2B.3

ASSESSING FORECAST UNCERTAINTY IN THE NATIONAL DIGITAL FORECAST DATABASE

Matthew R. Peroutka*, Gregory J. Zylstra, and John L. Wagner
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

1. INTRODUCTION

NOAA's National Weather Service (NWS) has implemented a National Digital Forecast Database (NDFD) that provides its customers and partners access to gridded forecasts of sensible weather elements (e.g., cloud cover, maximum temperature). As described by Glahn and Ruth (2003), the NDFD contains a mosaic of digital forecasts produced by NWS field offices working in collaboration with the National Centers for Environmental Prediction (NCEP). Table 1 lists the NDFD weather elements as well as their operational status at this time. Customers and partners are able to use NDFD forecasts to create a wide range of text, graphic, gridded, and image products of their own.

All NDFD weather elements except PoP12 represent single-value forecasts. The single-valued nature of the NDFD can be viewed as one of its limitations since all weather forecasts include some amount of uncertainty. The NWS recognizes this limitation, and the NWS Strategic Plan for 2005-2010 (NWS 2005) commits the agency to "including information on forecast uncertainty to enhance customer decision processes." Consistent with this goal, the Meteorological Development Laboratory (MDL) has been investigating techniques for assessing forecast uncertainty in the NDFD and generating products based on this information.

Figure 1 shows the basic structure of the Numeric Uncertainty Assessment of NDFD via Climatology and Ensembles (NUANCE). The NDFD forecast for a weather element, recent NDFD performance, and related guidance are all used to quantify the expected distribution of observations for that weather element. Initial efforts have focused on MaxT and MinT. This is because these two weather elements are accessed frequently by NDFD users, and because a considerable amount of data is available that describe their climatological behavior. MDL plans to use NUANCE to generate guidance products that allow NWS customers and partners to make better use of NDFD forecasts.

Table 1: NDFD Weather Elements

Operational
Maximum/Minimum Temperature (MaxT/MinT)
Probability of Precipitation (PoP12)
Dew Point
Temperature
Weather
Experimental
Sky Cover
Quantitative Precipitation Forecast (QPF)
Wind Direction and Speed
Snow Amount
Significant Wave Height
Apparent Temperature
Relative Humidity

2. METHODS

As with other guidance techniques, NUANCE will be implemented in two distinct phases, development and implementation.

The development process begins by amassing matched pairs of forecasts (denoted by f) and observations (denoted by x) to form a set of developmental data. The developmental data provide input to form a model from which the joint distribution of forecasts and observations, $p(f,x)$, is inferred. Additional diagnostic data (denoted by d) can be added to further refine the modeled distribution.

* *Corresponding author address:* Matthew R. Peroutka, W/OST24, Room 10376, National Oceanic and Atmospheric Administration, 1325 East West Highway, Silver Spring, MD 20910; e-mail: Matthew.Peroutka@noaa.gov.

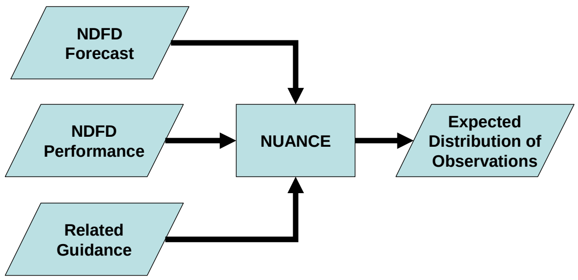

Figure 1: Overview of the NUANCE process

The implementation process uses the modeled distribution, *p(f,x,d)*, and current values of *x* and *d* to infer a conditional distribution of the observations given the forecast and diagnostic data, *p(x* | *f,d)*.

a. Data sources

The NDFD provides forecast values for regularly-spaced points on grids with mesh lengths that are close to 5 km. Efforts are underway within the NWS to routinely create a gridded Analysis of Record (AOR) on a similar spatial scale. The NDFD and the Analysis of Record are expected to be well-matched sources of *f* and *x*. Since the AOR is not operationally available, we have been using point data as sources of *f* and *x*. Observations are taken from hourly surface reports (generally encoded in METAR); forecasts are taken from the NDFD gridpoint nearest to the verifying surface observation.

Ensemble forecasting techniques seem likely to provide a useful source of diagnostic data. NCEP routinely provides a number of products that show the variability among the members of its ensemble NWP runs. Model Output Statistics (MOS; Glahn and Lowry 1972) generated from ensemble runs of the Global Forecast System (GFS; ENSMOS) can also provide diagnostic data for NUANCE.

The diagnostic data that have been studied to date for NUANCE are the ENSMOS forecasts for MaxT. Erickson (1996) describes the basic processes that are used to apply MOS equations to the individual members of an ensemble run. Archive files that contain the MOS forecasts generated from each ensemble member are available. These archive files include 11 separate ENSMOS bulletins for each model run. One set of forecasts is generated by applying the MOS equations to the so-called control (unperturbed) run of the model. Five of the ENSMOS forecasts come from the ensemble members that were perturbed in a “positive” way, and five of the ENSMOS forecasts come from the ensemble members that were perturbed in a “negative” way.

b. Transformation to percentiles

In prototyping NUANCE, it has proved useful to transform both the forecasts and the observations from their native values to climatological percentiles. This addresses a perennial problem in modeling *p(f,x)*, i. e., the lack of cases in the developmental data with extreme values of either *f* or *x*. One can expect this problem to be exacerbated by the relatively short length of the NDFD's archive of forecasts (little more than one year). The input provided by human forecasters into the NDFD may also lead to significant variations in the nature of *p(f,x)* among geographic regions as well as from one forecast to the next. We hope that transforming *f* and *x* to percentiles will enable us to combine data from multiple sites and multiple dates as *p(f,x)* is modeled, ultimately improving the model.

c. Diagnostic data

To date, two statistics derived from the ENSMOS MaxT guidance have been considered as diagnostic data for NUANCE. The first statistic is the standard deviation (SD) of the 11 MaxT forecasts contained in the ENSMOS guidance. This metric has been used successfully by NWS field forecasters to assess the quality of MOS guidance. The second statistic was named the “ensemble deviation” (ED). ED is computed by differencing each of the 10 perturbed forecasts with the control forecast, and computing the root mean square of these differences. By differencing the perturbed members with the control rather than the ensemble mean, we hoped to emphasize the role of the control member. This seemed desirable since the ENSMOS development sample did not include any of the perturbed ensemble members, but only the control member.

d. NUANCE development and implementation

Figure 2 shows the NUANCE development process. An archive of NDFD forecasts and their verifying observations are gathered and transformed from their native values to climatological percentiles. Related statistics from the associated ensemble forecasts are gathered as well. Together these data form a joint distribution model, *p(f,x,d)*, which can be used to assess the uncertainty of future NDFD forecasts. It is not clear whether *p(f,x,d)* should be developed once with the largest practical development data set or recomputed frequently using a smaller, “rolling” data set.

Figure 3 shows the NUANCE implementation process. The current NDFD forecasts are transformed from their native values to percentiles.

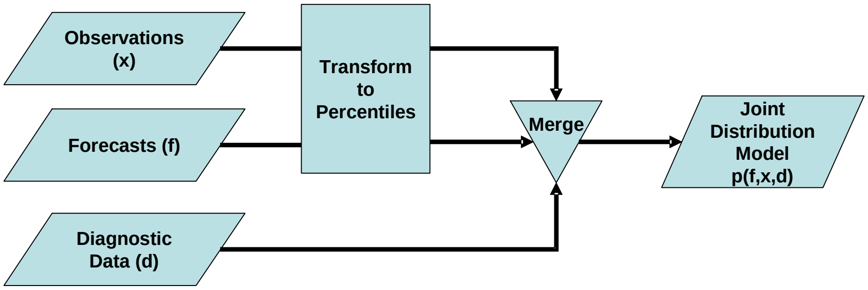

Figure 2: NUANCE development process

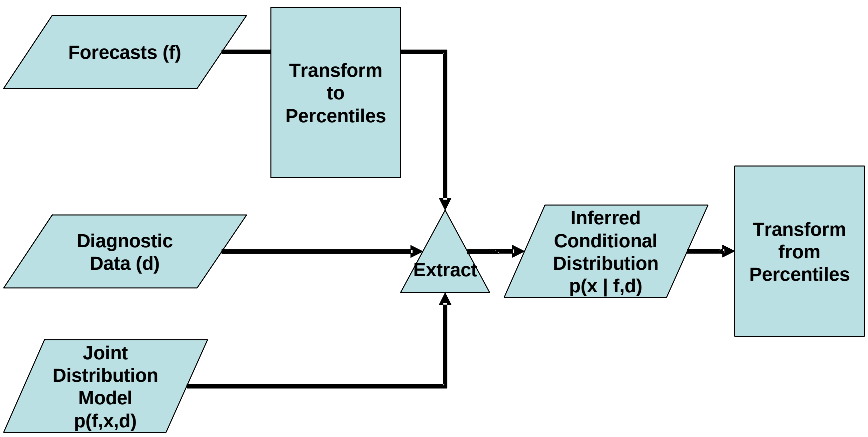

Figure 3: NUANCE implementation process

These forecasts and the current values of the associated diagnostic data are then used to infer the conditional distribution *p(x | f,d)* from the joint distribution model, *p(x,f,d)*.

3. RESULTS

We developed a prototype of the NUANCE process by using NDFD forecasts of MaxT at various forecast projection times. A set of 168 CONUS stations was selected that had long periods of record (~50 years) available in data sets provided by the U. S. Historical Climatology Network (USHCN; Karl, et al. 1990). Developmental data were taken from the period October 2004 to April 2005. Efforts focused on developing techniques that transformed MaxT forecasts and observations into climatological percentiles, qualitatively assessing the nature of the joint distribution, *p(f,x,d)*, and evaluating SD and ED as diagnostic data.

a. Transformation to percentiles

A technique was developed to produce percentile values for any station, value of MaxT, and day of the year as well as the inverse operation. For each station at five-day intervals throughout the year, ordered observations of MaxT, taken from USHCN, were used to compute an "observed" percentile function (MaxT as a function of accumulated probability) for each five-day interval.

Standard probability distributions were then used to model the distribution of MaxT for each station throughout the year. A cosine series based on the day of the year was used to form the parameters for each probability distribution. The use of a cosine series guaranteed that the final technique would yield results that were annually periodic. The magnitude and phase parameters of the cosine series were then adjusted fit the percentile function of this distribution to the observed percentile function derived from the USHCN data. The fits for each station were subjectively assessed, and the number of terms in the cosine series was adjusted to achieve a best fit. The quality of the fit was judged by using as few terms as possible, yet capturing the annual trends found in the observed distributions.

Eight standard probability distributions were tested for this technique. Table 2 provides the names of the distributions as well as a few subjective comments on their suitability. (The Binormal Distribution is described by Toth and Szentimrey (1990).) The distribution that fit the data best was the Generalized Lambda Distribution (GLD; Karian and Dudewicz 2000). The GLD is a powerful probability distribution that can take on a variety of shapes. This flexibility enables the GLD to model daily distributions of MaxT and MinT with suitable results. Four parameters define the GLD. The GLD is not well-behaved for all values of its four parameters, and this characteristic can lead to problems when fitting a GLD to observed data. Öztürk and Dale (1982) describe the use of GLD to model sunshine data.

Figure 4 shows sample results for this percentile transform technique, using GLD. The figure plots the five-day frequency data for the 5^{th}, 50^{th}, and 95^{th} percentiles at station BLH as well as curves produced by the technique. The curves are able to capture a number of subtle features that can be seen in the plotted data.

Table 2: Probability distributions tested for modeling daily distributions of MaxT and MinT

Distribution	Variable	Comment
Normal	MaxT	Poor fit "in the tails"
Normal	ln (MaxT)	Improved fit "in the tails"
Binormal	MaxT	Skewness improved fit for some stations. Poor fit "in the tails."
Logistic	MaxT	Better fit than either version of Normal
Laplace	MaxT	Worst fit
Gumbel	MaxT	Skewness improved fit for some stations.
Gumbel	-(MaxT)	Skewness improved fit for some stations.
Generalized Lambda		Best overall fit

Figure 5 compares the fitted curves from three CONUS stations that are located in drastically different climatological regimes. For each station, nine curves are plotted, one each for the 10^{th} through 90^{th} percentiles. Blythe, California, (BLH) is located in California's central valley, and it is subject to hot summertime temperatures. Baudette, Minnesota, (BDE) is located in the northern plains of the CONUS. Fort Lauderdale, Florida, (FLL) is a southern coastal station.

The percentile curves in Figure 5 clearly model a number of important characteristics of the climatology of MaxT at each station. KBLH is hot; the 90th percentile for MaxT approaches 115° F during the summer. During the winter and spring, MaxT shows increased variability. This variability can be seen in the increased spread in the percentile curves during those seasons. By contrast, KFLL shows relatively little seasonal variation. During July, the spread between the 10th and 90th percentiles is remarkably small. KBDE is, by far, the coldest station of the three. The 10th percentile for MaxT drops below minus 5° F during January. Note the large annual variation and the increased spread between the 10th and 90th percentile lines during January.

b. Modeling the joint distribution p(f,x,d)

Computationally modeling the joint distribution *p(f,x,d)* can be done in very straightforward ways since the NUANCE prototype is implemented for a small number of stations. Each MaxT forecast, its verifying observation, and associated ENSMOS metric is preserved within the application. Other techniques that are less memory-intensive will likely be needed before NUANCE can be implemented for gridded NDFD forecasts.

Scatter diagrams provide one tool for qualitatively assessing the nature of *p(f,x,d)*. Figure 6 compares scatter diagrams for NDFD MaxT forecasts for the day 1 (24h) and day 7 (168h) time projections. The scatter diagrams certainly show a difference in the characteristics of NDFD forecasts for these two time projections. In each diagram, the diagonal line that runs from the lower left to the upper right represents a perfect forecast. Data points that are coincident with or lie near that line verify best. The data-sparse regions on either side of the diagonal are the result of the one-degree (Fahrenheit) resolution of the forecast and observed data. Visual inspection quickly suggests that day 1 forecasts verify better than day 7 forecasts, as one might expect. The points on the day 7 scatter diagram cluster around the 0.50 forecast value more than the points on the day 1 diagram. This behavior coincides well with the tendency of human forecasters and objective forecasting techniques to be influenced by climatology more for later time projections. The day 7 diagram also shows fewer extreme forecasts than the day 1 diagram.

c. Diagnostic Data

We have already introduced the SD and ED of ENSMOS forecasts for MaxT as candidates for diagnostic data. Figure 7 uses scatter diagrams to show the effects of stratifying day 7 MaxT NDFD forecasts by the ED. The diagram on the left plots NDFD forecasts vs. observations for those cases where ED < 6° F. The diagram on the right shows the cases where ED ≥ 6° F. The union of all points plotted in the two diagrams in Figure 7 yields the diagram on the right in Figure 6. Comparing these three diagrams suggests that NDFD MaxT forecasts for day 7 are more skillful when ED < 6° F, but this relationship is not obvious.

Figures 8 and 9 provide some insight into the value that SD and ED may provide as diagnostic data. In all four graphs either SD or ED is used to stratify forecasts from the development sample. Intervals of SD/ED were established with a width of 0.1° F. NDFD forecasts and their verifying observations were assigned to these intervals.

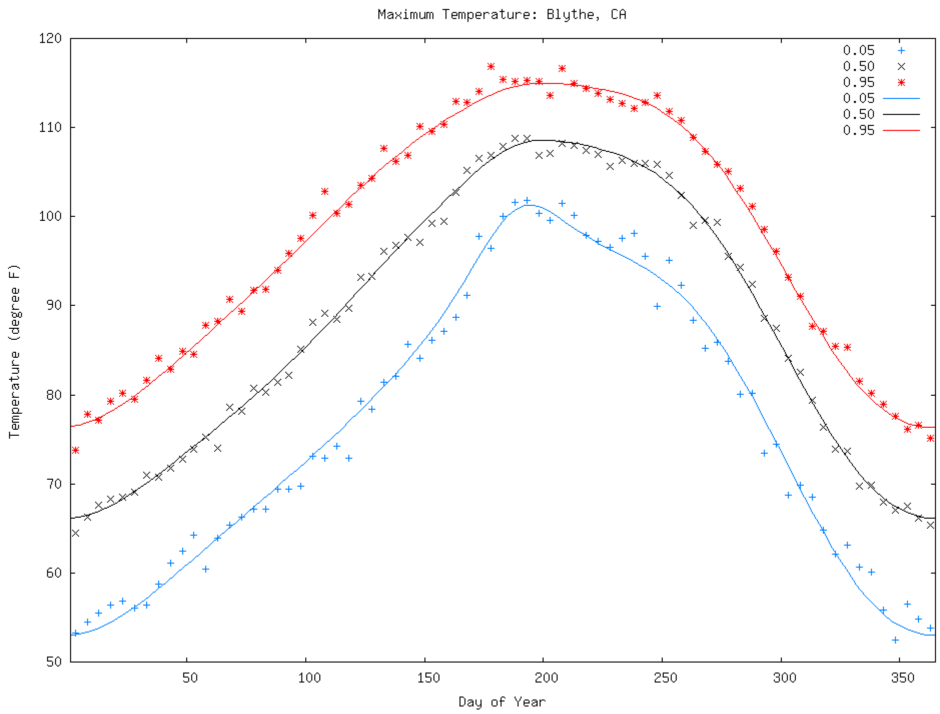

Figure 4: Results of percentile transform technique for Blythe, California. Data points show 5th, 50th, and 95th percentiles of climatological data. Curves show the same percentiles, produced by technique.

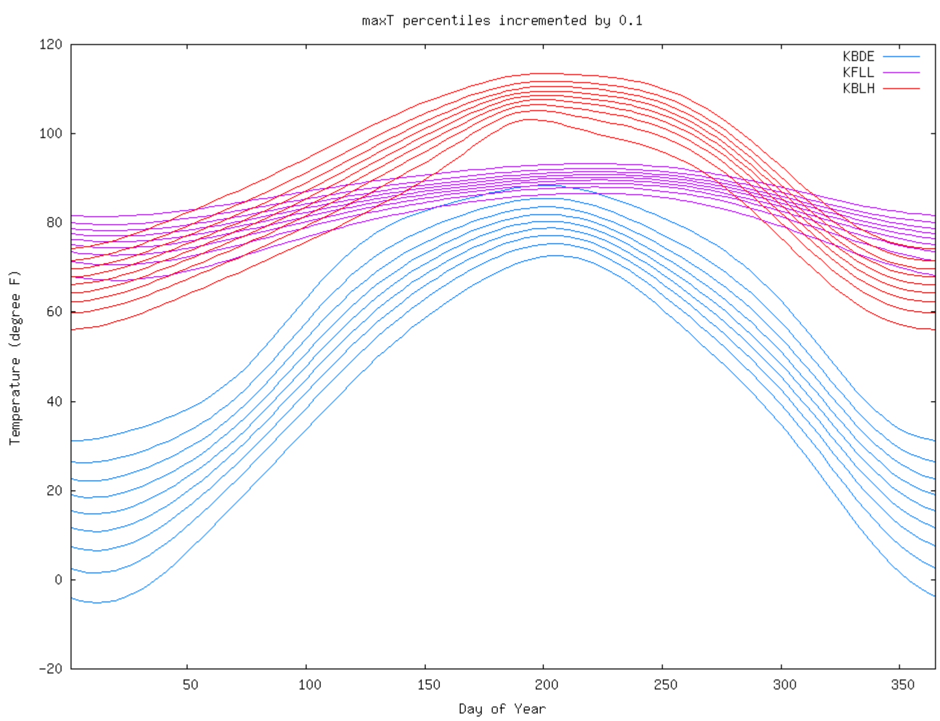

Figure 5: Comparison of MaxT percentiles for Baudette, Minnesota (KBDE); Fort Lauderdale, Florida (KFLL); and Blythe, California (KBLH)

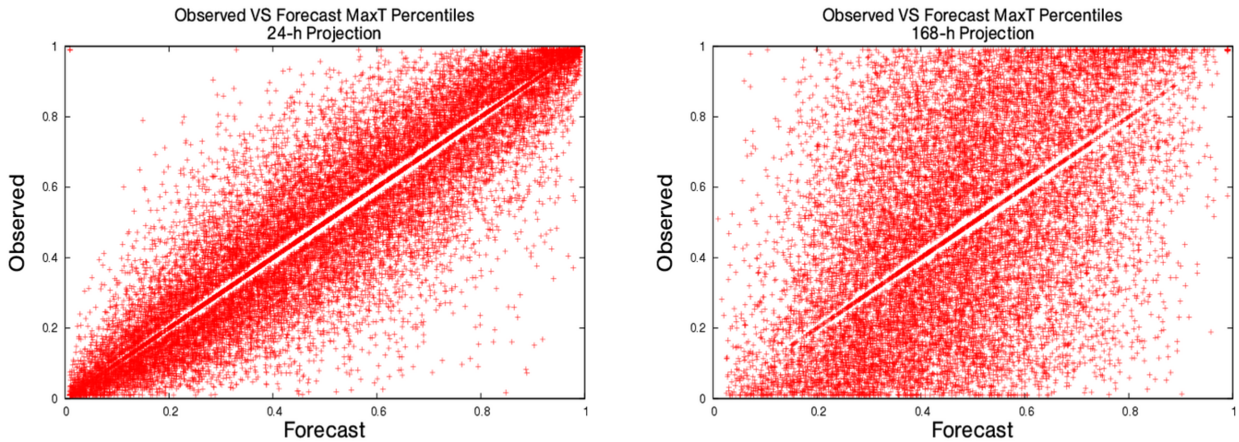

Figure 6: Comparison of scatter diagrams of NDFD forecasts/verifying observations for day 1 and day 7

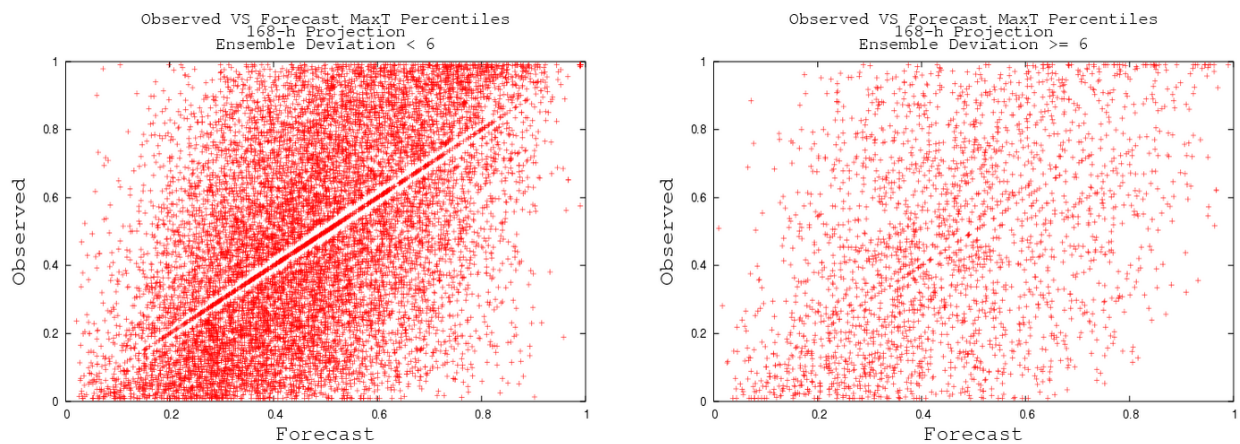

Figure 7: Comparison of scatter diagrams of NDFD forecasts/verifying observations for day 1, stratified by the ensemble deviation (defined in text) value computed from ENSMOS guidance

Mean Absolute Error (MAE) was then computed for those intervals that contained more than 30 forecasts. The two graphs in Figure 8 plot MAE for each bin vs. ED for day 1 (left) and day 7 (right). The two graphs of Figure 9 show the same information, but substitute SD for ED.

All four plots in Figures 8 and 9 associate larger MAE values with larger values of SD and ED. This relationship is stronger in data-rich regions at the center of each graph, and it is harder to identify where there are fewer cases. As one might expect, the values of MAE, SD, and ED are larger on day 7 than on day 1.

Of course, qualitative assessment of scatter diagrams is no substitute for a thorough statistical analysis of the data. These graphical results, however, indicate the potential value of the NUANCE technique.

4. Future Plans

Our next efforts will be focused on quantitatively assessing the information presented qualitatively here, assessing the behavior of MinT, working with grids, and producing experimental guidance products. Figures 8 and 9 suggest a relationship between NDFD errors and the ED and SD of ENSMOS forecasts. This relationship seems to be different for a day 1 forecast than for a day 7 forecast. MinT data are readily available for the same stations. Applying the techniques described here for MaxT to MinT should be straightforward. NDFD forecasts are grids, and, eventually, these NUANCE techniques will need to assess uncertainty at every gridpoint. This will require developing techniques that can assess the climatological distribution of weather elements at gridpoints.

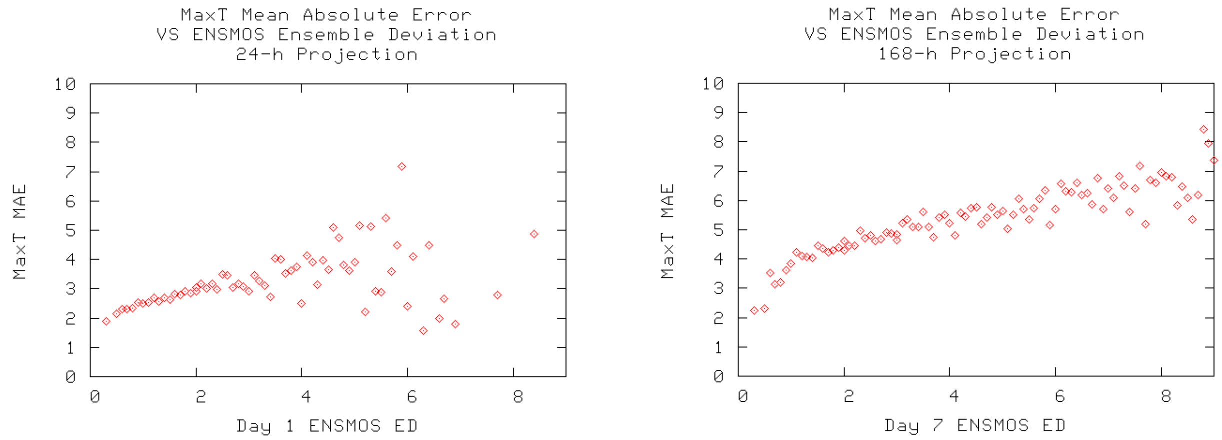

Figure 8: Comparison of NDFD MAE to various values of the ED of the ENSMOS for day 1 (left) and day 7 (right)

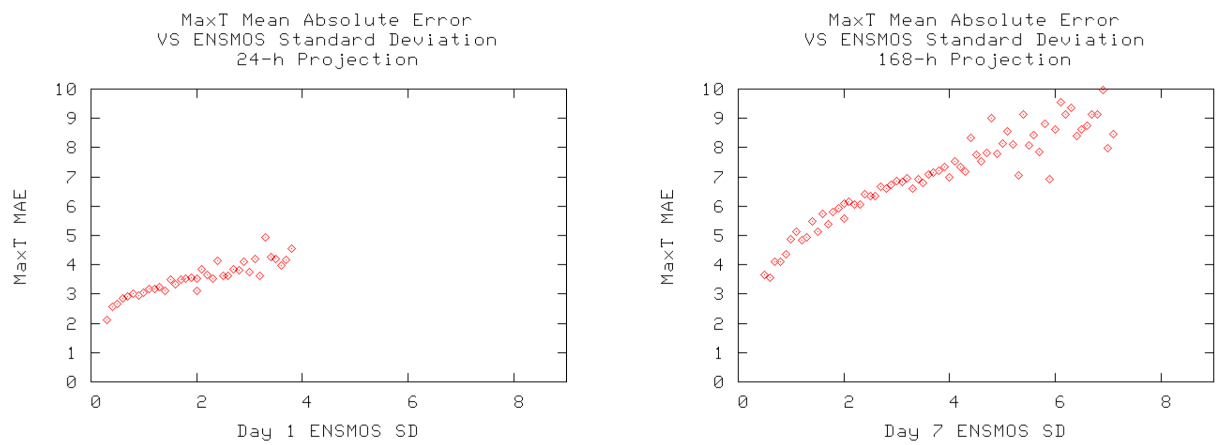

Figure 9: Same as Figure 8, but for SD rather than ED

A few efforts have been made to prototype products that take advantage of the conditional distribution *p(x | f,d)*. These have mostly taken the form of generating a 50% confidence interval around the NDFD MaxT forecast. Other products have been considered, including a probability density function (PDF). This PDF could be expressed as the boundary values for 10-percentile intervals. A PDF product could compliment the single-valued forecasts of the NDFD and offer NWS partners opportunities to develop additional products. Another candidate product would estimate the probability the value of a weather element will exceed certain key values (32° F, 100° F, etc.).

5. Conclusion

The NDFD is a resource of tremendous value. One possible use of these data is to compute uncertainty information that can augment the worth of the single-valued forecasts. The NUANCE technique may provide a number of tools that can derive additional value from the NDFD.

6. 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 Anal. and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J35-38.

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.

______, and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203-1211.

Karian, Z. A. and E. J. Dudewicz, 2000: *Fitting Statistical Distributions.* CRC Press, 438 pp.

Karl, T. R., C. N. Williams, Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

National Weather Service, cited 2005: Working together to save lives: National Weather Service Strategic Plan 2005-2010. [Available online at http://www.weather.gov/sp/NWS_strategic_plan_01-03-05.pdf.]

Öztürk, A. and R. F. Dale. 1982: A Study of fitting the Generalized Lambda Distribution to solar radiation data. *J. Appl. Meteor.*, **21**, 995–1004.

Toth, Z. and T. Szentimrey, 1990: The Binormal Distribution: a distribution for representing asymmetrical but normal-like weather elements. *J. Climate*, **3**, 128–136.