1. INTRODUCTION

In an age where reliance on operational forecasts is higher than ever by both the public and private sectors, the forecasts that are issued do not communicate a realistic sense of confidence to the end user. The current NWS graphically-based forecasts accessible to the public do not show this measure of uncertainty and thus communicate an often exaggerated sense of precision and confidence. The only forecast field that comes close to conveying a sense of confidence is precipitation, while other forecast fields are left as measures of deterministic forecasts without uncertainty.

Ensemble model data can provide a wealth of knowledge to forecasters especially in terms of forecast confidence. It has long been accepted that running an ensemble of numerical forecasts from slightly perturbed initial conditions can have a beneficial impact on the skill of the forecast (Toth and Kalnay 1997). A model run where members diverge corresponds to a low confidence forecast while a model run where members converge corresponds to a forecast of high confidence.

2. METHODOLOGY

One can supplement an ensemble forecast with a map of the standard deviation among the forecast members. When the spread among the forecasts is small, the average of the ensemble forecasts is probably accurate (Sivillo et al. 1997). The method that is being used here is similar, but it takes the previous method one step further.

Starting in August 2004 and extending into 2006, analysis of individual global GFS ensemble data is occurring as part of a COMET cooperative project with the NWS Office in Tallahassee. A climatology for each GFS ensemble member is being developed as a function of meteorological variable, location, time of year, and forecast length. Consistent with the work of Krishnamurti et al. (2000), a 45-day period centered on the initialization time is used to describe the climatology of a given day.

Once the normalized climatology distributions are calculated, forecast confidence/uncertainty measures can be developed from comparing the normalized spread of the real-time GFS ensemble members to the average spread of the GFS ensemble climatology. This normalized spread will also be compared to the typical spread for that time of year and location to arrive at a relative measure of forecast uncertainty. If the current model ensemble uncertainty is greater (less) than the uncertainty of the model ensemble climatology, then there is a lower (higher) than normal confidence.

3. CONFIDENCE IN A GRAPHICAL SENSE

Below is an image of the Binghamton, NY two-meter temperature confidence time series from the GFS run initialized on 1200 UTC 6 May 2004.
The blue line in the image shows the average GFS ensemble spread for this time of year. It is apparent that the average ensemble spread for this certain time of year increases with forecast length, from about 1F at initialization to around 6 to 7F at the 180 hour forecast. This is expected because forecast error generally increases with forecast length when averaged over long periods of time. The current GFS ensemble is shown as the black line. In this image, the forecast spread is generally near the normal line (blue line) until around May 8th.

From about May 8th until late on May 11th, the current forecast spread is LESS than the normal forecast spread for this time of year. The area between the normal spread and the current forecast spread is shaded in green. Thus, there is MORE agreement among the ensemble members than is normal for the time of year and forecast length. With all else equal, this means higher confidence in the forecast and less sensitivity to the uncertainty in the initial conditions. The yellow line shows the one standard deviation variability of the 25-year NCEP/DOE Reanalysis-2 (Kanamitsu et al. 2002) 2 meter temperature climatology for Binghamton. Before May 12, the GFS ensemble envelope is smaller than the climatology’s envelope, which is expected. However on May 12th and beyond, the spread of the GFS ensemble forecasts is actually larger than the observed climatology. It can be concluded that for May 12 and beyond, we should have no faith in the ensembles including their operational run since their envelope of solutions is broader than what a basic observed climatology gives us. In this case, it would probably be better to go with a climatology forecast on days 6 and 7.

It is found that the observed spread line (yellow line) will be higher in the winter months than in the summer months since the potential range of temperature in the winter is generally larger than that in the summer months. The only other forecast field for which confidence measures have been developed here is 10m wind. However, forecast confidence measures for precipitation and precipitation type are forthcoming.

4. RELATIONSHIP BETWEEN GFS CONFIDENCE AND HUMAN ERROR

It is implied from this research, although not yet proven, that human forecasters should perform less skillfully during times when the GFS forecast confidence is decreased (e.g. after 12 May in Figure 1). Such a relationship is necessary to produce a meaningful quantitative product on forecast confidence. Before continuing with this research, it is imperative to demonstrate such a relationship, if one indeed exists.

To determine the significance of the relationship between forecast confidence and forecast error, numerous NWS forecasts of two-meter temperature were verified in areas of low GFS confidence and high GFS confidence over the period February to July 2005. Thus far, approximately 254 NWS forecasts during below normal confidence were verified, while 131 NWS forecasts during above normal confidence were verified. The average NWS error for the below normal GFS confidence forecasts was 5.56°F while the average NWS error for the above normal GFS confidence forecasts was 3.64°F. Standard deviations of the errors were 3.31°F and 2.45°F, respectively.

After performing a student t-test on these values, it is found that there is a statistically significant difference to 95% confidence of the mean forecast error during low and high confidence GFS forecasts. That is, the mean human forecast error is significantly increased during times of low forecast confidence in the GFS ensemble. Therefore, forecasters have a-priori knowledge of the likely human forecast error when they see the GFS ensemble output—before the NWS forecast even verifies. Thus, it is suggested that forecasters use the GFS forecast confidence plots (e.g. Figure 1) to determine the periods of forecast time when more guidance analysis should occur. This relationship will be further examined as a function of forecast length to improve the significance of the results.

5. CASE STUDIES

A number of case studies have been performed in order to diagnose where areas of high and low confidence occur on the synoptic scale. Case studies commenced in mid March when a few nor’easters were examined to determine the behavior of forecast confidence. Other synoptic setups that were examined include strong springtime cold fronts and tropical cyclones Arlene and Dennis (2004). Several of these cases will be examined in detail.
6. SUMMARY AND FUTURE WORK

Only looking at the standard deviation of ensemble members will not give you a clear cut view of the actual confidence in a model forecast. The standard deviation of a forecast model must be compared with the standard deviation of the model climatology in order to take seasonal effects into account, as well as the natural degradation of forecast accuracy with forecast length. This comparison will lead to a more accurate view of forecast confidence and eventually a more valued forecast once the confidence values are shown graphically to the public. Indeed, it is expected in the coming year to convert the graphical output currently on display at the URL listed earlier into netCDF grids capable of being ingested into IFPS to produce NWS graphics of forecast confidence to accompany the deterministic forecasts already produced.

Early in this 3 year long project, only confidence values have been calculated for the GFS ensembles. Feedback from different NWS employees has suggested calculating confidence values for all the forecast models used in preparing a forecast. Essentially, this would give confidence values of the ever popular poor man's ensemble. A poor man's ensemble is a set of independent numerical weather prediction (NWP) model forecasts from several operational centers (Ebert 2001). Additional confidence values for different models, including the suggested “poor man's ensemble”, will be looked at in the coming years.

Further work will include expanding the GFS model climatology to produce a more accurate standard deviation climatology to compare to the real time GFS standard deviation. More than 2 years of ensemble data are needed to accurately estimate the spread distribution. In order to effectively utilize the information content present in the ensemble spread, a long record (10-15yr) of ensemble integration with the operational ensemble forecast system may be needed as suggested by Whitaker and Loughe (1998).

7. ACKNOWLEDGEMENTS/REFERENCES

This project is a joint COMET funded project (UCAR Award No. S04-44696) between The Florida State University and NWS Tallahassee. The first author would like to thank Robert Hart for invaluable insight in addition to Walt Drag, Irv Watson and the other SOOs whom have provided helpful feedback. Helpful insight into the use of the Graphical Forecast Editor (GFE) was given by Ken Gould of the NWS Tallahassee.


