# BAYESIAN ENSEMBLE FORECAST OF RIVER STAGES

# AND ENSEMBLE SIZE REQUIREMENTS

By

Henry D. Herr

Office of Hydrologic Development

National Weather Service

1325 East-West Highway

Silver Spring, MD 20910, USA

and

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–0801**

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

July 2008; Revised July 2009, January 2010

Published in *Journal of Hydrology*, **387**, 151–164 2010

---

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

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

**ABSTRACT**

The problem is to provide a short-term, probabilistic forecast of a river stage time series $\{H_1, ..., H_N\}$ based on a probabilistic quantitative precipitation forecast. The Bayesian forecasting system (BFS) for this problem is implemented as a Monte-Carlo algorithm that generates an ensemble of realizations of the river stage time series. This article (i) shows how the analytic-numerical BFS can be used as a generator of the Bayesian ensemble forecast (BEF), (ii) demonstrates the properties of the BEF, and (iii) investigates the sample size requirements for ensemble forecasts (produced by the BFS or by any other system).

The investigation of the ensemble size requirements exploits the unique advantage of the BFS, which outputs the exact, analytic, predictive distribution function of the stochastic process $\{H_1, ..., H_N\}$, as well as can generate an ensemble of realizations of this process from which a sample estimate of the predictive distribution function can be constructed. By comparing the analytic distribution with its sample estimates from ensembles of different sizes, the smallest ensemble size $M^*$ required to ensure a specified expected accuracy can be inferred. Numerical experiments in four river basins demonstrate that $M^*$ depends upon the kind of probabilistic forecast that is constructed from the ensemble. Three kinds of forecasts are constructed: (i) a probabilistic river stage forecast (PRSF), which for each time $n$ $(n = 1, ..., N)$ specifies a predictive distribution function of $H_n$; (ii) a probabilistic stage transition forecast (PSTF), which for each time $n$ specifies a family (for all $h_{n-1}$) of predictive one-step transition distribution functions from $H_{n-1} = h_{n-1}$ to $H_n$; and (iii) a probabilistic flood forecast (PFF), which for each time $n$ specifies a predictive distribution function of $\max\{H_1, ..., H_n\}$.

Overall, the experimental results demonstrate that the smallest ensemble size $M^*$ required for accurate estimation (or numerical representation) of these predictive distribution functions is

(i) insensitive to experimental factors and on the order of several hundreds for the PRSF and the PFF, and (ii) sensitive to experimental factors and on the order of several thousands for the PSTF. The general conclusions for system developers are that the ensemble size is an important design variable, and that the optimal ensemble size $M^*$ depends upon the purpose of the forecast: for dynamic control problems (which require a PSTF), $M^*$ is likely to be larger by a factor of 3–20 than it is for static decision problems (which require a PRSF or a PFF).

*Keywords*: Forecasting; Uncertainty; Bayesian analysis; Stochastic processes; Statistical analysis; Probability; Ensemble; Rivers; Floods

## TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1 Bayesian Forecasting System

Previous research formulated a general Bayesian theory of 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 systems 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 system 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, 2002). 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 whose product gives the predictive joint density function of the river stages $H_1, ..., H_N$ at times $t_1, ..., t_N$ (Krzysztofowicz and Maranzano, 2004b). As such, the PSTF provides a complete, analytic characterization of predictive uncertainty about the process $\{H_n : n = 1, ..., N\}$.

## 1.2 Bayesian Ensemble Forecast

The PSTF is analytic and has the form of input needed by stochastic control models. But many decision support systems for reservoir control, waterway operation, flood mitigation, or water quality management utilize deterministic simulation models which need input in the form of a time series — a realization of the process $\{H_1, ..., H_N\}$. For such systems, an ensemble of realizations can provide a numerical characterization of predictive uncertainty about the process $\{H_1, ..., H_N\}$. The degree to which such a numerical characterization is accurate (vis-a-vis the analytic characterization) depends upon the number of realizations — the *ensemble size*.

The objective of this article is threefold: (i) To show how the analytic-numerical BFS can be used as a generator of the *Bayesian ensemble forecast* (BEF) of river stages (discharges, or volumes). (ii) To demonstrate the properties of the BEF. (iii) To investigate the sample size requirements for ensemble forecasts (produced by the BFS or by any other system).

The investigation of the ensemble size requirements exploits the unique advantage of the BFS, which outputs the exact, analytic, predictive distribution function of the stochastic process $\{H_1, ..., H_N\}$, as well as can generate an ensemble of realizations of this process from which a sample estimate of the predictive distribution function can be constructed. By comparing the analytic distribution with its sample estimates from ensembles of different sizes, the accuracy of the numerical representation of the predictive uncertainty via an ensemble forecast can be ascertained.

Section 2 recalls the BFS. Section 3 explains the theory and the algorithm for the BEF generator. Sections 4, 5, and 6 present the procedures for estimating a PRSF, a PSTF, and a PFF (probabilistic flood forecast) from an ensemble forecast. Section 7 describes an experiment to ascertain the ensemble size required for accurate estimation of the predictive distributions that constitute the PRSF, the PSTF and the PFF. Section 8 reports the results of that experiment. Finally, Section 9 draws conclusions. (The appendix provides a glossary of mathematical symbols.)

### 1.3 Ensemble Bayesian Forecasting System

The usage of the analytic-numerical BFS as a generator of the ensemble forecast must be distinguished from the *ensemble Bayesian forecasting system* (EBFS), which implements the Bayesian theory of probabilistic forecasting entirely and exactly using Monte Carlo simulation. While the theoretical framework for the EBFS was published (Krzysztofowicz, 2001), a prototype system is just being developed and will be described in a future paper.

## 2. SYSTEM ELEMENTS

This section recalls (from Krzysztofowicz and Maranzano, 2004b) some elements of the analytic-numerical BFS necessary for the understanding of the BEF generator.

### 2.1 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 (Kelly and Krzysztofowicz, 2000). The hydrologic uncertainty processor quantifies hydrologic uncertainty under the hypothesis that there is no precipitation uncertainty (Krzysztofowicz and Maranzano, 2004a). Then the two uncertainties are optimally integrated to produce a PSTF.

### 2.2 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 $W$. Because the occurrence of precipitation alone is a significant predictor of river stages and a partial explainer of hydrologic uncertainty, it is treated explicitly as a predictand as follows.

Let $V$ denote an indicator of precipitation occurrence, with $V = 0 \Leftrightarrow W = 0$ and $V = 1 \Leftrightarrow$

$W > 0$. Hence the precipitation event is denoted $V = v$, where $v \in \{0, 1\}$. The first part of the PQPF specifies (i) the probability of precipitation occurrence during the period and over the basin

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

such that $0 \leq \nu \leq 1$, and (ii) the distribution function of the total precipitation amount $W$, conditional on the hypothesis that precipitation occurs, which for any $w > 0$ specifies $T_1(w) = P(W \leq w|V = 1)$. Then the (unconditional) distribution function of $W$ is specified for any $w \geq 0$ by

$$P(W \leq w) = (1 - \nu) + \nu T_1(w). \tag{2}$$

It is a mixed (binary-continuous) distribution function which assigns a probability mass of $(1 - \nu)$ to event $W = 0$ and spreads the remaining probability mass $\nu$ over the interval $(0, \infty)$. This structure of the mixture is mapped by the BFS to all the predictive distribution functions of river stages. Hence its importance.

A system that produces the PQPF is not part of the BFS. It must be developed separately and may employ any forecasting method. However, it must meet two requirements: (i) the PQPF must be in the specified format, and (ii) the PQPF system must be well calibrated (Krzysztofowicz, 1999).

In the examples reported throughout the article, the PQPF for a river basin was prepared judgmentally by a meteorologist in the Pittsburgh office of the US National Weather Service (NWS) according to a formal methodology (Krzysztofowicz et al., 1993; Krzysztofowicz and Pomroy, 1997; Krzysztofowicz and Sigrest, 1997). The calibration of the forecaster was verified statistically (Krzysztofowicz and Sigrest, 1999). The PQPF was for a 24-h period beginning at 1200 UTC (Universal Time Coordinated). The temporal disaggregation was into four 6-h subperiods. The spatial disaggregation was into sub-basins, as described in the next section.

### 2.3 Hydrologic Model

The BFS can be attached to any deterministic hydrologic model (lumped, semi-distributed, or distributed) which simulates the response of a river basin to a time series of precipitation amounts, and outputs a time series of model river stages at a forecast point.

The examples throughout the article are for the forecast point Eldred, Pennsylvania, located in the headwater of the Allegheny River and closing a drainage area of 550 miles$^2$ (1430 km$^2$). The time to peak of the unit hydrograph is 30 h. Forecasts are produced daily based on input data available at 1200 UTC. A forecast is for 3 days ahead, either in 6-h steps or in 24-h steps.

The conclusions at the end of the article are based on aggregate results for four forecast points, whose characteristics are listed in Table 1. Hydrologic models come from the NWS operational forecast system and have parameters estimated for each basin. For Eldred, it is a lumped model consisting of a continuous antecedent precipitation index, a unit hydrograph, a baseflow procedure, and a stage-discharge conversion (Sittner et al., 1969). For Dailey, Philippi, and Parsons, it is the Sacramento catchment model (Burnash, 1995) applied either in a lumped version (for Dailey) or in a semi-distributed version (for Philippi and Parsons). All input data come from the operational archives of the NWS. The precipitation input is in the form of a time series of 6-h spatially averaged precipitation amounts for a basin or each sub-basin. The hydrologic model for a forecast point outputs a time series of river stages at 6-h steps.

### 2.4 System Operation

The operation of the analytic-numerical BFS in real time is outlined below, borrowing from Sections 2–6 in Krzysztofowicz and Maranzano (2004b), where the reader can find the mathematical details. The purpose here is to explain, in concept, the genesis of Eq. (9) in Section 3, which provides the basis for the BEF generator.

*Hydrologic uncertainty processor* (HUP). Before real-time forecasting can begin, one must have a hydrologic model with parameters estimated for the river basin above the forecast point, and the HUP, formulated according to the Bayesian theory (Krzysztofowicz and Maranzano, 2004a), with parameters estimated according to the proper methodology (Krzysztofowicz and Kelly, 2000); this HUP is for the forecast point, for the PQPF time scale (e.g., 24 h divided into 6-h subperiods), and for the BEF time scale (e.g., 72 h in 24-h steps). For each lead time $n$ $(n = 1, ..., N)$ and for each hypothesized precipitation event $V = v\,(v = 0, 1)$, the HUP specifies a family (for all $s_n, h_{n-1}, h_0$) of conditional posterior (in the Bayesian sense) density functions of the actual river stage $H_n$:

$$p(h_n|s_n, h_{n-1}, h_0, V = v) = \phi_{nv}(h_n|s_n, h_{n-1}, h_0). \tag{3}$$

Above, $p$ is the generic symbol for a density function; $h_n$ is a realization of the actual river stage $H_n$ at lead time $n$, with $h_0$ being the observed river stage at the forecast time; $s_n$ is a realization of the model river stage $S_n$ at lead time $n$; and $\phi_{nv}\,(\cdot|s_n, h_{n-1}, h_0)$ is the operational notation for the conditional posterior density function, showing only the variables $(s_n, h_{n-1}, h_0)$ which define the family. Except for $s_n$, which must be included if the hydrologic model is informative at all, the list of the other conditioning variables (here $h_{n-1}, h_0, v$) may be modified, shortened or expanded, as appropriate. The above variables are sufficient for each of the four forecast points and were selected from a list of potential explanatory variables. A specialized statistical analysis of data that accomplishes this task was developed by Maranzano and Krzysztofowicz (2004).

*Real-time inputs*. Let $\mathbf{u}_0$ denote the vector 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, observed river stage $h_0$). The PQPF for the river basin supplies (i) the probability $\nu$ of precipitation occurrence $(V = 1)$; (ii) the distribution function $T_1$ of the total precipitation amount $W$, conditional on the hypothesis that $V = 1$; and (iii) a deterministic forecast of the spatio-temporal disaggregation of $W$, conditional on the hypothesis that $V = 1$, which is specified by the matrix $\boldsymbol{\xi}$ of expected disaggregation factors (e.g., for Parsons, it is a $4 \times 5$ matrix that disaggregates any total precipitation amount $W = w$ into 4 subperiods and 5 sub-basins). Details of this specification and its justification (in terms of the optimality-complexity trade-off and the deterministic equivalence principle) can be found in Kelly and Krzysztofowicz (2000, Section 6) and Krzysztofowicz (2002, Section 3).

*Precipitation uncertainty processor* (PUP). Seven quantiles of the total precipitation amount $W$ are judiciously selected from the conditional distribution function $T_1$; each quantile is disaggregated in time and space using the matrix $\boldsymbol{\xi}$; thereby seven precipitation inputs to the hydrologic model are created. Next, the hydrologic model is run seven times; the output is seven time series of model river stages. For each lead time $n \, (n = 1, ..., N)$, the PUP takes the seven output realizations and supplies two conditional output density functions of the model river stage $S_n$:

$$p(s_n|\mathbf{u}_0, V = 0) = \delta(s_n - s_{n0}), \tag{4a}$$

$$p(s_n|T_1, \boldsymbol{\xi}, \mathbf{u}_0, V = 1) = \pi_{n1}(s_n), \qquad s_n > s_{n0}. \tag{4b}$$

Above, $s_{n0}$ is the model river stage at lead time $n$ resulting from zero total precipitation amount; $\delta$ is the Dirac function; and $\pi_{n1}$ is the operational notation for the conditional output density function.

*Integrator* (INT). The precipitation uncertainty, which gives rise to the conditional output density functions of $S_n$, and the hydrologic uncertainty, which gives rise to the family of conditional posterior density functions of $H_n$, are integrated according to the prescription of the Bayesian theory. For each lead time $n \, (n = 2, ..., N)$, the total probability law yields a family (for all $h_{n-1}$) of conditional predictive density functions of the actual river stage $H_n$:

$$
\begin{aligned}
p(h_n|h_{n-1}, h_0, \mathbf{u}_0, V=0) &= \int_{-\infty}^{\infty} p(h_n|s_n, h_{n-1}, h_0, V=0) p(s_n|\mathbf{u}_0, V=0)\, ds_n \\
&= \int_{-\infty}^{\infty} \phi_{n0}(h_n|s_n, h_{n-1}, h_0)\delta(s_n - s_{n0})\, ds_n \\
&= \phi_{n0}(h_n|s_{n0}, h_{n-1}, h_0) \\
&= \theta_{n0}(h_n|h_{n-1}), \qquad (5)
\end{aligned}
$$

and

$$
\begin{aligned}
p(h_n|h_{n-1}, h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0, V=1) &= \int_{-\infty}^{\infty} p(h_n|s_n, h_{n-1}, h_0, V=1) p(s_n|T_1, \boldsymbol{\xi}, \mathbf{u}_0, V=1)\, ds_n \\
&= \int_{-\infty}^{\infty} \phi_{n1}(h_n|s_n, h_{n-1}, h_0)\pi_{n1}(s_n)\, ds_n \\
&= \theta_{n1}(h_n|h_{n-1}), \qquad (6)
\end{aligned}
$$

where $\theta_{nv}(\cdot|h_{n-1})$ is the operational notation for the conditional density function, showing only the variable $(h_{n-1})$ which defines the family. For the first lead time, $n = 1$, the conditioning simplifies, and thus the operational notation is

$$p(h_1|h_0, \mathbf{u}_0, V=0) = \psi_{10}(h_1), \qquad (7a)$$

$$p(h_1|h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0, V=1) = \psi_{11}(h_1). \qquad (7b)$$

### 2.5 Forecast Structure

With respect to the stochastic process $\{H_n : n = 1, ..., N\}$ being forecasted, the INT outputs, for each hypothesized event $V = v$ $(v = 0, 1)$, the *conditional predictive one-step transition density function* $\psi_{1v}$ 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 $H_1, ..., H_N$ takes the form

$$\psi_{1v}(h_1) \prod_{n=2}^{N} \theta_{nv}(h_n|h_{n-1}). \tag{8}$$

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. One can test this fact mentally as follows: Suppose that after the forecast time 1200 UTC, precipitation did occur within 24 h $(V = 1)$, and at some step $n$–1 $(n \geq 2)$ the actual river stage $h_{n-1}$ was observed. If the river stage process, conditional on $V = 1$, were Markov of order one, then the observer would be able to specify the predictive one-step transition density function $\theta_{n1}(\cdot|h_{n-1})$ of $H_n$. In fact, this task is impossible because to specify this function, the observer must also know $(h_0, T_1, \boldsymbol{\xi}, \mathbf{u}_0)$ and the hydrologic model.

## 3. BAYESIAN ENSEMBLE FORECAST

This section reviews the fundamentals of the BEF as they were first presented by Krzysztofowicz and Maranzano (2004b, Section 9). It shows how the output from the analytic-numerical BFS can be used to generate an ensemble of realizations of the river stage process $\{H_1, ..., H_N\}$.

### 3.1 Theoretic Basis

At the forecast time, the BFS outputs the predictive joint density function $\xi$ of river stages $H_1, ..., H_N$ for any $N \geq 2$ (Krzysztofowicz and Maranzano, 2004b, Eq. (30)) in the form

$$\xi(h_1, ..., h_N) = (1-\nu)\psi_{10}(h_1)\prod_{n=2}^{N}\theta_{n0}(h_n|h_{n-1}) + \nu\psi_{11}(h_1)\prod_{n=2}^{N}\theta_{n1}(h_n|h_{n-1}). \tag{9}$$

Above, $\nu$ is the probability of precipitation occurrence during the period covered by the PQPF and over the river basin, $\psi_{1v}$ is the conditional predictive one-step transition density function for lead time $n = 1$, and $\theta_{nv}(\cdot|h_{n-1})$ is the conditional predictive one-step transition density function for lead time $n \geq 2$. According to Eq. (9), the river stage process $\{H_1, ..., H_N\}$ is a mixture of two conditional Markov processes of order one. Thus, to sample a realization of the process, one can first generate a realization of $V$, either $V = 0$, or $V = 1$, which designates the branch of the mixture, and then recursively generate realizations of the river stages from the one-step transition distribution functions corresponding to the density functions in the designated branch. Thereby, Eq. (9) provides the theoretic basis for the BEF generator.

Three important facts to note are (i) that the predictive joint density function $\xi$ is conditional on the hydrologic model being used, all deterministic inputs to that model at the forecast time (e.g., initial model states, observed river stage $h_0$), and the PQPF; (ii) that the predictive uncertainty quantified by $\xi$ about the river stage process $\{H_1, ..., H_N\}$ is nonstationary; and (iii) that under $\xi$, the river stage process $\{H_1, ..., H_N\}$ is not Markov of any order.

### 3.2 Source Elements

The primary output from the BFS, which constitutes the input into the BEF generator, is the families of the conditional predictive one-step transition distribution functions (Krzysztofowicz and Maranzano, 2004b, Eq. (45)):

$$\{\Psi_{1v} : v = 0, 1\}\,, \tag{10a}$$

$$\{\Theta_{nv}(\cdot|h_{n-1}) : all\ h_{n-1}; v = 0, 1; n = 2, ..., N\}\,. \tag{10b}$$

Also required is the probability of precipitation occurrence, $\nu$, provided by the PQPF.

### 3.3 Ensemble Size

Given the source elements, the objective is to generate an ensemble of $M$ realizations of the river stage process, which characterizes faithfully the predictive uncertainty about the process. Assuming that $0 < \nu < 1$, let $M_v$ denote the number of realizations conditional on $V = v$, with $v \in \{0, 1\}$, so that $M = M_0 + M_1$. These numbers must satisfy two requirements.

First, $M_1/M$ must be (approximately) equal to $\nu$ so that the ensemble preserves (almost) exactly the probability of precipitation occurrence. Second, suppose $M^*$ is the *smallest sample size required* for accurate estimation from the ensemble of any predictive transition distribution function that a user might employ (explicitly or implicitly) in a decision model. (The determination of $M^*$ is studied experimentally in Section 8.) Then each $M_0$ and $M_1$ must be at least as large as $M^*$.

These two requirements can be met as follows. If $\nu = 0$, then $M = M_0 = M^*$ and $M_1 = 0$. If $\nu = 1$, then $M = M_1 = M^*$ and $M_0 = 0$. If $0 < \nu < 1$, then

$$M = M^* + \text{mod}\left(\frac{\max\ \{1-\nu, \nu\}}{\min\ \{1-\nu, \nu\}}M^* + \frac{1}{2}\right), \tag{11a}$$

$$M_1 = \mathrm{mod}\left(\nu M + \frac{1}{2}\right), \tag{11b}$$

$$M_0 = M - M_1, \tag{11c}$$

where mod$(m)$ denotes the largest integer not exceeding $m$.

It follows that for a given $M^*$, the *ensemble size* $M$ and its summands, $M_1$ and $M_0$, are functions of the probability of precipitation occurrence $\nu$. Figure 1 depicts these functions for $M^* = 100$. At $\nu = 0$ and $\nu = 1$, $M = 100$. For all $0 < \nu < 1$, $M$ is a convex function of $\nu$, attaining a minimum at $\nu = 0.5$, and being unbounded, $M \to \infty$, as either $\nu \to 0$ or $\nu \to 1$.

[This theoretic fact sends a message of caution to the forecasters: to hold the computation cost (which increases with $M$) down, it is advisable not to be almost certain (e.g., $\nu = 0.001$ or $\nu = 0.999$), but rather somewhat uncertain (e.g., $\nu = 0.1$ or $\nu = 0.9$) or completely certain ($\nu = 0$ or $\nu = 1$) about the occurrence of precipitation. This later forecast violates the Cromwell's rule (never assign probability zero or one to an uncertain event because such a probability cannot be revised by Bayes theorem; Lindley, 1985, p. 104), but cloudless skies or rains already pouring may qualify as exceptions even to the most ardent probabilists.]

In summary, the uncertainty associated with the intermittence of the precipitation process has an important implication: to maintain a desired level of accuracy of the predictive transition distribution functions, of which the ensemble is a numerical representation, the ensemble size $M$ must not be constant, but must vary in the prescribed manner with $\nu$. [Of course, in real-time forecasting the ensemble size could be kept constant, provided it were equal to the largest number $M$ calculated from (11a) using a given $M^*$ and the greatest $\nu$ or $1 - \nu$ $(0 < \nu < 1)$ ever to be assigned.]

### 3.4 Ensemble Generator

Based on expression (9), and using input (10), a single realization $(h_1, ..., h_N)$ of the river stage process $\{H_1, ..., H_N\}$, conditional on $V = v$, with $v \in \{0, 1\}$, is generated according to the following algorithm.

1. Generate $N$ random numbers $(p_1, ..., p_N), 0 < p_n < 1, n = 1, ..., N$.
2. Find realization $h_1$ such that $p_1 = \Psi_{1v}(h_1)$.
3. Proceed recursively on $n$ $(n = 2, ..., N)$; given realization $h_{n-1}$, find realization $h_n$ such that $p_n = \Theta_{nv}(h_n|h_{n-1})$.

### 3.5 Example of BEF

The example uses the input data from Krzysztofowicz and Maranzano (2004b). In particular, the PQPF specifies (i) a probability of precipitation occurrence during the 24-h period and over the basin, $\nu = 0.81$; (ii) a distribution function of the 24-h basin average precipitation amount conditional on the hypothesis that precipitation occurs, a Weibull distribution with scale parameter $\alpha = 1.807$ and shape parameter $\beta = 1.378$; Figure 2 shows the resultant unconditional distribution function; and (iii) a vector of expected fractions $\boldsymbol{\xi} = (0.0, 0.1, 0.4, 0.5)$ for temporal disaggregation of any total precipitation amount into 6-h subperiods. The observed river stage is $h_0 = 7.9$ ft. The BFS outputs the family of distribution functions (10). This family is employed in the ensemble generator to obtain 100 members, each consisting of 12 river stages at 6-h time steps $(n = 1, ..., 12)$.

Figure 3 shows a spaghetti plot of the BEF, with members conditional on no rain displayed in gray and members conditional on rain displayed in black. Three features are of note: (i) the distinct nature of uncertainty under no rain and rain conditions, with little mixing of the conditional realizations; (ii) the nonstationarity of uncertainty, both conditional and total, as the ensemble spread increases with the lead time until, at the lead time long enough, it stabilizes and conforms to the marginal prior (climatic) distribution function of the river stage; (iii) the relatively sharp delineation of the minimal river stages (where the hydrologic uncertainty dominates) and the relatively fuzzy delineation of the maximal river stages (where the precipitation uncertainty dominates). This last feature suggests that a 100-member ensemble is too small to reliably estimate the distribution functions of the flood crest and the time to crest.

Figure 4 shows the predictive $n$-step transition distribution functions $\hat{\Psi}_n$ estimated from the ensemble for $n = 1, ..., 12$. Note the inflection point between rain and no rain ensemble members, and the resultant bimodality of the distribution functions. This reflects the fact that each of these distribution functions is a mixture of two functions: one conditional on rain and another conditional on no rain.

## 4. ESTIMATING PRSF FROM ENSEMBLE

For all analyses that follow, the BFS was reset to produce forecasts on a coarser time scale: $n = 1, 2, 3$ at 24-h steps, so that the lead times are 24, 48, 72 h.

The PRSF consists of a sequence of predictive $n$-step transition distribution functions $\Psi_n$ $(n = 1, 2, 3)$, each of which is a mixture (Krzysztofowicz and Maranzano, 2004b, Eq. (43)):

$$\Psi_n(h_n) = (1 - \nu)\Psi_{n0}(h_n) + \nu\Psi_{n1}(h_n). \tag{12}$$

[Function $\Psi_n$ is the $n$-step transition distribution function because it characterizes the transition from the river stage $H_0 = h_0$ observed at the forecast time to the uncertain river stage $H_n$ at the lead time of $n$ steps. Because $h_0$ is known and fixed on any particular forecasting occasion, it is not shown explicitly in (12).]

The ensemble generator was run multiple times, and each time the empirical distribution functions $\hat{\Psi}_{nv}$ were estimated from the ensemble and compared with the analytic distribution functions $\Psi_{nv}$ output from the BFS $(v = 0, 1; n = 1, 2, 3)$. The goodness of the ensemble-based estimate was measured by the *maximum absolute difference*

$$MAD = \max_{h_n} |\Psi_{nv}(h_n) - \hat{\Psi}_{nv}(h_n)|. \tag{13}$$

For $v = 0, 1$, Fig. 5 compares $\Psi_{3v}$ with three $\hat{\Psi}_{3v}$, each estimated from a different ensemble of size $M = 100$, and reports the maximum $MAD$ across the three estimates. The example demonstrates that the estimation error may be phenomenal — up to 0.20. This implies that in a particular forecasting situation, two consecutive ensembles may yield estimates of the same predictive $n$-step transition distribution function $\Psi_n$ that differ by as much as 0.40 when evaluated at some particular river stage. A user who took the PRSF at its face value, would interpret this difference as a change in uncertainty, whereas in fact the change resulted from the sampling variability.

Whereas the estimates $\hat{\Psi}_{3v}$ in Fig. 5 were selected to show the range of the sampling variability, not the "average" variability, they do illustrate the problem: The $MAD$ on the order of 0.20 is possible, which is obviously unacceptable for real-time forecasting because a forecast with such large $MAD$ deceives the users. Clearly then, the ensemble size $M = 100$ is too small to produce acceptable estimates of predictive probabilities. What the required ensemble size should be is researched in Section 7.

## 5. ESTIMATING PSTF FROM ENSEMBLE

### 5.1 Estimation Problem

The PSTF consists of the predictive distribution function $\Psi_1$ of $H_1$, conditional on the observed river stage $H_0 = h_0$ (which is the forecast for the first time step, $n = 1$), and a sequence of families of the predictive one-step transition distribution functions:

$$\{\Theta_n(\cdot|h_{n-1}, ..., h_1) : all\ h_1, ..., h_{n-1}; n = 2, ..., N\}. \tag{14}$$

These transition distribution functions are not conditional Markov of order one, but they may be constructed from $\nu$, $\Psi_{10}$, $\Psi_{11}$, and a sequence of families of the conditional predictive one-step transition distribution functions, which are conditional Markov of order one:

$$\{\Theta_{nv}(\cdot|h_{n-1}) : all\ h_{n-1}; v = 0, 1; n = 2, ..., N\}. \tag{15}$$

Therefore, to ascertain whether or not an ensemble of a particular size yields accurate estimates of all one-step transition distribution functions, it is sufficient to validate that each family of the distribution functions in (15) can be estimated accurately. This means that for each $v \in \{0, 1\}$ and $n \in \{2, ..., N\}$, and at every point $h_{n-1}$ in the domain of variate $H_{n-1}$, one must estimate the distribution function $\Theta_{nv}(\cdot|h_{n-1})$ of variate $H_n$.

An obstacle to such a validation is that there are infinitely many points $h_{n-1}$ in the domain of $H_{n-1}$ and that $P(H_{n-1} = h_{n-1}) = 0$ because $H_{n-1}$ is a continuous variate. Hence, it is impossible to estimate function $\Theta_{nv}(\cdot|h_{n-1})$ at the point $h_{n-1}$ from a finite ensemble, not to mention the infinite number of such functions, at all points $h_{n-1}$ as required for (15).

One solution to this estimation problem is to create a window about $h_{n-1}$ of a fixed size and to specify a sampling rule: if an ensemble member generated conditional on $V = v$ contains a realization of $H_{n-1}$ that falls within the window, then this member is included in the sample from which $\Theta_{nv}(\cdot|h_{n-1})$ is estimated. Formally, let $d_{nv}$ be the size of the window about $h_{n-1}$ for lead time $n$ and $V = v$, and let $h_{n-1,k}$ be the realization of $H_{n-1}$ in the $k^{th}$ ensemble member, conditional on $V = v$. If $h_{n-1,k} \in [h_{n-1} - d_{nv}, h_{n-1} + d_{nv}]$, then the $k^{th}$ ensemble member is included in the sample. The selection of $d_{nv}$ is described next.

### 5.2 Sampling Window

The empirical distribution function $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ estimated from the ensemble was compared with the analytic distribution function $\Theta_{nv}(\cdot|h_{n-1})$ output from the BFS in terms of

$$MAD = \max_{h_n} |\Theta_{nv}(h_n|h_{n-1}) - \hat{\Theta}_{nv}(h_n|h_{n-1})|, \tag{16}$$

for a given $h_{n-1}$, and a fixed $n$ and $v$. The window size $d_{nv}$ was selected to minimize the average $MAD$ for the estimate $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ across points $h_{n-1} = 5, 6, 7, 8, 9, 10, 11, 12$ ft. The following values for $d_{nv}$ were considered: $0.5$, $0.4$, $0.3$, $0.2$, $0.1$, $0.05$ ft. Table 2 presents the results for the ensemble sizes used in the experiment. It is clear that the window size minimizing the average $MAD$ is, in general, larger for $v = 1$ than for $v = 0$. In fact, it is usually as large as $0.5$ ft. For $v = 0$ and $n = 3$, the optimal window size becomes smaller as the ensemble size increases, down to a minimum of $0.2$ ft. This pattern is broken for large values of $h_2$ and small ensemble sizes $M$, but this is an artifact of small samples collected by the windows regardless of the value of $d_{30}$. For $v = 0$ and $n = 2$, only $h_1 = 6$ ft and $h_1 = 7$ ft yield any ensemble members. For $h_1 = 7$, the average sample size is too small to draw any conclusion. However, for $h_1 = 6$, all samples include most of the ensemble members for every value of $d_{20}$, and it is this case where the tendency for the optimal value of $d_{nv}$ to decrease with ensemble size is the most prevalent.

### 5.3 Explanation of Behavior

The optimal window sizes $d_{nv}$ being smaller for $v = 0$ than for $v = 1$ may be explained by examining the impact of the window size on the estimate of the conditional distribution function. Suppose the $k^{th}$ ensemble member, generated conditional on $V = v$, has river stage $h_{n-1,k}$ at time $n-1$. This member is included in the estimation sample for $\Theta_{nv}(\cdot|h_{n-1})$ if $h_{n-1} - d_{nv} \leq h_{n-1,k} \leq h_{n-1} + d_{nv}$. This implies that the acquired estimation sample is actually generated from a mixture of conditional distribution functions that belong to the family:

$$\left\{ \Theta_{nv}(\cdot|h^*_{n-1}) : all\ h^*_{n-1} \in [h_{n-1} - d_{nv}, h_{n-1} + d_{nv}] \right\}. \tag{17}$$

Hence, the greater the diversity of this family, the worse the empirical estimate $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ obtained from the ensemble is. For our purposes, the diversity is measured by the $MAD$ between the members of the family and the distribution we wish to estimate, $\Theta_{nv}(\cdot|h_{n-1})$.

Figure 6 shows select members of the families when the task is to estimate $\Theta_{20}(\cdot|h_1 = 6)$ and $\Theta_{21}(\cdot|h_1 = 8)$ from samples acquired with $d_{20} = 0.1$ ft and $d_{21} = 0.5$ ft, respectively. Clearly, the first family is more diverse than the second family, as measured by the $MAD$. Hence, one may suspect that the optimal window size $d_{2v}$ is smaller for $\Theta_{20}(\cdot|h_1 = 6)$ than for $\Theta_{21}(\cdot|h_1 = 8)$, and this is indeed the case as shown in Table 2.

### 5.4 Choice of Window Size

In the remainder of the analyses reported herein, the following window sizes are used: $d_{20} = 0.1$, $d_{30} = 0.3$, $d_{21} = 0.5$, and $d_{31} = 0.5$ ft. These sizes were selected in order to minimize the average $MAD$ in the experiment that should yield the smallest average $MAD$, which is when $M = 10000$.

## 6. ESTIMATING PFF FROM ENSEMBLE

Let $t_0$ denote the forecast time, $t_n$ denote the time instance at which $H_n$ is observed, and let $Z_n$ denote the maximum river stage within time interval $(t_0, t_n]$, practically $Z_n = \max\{H_1, ..., H_n\}$. The PFF consists of a sequence of predictive distribution functions $F_n$ $(n = 1, 2, 3)$ such that $F_n(h) = P(Z_n \leq h)$ for all $h$. Each of these distribution functions is a mixture

$$F_n(h) = (1 - \nu)F_{n0}(h) + \nu F_{n1}(h), \tag{18}$$

wherein $F_{nv}(h) = P(Z_n \leq h | V = v)$ for $v = 0, 1$. For $n = 1$, obviously $F_1 = \Psi_1$, which is defined by Eq. (12). For $n = 2, 3$, function $F_n$ is derived from Eq. (9), as shown by Krzysztofowicz (2008). It follows that $F_n$ is conditioned on the same elements on which the predictive joint density function of river stages $H_1, ..., H_n$ is conditioned, as explained in Section 2. In essence, the PFF is a byproduct of the PSTF.

The ensemble generator was run multiple times, and each time the empirical distribution functions $\hat{F}_{nv}$ were estimated from the ensemble and compared with the analytic distribution functions $F_{nv}$ output from the BFS $(v = 0, 1; n = 2, 3)$. The goodness of the ensemble-based estimate was measured by the *maximum absolute difference*

$$MAD = \max_h |F_{nv}(h) - \hat{F}_{nv}(h)|. \tag{19}$$

## 7. EXPERIMENTAL DESIGN

Define $M^*$ to be the smallest ensemble size required for the BEF-based empirical predictive distributions to accurately approximate the BFS analytical predictive distributions. Because the BFS output is used to generate the ensemble members, the convergence of the BEF-based empirical distributions to the BFS analytical distributions is assured as the ensemble size increases. The objective of the experiments is to determine $M^*$.

The example presented in Section 3.5 served as the source of data for the experiment at Eldred. Parallel sources of data were used for the experiments at Dailey, Philippi, and Parsons. Aside drainage areas and hydrologic models (Table 1), the main difference in these experiments was the PQPF (Fig. 2): if the incoming storm passed over the river basin (probability $\nu = 0.81$), then it would produce almost certainly $1.0$–$2.5$ in of basin average precipitation amount.

At each forecast point, the BEF generator was run under two scenarios, $v = 0$ and $v = 1$, and for $M_v = 50, 100, 150, 200, 500, 750, 1000, 1500, 2000, 5000, 7500, 10000$. For each combination of $v$ and $M_v$, the generator was executed $500$ times. After every execution, each distribution function in the following families was estimated from the ensemble and compared in terms of the $MAD$ to the corresponding analytical distribution function:

$$\{\Psi_{nv} : v = 0, 1; n = 1, 2, 3\}\,, \tag{20}$$

$$\{\Theta_{nv}(\cdot|h_{n-1}) : all\ h_{n-1}; v = 0, 1; n = 2, 3\}, \tag{21}$$

$$\{F_{nv} : v = 0, 1; n = 2, 3\}. \tag{22}$$

Then the expected $MAD$ (across 500 ensembles) was calculated to measure the accuracy of the BEF-based estimate relative to the BFS analytical distribution. Finally, the relationship between the expected $MAD$ and the ensemble size $M_v$ for every $v, n,$ and forecast point was analyzed to infer the smallest ensemble size $M^*$ required to achieve a specified level of accuracy.

# 8. EXPERIMENTAL RESULTS

## 8.1 Results for Predictive n-Step Transition Distributions

The results turned out to be nearly identical across the four forecast points, as well as independent of the precipitation event $(v = 0, 1)$ and the lead time $(n = 1, 2, 3)$. To convey these three invariance properties, Table 3 reports the expected $MAD$s for selected ensemble sizes $M_v$. The largest difference in the corresponding $MAD$s across the forecast points is $0.005$; the largest difference in the corresponding $MAD$s across the precipitation events $(v = 0, 1)$ and across the lead times $(n = 1, 2, 3)$ is $0.003$. These largest differences occur when the ensemble size is the smallest, $M_v = 50$; and $0.005$ is less than $14\%$ of the standard deviation of the $MAD$ due to sampling (sample size of $500$); viz, all the differences are statistically insignificant and visually imperceptible.

With the nearly identical results, the average of the 24 expected $MAD$s was calculated (across four forecast points, two precipitation events, three lead times) and then graphed against the ensemble size $M$. Figure 7 shows the graph.

We conclude that within the scope of the experimental factors (four river basins of 484–2372 km$^2$, two hydrologic models, two PQPFs, rain and no rain events, and three lead times up to 72 h), there exists a general graph of the expected $MAD$ (between the predictive $n$-step transition distribution functions $\hat{\Psi}_n$ and $\Psi_n$) against the ensemble size $M$. This graph (Fig. 7) allows one to select the smallest ensemble size $M^*$ required to maintain a particular expected $MAD$. For example, if one is willing to allow the expected $MAD$ of $0.06$, then the ensemble size of about $200$ members is required. If the allowable expected $MAD$ is decreased to $0.04$, then the ensemble size must be about $500$; decreasing the allowable expected $MAD$ further to $0.02$, requires the ensemble size of about $2000$.

## 8.2 Results for Predictive One-Step Transition Distributions

### *8.2.1 One Forecast Point*

The results are much more voluminous than the previous ones because of the additional factor — the preceding river stage $h_{n-1}$, which conditions every predictive one-step transition distribution function $\Theta_{nv}(\cdot|h_{n-1})$ for $v = 0, 1$ and $n = 2, 3$. The number of different stages $h_{n-1}$ considered is $8$ at Eldred and Dailey and 9 at Philippi and Parsons.

To gain insight into the nature of the results, the case of $(v = 1, n = 3, d_{31} = 0.5$ ft$)$ at Eldred is examined in detail. Table 4 reports the expected $MAD$ for every ensemble size $M_1$ and for $h_2 = 6, 8, 10, 12$ ft. (The case with $h_2 = 10$ ft is plotted in Fig. 8.) These results imply that the smallest ensemble size $M^*$ required to ensure a particular expected $MAD$ depends on the conditioning river stage $h_2$. For instance, to ensure the expected $MAD$ less than $0.06$ for $\hat{\Theta}_{31}(\cdot|H_2 = 6)$ one needs $M^* = 7500$, but for $\hat{\Theta}_{31}(\cdot|H_2 = 10)$, one needs only $M^* = 1500$. The difference can be explained by the predictive probabilities of the windows that "catch" the ensemble members into sub-samples from which $\Theta_{31}(\cdot|h_2)$ are estimated; apparently $P(H_2 \in [5.5, 6.5]) < P(H_2 \in [9.5, 10.5])$.

Table 4 reports also the standard deviation of each $MAD$. For a fixed ensemble size $M_1$, the smallest standard deviation is associated with the smallest expected $MAD$, which is at $h_2 = 10$ ft. Not surprisingly, the sampling window (with $d_{31} = 0.5$ ft) about this $h_2$ catches, on average, the largest number of ensemble members that form the sub-sample from which an empirical estimate $\hat{\Theta}_{31}(\cdot|h_2)$ is obtained (Table 5).

The above pattern of results prevails (with only a few minor perturbations) for all $v \in \{0, 1\}$, $n = \{2, 3\}$, and forecast points. Therefore, to reduce the volume of numbers, only one column from each of the sixteen tables (like Table 4) is reported; the one column is for the conditioning river stage $h_{n-1}$ at which the standard deviation of the $MAD$ is the smallest (across all, or most, of the twelve ensemble sizes $M_v$). Table 6 lists these stages. The implication is that the results reported henceforth are "the best cases" (across all $h_{n-1}$ stages for each $v, n$, and forecast point). Consequently, any required ensemble size inferred from these results should be viewed as a lower bound on $M^*$.

*8.2.2 Four Forecast Points*

With the conditioning river stages $h_{n-1}$ fixed (Table 6), the results are reported in Table 7, which has the same format as Table 3 to facilitate comparisons. The table reports the expected $MAD$ for selected ensemble sizes $M_v$, for both precipitation events $(v = 0, 1)$, both lead times $(n = 2, 3)$, and the four forecast points. The largest difference in the corresponding $MAD$s is $0.133$ across the forecast points, $0.096$ across precipitation events, and $0.123$ across the lead times. These differences exceed the corresponding standard deviations of the $MAD$ due to sampling (sample size of $500$). Clearly, the expected $MAD$s in Table 7 differ significantly from those in Table 3 in terms of both magnitude (at least twice as large on average) and variability (significantly greater).

For each forecast point, a graph of the expected $MAD$ against the ensemble size $M_v$ was made for $v = 0, 1$ and $n = 2, 3$. The most diverse graphs were those at Eldred and Philippi; they are shown in Figs. 8 and 9, respectively. A comparison of Figs. 8 and 9 with Fig. 7 makes it clear that the ensemble size required to maintain the expected $MAD$ of a particular value is significantly larger for a predictive one-step transition distribution than for a predictive $n$-step transition distribution. For example, consider Fig. 9 for Philippi: to ensure the expected $MAD$ between $0.06$ and $0.04$ when it rains $(v = 1)$ requires $4000$–$8000$ ensemble members for $n = 2$, and $5000$–$10000$ ensemble members for $n = 3$; the expected $MAD$ of $0.02$ is reachable with about $7000$ members in only one case $(v = 0, n = 2)$ out of four.

These results are not surprising because only a fraction of the ensemble can be used in the estimation of $\Theta_{nv}(\cdot|h_{n-1})$ for any particular $h_{n-1}$. Furthermore, the required ensemble size when $v = 1$ is different than that when $v = 0$. This is explained by the diversity of the family of the distribution functions making up the estimates (see Section 5.3). Also, the expected $MAD$ when $v = 0$ and $n = 3$ is never below $0.04$ regardless of the ensemble size. This is explained by the absence of precipitation uncertainty in the forecast for day 1, which results in a very steep (concentrated) predictive one-step transition distribution functions. Such functions are notoriously difficult to estimate because of the inherent tradeoff between the size of the sampling window (which one wants to minimize) and the expected number of ensemble members caught by the window (which one wants to maximize).

8.*2.3 Stable Sampling Requirement*

The example for Eldred (Table 5) reveals that while the smallest ensemble size $M^*$ required for the expected $MAD$ to be below $0.06$ varies with the conditioning river stage $h_2$, the expected number of ensemble members that form the sub-sample from which $\Theta_{31}(\cdot|h_2)$ is estimated remains fairly stable. It is less than $277, 228, 202, 214$ for $h_2 = 6, 8, 10, 12$ ft respectively.

The generality of this observation is confirmed in Table 8, which shows the expected number of ensemble members that must fall within the sampling window about $h_{n-1}$ in order to obtain an empirical estimate $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ with the expected $MAD$ less than $0.06$, for $v = 0, 1$, and $n = 2, 3$, for all stages $h_{n-1}$ and all forecast points. The grand average is about $210$ ensemble members per sampling window $[h_{n-1} - d_{nv}, h_{n-1} + d_{nv}]$.

The implication is immediate: If most of the variability of $H_{n-1}$ occurs within $L$ sampling windows (of optimal size, not arbitrary size), then a lower bound on the required ensemble size is $M^* = 210 \times L$. Thus, even with a crude discretization of the sample space of the river stage into

$L = 10$ windows of 1 ft width, the minimum required ensemble size is about $M^* = 2100$.

*8.2.4 General Sampling Requirement*

The above analyses suggest a general conclusion: in order that an ensemble forecast represents reliably the predictive one-step transition distribution functions, say, with the expected $MAD$ of less than $0.06$, the required ensemble size $M^*$ is on the order of $1000$–$5000$. Decreasing the expected $MAD$ threshold to $0.04$, increases the range of $M^*$ to $2000$–$10000$.

## 8.3 Results for Maximum Stage Distributions

The results turned out to be nearly identical across the four forecast points, with a tendency for the expected $MAD$ to increase slightly with the precipitation occurrence index $(v = 0, 1)$ and the lead time $(n = 2, 3)$. The average, the minimum, and the maximum of the expected $MAD$s were calculated (across four forecast points, two precipitation events, two lead times) and then graphed against the ensemble size $M$. Figure 10 shows the graph.

We conclude that within the scope of the experimental factors, there exists a general graph of the expected $MAD$ (between the predictive distribution functions $\hat{F}_n$ and $F_n$ of the flood crest) against the ensemble size $M$. This graph (Fig. 10) is nearly identical with the graph of the expected $MAD$ for the predictive $n$-step transition distribution functions $\hat{\Psi}_n$ and $\Psi_n$ (Fig. 7).

# 9. SUMMARY AND CONCLUSION

## 9.1 Bayesian Ensemble Forecast

The BEF, introduced first in concept by Krzysztofowicz (1999, 2001), can be produced in two ways, either by attaching an ensemble generator to an analytic-numerical BFS (Krzysztofowicz and Maranzano, 2004b) or by implementing the BFS entirely using Monte Carlo simulation (Krzysztofowicz, 2001). Herein, an analytic-numerical BFS for short-term PSTF was coupled with an ensemble generator to produce the first BEF of the river stage process. This BEF quantifies the total uncertainty, which represents the integration of precipitation uncertainty and hydrologic uncertainty. In particular, this BEF characterizes the stochastic dependence in the future river stage process, which is nonstationary and complex, especially in the presence of uncertainty about the occurrence of precipitation: the predictive $n$-step transition distribution functions are bimodal; and the predictive one-step transition distribution functions depend on all preceding river stages, beginning with the river stage observed at the forecast time.

The analytic-numerical BFS offers a unique advantage for research purposes because it allows one to obtain each predictive distribution function via two methods of computation: (i) the analytic-numerical integration, which supplies the exact solution, and (ii) the empirical estimation from a sample provided by the ensemble forecast, which supplies an approximate solution whose expected accuracy depends on the ensemble size. A comparative analysis of the two solutions for different ensemble sizes offers a means of determining the optimal ensemble size for a particular purpose.

## 9.2 Optimal Ensemble Size

Through experimentation, it was discovered that to ensure reasonable accuracy of the empirical distribution functions estimated from an ensemble, hundreds or possibly thousands of ensemble members are needed. These results are pertinent to any ensemble forecasting system. And they serve as a caution to system developers, reminding us that (i) the ensemble size is an important design variable, and (ii) the optimal ensemble size $M^*$ depends upon the purpose of the forecast.

For those purposes that require forecasts of the river stages at time instances (e.g., barge operation), or forecasts of the maximum river stages within time intervals (e.g., flood forecasts, public warnings), it is sufficient to quantify the total uncertainty in terms of the predictive $n$-step transition distribution functions (essentially the conditional marginal distribution functions of variates $H_1, ..., H_N$ or of their sequential maxima). The ensemble size $M^*$ required for accurate estimation of these univariate distribution functions is on the order of several hundreds. This requirement on $M^*$ is basic and possibly general because it is independent of the factors varied in the present experiments. It is depicted by the two nearly-identical graphs in Figs. 7 and 10, from which one can read $M^*$ as a function of the expected $MAD$. Table 9 offers a brief guide.

For those purposes that require forecasts of the river flow process (e.g., forecasts of reservoir inflows, control of storage reservoirs, or operation of waterways), it is necessary to quantify the total uncertainty not only about the river flows at time instances, but also about the evolution of the river flow from one time instance to the next. For this purpose, the families of the predictive one-step transition distribution functions for the process $\{H_1, ..., H_N\}$ are necessary. To ensure that every conditional distribution function in each of the families is represented adequately in the ensemble forecast, the ensemble size $M^*$ required may be on the order of several thousands. This requirement on $M^*$ is more stringent and variable with factors such as the probability of precipitation occurrence, the distribution function of precipitation amount, the lead time, the river basin, and possibly others. The range of the variability of $M^*$ as a function of the expected $MAD$ is conveyed by the eight graphs in Figs. 8 and 9. Table 9 offers a brief guide. At the minimum, this more stringent requirement on $M^*$ is about three times larger than the basic requirement; but the multiplier may be as high as 20.

### 9.3 Future Research

While this study points out the importance of the ensemble size as a design variable, it is confined to four case studies. Many more case studies in river basins of different sizes and characteristics are needed to establish general design guidelines.

The ensemble size requirements identified herein exceed substantially the size of the largest meteorological ensembles (of precipitation and temperature) available currently in real-time forecasting. Therefore, a more efficient ensemble generator for hydrological forecasting is desirable in order to generate the hundreds or the thousands of members operationally in a reasonable amount of time. To this end, the theory of the ensemble BFS (Krzysztofowicz, 2001) could be explored and implemented.

**ACKNOWLEDGMENT**

This material is based upon work supported by the National Science Foundation under Grant No. ATM-0641572, “New Statistical Techniques for Probabilistic Weather Forecasting”.

# APPENDIX: GLOSSARY OF MATHEMATICAL SYMBOLS

## A.1 Variables and Parameters

| | |
|---|---|
| $d_{nv}$ | size of the window about $h_{n-1}$, conditional on $V = v$ |
| $h_0$ | observed river stage at time $t_0$ |
| $M$ | ensemble size, number of ensemble members (realizations) |
| $M^*$ | the smallest ensemble size required for accurate estimation of predictive distribution functions, optimal ensemble size |
| $M_v$ | number of ensemble members, conditional on $V = v$ |
| $MAD$ | maximum absolute difference between two distribution functions |
| $n$ | index of time steps, index of lead times |
| $N$ | last time step, last lead time |
| $p_n$ | probability number |
| $s_{n0}$ | model river stage at time $t_n$ resulting from $W = 0$ |
| $t_0$ | forecast time |
| $t_n$ | time instance |
| $\mathbf{u}_0$ | vector of deterministic inputs to the hydrologic model at time $t_0$ |
| $\alpha$ | scale parameter of a Weibull distribution |
| $\beta$ | shape parameter of a Weibull distribution |
| $\nu$ | probability of precipitation occurrence |
| $\boldsymbol{\xi}$ | matrix of expected disaggregation factors of $W$, conditional on $V = 1$; vector of expected fractions of $W$, conditional on $V = 1$ |

## A.2 Variates and Realizations

| | | |
|---|---|---|
| $H_n$, | $h_n$ | actual river stage at time $t_n$ |
| | $h^*_{n-1}$ | realization of $H_{n-1}$ falling within a specified window |

$S_n$, $s_n$ model river stage at time $t_n$

$V$, $v$ indicator of precipitation occurrence

$W$, $w$ basin average precipitation amount

$Z_n$, $h$ maximum river stage within time interval $(t_0, t_n]$

### A.3 Distribution and Density Functions

$D.F.$, $d.f.$ distribution function, density function

$F_n$ predictive $D.F.$ of $Z_n$

$F_{nv}$ conditional predictive $D.F.$ of $Z_n$

$\hat{F}_{nv}$ empirical, ensemble-based estimate of $F_{nv}$

$P$ generic probability function

$p$ generic density function

$T_1$ conditional $D.F.$ of $W$

$\delta$ Dirac function

$\Theta_n$ predictive one-step transition $D.F.$

$\Theta_{nv}$, $\theta_{nv}$ conditional predictive one-step transition $D.F.$, $d.f.$

$\hat{\Theta}_{nv}$ empirical, ensemble-based estimate of $\Theta_{nv}$

$\xi$ predictive joint $d.f.$ of $(H_1, ..., H_N)$

$\pi_{n1}$ conditional output $d.f.$ of $S_n$

$\phi_{nv}$ conditional posterior $d.f.$ of $H_n$

$\Psi_n$ predictive $n$-step transition $D.F.$

$\Psi_{nv}$, $\psi_{nv}$ conditional predictive $n$-step transition $D.F.$, $d.f.$

$\hat{\Psi}_{nv}$ empirical, ensemble-based estimate of $\Psi_{nv}$

## REFERENCES


Burnash, R.J.C., 1995. The NWS river forecast system: catchment modeling. In Singh, V.P., (Ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, Colorado, Chapter 10, pp. 311–366.

Kelly, K.S., Krzysztofowicz, R., 2000. Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36**(9), 2643–2653.

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

Krzysztofowicz, R., 2001. Reply. *Water Resources Research* **37**(2), 441–442.

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

Krzysztofowicz, R., 2008. Probabilistic flood forecast: exact and approximate predictive distributions. Research Paper RK–0802, University of Virginia, September 2008.

Krzysztofowicz, R., Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36**(11), 3265–3277.

Krzysztofowicz, R., Maranzano, C.J., 2004a. Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology* **293**(1–4), 57–73.

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

Krzysztofowicz, R., Pomroy, T.A., 1997. Disaggregative invariance of daily precipitation. *Journal of Applied Meteorology* **36**(6), 721–734.

Krzysztofowicz, R., Sigrest, A.A., 1997. Local climatic guidance for probabilistic quantitative precipitation forecasting. *Monthly Weather Review* **125**(3), 305–316.

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

Krzysztofowicz, R., Drzal, W.J., Drake, T.R., Weyman, J.C., Giordano, L.A., 1993. Probabilistic quantitative precipitation forecasts for river basins. *Weather and Forecasting* **8**(4), 424–439.

Lindley, D.V., 1985. *Decision Making*. Wiley, New York.

Maranzano, C.J., Krzysztofowicz, R., 2004. Identification of likelihood and prior dependence structures for hydrologic uncertainty processor. *Journal of Hydrology* **290**(1–2), 1–21.

Sittner, W.T., Schauss, C.E., Monro, J.C., 1969. Continuous hydrograph synthesis with an API-type hydrologic model. *Water Resources Research* **5**(5), 1007–1022.

Table 1. Forecast points, river basins, and hydrologic models.

| Forecast point | River | State[a] | Drainage area | | Hydrologic model | Number of sub-basins |
|---|---|---|---|---|---|---|
| | | | sq. miles | km$^2$ | | |
| Eldred | Allegheny | PA | 550 | 1430 | API-based[b] | 1 |
| Dailey | Tygart Valley | WV | 187 | 484 | Sacramento | 1 |
| Philippi | Tygart Valley | WV | 916 | 2372 | Sacramento | 5 |
| Parsons | Cheat | WV | 718 | 1859 | Sacramento | 5 |

[a]/PA, Pennsylvania; WV, West Virginia.

[b]/API, antecedent precipitation index.

Table 2. An analysis of the sampling window size $d_{nv}$ $(v = 0, 1; n = 2, 3)$ for estimating the conditional predictive one-step transition distribution functions $\Theta_{nv}(\cdot|h_{n-1})$ from an ensemble. A cell entry is $10 \cdot d_{nv}$ [ft], which minimizes the expected $MAD$ defined by Eq. (16); Eldred.

| | | $n = 2$ | | | | | | | | $n = 3$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $h_1$ $[ft]$ | | | | | | | | $h_2$ $[ft]$ | | | | | | | |
| | $M$ | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $v = 0$ | 50 | – | 4 | 5 | – | – | – | – | – | 4 | 5 | 5 | 2 | 3 | 5 | .5 | .5 |
| | 100 | – | 3 | 4 | – | – | – | – | – | 4 | 5 | 5 | 5 | 4 | 1 | 5 | 2 |
| | 150 | – | 3 | 4 | – | – | – | – | – | 4 | 5 | 5 | 5 | .5 | .5 | .5 | .5 |
| | 200 | – | 2 | 5 | – | – | – | – | – | 4 | 5 | 5 | 5 | 3 | .5 | .5 | .5 |
| | 500 | – | 2 | 4 | – | – | – | – | – | 3 | 5 | 5 | 5 | 5 | 1 | .5 | .5 |
| | 750 | – | 2 | 5 | – | – | – | – | – | 3 | 5 | 5 | 5 | 5 | 1 | 5 | 5 |
| | 1000 | – | 1 | 3 | – | – | – | – | – | 3 | 4 | 4 | 5 | 5 | .5 | .5 | .5 |
| | 1500 | – | 1 | 4 | – | – | – | – | – | 2 | 4 | 4 | 5 | 5 | 5 | 1 | .5 |
| | 2000 | – | 1 | 3 | – | – | – | – | – | 2 | 4 | 4 | 5 | 5 | 5 | 1 | .5 |
| | 5000 | – | 1 | 3 | – | – | – | – | – | 2 | 3 | 3 | 4 | 5 | 5 | .5 | .5 |
| | 7500 | – | 1 | 3 | – | – | – | – | – | 2 | 3 | 3 | 4 | 5 | 5 | 5 | .5 |
| | 10000 | – | 1 | 5 | – | – | – | – | – | 2 | 3 | 3 | 4 | 5 | 5 | 5 | .5 |
| $v = 1$ | 50 | 2 | 5 | 5 | 5 | 5 | .5 | 1 | 3 | .5 | .5 | .5 | 5 | .5 | 5 | 5 | 5 |
| | 100 | 2 | 5 | 5 | 5 | 5 | 5 | 1 | .5 | 1 | .5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 150 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | .5 | 1 | .5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 200 | 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 500 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 750 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 1000 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | .5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 1500 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 2000 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 5000 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 7500 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 10000 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

Table 3. The expected (across 500 samples) $MAD$ between $\hat{\Psi}_{nv}$ and $\Psi_{nv}$ as a function of the ensemble size $M_v$, precipitation indicator $v$, and lead time $n$, at four forecast points; the last column reports the average (across four forecast points) standard deviation of $MAD$.

| $v$ | $n$ | $M_v$ | Eldred | Dailey | Philippi | Parsons | Std. Dev. |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 50 | 0.114 | 0.110 | 0.112 | 0.112 | 0.0365 |
| | | 200 | 0.058 | 0.057 | 0.059 | 0.059 | 0.0185 |
| | | 2000 | 0.019 | 0.019 | 0.019 | 0.019 | 0.0058 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.011 | 0.0030 |
| | 2 | 50 | 0.112 | 0.111 | 0.110 | 0.110 | 0.0369 |
| | | 200 | 0.058 | 0.059 | 0.059 | 0.059 | 0.0185 |
| | | 2000 | 0.019 | 0.019 | 0.019 | 0.019 | 0.0057 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.010 | 0.0030 |
| | 3 | 50 | 0.112 | 0.108 | 0.111 | 0.110 | 0.0362 |
| | | 200 | 0.058 | 0.059 | 0.059 | 0.058 | 0.0181 |
| | | 2000 | 0.019 | 0.019 | 0.020 | 0.019 | 0.0060 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.010 | 0.0029 |
| 1 | 1 | 50 | 0.111 | 0.110 | 0.111 | 0.112 | 0.0383 |
| | | 200 | 0.058 | 0.057 | 0.059 | 0.060 | 0.0183 |
| | | 2000 | 0.019 | 0.019 | 0.019 | 0.019 | 0.0057 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.010 | 0.0030 |
| | 2 | 50 | 0.109 | 0.112 | 0.112 | 0.114 | 0.0367 |
| | | 200 | 0.058 | 0.059 | 0.059 | 0.058 | 0.0188 |
| | | 2000 | 0.019 | 0.019 | 0.020 | 0.019 | 0.0058 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.010 | 0.0031 |
| | 3 | 50 | 0.111 | 0.109 | 0.111 | 0.111 | 0.0380 |
| | | 200 | 0.058 | 0.059 | 0.058 | 0.059 | 0.0183 |
| | | 2000 | 0.019 | 0.019 | 0.019 | 0.019 | 0.0059 |
| | | 7500 | 0.010 | 0.010 | 0.010 | 0.010 | 0.0031 |

Table 4. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{31}(\cdot|h_2)$ and $\Theta_{31}(\cdot|h_2)$ as a function of the ensemble size $M_1$ and the conditioning river stage $h_2$; and the corresponding standard deviation of $MAD$; Eldred.

| | | $h_2$ [ft] | | | |
|---|---|---|---|---|---|
| Statistic | $M_1$ | 6 | 8 | 10 | 12 |
| Expectation | 50 | 0.265 | 0.254 | 0.247 | 0.249 |
| | 100 | 0.268 | 0.209 | 0.192 | 0.210 |
| | 150 | 0.250 | 0.180 | 0.164 | 0.180 |
| | 200 | 0.241 | 0.153 | 0.143 | 0.157 |
| | 500 | 0.176 | 0.103 | 0.096 | 0.110 |
| | 750 | 0.147 | 0.088 | 0.081 | 0.088 |
| | 1000 | 0.131 | 0.078 | 0.073 | 0.078 |
| | 1500 | 0.108 | 0.062 | 0.058* | 0.065 |
| | 2000 | 0.096 | 0.057* | 0.051* | 0.057* |
| | 5000 | 0.067 | 0.035* | 0.032* | 0.036* |
| | 7500 | 0.058* | 0.030* | 0.027* | 0.030* |
| | 10000 | 0.053* | 0.025* | 0.023* | 0.026* |
| Std. dev. | 50 | 0.137 | 0.108 | 0.106 | 0.116 |
| | 100 | 0.127 | 0.082 | 0.072 | 0.084 |
| | 150 | 0.109 | 0.068 | 0.063 | 0.068 |
| | 200 | 0.101 | 0.056 | 0.053 | 0.058 |
| | 500 | 0.067 | 0.036 | 0.033 | 0.038 |
| | 750 | 0.054 | 0.030 | 0.028 | 0.028 |
| | 1000 | 0.047 | 0.026 | 0.024 | 0.026 |
| | 1500 | 0.038 | 0.020 | 0.018 | 0.020 |
| | 2000 | 0.034 | 0.017 | 0.017 | 0.018 |
| | 5000 | 0.022 | 0.010 | 0.010 | 0.011 |
| | 7500 | 0.020 | 0.009 | 0.008 | 0.009 |
| | 10000 | 0.018 | 0.007 | 0.006 | 0.008 |

Asterisks indicate values less than 0.06.

Table 5. The expected (across 500 samples) number of ensemble members falling into the sampling window ($d_{31} = 0.5$ ft) about $h_2$, and forming the sub-sample from which an empirical estimate $\hat{\Theta}_{31}(\cdot|h_2)$ of the conditional predictive one-step transition distribution function is obtained, as a function of the ensemble size $M_1$ and the conditioning river stage $h_2$; Eldred.

| | $h_2$ [ft] | | | |
|---|---|---|---|---|
| $M_1$ | 6 | 8 | 10 | 12 |
| 50 | 2 | 6 | 7 | 5 |
| 100 | 4 | 11 | 14 | 11 |
| 150 | 6 | 17 | 20 | 16 |
| 200 | 7 | 23 | 27 | 21 |
| 500 | 18 | 56 | 68 | 53 |
| 750 | 28 | 85 | 101 | 80 |
| 1000 | 37 | 114 | 135 | 106 |
| 1500 | 55 | 171 | 202* | 161 |
| 2000 | 74 | 228* | 270* | 214* |
| 5000 | 185 | 569* | 676* | 533* |
| 7500 | 277* | 851* | 1011* | 800* |
| 10000 | 368* | 1135* | 1348* | 1066* |

Asterisks indicate sample sizes that yield the average $MAD$ less than $0.06$ (see Table 4).

Table 6. The initial (observed) river stage $h_0$, and the river stages $h_{n-1}$ $(n = 2, 3)$ that condition the predictive one-step transition distribution functions $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ compared in terms of the expected $MAD$ in Table 7; for these $h_{n-1}$ values, the standard deviation of $MAD$ is the smallest across (almost) all of the ensemble sizes.

| | | $h_{n-1}$ [ft] | | | |
|---|---|---|---|---|---|
| $v$ | $n$ | Eldred | Dailey | Philippi | Parsons |
| 0 | 1 | 8 | 3 | 6 | 5 |
| | 2 | 6 | 3 | 8 | 5 |
| | 3 | 6 | 3 | 8 | 4 |
| 1 | 1 | 8 | 3 | 6 | 5 |
| | 2 | 7 | 8 | 14 | 10 |
| | 3 | 10 | 6 | 16 | 7 |

Table 7. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ as a function of the ensemble size $M_v$, precipitation indicator $v$, and lead time $n$, at four forecast points; stages $h_{n-1}$ are listed in Table 6; results for Eldred and Philippi are graphed in Figs. 8 and 9; the last column reports the average (across four forecast points) standard deviation of $MAD$.

| $v$ | $n$ | $M_v$ | Eldred | Dailey | Philippi | Parsons | Std. Dev. |
|---|---|---|---|---|---|---|---|
| 0 | 2 | 50 | 0.209 | 0.221 | 0.187 | 0.246 | 0.0920 |
| | | 200 | 0.120 | 0.135 | 0.102 | 0.140 | 0.0442 |
| | | 2000 | 0.046 | 0.047 | 0.035 | 0.049 | 0.0145 |
| | | 7500 | 0.030 | 0.025 | 0.018 | 0.025 | 0.0077 |
| | 3 | 50 | 0.184 | 0.211 | 0.274 | 0.163 | 0.0852 |
| | | 200 | 0.101 | 0.119 | 0.225 | 0.092 | 0.0486 |
| | | 2000 | 0.036 | 0.041 | 0.086 | 0.033 | 0.0154 |
| | | 7500 | 0.021 | 0.022 | 0.047 | 0.021 | 0.0084 |
| 1 | 2 | 50 | 0.192 | 0.240 | 0.274 | 0.245 | 0.1014 |
| | | 200 | 0.109 | 0.143 | 0.198 | 0.138 | 0.0541 |
| | | 2000 | 0.037 | 0.050 | 0.073 | 0.049 | 0.0163 |
| | | 7500 | 0.019 | 0.026 | 0.040 | 0.026 | 0.0086 |
| | 3 | 50 | 0.247 | 0.249 | 0.273 | 0.241 | 0.1133 |
| | | 200 | 0.143 | 0.153 | 0.225 | 0.147 | 0.0655 |
| | | 2000 | 0.051 | 0.053 | 0.085 | 0.049 | 0.0189 |
| | | 7500 | 0.027 | 0.028 | 0.047 | 0.027 | 0.0101 |

Table 8. The expected (across 500 samples) number of ensemble members falling into the sampling window about $h_{n-1}$, and forming the sub-sample from which an empirical estimate $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ is obtained having the expected $MAD$ less than $0.06$.

| Forecast point | $n$ | $v$ | Index[1/] of $h_{n-1}$: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eldred | 2 | 0 | – | 211 | – | – | – | – | – | – | | 211 |
| | 2 | 1 | – | 206 | 193 | 198 | 191 | 191 | 215 | 191 | | 198 |
| | 3 | 0 | – | 199 | 203 | 203 | – | – | – | – | | 202 |
| | 3 | 1 | – | 257 | 222 | 189 | 188 | 195 | 193 | 191 | | 205 |
| Dailey | 2 | 0 | – | 208 | – | – | – | – | – | – | | 208 |
| | 2 | 1 | – | – | 222 | 223 | 188 | 198 | 184 | 215 | | 205 |
| | 3 | 0 | 379 | 194 | 225 | – | – | – | – | – | | 266 |
| | 3 | 1 | – | – | 192 | 188 | 185 | 202 | 225 | 189 | | 197 |
| Philippi | 2 | 0 | – | 195 | – | – | – | – | – | – | – | 195 |
| | 2 | 1 | – | – | 191 | 222 | 220 | 221 | 207 | 197 | – | 210 |
| | 3 | 0 | – | 212 | – | – | – | – | – | – | – | 212 |
| | 3 | 1 | – | – | 193 | 195 | 206 | 198 | 192 | 193 | 190 | 195 |
| Parsons | 2 | 0 | – | 192 | – | – | – | – | – | – | – | 192 |
| | 2 | 1 | – | – | – | 202 | 183 | 191 | 196 | 222 | 191 | 198 |
| | 3 | 0 | 207 | 220 | 245 | 212 | – | – | – | – | – | 221 |
| | 3 | 1 | – | 227 | 206 | 202 | 193 | 209 | 212 | 194 | – | 206 |
| | | | | | | | | | Grand average | | | 208 |

[1]/ The stages [in ft] corresponding to the indices 1, 2, 3, 4, 5, 6, 7, 8, 9 are as follows:

| | |
|---|---|
| Eldred | 5, 6, 7, 8, 9, 10, 11, 12, none |
| Dailey | 2, 3, 4, 5, 6, 8, 10, 12, none |
| Philippi | 6, 8, 10, 12, 14, 16, 20, 24, 28 |
| Parsons | 4, 5, 6, 7, 8, 10, 12, 14, 16 |

[2]/ A dash means that there is no empirical estimate with the average $MAD$ less than $0.06$ for that case.

Table 9. A brief empirical guide: the smallest ensemble size $M^*$ required to ensure the expected $MAD$ less than a threshold, depending on the kind of predictive distribution functions needed for decision making.

| Distribution kind | Expected $MAD$ threshold | Required ensemble size $M^*$ |
|---|---|---|
| Marginal | $0.06$ | $200$ |
| | $0.04$ | $500$ |
| | $0.02$ | $2000$ |
| Transition | $0.06$ | $700$–$5000$ |
| | $0.04$ | $1500$–$10000$ |
| | $0.02$ | $7000$– |

For transition, to achieve the expected $MAD = 0.06$ when the preceding river stage is discretized into $L$ windows, the lower bound on $M^*$ is $210 \cdot L$.

## FIGURE CAPTIONS

Figure 1. The ensemble size $M$ (black line and dots) and the number of realizations conditional on precipitation occurrence $M_1$ (gray line and crosses) needed to preserve the probability of precipitation occurrence $\nu$ when the smallest required sample size is $M^* = 100$.

Figure 2. The PQPF specifying the distribution function of the 24-h basin average precipitation amount for Eldred (upper graph) and for Dailey, Philippi, Parsons (lower graph).

Figure 3. Spaghetti plot of the 100-member Bayesian ensemble forecast of river stages generated using the PQPF shown in Figure 2; Eldred.

Figure 4. Predictive $n$-step transition distribution functions $\hat{\Psi}_n$ of river stages $H_n$ for $n = 1, ..., 12$, estimated from the 100-member Bayesian ensemble forecast shown in Figure 3; Eldred.

Figure 5. Comparison of the PRSFs: the conditional predictive 3-step transition distribution function $\Psi_{3v}$ calculated analytically from the BFS and its three empirical estimates $\hat{\Psi}_{3v}$ obtained from three different ensembles of size $M = 100$. In each case $(v = 0, 1)$, one estimate is near perfect, one is biased leftward, and one is biased rightward; Eldred.

Figure 6. Select conditional predictive one-step transition distribution functions that comprise the family which generates the ensemble members from which the function $\Theta_{2v}(\cdot|h_1)$ is actually estimated when the sampling window size is $d_{2v}$; $h_1 = 6$ ft for $v = 0$, and $h_1 = 8$ ft for $v = 1$; Eldred.

Figure 7. The average (across four forecast points, two precipitation events, three lead times) of the expected (across 500 samples) $MAD$ between $\hat{\Psi}_{nv}$ and $\Psi_{nv}$.

Figure 8. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ as a function of the ensemble size $M_v$ for $v = 0, 1$ and $n = 2, 3$; Eldred.

Figure 9. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ as a function of the ensemble size $M_v$ for $v = 0, 1$ and $n = 2, 3$; Philippi.

Figure 10. The average, the minimum, and the maximum (across four forecast points, two precipitation events, two lead times) of the expected (across 500 samples) $MAD$ between $\hat{F}_{nv}$ and $F_{nv}$.

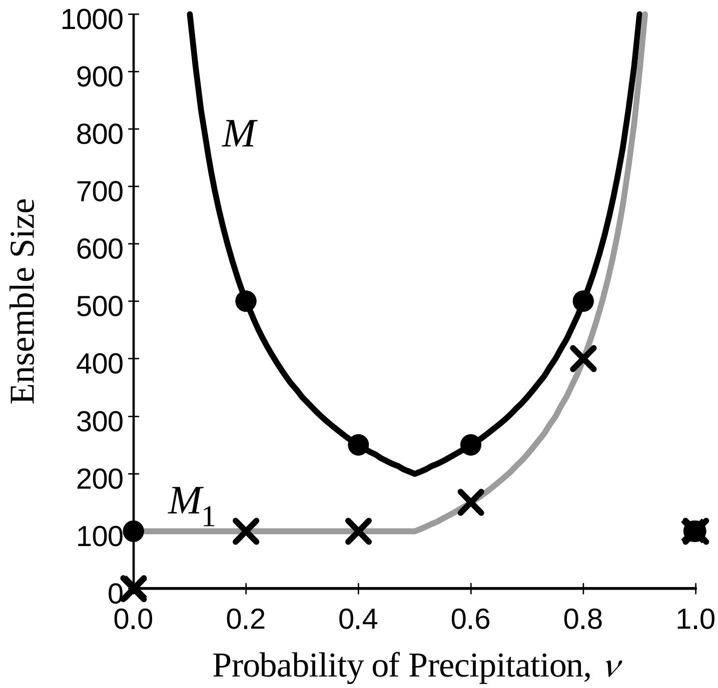
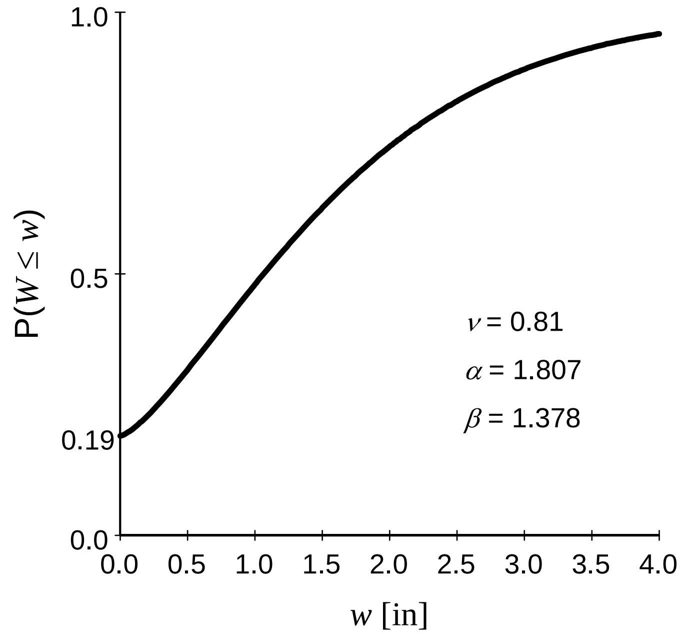



Figure 1. The ensemble size $M$ (black line and dots) and the number of realizations conditional on precipitation occurrence $M_1$ (gray line and crosses) needed to preserve the probability of precipitation occurrence $\nu$ when the smallest required sample size is $M^* = 100$.



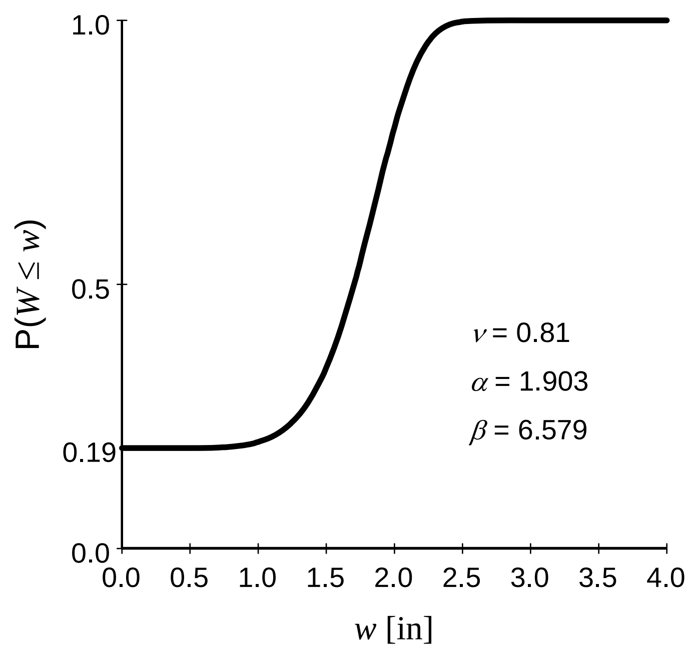




Figure 2. The PQPF specifying the distribution function of the 24-h basin average precipitation amount for Eldred (upper graph) and for Dailey, Philippi, Parsons (lower graph).

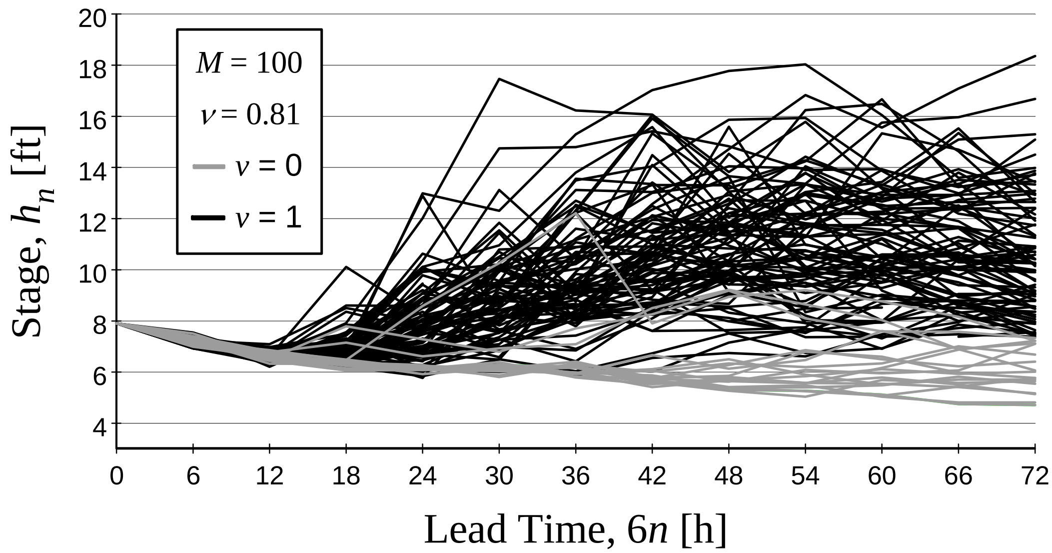



Figure 3. Spaghetti plot of the 100-member Bayesian ensemble forecast of river stages generated using the PQPF shown in Figure 2; Eldred.

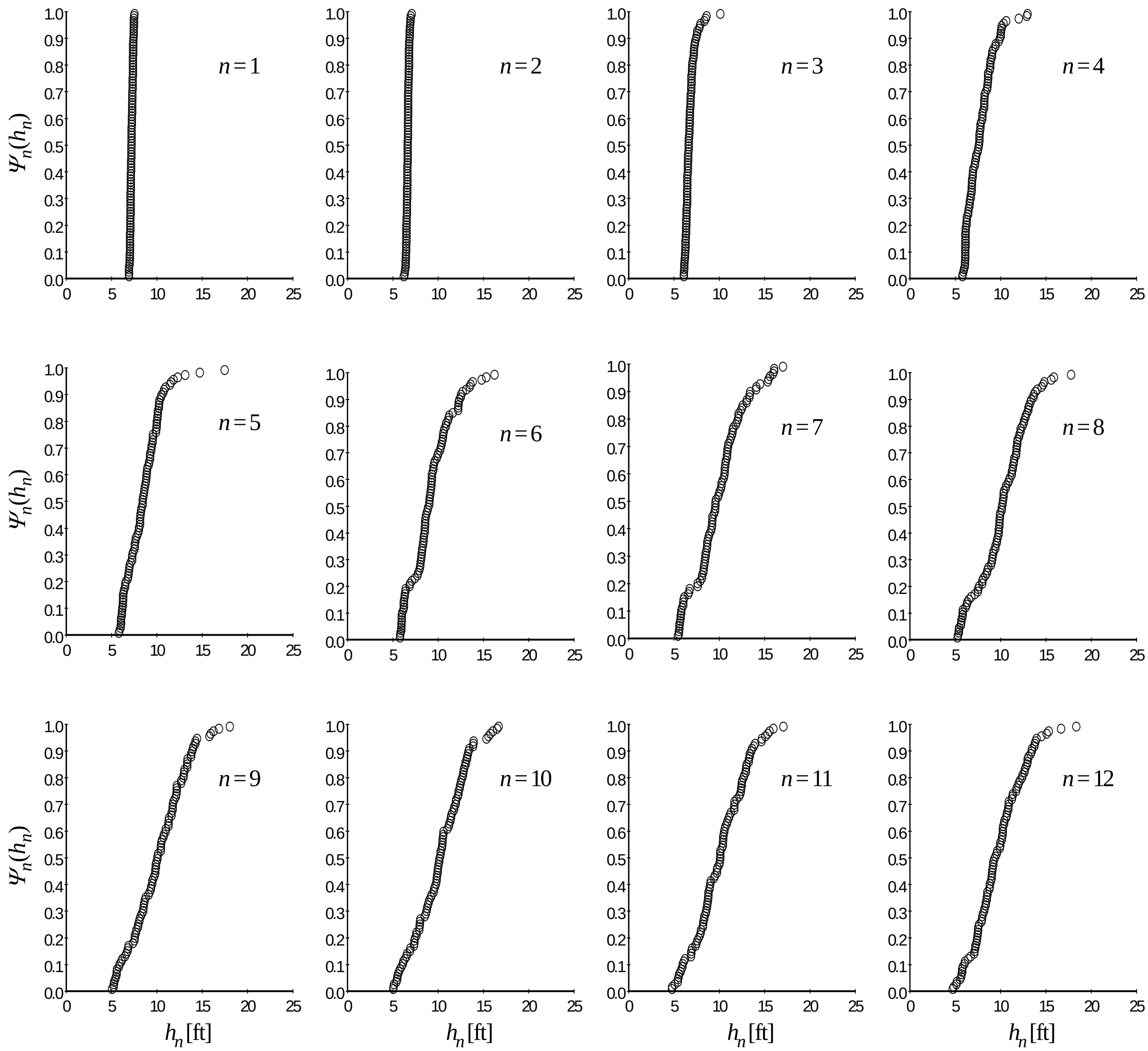



Figure 4. Predictive $n$-step transition distribution functions $\hat{\Psi}_n$ of river stages $H_n$ for $n = 1, ..., 12$, estimated from the 100-member Bayesian ensemble forecast shown in Figure 3; Eldred.

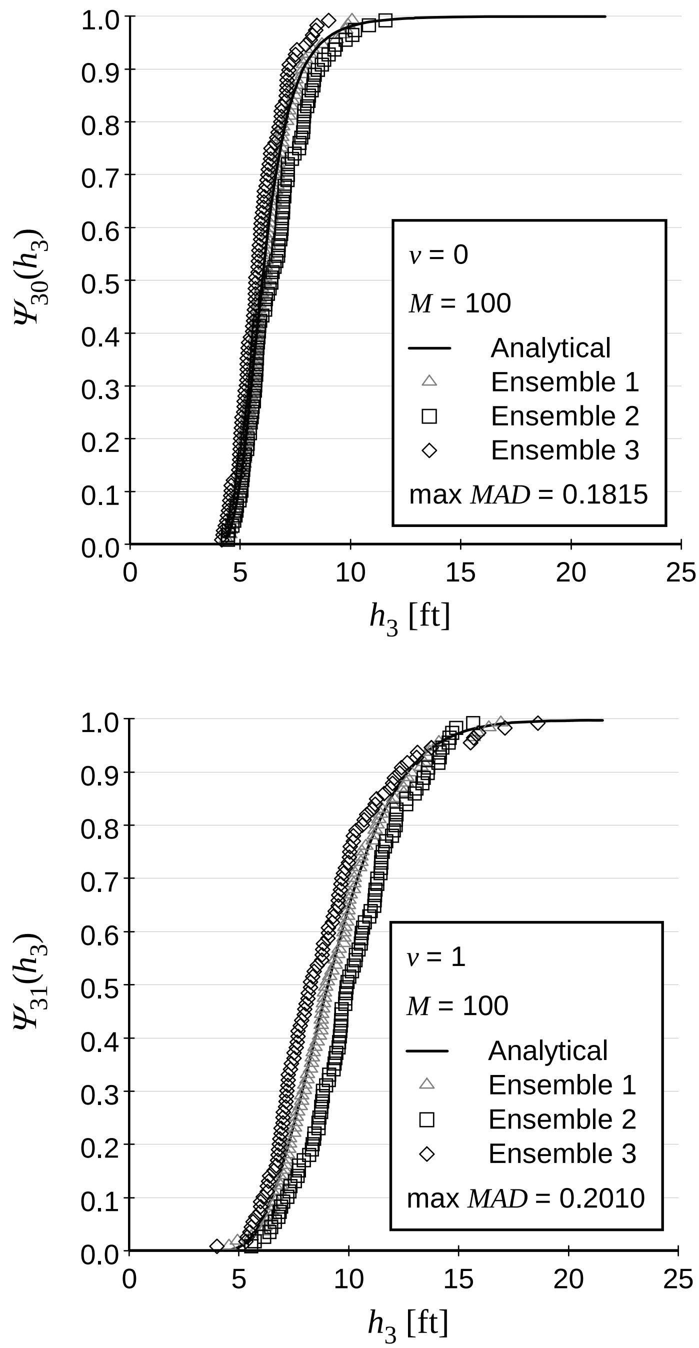
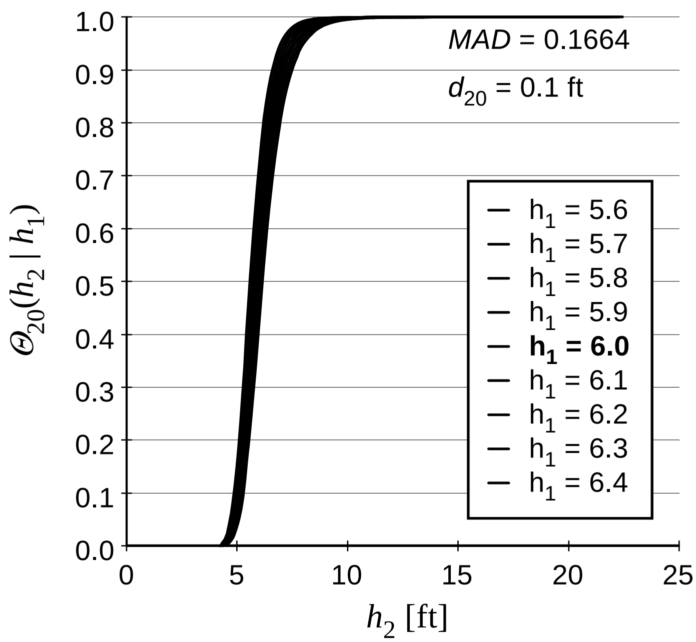



Figure 5. Comparison of the PRSFs: the conditional predictive 3-step transition distribution function $\Psi_{3v}$ calculated analytically from the BFS and its three empirical estimates $\hat{\Psi}_{3v}$ obtained from three different ensembles of size $M = 100$. In each case $(v = 0, 1)$, one estimate is near perfect, one is biased leftward, and one is biased rightward; Eldred.



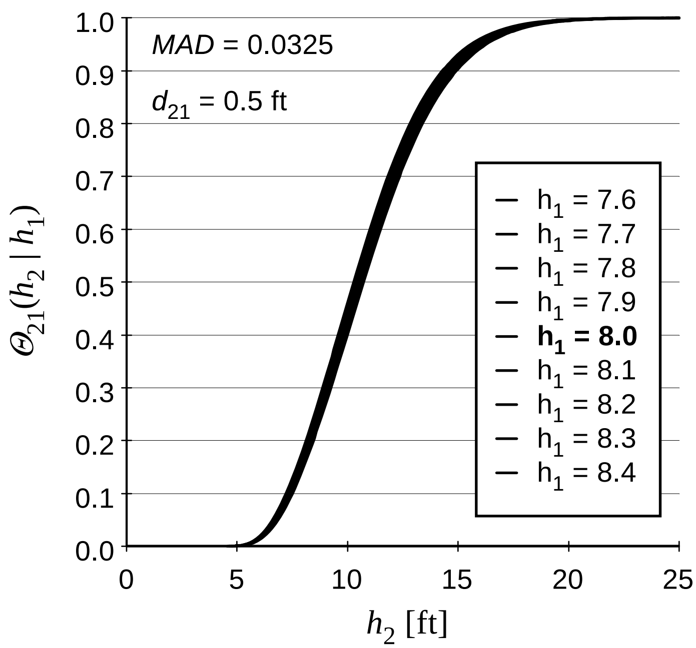




Figure 6. Select conditional predictive one-step transition distribution functions that comprise the family which generates the ensemble members from which the function $\Theta_{2v}(\cdot|h_1)$ is actually estimated when the sampling window size is $d_{2v}$; $h_1 = 6$ ft for $v = 0$, and $h_1 = 8$ ft for $v = 1$; Eldred.

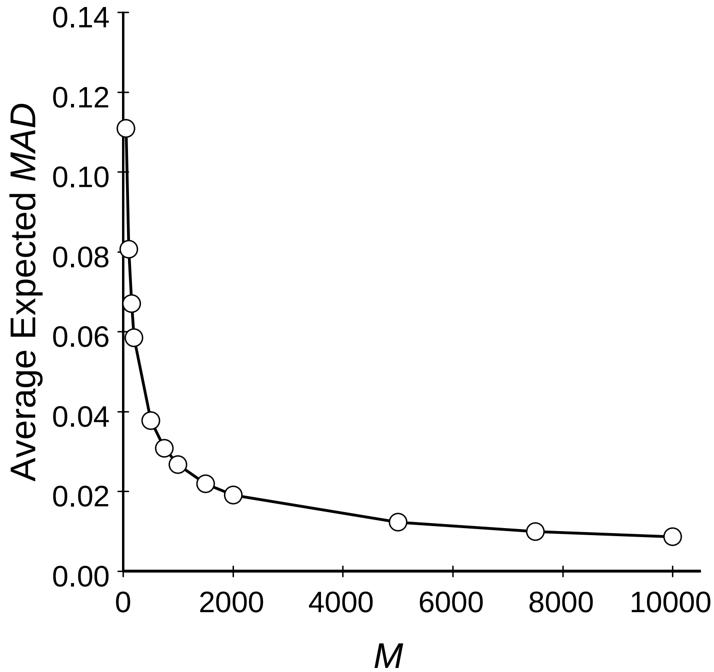



Figure 7. The average (across four forecast points, two precipitation events, three lead times) of the expected (across 500 samples) $MAD$ between $\hat{\Psi}_{nv}$ and $\Psi_{nv}$.

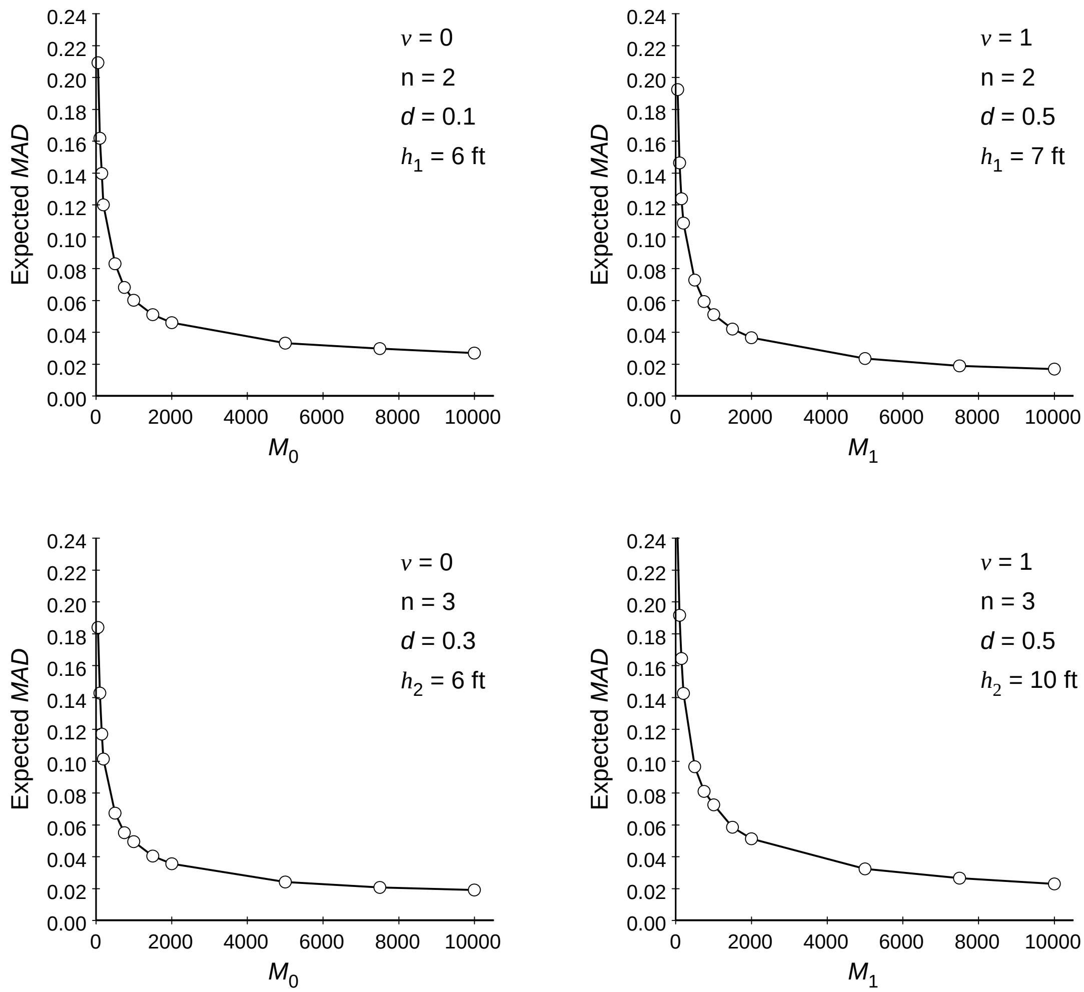



Figure 8. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ as a function of the ensemble size $M_v$ for $v = 0, 1$ and $n = 2, 3$; Eldred.

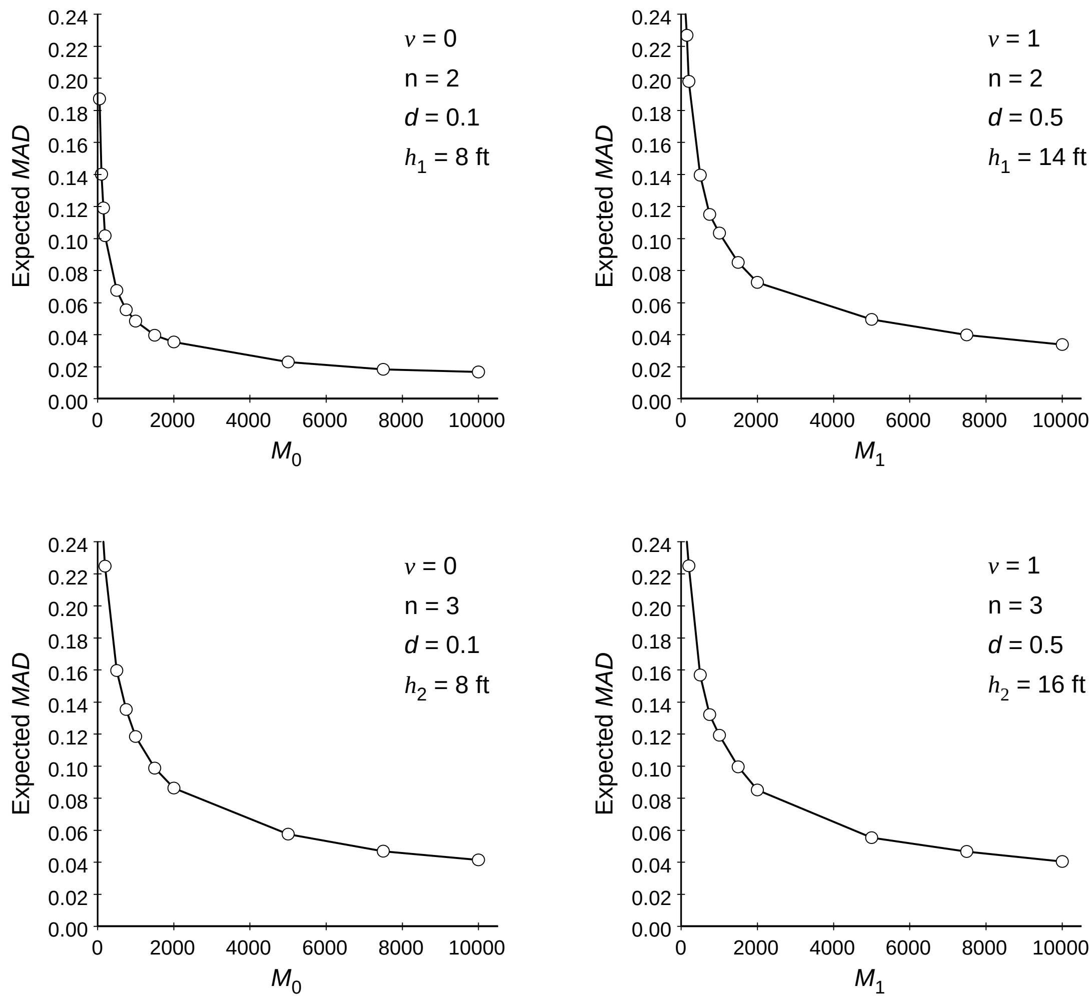



Figure 9. The expected (across 500 samples) $MAD$ between $\hat{\Theta}_{nv}(\cdot|h_{n-1})$ and $\Theta_{nv}(\cdot|h_{n-1})$ as a function of the ensemble size $M_v$ for $v = 0, 1$ and $n = 2, 3$; Philippi.

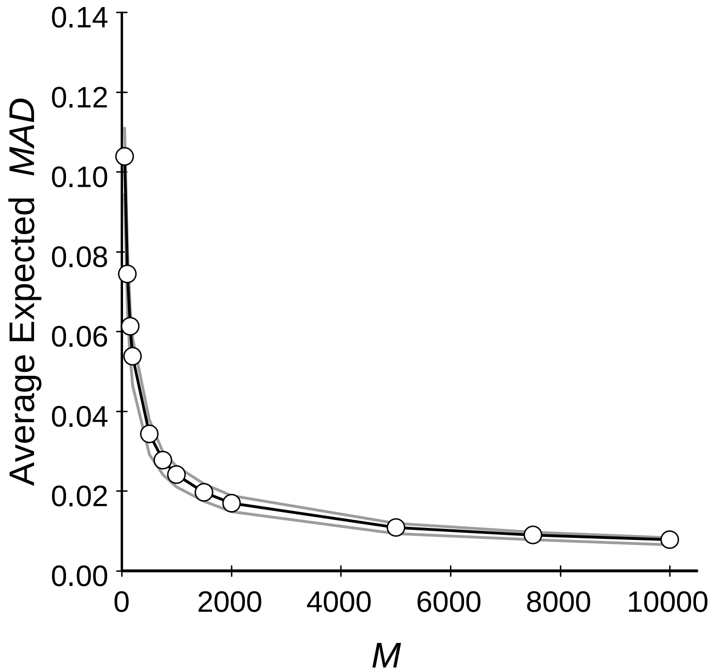



Figure 10. The average, the minimum, and the maximum (across four forecast points, two precipitation events, two lead times) of the expected (across 500 samples) $MAD$ between $\hat{F}_{nv}$ and $F_{nv}$.