ENSEMBLE FORECAST SYSTEMS AND MOS

Matthew R. Peroutka
NOAA/National Weather Service, Silver Spring, Maryland

1. INTRODUCTION

In the early 1990s, the Model Output Statistics (MOS; Glahn and Lowry 1972) technique and its forecasts were a part of the operational mainstream of NOAA's National Weather Service (NWS) and of a number of countries outside the US. In December 1992, the National Meteorological Center (NMC, predecessor to NOAA's National Centers for Environmental Prediction or NCEP) implemented the NWS's first Ensemble Forecast System (EFS; Toth and Kalnay 1997).

As both the MOS and EFS evolved through the 1990s, there were repeated calls for an "Ensemble MOS" product. The Meteorological Development Laboratory (MDL) of the NWS developed a product that applied MOS equations based on an operational deterministic model to the individual members of the operational EFS (Erickson 1996). Fig. 1 shows a typical visualization of one of these forecasts, as a box-and-whisker diagram. This method produced a distribution of forecasts with predictable strengths and weaknesses. The center of the MOS distribution was largely consistent with other deterministic MOS forecasts. The spread of the distribution provided some insight into the spread forecast by the EFS. Temperature forecasts for long lead times, however, tended to exhibit less spread than those at earlier lead times, and they were generally underdispersive. In short, this method provided no mechanism for spread calibration. This behavior is understandable. MOS forecasts generally trend toward the mean of the development sample as lead time increases and the forecast skill of the underlying model diminishes.

During the mid-2000s, Harry R. Glahn, the Director of MDL, formed a team to develop statistical postprocessing methods that were more suitable for EFS. The goal was to statistically postprocess EFS output to yield accurate and reliable probabilistic forecasts of sensible weather elements. Glahn, et al. (2009) documents many of the methods used and the results achieved by this team. This extended abstract will briefly summarize the work documented in Glahn, et al. (2009), note some of the improvements introduced after its publication, and provide some insight into MDL's ongoing work with the statistical postprocessing of EFS output.

2. METHOD

2.1 Application of MOS to EFS

Fig. 2 is a graphical depiction of the Ensemble Kernel Density MOS technique (EKDMOS). Glahn, et al. (2009) described the original technique as it was applied to the NCEP's Global Ensemble Forecast System (GEFS). Wagner and Glahn (2010) described enhancements that support the combination of multiple EFS with similar skill. Glahn, et al. (2009) described the need for a spread calibration step, and introduced a simple one. Veenhuis and Wagner (2012) and Veenhuis (2013) introduced an improved spread calibration technique that is more responsive to day-to-day changes in the component EFS.

The top portion of Fig. 2 shows the initial EKDMOS development step. As with other variations of MOS, developers gather a representative sample of EFS forecasts and observations of sensible weather parameters. The developers then use forward-screening least squares regression using the mean from each EFS. This regression step is repeated for each EFS. Like other MOS techniques, this step yields a forecast equation for each station/gridpoint, time projection, and weather element. Unlike other MOS techniques, this step also yields a confidence estimate that is computed directly as part of the regression process. This step expresses the confidence in a new forecast for a case drawn from the same population as the initial development. As one might expect, confidence in a new forecast is higher when that forecast is near the mean of the development sample, and that confidence is also higher when the spread of the development sample about the regression line is small.

Fig. 2 also shows two EFS, the GEFS and the Canadian Meteorological Center's EFS (CMCE). The technique has been applied to 3 EFS and more. Additional adaptations are required when the component EFS exhibit dissimilar skill.

The middle portion of Fig. 2 illustrates the spread calibration step which is a follow-on to the development step. Again, a representative sample of ensemble forecasts and observations are used to develop a spread-skill relationship for each EFS. In general, we have found that the spread taken directly from EFS is not a reliable indicator of day-to-day skill *when used directly*. We have found, however, that the EFS spread is *informative* in this regard. It is interesting to note that the relationship between EFS skill and EFS spread varies considerably with climatological regime.

The last step in the EKDMOS process, shown in the lower third of Fig. 2, is the implementation. This is where output from a single run of each EFS is gathered, postprocessed, and used to create forecast Probability Density Functions (PDF). In this step, the regression equations that were developed using the ensemble means are applied to each member of the EFS, yielding a number of varying forecasts. Kernel Density Estimation (KDE) is used to combine the various forecasts generated from all EFS into a single PDF. For this step, we chose to use kernels that have the shape of a Gaussian Distribution, and to use the confidence estimate from the regression to set the variance of each kernel. Once the forecast PDF is formed and normalized, the spread-skill relationship is used to adjust the spread of the PDF.

2.2 Presentation of probability forecasts

One of the many challenges associated with creating probabilistic forecasts of sensible weather elements is developing suitable methods to disseminate them. Fig. 3 shows three different ways to graphically depict a temperature forecast that is a probability distribution. Fig. 3(a) is a PDF. PDFs are intuitively appealing to many, and they quickly provide a user information about location, spread, and skew of the forecast distribution. That said, users are typically not prepared to extract quantitative information from a PDF. I.e., PDFs do not readily answer questions such as “What is the probability of freezing tonight?” “What is the probability of reaching 500 accumulated growing degree days tomorrow?” and “What is the probability of wind speeds too fast to operate a wind turbine next week?”

Fig. 3(b) shows the same distribution as a Cumulative Distribution Function (CDF), the integral of the PDF. While the CDF cannot be interpreted visually as easily as the PDF, a user can readily use it to answer the questions posed above. One can read the values of the weather element on the x-axis and the probability values that run from 0 to 1 along the y-axis.

Fig. 3(c) shows the same distribution as a Quantile Function. The Quantile Function is the same as the CDF with axes adjusted to increase its utility. The squares along the curve are placed at the probabilities 0.05, 0.10, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.75, 0.80, 0.90, and 0.95. EKDMOS forecasts are disseminated for almost all of these probability values. (We chose to omit 0.25 and 0.75 for our operational implementations. Testing revealed that a user could almost always extract sufficient information by interpolating between the two neighboring probability values.)

As of this writing the NWS produces operational EKDMOS forecasts twice daily (0000 and 1200 UTC model runs) for the North American Ensemble Forecast System (NAEFS; Toth, et al. 2006). The primary means of dissemination is grids of the percentile values described above. The EKDMOS grids are compatible with the National Digital Forecast Database (NDFD; Glahn and Ruth 2003), and the CONUS grids have a resolution of 2.5 km. Gridded forecasts for seven weather elements are issued. EKDMOS forecasts of temperature, dew point, daytime maximum temperature, and nighttime minimum temperature are classified as operational. EKDMOS forecasts of apparent temperature (a combination of wind chill temperature, temperature, and heat index), wind speed, and Quantitative Precipitation Forecast (QPF) are classified as experimental. Lead times vary among the weather elements. The longest lead times are for daytime maximum and nighttime minimum temperatures which extend to 16 days.

Fig. 4 and Fig 5 are both examples of graphical depictions of probabilistic weather forecasts made with EKDMOS. As noted above, EKDMOS forecasts are disseminated as 11 fixed percentile points along the quantile function. Ideally, users and partners will analyze these percentiles, interpolate as needed, and obtain reliable and accurate answers to their own questions. It is interesting (and, perhaps, sad) to note that almost all visualizations of probabilistic weather forecasts are forced to abandon eight of the 11 percentile points and use only the 10th, 50th, and 90th percentiles.

Fig. 4 shows a sample gridded probabilistic forecast of wind speed, plotted on the NDFD CONUS grid. The frames show the 10th, 50th, and 90th percentiles of forecasted wind speed at the 24 h lead time. Fig. 5 shows forecasts of daytime maximum temperature and wind speed as meteograms. Three colors encode the 10th, 50th, and 90th percentiles. The center portion of Fig. 5 uses stacked histograms to show the probability of pre-selected QPF categories.

2.3 Methods of evaluation

The evaluation of a single probabilistic weather forecast is a long-standing challenge. In one sense of the word, a probabilistic forecast cannot be “wrong” since the observed outcome was assigned some non-zero probability of occurrence. Of course, given a sufficient sample of forecast distributions and verifying observations, we can assess the performance of EKDMOS or any other probabilistic forecasting technique. As our team developed the EKDMOS technique, we also spent substantial effort developing appropriate evaluation techniques. These techniques focused on evaluating the first two moments of the forecast probability distribution--location and spread. Higher moments, such as skew and kurtosis, seemed to have less intuitive appeal and practical application for users. Moreover, our goal was to provide accurate and reliable forecast distributions. Location and spread both seemed to contribute directly to these two goals.

Glahn, et al. (2009) provides a thorough description of these evaluation techniques, useful references, and more than a few examples of their use. A few methods will be presented briefly here.

Assessing the first moment of the forecast distribution is quite straightforward. One selects a suitable metric for the location of the distribution and applies the same techniques typically used for single-valued forecasts. The mean and the median of the forecast distribution are the two obvious metrics for location. There seemed to be little difference between the two when evaluating EKDMOS and similar probabilistic forecasts. Mean Absolute Error (MAE) and bias are scores that seemed to be both simple and informative for probabilistic forecasts.

Assessing the second moment of the EKDMOS forecasts proved to be more challenging. Fig. 6 shows an assessment of forecast spread called a Cumulative Reliability Diagram (CRD). This plots the cumulative forecast probability against the cumulative observed relative frequency for a representative sample of forecasts. The CRD can visually identify departures from reliability. Ideally, the CRD curve will remain close to the diagonal reference line.

Fig. 7 shows another tool named the Probability Integral Transform (PIT) histogram. The PIT is the probability value of the CDF at the value of the verifying observation. For example, if the forecast CDF called for a probability of 0.20 of a temperature of 36 degrees Fahrenheit or lower, and the verifying observation was 36, then the PIT for this case is 0.20. To form a PIT histogram, one gathers a representative sample of cases, bins them according to PIT values, and computes the proportion of cases observed to the number of cases expected in each bin. Ideally, the histogram bars will align with the horizontal unity reference line.

Fig. 8 illustrates the computation of the Continuous Ranked Probability Score (CRPS). The left half illustrates the comparison of the forecast CDF (red line) with a unit step function of the verifying observation (green line). The shaded area in the right half is the difference between the two. The CRPS for a single case is the square of the shaded area. CRPS values for a representative sample of cases are averaged together. One can consider the unit step function an absolutely sharp and accurate forecast. CRPS is negatively oriented and positive. A perfect CRPS of zero means that all forecasts were accurate and sharp. There is no upper limit.

3. ONGOING WORK

Since 2013, much of the NWS’s effort in statistical post processing of EFS has been focused on the National Blend of Global Models Project (Gilbert, et al. 2015). This ambitious project is developing techniques that will produce a nationally consistent and skillful suite of calibrated forecast guidance from a blend of both NWS and non-NWS models. “The blend” will be useful to national centers, local field offices, and the private sector. Most of the requirements for this blend (e.g., lead times, forecast grids, weather element definitions) have been chosen to suit the needs of the NDFD.

It was noted above that Wagner and Glahn (2010) described an adaptation of the EKDMOS technique that could combine forecasts from EFS of similar skill. Veenhuis (2014) described a technique, named Updatable Bayesian Model Averaging (UBMA), which can combine forecasts from EFS with different skills. UBMA assigns weights to each EFS and creates a PDF from the members. Like EKDMOS, UBMA dresses each member with a Gaussian kernel. The concepts and infrastructure of EKDMOS and UBMA contribute much to the NBM project.

Acknowledgments

Many hands contributed to the original development of EKDMOS and its subsequent evolution. Many more hands are contributing to the NBM at this time. The author acknowledges his role as a manager and chronicler, and little more. It is especially gratifying to note that a number of student trainees who worked on EKDMOS have now gone onto successful NWS careers where they routinely use these guidance products.

REFERENCES

Erickson, M. C., 1996: Medium-range prediction of PoP and Max/Min in the era of ensemble model output. *Preprints, 15th Conf. on Weather Analysis and Forecasting*, Norfolk, Virginia, Amer. Meteor. Soc., J35–J38.

Gilbert, K. K. and J. P. Craven, D. R. Novak, T. M. Hamill, J. Sieveking, D. P. Ruth, and S. J. Lord, 2015: An Introduction to the National Blend of Global Models Project. *Special Symposium on Model Postprocessing and Downscaling*, Phoenix, Arizona, Amer. Meteor. Soc.

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

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.

Glahn, H.R. and D. P. Ruth, 2003: The New Digital Forecast Database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–201.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. Mon. Wea. Rev., **125**, 3297–3319.

Toth, Z., and L. Lefaivre, G. Brunet, P. L. Houtekamer, Y. Zhu, R. Wobus, Y. Pelletier, R. Verret, L. Wilson, B. Cui, G. Pellerin, B. A. Gordon, D. Michaud, E. Olenic, D. Unger, S. Beauregard, 2006: The North American Ensemble Forecast System (NAEFS). *Preprints, 18th Conf. on Probability and Statistics in the Atmospheric Sciences,* Atlanta, Georgia, Amer. Meteor. Soc., 4.1.

Veenhuis, B. A., 2013: Spread Calibration of Ensemble MOS Forecasts. *Mon. Wea. Rev.*, **141**, 2467–2482.

_____, 2014: A Practical Model Blending Technique Based on Bayesian Model Averaging. *Extended Abstract, 22nd Conf. on Prob. and Stat. in the Atmos. Sci.*, Atlanta, Georgia, Amer. Meteor. Soc.

_____, and J. L. Wagner, 2012: Second moment calibration and comparative verification of ensemble MOS forecasts. *21st Conf. on Prob. and Stat. in the Atmos. Sci.* New Orleans, Louisiana, Amer. Meteor. Soc.

Wagner, John and B. Glahn, 2010: Ensemble MOS forecasts from multiple models. *Preprints, 20th Conf. on Probability and Statistics in the Atmospheric Sciences,* Atlanta, Georgia, Amer. Meteor. Soc.

FIGURES

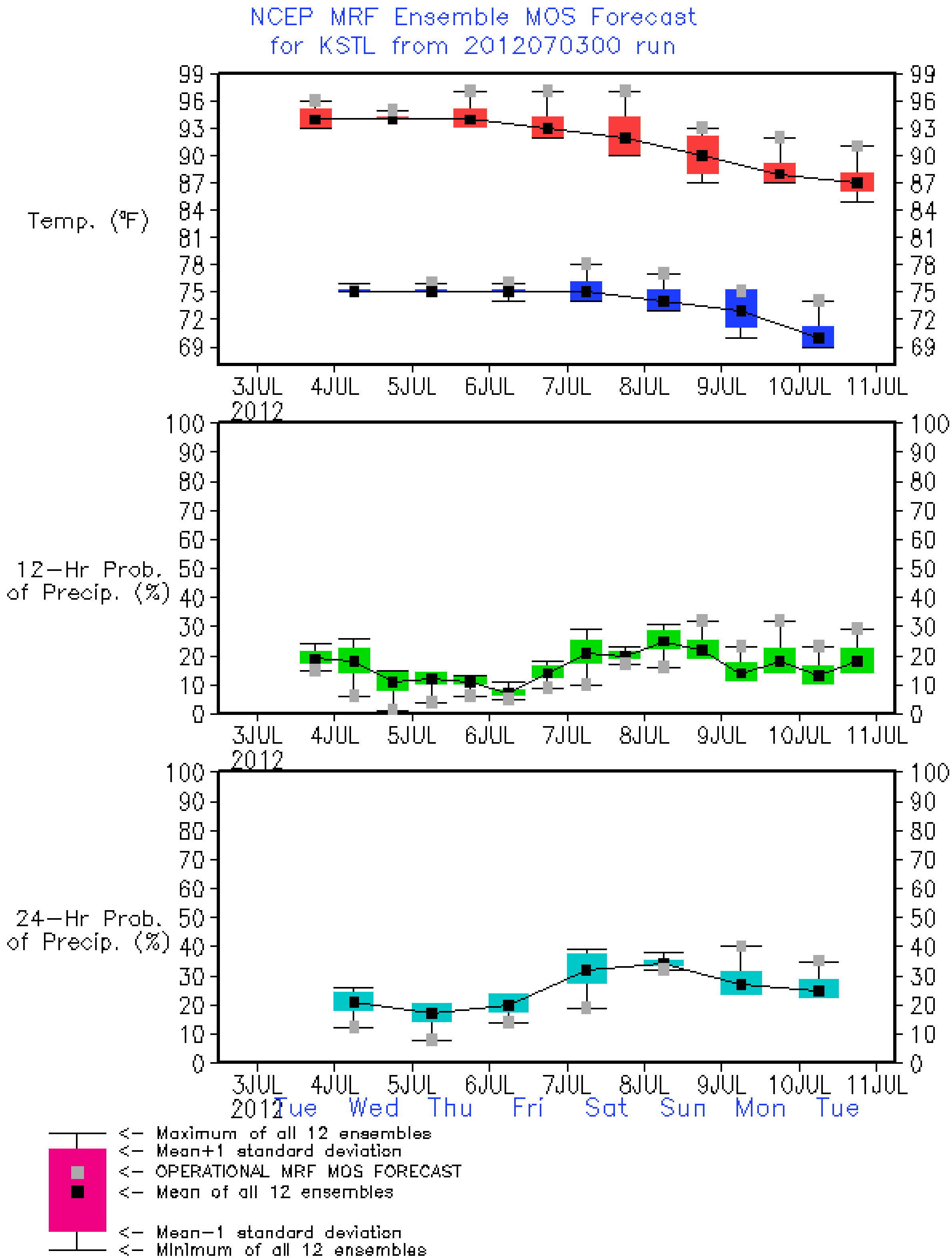

Fig. 1. Sample visualization of a MOS forecast generated from EFS

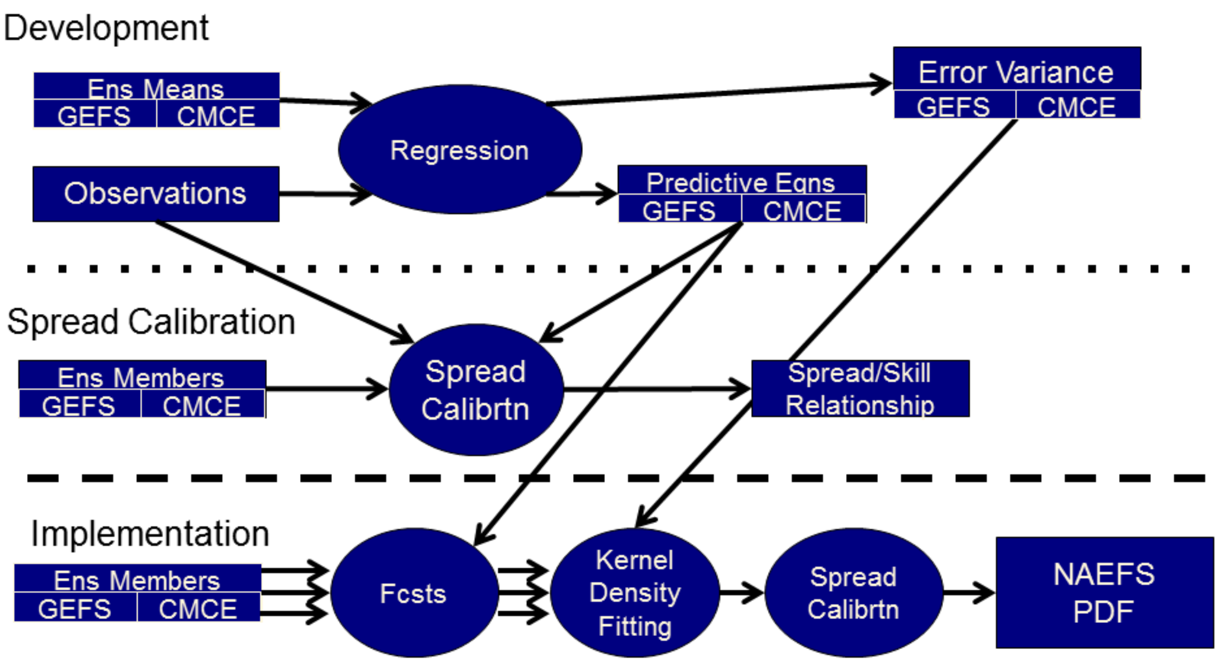

Fig. 2. Graphic depiction of EKDMOS technique

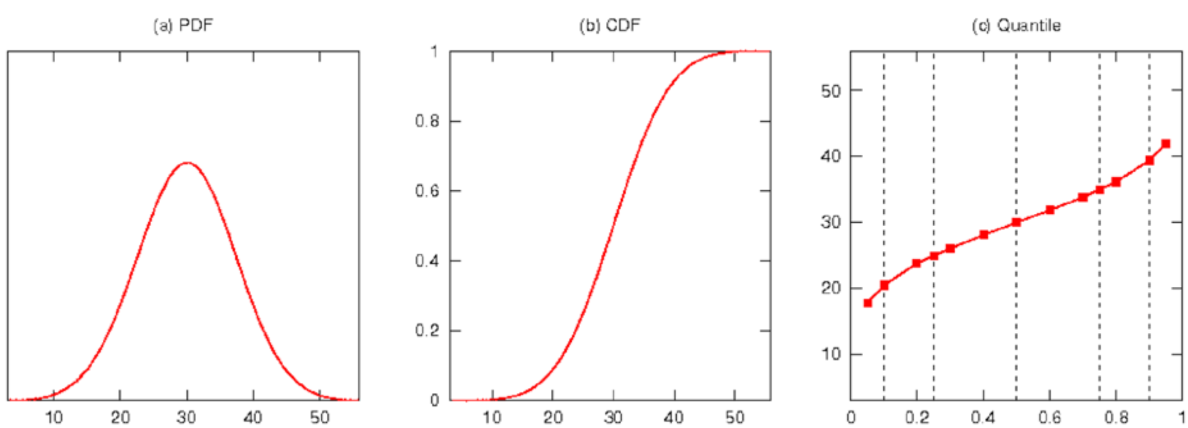

Fig. 3. An example of an (a) Probability Density Function (PDF), (b) Cumulative Distribution Function (CDF), and (c) a quantile function

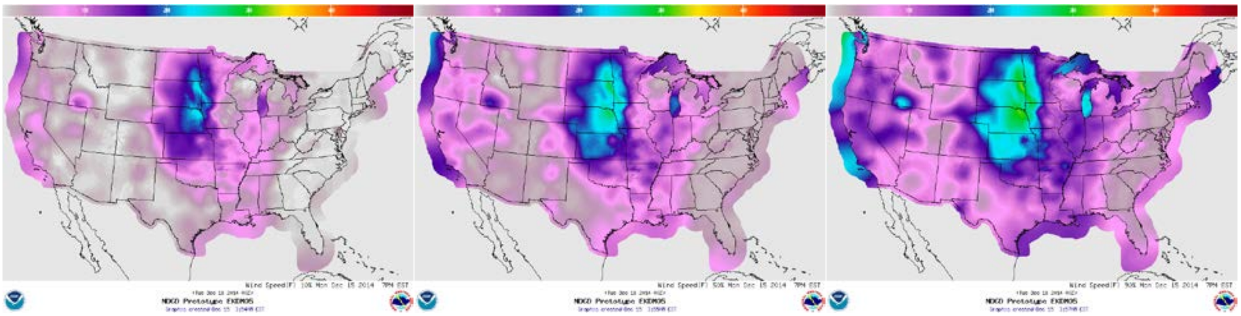

Fig. 4. An example of a gridded probability forecast. These 3 images can be presented as an animation loop. The images depict the 10th (left), 50th (middle), and 90th (right) percentiles of the forecast distribution.

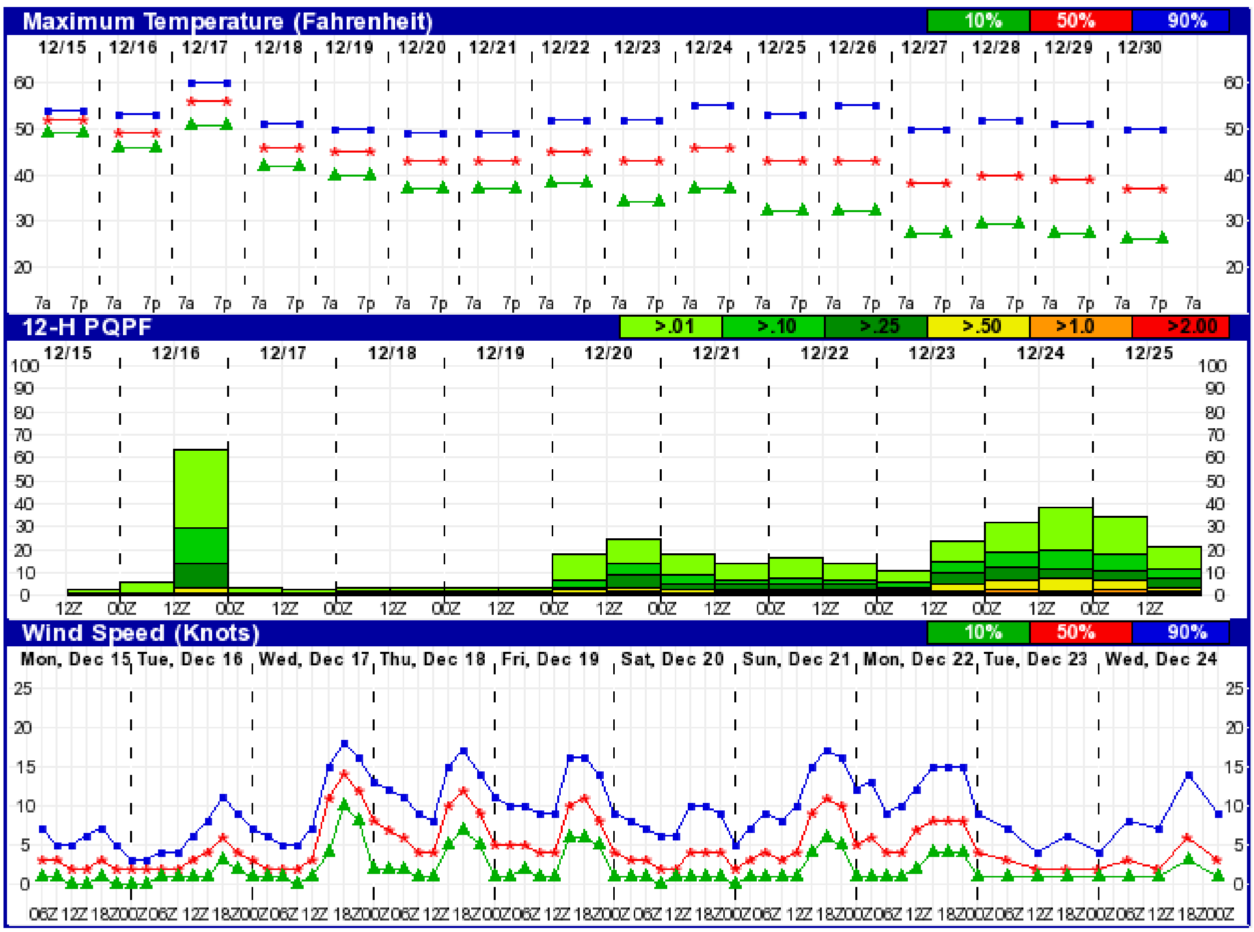

Fig. 5. Sample probabilistic forecasts for daytime maximum temperature, 12-h probabilisitic quantitative precipitation forecast, and wind speed visualized as a mixture of 2 meteograms and a stacked histogram.

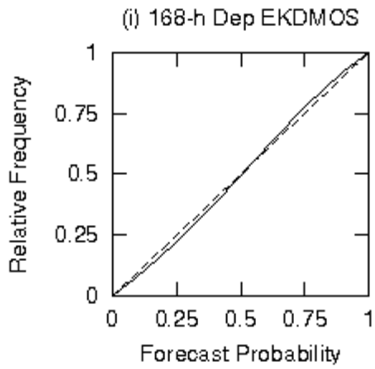

Fig. 6. Sample Cumulative Reliability Diagram

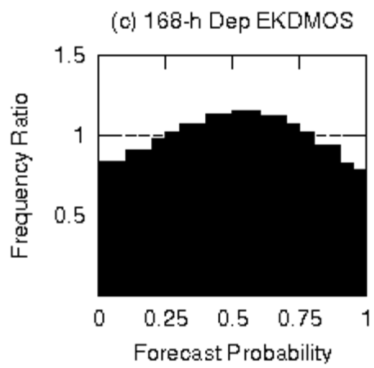

Fig. 7. Sample Probability Integral Transform (PIT) histogram

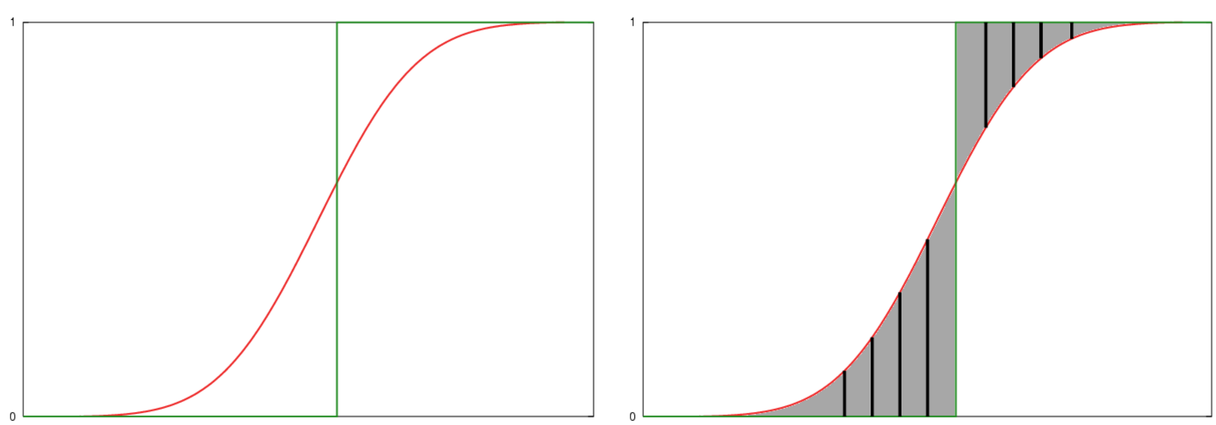

Fig. 8. An illustration of the computation of Continuous Rank Probability Score for a single forecast case. The red curve is the forecast CDF. The green line is a unit step function located at the value of the verifying observation.