15B.3

PROBABILISTIC TEMPERATURE FORECASTING: A PROOF OF CONCEPT

Steve Amburn* and Glenn Wiley
National Weather Service
Tulsa, Oklahoma

1. INTRODUCTION

Communicating uncertainty in weather forecasts is becoming more important as decision makers become better able to apply uncertainty in their businesses practices and operations. Probabilistic precipitation forecasts have effectively communicated useful uncertainty for many decades. The National Weather Service Forecast Office in Tulsa, Oklahoma (NWS WFO Tulsa) is now studying how probabilities may be applied to maximum and minimum (max/min) temperature forecasts and how those uncertainties may be communicated to customers. Several approaches are available and are presented from the WFO perspective as a proof of concept.

It is common at forecast offices to hear forecasters describe a variety of scenarios the weather presents for the following days. Temperature forecasting can present a particularly difficult issue when the speed of a cold front is in question or when precipitation duration may limit the diurnal heating on a particular day. In these cases, the forecast maximum temperature might be 90°F or 75°F. Today at WFOs, a forecaster must pick one number or the other, or try to minimize his or her potential error by picking a number in the middle which has even less likelihood of being correct. Unfortunately, that single forecast value fails to communicate the range of possible max/min temperatures known to the forecaster.

It would be desirable to communicate an appropriate degree of uncertainty while not simply communicating every possible model outcome. How best to do that is the issue.

After forecasters review and assess the model fields in preparation to make their temperature forecasts, they finally review and assess the MOS and previous official forecasts for each forecast period. In their process, two thoughts are generally considered: 1) the change if any from the previous forecast, 2) inter-agreement between the new MOS and their own forecast idea. A third consideration might be the general performance accuracy of the particular forecaster. Although potentially important, it will not be addressed here.

This paper is an investigation of how uncertainty in max/min temperatures might be accomplished at a typical WFO today, given the forecast process identified above. The concept is straight-forward and gives forecasters maximum control over the final probabilistic temperature forecast product. At this level, only limited tools would be required and additional workload on the forecaster could be minimal. Inputs for this investigation included Model Output Statistics (MOS) from the Global Forecast System (GFS MAV MOS) and the North American Model (NAM MET MOS) and the WFO official forecast (CCF).

The method to estimate the final probability distribution assumes that each individual forecast (MAV, MET, CCF, etc) represents the mean temperature of a Gaussian distribution that might be expected from an infinite number of similar events. Appropriate standard deviation for each forecast (MAV, MET and CCF) can be determined from previous seasonal forecast accuracy, from the inter-consistency of the guidance products and CCF forecasts, or from the MOS and CCF consistency from one forecast cycle to the next. The resulting individual distributions may then be appropriately combined, with or without additional weighting, to arrive at the final expected distribution of the forecast temperature. The question then becomes how to determine the appropriate standard deviations for individual guidance products and the CCF.

One method of combining a variety of forecast distributions has been developed by Wilks (2005) and employed by Glahn (2009) in the development of Ensemble-Kernel Density-MOS also known as EKDMOS. That method is far superior to anything that could be presented here. Therefore, only simple averaging was used in this paper to illustrate samples of possible forecast distributions.

The method to create forecast temperature distributions suggested in this paper includes forecasters' judgment into the final product, thereby allowing them to weight or completely discard selected MOS or model forecasts for all or part of their forecast area. Standard deviations were generally based on typical forecast errors and model consistencies, not a variety of perturbed model runs.

Corresponding author address: Steve Amburn, National Weather Service, 10159 East 11th Street, Suite 300, Tulsa, OK, 74128; e-mail: Steve.Amburn@noaa.gov

The results of this proof of concept indicate that uncertainty of max/min temperature forecasts can be addressed very well from the WFO perspective. Equipped with their knowledge, experience and model guidance, forecasters can easily provide lower uncertainty than that derived from the climatological range. Yet, forecasters can still communicate considerable uncertainty for appropriate forecast periods where model solutions are quite varied and forecaster uncertainty is high. Most of the tools are currently available at WFOs to make probabilistic temperatures forecasts now. The question then becomes whether to proceed or not.

2. COMPUTED MEANS AND STANDARD DEVIATIONS

Computed means and standard deviations provide sufficient information to start making probabilistic temperature forecasts. The climatological distribution of observed temperatures about a mean for the specific date and the distribution of forecast errors about the observed temperature suggest a Gaussian distribution is appropriate. However, combining individual Gaussian distributions will normally yield a non-Gaussian solution. An analysis of forecast errors from 2004 through 2008 at WFO Tulsa indicates that forecaster and model errors are generally distributed evenly at the two primary forecast verification sites for WFO Tulsa (Figs. 1 and 2). These data include all forecast cycles for the entire period of time, therefore may not be as representative for the extreme seasons of summer and winter. However, the error distributions indicate only a small bias, so it seemed reasonable to assume that most forecasts will have near-Gaussian distributions.

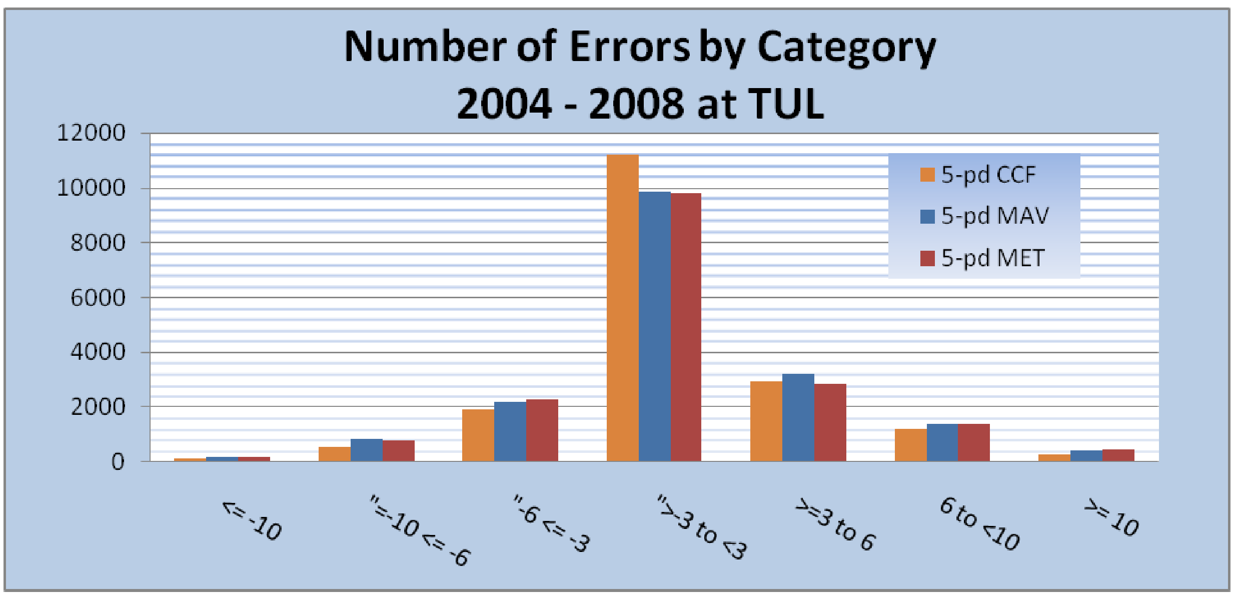

Figure 1. Forecast errors by category for TUL from 1 January 2004 through 31 December 2008.

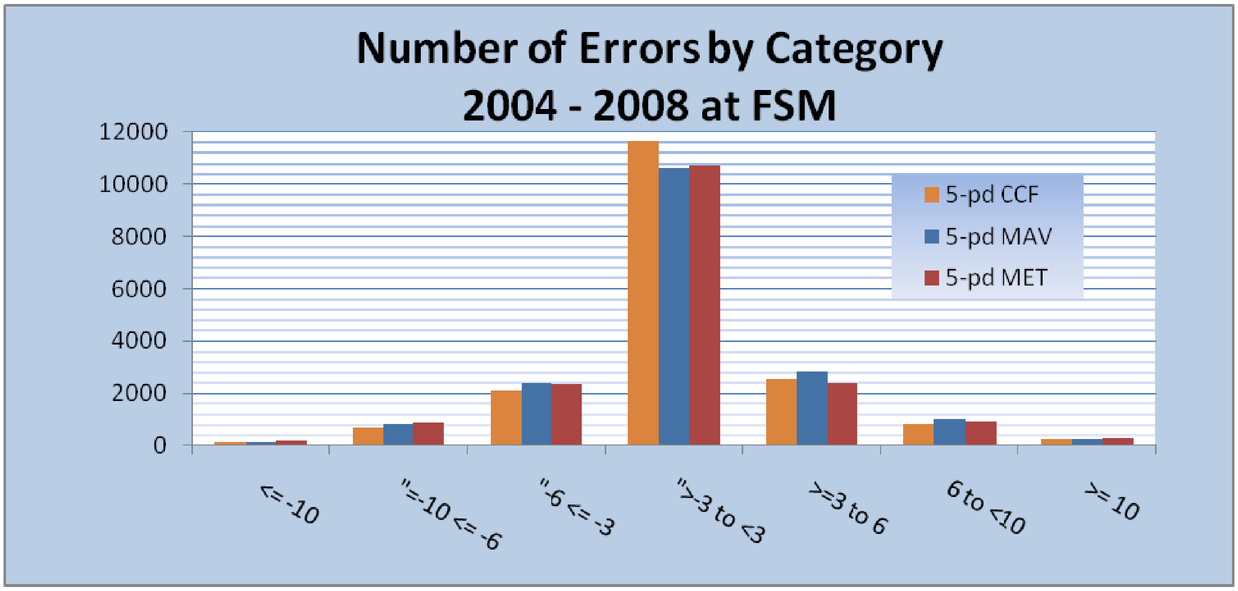

Figure 2. Same as Figure 1 but for FSM.

Figures 3 and 4 show the climatologically expected maximum temperature distribution for January 1 and June 30, respectively, at TUL (Tulsa International Airport). These distributions were computed from twelve years of data from 1996 through 2008. Data for each date were increased by including four days on either side of the date. This provided a smoother transition in the daily means and resulted in a better approximation of the means and standard deviations.

The standard deviation about the mean on January 1 is over 13°F while it is only 5°F on June 30. Clearly, a system should be developed to narrow this degree of climatological uncertainty, yet provide an appropriate level of uncertainty that will be useful to customers and partners.

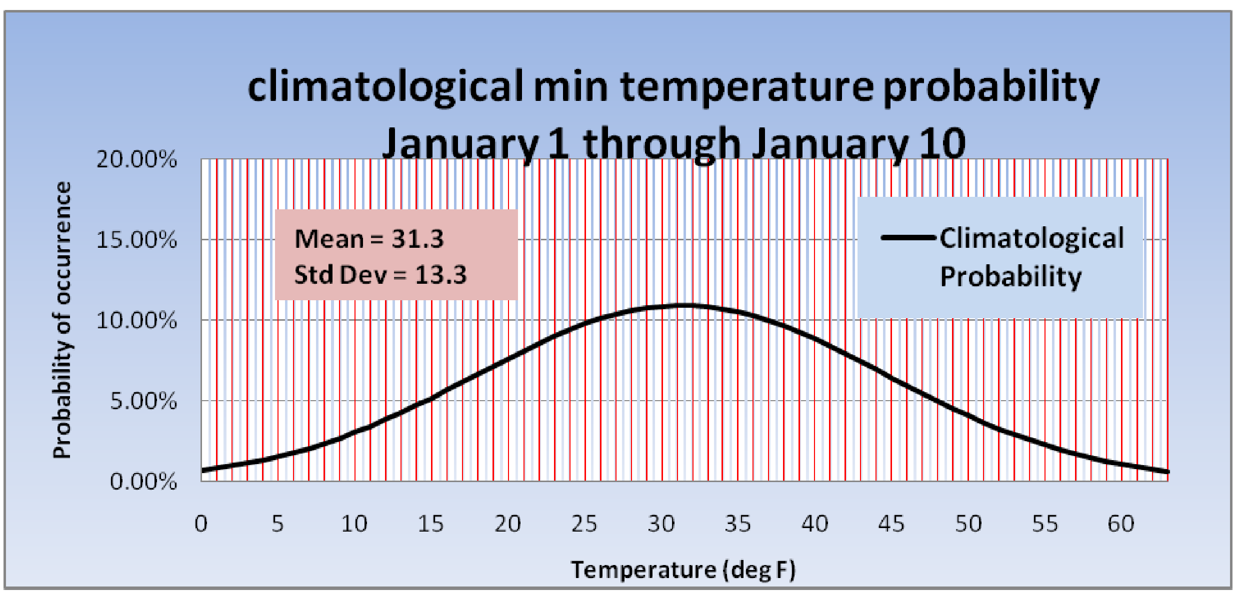

Figure 3. Climatological probability of minimum temperature computed for January 1 for TUL. Period of data was from 1996 through 2008.

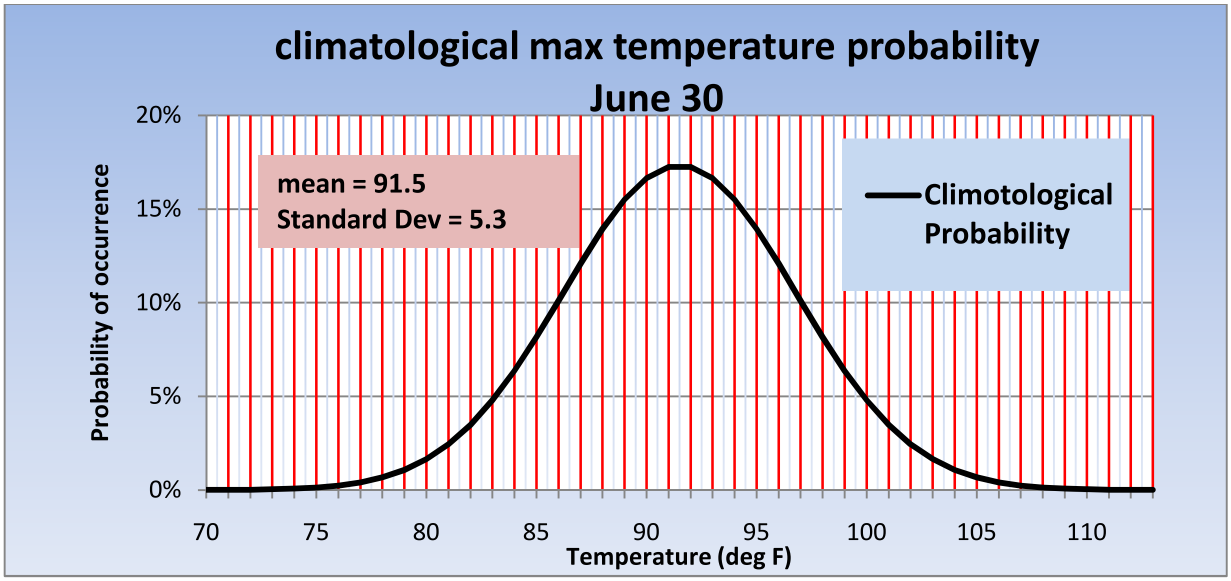

Figure 4. Climatological probability of maximum temperature computed for June 30 for TUL. Period of data was from 1996 through 2008.

3. STANDARD DEVIATIONS BASED ON VERIFICATION ERRORS

At the WFO, key components in providing the appropriate level of uncertainty in a temperature forecast would be the MOS guidance and the previous cycle's CCF. These provide the forecaster with not only a good first approximation of the expected max/min temperature, but often the final number used in the official forecast. Standard deviations can be computed from the errors of the MOS and CCF to estimate a baseline uncertainty. However, standard deviations can also be computed based on how well the CCF and MOS guidances agree.

Table 1 contains standard deviations (Std Devs) and Mean Absolute Errors (MAEs) for the short term CCF and guidance for periods one through five. As expected these show increases in both statistics at greater forecast ranges. Such information would provide a reasonable first approximation for uncertainty if desired.

May 2008 – April 2009	Pd 1	Pd 2	Pd 3	Pd 4	Pd 5
CCF Std Dev (MAE)	**3.1** (2.3)	**3.5** (2.7)	**3.8** (2.9)	**4.2** (3.2)	**4.5** (3.4)
MAV Std Dev (MAE)	**3.6** (2.9)	**4.0** (3.2)	**4.3** (3.3)	**4.7** (3.6)	**5.1** (3.9)
MET Std Dev (MAE)	**3.4** (2.6)	**3.9** (3.0)	**4.3** (3.3)	**4.7** (3.6)	**5.2** (4.1)
UKMET Std Dev (MAE)	**4.1** (3.0)	**4.4** (3.3)	**4.7** (3.5)	**4.8** (3.7)	**5.2** (4.0)

Table 1. MAEs and standard deviations for periods 1-5 for CCF, MAV, MET and UKMET forecast maximum temperature. Data set is from 5/1/2008 through 4/30/ 2009 for highs and lows combined.

Table 2 shows seasonal standard deviations for the indicated forecasts and guidance (MAEs omitted). The statistics for the Dec-Jan-Feb periods are similar to those for the Mar-Apr-May period, indicating the strong variability of the weather at TUL and also the uncertainty of the forecasts and guidances. However, those standard deviations drop significantly in the summer and fall seasons, indicating the higher degree of forecast and guidance certainty

MONTH	Dec-Jan-Feb	Mar-Apr-May	Jun-Jul-Aug	Sep-Oct-Nov
CCF Std Dev	**3.6**	**3.6**	**2.5**	**2.5**
MAV Std Dev	**4.2**	**4.1**	**2.9**	**2.8**
MET Std Dev	**4.0**	**3.8**	**2.7**	**2.8**
UKMET Std Dev	**5.3**	**4.1**	**3.0**	**3.7**

Table 2. MAEs and standard deviations for representative months for CCF forecast temperatures. Data set is for first period forecasts only, from 5/1/2008 through 4/30/2009 for highs and lows combined.

Tables 3 and 4 show standard deviations computed from one full year of max/min data for the period 5/1/2008 through 4/30/2009. These data compare the standard deviations of the CCF/MAV and CCF/MET for periods one and five based on how closely the CCF matches either of the two guidance products. For Table 3, the first row shows standard deviations when the CCF and MAV guidance either matched or differed by only one degree. The second row shows the standard deviations when the CCF/MAV difference was five degrees or greater. Table 4 presents the same computations but for the CCF and MET guidance.

The point to be made with the information in Tables 3 and 4 is that meteorologists' input into the forecast process can have a positive effect on the resulting standard deviations and therefore the uncertainty of a final probabilistic temperature product. The differences in standard deviation for the CCF and MAV for period 1 are not dramatic, but the differences in standard deviations from period one to period five are significant for both the CCF/MAV and also for the CCF/MET.

CCF/MAV differences	Pd 1 - CCF	Pd 1 - MAV	Pd 5 - CCF	Pd 5 - MAV
|CCF-MAV| = 0,1	2.99	3.06	4.32	4.37
|CCF-MAV| > =5	4.74	5.25	5.72	8.49

Table 3. Standard deviations based on CCF/MAV agreement. Data set is from 5/1/2008 through 4/30/2009 for period one and period five forecasts.

CCF/MET differences	Pd 1 - CCF	Pd 1 - MET	Pd 5 - CCF	Pd 5 - MET
|CCF-MET| = 0,1	2.97	3.06	4.30	4.38
|CCF-MET| > =5	3.76	5.32	5.29	6.79

Table 4. MAEs and standard deviations based on error for representative months for official WFO minimum temperature. Data set is from 2001 through 2008.

4. Forecaster Inputs to Uncertainty

Unfortunately, forecasters are frequently forced to choose a specific maximum or minimum temperature forecast even though that forecaster may feel the range of outcomes could be potentially large. In contrast, a forecaster may have high confidence in the particular forecast. Without a means to convey that information, the forecaster leaves the public or other customers without benefit of his/her knowledge. The basic mechanics of a system to convey that knowledge are already available at WFOs.

In the current forecasting process, meteorologists effectively weight the models and MOS in their effort to arrive at a final temperature forecast. Frequently, forecasters will simply average MOS guidance or "lean" more heavily toward one or the other for a variety of reasons. These might include knowledge of model or MOS biases, data that may have arrived too late for input to the model, or other reasons.

Sometimes, a meteorologist may find the weather to be in a relatively stagnant pattern with very similar maximum and minimum temperatures for several days. He or she may also find that the MOS guidance has been very accurate for several days and is also relatively unchanged for the next periods' forecasts. Perhaps the MOS guidance simply agrees with the current official forecast. In such instances, the meteorologists may have very high confidence in the forecast.

At other times, two separate models may simply differ on the forecast solution, implying a higher degree of uncertainty. The resulting MOS differences are a primary factor in forecaster uncertainty. The differences may be due to the speed at which different models move a synoptic system, or a variety of other factors. These MOS differences are easily interpreted. A forecaster can combine and/or weight the MAV and MET guidances where necessary to better depict the uncertainty of a forecast and its range of possible outcomes. The Gridded Forecast Editor or GFE (Global Systems Division, http://www-md.fsl.noaa.gov/eft/) provides several graphical user interfaces (GUIs) to assign weights to forecasts. These GUIs are used routinely at most WFOs and are shown in Figures 5 through 7. These GFE GUIs not only allow meteorologists to create forecasts by selecting and assigning weights to MOS and model guidance but can also be used in creating the final of probability distributions.

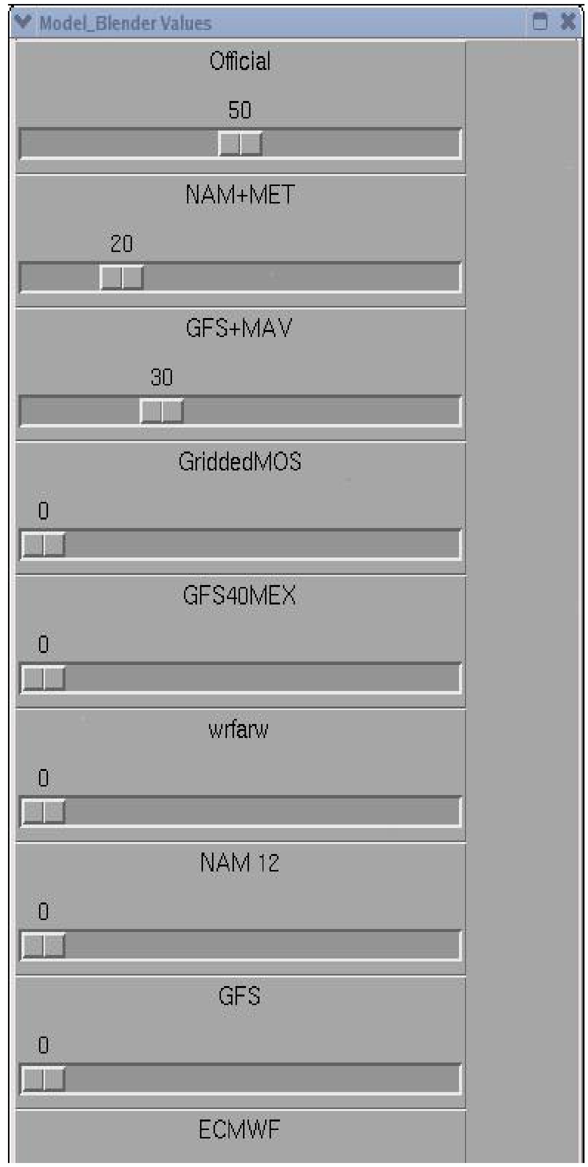

Figure 5. GFE Model blender GUI or "slider bar" can be used to assign weights to the current official forecast or MOS or model values.

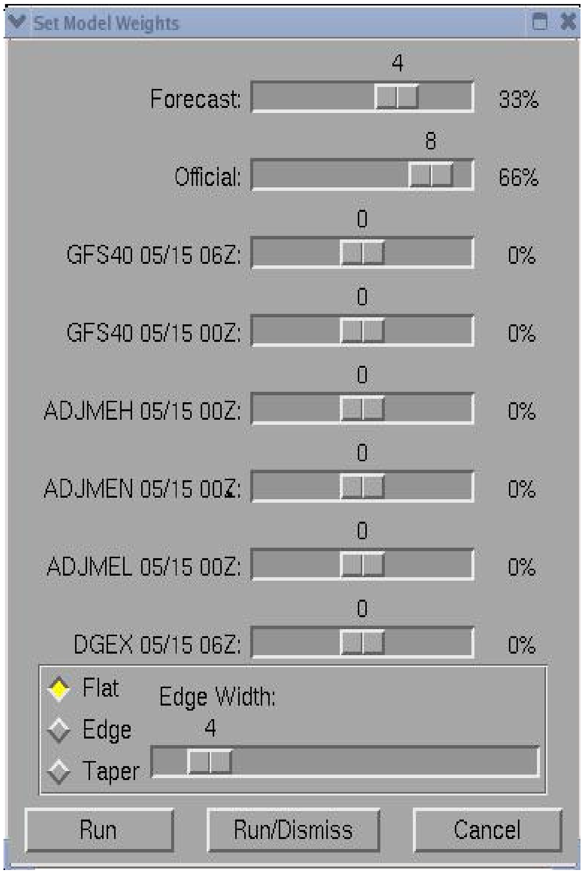

Figure 6. Another GFE model blender tool with slider bars to assign specific weights to model, MOS and official forecasts.

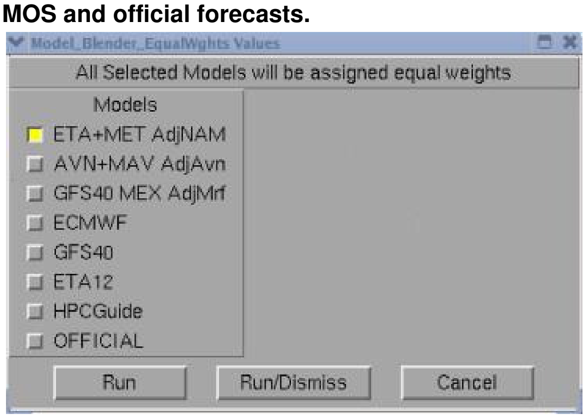

Figure 7. GFE model/MOS blender that assigns equal weights to each selection.

5. More on Standard Deviations

The value of the standard deviations used in calculating the temperature probabilities is critical in determining the uncertainty of the final forecast. Examples of standard deviations based on CCF and MOS forecast accuracy were presented in Tables 1 and 2 while those based on CCF/MOS agreement were shown in Tables 3 and 4. However, the certainty of a forecast may one day be estimated by other factors. At present, using error-derived standard deviations to estimate uncertainty appears most practicable while still allowing forecasters to add value through the GFE GUIs.

5.1 Standard deviations from guidance and forecast

Perhaps the most straight-forward method to estimate uncertainty is to compute the standard deviation from the differences between MOS and/or model guidance products for the period in question. Table 6 is an example of two different forecasts, one for highs in the lower 80s, and another for lows in the 40s. This technique could easily be automated and used as a default unless the meteorologist wished to make changes or assign weights.

The example in Table 6 shows a set of forecasts from the CCF and four guidance products. Means and standard deviations were computed using a spreadsheet. The standard deviations appear smaller than what might be suggested from the data in Tables 1 and 2 above. Table 7 shows examples of standard deviations for a sampling of forecast ranges above and below 5°F. The differences in Table 6 and Table 7 suggest that simply calculating standard deviations from the range of CCF and guidance products for a given day may underestimate of uncertainty.

CCF	MAV	MET	UKMET	Mean	Forecast range	Standard Deviation
82	84	81	81	82.0	3	1.225
46	44	49	41	45.0	8	2.915

Table 6. Example of single period forecasts to derive the mean and standard deviation of CCF and guidance shown. The range of forecast values implies high certainty.

Period 1 Forecasts	CCF Std Dev	MAV Std Dev	MET Std Dev	UKMET Std Dev
Total Fcst range <5°F	2.55	3.08	3.02	3.03
Total Fcst range ≥5°F	3.53	4.51	3.98	5.53

Table 7. Standard Deviations computed from forecast errors when the total range of CCF and guidance products were less than 5°F and equal to or greater than 5°F.

5.2 Other possibilities for calculating standard deviations

Sometimes, a meteorologist may find the weather to be in a relatively stagnant pattern with very similar maximum and minimum temperatures for several days. He or she may also find that the MOS guidance has been very accurate for several days and is also relatively unchanged for the next periods' forecasts. The forecaster may find that the MOS guidance agrees with the current official forecast. In such instances, the meteorologists may have very high confidence in his/her predictions.

Table 8 shows first period standard deviations computed from first period errors. These were stratified based on the consistency of the forecasts from period 5 through period 1. The high and low forecasts from period 5 through period one determine the "delta range." If the forecast remained constant, regardless of accuracy, the delta range would be zero. If the forecast swung from 52 to 58, the delta range would be 6. Two sets of computations were made for each the CCF and MAV with delta ranges of zero through two and also six through 10.

6. Examples of Combined Distributions

Three keys assumptions must be made to allow WFO forecasters to create probabilistic temperature forecasts. First, it must be assumed that the CCF and guidance will be the mean that one might expect from an infinite number of similar events, and that the distribution around that mean is Gaussian. Second, appropriate standard deviations will be required for each forecast period, based on such inputs as forecaster and model error, consistency of CCF and guidance per period and from one cycle to the next. Third, information from the GFE GUIs should supply forecaster-determined weights into the computation of the final distribution. Several simple Excel spreadsheet examples are shown here to illustrate some of the possibilities.

Figure 8 shows an example of a probabilistic maximum temperature forecast that might be expected, given the MOS guidance (MAV and MET) and the CCF, for what might be a common, relatively stagnant weather pattern. For this example, each forecast was assigned the same standard deviation. The Gaussian distributions were then calculated and averaged to arrive at the final expected distribution.

	CCF Std Dev	MAV Std Dev
Delta range 0, 1, 2	**2.94**	**3.08**
Delta range 5-10	**3.53**	**3.81**

Table 8. Standard deviations for first period errors when the cycle-to-cycle "delta range" was as shown on left.

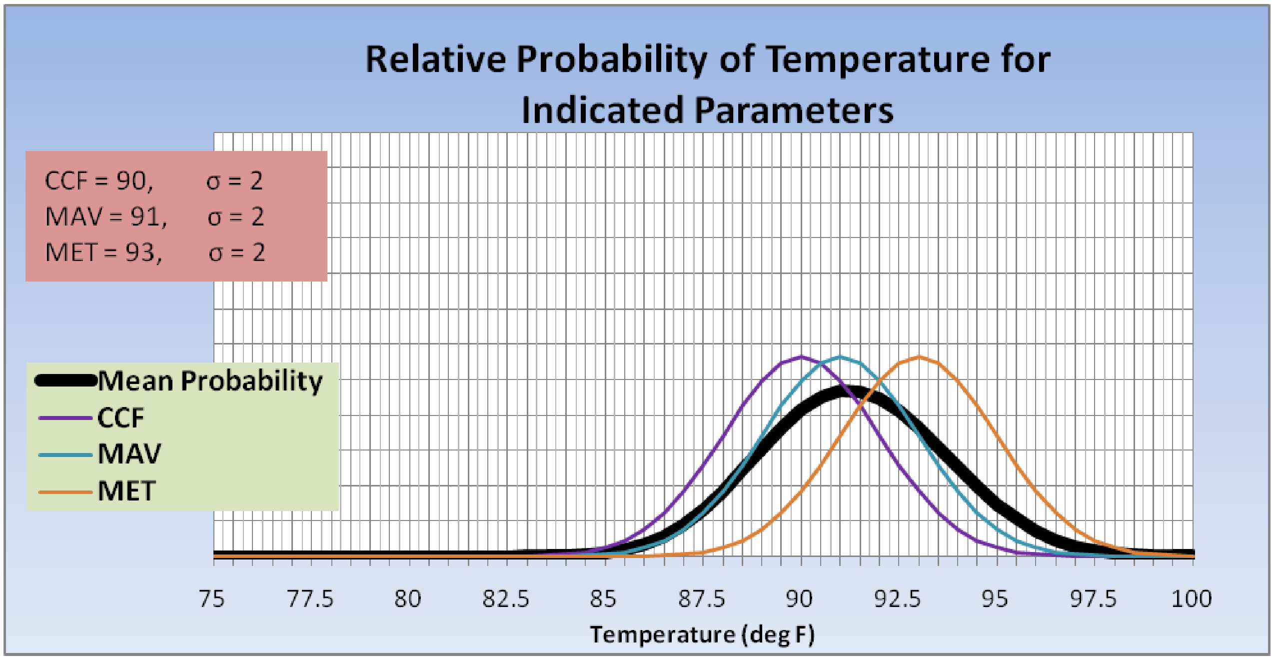

Figure 8. Simple individual and mean distributions for high temperature forecast. CCF is the WFO forecast. This might represent a quiescent weather pattern

A more diverse case is shown in Figure 9 where there is more disagreement among the guidance products. The resulting forecast has less certainty as indicated by the broader relative distribution.

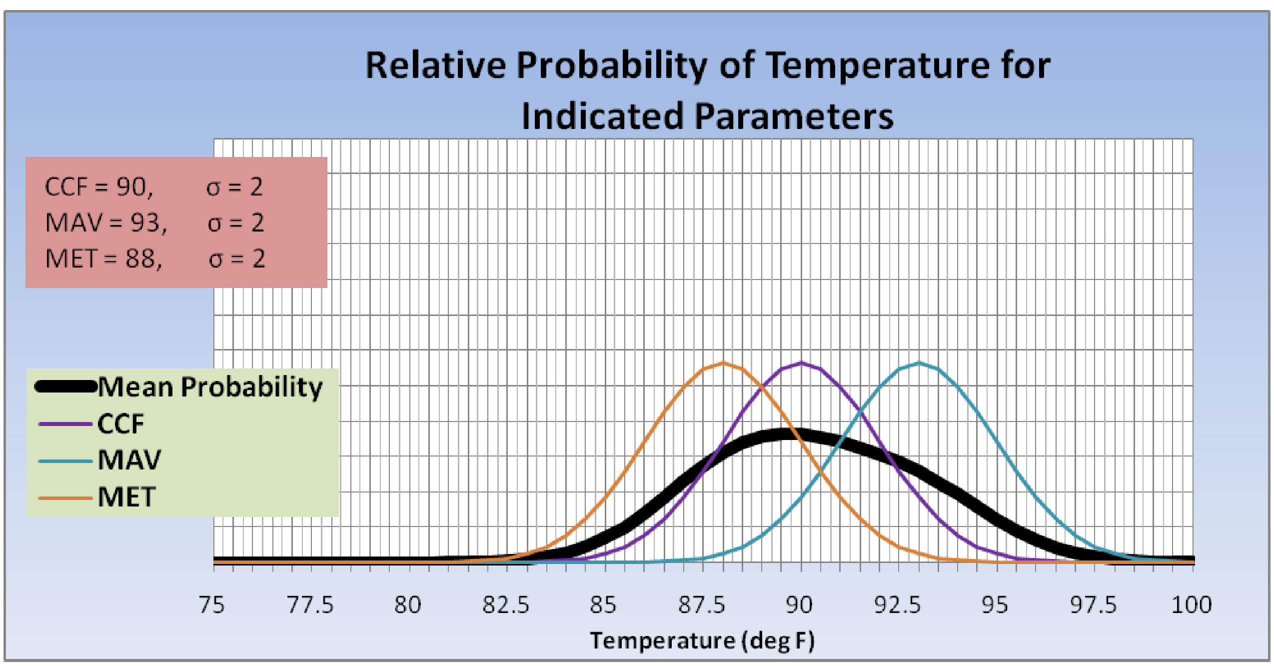

Figure 9. Same as Figure 5 but for a broader range of MOS and CCF solutions. Standard deviations (σ) all equal 1.

Figure 10 is an interesting example where there is significant model disagreement, but the forecaster apparently prefers the warmer guidance. The resultant distribution is bimodal, indicating to the astute user that there are two distinctly different outcomes possible in this forecast. This particular case might be the result of one model's development of a convective complex that will keep temperatures much cooler than the other model. The CCF indicates the forecaster hedged but preferred the warmer forecast.

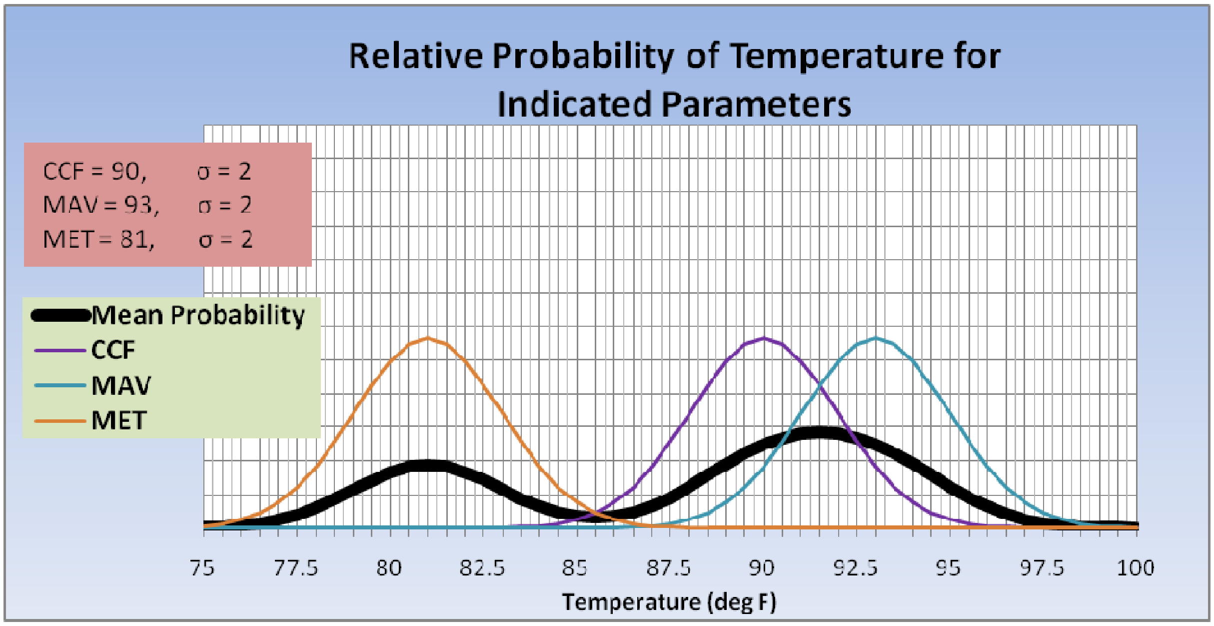

Figure 10. Simple example where the MET differs considerably from the MAV and CCF.

Figure 11 is an example where weighting factors were used. In this example, the forecaster doubled the weight of his/her own forecast, and halved the weight on the MET MOS, leaving the MAV with a weighting factor of 1.

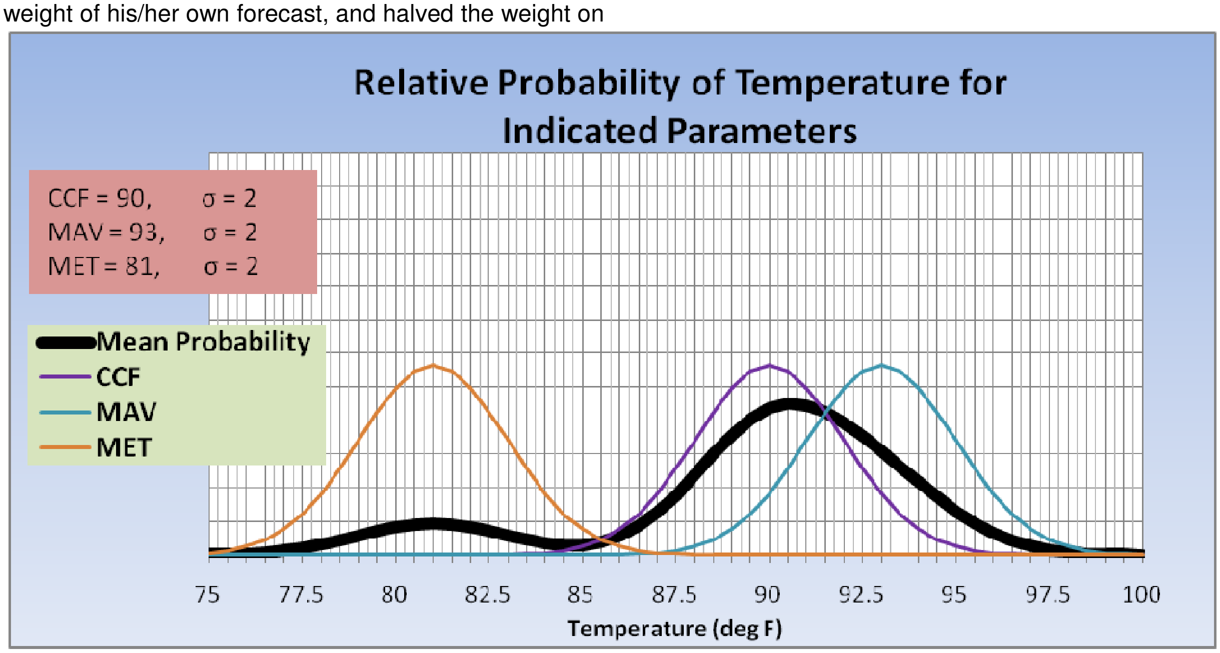

Figure 11. This depiction allows the forecaster to "double-weight" his/her own forecast, single weight the MAF and halve the weight of the MET.

Figure 12 is an example where the standard deviations for CCF and guidance are different. This generates a more complex distribution. It is interesting to note that both guidance products are warmer than the CCF, resulting in a distribution that suggests higher temperature may be expected.

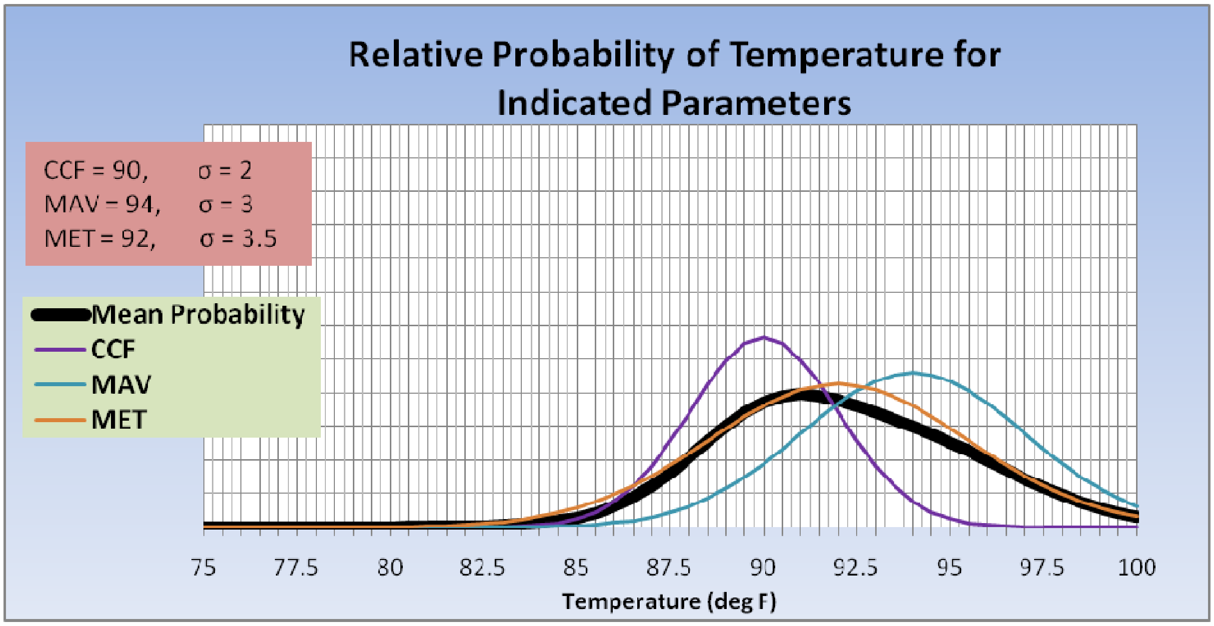

Figure 12. The standard deviations (σ) are different in this example and are as shown in the graphic, resulting in a noticeably different shape to the final mean probability curve.

7. Conclusion

This WFO approach to computing the uncertainty of forecast temperatures assumes each contributing forecast (MOS, model, or CCF) has a Gaussian distribution and that each contributing forecast is for the mean of that distribution. The standard deviations of those distributions are unique and can be approximated with relative accuracy by a variety of means. The forecast distributions can then be combined, with or without forecaster-assigned weights, to estimate the final expected distribution of max/min temperature.

The method shown here suggests that uncertainty could be based on standard deviations derived from general seasonal forecast error, MOS/CCF differences for a specific forecast, and the degree to which CCF forecasts and MOS guidance change from one cycle to the next. In addition, it may be important to identify and include the specific error-derived standard deviations of each forecaster, given that some forecasters are more skillful than others.

The overall concept provides forecasters with as much or as little control over the final forecast distribution as they wish. This is important because most forecasters routinely add value to guidance products. Also, forecasters must deal with sub-synoptic forecast issues which may be below the scale an ensemble approach could address. Forecasters can also apply local biases, on the sub-synoptic scale which ensemble members might not resolve.

Forecaster control is also important because they are often able to discern which model(s) are more likely to be accurate, which may not always include the GFS or NAM or their ensemble members. Max/min temperature forecasts are also available from the European Centre for Medium-Range Weather Forecasts, the United Kingdom Meteorological Office model and WFO locally run models and others. With WFO local control, any or all of these could be ingested into the Gridded Forecast Editor and used to compute the final expected temperature distributions.

The current configuration at NWS WFOs should be sufficient to allow the method described in this paper to be implemented. The Gridded Forecast Editor contains most, if not all of the “tools” required. Forecasters have already learned to use weighting techniques such as “slider bars” and have effectively been weighting forecasts since they began using MOS.

Communicating maximum and minimum temperature forecasts continues to be problematic in that only one number is currently provided. This is unfortunate since meteorologists are frequently aware of the changing uncertainty from one forecast period to another. Probabilistic temperature forecasts should allow meteorologists to convey that uncertainty in addition to a "best guess" number. Graphical depictions, such as those shown here could be made available through internet web sites for the discerning partner or customer. The method described here is practicable, efficient, and could be implemented with relative ease.

8. References

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

Wilks, D.S., 2005: Statistical Methods in the Atmospheric Sciences, Second Edition. Elsevier, Academic Press, New York, NY, 627 pp.