7.5 ENSEMBLE MOS FORECASTS FROM MULTIPLE MODELS

John Wagner* and Bob Glahn
Meteorological Development Laboratory
Office of Science and Technology
NOAA/National Weather Service
Silver Spring, Maryland

1. Introduction

Probabilistic forecasts are used to convey the uncertainty associated with the forecast of a given weather element. Several techniques have been described recently to express forecasts in terms of probability, including Hamill's analog forecasting technique (Hamill et al. 2006; Hamill and Whitaker 2006) and Bayesian Model Averaging (Mass et al 2009). The Meteorological Development Laboratory (MDL) of the National Oceanographic and Atmospheric Administration's (NOAA) National Weather Service (NWS) has developed a technique that creates probability density functions (PDF) and cumulative distribution functions (CDF) from an ensemble of members from a single model or from multiple models.

This technique, known as Ensemble Kernel Density Model Output Statistics (EKDMOS) (Glahn et al. 2009), uses the MOS technique (Glahn and Lowry 1972) to create forecast equations and error estimates for a particular weather element and forecast projection associated with a set of ensemble members. These errors and the single value forecasts from the members can be used to create an ensemble of distribution functions. These individual functions can be combined with kernel density fitting to arrive at the PDF and CDF for that weather variable and time projection.

The EKDMOS technique has been applied to the American and Canadian components of the North American Ensemble Forecast System (NAEFS). A sample was obtained that consisted of the control member and 20 perturbations from each component model of the ensemble. This paper shows the similarities and differences when the EKDMOS technique is applied to these two ensemble components and provides a comparison of their reliability and accuracy. In addition, the EKDMOS forecasts for both components of NAEFS have been combined into a single probabilistic forecast and the gain in accuracy and reliability achieved by using both instead of one independently are shown.

* *Corresponding author's address:* John Wagner, NOAA, 1325 East-West Hwy, Room 10352, W/OST24, Silver Spring, MD 20910; email: john.l.wagner@noaa.gov

2. Data

The NAEFS is a multinational ensemble prediction system (EPS) that combines the National Centers for Environmental Prediction's (NCEP) Global Ensemble Forecast System (GEFS) with the Canadian Meteorological Centre's ensemble prediction system (CMCE). Since July 2007, both models consist of a control member and 20 perturbed members. Both models contain a unified set of output fields so that they could be readily compared and combined.

In order to keep the number of members consistent for both models, a data set of available weather elements for use as predictors was obtained starting on July 11, 2007, and running through September 30, 2009. For development purposes, these data were then divided into 6-month seasons consistent with other MOS developments: two cool seasons (October 2007-March 2008 and October 2008-March 2009) and roughly two and a half warm seasons (July 2007-September 2007, April 2008-September 2008, and April 2009-Septempber 2009). All data were taken from the 0000 UTC runs of the GEFS and CMCE. These data were also converted into a format suitable for use with MDL software and comprise our dependent sample. Through previous testing, we have found that at least two seasons of dependent data are needed to generate reliable forecast equations. Since a third season of data was not available in time for this publication, independent data will not be shown.

Observations were obtained from an archive maintained by MDL. A set of 2280 stations was chosen for testing. These stations are located throughout the conterminous United States (CONUS), Canada, Alaska, Hawaii, and the U.S. territories.

3. The EKDMOS Technique

The EKDMOS technique uses screening multiple linear regression to generate forecast equations based on the mean of the ensemble members for the given weather element. The regression also provides an estimate of error variance. In this instance, temperature and dewpoint equations were developed simultaneously for each of the GEFS and the CMCE. Simultaneous development is a term used to denote selecting predictors in a way that ensures the same predictors are chosen for both elements, enhancing the meteorological consistency of the forecasts (National Weather Service 1985). The most frequently chosen predictors from both models include 2-m temperature and dewpoint, low level wind components and speed, thickness fields, and 2-m relative humidity. The predictands were temperature and dewpoint observations. Equations were made at 3-h intervals from 6 to 192-h.

EKDMOS forecasts were then generated for each ensemble member by applying the regression equation for each model to the 21 ensemble members associated with that model. Kernel density estimation (KDE) was then used to combine the forecasts with their corresponding error variance into a PDF. For verification purposes, each PDF was converted into a CDF.

As stated earlier, the regression equations and error variances for each model are based on the mean of the ensemble members. When they are applied to all 21 ensemble members of each model, the resulting PDF contains more spread than was accounted for in the initial regression. Figure 1a shows the average 80% credible interval (CI) for each model over two cool seasons. The average 80% CI is the average difference between the temperature associated with the 90th percentile (T90) on our CDF and the temperature associated with the 10th percentile (T10) on our CDF. The amount of dispersion is shown to be greater in the CMCE than the GEFS. To correct for the extra dispersion caused by applying the equation to each member, a spread adjustment factor (Glahn et al. 2009) was applied to each forecast. Since the GEFS and CMCE are shown to have different dispersions, different spread factors were needed to correct their dispersions. Through testing, values of 0.4 for the GEFS and 0.2 for the CMCE were found to provide the best adjustment. Figure 1b shows the average 80% CI for each model after the spread adjustment has been applied. At this point, the forecasts from both models contain approximately the same amount of spread.

4. Methods of Evaluation

There are several techniques for verifying the accuracy and reliability of a probabilistic forecast. The probability integral transform (PIT) histogram and cumulative reliability diagram (CRD) can be computed to show the reliability of a set of forecasts. A PIT is the percentile location of the verifying observation on the forecast CDF. A histogram, similar to a Talagrand diagram, can be created from these values, the shape of which can be used to identify biases, over-dispersion, and under-dispersion (Hamill 2001).

A CRD is a visual representation of cumulative departures of the probability forecasts from the observed relative frequencies associated with a probabilistic forecast. A line running diagonally across the plot acts as a reference of perfect reliability. It is evaluated like a reliability diagram in that relative frequencies (RFs) that fall to the left of the diagonal line denote under-forecasting (probability forecasts less than the observed relative frequencies) while RFs that fall to the right of the line show over-forecasting.

While PIT histograms and CRDs make it easy to visualize the reliability of a single model, it is not always easy to use them to compare the reliability of multiple models. Reliability can be more easily compared by calculating the square bias of the RF. This is done by computing the squared difference between the RF of each bin in a PIT histogram and unity, weighted by the width of the probability bin, summed over the entire range of probabilities (Glahn et al. 2009). These values can then be plotted by projection time, making it easy to compare several models.

The continuous ranked probability score (CRPS) is another single-valued score that can be computed to show accuracy. The CRPS is a squared measure of the difference between the CDF and the verifying observation (Unger 1985; Hersbach 2000). It takes into account both the accuracy of the median of the distribution (T50) and the amount of uncertainty in the forecast.

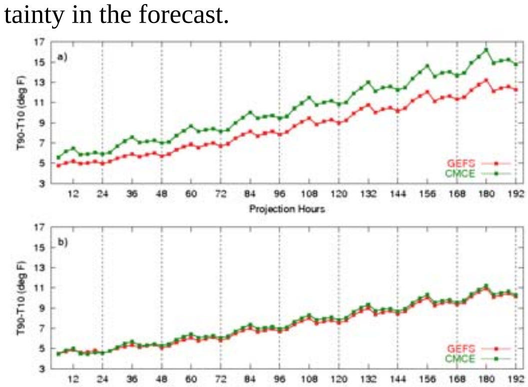

Fig. 1. (a) The average 80% credible interval (CI) without spread adjustment for the GEFS and the CMCE and (b) the average 80% CI after spread adjustment has been applied. The spread adjustment factor is 0.4 for the GEFS and 0.2 for the CMCE.

5. GEFS and CMCE Comparison

Plots a and b in figures 2, 4 and 6 show CRDs for the GEFS and the CMCE for cool season temperatures for 48-h, 120-h, and 168-h, respectively, which represent forecast days 2, 5, and 7. Plots d and e in figures 2, 4, and 6 show the same results for cool season dewpoint. Figures 3, 5, and 7 show PIT histograms for the same elements and projection hours. The CRDs show that applying the EKDMOS technique to the GEFS and CMCE yield very reliable forecasts, with a maximum departure from perfect reliability of 3.3% occurring at 120-h for the GEFS cool season dewpoint, as seen in Fig. 4d. The PIT histograms provide more information about where the differences between the RFs of the forecasts and perfect reliability exist. Perfect reliability on the PIT histogram is denoted by a dashed horizontal line at 1. Most histograms show a slight hump to the right of center, denoting some over-dispersion as well as a slight cool bias to the forecast. Images for the warm season are not shown here but were very similar.

The square bias in RF for the GEFS and the CMCE is compared in figures 8 and 9 for temperature and dewpoint, respectively. Note that all values are very low, showing little bias in agreement with the CRDs. The differences in the square biases of the two models differ by projection hour and element and are overall very similar. With the exception of the high values at the early projections for cool season dewpoint in figure 9a, the CMCE results are slightly lower in terms of square bias and therefore slightly more reliable than the GEFS.

The CRPS values for the GEFS and the CMCE are compared in figures 10 and 11 for temperature and dewpoint, respectively. This time the GEFS is slightly better than the CMCE at most projections. This is more noticeable for the cool season temperature and dewpoint than the warm season. Overall, the EKDMOS technique produces slightly more accurate forecasts when applied to the GEFS.

6. NAEFS

The NAEFS probabilistic forecasts were produced using kernel density estimation to combine the 21 GEFS ensemble members with the 21 CMCE ensemble members. Through testing, it was found that a spread factor of 0.2 was needed to correct for over-dispersion. Figures 12a and b show the temperature forecasts for the GEFS and the CMCE ensemble members for a 72-h forecast for Baltimore, Maryland with the PDF that was generated for each of them. No spread adjustment has been applied to the PDFs shown here. As discussed with Fig. 1a, the GEFS ensemble forecasts show less dispersion and uncertainty as compared to the CMCE forecasts, which have more dispersion and result in a PDF with more spread. Figure 12c shows how the same ensemble members were combined to form the NAEFS PDF. The result is a PDF that takes on the characteristics of both of the EPSs used to create it. Figures 12d, e, and f show the same results for Atlanta, Georgia.

Figure 2 compares the CRDs for 48-h cool season temperature and dewpoint for the GEFS, CMCE, and NAEFS while Fig. 3 shows the PIT histograms for the same data. Both figures show the NAEFS to have the same reliable forecasts that the GEFS and CMCE had individually. Figures 4 through 7 show that this reliability also exists at projection hours 120 and 168. Results for the warm season were very similar.

Figures 8 and 9 compare the square bias in RF for temperature and dewpoint respectively. Here again the NAEFS is shown to be as reliable as the GEFS and the CMCE. At some projection hours, the NAEFS is more reliable. Note again that all values are extremely low, indicating very little bias.

The CRPS scores are shown for the GEFS, the CMCE, and the NAEFS in figures 10 and 11 for temperature and dewpoint, respectively. Here the NAEFS forecasts are shown to be more accurate than the GEFS and the CMCE at every projection hour. This increase in accuracy is show for both temperature (Fig. 10) and dewpoint (Fig. 11).

7. Conclusions

The EKDMOS technique has been shown in the past to produce accurate and reliable probabilistic forecasts when applied to the GEFS (Glahn et al. 2009). When the same technique is applied to another EPS, the results are very similar. PIT histograms, CRDs, and square bias in RF and CRPS plots have all been used to show that the CMCE EKDMOS forecasts are reliable and accurate, with the GEFS being slightly more accurate and the CMCE being slightly more reliable.

The benefits of combining NCEP's GEFS with the CMCE to gain a more skillful probabilistic forecast have already been shown (Candille 2009). These benefits are also present when the EKDMOS technique is applied to the NAEFS forecasts. When these two ensemble systems are combined into a single EPS, the strengths that each system possesses are carried over into the combined system. The end result is a probabilistic forecast that is as reliable as and more accurate than the

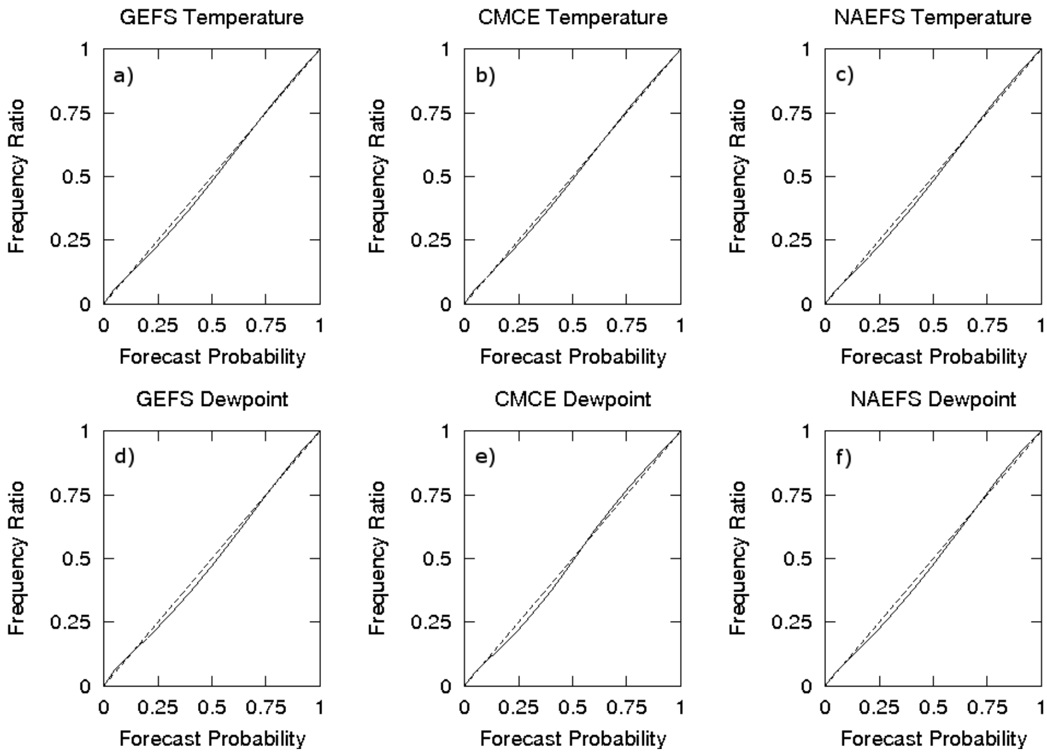

Fig. 2. 48-h CRDs for cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

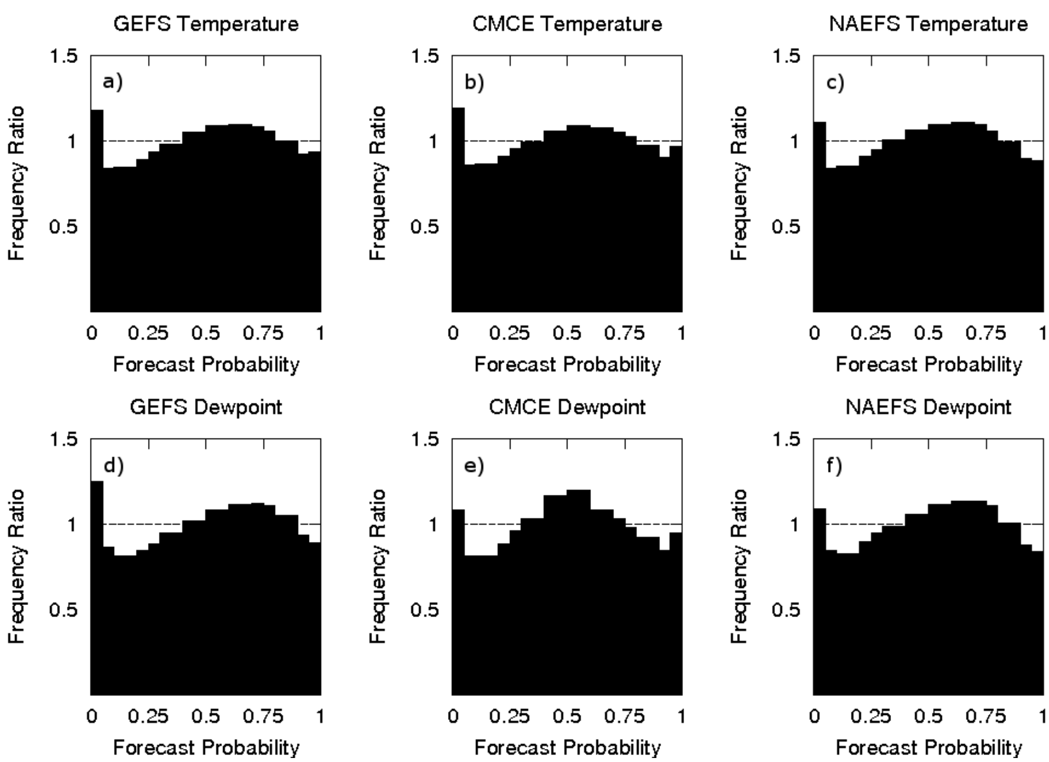

Fig. 3. 48-h PIT histograms for cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

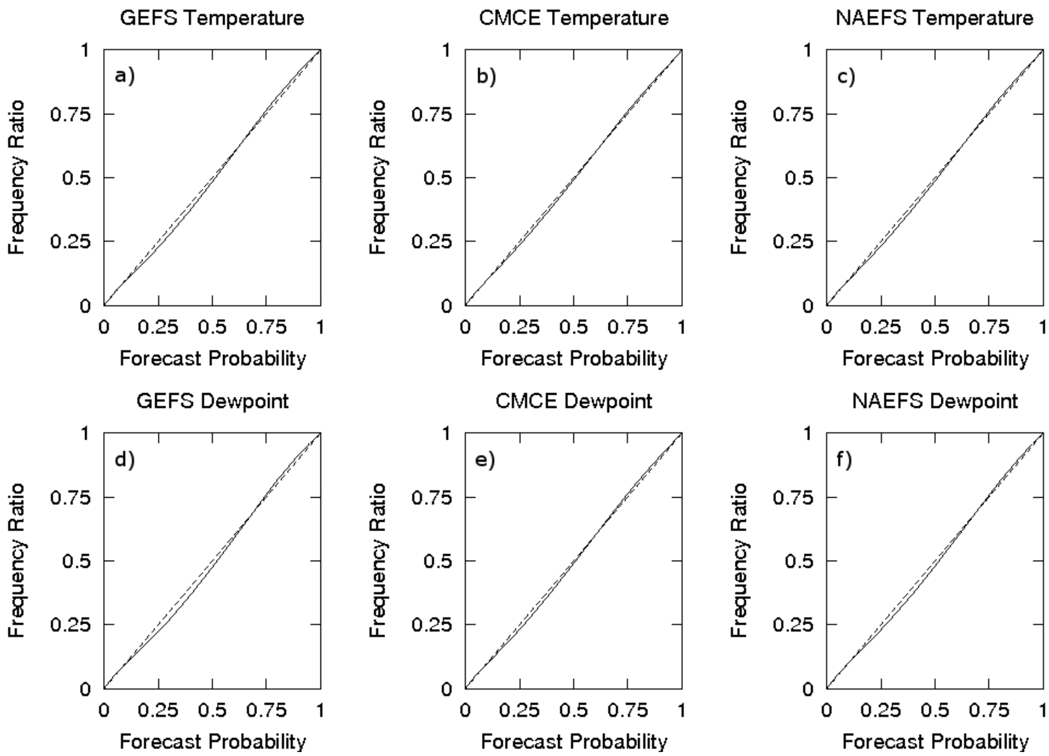

Fig. 4. 120-h CRDs for the cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

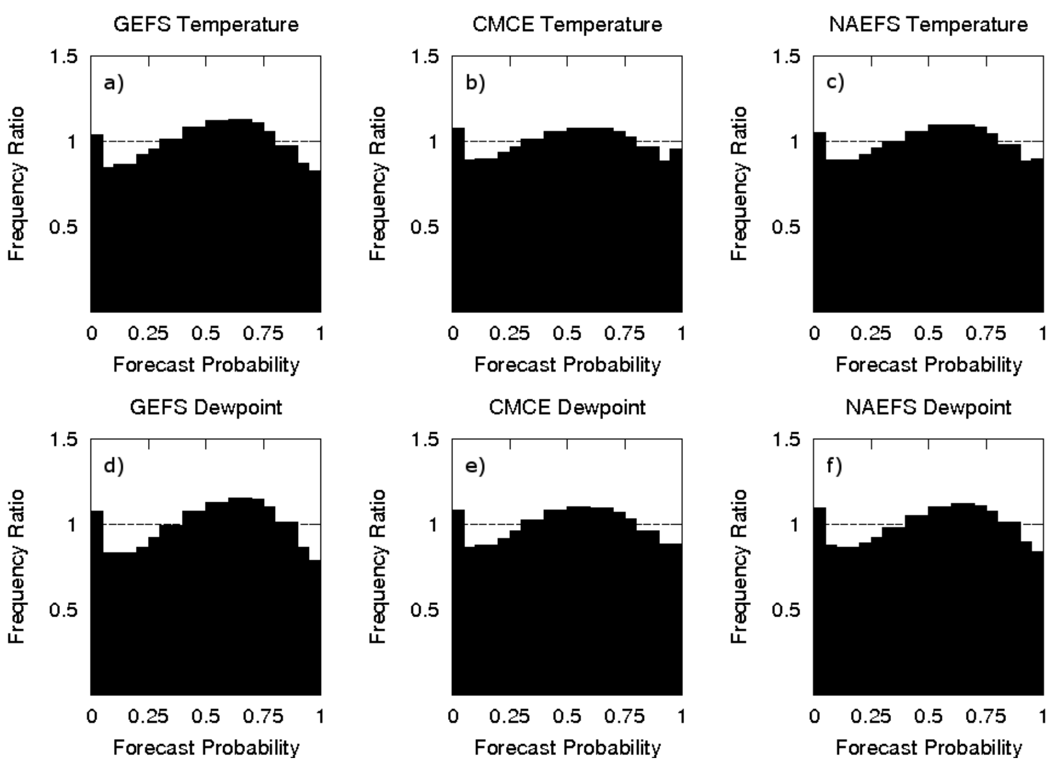

Fig. 5. 120-h PIT histograms for cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

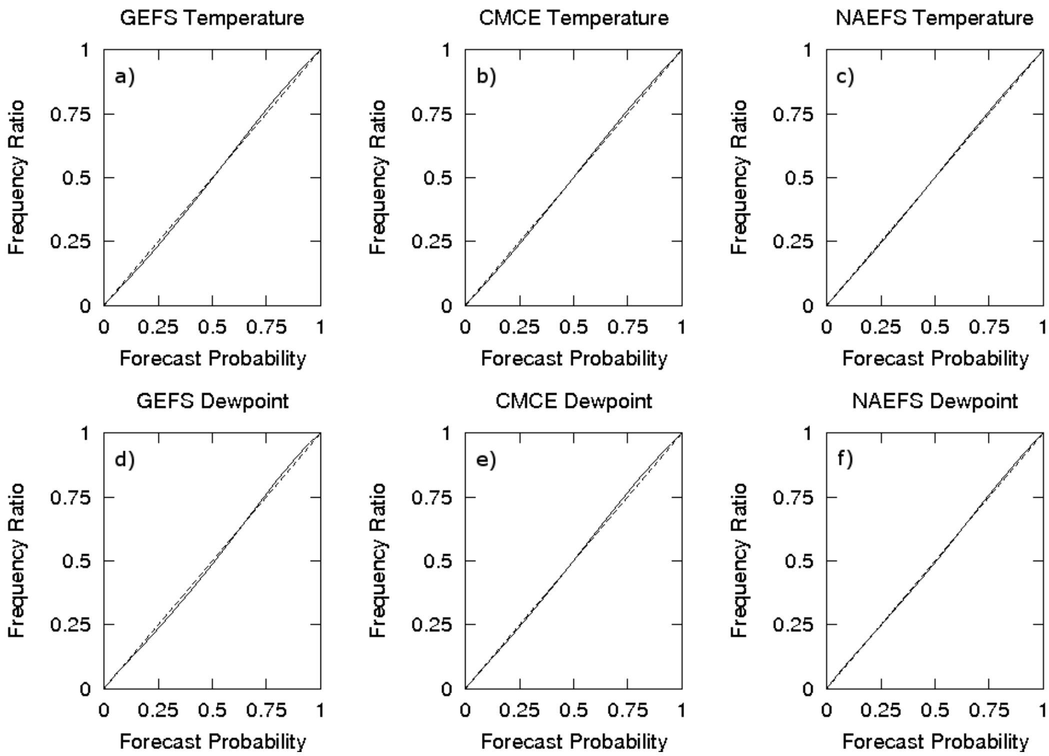

Fig. 6. 168-h CRDs for cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

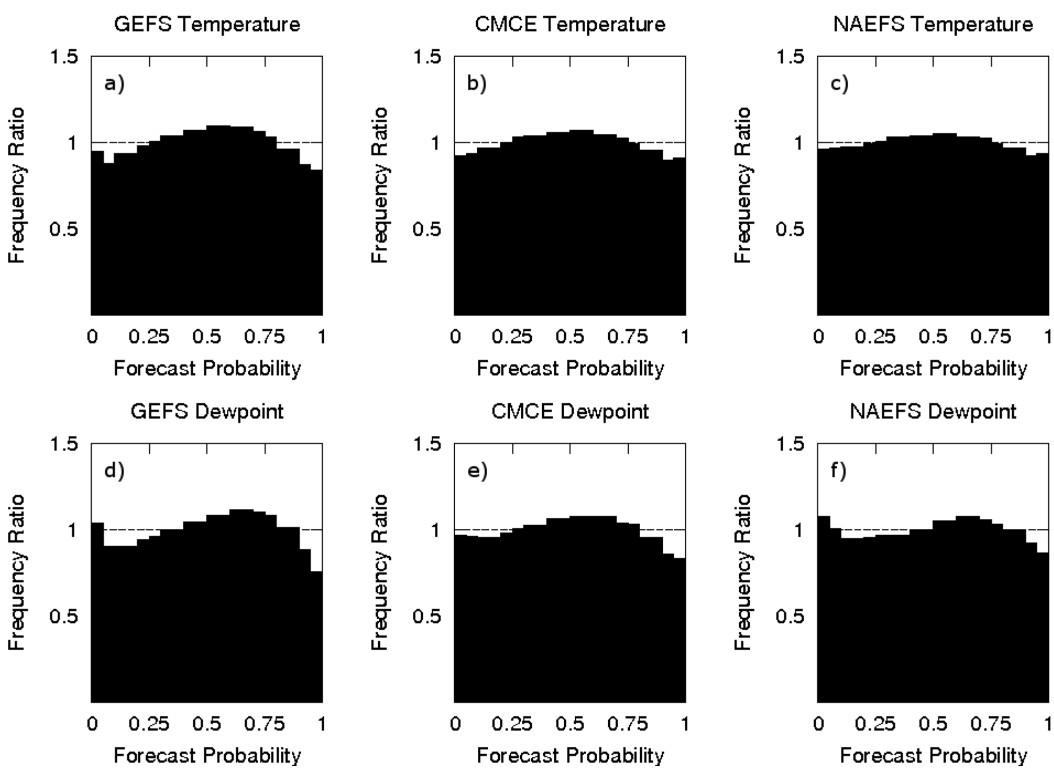

Fig. 7. 168-h PIT histograms for cool season temperature and dewpoint for the GEFS, the CMCE, and their combination into the NAEFS.

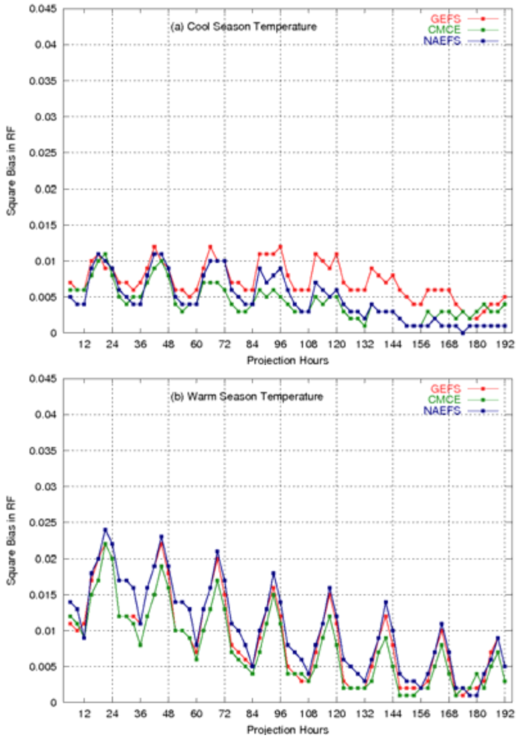

Fig. 8. Square bias in RF for (a) cool season and (b) warm season temperature for the GEFS, the CMCE, and the NAEFS.

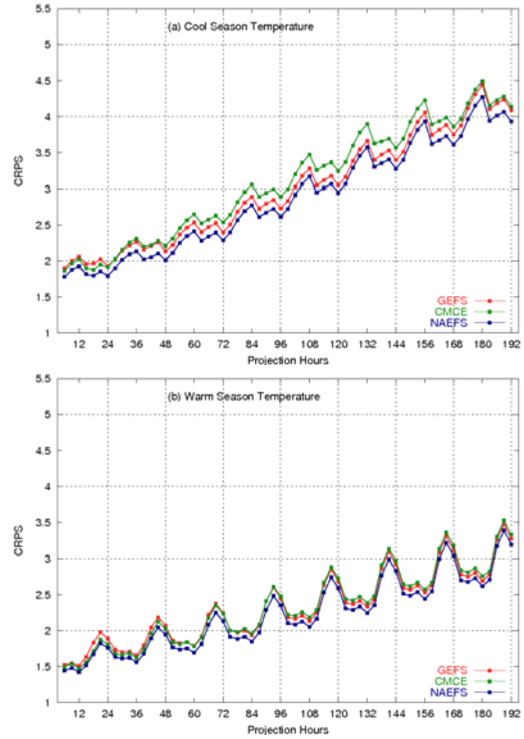

Fig. 10. CRPS in degrees F for (a) cool season and (b) warm season temperature for the GEFS, the CMCE, and the NAEFS.

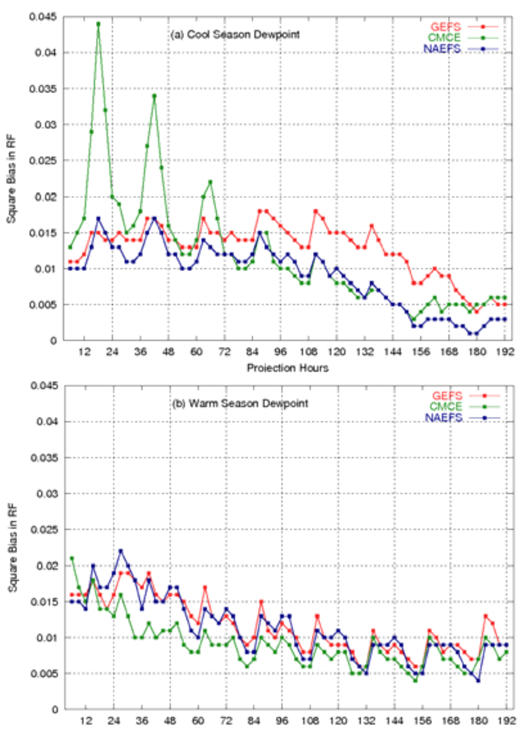

Fig. 9. Square bias in RF. Same as Fig. 8, except for dewpoint.

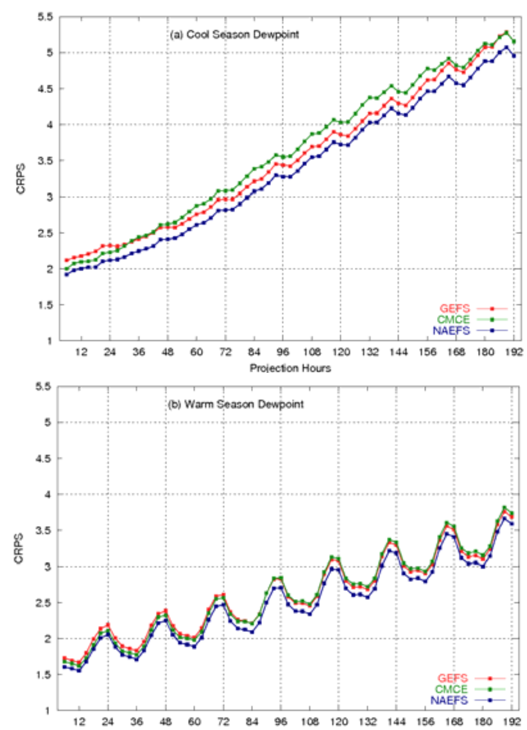

Fig. 11. CRPS in degrees F. Same as Fig. 10, except for dewpoint.

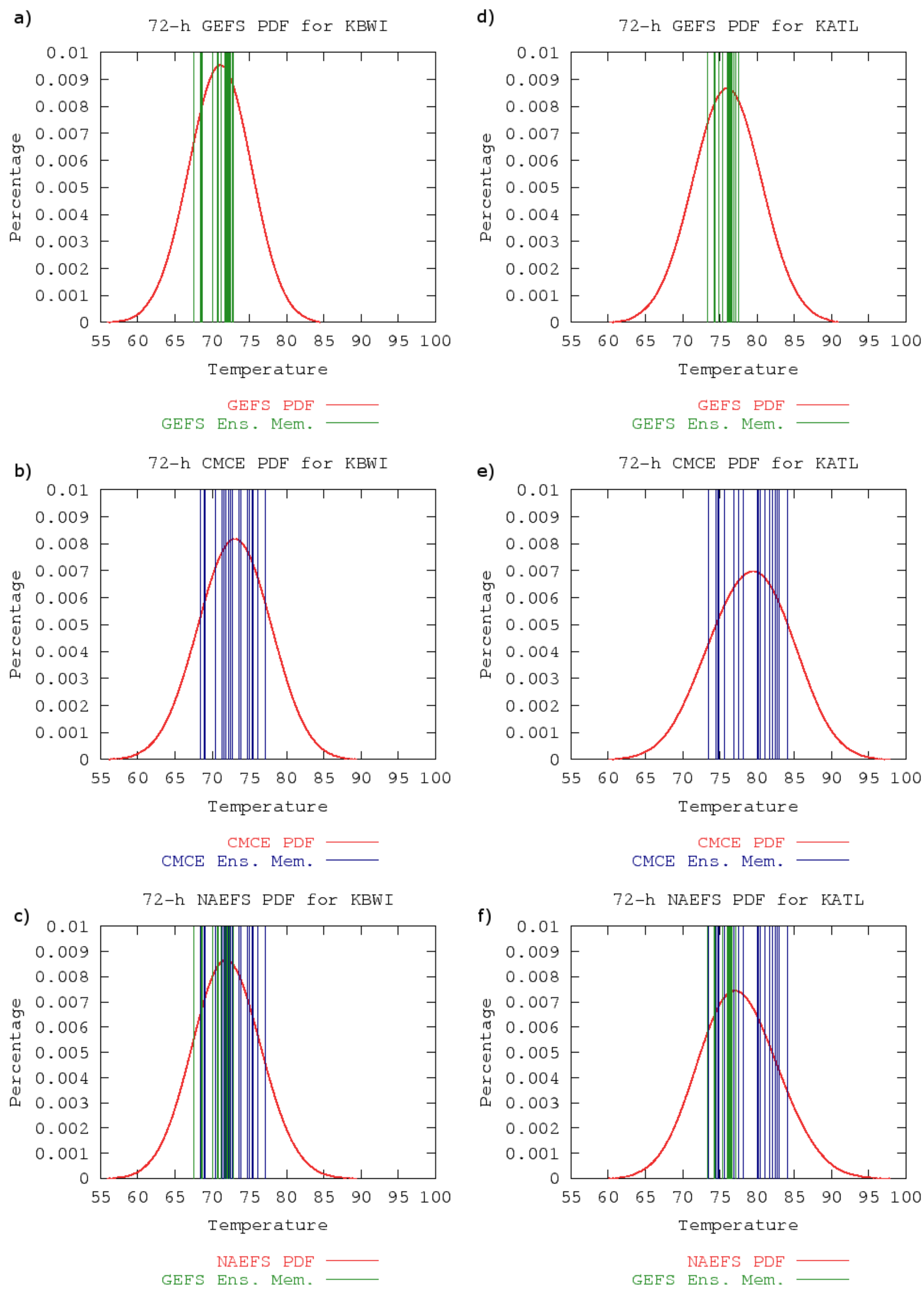

Fig. 12. PDFs for the GEFS, CMCE, and NAEFS for Baltimore, Maryland (KBWI) and Atlanta, Georgia (KATL) for the 72-h forecast on October 1, 2007.

individual ensemble forecast systems from which it was created.

8. Future Work

Results shown in the paper were for dependent data only. The next step will be to apply the equations and techniques to independent data. These data will start to become available in April 2010. Results were also shown for the 0000 UTC run of the NAEFS only. These same techniques can also be applied to the 1200 UTC run of the NAEFS. Glahn, et al. (2009) also showed the EKDMOS technique to be successful for maximum and minimum temperature forecasts. Forecasts for these elements will also be produced from the NAEFS model. The intent is to make these forecasts available to field forecasters through the National Digital Guidance Database.

9. References

Candille, G., 2009: The Multiensemble Approach: The NAEFS Example. *Mon. Wea. Rev.*, **137**, 1655–1665.

Glahn, H.R., and D. A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

----, M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS Uncertainty Estimates in an Ensemble Framework. *Mon. Wea. Rev.*, **137**, 246–268.

Hamill, T. M., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129**, 550–560.

----, J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: An Important Dataset for Improving Weather Predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.

----, and J. S. Whitaker, 2006: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application. *Mon. Wea. Rev.*, **134**, 3209–3229.

Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Wea. Forecasting*, **15**, 559–570.

Mass, C., S. Joslyn, J. Pyle, P. Tewson, T. Gneiting, A. Raftery, J. Baars, J. M. Sloughter, D. Jones, and C. Fraley, 2009: PROBCAST: A Web-Based Portal to Mesoscale Probabilistic Forecasts. *Bull. Amer. Meteor. Soc.*, **90**, 1009–1014.

National Weather Service, 1985: Automated daytime maximum, nighttime minimum, 3 hourly surface temperature and 3-hourly surface dewpoint guidance. NWS Tech. Procedures Bull., 356, NOAA/NWS, Silver Spring, MD, 14pp.

Unger, D. A., 1985: A method to estimate the Continuous Ranked Probability Score. Preprints, *Ninth Conf. on Probability and Statistics in the Atmospheric Sciences,* Virginia Beach, VA, Amer. Meteor. Soc., 206-213