Journal of Hydrology 517 (2014) 643–651

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Probabilistic flood forecast: Exact and approximate predictive distributions

Roman Krzysztofowicz *

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

ARTICLE INFO

Article history:
Received 4 December 2013
Received in revised form 16 April 2014
Accepted 20 April 2014
Available online 9 June 2014
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief

Keywords:
Stochastic processes
Statistical analysis
Probability
Rivers
Floods

SUMMARY

For quantification of predictive uncertainty at the forecast time t_0, the future hydrograph is viewed as a discrete-time continuous-state stochastic process $\{H_n: n = 1, \ldots, N\}$, 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 $\{\overline{F}_n: n = 1, \ldots, N\}$ such that $\overline{F}_n(h) = P(Z_n > h)$, where P stands for probability, and Z_n is the maximum river stage within time interval $(t_0, t_n]$, practically $Z_n = \max\{H_1, \ldots, H_n\}$. 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 $\overline{F}_n$, 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 $\overline{F}_n$, which can be constructed from the bounds via a *recursive linear interpolator* (RLI) without information about the stochastic dependence in the process $\{H_1, \ldots, H_n\}$, 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.

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

* Address: Department of Systems Engineering, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747, USA. Tel.: +1 434 982 2067; fax: +1 434 982 2972.
E-mail address: rk@virginia.edu
URL: http://www.faculty.virginia.edu/rk/

http://dx.doi.org/10.1016/j.jhydrol.2014.04.050
0022-1694/

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 $\{H_n: \ n = 1, \ldots, N\}$ 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):

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

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 $\{H_n: \ n = 1, \ldots, N\}$, 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 $H_1, \ldots, H_N$.

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

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

where

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

That is, $\theta_n(\cdot | h_{n-1}, \ldots, h_1)$ is the predictive density function p of river stage H_n at time t_n, **conditional on** (i) the river stages $H_{n-1} = h_{n-1}, \ldots, H_1 = h_1$ 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 $\{H_n: \ n = 1, \ldots, N\}$, function $\theta_n(\cdot | h_{n-1}, \ldots, h_1)$ characterizes the uncertainty in the *one-step transition* from the observed (initial) river stage $H_0 = h_0$ and the hypothesized (preceding) river stages $H_1 = h_1, \ldots, H_{n-1} = h_{n-1}$, 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 $H_1, \ldots, H_N$:

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

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 $\{H_n: \ n = 1, \ldots, N\}$.

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 $(n = 1, \ldots, N)$ 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 is the *river stage* at time t_n; it is a continuous variate which may take any value above the gauge datum.

Z_n is the *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

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

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

$$\overline{F}_n(h) = P(Z_n > h) = 1 - P(Z_n \leqslant h) = 1 - P(H_1 \leqslant h, \ldots, H_n \leqslant h); \tag{2}$$

that is, $\overline{F}_n(h)$ 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

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

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