ELSEVIER

Journal of Hydrology 268 (2002) 41–55

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Probabilistic flood forecast: bounds and approximations

Roman Krzysztofowicz*

Department of Systems Engineering and Department of Statistics, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747, USA

Received 24 July 2001; revised 25 April 2002; accepted 3 May 2002

Abstract

The probabilistic river stage forecast (PRSF) specifies a sequence of exceedance functions $\{\bar{\Psi}_n : n = 1, \ldots, N\}$ such that $\bar{\Psi}_n(h_n) = P(H_n > h_n)$, where H_n is the river stage at time instance t_n, and P stands for probability. The probabilistic flood forecast (PFF) should specify a sequence of exceedance functions $\{\bar{F}_n : n = 1, \ldots, N\}$ such that $\bar{F}_n(h) = P(Z_n > h)$, where Z_n is the maximum river stage within time interval $(t_0, t_n]$, practically $Z_n = \max\{H_1, \ldots, H_n\}$. In the absence of information about the stochastic dependence structure of the process $\{H_1, \ldots, H_N\}$, the PFF cannot be derived from the PRSF. This article presents simple methods for calculating bounds on $\bar{F}_n$ and approximations to $\bar{F}_n$ using solely the marginal exceedance functions $\bar{\Psi}_1, \ldots, \bar{\Psi}_n$. The methods are illustrated with tutorial examples and a case study for a 1430 km^2 headwater basin wherein the PRSF is for a 72-h interval discretized into 6-h steps.

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

1. Introduction

1.1. Forecast purpose

A probabilistic river stage forecast (PRSF) typically specifies exceedance functions of river stages at time instances that form the discrete time scale of the hydrograph (Lardet and Obled, 1994; Krzysztofowicz, 2002). A probabilistic flood forecast (PFF) should specify exceedance functions of maximum river stages within time intervals that form a nested set. The purpose of the PFF is to support flood warning and response decisions (Krzysztofowicz and Davis, 1983; Krzysztofowicz, 1993).

In order to rigorously construct a PFF, one would need a PRSF and a real-time characterization of the stochastic dependence structure of the river stage process within the time horizon of the PRSF. Of course, it is possible to produce such a characterization, but it requires a more complex forecasting system than that producing the PRSF alone.

This article shows how to construct bounds on and approximations to the PFF from a given PRSF alone. The construction methods are simple and thus easily implementable in real-time forecasting. They can be attached to any forecasting system that produces a PRSF in the specified format. Herein, these methods are illustrated in conjunction with the Bayesian Forecasting System (BFS) that produces a short-term PRSF for a headwater basin (Krzysztofowicz, 2002).

1.2. Case study

The case study reported throughout the article is

* Tel.: +1-434-924-5393; fax: +1-434-982-2972.
E-mail address: rk@virginia.edu (R. Krzysztofowicz).

0022-1694/02/$ - see front matter
PII: S0022-1694(02)00107-5

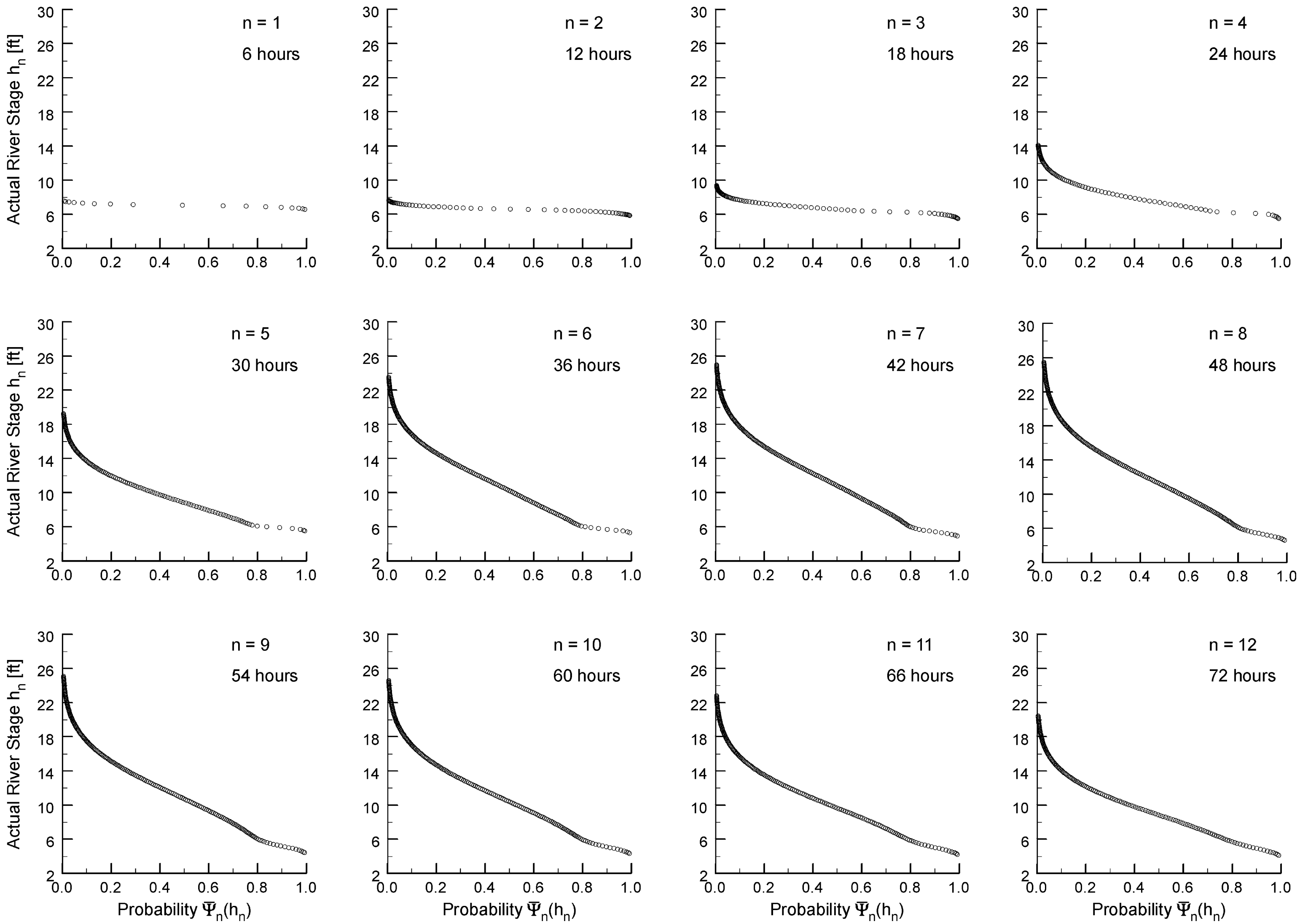

Fig. 1. A PRSF produced by the BFS: sequence of exceedance functions $\{\bar{\Psi}_n : n = 1, \ldots, 12\}$ of river stages at time instances t_n ($n = 1, \ldots, 12$); time step $\Delta t = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

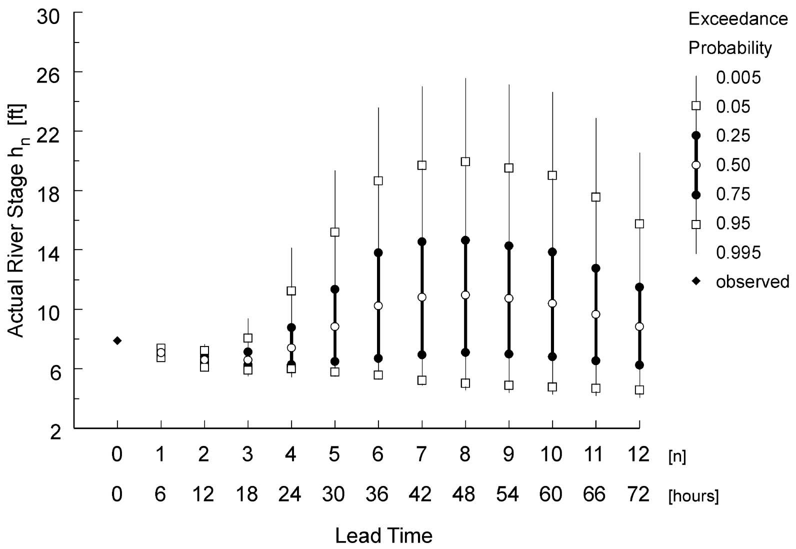

Fig. 2. Isoprobability time series $\{h_{np} : n = 1, \ldots, 12\}$ of quantiles of the river stages having the exceedance probability $p = \bar{\Psi}_n(h_{np})$, for seven values of p.

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). In the pilot testing of the BFS, forecasts are produced daily. The probabilistic quantitative precipitation forecast (PQPF) is prepared for a 24-h period beginning at 1200 UTC (Universal Time Coordinated), divided into four 6-h subperiods. The PRSF is prepared based on the PQPF and other input data available at 1200 UTC. The case study uses real-time input data from the archives of the National Weather Service (NWS). The PRSF is produced for 72 h ahead in 6-h steps.

1.3. Overview

Section 2 formally defines the PRSF and the PFF. Section 3 derives two sets of bounds on the PFF: the Fréchet bounds which hold without any assumptions, and tighter bounds which hold under the assumption of positive quadrant dependence; theoretical arguments and empirical evidence in support of this assumption for the river stage process are discussed. Section 4 presents two estimators of the PFF based on the bounds: a direct linear interpolator (DLI) and a recursive linear interpolator (RLI); each interpolator uses the PRSF and a single parameter. Section 5 comprises tutorial examples to help in the understanding of the bounds and the estimators. Section 6 reports a case study.

2. Probabilistic forecasts

2.1. Probabilistic river stage forecast

Let t_0 denote the forecast time, and let t_n $(n = 1, \ldots, N)$ denote the time at which the river stage is forecasted and then 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 that may take any value above the gauge datum.
$\bar{\Psi}_n$—exceedance function of river stage H_n, such that for any level h

$$\bar{\Psi}_n(h) = P(H_n > h); \tag{1}$$

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 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. For simplicity, the information conditioning the exceedance function is not shown. This information 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

$$\{\bar{\Psi}_n : n = 1, \ldots, N\}. \tag{2}$$

Fig. 1 displays a sequence $\{\bar{\Psi}_n : n = 1, \ldots, 12\}$ with a constant time step $\Delta t = t_n - t_{n-1} = 6$ h. The orientation of the axes, with the river stage on the vertical axis and the exceedance probability on the horizontal axis, is the same as in the displays that communicate the PRSF to decision makers. A partial information extracted from the exceedance functions is displayed in a different format in Fig. 2. It shows the isoprobability time series $\{h_{np} : n = 1, \ldots, 12\}$ of quantiles of the river stages having the exceedance probability $p = \bar{\Psi}_n(h_{np})$; there are seven time series corresponding to $p = 0.005$, 0.05, 0.25, 0.50, 0.75, 0.95, 0.995. Together, Figs. 1 and 2 convey to the decision maker the evolution of uncertainty in time.

2.2. Probabilistic flood forecast

Whereas the exceedance functions of instantaneous river stages $H_1, \ldots, H_N$ provide information useful for some decision problems, they do not provide information needed by a flood warning system (Krzysztofowicz, 1993). In order to optimally decide whether or not to issue a flood warning for a zone of the floodplain, the decision maker needs a distribution of the maximum river stage within a time interval. The notion of the time interval is essential for two reasons. First, it is needed to quantify the total risk of flooding from a rainfall event (or a portion thereof that is covered by a probabilistic quantitative precipitation forecast upon which the PRSF is based). Second, it is needed to capture the uncertainty about the timing of the flood crest.

The purpose of the PFF is to provide information needed by a flood warning system. Toward this end define

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 $\{H_1, \ldots, H_n\}$ is defined as

$$Z_n = \max\{H_1, \ldots, H_{n-1}, H_n\}. \tag{3}$$

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

$$\begin{aligned}\bar{F}_n(h) &= P(Z_n > h) \\ &= 1 - P(Z_n \le h) \\ &= 1 - P(H_1 \le h, \ldots, H_{n-1} \le h, H_n \le h);\end{aligned} \tag{4}$$

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 $H_1, \ldots, H_n$ exceeding level h.

The PFF is defined henceforth as a sequence of exceedance functions

$$\{\bar{F}_n : n = 1, \ldots, N\}. \tag{5}$$

The key difference between the PRSF and the PFF boils down to the time scale: $\bar{\Psi}_n(h)$ is the probability of the variate H_n exceeding level h at time instance t_n; $\bar{F}_n(h)$ is the probability of the process $\{H_1, \ldots, H_n\}$ exceeding level h within time interval $(t_0, t_n]$. Given the PFF, the probability distributions needed for a flood warning system can readily be obtained (Kelly and Krzysztofowicz, 1994).

3. Bounds on flood forecast

It is apparent from Eq. (4) that the PFF cannot be derived from the PRSF because Eqs. (1) and (2) do not specify the joint distribution of $(H_1, \ldots, H_n)$. It turns out, however, that certain theoretical relations between marginal probabilities exist and that they can be exploited to construct (i) bounds on probability $\bar{F}_n(h)$ and (ii) approximations to probability $\bar{F}_n(h)$. Both constructs are obtained solely in terms of probabilities $\bar{\Psi}_1(h), \ldots, \bar{\Psi}_n(h)$. The objective of this article is to present these potentially useful results.

3.1. Solution framework

For $n = 1$, the solution is $\bar{F}_1(h) = \bar{\Psi}_1(h)$. For any

$n \geq 2$, the solution framework must be based on Eq. (4), which implies

$$P(Z_n \leq h) = P(H_1 \leq h, \ldots, H_{n-1} \leq h, H_n \leq h)$$
$$= P(Z_{n-1} \leq h, H_n \leq h). \quad (6)$$

In other words, the event on the left side is equivalent to two subevents on the right side; symbolically,

$$\{Z_n \leq h\} \Leftrightarrow \{Z_{n-1} \leq h\} \text{ and } \{H_n \leq h\}. \quad (7)$$

We are interested in the complements whose marginal probabilities are $P(Z_n > h) = \bar{F}_n(h)$, $P(Z_{n-1} > h) = \bar{F}_{n-1}(h)$, and $P(H_n > h) = \bar{\Psi}_n(h)$.

3.2. Fréchet bounds

Probability theory imposes a coherence condition that the marginal probabilities of the equivalent events (7) must satisfy (Krzysztofowicz, 1999). For $n = 1$,

$$\bar{F}_1(h) = \bar{\Psi}_1(h), \quad (8)$$

and for $n \geq 2$

$$L_n(h) \leq \bar{F}_n(h) \leq U_n(h), \quad (9)$$

where the lower bound and the upper bound are, respectively,

$$L_n(h) = \max\{\bar{F}_{n-1}(h), \bar{\Psi}_n(h)\}, \quad (10a)$$

$$U_n(h) = \min\{\bar{F}_{n-1}(h) + \bar{\Psi}_n(h), 1\}. \quad (10b)$$

Beginning with $L_2(h)$, $U_2(h)$ and initial condition (8), the recursive substitutions $\bar{F}_{n-1}(h) = L_{n-1}(h)$ and $\bar{F}_{n-1}(h) = U_{n-1}(h)$ for $n = 3, 4, \ldots$ yield explicit expressions

$$L_n(h) = \max\{\bar{\Psi}_1(h), \ldots, \bar{\Psi}_n(h)\}, \quad (11a)$$

$$U_n(h) = \min\{\bar{\Psi}_1(h) + \cdots + \bar{\Psi}_n(h), 1\}. \quad (11b)$$

Bounds of this form are commonly known as the Fréchet bounds (Fréchet, 1935). These are the most general bounds as their derivation requires no assumptions about the process $\{H_1, \ldots, H_n\}$ or its joint distribution.

3.3. Stochastic dependence

To obtain tighter bounds it is necessary to characterize the stochastic dependence between the subevents. The weakest, and therefore most general, characterization of stochastic dependence between the subevents rests on the following definition adapted from Lehmann (1966).

Definition. *The subevents are said to be*

(i) *stochastically independent if and only if*

$$P(H_n > h | Z_{n-1} > h) = P(H_n > h); \quad (12a)$$

(ii) *positively quadrant dependent if and only if*

$$P(H_n > h | Z_{n-1} > h) > P(H_n > h); \quad (12b)$$

(iii) *negatively quadrant dependent if and only if*

$$P(H_n > h | Z_{n-1} > h) < P(H_n > h). \quad (12c)$$

In positive quadrant dependence, observing an exceedance of level h within time interval $(t_0, t_{n-1}]$ increases the probability of exceedance at time t_n. In negative quadrant dependence, observing an exceedance of level h within time interval $(t_0, t_{n-1}]$ decreases the probability of exceedance at time t_n. An equivalent definition is obtained by replacing subevents $\{Z_{n-1} > h\}$ and $\{H_n > h\}$ in Eqs. (12a)–(12c) with subevents $\{Z_{n-1} \leq h\}$ and $\{H_n \leq h\}$, respectively (Lehmann, 1966, Lemma 1).

To exploit this definition, define for $n \geq 2$

$$M_n(h) = \bar{F}_{n-1}(h) + \bar{\Psi}_n(h) - \bar{F}_{n-1}(h)\bar{\Psi}_n(h). \quad (13)$$

Then beginning with $M_2(h)$ and initial condition (8), the recursive substitution $\bar{F}_{n-1}(h) = M_{n-1}(h)$ for $n = 3, 4, \ldots$ yields an explicit expression

$$M_n(h) = 1 - \prod_{k=1}^{n} [1 - \bar{\Psi}_k(h)]. \quad (14)$$

It is now possible to state the following (Krzysztofowicz, 1999).

Theorem 1. *The subevents are*

(i) *stochastically independent if and only if*

$$\bar{F}_n(h) = M_n(h); \quad (15a)$$

(ii) *positively quadrant dependent if and only if*

Table 1
Bounds on the flood occurrence probability $\bar{F}_n(h)$ obtained from the PRSF, and the DLI estimate $\bar{F}_n^*(h)$

No.	n	PRSF	Bounds[a] on $\bar{F}_n(h)$			DLI estimate[b]
		$\bar{\Psi}_n(h)$	$L_n(h)$	$M_n(h)$	$U_n(h)$	$\bar{F}_n^*(h)$
I	1	0.1				0.10
	2	0.1	0.1	0.19	0.2	0.12
	3	0.1	0.1	0.27	0.3	0.14
II	1	0.5				0.50
	2	0.5	0.5	0.75	1.0	0.56
	3	0.5	0.5	0.88	1.0	0.59
III	1	0.1				0.10
	2	0.2	0.2	0.28	0.3	0.22
	3	0.6	0.6	0.71	0.9	0.63
IV	1	0.1				0.10
	2	0.7	0.7	0.73	0.8	0.71
	3	0.2	0.7	0.78	1.0	0.72

[a] Calculated values of $M_n(h)$ and $\bar{F}_n^*(h)$ are rounded off to two decimal places.
[b] The weight is $v = 0.75$.

$$L_n(h) \le \bar{F}_n(h) < M_n(h); \tag{15b}$$

(iii) *negatively quadrant dependent if and only if*

$$M_n(h) < \bar{F}_n(h) \le U_n(h). \tag{15c}$$

Hence, if it were possible to determine the sign of the stochastic dependence between the subevents, then the exceedance probability $\bar{F}_n(h)$ could be bounded according to either Eq. (15b) or Eq. (15c). These bounds are tighter than the Fréchet bounds (9).

3.4. Dependence sign

Our strategy is to find a logical link between some familiar properties of the river stage process $\{H_1, \dots, H_N\}$ and the sign of the stochastic dependence between the subevents.

Theorem 2. *If the river stage process* $\{H_1, \dots, H_N\}$ *is an independent process, then Eq.* (15a) *holds for all h.*

The proof is immediate: when the left side of Eq. (15a) is replaced by Eq. (4) and the right side of Eq. (15a) is replaced by Eq. (14), one obtains the definition of the independent process.

Theorem 3. *If the river stage process* $\{H_1, \dots, H_N\}$ *is a Markov process of order one and for each n* ($n = 2, \dots, N$) *the variates* (H_{n-1}, H_n) *are positively quadrant dependent, then Eq.* (15b) *holds for all h.*

The proof is given in Appendix A. Whereas the hypothesis of the Markov process has been tested and verified empirically (Krzysztofowicz and Herr, 2001), the hypothesis of positive quadrant dependence has not been tested directly to the best of our knowledge. The following result due to Lehmann (1966, Lemma 3 and Corollary 1) offers a partial recourse.

Theorem 4. *If the variates* (H_{n-1}, H_n) *are positively quadrant dependent, then the Pearson's product-moment correlation coefficient and the Spearman's rank correlation coefficient are both positive.*

Inasmuch as the Spearman's rank correlation coefficient for each pair (H_{n-1}, H_n), on the time step of 24 h within the time interval of 72 h, has been found decisively positive (Krzysztofowicz and Herr, 2001), the necessary condition for positive quadrant dependence has a strong empirical support. It is therefore plausible to conjecture that the hypotheses of Theorem 3 are realistic for short-term flood forecasting. Consequently, among the three cases covered by Theorem 1, the most plausible is case (15b).

In summary, for any $n \ge 2$, the exceedance function $\bar{F}_n$ of the maximum river stage within time interval $(t_0, t_n]$ has bounds specified by Eqs. (11a), (14) and (15b). These bounds are constructed solely in terms of the exceedance functions $\bar{\Psi}_1, \dots, \bar{\Psi}_n$ of the river stages at time instances $t_1, \dots, t_n$.

3.5. Examples

Table 1 shows four examples in which, for a fixed level h, the stage exceedance probabilities $\bar{\Psi}_1(h)$, $\bar{\Psi}_2(h)$, and $\bar{\Psi}_3(h)$ are transformed into the bounds on the flood occurrence probabilities $\bar{F}_2(h)$ and $\bar{F}_3(h)$. For each $\bar{F}_n(h)$, $n = 2$, 3, there are three bounds, $L_n(h) < M_n(h) < U_n(h)$. The probability intervals specified by the bounds may be viewed as imprecise measures of uncertainty about flood occurrence within time interval $(t_0, t_n]$. With the exception of Example II, the interval $[L_n(h), U_n(h)]$ specified by the Fréchet bounds may be tight enough to be useful for some decisions. The interval $[L_n(h), M_n(h)]$, specified by the bounds under the positive quadrant dependence hypothesis, is even tighter; it may 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.

4. Approximations to flood forecast

It is tempting to forge an estimator of $\bar{F}_n$ using the bounds. If the bounds are tight (and Table 1 shows that oftentimes they are) and if some rationale can be found for $\bar{F}_n$ to be closer to one of the bounds, either L_n or M_n, then a reasonable approximation to $\bar{F}_n$ may exist. Not only it may be useful, but it is certainly cheap: the calculation of the bounds is a simple task.

4.1. Behavior of bounds

Before constructing an estimator of $\bar{F}_n$, it is desirable to understand the behavior of $\bar{F}_n$ relative to its bounds L_n and M_n.

First, suppose the time step $\Delta t = t_n - t_{n-1}$ is constant for $n = 2, \ldots, N$. For an infinitesimal Δt, variates (H_n, H_{n-1}) are extremely positively dependent for all $n \geq 2$; consequently, $\bar{F}_n$ tends toward the lower Fréchet bound L_n. For some large Δt, variates (H_{n-1}, H_n) are stochastically independent for all $n \geq 2$; consequently, $\bar{F}_n$ tends toward the middle bound M_n.

Second, the PRSF provides a discrete-time approximation to the continuous-time random hydrograph within time interval $(t_0, t_0 + L]$. Let $t_N = t_0 + L$ and $\Delta t = t_n - t_{n-1}$ for $n = 2, \ldots, N$. When the number of time steps N is increased, the approximation is expected to improve. Thus, as $N \to \infty$, $\bar{F}_N$ converges to the exceedance function $\bar{F}$ of the maximum of the continuous-time random hydrograph within time interval $(t_0, t_0 + L]$. Ideally, it is $\bar{F}$ that one would want to know. We know only the bounds on $\bar{F}_N(h)$, which behave as follows: as $N \to \infty$, $M_N(h) \to 1$ and $L_N(h) \to \lim \max\{\bar{\Psi}_1(h), \ldots, \bar{\Psi}_N(h)\}$, which is a well defined function of h. If $\bar{F}_N$ behaved like M_N, then the probability of flood occurrence within time interval $(t_0, t_0 + L]$ would be increasing with the number of time steps N, clearly an artifact of the discretization. Therefore, $\bar{F}_N$ must behave like L_N: as the number of time steps N increases, $\bar{F}_N$ must converge to some fixed though unknown function $\bar{F}$.

Third, suppose the maximum river stage is certain to occur at time t_m, $m \in \{1, \ldots, N\}$, and only its magnitude is uncertain. Then $\bar{F} = \bar{F}_N = \bar{\Psi}_m$ and $L_N = \bar{\Psi}_m$; in other words, the lower Fréchet bound constitutes the solution.

In summary, the exceedance function $\bar{F}_n$ departs from its lower Fréchet bound as the time step Δt increases and as the degree of uncertainty about the timing of the miximum river stage increases. Any approximation to $\bar{F}_n$ must behave in this way.

4.2. Direct linear interpolator

The simplest estimator $\bar{F}_n^*$ of $\bar{F}_n$ from the bounds is the DLI

$$\bar{F}_n^*(h) = vL_n(h) + (1 - v)M_n(h), \tag{16}$$

where $L_n(h)$ is given by Eq. (11a), $M_n(h)$ is given by Eq. (14), and v is a weight bounded by $0 < v < 1$. Based on the conclusion of Section 4.1, the weight depends, in general, on the time step Δt of the PRSF and the degree of uncertainty about the timing of the maximum river stage relative to the time instances $t_1, \ldots, t_N$. As an explicit function $v = v(\Delta t)$, the weight v is a decreasing function of Δt, with $v(0) = 1$ and $v(\Delta t) \to 0$ as Δt increases. In application, v could be invariant for a given forecast point, hydrologic season, and Δt.

The last column of Table 1 shows the estimates $\bar{F}_n^*(h)$ calculated from bounds $L_n(h)$ and $M_n(h)$ according to Eq. (16).

4.3. Recursive linear interpolator

A more refined estimator $\bar{F}_n^*$ of $\bar{F}_n$ from the bounds is the RLI. It is based on recursive equations (10a), (10b) and (13) and the initial condition (8). The initial

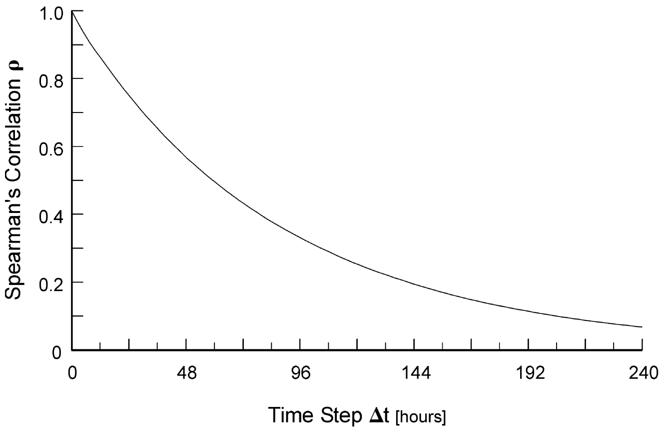

Fig. 3. Spearman's rank correlation coefficient ρ between river stages H_n and H_{n-1} as a function of the time step $\Delta t = t_n - t_{n-1}$. Values of $\rho(\Delta t)$ are average estimates for different n, conditional on the occurrence of precipitation in the 24-h period beginning at 1200 UTC on the forecast day; Eldred, Pennsylvania.

condition gives

$$\bar{F}_1^*(h) = \bar{\Psi}_1(h). \tag{17}$$

Then proceeding recursively for $n = 2, \ldots, N$, the one-step-ahead estimators of the bounds are

$$L_n^*(h) = \max\{\bar{F}_{n-1}^*(h), \bar{\Psi}_n(h)\}, \tag{18a}$$

$$M_n^*(h) = \bar{F}_{n-1}^*(h) + \bar{\Psi}_n(h) - \bar{F}_{n-1}^*(h)\bar{\Psi}_n(h), \tag{18b}$$

$$U_n^*(h) = \min\{\bar{F}_{n-1}^*(h) + \bar{\Psi}_n(h), 1\}, \tag{18c}$$

and the estimator of the exceedance probability is

$$\bar{F}_n^*(h) = \omega L_n^*(h) + (1 - \omega)M_n^*(h), \tag{19}$$

where ω is a weight bounded by $0 < \omega < 1$. The weight ω has the same general properties that the weight v in Eq. (16) has, but may have a different magnitude. In general, $\omega = \omega(\Delta t)$ with $\omega(0) = 1$ and $\omega(\Delta t) \to 0$ as Δt increases. In application, ω could be invariant for a given forecast point, hydrologic season, and Δt.

Imbedded in the algorithm is a trade-off. On one hand, the one-step-ahead bounds L_n^* and M_n^* are only approximate as they depend on the estimate $\bar{F}_{n-1}^*$. On the other hand, the one-step-ahead bounds L_n^* and M_n^* are tighter than the overall bounds L_n and M_n for $n \geq 3$. Consequently, the linear interpolation (19) has a greater degree of validity than the linear interpolation (16).

Table 2 shows two examples of the RLI estimates. The given stage exceedance probabilities $\bar{\Psi}_1(h)$, $\bar{\Psi}_2(h)$, and $\bar{\Psi}_3(h)$ are the same as those in Table 1, where the DLI estimates are reported.

There are no differences for $n = 1, 2$. But for $n = 3$, the one-step-ahead bounds are tighter and the estimates are somewhat different.

5. Tutorial examples

5.1. Weighting function

Fig. 3 shows a plot of the Spearman's rank correlation coefficient ρ between river stages H_n and H_{n-1} as a function of the time step $\Delta t = t_n - t_{n-1}$.

Table 2
One-step-ahead bounds on the flood occurrence probability $\bar{F}_n(h)$ obtained from the PRSF, and the RLI estimate $\bar{F}_n^*(h)$

No.	n	PRSF	One-step-ahead bounds[a] on $\bar{F}_n(h)$			RLI estimate[b]
		$\bar{\Psi}_n(h)$	$L_n^*(h)$	$M_n^*(h)$	$U_n^*(h)$	$\bar{F}_n^*(h)$
I	1	0.1				0.10
	2	0.1	0.1	0.19	0.2	0.12
	3	0.1	0.12	0.21	0.21	0.14
II	1	0.5				0.50
	2	0.5	0.5	0.75	1.0	0.56
	3	0.5	0.56	0.78	1.0	0.62

[a] Calculated values of the bounds and the estimate are rounded off to two decimal places.
[b] The weight is $\omega = 0.75$.

Table 3
Effect of the time step Δt of the PRSF upon the bounds of the flood occurrence probability $\bar{F}_n(h)$ and upon the RLI estimate $\bar{F}^*_n(h)$

Δt	$\omega(\Delta t)$	n	t_n	PRSF	Bounds[a] on $\bar{F}_n(h)$			RLI estimate
				$\bar{\Psi}_n(h)$	$L_n(h)$	$M_n(h)$	$U_n(h)$	$\bar{F}^*_n(h)$
24	0.57	1	24	0.2				0.20
		2	48	0.4	0.4	0.52	0.6	0.45
12	0.75	1	12	0.1				0.10
		2	24	0.2	0.2	0.28	0.3	0.22
		3	36	0.3	0.3	0.50	0.6	0.34
		4	48	0.4	0.4	0.70	1.0	0.45
6	0.80	1	6	0.05				0.05
		2	12	0.1	0.1	0.15	0.15	0.11
		3	18	0.15	0.15	0.27	0.3	0.17
		4	24	0.2	0.2	0.42	0.5	0.23
		5	30	0.25	0.25	0.56	0.75	0.28
		6	36	0.3	0.3	0.69	1.0	0.34
		7	42	0.35	0.35	0.80	1.0	0.39
		8	48	0.4	0.4	0.88	1.0	0.45

[a] Calculated values of $M_n(h)$ and $\bar{F}^*_n(h)$ are rounded off to two decimal places.

The smooth function interpolates values of $\rho(\Delta t)$ for $\Delta t = 6, 12, 24, 48, \ldots, 240$ h. For each Δt, the value $\rho(\Delta t)$ was obtained as an average of estimates for different n, calculated conditional on the occurrence of precipitation in the 24-h period beginning at 1200 UTC on the forecast day. Inasmuch as $\rho(\Delta t)$ conditional on the occurrence of precipitation is smaller than $\rho(\Delta t)$ conditional on the nonoccurrence of precipitation, the former is more indicative of the behavior of the river stage process during a flood event.

Two observations can be drawn from Fig. 3. First, the Spearman's rank correlation coefficient is decisively positive, which supports the rationale expounded in Section 3.4 for choosing M_n as the upper bound on $\bar{F}_n$. Second, the behavior of ρ in-the-large is similar to the postulated behavior of the weighting functions v and ω. Hence $\rho(\Delta t)$, or some monotone transformation thereof, may offer a reasonable value of the weight.

5.2. Effect of time step

Table 3 reports an example whose purpose is to support the theoretical considerations in Section 4.1, in particular (i) the effect of the time step Δt on the bounds and (ii) the plausibility of the weighting function ω. The example is constructed as follows. First, the stage exceedance probability $\bar{\Psi}_n(h)$ for $n = 1, \ldots, N$ increases linearly with time t_n in order to prevent estimation errors due to the discretization; such errors occur when the PRSF with large Δt misses a significant nonlinear (with time) change in the stage exceedance probabilities specified by the PRSF with small Δt. Second, the PRSF is for time interval of 48 h. Third, three alternative discretizations of this time interval are considered, with $\Delta t = 24$, 12, 6 h, so that the number of time steps is $N = 2$, 4, 8, respectively. Fourth, the weighting function is chosen without any optimization as $\omega(\Delta t) = [\rho(\Delta t)]^{1/2}$, where $\rho(\Delta t)$ comes from Fig. 3.

The results support two observations. (i) The lower bound $L_N(h)$ is the same regardless of N, whereas the middle bound $M_N(h)$ increases with N, as conjectured. (ii) The estimate $\bar{F}^*_N(h)$ is nearly identical regardless of N, which is a desired result. Also, nearly identical are $\bar{F}^*_1(h) = 0.20$ when $\Delta t = 24$, $\bar{F}^*_2(h) = 0.22$ when $\Delta t = 12$, and $\bar{F}^*_4(h) = 0.23$ when $\Delta t = 6$, as they should be. Likewise, practically identical are $\bar{F}^*_3(h) = 0.34$ when $\Delta t = 12$, and $\bar{F}^*_6(h) = 0.34$ when $\Delta t = 6$, as they should be.

In conclusion, it is possible to select the weighting function ω for the RLI (and also for the DLI) so that the resultant estimates of the flood occurrence probabilities behave reasonably and are consistent regardless of the time step. Of course, one could set up a formal estimation problem to find an optimal value of $\omega(\Delta t)$, but this is done best in the context of a specific implementation when samples of PRSFs and actual flood hydrographs are available.

5.3. Effect of partitioning time interval

An ensemble streamflow prediction system tested by the NWS (Adams et al., 1999) outputs random realizations of the discrete-time hydrograph at 6-h steps for 72 h ahead; thus $N = 12$. Recently, some analysts proposed to estimate the exceedance functions of the maximum river stages within three nonoverlapping time intervals $(t_0, t_4]$, $(t_4, t_8]$, $(t_8, t_{12}]$. The presumption was that such exceedance functions constitute a useful flood forecast. According to the theory presented herein and elsewhere

Table 4
Example V of the flood occurrence probabilities calculated for time interval $(t_0, t_4]$, and example VI of the flood occurrence probabilities calculated independently for each of the two nonoverlapping time subintervals $(t_0, t_2]$ and $(t_2, t_4]$

No.	n	PRSF	Bounds[a] on $\bar{F}_n(h)$			RLI estimate[b]
		$\bar{\Psi}_n(h)$	$L_n(h)$	$M_n(h)$	$U_n(h)$	$\bar{F}_n^*(h)$
V	1	0.2				0.20
	2	0.7	0.7	0.76	0.9	0.72
	3	0.3	0.7	0.83	1.0	0.74
	4	0.4	0.7	0.90	1.0	0.76
VI	1	0.2				0.20
	2	0.7	0.7	0.76	0.9	0.72
	3[c]	0.3				0.30
	4[c]	0.4	0.4	0.58	0.7	0.45

[a] Calculated values of $M_n(h)$ and $\bar{F}_n^*(h)$ are rounded off to two decimal places.
[b] The weight is $\omega = 0.75$.
[c] In the calculation of the bounds and the estimate these times are treated as $n = 1, 2$.

(Krzysztofowicz, 1993), a flood forecast (when limited to three functions) would consist of the exceedance functions of the maximum river stages within three nested time intervals $(t_0, t_4]$, $(t_0, t_8]$, $(t_0, t_{12}]$. The following simplified examples contrast the two schemes and illustrate the pitfall of the former.

Table 4 shows two examples, each with four time steps ($N = 4$). In Example V, the flood occurrence probabilities are calculated according to the theory presented herein. In Example VI, the flood occurrence probabilities are calculated according to the scheme wherein the time interval $(t_0, t_4]$ is partitioned into two nonoverlapping subintervals $(t_0, t_2]$ and $(t_2, t_4]$. Clearly, the two forecasts convey very different information and may create a very different perception of risk. The first forecast conveys the total probability of flood occurrence within each of the successively longer time intervals; this probability compounds over time as the river stage process evolves. The second forecast arbitrarily partitions the river stage process into independent subprocesses; then the probability of flood occurrence is calculated independently for each of the subprocesses. Although the latter probabilities are properly defined, one may question (i) their usefulness for making rational decisions about issuing warning or ordering evacuation and (ii) their veracity as measures of the flood risk for the general public.

In conclusion, the PFF should be defined on a nested set of time intervals, the longest of which is the time interval for which the PRSF is available.

6. Case study

6.1. Bounds on PFF

Given the sequence of exceedance functions $\{\bar{\Psi}_n : n = 1, \ldots, 12\}$ shown in Fig. 1, Eqs. (11a), (14), and (11b) were used to calculate the sequence of bounds $\{L_n, M_n, U_n : n = 1, \ldots, 12\}$, which is displayed in Fig. 4. Up to $n = 4$, the three bounds essentially overlay each other; for $n > 4$ the bounds define two regions, each increasing with n. Of primary interest is the region defined by the lower bound L_n and the middle bound M_n between which lies the unknown exceedance function $\bar{F}_n$ of the maximum river stage Z_n. It is apparent that the ability to infer the sign of the stochastic dependence, and therefore to select one of the two regions, reduces the imprecision with which the exceedance function $\bar{F}_n$ is specified by the bounds. It is also apparent that the imprecision increases with lead time. Thus the bounds may be useful for flood warning decisions with lead times shorter than the time interval for which the PRSF is available.

The decision maker can use the bounds by interpreting them as follows: the probability of level h being exceeded within time interval $(t_0, t_n]$ is at least $L_n(h)$ and at most $M_n(h)$. If $L_n(h)$ is greater than a threshold probability for taking action, then it is optimal to initiate action, and the knowledge of the exceedance probability $\bar{F}_n(h)$ is redundant.

Fig. 5 depicts the evolution of each bound with time. Clearly, each bound evolves monotonically as $L_{n-1} \le L_n$, $M_{n-1} \le M_n$, and $U_{n-1} \le U_n$ for $n = 2, \ldots, 12$. The lower bound L_n reaches its steady shape at $n = 8$; the other two bounds, M_n and U_n, continue to evolve until $n = 12$. Fig. 5 shows them only up to $n = 7$ and $n = 11$, respectively, in order to avoid blurring of lines.

6.2. Estimate of PFF

The PRSF $\{\bar{\Psi}_n : n = 1, \ldots, 12\}$ shown in Fig. 1 was

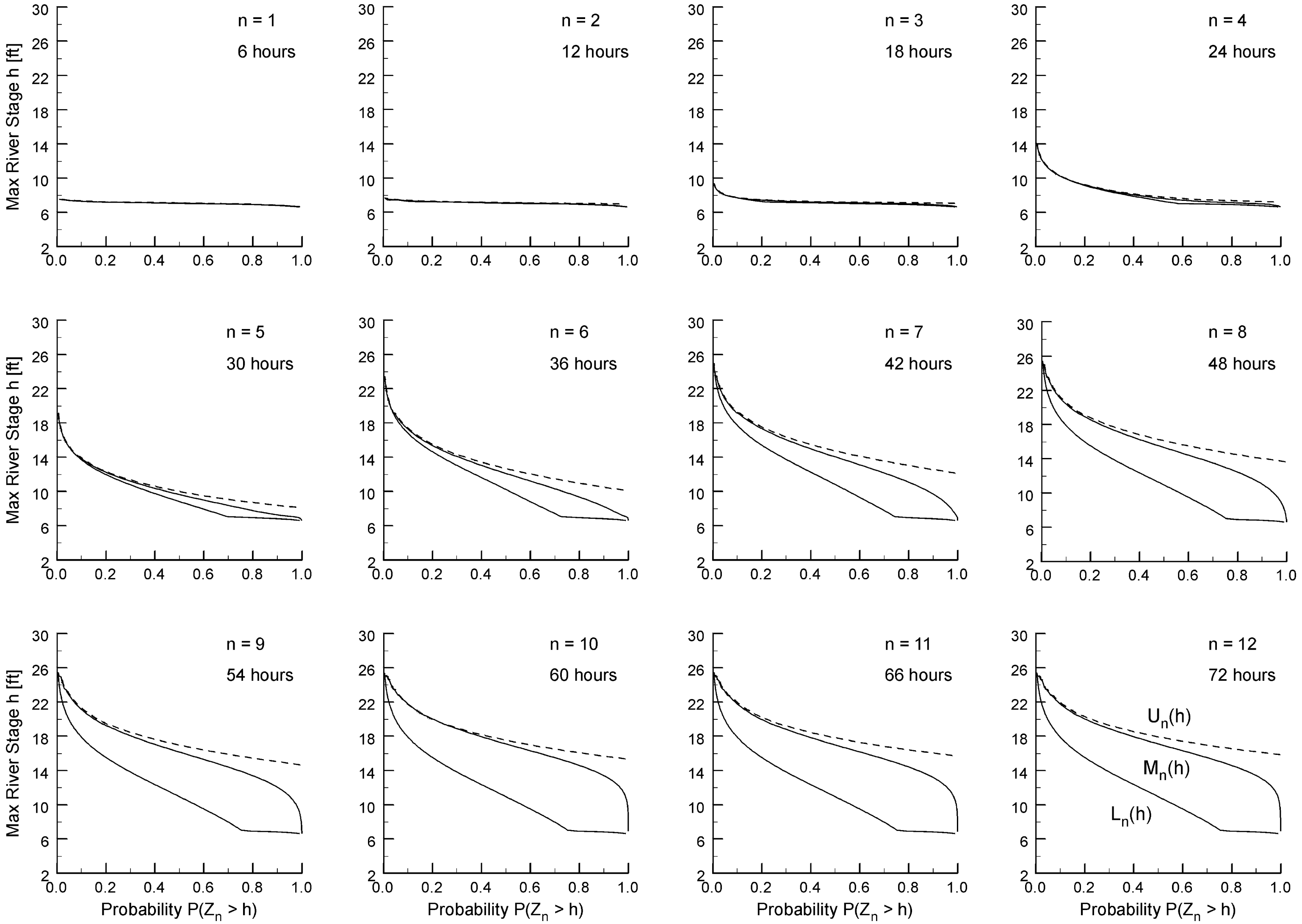

Fig. 4. Bounds on the exceedance function $\bar{F}_n$ of the maximum river stage Z_n within time interval $(t_0, t_n]$: lower bound L_n, middle bound M_n, and upper bound U_n; time step $\Delta t = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

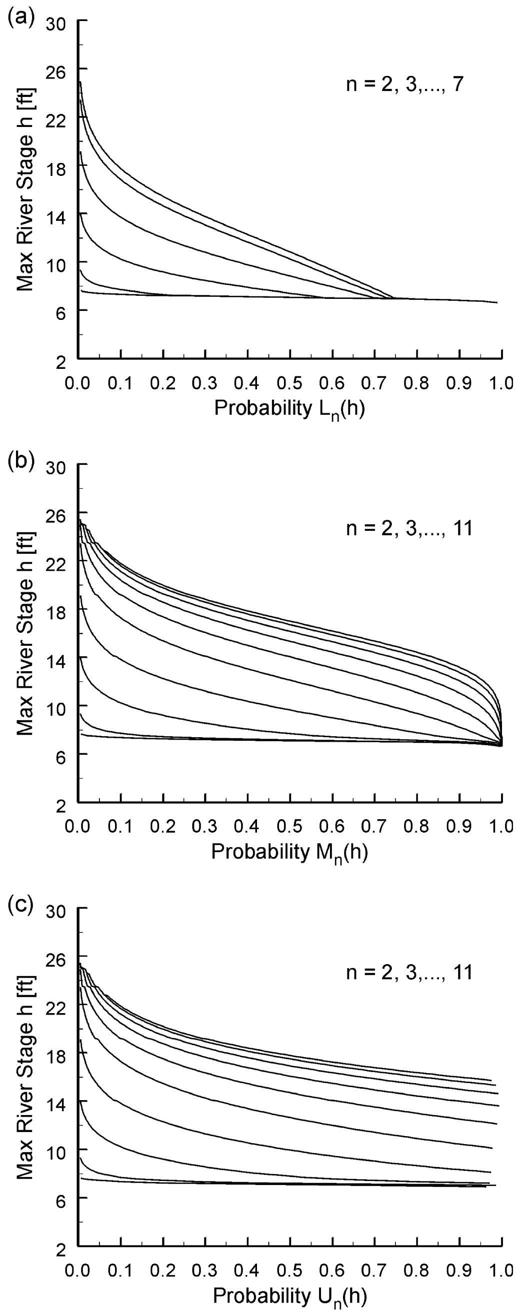

Fig. 5. Evolution of (a) lower bound L_n with n ($n = 2, \ldots, 7$); (b) middle bound M_n with n ($n = 2, \ldots, 11$); and (c) upper bound U_n with n ($n = 2, \ldots, 11$).

transformed into a PFF $\{\bar{F}_n : n = 1, \ldots, 12\}$ shown in Fig. 6. Specifically, Fig. 6 shows the estimates $\bar{F}_n^*$ obtained via the RLI with the weight $\omega = 0.8$. As postulated in Section 4.1, the shape of the exceedance function $\bar{F}_n^*$ follows closer the shape of the lower bound L_n than the shape of the middle bound M_n.

Like the bounds, the exceedance function $\bar{F}_n$ evolves monotonically with time: $\bar{F}_{n-1} \leq \bar{F}_n$ for $n = 2, \ldots, 12$. 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. This property is vivid in Fig. 7. It shows the isoprobability time series $\{z_{np} : n = 1, \ldots, 12\}$ 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. As can be seen in Fig. 7, $z_{n-1,p} \leq z_{np}$ for $n = 2, \ldots, 12$ and for each p.

6.3. 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 instance at which river stage process $\{H_1, \ldots, H_N\}$ exceeds level h for the first time; it is a discrete variate taking values in the set $\{t_1, \ldots, t_N\}$. 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

$$P(T(h) \leq t_n) = \bar{F}_n(h); \tag{20}$$

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 RLI estimates $\bar{F}_n^*(h)$ of $\bar{F}_n(h)$ shown in Fig. 6, the distribution function of $T(h)$ is plotted in Fig. 8 for $h = 10$, 14, 18 ft. 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 lead time for a given zone of the floodplain. (Yet such a display cannot be produced based on forecasts which arbitrarily partition the time scale into nonoverlapping time intervals, as detailed in Section 5.3.)

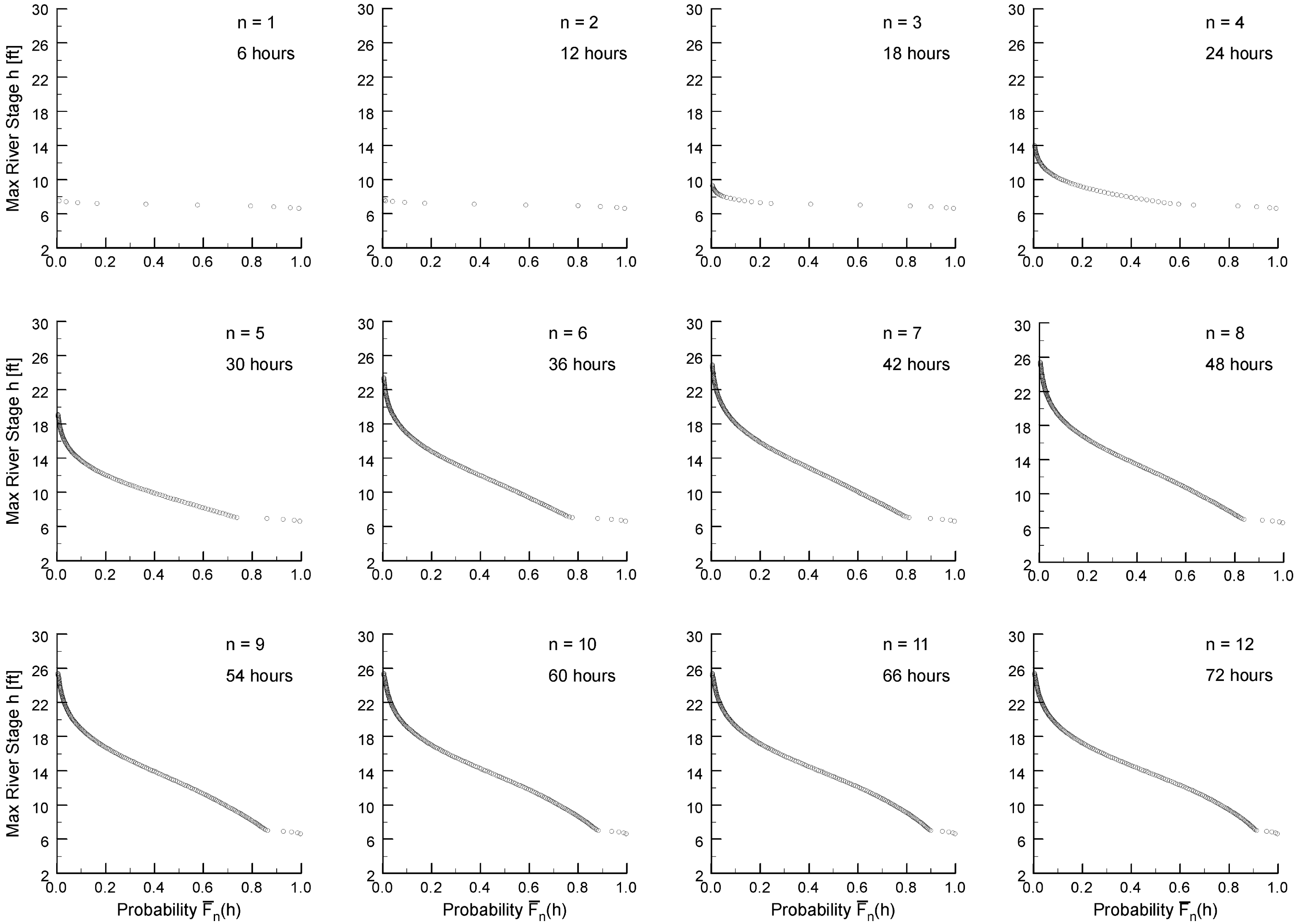

Fig. 6. A PFF produced from the PRSF via the RLI estimator: sequence of exceedance functions $\{\bar{F}_n : n = 1, \ldots, 12\}$ of maximum river stages within time intervals $(t_0, t_n]$ $(n = 1, \ldots, 12)$; time step $\Delta t = 6$ h, counting from 1200 UTC on the forecast day; Eldred, Pennsylvania.

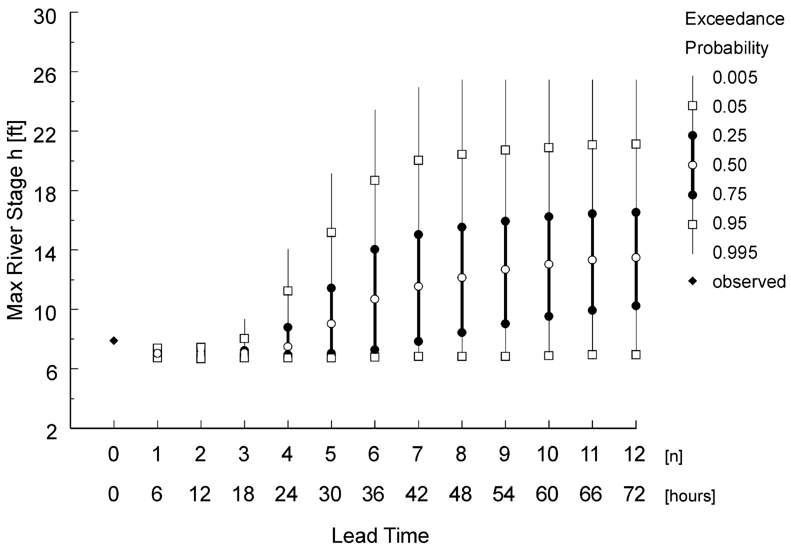

Fig. 7. Isoprobability time series $\{z_{np} : n = 1, \ldots, 12\}$ of quantiles of the maximum river stages having the exceedance probability $p = \bar{F}_n^*(z_{np})$, for seven values of p; $\bar{F}_n^*$ is the RLI estimate of the exceedance function $\bar{F}_n$.

7. Closure

A PFF is needed to support flood warning and response decisions. Given a PRSF, it is possible to construct bounds on and estimators of the PFF. The Fréchet bounds hold without any assumptions; the tighter bounds hold under the assumption of positive quadrant dependence, which is likely to be satisfied by probabilistic forecasts of river stage processes. The tighter bounds are used to construct two simple estimators of the PFF. Each of the estimators requires a single parameter. An analysis of the behavior of the bounds suggests that the value of this parameter could be invariant for a given forecast point, hydrologic season, and time step.

The simplicity of the construction methods makes

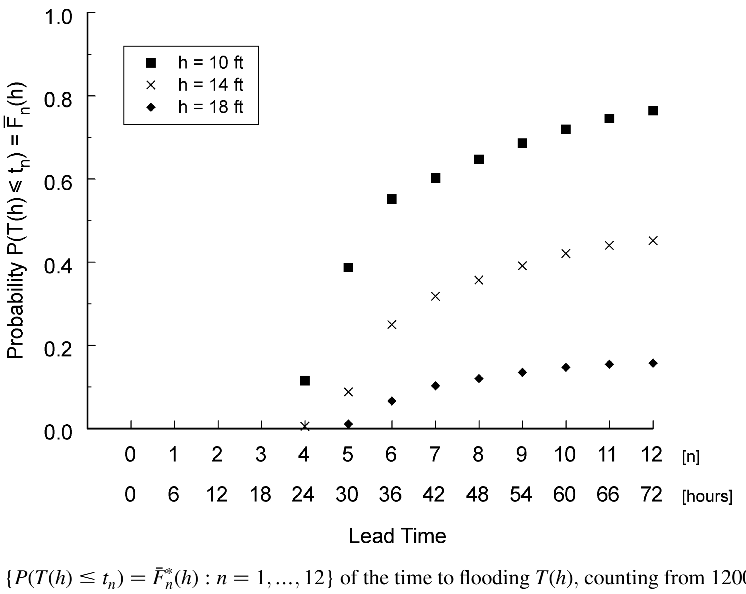

Fig. 8. Distribution function $\{P(T(h) \leq t_n) = \bar{F}_n^*(h) : n = 1, \ldots, 12\}$ of the time to flooding $T(h)$, counting from 1200 UTC on the forecast day, for three levels $h = 10$, 14, 18 ft; $\bar{F}_n^*$ is the RLI estimate of the exceedance function $\bar{F}_n$; Eldred, Pennsylvania.

them potentially attractive for real-time forecasting. Essentially, an approximate PFF can be constructed from the PRSF at no additional cost. Whereas the bounds have theoretical justification, the estimators have to be validated experimentally for each application.

Acknowledgments

Coire Maranzano performed the numerical calculations and made the drawings. Research leading to this article was funded by the National Oceanic and Atmospheric Administration under award No. NA07WD0097 'Probabilistic Hydrometeorological Forecast System' to the University of Virginia.

Appendix A. Proof of Theorem 3

The hypotheses of the theorem can be stated formally as follows. (i) The process $\{H_1, \ldots, H_N\}$ is Markov of order one if and only if

$$P(H_1 \le h_1, \ldots, H_N \le h_N) = \prod_{k=2}^{N} P(H_k \le h_k | H_{k-1} \le h_{k-1}) P(H_1 \le h_1) \quad \text{(A1)}$$

for all $h_1, \ldots, h_N$. (ii) For each n $(n = 2, \ldots, N)$ the variates (H_{n-1}, H_n) are positively quadrant dependent if and only if

$$P(H_n \le h_n | H_{n-1} \le h_{n-1}) > P(H_n \le h_n) \quad \text{(A2)}$$

for all h_n, h_{n-1}.

It suffices to show that the above two hypotheses imply that for any n $(n = 2, \ldots, N)$ and all h, the subevents $\{Z_{n-1} \le h\}$ and $\{H_n \le h\}$ are positively quadrant dependent. Beginning with Eq. (6) and making use of Eq. (A1) gives

$$\begin{aligned} P(Z_{n-1} \le h, H_n \le h) &= P(H_1 \le h, \ldots, H_n \le h) \\ &= \prod_{k=2}^{n} P(H_k \le h | H_{k-1} \le h) P(H_1 \le h) \\ &= P(H_n \le h | H_{n-1} \le h) P(H_1 \le h, \ldots, H_{n-1} \le h) \\ &= P(H_n \le h | H_{n-1} \le h) P(Z_{n-1} \le h). \end{aligned} \quad \text{(A3)}$$

Now making use of Eq. (A2) on the right side gives

$$P(Z_{n-1} \le h, H_n \le h) > P(H_n \le h) P(Z_{n-1} \le h), \quad \text{(A4)}$$

which when divided by $P(Z_{n-1} \le h)$ yields

$$P(H_n \le h | Z_{n-1} \le h) > P(H_n \le h). \quad \text{(A5)}$$

By Lemma 1 in Lehmann (1966), this is equivalent to Eq. (12b).

References

Adams, T.E., Welles, E., Schaake, J., 1999. Operational experience with ensemble streamflow prediction in the Upper Monongahela probabilistic forecast demonstration project. Preprints, 14th Conference on Hydrology, American Meteorological Society, Boston, MA, pp. 8–11.

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

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 Resour. Res. 30 (4), 1145–1152.

Krzysztofowicz, R., 1993. A theory of flood warning systems. Water Resour. Res. 29 (12), 3981–3994.

Krzysztofowicz, R., 1999. Probabilities for a period and its subperiods: theoretical relations for forecasting. Monthly Weather Rev. 127 (2), 228–235.

Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. J. Hydrol. 4546 – this issue.

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

Krzysztofowicz, R., Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. J. Hydrol. 249 (1–4), 46–68.

Lardet, P., Obled, Ch., 1994. Real-time flood forecasting using a stochastic rainfall generator. J. Hydrol. 162, 391–408.

Lehmann, E.L., 1966. Some concepts of dependence. Ann. Math. Stat. 37, 1137–1153.