J1.17

SIMPLIFIED SHORT TERM PRECIPITATION ENSEMBLE FORECASTS: THEORY

Hank Herr, Edwin Welles, Mary Mullusky, Limin Wu, John Schaake
The Hydrology Laboratory, National Weather Service, Silver Spring, MD 20910

1. ABSTRACT

The National Weather Service (NWS) has been requested by a variety of users to provide hydrologic forecasts that explicitly account for the uncertainty in a forecast. A primary source of uncertainty is the precipitation and temperature forecasts used as input to produce hydrologic forecasts. While several methods have been explored for quantifying the uncertainty, it has been found that ensemble methods best satisfy the complex mix of operational and scientific requirements. Presented herein is an ensemble approach for generating hydrologic forecasts that account for uncertainty in forecast precipitation and temperature. It uses existing NWS data streams of quantitative precipitation forecasts (QPFs) and observed precipitation, and is relatively easy to calibrate.

2. INTRODUCTION

At the request of hydrologic forecast users, the National Weather Service (NWS) has been developing techniques for computing probabilistic forecasts which assess the uncertainty of possible events. Numerous methods exist for computing such forecasts and for communicating the probabilities to a user. The NWS has used the method of ensembles for many years and finds that this method best provides for the ability to handle the varied hydrologic conditions across the nation while being the most amenable to interfacing with input meteorological forecasts.

There are two primary sources of uncertainty in a river forecast: the future meteorological conditions and the hydrologic modeling. Both sources of uncertainty must be addressed to effectively define the probability of future events. The focus of this paper is generating ensembles of the meteorological input variables to capture the uncertainty in the forecast precipitation. Additional work is required to handle temperature forecasts. A number of methods have been proposed for computing the precipitation ensembles, and several have been implemented into the NWS operational software. However, the existing methods either apply to long range forecasts, or require a substantial implementation effort. The NWS needs an effective method for computing short term precipitation (and temperature) ensembles which can be implemented easily.

Corresponding author address: Edwin Welles, Hydrology Laboratory, National Weather Service, Silver Spring, MD 20912; e-mail: Edwin.Welles@noaa.gov

3. FORMULATION

3.1 Marginal Distributions

Let X be the observed precipitation amount with realization x, and Y be the forecasted precipitation amount with realization y. Let f_X be the density of X and f_Y be the density of Y. In order to account for the probability associated with the observed or forecasted precipitation amount being zero, density f_X incorporates the Dirac delta function, δ, (Edwards and Penney, 1994) as follows:

$$f_X(x) = (1 - p_{0X})\delta(x) + p_{0X}f_{XC}(x \mid x > 0),$$

where p_{0X} is the observed probability of precipitation and f_{xc} is the conditional density function of X where x > 0. Hence, the cumulative distribution function, F, has the form

$$F_X(x) = 1 - p_{0X} + p_{0X}F_{XC}(x \mid x > 0)$$

where F_{XC} is the conditional distribution function of X where x>0.

The forecast precipitation density f_Y and cdf F_Y have similar forms. Figure 1 provides examples of the cdfs F_X and F_Y.

3.2 Characteristics of F

The bivariate distribution function F(x, y), with marginals F_X and F_Y, is an interesting distribution in that both F_X and F_Y have a discrete component associated with the probability of no precipitation and a continuous component associated with the event that precipitation occurs. This means that there may be distinct probability that X is zero and

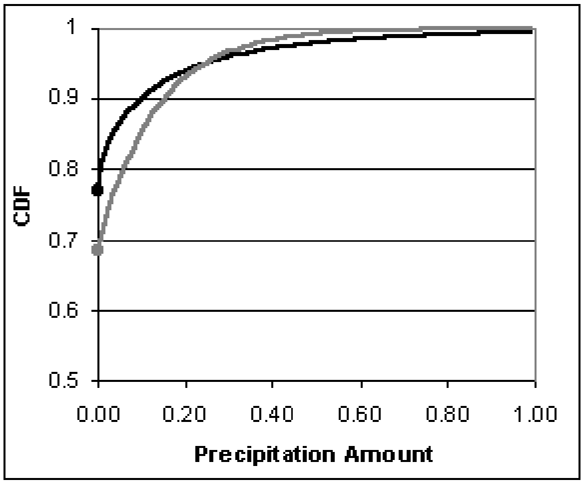

Figure 1: Example observed (black) and forecasted (gray) cdfs with circles displaying value of 1 - p_0.

Y is positive and vice-versa. Any method used to model F must be able to account for this.

3.3 Bivariate Normal Distribution Φ

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, or 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.

3.4 Characteristics of ρ

The correlation coefficient ρ is dependent on the spatial scale of the forecast, the width of the time interval of the forecast, and the lead time of the forecast. Kelly and Krzysztofowicz (1997) have also shown that ρ is the Spearman's rank correlation coefficient between X and Y in the original space, and serves as a measure of the skill of the forecaster, being 1 for a perfect forecast and 0 for a completely unskilled forecast.

3.5 Conditional Distribution

Modeling ϕ with a bivariate normal density allows for the conditional density function $f_C(x \mid Y = y)$ to be computed as the conditional density $\phi_C(z_X \mid Z_Y = z_Y)$ 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.

4. CALIBRATION

4.1 Smoothed Climatology

The distributions of observed and forecasted precipitation amounts are noisy at the daily time step, meaning that the distribution for one day may differ markedly from the distribution on the next day. This is caused by severe storms present in the historical record on only a few days, thus skewing the distribution for those days.

Hence, in order to use the historical data to construct distributions F_x and F_y, the data is first smoothed. Three statistics are the object of the smoothing: (1) the daily probability of precipitation (p_{0X} or p_{0Y}), (2) the daily average of all non-zero events (the conditional average or cavg), and (3) the daily coefficient of variation of all non-zero events (the conditional coefficient of variation or ccv). These three statistics are smoothed with a three component Fourier series, and then the cavg and ccv are used to estimate a distribution, which, when combined with p_{0X} or p_{0Y}, define F_x or F_y. Distributions currently used include the Gamma and the Weibull distributions, both described in Evans et al. (1993), although any reasonable distribution is possible. Figure 2 provides an example of smoothed probability of precipitation data.

4.2 Calibrated Parameters

Calibration of this process consists of (1)

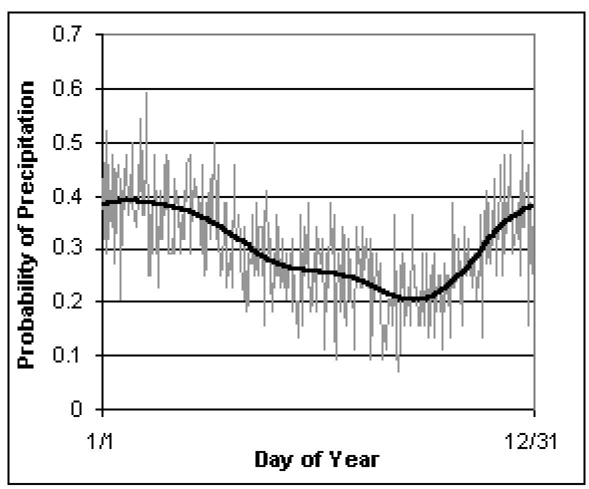

Figure 2: Example of smoothed (black) and unsmoothed (gray) probability of precipitation.

computing the smoothed daily statistics, as described above and (2) computing the correlation coefficient ρ between Z_X and Z_Y. The smoothed statistics are then used at run time to estimate F_x and F_y. 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.

5. APPLICATION

5.1 Constructing an Ensemble

An ensemble of precipitation amounts is constructed using the climatological record, where each year of data corresponds to one time series in the ensemble. For a given day and a given year, k, the process is as follows:

1. The year is ranked according to the amount of precipitation that occurred on that day in that year relative to other years. For zero events, the ranks are assigned randomly. However, a “nearest neighbor” technique is currently being researched that will assign the ranks based on how “close” the zero event is to a non-zero event, both temporally and spatially.

2. The year has a probability, p_K, assigned to it based upon its rank.

3. The year has a value for variate Z_X assigned to it, which is computed as the inverse of the conditional distribution of the density function ϕ_C, described in section 3.5, or $z_{XK} = \Phi_c^{-1}(p_K; \mu, \sigma)$.

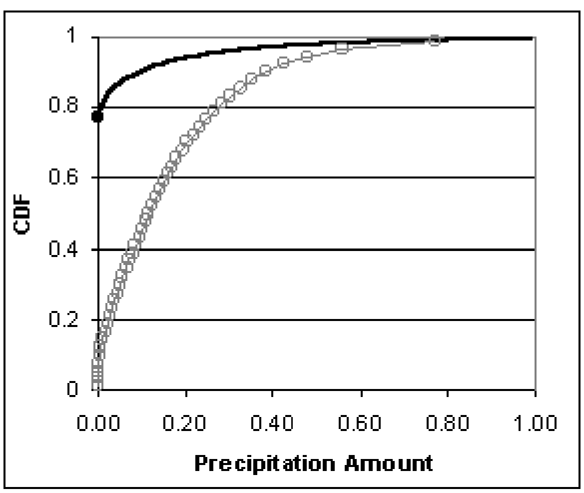

Figure 3: Example of the PQPF (gray line) and computed ensemble points (gray circles) relative to the distribution of observed precipitation (black).

4. The year has a precipitation amount assigned to it by performing the inverse of the NQT, or $x_K = F_{X|Y}^{-1}(p_K)$ where $F_{X|Y}$ is the conditional distribution function of X given Y = y computed as $F_{X|Y}(x|y) = \Phi_c(z_x; \mu, \sigma)$.

Figure 3 is an example of a possible output PQPF, providing the computed ensemble points as well as the historical distribution of observed precipitation.

5.2 Characteristics of this Ensemble Approach

By using the historical record to construct the ensembles, the spatial and temporal characteristics of the rainfall is captured. For example, if precipitation over two basins is highly correlated, this characteristic will be captured in the climatological record, so that the ensembles that are constructed will also capture this characteristic. Furthermore, by ranking the zeros and shifting the entire distribution the intermittent character of precipitation is preserved. When more rain falls, it falls on more days and not just in larger amounts.

6. CONCLUSION

A new method for generating a PQPF has been developed that depends only upon existing NWS data streams. In addition, a new ensemble generation scheme has been developed that effectively links the physical characteristics of the local precipitation fields and the statistically generated precipitation amounts. A simple calibration is required to extract parameters that define the relationship between the observed and the forecast precipitation. The operational mode then generates a PQPF, samples that PQPF, and assigns values to the time series based on the ranks of observed record. Paper JP1.19 of this conference presents the results of an operational demonstration of this approach.

REFERENCES

Edwards, Jr., C. H. and D. E. Penney, *Elementary Differential Equations*, Prentice Hill, Englewood Cliffs, NJ, 1994.

Evans, M., N. Hastings, and B Peacock, *Statistical Distributions*, Second Edition, John Wiley & Sons, New York, NY, 1993.

Kelly, K. S., and R. Krzysztofowicz, “A Bivariate Meta-Gaussian Density for Use in Hydrology,”, *Stochastic Hydrology and Hydraulics*, 11, 17-31, 1997.