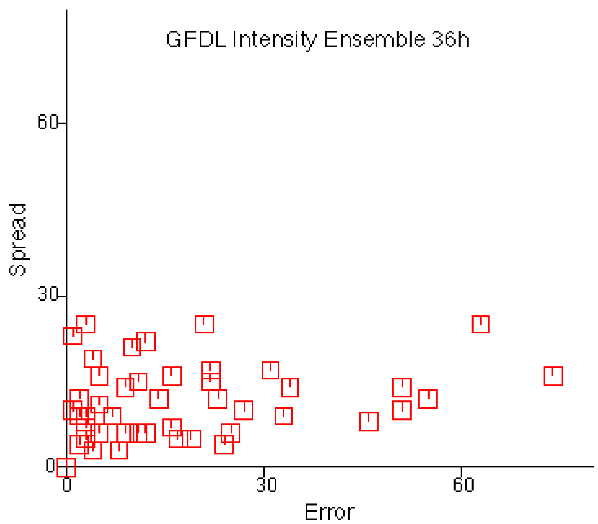

FIG 4. Relationship between forecast error (absolute value of the difference between the forecast and the verifying intensity) and the maximum spread amont he ensemble members for the GFDL intensity ensembles at 36 h.

5. CONCLUSIONS

Ensemble forecasting with the GFDL hurricane model shows great promise in improving tropical cyclone intensity forecasts. While the ensembles do not show forecast skill until 36 h, they are able to quantify the potential errors in each forecast since the ensemble spread and the errors seem to be related. Unfortunately, the spread among the ensemble members is not large enough to encompass the actual evolution of the tropical cylone intensity, and the model itself has deficiencies in intensity prediction, especially in cases of rapid intensification. Planned improvements to the vortex initialization in the GFDL model, and to the NCEP ensembles to increase the spread among the ensemble members in the tropics, will certainly help to improve the 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.

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.

DeMaria, M., M. B. Lawrence, and J. T. Kroll, 1990: An error analysis of Atlantic tropical cyclone track guidance models. *Wea. Forecasting*, **5**, 47-61.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecast of tropical cyclone intensity, NOAA Tech. Memo. NS NHC-10, 22pp.

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

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.