PROBABILISTIC FLOOD FORECAST:

EXACT AND APPROXIMATE PREDICTIVE DISTRIBUTIONS

By

Roman Krzysztofowicz*

Department of Systems Engineering and Department of Statistics

University of Virginia

P.O. Box 400747

Charlottesville,Virginia 22904–4747, USA

Research Paper RK–0802

http://www.faculty.virginia.edu/rk/

September 2008

Revised December 2009

* Tel.: +1-434-982-2067; fax: +1-434-982-2972.

E-mail address: rk@virginia.edu (R. Krzysztofowicz).

ABSTRACT

For quantification of predictive uncertainty at the forecast time t_0, the future hydrograph is viewed as a discrete-time continuous-state stochastic process $\{H_n : n = 1, ..., N\}$, where H_n is the river stage at time instance $t_n > t_0$. The *probabilistic flood forecast* (PFF) should specify a sequence of exceedance functions $\{\bar{F}_n : n = 1, ..., N\}$ such that $\bar{F}_n(h) = P(Z_n > h)$, where P stands for probability, and Z_n is the maximum river stage within time interval $(t_0, t_n]$, practically $Z_n = \max\{H_1, ..., H_n\}$. This article presents a method for deriving the exact PFF from a *probabilistic stage transition forecast* (PSTF) produced by the *Bayesian forecasting system* (BFS). It then recalls (i) the bounds on $\bar{F}_n$, which can be derived cheaply from a *probabilistic river stage forecast* (PRSF) produced by a simpler version of the BFS, and (ii) an approximation to $\bar{F}_n$, which can be constructed from the bounds via a *recursive linear interpolator* (RLI) without information about the stochastic dependence in the process $\{H_1, ..., H_n\}$, as this information is not provided by the PRSF. The RLI is substantiated by comparing the approximate PFF against the exact PFF. Being reasonably accurate and very simple, the RLI may be attractive for real-time flood forecasting in systems of lesser complexity. All methods are illustrated with a case study for a 1430 km^2 headwater basin wherein the PFF is produced for a 72-h interval discretized into 6-h steps.

Keywords: Stochastic processes; Statistical analysis; Probability; Rivers; Floods

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Probabilistic Flood Forecast

Flood warning-response systems are designed and operated to mitigate the consequences of extreme river stages induced by heavy rainfall or rapid snowmelt. They achieve their purpose by supporting decisions of emergency managers and floodplain dwellers (Krzysztofowicz and Davis, 1983, 1984). The key one is to decide whether or not to issue a flood warning for a zone of the floodplain (Krzysztofowicz, 1993). To make such a decision optimally, the system needs quantitative information about the predictive uncertainty (Krzysztofowicz, 2001) associated with the maximum river stage within a time interval. To implement the optimal decision effectively, the system needs sufficient lead time. This, in turn, requires that a hydrologic forecast be based on a meteorological forecast (Georgakakos, 1986; Lardet and Obled, 1994), and that the uncertainty in both forecasts be integrated.

The purpose of the *probabilistic flood forecast* (PFF) is to provide information needed by a flood warning system. As such, the PFF should specify a sequence of exceedance functions of maximum river stages within time intervals that form a nested set (Hoffman, 1975): $(t_0, t_1] \subset (t_0, t_2] \subset \cdots \subset (t_0, t_N]$, where t_0 is the forecast time and $t_1 < t_2 < \cdots < t_N$ are future times. The notion of the nested time intervals is essential for two reasons. First, it is needed to quantify the total risk of flooding from a rainfall event or a snowmelt event (or a portion thereof that is covered by a rainfall or temperature forecast upon which the PFF is based). Second, it is needed to capture the uncertainty about the timing of the flood crest and thereby to allow the decisions to be made dynamically and adaptively (rather than statically). This article shows how to construct the PFF, exactly or approximately, from the outputs of a Bayesian forecasting system (Krzysztofowicz and Maranzano, 2004).

1.2 Bayesian Forecasting System

The Bayesian theory provides a general mathematical and methodological framework for probabilistic forecasting of river processes (time series of stages, discharges, or volumes) via a deterministic hydrologic model of any complexity (Krzysztofowicz, 1999). For short-term forecasting in small-to-medium headwater basins, the theory was implemented as an analytic-numerical *Bayesian forecasting system* (BFS). Two versions of this system have been developed to date for forecasting a discrete-time, continuous-state stochastic process $\{H_n : n = 1, ..., N\}$ with lead time of N time steps. Each version takes a *probabilistic quantitative precipitation forecast* (PQPF) as input and employs a deterministic hydrologic model to calculate the response of a river basin to precipitation.

The first BFS outputs a *probabilistic river stage forecast* (PRSF) in the form of a sequence of predictive n-step transition density functions (Krzysztofowicz, 2002a):

$$\{\psi_n : n = 1, ..., N\},$$

where

$$\psi_n(h_n) = p(h_n|h_0, PQPF, \mathbf{u}_0)\,.$$

That is, ψ_n is the predictive density function p of river stage H_n at time t_n, **conditional on** (i) the river stage $H_0 = h_0$ observed at the forecast time t_0, (ii) the $PQPF$ for the river basin, (iii) the vector $\mathbf{u}_0$ of deterministic inputs to the hydrologic model (except future precipitation), which are needed to produce a deterministic forecast and whose values vary from one forecast time to the next (e.g., initial model states), and (iv) the hydrologic model of the river basin (implicit in $\mathbf{u}_0$) with parameters estimated for the given forecast point. With respect to the stochastic process $\{H_n : n = 1, ..., N\}$, function ψ_n characterizes the uncertainty in the n-*step transition* from the observed (initial) river stage $H_0 = h_0$ to the future river stage H_n, conditional on $PQPF$ and $\mathbf{u}_0$.

The conditioning on $(h_0, PQPF, \mathbf{u}_0)$, whose value is fixed at the forecast time, is suppressed in the operational notation ψ_n. But it is crucial for the understanding: this conditioning shows that the river stage process is forecasted as a fully dependent stochastic process (of order N). Still the PRSF does not provide a complete characterization of this process: it does not provide the predictive joint density function of the river stages $H_1, ..., H_N$.

The second BFS outputs a *probabilistic stage transition forecast* (PSTF) in the form of a sequence of families of predictive one-step transition density functions (Krzysztofowicz and Maranzano, 2004): ψ_1 and

$$\{\theta_n(\cdot|h_{n-1}, ..., h_1) : \text{ all } h_1, ..., h_{n-1};\ n = 2, ..., N\},$$

where

$$\theta_n(h_n|h_{n-1}, ..., h_1) = p(h_n|h_{n-1}, ..., h_1, h_0, PQPF, \mathbf{u}_0).$$

That is, $\theta_n(\cdot|h_{n-1}, ..., h_1)$ is the predictive density function p of river stage H_n at time t_n, **conditional on** (i) the river stages $H_{n-1} = h_{n-1}, ..., H_1 = h_1$ hypothesized at the preceding times, (ii) the river stage $H_0 = h_0$ observed at the forecast time, (iii) the $PQPF$, (iv) the vector $\mathbf{u}_0$ of deterministic inputs to the hydrologic model, and (v) the hydrologic model (implicit in $\mathbf{u}_0$). With respect to the stochastic process $\{H_n : n = 1, ..., N\}$, function $\theta_n(\cdot|h_{n-1}, ..., h_1)$ characterizes the uncertainty in the *one-step transition* from the observed (initial) river stage $H_0 = h_0$ and the hypothesized (preceding) river stages $H_1 = h_1, ..., H_{n-1} = h_{n-1}$, to the next river stage H_n, conditional on $PQPF$ and $\mathbf{u}_0$. As before, the conditioning on $(h_0, PQPF, \mathbf{u}_0)$, whose value is fixed at the forecast time, is suppressed in the operational notation θ_n.

The PSTF is exact in the sense that the product of the predictive one-step transition density functions gives the predictive joint density function of the river stages $H_1, ..., H_N$:

$$\xi_N(h_1, ..., h_N) = \psi_1(h_1) \prod_{n=2}^{N} \theta_n(h_n | h_{n-1}, ..., h_1).$$

Two properties of ξ_N are evident: (i) that it is conditional on $(h_0, PQPF, \mathbf{u}_0)$, and (ii) that it predicts the river stage process as a fully dependent stochastic process (of order N). Therefore, the PSTF provides a complete, analytic characterization of predictive uncertainty about the river stage process $\{H_n : n = 1, ..., N\}$.

A previous article (Krzysztofowicz, 2002b) showed how to construct bounds on and approximations to the PFF from a PRSF alone. This article shows how to construct the exact PFF from a PSTF — more specifically, from the source elements which are output by the BFS and which are used to construct the PSTF (and which can, as well, be used to construct the PRSF).

1.3 Overview

Section 2 formally defines the PFF. Section 3 derives the theoretical relation between the PSTF and the PFF and presents a numerical algorithm for efficient calculation of the PFF. Section 4 reports a case study, explains three kinds of forecast products, and describes an analytic procedure for updating the PFF based on a partially updated PQPF. Section 5 reviews the theory of bounds on the PFF that can be derived from the PRSF. Section 6 recalls a recursive linear interpolator (RLI) based on the bounds, which processes a PRSF into an estimate of the PFF; then by comparing this estimate with the exact PFF derived from the PSTF, it reports the first empirical substantiation of the RLI.

2. DEFINITION OF PROBABILISTIC FLOOD FORECAST

Let t_0 denote the *forecast time,* and let t_n $(n = 1, ..., N)$ denote the time at which the river stage being forecasted will be observed. The *lead time* of the forecast prepared at time t_0 for time t_n is $t_n - t_0$. For simplicity, index n itself will sometimes be referred to as lead time. Next define

H_n — *river stage* at time t_n; it is a continuous variate which may take any value above the gauge datum.

Z_n — *maximum river stage* within time interval $(t_0, t_n]$; it is a continuous variate which for a discrete-time river stage process $\{H_1, ..., H_n\}$ is defined as

$$Z_n = \max\{H_1, ..., H_n\}. \tag{1}$$

$\bar{F}_n$ — *exceedance function* of maximum river stage Z_n, such that for any level h

$$\begin{aligned} \bar{F}_n(h) &= P(Z_n > h) \\ &= 1 - P(Z_n \leq h) \\ &= 1 - P(H_1 \leq h, ..., H_n \leq h); \end{aligned} \tag{2}$$

that is, $\bar{F}_n(h)$ is the probability of the maximum river stage Z_n within time interval $(t_0, t_n]$ exceeding level h. Alternatively, it is the probability of at least one among the n river stages $H_1, ..., H_n$ exceeding level h.

The PFF is defined henceforth as a sequence of exceedance functions

$$\{\bar{F}_n : n = 1, ..., N\}. \tag{3}$$

Given the PFF, the probability distributions needed for a flood warning system can readily be obtained (Kelly and Krzysztofowicz, 1994).

3. THEORY OF PROBABILISTIC FLOOD FORECAST

3.1 Uncertainty Processors

In the BFS, the total uncertainty is decomposed into precipitation uncertainty and hydrologic uncertainty. *Precipitation uncertainty* is associated with the total basin average precipitation amount during the period covered by the PQPF. *Hydrologic uncertainty* is the aggregate of all uncertainties arising from sources other than the total basin average precipitation amount.

The two sources of uncertainty are quantified independently and then are integrated. For this purpose, two processors are attached to a deterministic hydrologic model. The precipitation uncertainty processor maps precipitation uncertainty (input uncertainty quantified by the PQPF) into output uncertainty under the hypothesis that there is no hydrologic uncertainty. The hydrologic uncertainty processor quantifies hydrologic uncertainty under the hypothesis that there is no precipitation uncertainty. Then the two uncertainties are optimally integrated to produce a PSTF.

3.2 Precipitation Forecast

The PQPF for a river basin consists of two parts: (i) a probabilistic forecast of the basin average precipitation amount to be accumulated during the period, W (the total precipitation amount, for short), and (ii) a deterministic forecast of the spatio-temporal disaggregation of W. Specifically, the PQPF specifies three elements: (i) the probability of precipitation occurrence during the period and over the basin,

$$\nu = P(V = 1), \tag{4}$$

such that $0 \leq \nu \leq 1$, where V is an indicator such the $V = 1$ if and only if precipitation occurs $(W > 0)$ and $V = 0$ otherwise; (ii) the distribution function T_1 of the total precipitation amount W, conditional on the hypothesis that $V = 1$, which for any $w > 0$ specifies $T_1(w) = P(W \leq w|V = 1)$; and (iii) the matrix $\boldsymbol{\xi}$ of expected disaggregation factors, conditional on the hypothesis that $V = 1$ (for instance, a 4×5 matrix disaggregates any total precipitation amount w into 4 sub-periods and 5 subbasins).

The uncertainty associated with the intermittence of the precipitation process plays an important role in small river basins because the indicator V turns out to be one of the predictors explaining (in part) hydrologic uncertainty. For this reason, the precipitation event $V = v$, $v \in \{0, 1\}$, conditions various elements.

3.3 Source Elements

The primary output from the BFS-PSTF is the families of the *conditional predictive one-step transition density functions* (Krzysztofowicz and Maranzano, 2004, Eq. (45)):

$$\{\psi_{1v} : v = 0, 1\}, \tag{5a}$$

$$\{\theta_{nv}(\cdot|h_{n-1})\text{: all } h_{n-1}; v = 0, 1; n = 2, ..., N\}. \tag{5b}$$

The full interpretation of these functions is this:

$$\psi_{10}(h_1) = p(h_1|h_0, \mathbf{u}_0, V = 0),$$

$$\psi_{11}(h_1) = p(h_1|h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0, V = 1),$$

$$\theta_{n0}(h_n|h_{n-1}) = p(h_n|h_{n-1}, h_0, \mathbf{u}_0, V = 0),$$

$$\theta_{n1}(h_n|h_{n-1}) = p(h_n|h_{n-1}, h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0, V = 1).$$

The operational notation on the left side does not show the conditioning on $(h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0)$ whose value is fixed at the forecast time. Thus the structure of $\theta_{nv}(h_n|h_{n-1})$ is Markov only in appearance.

With respect to the stochastic process $\{H_n : n = 1, ..., N\}$ being forecasted, the source elements (5) specify, for each hypothesized event $V = v$ $(v = 0, 1)$, the conditional predictive one-step transition density function ψ_{1v} for lead time $n = 1$, and a family (for all h_{n-1}) of the conditional predictive one-step transition density functions $\theta_{nv}(\cdot|h_{n-1})$ for every lead time $n \in \{2, ..., N\}$. Thus under the hypothesis that the precipitation event is $V = v$, the conditional predictive *joint* density function of the river stages $H_1, ..., H_n$, for any $n \in \{2, ..., N\}$, takes the form

$$\psi_{1v}(h_1) \prod_{k=2}^{n} \theta_{kv}(h_k|h_{k-1}).$$

For brevity, such a probabilistic forecast is said to have a *conditional Markov structure* (of order one). The adjective "conditional" is crucial and must not be omitted: For as the full generic notation shows, the stochastic dependence structure of the process $\{H_1, ..., H_n\}$ under the Bayesian forecast is not Markov at all.

The source elements (5) are used to construct several kinds of probabilistic river forecasts. The article on the BFS-PSTF presents theory and methods for constructing the PSTF, the Markovian PSTF, and the PRSF. This article presents theory and methods for constructing the PFF.

3.4 Exceedance Function

At any lead time $n \geqslant 2$, the predictive joint density function ξ_n of the river stages $H_1, ..., H_n$ is specified by (Krzysztofowicz and Maranzano, 2004, Eq. (30))

$$\xi_n(h_1, ..., h_n) = (1-\nu)\psi_{10}(h_1) \prod_{k=2}^{n} \theta_{k0}(h_k|h_{k-1}) + \nu\psi_{11}(h_1) \prod_{k=2}^{n} \theta_{k1}(h_k|h_{k-1}) \,. \tag{6}$$

With Ξ_n denoting the corresponding predictive joint distribution function and h being any level,

$$F_n(h) = P(H_1 \leq h, ..., H_n \leq h) = \Xi_n(h, ..., h) \,. \tag{7}$$

This probability is given by the integral of (6) over the n-dimensional box defined by the Cartesian product of n intervals $(-\infty, h]$. The result can be expressed as follows: for $v = 0, 1$, let

$$F_{nv}(h) = \int_{-\infty}^{h} \cdots \int_{-\infty}^{h} \psi_{1v}(h_1) \prod_{k=2}^{n} \theta_{kv}(h_k|h_{k-1}) \ dh_1 \cdots dh_n; \tag{8}$$

then

$$F_n(h) = (1-\nu)F_{n0}(h) + \nu F_{n1}(h). \tag{9}$$

In other words, the distribution function F_n of the maximum river stage Z_n within time interval $(t_0, t_n]$ is a mixture of two functions: F_{n0} — the distribution function of Z_n conditional on the nonoccurrence of precipitation, $V = 0$; and F_{n1} — the distribution function of Z_n conditional on the occurrence of precipitation, $V = 1$. Thereby, the exceedance function of Z_n is specified by

$$\bar{F}_n(h) = 1 - (1-\nu)F_{n0}(h) - \nu F_{n1}(h). \tag{10}$$

In summary, given the probability of precipitation occurrence ν specified by the PQPF and the source elements (5) output from the BFS-PSTF, the PFF can be constructed via (8) and (10). The construction involves the evaluation of two n-dimensional integrals (8) for every level h.

The exceedance function $\bar{F}_n$ so constructed is predictive in a Bayesian sense: it quantifies the total uncertainty about the maximum river stage Z_n within time interval $(t_0, t_n]$, given all information utilized by the forecasting system at time t_0. This includes information utilized by the deterministic hydrologic model, which simulates the response of the river basin to precipitation, and information utilized by the BFS-PSTF, which processes the precipitation uncertainty, quantifies the hydrologic uncertainty, and integrates them. Specifically, the information $(h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0)$, which conditions the source elements (5), conditions also the exceedance function (10). Thus, strictly speaking, $\bar{F}_n$ is the *predictive conditional exceedance function* of Z_n.

3.5 Numerical Algorithm

To calculate the PFF for a given level h, an efficient numerical algorithm can be designed that exploits the structure of the integral (8) and the fact that the BFS-PSTF outputs not only the density functions (5) but also the corresponding distribution functions.

Step 0. Given are the probability of precipitation occurrence ν and arrays, on the discretized sample space of $(H_1, ..., H_N)$, with values of the density functions

$$\psi_{1v}(h_1), \{\theta_{nv}(h_n|h_{n-1}) : n = 2, ..., N-1\}, \qquad v = 0, 1,$$

and values of the distribution functions

$$\Psi_{1v}(h_1), \{\Theta_{nv}(h_n|h_{n-1}) : n = 2, ..., N\}, \qquad v = 0, 1.$$

Step 1. For $v = 0, 1$, calculate

$$F_{1v}(h) = \Psi_{1v}(h),$$

$$F_{2v}(h) = \int_{-\infty}^{h} \Theta_{2v}(h|h_1)\, \psi_{1v}(h_1)\, dh_1,$$

and for all values h_2 in the discretized sample space, calculate

$$I_{2v}(h_2, h) = \int_{-\infty}^{h} \theta_{2v}(h_2|h_1)\, \psi_{1v}(h_1)\, dh_1.$$

Step 2. Calculate sequentially for $n = 3, ..., N-1$, for $v = 0, 1$, and for all values h_n in the discretized sample space

$$I_{nv}(h_n, h) = \int_{-\infty}^{h} \theta_{nv}(h_n|h_{n-1})\, I_{n-1,v}(h_{n-1}, h)\, dh_{n-1}.$$

Step 3. Calculate sequentially for $n = 3, ..., N$ and for $v = 0, 1$

$$F_{nv}(h) = \int_{-\infty}^{h} \Theta_{nv}(h|h_{n-1}) \, I_{n-1,v}(h_{n-1}, h) \, dh_{n-1}.$$

Step 4. Calculate for $n = 1, ..., N$

$$\bar{F}_n(h) = 1 - (1 - \nu) F_{n0}(h) - \nu F_{n1}(h).$$

The algorithm requires only one-dimensional integrations, of which there are $2(2N - 3)$. The output is a sequence of exceedance probabilities $\left\{\bar{F}_n(h) : n = 1, ..., N\right\}$, which constitute the PFF for the given level h.

4. PROBABILISTIC FLOOD FORECAST PRODUCTS

Once the PFF is calculated, three kinds of products can be formatted for displaying and conveying the uncertainty to a decision maker. These products are defined below and illustrated with results of the following case study.

4.1 Case Study

The case study is for the forecast point Eldred, Pennsylvania, located in the headwater of the Allegheny River and closing a drainage area of 550 square miles (1430 km^2). It uses real-time input data from the archives of the U.S. National Weather Service. Forecasts are produced daily. The PQPF is prepared for a 24-h period beginning at 1200 UTC (Universal Time Coordinated), divided into four 6-h subperiods. The PRSF and PSTF are prepared for 72 h ahead in 6-h steps based on the PQPF and other input data available at 1200 UTC.

(Because the time to peak of the unit hydrograph is 30 h, the maximum reasonable lead time of the PRSF and PSTF might seem to be 54 h; however, the hydrologic uncertainty processor of the BFS takes into account the prior (climatic) uncertainty about precipitation occurrence and amount beyond the period covered by PQPF, which is here 24–72 h. Therefore, well-calibrated PRSF and PSTF can be produced up to 72 h ahead. This is, in fact, one of the advantages of the BFS: it combines seamlessly a real-time forecast for a short period, 0–24 h, with a climatic forecast for an extended period, 24–72 h).

The source elements (5) and the probability of precipitation occurrence, $\nu = 0.81$, from the Eldred case study reported by Krzysztofowicz and Maranzano (2004) were used to calculate the PFF on the nested time intervals $\{(t_0, t_n] : n = 1, ..., 12\}$ with a fixed time step $\Delta t = t_n - t_{n-1} = 6$ h and the maximum lead time $t_{12} - t_0 = 72$ h, counting from the forecast time t_0 at 1200 UTC.

4.2 Exceedance Functions

Figure 1 shows the sequence of exceedance functions $\{\bar{F}_n : n = 1, ..., 12\}$. For any level h and any lead time t_n, the decision maker may read $\bar{F}_n(h) = P(Z_n > h)$, the probability of the maximum river stage Z_n within time interval $(t_0, t_n]$ exceeding level h. For example, for a structure located at level $h = 6$ ft, the probability of being flooded within 48 h is $\bar{F}_8(6) = 0.99$, whereas for a structure located at level $h = 10$ ft, the probability of being flooded within 48 h is $\bar{F}_8(10) = 0.59$.

The exceedance functions evolve monotonically with time: $\bar{F}_{n-1} \leq \bar{F}_n$ for $n = 2, ..., 12$. That is, for any level h, the probability $\bar{F}_n(h)$ of that level being exceeded within time interval $(t_0, t_n]$ is a nondecreasing function of time t_n. Loosely speaking, the probability of flood occurrence after a fixed time instant t_0 compounds over time.

4.3 Isoprobability Time Series

The monotone evolution of the exceedance probability is vivid in Fig. 2. It shows the isoprobability time series $\{z_{np} : n = 1, ..., 12\}$ of quantiles of the maximum river stages having the exceedance probability $p = \bar{F}_n(z_{np})$; there are seven time series corresponding to $p = 0.005$, 0.05, 0.25, 0.50, 0.75, 0.95, 0.995. Clearly, $z_{n-1,p} \leq z_{np}$ for $n = 2, ..., 12$ and for each p.

The isoprobability time series provide information for the following type of decision. Suppose a temporary dike should protect a town with reliability not smaller than 0.95; equivalently, the probability of overtopping should be not greater than 0.05. What should the minimum height of the dike be? The isoprobability time series $\{z_{n,0.05} : n = 1, ..., 12\}$, marked with the upper squares in Fig. 2, provides the answer: the dike should reach height of at least $z_{4,0.05} = 9.7$ ft within 24 h, at least $z_{8,0.05} = 15.1$ ft within 48 h, and at least $z_{12,0.05} = 16.6$ ft within 72 h. Such a progress of work will ensure a uniform risk throughout the duration of flood.

4.4 Distribution of Time to Flooding

Information derivable from the PFF and useful for flood warning decisions is the distribution of the time to flooding. Let

$T(h)$ — time instant at which river stage process $\{H_1, ..., H_N\}$ exceeds level h for the first time; it is a discrete variate taking values in the set $\{t_1, ..., t_N\}$. When $t_0 = 0$, variate $T(h)$ is a discrete measure (an approximation) of the time to flooding level h, measured from the forecast time t_0.

The distribution function of the time to flooding $T(h)$ is defined by

$$P(T(h) \leq t_n) = \bar{F}_n(h); \tag{11}$$

that is, $\bar{F}_n(h)$ is the probability of the time to flooding $T(h)$ being t_n or shorter (Karlin and Taylor, 1975).

Using the exceedance probabilities $\bar{F}_n(h)$ shown in Fig. 1, the distribution function of $T(h)$ is plotted in Fig. 3 for $h = 10, 14, 18$ ft. To illustrate its usage, suppose the dike protecting a town has height $h = 10$ ft. Then the probability of the dike being overtopped in 24 h or less is $\bar{F}_4(10) = 0.02$, in 48 h or less it is $\bar{F}_8(10) = 0.60$, and in 72 h or less it is $\bar{F}_{12}(10) = 0.72$. In general, the plot informs the decision maker about the temporal evolution of the risk of flooding for a zone of the floodplain extending upward from level h. It is the most relevant display for flood warning decisions: it conveys the trade-off between the flood risk and the lead time for a given zone of the floodplain.

4.5 Forecast Updating

An operational advantage of the analytic-numerical BFS is the ease of updating the PFF whenever the probability of precipitation occurrence ν is updated between the scheduled forecast times. This updating does not require re-running the hydrologic model and re-computing the source elements of the PSTF, but can be performed directly via Eq. (10). At the forecast time, when executing Step 4 of the algorithm from Section 3.5, one should store values $F_{n0}(h)$ and $F_{n1}(h)$. Then, given the updated value of ν, one needs only to repeat Step 4 of the algorithm to obtain the updated PFF.

Figure 4 shows the updated exceedance functions $\bar{F}_n$ for $n = 8$ based on six different updates of the probability of precipitation occurrence ν. It is apparent that ν alone exerts a phenomenal influence on the shape of $\bar{F}_n$. For instance, within 48 h since the forecast time, the probability of flooding level h is one for $h = 6$ ft regardless of ν, is slightly lower than ν for $h = 8$ ft, and decreases non-linearly with ν as h increases.

The rapid updating may be especially useful in forecasting hurricane-induced or convection-induced floods, when the degree of uncertainty about the storm track changes rapidly. For example, when a severe thunderstorm (a supercell) is developing, there may be little uncertainty regarding a copious amount of rainfall it will produce (somewhere, sometime), but there may be huge uncertainty about its track. This translates into huge uncertainty regarding the occurrence of rainfall over any of the small river basins lying within the reach of the storm. In effect, ν for a particular basin should be updated rapidly as the projection of the storm track is updated based on real-time observations. Equation (10) makes it feasible to update the PFF simultaneously with ν.

5. THEORY OF BOUNDS ON EXCEEDANCE FUNCTION

There exist certain theoretical order relations between the marginal distribution functions of the river stage process $\{H_1, ..., H_N\}$ and the PFF. They are instructive and can be exploited to construct (i) bounds on probability $\bar{F}_n(h)$ and (ii) approximations to probability $\bar{F}_n(h)$. These approximations are useful when PRSF is available but PSTF is not.

5.1 Probabilistic River Stage Forecast

Let $\bar{\Psi}_n$ denote the exceedance function of river stage H_n, such that for any level h

$$\bar{\Psi}_n(h) = P(H_n > h); \tag{12}$$

that is, $\bar{\Psi}_n(h)$ is the probability of river stage H_n exceeding level h at time t_n. The exceedance function $\bar{\Psi}_n$ produced by the BFS-PSTF is predictive in a Bayesian sense: it quantifies the total uncertainty about river stage H_n, given all information utilized by the forecasting system at time t_0. Whereas the information conditioning $\bar{\Psi}_n$ is not shown, for simplicity, it includes observed river stage $H_0 = h_0$ at time t_0. Thus $\bar{\Psi}_n(h)$ can also be interpreted as the predictive n-step transition probability $P(H_n > h | H_0 = h_0)$.

The PRSF is defined henceforth as a sequence of exceedance functions

$$\{\bar{\Psi}_n : n = 1, ..., N\}. \tag{13}$$

These functions can be derived from the source elements (5) output by the BFS-PSTF, as shown by Krzysztofowicz and Maranzano (2004); their approximations can be produced directly by the BFS-PRSF, as described by Krzysztofowicz (2002a).

5.2 Bounds on Probabilistic Flood Forecast

With respect to the river stage process $\{H_1, ..., H_N\}$, the PRSF does not specify the stochastic dependence, but only the marginal distributions. Hence the PRSF is insufficient for deriving the PFF; however, it is sufficient for deriving bounds in the PFF. After Krzysztofowicz (2002b), for $n = 1$,

$$\bar{F}_1(h) = \bar{\Psi}_1(h), \tag{14}$$

and for $n \geqslant 2$, there exist the lower bound, the middle bound, and the upper bound, respectively,

$$L_n(h) = \max\{\bar{\Psi}_1(h), ..., \bar{\Psi}_n(h)\}, \tag{15a}$$

$$M_n(h) = 1 - \prod_{k=1}^{n} \left[1 - \bar{\Psi}_k(h)\right], \tag{15b}$$

$$U_n(h) = \min\{\bar{\Psi}_1(h) + \cdots + \bar{\Psi}_n(h), 1\}, \tag{15c}$$

such that the Fréchet bounds (Fréchet, 1935), which hold always, are

$$L_n(h) \leq \bar{F}_n(h) \leq U_n(h). \tag{16}$$

Moreover, when the stochastic dependence in the process $\{H_1, ..., H_N\}$ can be characterized in a particular way, tighter bounds may be inferred thusly: (i) if the process is independent, then

$$\bar{F}_n(h) = M_n(h); \tag{17a}$$

(ii) if the process is Markov of order one with positive quadrant dependence, then

$$L_n(h) \leq \bar{F}_n(h) < M_n(h); \tag{17b}$$

(iii) if the process is Markov of order one with negative quadrant dependence, then

$$M_n(h) < \bar{F}_n(h) \leq U_n(h). \tag{17c}$$

[For definitions and derivations, see Krzysztofowicz (2002b, Section 3).]

Figure 5 shows the three bounds for $n = 1, ..., 12$ and all h. The probability intervals specified by the bounds may be viewed as imprecise measures of uncertainty about flood occurrence within time interval $(t_0, t_n]$. The interval may even be tight enough to be useful for flood warning decisions. Typically, the optimal warning rule is of the threshold type (Krzysztofowicz, 1993). Thus if the middle bound $M_n(h)$ exceeds the optimal threshold for lead time $t_n - t_0$, then taking action is possibly optimal; and if the lower bound $L_n(h)$ exceeds the optimal threshold for lead time $t_n - t_0$, then taking action is surely optimal, in which case the knowledge of the exceedance probability $\bar{F}_n(h)$ is redundant.

6. APPROXIMATION TO PROBABILISTIC FLOOD FORECAST

6.1 Recursive Linear Interpolator

By analyzing theoretically the behavior of $\bar{F}_n$ relative to its bounds and by examining empirically the dependence characteristics of the river stage process, Krzysztofowicz (2002b, Section 4) conjectured that bounds (17b) are the most plausible. Based on this premise, he formulated a *recursive linear interpolator* (RLI) that maps the PRSF $\{\bar{\Psi}_n : n = 1, ..., N\}$ into an estimate of the PFF $\{\bar{F}_n^* : n = 1, ..., N\}$. For $n = 1$, the estimate is

$$\bar{F}_1^*(h) = \bar{\Psi}_1(h). \tag{18}$$

Then proceeding recursively for $n = 2, ..., N$, the one-step-ahead estimates of the bounds are

$$L_n^*(h) = \max\{\bar{F}_{n-1}^*(h), \bar{\Psi}_n(h)\}, \tag{19a}$$

$$M_n^*(h) = \bar{F}_{n-1}^*(h) + \bar{\Psi}_n(h) - \bar{F}_{n-1}^*(h)\bar{\Psi}_n(h), \tag{19b}$$

and the estimate of the exceedance probability is

$$\bar{F}_n^*(h) = wL_n^*(h) + (1 - w)M_n^*(h), \tag{20}$$

where w is a weight bounded by $0 < w < 1$. A plausible value of the weight w for a fixed time step $\Delta t = t_n - t_{n-1}$ is the average (over n) value of the Spearman's rank correlation coefficient between river stages H_n and H_{n-1}, estimated conditional on the occurrence of precipitation in the period covered by the PQPF. [For the properties of w and the rationale behind this estimate, see Krzysztofowicz (2002b, Section 5).]

6.2 Empirical Substantiation of the RLI

At the time of its development, the RLI could not be substantiated empirically because the exact PFF was unavailable. In this article, having derived and calculated the exact PFF, we have the first opportunity to assess the performance of the RLI empirically.

Figure 6 shows the bounds L_n and M_n, the exact exceedance functions $\bar{F}_n$, and the estimates $\bar{F}_n^*$ obtained via the RLI with the weight $w = 0.8$. Three observations are noteworthy. First, $L_n < \bar{F}_n < M_n$ for $n = 1, ..., 12$; this substantiates the conjecture that bounds (17b), which are tighter than the Fréchet bounds (16), hold for the PFF. Second, the exact exceedance function $\bar{F}_n$ behaves more like the lower bound L_n than the middle bound M_n, precisely as conjectured (Krzysztofowicz, 2002b, Section 4). Third, the estimate $\bar{F}_n^*$ is astonishingly close to the exact exceedance function $\bar{F}_n$, especially for $n \leq 8$; but even for $n = 9, 10, 11, 12$, the maximum absolute differences $\left|\bar{F}_n(h) - \bar{F}_n^*(h)\right|$ occur all at the low levels h (where the exact exceedance function $\bar{F}_n$ is nearly horizontal as it moves away from the lower bound L_n towards the middle bound M_n), and their values 0.106, 0.089, 0.081, 0.073 are not unlike the calibration scores of some real-time probabilistic forecasts (Krzysztofowicz and Sigrest, 1999).

In summary, the RLI estimator of the PFF from the PRSF appears to be reasonably accurate. It is also fast and cheap, as the execution of Eqs. (18)–(20) is straightforward. Therefore, it offers an attractive means of obtaining a PFF in simpler forecasting systems which produce only the PRSF, but not the PSTF.

7. CLOSURE

The purpose of a PFF is to support flood warning and response decisions. The PFF for N time steps ahead consists of N predictive exceedance functions of the maximum river stages within N nested time intervals, all beginning at the forecast time. It can be derived from the source elements outputted by the BFS-PSTF, a Bayesian forecasting system attached to a deterministic hydrologic model of the river basin above the forecast point of interest. The PFF so derived is exact and may be updated rapidly (essentially at no additional cost) whenever the probability of precipitation occurrence over the basin is updated — the capability especially useful in forecasting hurricane-induced or convection-induced floods in small basins, when the degree of uncertainty about the storm track changes rapidly. The PFF may be displayed to a decision maker in three formats, as (i) the sequence of exceedance functions of the maximum river stages, (ii) the isoprobability time series of quantiles of the maximum river stages, and (iii) the distribution functions of the time to flooding for specified levels.

Bounds on the PFF can be constructed from the PRSF — a product much simpler than the PSTF, as it specifies only the exceedance functions of river stages at discrete times, but not the transitions between them. The Fréchet bounds hold without any assumptions; the tighter bounds hold under the assumption of positive quadrant dependence in the river stage process. Whereas the validity of this assumption was conjectured in an earlier article, this article reports the first empirical evidence supporting the conjecture. It also reports the first empirical substantiation of the RLI — a method that constructs an approximation to the PFF from the PRSF alone. Being reasonably accurate and very simple, the RLI seems potentially attractive for real-time flood forecasting in systems of lesser complexity.

ACKNOWLEDGMENTS

Coire Maranzano performed the numerical calculations and made the drawings. This material is based upon work supported by the National Science Foundation under Grant No. ATM-0641572, "New Statistical Techniques for Probabilistic Weather Forecasting".

REFERENCES

Fréchet, M., 1935. Généralisations du théorème des probabilités totales. *Fund. Math.* **25**, 379–387.

Georgakakos, K.P., 1986. A generalized stochastic hydrometeorological model for flood and flash-flood forecasting, Part 1: Formulation. *Water Resources Research* **22**(13), 2083–2095.

Hoffman, K., 1975. *Analysis in Euclidean Space*, Prentice-Hall, Englewood Cliffs, New Jersey.

Karlin, S., Taylor, H.M., 1975. *A First Course in Stochastic Processes*, Academic Press, New York.

Kelly, K.S., Krzysztofowicz, R., 1994. Probability distributions for flood warning systems. *Water Resources Research* **30**(4), 1145–1152.

Krzysztofowicz, R., 1993. A theory of flood warning systems. *Water Resources Research* **29**(12), 3981–3994.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research* **35**(9), 2739–2750.

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. *Journal of Hydrology* **249**(1–4), 2–9.

Krzysztofowicz, R., 2002a. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology* **268**(1–4), 16–40.

Krzysztofowicz, R., 2002b. Probabilistic flood forecast: bounds and approximations. *Journal of Hydrology* **268**(1–4), 41–55.

Krzysztofowicz, R., Davis, D.R., 1983. A methodology for evaluation of flood forecast-response systems, Part 2: Theory. *Water Resources Research* **19**(6), 1431–1440.

Krzysztofowicz, R., Davis, D.R., 1984. Toward improving flood forecast-response systems. *Interfaces* **14**(3), 1–14.

Krzysztofowicz, R., Maranzano, C.J., 2004. Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology* **299**(1–2), 15–44.

Krzysztofowicz, R., Sigrest, A.A., 1999. Calibration of probabilistic quantitative precipitation forecasts. *Weather and Forecasting* **14**(3), 427–442.

Lardet, P., Obled, Ch., 1994. Real-time flood forecasting using a stochastic rainfall generator. *Journal of Hydrology* **162**, 391–408.

FIGURE CAPTIONS

Figure 1. The exact PFF calculated from the PSTF: sequence of exceedance functions $\{\bar{F}_n : n = 1, ..., 12\}$ of the maximum river stages within time intervals $\{(t_0, t_n] : n = 1, ..., 12\}$; time step $\Delta t = t_{n-1} - t_n = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

Figure 2. Isoprobability time series $\{z_{np} : n = 1, ..., 12\}$ of quantiles of the maximum river stages having the exceedance probability $p = \bar{F}_n(z_{np})$, for seven values of p; $\bar{F}_n$ is the exact exceedance function calculated from the PSTF; Eldred, Pennsylvania.

Figure 3. Distribution function $\{P(T(h) \le t_n) = \bar{F}_n(h) : n = 1, ..., 12\}$ of the time to flooding $T(h)$, counting from 1200 UTC on the forecast day, for three levels $h = 10, 14, 18$ ft; $\bar{F}_n$ is the exact exceedance function calculated from the PSTF; Eldred, Pennsylvania.

Figure 4. Sensitivity of the exceedance function $\bar{F}_n$ of the maximum river stage Z_n at lead time $n = 8$ $(48$ h$)$ to the probability ν of precipitation occurrence over the basin in the 24-h period counting from the forecast time; Eldred, Pennsylvania.

Figure 5. Bounds on the exceedance function $\bar{F}_n$ of the maximum river stage Z_n within time interval $(t_0, t_n]$ calculated from the PRSF produced by the BFS-PSTF: lower bound L_n, middle bound M_n, and upper bound U_n; time step $\Delta t = t_{n-1} - t_n = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

Figure 6. Comparison of the exact PFF (exceedance functions $\bar{F}_n$ calculated from the source elements output by the BFS-PSTF) with the approximate PFF (exceedance functions $\bar{F}_n^*$ obtained from the PRSF via the RLI estimator); the PRSF was calculated from the source elements output by the BFS-PSTF; the bounds L_n and M_n are those shown in Fig. 5. The closeness of $\bar{F}_n^*$ to $\bar{F}_n$ and their staying within the bounds L_n and M_n provides a check on the coherence of the BFS (both theory and computations).

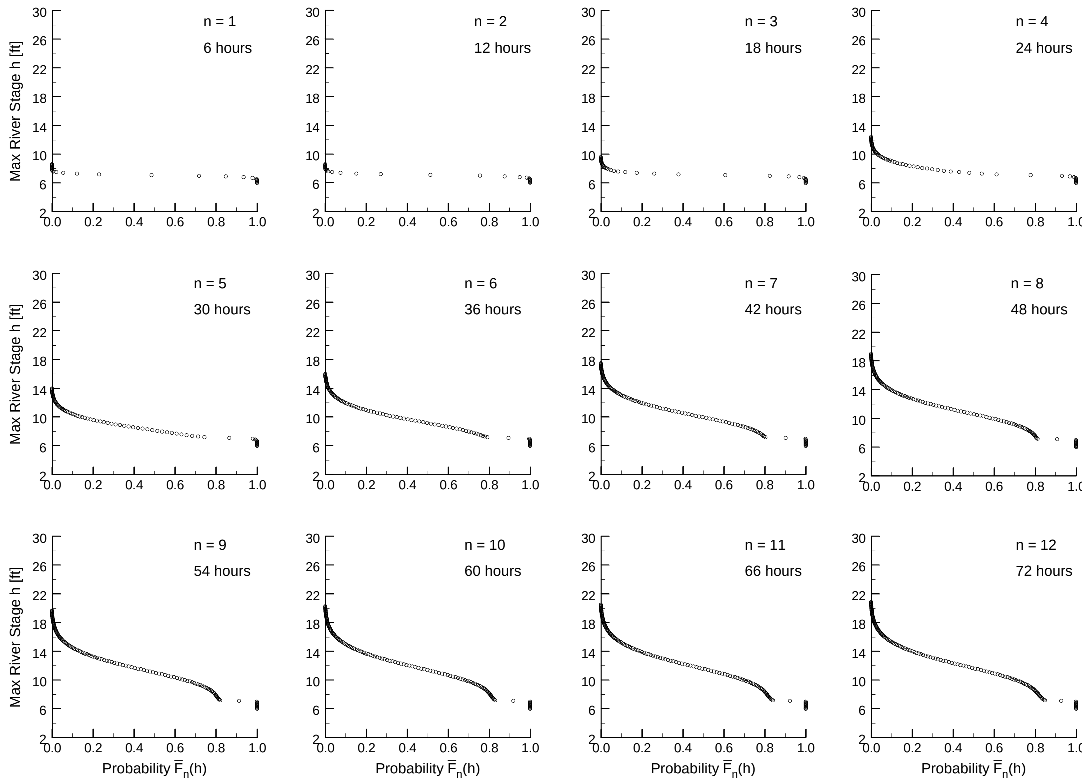

Figure 1. The exact PFF calculated from the PSTF: sequence of exceedance functions $\{\bar{F}_n : n = 1, ..., 12\}$ of the maximum river stages within time intervals $\{(t_0, t_n] : n = 1, ..., 12\}$; time step $\Delta t = t_{n-1} - t_n = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

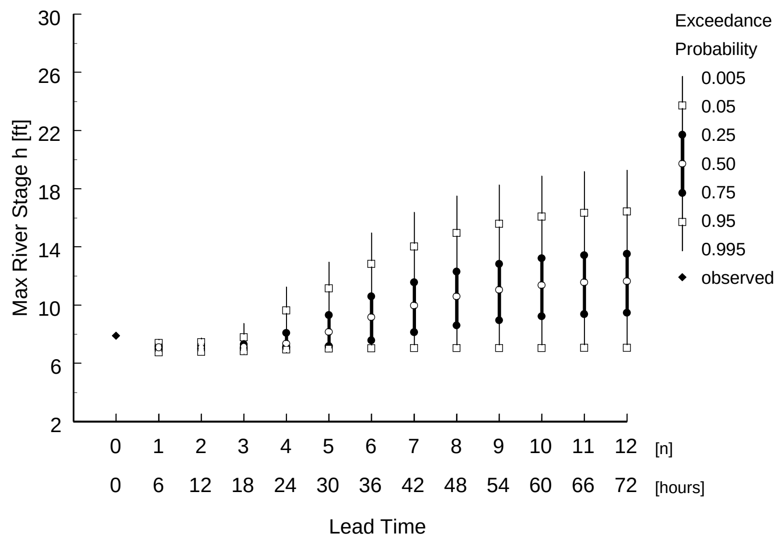

Figure 2. Isoprobability time series $\{z_{np} : n = 1, ..., 12\}$ of quantiles of the maximum river stages having the exceedance probability $p = \bar{F}_n(z_{np})$, for seven values of p; $\bar{F}_n$ is the exact exceedance function calculated from the PSTF; Eldred, Pennsylvania.

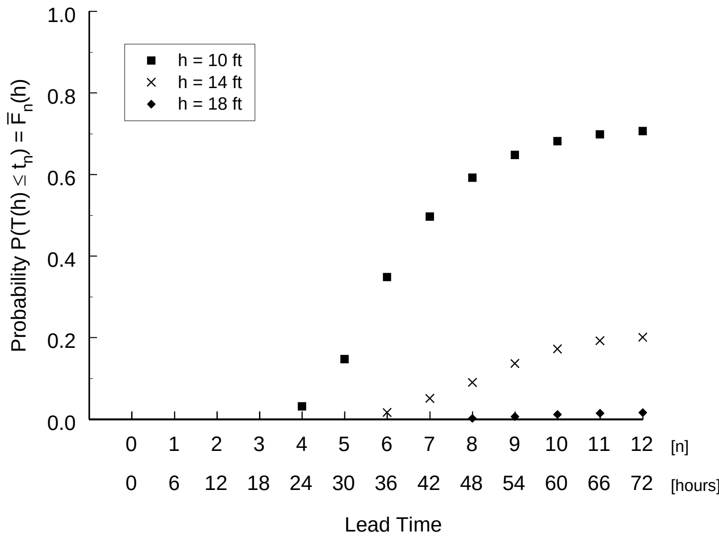

Figure 3. Distribution function $\{P(T(h) \leq t_n) = \bar{F}_n(h) : n = 1, ..., 12\}$ of the time to flooding $T(h)$, counting from 1200 UTC on the forecast day, for three levels $h = 10, 14, 18$ ft; $\bar{F}_n$ is the exact exceedance function calculated from the PSTF; Eldred, Pennsylvania.

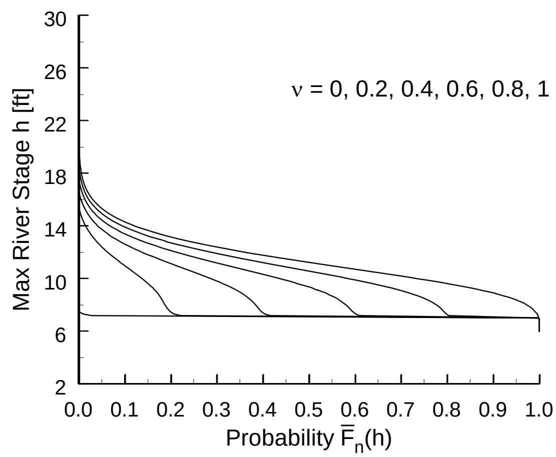

Figure 4. Sensitivity of the exceedance function $\bar{F}_n$ of the maximum river stage Z_n at lead time $n = 8$ (48 h) to the probability ν of precipitation occurrence over the basin in the 24-h period counting from the forecast time; Eldred, Pennsylvania.

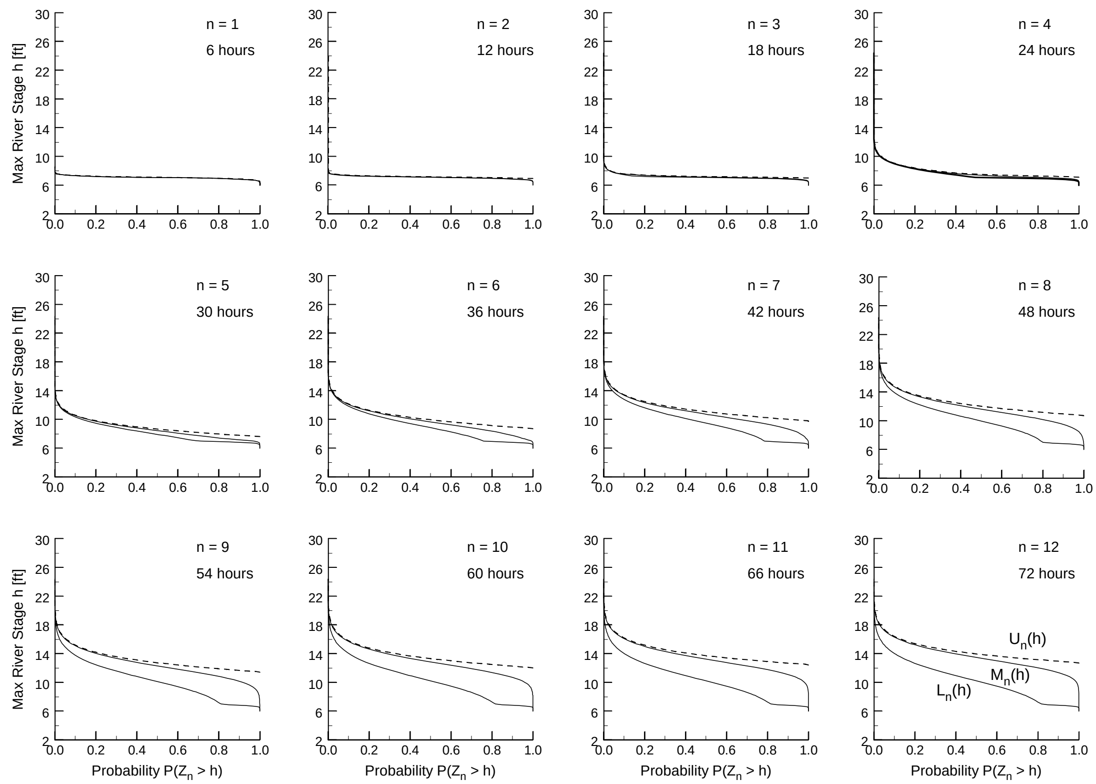

Figure 5. Bounds on the exceedance function $\bar{F}_n$ of the maximum river stage Z_n within time interval $(t_0, t_n]$ calculated from the PRSF produced by the BFS-PSTF: lower bound L_n, middle bound M_n, and upper bound U_n; time step $\Delta t = t_{n-1} - t_n = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

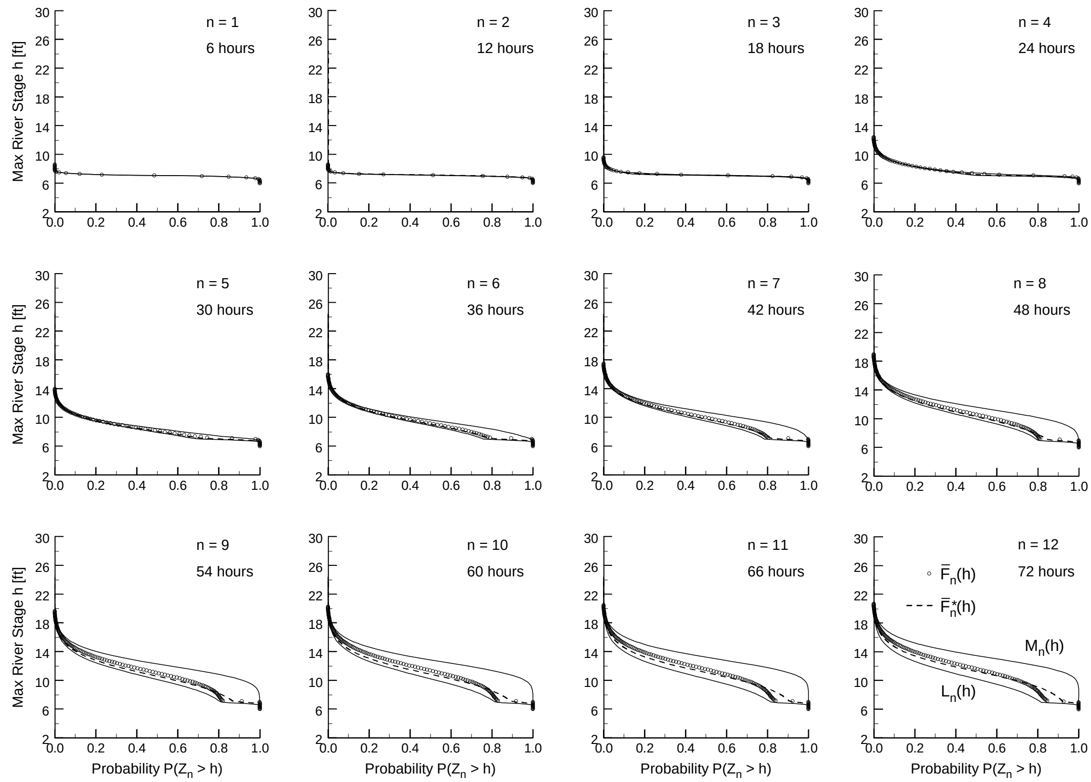

Figure 6. Comparison of the exact PFF (exceedance functions $\bar{F}_n$ calculated from the source elements output by the BFS-PSTF) with the approximate PFF (exceedance functions $\bar{F}_n^*$ obtained from the PRSF via the RLI estimator); the PRSF was calculated from the source elements output by the BFS-PSTF; the bounds L_n and M_n are those shown in Fig. 5. The closeness of $\bar{F}_n^*$ to $\bar{F}_n$ and their staying within the bounds L_n and M_n provides a check on the coherence of the BFS (both theory and computations).