University: Florida State University

Name of University Researcher Preparing Report: Robert Hart

NWS Office: Tallahassee

Name of NWS Researcher Preparing Report: Andrew I. Watson

Type of Project (Partners or Cooperative): Cooperative

Project Title: The development of forecast confidence measures using NCEP ensembles and their real-time implementation within NWS web-based graphical forecasts

UCAR Award No.: S04-44696

Date: 12 February 2008

Section 1: Summary of Project Objectives

The proposed research sought to produce scientifically and practically useful measures of forecast uncertainty/confidence for use by both NWS forecasters and the public. These measures were planned for the forecast fields of temperature, wind, vorticity, and if resources permitted, precipitation. Through the quantification of the forecast confidence, forecasters would be able to spend their shifts more efficiently, through focusing of the forecast refinement on periods when the forecast confidence was less than "normal" – through application of their educational and forecaster training. In short, there is no point in focusing the forecast shift on a forecast where the entire model ensemble agrees to a high degree when compared to a climatology of model spread. Forecasters should be able to focus their time on forecast periods when the model guidance was in less agreement, although not extreme disagreement.

Section 2: Project Accomplishments and Findings

Initially, several measures of confidence were tested for anticipation of human and guidance forecast error. Among these were the simple standard deviation of the GFS ensemble, comparison of this to the climatological model spread, and then de-biased versions. Ultimately, the best predictor for the future forecast error was the normalized ensemble standard deviation, as compared to model ensemble climatology. As a result of the initial research, as defined at length in Andrew Durante's thesis (Durante 2007), the formula for confidence is the simple subtraction of the current de-biased model ensemble standard deviation from the ensemble climatological standard deviation. The result is a measure of confidence that is the number of standard deviations from "normal spread" . "Normal spread" is defined as the average standard deviation of the GFS ensemble for that gridpoint, day of year, and forecast length (e.g., 0hr, 120hr) using a 45-day window centered on the day. A 45-day window centered on the initialization time is necessary as we have access to only a few years of ensemble data. Forty-five days is sufficiently long to adequately measure the variability of "weather" for that time of year, but not so long as to begin overlapping into other seasons that may not represent the nature of weather patterns around that time. This is comparable to the temporal extension employed by Krishnamurti et al. (2000) for Superensemble optimization, also given the limited archive of model performance as well as the evolving nature of model physics and resolution over time. The development of the confidence index reveals that in order to optimize your prediction of forecast error, it is necessary to take into account the climatology for the region. What is viewed as a “high confidence” or “low confidence” weather regime is dependent on the time of year, forecast length, and most importantly, location. Given forecaster experience in said location, this experience could be exploited in light of these confidence measures to improve upon simple climatology forecasts in the medium range.

These confidence products are currently produced for 2-meter temperature, 10-meter wind, 10-meter vorticity, and (nearing completion this month) precipitation thresholds, such as 0.25”, 0.5”, 1”. For every model run of the GFS ensemble, maps and movies of this confidence measure are provided in near-real-time at http://moe.met.fsu.edu/confidence. An example of such a map is given below:

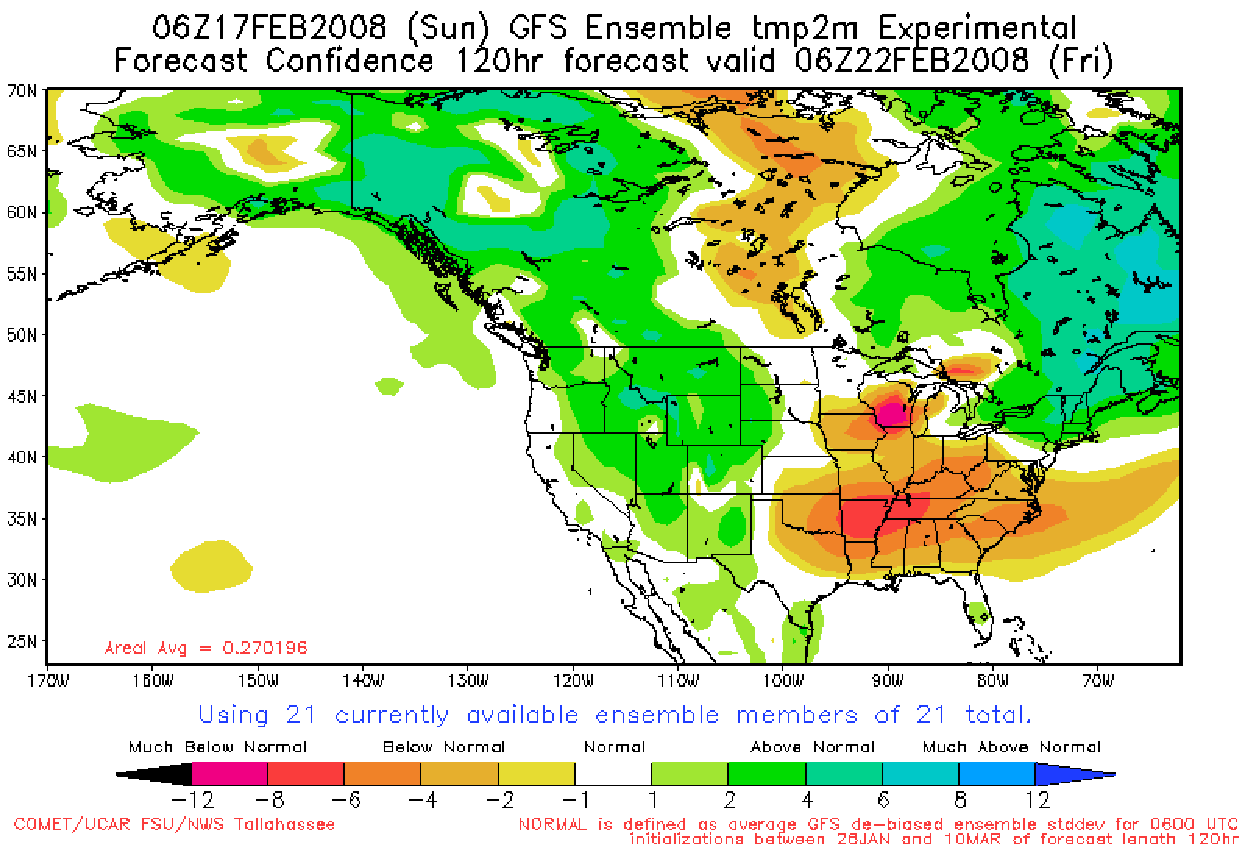

Figure 1: 5-day forecast confidence for 2-meter temperature. Areas of cool colors have considerably above normal confidence while areas in warm colors have considerably below normal confidence. For

example, in regions of the Southeast U.S., the model ensemble spread is unusually broad when compared to 5-day forecasts over 3 years using a window of 45 days centered on February 17.

In addition to the maps provided above, there are also timeseries produced for dozens of cities nationally and internationally. An example of this timeseries is provided below, for the core of the 5-day uncertainty in the map above, at Little Rock, AR:

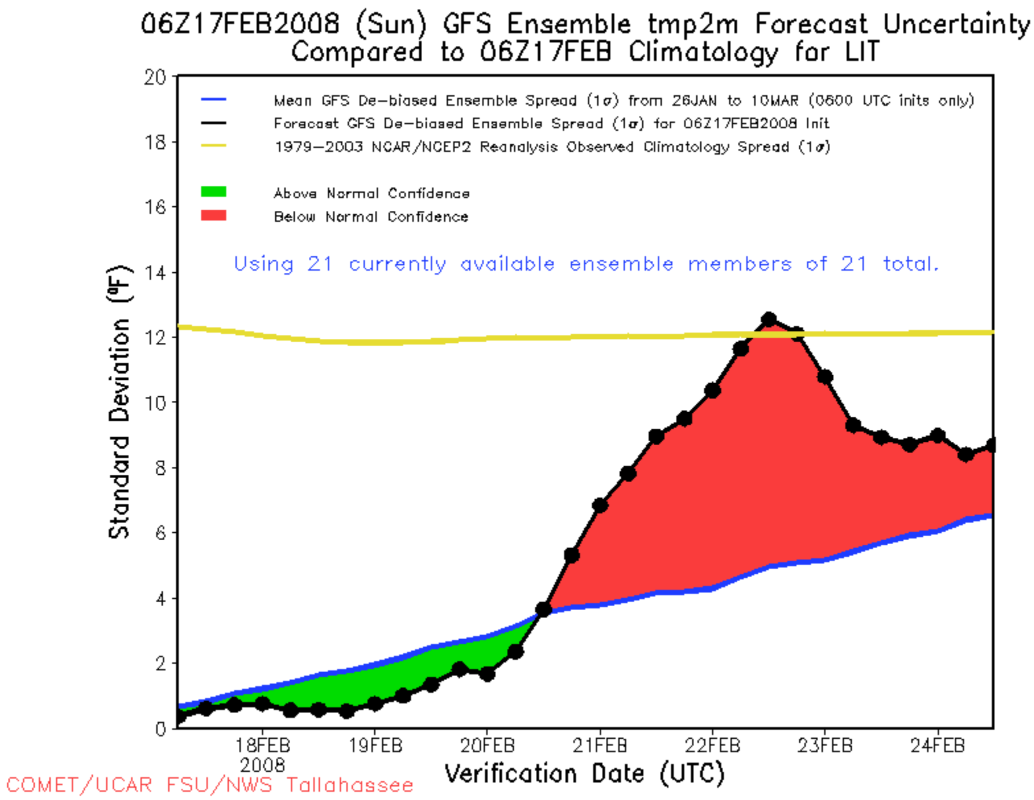

Figure 2: Timeseries of 2-meter temperature forecast confidence for Little Rock, AR. The black line represents the standard deviation of the GFS model ensemble initialized at 0600UTC 17 February 2008. The blue line is the climatology of the GFS ensemble standard deviation, which expectedly shows increasing spread with forecast length. The regions of green are above-normal confidence, while regions of red are below-normal confidence. The yellow line is the simple standard deviation that results from examining 30-years of historical observations for Little Rock. The role of this threshold in the forecast setting is discussed later. This forecast shows that there is unusual agreement among the ensemble members for the first 72hr of the forecast, which rapidly evolves into extreme disagreement through 168hr.

It is worth noting that during the period of this project, the foundation of ensemble generation at NCEP was changed. Initially, the breeding method (Kalnay and Toth 1993, 1996) was used to produce approximately a dozen ensemble members. With the breeding method, ensemble members (e.g. n2 or p2) and their forecasts could be de-biased since the perturbations had the same approach from run to run. However, with the shift to the Ensemble-Transform Kalman Filter method (Bishop et al 1999) halfway through the project, this was no longer true. A bias no longer had meaning since the resulting perturbations were not produced in the same manner from run-to-run for a given ensemble member (e.g., p0 or p17). Further, the number of ensemble members increased to 20. As a consequence of this, the measure of climatological ensemble forecast confidence in the confidence equation changed over the period of the project. However, preliminary tests on using a two-year period of the new ensemble generation approach (2006-2007) shows that the climatological GFS ensemble confidence is altered by less than 10% in most areas. Thus, the existing model ensemble database from 2004 through 2006 adequately quantifies the climatology of the current GFS ensemble system. The GFS ensemble climatology will continue to be revised as additional model output is archived, however, providing an increasingly robust measure of climatological forecast confidence.

The timeseries in Figure 2 shows a rare, but important, circumstance where the forecast confidence actually is sufficiently poor that the ensemble spread (black line) exceeds the observational history spread (yellow line). In these situations, one must ask whether there is so much uncertainty that forecasting observational climatology (20 or 30-year mean) would lead to a more accurate forecast than the ensemble mean. For approximately 40 cases of such events, this question was addressed also as part of Durante's thesis. On average, the forecast error when using the 27-year observational mean (in actuality, the 1979-2005 NCEP/NCAR2 reanalysis) as forecast was 9.6F for the 40 cases. In comparison, when the GFS ensemble mean was used as the forecast for these 40 cases, the average error was 13.2F. Although additional cases are needed to further refine this result, it does appear that an additional important result from this research is that there is a threshold beyond which it is best for the forecaster to simply use observational climatology as a forecast – unless the forecaster has unique and confident knowledge that the verification is likely to be one side of the ensemble mean. Further, the difference in error just described (9.6F vs 13.2F) argues that the threshold may in fact be a fraction of the 30-year observational spread. Further study on this temporal "window" of confidence is ongoing and is likely to further optimize the operational forecast environment.

We have received numerous emails over the past three years concerning the research, with requests to expand the web page geographically and meteorologically. As of February 2008, list of NWS Forecast Offices utilizing the online research results based upon emails, web requests, and web logs:
Tallahassee, FL
State College, PA
Taunton, MA
Caribou, ME
Portland, OR
Wilmington, NC
Salt Lake City, UT
Fairbanks, AK
Gray, ME
Charleston, WV
Charleston, SC

Binghamton, NY
Denver/Boulder, CO
Juneau, AK
Columbia, SC
Burlington, VT
Reno, NV
Buffalo, NY
Green Bay, WI
Wichita, KS
Goodland, KS
Cheyenne, WY

Regional Offices/Centers:
Forecast Systems Lab (FSL)
Storms Protection Center (SPC) / National Severe Storms Lab (NSSL)
Hydrologic Prediction Center (HPC)

International Organizations:
Environment Canada

Not e that over half of the above offices sent email to Hart, Durante, or Watson, requesting their regions be added to the timeseries output on the web page. Further, several have requested other fields on the confidence output (such as precipitation amount), and thus the pending precipitation confidence graphics will be in heavy use when implemented this Spring.

As an additional part of his thesis, Andrew examined a half-dozen cases within the confidence analysis. These included a mesoscale convective complex in the Midwest which lead to a plume of decreased forecast confidence of growing areal coverage, a front in the Southeast US, a Nor'Easter whose track uncertainty was well-captured by the vorticity confidence, and a hurricane in the Gulf of Mexico whose bi-modal track solutions were also captured well by wind and vorticity confidence plots. In each instance, some measure of forecast confidence proved useful, and well-summarized dozens of ensemble runs into one graphic. These cases are examined in detail as part of his thesis, and will not be restated here for the sake of brevity.

As part of one of the case studies, however, Andrew ran an ensemble of MM5 runs using the initial conditions from each of the GFS ensembles. This was done as a sensitivity test to see if the measure of forecast confidence increased or decreased by changing the model physics, and keeping the initial conditions the same as the GFS counterparts. The results, which are examined in detail in the M.S. thesis, show that a shift in the measure of forecast confidence is achieved by changing the model physics. Although one case does not prove any conclusion, it may argue for further evolutions of the SREF approach at NCEP, to provide further regional ensembles to further test the sensitivity of model forecast to various changes in initialization, physics, resolution.

Although the value of forecast confidence measures proved themselves routinely over the past three years of the project, as evidenced above and in Andrew's thesis, the potential value seemed to vary considerably from WFO to WFO. An examination of the human forecast performance at WFO in high vs. low confidence settings for a several-month period revealed considerable variability from office to office. Some component of this variability may be attributable to the aggregate forecaster experience within a WFO (not shown). However, this relationship was not consistent. In many cases, even with substantially above normal forecaster experience, the challenge in low-confidence forecast settings was not overcome. This suggested, in part, that some forecast regions have potentially greater challenges meteorologically than others. Such challenges could result from strong mesoscale forcing that may not be adequately captured by coarser resolution global models. Examples include Lake-Effect precipitation, sea-breezes, enhanced radiational cooling, and urban heat-island effects. These drawbacks may be alleviated by production of forecast confidence measures using higher resolution ensembles, such as the SREF, as has been pursued by Richard Grumm (SOO, CTP) at http://eyewall.met.psu.edu.

In early 2008, NCEP discontinued the production of GRIB1 format, and now exclusively produces GRIB2 format for its grid dissemination. Given we use GrADS for the real-time web generation, changes had to be implemented to ensure the confidence products could continue to be produced—as all versions of GrADS to that point could not read GRIB2 format. In late 2007, we installed and implemented scripts to convert the GRIB2 format to GRIB1, in particular using the cnvgrib program from NCEP. Although this worked fine, it was a huge resource hog given that we had to convert all 21 ensembles four times a day, and out to 180hr forecast. Nonetheless, the conversion was successful and confidence products continued without interruption in 2008. In January 2008, an alpha version of GrADS2 (which reads GRIB2 data) was made available from COLA/IGES, announced by Ms. Jennifer Adams. After a successful week-long rigorous test of GrADS2, the real-time confidence products and scripting were converted to use the GRIB2 data natively through GrADS2. As of late February 2008, the web page is now exclusively using GRIB2 data. This advance removes the need to convert the GRIB2 data to GRIB1, and frees up nearly 300GB daily of converted GRIB2 to GRIB1 data. This freeing of space also provides the opportunity (finally) to produce real-time precipitation forecast confidence products, as the climatological calculations for precipitation confidence require extensive disk space. As mentioned earlier, the precipitation confidence implementation is expected in early Spring.

Finally, netCDF and GRIB1 confidence grids are being produced in real-time as part of this project with the goals of ingesting into D2D for the forecaster environment. We have been working to correctly ingest these grids so that forecasters can graphically overlay the forecast confidence with other fields, in their native graphical environment. Once success is reached in the coming weeks, a README file will be added to the grids directory (http://moe.met.fsu.edu/confidence/grids) to explain how this can be performed for other web page users.

Section 3: Benefits and Lessons Learned: Operational Partner Perspective

The NWS continues to make deterministic forecasts through the IFPS/GFE system. However, there is more and more discussion from local offices regarding the need to express a measure of confidence in our forecasts. The SREF and MREF ensemble members and their various statistical products have given us a look at the variability of model output. More and more, our office is drifting away from comparing the individual operational models, to using ensemble and bias corrected model data. When possible, we include model guidance with bias corrections in GFE, based on recent weather during the past 30 days.

With the direction our office (as well as NWS other offices) is going, the Hart/Durante forecast confidence measures are just one more tool required to alert the forecaster of periods of both low confidence and high confidence. Forecasters continue to use the confidence measures via the web, and occasionally mention their usefulness in their Area Forecast Discussions.

We continue to strive to assimilate the Hart/Durante confidence measures into AWIPS. Because of difficulty in decoding netCDF files, Dr. Hart has graciously started to provide us with grib files, and our IT is now making good progress. Hopefully, he will be successful shortly.

Section 4: Benefits and Lessons Learned: University Partner Perspective

The collaborative relationship between FSU and NWS TAE continued to build upon the already strong foundation during the past three years. No less than a half dozen NWS employees gave seminars and invited class presentations at the Department of Meteorology, including Richard Pasch (TPC), Richard Knabb (TPC), Irv Watson (TAE), Jack Beven (TPC), James Franklin (TPC), Naomi Surgi (EMC), and Steve Lord (EMC). FSU Meteorology has benefited considerably from the co-instructoring of the Operational Meteorology Class taught by Irv Watson in Spring semesters. Further, FSU Meteorology has benefited considerably from the interaction with NWS employees nationally through email correspondence with these forecasters on the use of the confidence products. Prof. Hart has given COMET labs at the NORLAT conference three of the past four years in October, in part to present the results of this research. FSU Meteorology has further improved upon its already solid reputation in applied research as a result of this interaction. Also as a result of this successful collaboration, internal collaboration within FSU meteorology has occurred, with Hart and Fuelberg seeking collaborative funding on other NOAA-based grants, such as CSTAR and JHT. Further, the success of this project has lead, in part, to Hart being nominated as a member of the JHT steering committee.

Section 5: Publications and Presentations

The project resulted in one Master's thesis, one in-review Weather and Forecasting publication, and numerous conference presentations (both poster and talks) as outlined below. These documents are all available on the Hart lab website, http://moe.met.fsu.edu.

<u>Thesis:</u>

Durante, Andrew V, **2007**: "The Development of Forecast Confidence Measures Using NCEP Ensembles", M.S. Thesis, Florida State University, 117pp. Available at http://etd.lib.fsu.edu/theses/available/etd-07072006-115847

Refereed Publication:

Hart, R.E., A.V. Durante, and A. Watson, **2008**: The use and application of normalized ensemble spread to incorporate forecast confidence. *Weather and Forecasting*, Submitted March 2008, in review.

Presentations:

Hart, R.E., A.V. Durante, A. Watson, **2008**: The development of forecast confidence measures using NCEP ensembles. 19th Conference on Probability and Statistics. American Meteorological Society, New Orleans, LA, January 2008.

Hart, R.E., A.V.Durante, A. Watson, **2007**: Forecast confidence and a means to define the limits of predictability. 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction. Park City, UT, June 2007.

Hart, R.E., A.V. Durante, A. Watson, W. Drag, R. Grumm, **2006**: The development of forecast confidence measures using NCEP ensembles. 7th Southern New England Weather Conference, 24 October 2006.

Durante, A.V., **2006**: Forecast confidence. Conference Call, Southern Region. Audience: Robert Hart (FSU), Irv Watson (TAE), Richard Grumm (CTP), Walter Drag (BOX), David Novak (ER), Joshua Watson (ER), Bernard Meisner (SR), Henry Fuelberg (FSU), among others.

Durante, A.V., R. Hart, A. I. Watson, R. H. Grumm, and W. Drag, **2006**: The development of forecast confidence measures using NCEP ensembles and their real—time implementation within NWS web—based graphical forecasts. Preprints, 18th Conference on Probability and Statistics in the Atmospheric Sciences, Amer. Meteor. Soc., January 2006, Atlanta, GA.

Durante, A.V., R. Hart, A. I. Watson, R. H. Grumm, and W. Drag, **2005**: The development of forecast confidence measures using NCEP ensembles and their real—time implementation within NWS web—based graphical forecasts. Preprints, 21st Conference on Weather Analysis and Forecasting, Amer. Meteor. Soc., August 2005, Washington, DC.

Section 6: Summary of University/Operational Partner Interactions and Roles

Describe the responsibilities of the various project participants over the course of the entire project.

Robert Hart: Principal Investigator on the University Component. Responsible for advising Andrew Durante, guiding the conceptual approach to the project, setting up MM5 runs for case studies, and further development after Andrew's graduation in early 2007. Primary author of presentations after 2007 and of the manuscript submitted to *Weather and Forecasting* in 2008.

Andrew Durante: M.S. Student Advised on the project. Responsible for development and optimization of code to calculate the climatological confidence, web implementation, case study analysis, thesis writing and presentations prior to 2007.

Andrew (Irv) Watson (TAE): Principal Investigator on the Operational Partner Component. Responsible for encouraging forecasters to use the products and coordinator of data ingest into AWIPS.

Walter Drag (BOX) & Richard Grumm (CTP): Real-time evaluation and feedback of confidence products, suggestions for future improvement, and feedback on coauthored presentations. As a result of the success of this collaborative research, Grumm has implemented a regional SREF confidence page at NWS CTP that calculates and display higher-resolution but more limited-in-space measures of forecast confidence.