BAYESIAN THEORY of ENSEMBLE FORECASTING

Roman Krzysztofowicz
University of Virginia
USA

Presented at the
HEPEX Meeting on
Uncertainty Post-Processing for Streamflow Forecasts
Delft, The Netherlands
23 – 25 June 2008

Acknowledgments:
National Science Foundation, Grant No. ATM–0641572
Henry D. Herr, Coire J. Maranzano
BAYESIAN THEORY
(Krzysztofowicz, 1999)

- Derived from: Total Probability Law
  Bayes Theorem

- Applicable to any deterministic hydrologic model
- Fuses optimally information  
  (=) informativeness
- Quantifies total uncertainty
- Outputs probabilistic forecast  
  (=) calibration

- Basis for: system decomposition
  mathematical models
  computation methods

1. Bayesian Forecasting System (BFS)
   Analytic – Numerical
   For headwater basin

2. Ensemble BFS (EBFS)
   Ensemble / Monte-Carlo simulation
   For any basin
DECOMPOSITION OF UNCERTAINTY

OPERATIONAL UNCERTAINTY (exterior to forecasting theory)
- Missing data
- Processing errors
- ...

✓ INPUT UNCERTAINTY (dominant uncertainty)
- Future precipitation

✓ HYDROLOGIC UNCERTAINTY (all other uncertainties)
- Models
- Parameters
- Initial conditions
- Deterministic inputs
- ...

TOTAL: INPUT and HYDROLOGIC
Decomposition Theorem

\[
\begin{bmatrix}
\text{Total Uncertainty} \\
\text{Uncertainty}
\end{bmatrix} = \int \begin{bmatrix}
\text{Hydrologic Uncertainty} \\
\text{Uncertainty}
\end{bmatrix} \cdot \begin{bmatrix}
\text{Input Uncertainty}
\end{bmatrix}
\]

[Bayesian HUP]  

Calibration Theorem

In a *Bayesian Forecasting System* (BFS) with any deterministic hydrologic model, if the input distribution is *well calibrated*, then the output distribution is *well calibrated*.

Inferred Requirements on Input

1. Ensemble QPF must be *well calibrated*.
2. Ensemble size must be *large enough*.
ENSEMBLE (NAIVE) SYSTEM

Precipitation Amounts \( \{w_1, \ldots, w_M\} \)

Model River Stages \( \{s_1, \ldots, s_N\} \)

Deterministic Hydrologic Model

How to implement Bayesian Theory?
Precipitation Ensembles Processor (Monte Carlo) PEP

Precipitation Amounts \{\{w_1, \ldots, w_M\}\}

Deterministic Hydrologic Model

Model River Stages \{\{s_1, \ldots, s_N\}\}

Hydrologic Uncertainty Processor (Monte Carlo) HUP

Actual River Stages \{\{h_1, \ldots, h_N\}\}

Real-Time Processing

Off-line Simulation
NOTATION

EQPF Period

0 1 2 3  n

\( h_0 \quad h_1 \quad h_2 \quad h_3 \) Actual
\( s_1 \quad s_2 \quad s_3 \) Model

\( \nu \) – indicator of precip. occurrence during EQPF period:
\[ \nu = 1 \quad \text{yes,} \quad \nu = 0 \quad \text{no} \]

\( n \) – index of times
\( h_n \) – actual river stage
\( s_n \) – model river stage

induced by actual precipitation during EQPF period
→ no precip. uncertainty
Process: Two-branch, Non-stationary, Markov

Two families of joint conditional densities:

$$\phi_v(h_N|s_N, h_0) \quad v = 0, 1$$

Bayesian Formulation

Meta-Gaussian Model:

- Marginal distributions: Any form
- Dependence structure: Non-linear
  Heteroscedastic
HYDROLOGIC UNCERTAINTY PROCESSOR

Prior 1-step Transition Density
\[ r_{nv}(h_n|h_{n-1}) \quad n = 1, \ldots, N \quad v = 0, 1 \]

- stochastic model of river stage process

Likelihood Function
\[ f_{nv}(s_n|h_n, h_{n-1}, h_0) \quad n = 1, \ldots, N \quad v = 0, 1 \]

- stochastic representation of deterministic model

\[ \rightarrow \] Posterior 1-step Transition Density
\[ \phi_{nv}(h_n|s_n, h_{n-1}, h_0) \propto f_{nv} \cdot r_{nv} \quad n = 1, \ldots, N \quad v = 0, 1 \]

- Joint Conditional Density
\[ \phi_v(h_N|s_N, h_0) = \prod_{n=1}^{N} \phi_{nv}(h_n|s_n, h_{n-1}, h_0) \quad v = 0, 1 \]
Conditional Expected Density

\[ \kappa_{nv}(s_n|h_{n-1}, h_0) = \int_{-\infty}^{\infty} f_{nv}(s_n|h_n, h_{n-1}, h_0) \ r_{nv}(h_n|h_{n-1}) \ dh_n \]

\[ n = 1, \ldots, N \quad v = 0, 1 \]

Posterior 1–step Transition Density

\[ \phi_{nv}(h_n|s_n, h_{n-1}, h_0) = \frac{f_{nv}(s_n|h_n, h_{n-1}, h_0) \ r_{nv}(h_n|h_{n-1})}{\kappa_{nv}(s_n|h_{n-1}, h_0)} \]

\[ n = 1, \ldots, N \quad v = 0, 1 \]
EQPF Period

\[ v = 1 \quad \text{or} \quad v = 0 \]

\[ h_0 \quad h_1 \quad h_2 \quad h_3 \]
\[ s_1 \quad s_2 \quad s_3 \]

Actual Model

Prior Distribution \(\leftarrow\) historical record (long)
\[ \{(v; h_0, h_1, h_2, h_3)\} \]

Likelihood Function \(\leftarrow\) simulation experiment (short)
\[ \{(v; s_1, s_2, s_3, h_0, h_1, h_2, h_3)\} \]

Hydrologic model re-initialized (as in real-time)
Real-time inputs, except precipitation
Actual precip. during EQPF period
\[ \rightarrow\text{ No precip. uncertainty} \]
Marginal Distribution Functions of $H_1$

$V = 1$

$V = 0$
Prior Dependence Structure: \( H_n|H_{n-1}, \quad V = 1 \)

\( n = 1, \quad 24 \text{ h} \)
\( n = 2, \quad 48 \text{ h} \)
\( n = 3, \quad 72 \text{ h} \)

(a)

\( c_{11} = 0.702 \)

\( c_{21} = 0.810 \)

\( c_{31} = 0.789 \)

(b)

\( \rho_{11} = 0.685 \)

\( \rho_{21} = 0.796 \)

\( \rho_{31} = 0.774 \)
**Prior Dependence Structure:**

\[ H_n | H_{n-1}, \quad V = 0 \]

\[ n = 1, \quad 24 \text{ h} \quad n = 2, \quad 48 \text{ h} \quad n = 3, \quad 72 \text{ h} \]

(a) \[ c_{10} = 0.948 \]

(b) \[ \rho_{10} = 0.944 \]

\[ c_{20} = 0.797 \]

\[ \rho_{20} = 0.783 \]

\[ c_{30} = 0.813 \]

\[ \rho_{30} = 0.800 \]
Marginal Distribution Functions of $H_2$ and $S_2$
Likelihood Dependence Structure: $S_n | H_n, V = 1$

$n = 1, \ 24 \text{ h} \quad n = 2, \ 48 \text{ h} \quad n = 3, \ 72 \text{ h}$

(a)

\[
\begin{align*}
a_{11} &= 0.946 \\
a_{21} &= 0.838 \\
a_{31} &= 0.682
\end{align*}
\]

(b)

\[
\begin{align*}
\rho_{11} &= 0.941 \\
\rho_{21} &= 0.826 \\
\rho_{31} &= 0.665
\end{align*}
\]
Likelihood Dependence Structure: $S_n | H_n, \ V = 0$

$n = 1, \ 24 \ h \quad n = 2, \ 48 \ h \quad n = 3, \ 72 \ h$

(a) 

$w_1 \quad w_2 \quad w_3$

$x_1 = 0.998 \quad x_2 = 0.901 \quad x_3 = 0.811$

(b) 

$\rho_{10} = 0.998 \quad \rho_{20} = 0.893 \quad \rho_{30} = 0.797$
Prior and Posterior 1-Step Transition Densities

\( n = 1 \)
24 h

\( n = 2 \)
48 h

\( n = 3 \)
72 h
Posterior 1-Step Transition Densities

\( n = 1 \)
24 h

\( n = 2 \)
48 h

\( n = 3 \)
72 h
ENGLISH BAYESIAN SYSTEM

Precipitation Ensemble Processor
PEP (Monte Carlo)

Precipitation Amounts \{ (w_1, \ldots, w_M) \}

Deterministic Hydrologic Model

Model River Stages \{ (s_1, \ldots, s_N) \}

Hydrologic Uncertainty Processor
HUP

Integrator
INT (Monte Carlo)

Actual River Stages \{ (h_1, \ldots, h_N) \}

Real-Time Processing

Off-line Simulation
INTEGRATOR (Monte Carlo)

- **Input:** one realization

  Model River Stages $\{s_1, s_2, s_3, \ldots, s_N\}$
  Indicator of Precip. $v$
  Initial River Stage $h_0$

<table>
<thead>
<tr>
<th>Given</th>
<th>Generate</th>
<th>From Posterior Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>$h_1$</td>
<td>$\Phi_{1v}(\cdot</td>
</tr>
<tr>
<td>$s_2, h_1$</td>
<td>$h_2$</td>
<td>$\Phi_{2v}(\cdot</td>
</tr>
<tr>
<td>$s_3, h_2$</td>
<td>$h_3$</td>
<td>$\Phi_{3v}(\cdot</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Output:** multiple realizations

  Actual River Stages $\{(h_1, h_2, h_3, \ldots, h_N)\}$
Bayesian Ensemble Forecast of River Stages
Predictive $n$-Step Transition Distributions

Empirical Distributions

Ensemble Size $M = 100$
ERROR IN DISTRIBUTION Due to Ensemble Size

$v = 0$
$M = 100$

$\Psi_{30}(h_3)$

$MAD_{max} = 0.1815$

$v = 1$
$M = 100$

$\Psi_{31}(h_3)$

$MAD_{max} = 0.2010$
MAD Between Ensemble $\hat{\Psi}_{nv}(h_n)$ and Analytical $\Psi_{nv}(h_n)$

Average MAD across 500 ensembles
MAD Between Ensemble $\hat{\Theta}_{nv}(h_n|h_{n-1})$ and Analytical $\Theta_{nv}(h_n|h_{n-1})$

Average MAD across 500 ensembles
Ensemble Bayesian Forecasting System (EBFS)

0. Deterministic Hydrologic Model
   1. Precipitation Ensemble Processor
      • calibrated ensemble QPF

2. Hydrologic Uncertainty Processor
   • all hydrologic uncertainties
   • correct theoretic structure (Bayesian)
   • model fitting data (meta-Gaussian)

3. Integrator
   • Monte Carlo re-generation (ensemble size)

Forecast Products
   • calibrated distributions
REFERENCES


