Journal of Hydrology 524 (2015) 789–802

Contents lists available at ScienceDirect

Journal of Hydrology

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

# Ensemble Bayesian forecasting system Part I: Theory and algorithms

Henry D. Herr [a,*], Roman Krzysztofowicz [b,c]

[a] Office of Hydrologic Development, National Weather Service, Building SSMC II, 1325 East West Highway, Silver Spring, MD 20910, USA
[b] Department of Systems Engineering, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747, USA
[c] Department of Statistics, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747, USA

ARTICLE INFO

*Article history:*
Available online 2 December 2014

*Keywords:*
Forecasting
Bayesian analysis
Stochastic processes
Probability
Ensemble
Rivers

SUMMARY

The ensemble Bayesian forecasting system (EBFS), whose theory was published in 2001, is developed for the purpose of quantifying the total uncertainty about a discrete-time, continuous-state, non-stationary stochastic process such as a time series of stages, discharges, or volumes at a river gauge. The EBFS is built of three components: an input ensemble forecaster (IEF), which simulates the uncertainty associated with random inputs; a deterministic hydrologic model (of any complexity), which simulates physical processes within a river basin; and a hydrologic uncertainty processor (HUP), which simulates the hydrologic uncertainty (an aggregate of all uncertainties except input). It works as a Monte Carlo simulator: an ensemble of time series of inputs (e.g., precipitation amounts) generated by the IEF is transformed deterministically through a hydrologic model into an ensemble of time series of outputs, which is next transformed stochastically by the HUP into an ensemble of time series of predictands (e.g., river stages). Previous research indicated that in order to attain an acceptable sampling error, the ensemble size must be on the order of hundreds (for probabilistic river stage forecasts and probabilistic flood forecasts) or even thousands (for probabilistic stage transition forecasts). The computing time needed to run the hydrologic model this many times renders the straightforward simulations operationally infeasible. This motivates the development of the ensemble Bayesian forecasting system with randomization (EBFSR), which takes full advantage of the analytic meta-Gaussian HUP and generates multiple ensemble members after each run of the hydrologic model; this auxiliary randomization reduces the required size of the meteorological input ensemble and makes it operationally feasible to generate a Bayesian ensemble forecast of large size. Such a forecast quantifies the total uncertainty, is well calibrated against the prior (climatic) distribution of predictand, possesses a Bayesian coherence property, constitutes a random sample of the predictand, and has an acceptable sampling error—which makes it suitable for rational decision making under uncertainty.



## 1. Introduction

### 1.1. Background

The need for probabilistic forecasting in hydrology has been recognized (Krzysztofowicz, 2001a), and steps toward routine production of probabilistic forecasts are being taken (Schaake et al., 2007; Demargne et al., 2014). Because modeling hydrologic processes in large basins, wherein dependencies between a basin and neighboring or upstream basins must be accounted for, is very complex, specialized techniques for producing probabilistic forecasts through a deterministic hydrologic model must be developed. However, no existing techniques are wholly adequate: either they fail to satisfy important theoretic properties or they do not meet the needs of all users in all basins.

Probabilistic forecast of a stochastic process $\{H_1, \ldots, H_N\}$ may take one of two forms: analytical or ensemble. An analytical forecast provides a predictive joint distribution function of the process $\{H_1, \ldots, H_N\}$, which directly quantifies uncertainty about all predictands. Such a forecast is most appropriate for users employing analytical decision systems which require distribution functions (e.g., warning-response models and stochastic control models). However, for users employing simulation-based decision systems, the required format of the forecast is an ensemble of possible realizations of the process $\{H_1, \ldots, H_N\}$. This ensemble may be used as input to a decision system and as a sample for estimating empirical distribution functions of desired predictands.

Starting from normative requirements of rational deciders, Krzysztofowicz (1999) formulated a Bayesian theory of probabilistic

* Corresponding author. Tel.: +1 301 713 0640x166.
*E-mail address:* Hank.Herr@noaa.gov (H.D. Herr).

http://dx.doi.org/10.1016/j.jhydrol.2014.11.072
0022-1694/Published by Elsevier B.V.

**Nomenclature**

*General*

| | |
|---|---|
| HUP | hydrologic uncertainty processor |
| IEF | input ensemble forecaster |
| INT | integrator |
| IUP | input uncertainty processor |
| MCG | Monte-Carlo generator |
| NQT | normal quantile transform |
| *Cor* | Pearson's product-moment correlation function |
| *r* | response function representing deterministic hydrologic model |

*Variables and parameters*

| | |
|---|---|
| $a_{nv}$, $b_{nv}$, $d_{nv}$, $e_{nv}$ | (HUP) likelihood conditional regression coefficients |
| $A_{nv}$, $B_{nv}$, $D_{nv}$, $E_{nv}$, $T_{nv}$ | (HUP) posterior parameters |
| $c_{nv}$ | (HUP) prior conditional correlation coefficient |
| $\boldsymbol{h}_0$ | observations of $\boldsymbol{H}$ up to forecast time |
| $k$ | index of realizations in a sample |
| $K$ | likelihood sample size |
| $K'$ | prior sample size |
| $M$ | ensemble size, number of ensemble members (realizations) |
| $M_P$ | number of runs of the hydrologic model with $\boldsymbol{\omega} > 0$ |
| $M_v$ | number of ensemble members, conditional on $V = v$ |
| $n$ | index of time steps, index of lead times |
| $N$ | last time step, last lead time |
| $p_n$ | probability number |
| $R$ | randomization factor |
| $t_0$ | forecast time |
| $t_n$ | time instance |
| $T$ | lead time of the forecast of $\boldsymbol{\Omega}$ |
| $\boldsymbol{u}$ | deterministic inputs to the hydrologic model |
| $\boldsymbol{v}$ | information predicting $\boldsymbol{\Omega}$ |
| $\nu$ | probability of precipitation occurrence |
| $\boldsymbol{y}$ | states that partially explain hydrologic uncertainty |
| $\Theta_{nv}$ | (HUP) residual variate from likelihood conditional regression |
| $\Xi_{nv}$ | (HUP) residual variate from prior conditional regression |
| $\sigma^2_{nv}$ | (HUP) variance of $\Theta_{nv}$ |
| $\tau^2_{nv}$ | (HUP) variance of $\Xi_{nv}$ |

*Variates and realizations*

| | |
|---|---|
| $\boldsymbol{H}$, $\boldsymbol{h}$ | predictands (variates being forecasted) |
| $H_n$, $h_n$ | actual river stage at time $t_n$ |
| $\boldsymbol{S}$, $\boldsymbol{s}$ | outputs from the hydrologic model |
| $S_n$, $s_n$ | model river stage at time $t_n$ |
| $V$, $v$ | indicator of precipitation occurrence |
| $W_n$, $w_n$ | NQT of $H_n$, $h_n$ |
| $X_n$, $x_n$ | NQT of $S_n$, $s_n$ |
| $\boldsymbol{\Omega}$, $\boldsymbol{\omega}$ | inputs to the hydrologic model forecasted probabilistically |

*Distribution and density functions*

| | |
|---|---|
| *DF*, *df* | distribution function, density function |
| $f$ | conditional *df* of $\boldsymbol{S}$, likelihood function of $\boldsymbol{H}$ |
| $g$ | prior conditional *df* of $\boldsymbol{H}$ |
| $P$ | generic probability function |
| $Q$, $Q^{-1}$ | standard normal *DF*, inverse of *DF* |
| $\Gamma_{nv}$ | (HUP) marginal *DF* of $H_n$ within HUP |
| $\eta$ | generalized *df* of $\boldsymbol{\Omega}$ |
| $\bar{\Lambda}_{nv}$ | (HUP) marginal *DF* of $S_n$ within HUP |
| $\xi$ | generalized predictive *df* of $\boldsymbol{H}$ |
| $\pi$ | generalized *df* of $\boldsymbol{S}$ |
| $\phi$ | generalized posterior *df* of $\boldsymbol{H}$ |
| $\Phi_{nv}$ | (HUP) conditional posterior one-step transition *DF* |

forecasting via deterministic hydrologic model. From that theory, three analytic-numerical *Bayesian forecasting systems* (BFS) were developed. The most complex one, the analytic-numerical BFS for probabilistic stage transition forecasting (Krzysztofowicz and Maranzano, 2004b), was recently deployed as a Monte-Carlo generator of the Bayesian ensemble forecast of the river stage time series (Herr and Krzysztofowicz, 2010): the ensemble is generated by recursively sampling from the family of analytical one-step transition distributions output by the BFS. This ensemble generator satisfies important theoretic properties, is very efficient computationally, and meets the needs of users who require an ensemble forecast. However, it inherits the limitation of the analytic-numerical BFS: it is suitable for small-to-medium headwater basins only because the uncertainty in the spatio-temporal disaggregation of the total precipitation amount can be modeled analytically only in an approximate fashion.

### 1.2. Objective

The research reported herein works toward a general Bayesian technique for ensemble forecasting, which satisfies the key theoretic properties and can be easily scaled up to large basins, so long as an appropriate source of an ensemble of future inputs to a hydrologic model is available. The basic technique, the *ensemble Bayesian forecasting system* (EBFS), employs a Monte Carlo generator which outputs a random sample of the future hydrologic time series. However, as will be shown experimentally, this technique may be operationally infeasible due to computing time needed to meet the ensemble size requirements established by Herr and Krzysztofowicz (2010). This finding motivates the development of a refined technique, the *ensemble Bayesian forecasting system with randomization* (EBFSR), which can increase a given ensemble size without additional runs of the hydrologic model; this technique makes generation of large ensembles operationally feasible.

The research is reported in two parts. Part I presents the theory, the models, and the forecasting algorithms. Part II (Herr and Krzysztofowicz, 2015) reports numerical experiments whose purpose is to validate the EBFS and EBFSR, to illustrate their properties, and to establish guidelines for choosing the randomization factor in the EBFSR.

### 1.3. Required system properties

From the viewpoint of Bayesian forecast-decision theory (Krzysztofowicz, 1983, 1999), there are three properties required of any probabilistic forecasting system intended to provide information for rational decision making under uncertainty:

1. it must quantify all sources of uncertainty pertaining to the predictand;
2. it must possess a self-calibration property, wherein, in the long run, the probabilistic forecasts preserve the prior (climatic) distribution of the predictand; and
3. it must possess a coherence property, wherein the economic value of the forecast is never negative, relative to the value of the prior distribution of the predictand.