J5.5

SHORT TERM ENSEMBLE RIVER STAGE FORECASTS: APPLICATION

Xiaobiao Fan[1*], Mary Mullusky[1], Limin Wu[1], Edwin Welles[1], Joseph Ostrowski[2], Ned Pryor[2], and John Schaake[1]

[1] Hydrology Laboratory, Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, MD
[2] Middle Atlantic River Forecast Center, National Weather Service, NOAA, State College, PA

1. ABSTRACT

A new software system for assessing the uncertainty in short term river stage forecasts has been developed at the Office of Hydrologic Development (OHD) of the National Weather Service (NWS). This method explicitly accounts for the uncertainty in the future precipitation and temperature conditions, a major portion of the uncertainty in the river stage forecasts. The method requires deriving a forecast distribution from a given deterministic forecast using the joint
distribution of historical forecasts and observations. Once the forecast distribution is known, the climatological distribution is mapped into the new forecast distribution to create ensembles of precipitation and ensembles of temperature. The existing Ensemble Streamflow Prediction technique within the NWS River Forecast System (NWSRFS) is used to generate stream flow ensembles from the precipitation and temperature ensembles. Using the system, the Mid-Atlantic River Forecast Center (MARFC) has been producing forecast ensembles for the Juniata river basin in Pennsylvania on a daily basis since July 2002. Presented here are some the results for the Juniata basin obtained during the Summer and Winter.

2. INTRODUCTION

There are two primary sources of uncertainty in a river forecast system; the future meteorological conditions and the hydrological modeling. These uncertainties drive the need for probabilistic forecasting.

The portion of the NWSRFS that produces probabilistic forecasts of streamflow and streamflow-related variables is called the Ensemble Streamflow Prediction (ESP) system. The original version of ESP applied to long range forecasts by creating an ensemble of streamflow traces using multiple years of historically observed precipitation and temperature time series as possible future meteorological realizations (Day, 1985). The long-range forecasts were later enhanced by integrating meteorological forecasts and climate outlooks into the ESP System (Perica, S., 1998). OHD has now created a new method to generate short term ensembles to implicitly incorporate the skill of the short term meteorological forecast (Herr et al., 2002). This work was originally applied to precipitation forecasts. Additional work was required to handle short term temperature ensembles which drive the winter snow hydrology operations. Once the precipitation and temperature ensembles are both generated with this new method, the existing ESP technique can be used to generate short term streamflow ensembles. This paper discusses the application of this new method to generate the short term temperature ensembles and the resulting streamflow ensembles for the Juniata river basin in Pennsylvania. This method, in collaboration with the forecasters at MARFC, has been implemented for 10 forecast points on the Juniata River since July 2002 and is currently being tested on a daily basis.

2. FORMULATION

The goal of this method is to create an ensemble of meteorological inputs from a given deterministic forecast of that meteorological variable. The theory and formulation of this method is described in detail for precipitation in Herr et al. (2002) and is summarized here for temperature.

Let X be the observed temperature with realization x, and Y be the forecasted temperature with realization y. Let f_x be the density of X with a cumulative density function F_x and f_y be the density of Y with a cumulative density function F_y. The goal is to be able to compute the distribution of X given some forecast Y=y. To begin, variates X and Y are transformed into normal space. That is, variates Z_x and Z_y are defined so that
$z_x = Q^{-1}(F_x(x))$ and $z_y = Q^{-1}(F_y(y))$, where Q is the standard normal distribution. This transformation is referred to as the normal quantile transform (NQT). Next, the density $\phi(z_x,z_y)$ is modeled as bivariate normal with standard normal marginals and with parameter ρ, which is the Pearson's correlation coefficient between Z_x and Z_y.

Modeling ϕ with a bivariate normal density allows for the conditional density function $f_c(x|Y=y)$ to be computed as the conditional density $\phi_c(z_{X/}Z_Y=z_Y)$ which is known to be normal with mean $\mu=\rho z_Y$ and variance $\sigma^2= (1-\rho^2)$. This form of the distribution can be viewed as the climatology being shifted by the information contained in the forecast, so that as the skill of the forecast decreases (i.e., as ρ goes to 0), the conditional density $\phi_{X|Y}$ is just the marginal distribution F.

* *Corresponding author address*: Xiaobiao Fan, Hydrology Laboratory, National Weather Service, Silver Spring, MD 20910; e-mail: Xiaobiao.Fan@noaa.gov

3. APPLICATION

The method to implement the formulation described in Section 2 to create short term streamflow ensembles is threefold. First, in order to use historical data to construct the distributions of F_x and F_y parameters describing the distributions must be derived from the temperature data. Second, once the calibrated parameters are in place the method can be used to create temperature and precipitation ensembles. Third, using the existing ESP system with the new ensembles of temperature and precipitation, streamflow ensembles can be generated. The calibration of temperature parameters, creation of temperature ensembles and creation of streamflow ensembles are described in the following sections.

3.1 Calibration of Temperature Parameters

The calibration of the temperature parameters needed to construct the distributions of F_x and F_y was complicated by the types of forecasted and observed temperature data archived for the Juniata. Archives of forecast temperatures included only point forecasts of daily maximum and daily minimum temperatures for a sparse network. Archived observed data included a 6-hour time series of mean areal temperatures for each of the basins of interest.

The observed time series of 6-hour mean areal temperature was derived from a mean areal maximum and a mean areal minimum temperature with a fixed diurnal cycle. Since this fixed diurnal cycle was known, a simple back computation was applied to the 6-hour time series of mean areal temperatures to determine the mean areal maximum and the mean areal minimum. Now, both the forecast and observed temperature data had daily maximum and minimum value, however the forecast archive was of point values and the observed archive was of mean areal values. To resolve this question, a hypothetical station was created at the centroid of each basin and using the NWSRFS methods of estimating missing data a time series of maximum and minimum daily temperature was created at this point. Point data were compared to the mean areal maximum and minimum temperature for that basin. This comparison is shown in Figure 1 for the minimum daily temperature and Figure 2 for the maximum daily temperature for Huntingdon, PA. The x-axis represents the maximum/minimum temperature derived as a point estimate and the y-axis is the maximum/minimum temperature derived from the mean areal temperature. Because of the linearity around the 45° line it was decided to use the maximum and minimum temperature derived from the mean areal temperature as this was an existing data source already available at the RFCs.

In order to use historical data to construct distributions Fx and Fy two statistics are needed: (1) the daily average and (2) the daily coefficient of variation. Because the statistics are noisy on a daily time step the

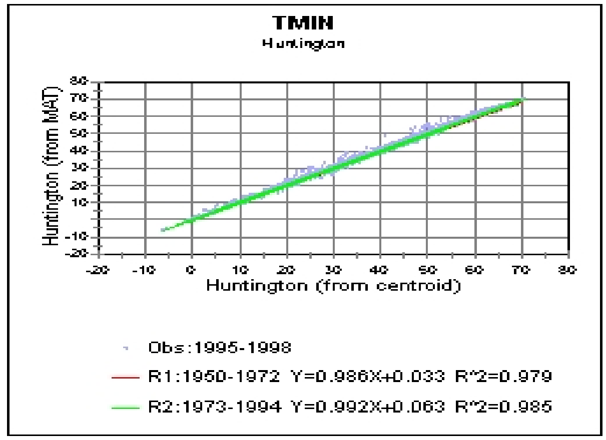

Figure 1: Comparison Between Areal Minimum Temperature and Point Minimum Temperature for Huntington.

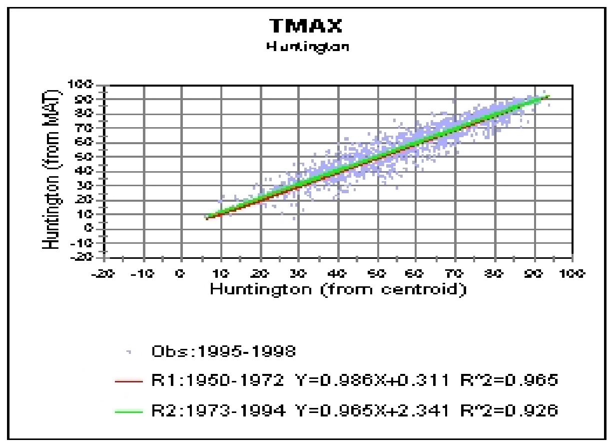

Figure 2: Comparison Between Areal Maximum Temperature and Point Maximum Temperature for Huntington.

statistics are smoothed with a three component Fourier series, then the average and coefficient of variation can be used to estimate a distribution. For temperature the distribution is assumed to be Gaussian. Figure 3 shows an example of the smoothed coefficient of variation. The final parameter needed for the calibration is the correlation coefficient ρ between Z_x and Z_y.

The historical data now consists of a forecasted and observed maximum and minimum temperature value. The formulation described in Section 2 can then be followed for both variables thus creating a set of maximum ensembles and a set of minimum ensembles. The two sets of ensembles are then temporally disaggregated to create a 6-hour time series of temperature ensembles. The calibration statistics needed to create an ensemble of maximum temperatures include; average maximum observed temperature, coefficient of variation of maximum observed temperature, average maximum forecasted temperature, coefficient of variation of maximum forecasted temperature, and the correlation between maximum observed temperature and the maximum forecasted temperature. The same set of statistics are also needed for the minimum temperature. Although this is a large number of parameters, since there are 365 sets of statistics, they are computed off-line prior to forecast time and the process is fully automated.

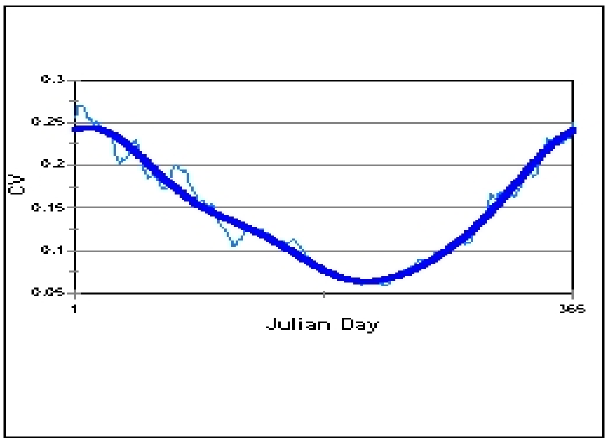

Figure 3: Example of Smoothed vs Unsmoothed Coefficient of Variation

3.2 Deriving Temperature Ensembles

The first step in deriving the temperature ensembles is to compute a deterministic forecast of maximum and minimum temperatures from the forecasted mean areal temperature for the basin of interest. This is accomplished by back calculating the known fixed diurnal cycle to the 6-hour time series to compute the forecasted daily maximum and daily minimum.

The second step is to calculate from the calibration statistics and the deterministic forecast for both the maximum and minimum temperature the conditional distribution function $F_{X|Y}$. $F_{X|Y}$ is the conditional distribution function of X given the deterministic forecast Y=y computed as $F_{X|Y}(x|y)=\phi_c(z_x;\ \mu,\ \sigma)$ where ϕ_c is the conditional density of the normal bivariate which is known to be normal with mean $\mu=\rho z_Y$ and variance $\sigma^2= (1-\rho^2)$.

The third step is to use a distribution mapping technique to map the two conditional distributions into an ensemble of maximums and an ensemble of minimums. The distribution mapping technique works as follows for maximum temperatures; the maximum temperatures are ranked and assigned a probability value based upon its rank, the year is then assigned a new temperature value by performing the inverse of the normal quantile transform for that probability level on the conditional distribution function.

After these steps are performed to create an ensemble of maximums and an ensemble of minimums, the final step is to temporally disaggregate the maximum and minimum ensembles to produce one ensemble of a 6-hour time series of temperature. The temporal disaggregation is accomplished with user defined parameters that drive a diurnal cycle as a function of maximum and minimum temperature. A graphical user interface was developed to assist users in determining these parameters and is shown in Figure 4. At each time step the user defines the temperature parameter TP and the 6-hour temperature is calculated as $T_6= T_{min} + TP(T_{max}-T_{min})$.

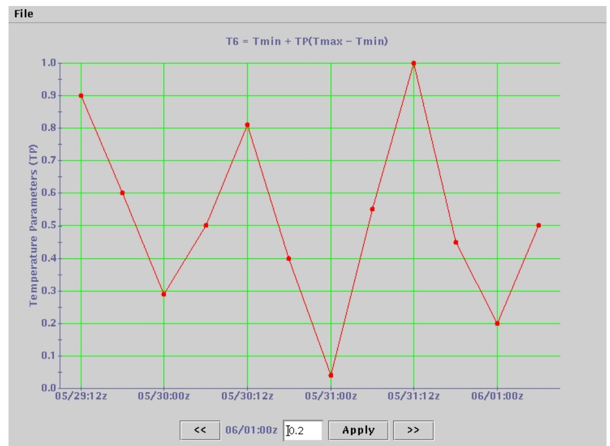

Figure 4: Temporal Disaggregation Graphical User Interface.

3.3 Deriving Streamflow Ensembles

Once the temperature and precipitation ensembles have been created the existing ESP system in NWSRFS can be used to compute streamflow ensembles. The resulting streamflow ensembles are conditioned on the initial conditions of the model and implicitly incorporate the uncertainty of the deterministic meteorological forecasts.

4. EXPERIENCES

The MARFC has been producing daily short term probabilistic forecasts for 10 basins on the Juniata River since June of 2001. Originally, the streamflow forecasts were computed using this technique for precipitation. Since July 2002 the short term probabilistic forecasts account for the meteorological uncertainty of both precipitation and temperature. Since the temperature ensembles are only used in the snow models of NWSRFS, the true value of the temperature ensembles will be assessed in the coming winter. The daily short-term forecasts are available at http://www.erh.noaa.gov/er/marfc/AHPS/5-day.htm .

To begin to understand the uncertainty in a temperature forecast, temperature ensembles were derived for winter and summer conditions and are shown in Figures 5 and 6, respectively. The variability of the ensembles provide information about the confidence of the forecast, the large variability indicating more uncertainty in a given forecast. Figures 5 and 6 show that the uncertainty in the winter temperature forecasts is larger than the summer temperature forecasts. Additionally, the winter ensembles show one trace having an opposite diurnal cycle than all of the other traces even with the user controlled diurnal cycle. This initially was cause for concern and a thorough evaluation of the procedure was performed. The historical data that defines the joint distribution that approximately 2.5% of the time the maximum temperature was less than the minimum temperature. This was done in an effort to simulate inversions with the fixed diurnal cycle. Since this was a part of the joint distribution it is actually correct that in some resulting ensembles the maximum temperature will be less than the minimum causing a chance of an inversion occurring with a given forecast of temperature.

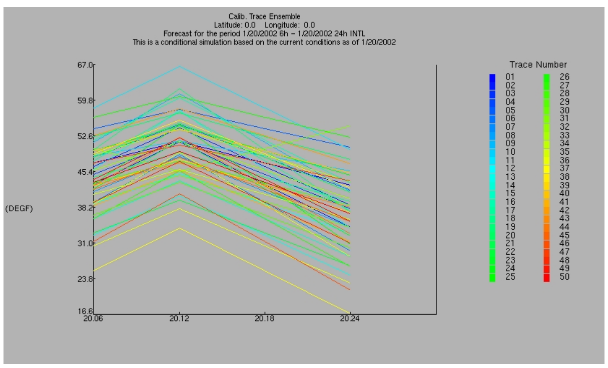

Figure 5: Temperature Ensembles for Winter Conditions at Huntingdon.

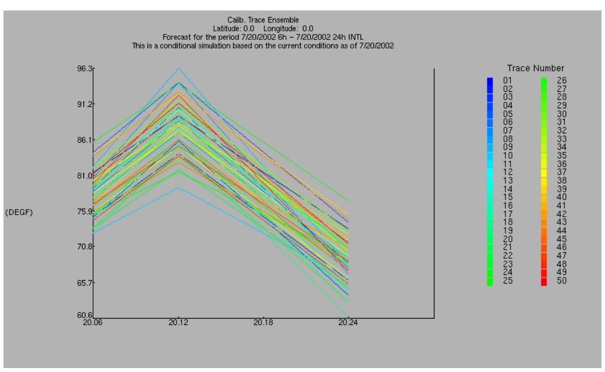

Figure 6: Temperature Ensembles for Summer Conditions at Huntingdon.

5. CONCLUSIONS

In this paper, we described a process that assesses the uncertainty in short term streamflow forecasts. This method requires deriving a forecast distribution for a given deterministic forecast from the joint distribution of historical forecasts and observations, and can effectively capture the uncertainty in the future meteorological inputs which include precipitation and temperature conditions.

The system has been operational at the MARFC over ten forecast basins and the process appears to be working well. Additional testing and vetting of the science are required before the process is implemented nationwide.

6. REFERENCES

Day, G.N., 1985: "Extended Streamflow Forecasting using NWSRFS," Journal of Water Resources Planning and Management, 111(2), 157-170.

Herr, H., E. Welles, M. Mullusky, L. Wu, J. Schaake, 2002: "Simplified Short Term Precipitation Ensemble Forecasts: Theory," Preprints of the Symposium on Observations, data Assimilation and Probabilistic Prediction, 82nd AMS Annual Meeting, Orlando, Florida.

Perica S., 1998: Integration of Meteorological Forecasts/Climate Outlooks into an Ensemble Streamflow Prediction System. Preprints of the 14th Conference on Probability and Statistics in the Atmospheric Sciences, 78th AMS Annual Meeting, Phoenix, Arizona, 130-133.