Ensemble Verification in the Hydrology Program of the NOAA/National Weather Service

James D Brown¹,², Julie Demargne¹,², Yuqiong Liu¹,³, Dong-Jun Seo¹,²

Hydrologic Ensemble Prediction Group (HEP)
¹NOAA/NWS/Office of Hydrologic Development (OHD)
²University Corporation for Atmospheric Research
³Riverside Technology, Inc.

NCAR DTC Verification Workshop
Boulder CO, April 16-18, 2008
Outline

- Hydrologic verification: the problem
  - Most techniques pioneered in meteorology
  - Difficult to apply in operational hydrology
    - multiple scales and multiple users
    - real-time setting

- Recent work on diagnostic metrics
  - Simpler, user-friendly metrics
  - Ensemble Verification System (EVS)

- Real-time (“prognostic”) verification
  - Assessing how “similar” past ensemble forecasts performed
Hydrologic forecasting: a multi-scale problem

National
Major River System
River Basin with River Forecast Points
Forecast Group
Headwater Basin and Radar Rainfall Grid
High Resolution Flash Flood Basins

April 16-18, 2008
NCAR DTC Verification Workshop, Boulder CO
Seamless probabilistic water forecasts are required for all lead times and all users; so is verification information.
Hydrologic verification: from weather and climate forecasts to river and water forecasts

Goal: improve reliability and skill and accurately account for uncertainties
Recent Work on Diagnostic Verification
User Friendly Reliability

Issue: traditional reliability diagram is rigorous but complex ...

User Friendly Reliability

- Where do observations fall in forecast?
- Split the forecast into probability windows
- E.g. 10% of obs. should fall in each decile
- Similar to Talagrand diagram ...

\[ REL(p_w) = \frac{1}{m} \sum_{j=1}^{m} I_j(p_w) \quad \forall p_w \]

where

\[ I_j(p_w) = \begin{cases} 1, x_j \in p_w \\ 0, \text{otherwise} \end{cases} \]

\[ REL(p_w) = p_w, \text{then perfectly reliable} \]

\( X = \) observed \( Y = \) forecast

\( p_w = p_w(y) = \) a probability window in the forecast

\( m = \) number of forecast/observation pairs
User-friendly reliability: example

GFS-EPP precipitation ensembles (w/o zero observed)

"Hit rate" = 90%

40% of time, observation should fall in window ±20%

"Underspread"
User-friendly Probability Score

Issue: Forecast errors not expressed in real terms (e.g., in CRPS)

Probability Score of Wilson et al. (1999)

- Integrate forecast in window, $w$, around $X$
- $w$ can be viewed as an ‘acceptable error’ in real terms (useful in operational hydrology)

$$PS(f(y), w) = \int_{x-0.5w}^{x+0.5w} f(y)dy$$
Wilson PS (although improper) is related to CRPS (proper) as follows:

\[
\text{CRPS} = PS - 2E[YF_Y(y)] + E[Y]
\]

Mean Capture Rate (MCR) Diagram

- Extends the PS to all windows, \( w \)
- Average over \( m \) forecasts
- Prob. of outcome being within given error window (\( w \))

\[
\text{MCR} (F_Y(y), w) = \frac{1}{m} \sum_{i=1}^{m} \left[ F_Y(x_i + 0.5 \ w) - F_Y(x_i - 0.5 \ w) \right] \ \forall \ w
\]
MCR Diagram: example

GFS-EPP precipitation ensembles (w/o zero observed)

“5% chance of exceeding ± 20mm error, on average.”
Currently Available Features

- Java tool with structured GUI
- Only verifies numerical time-series for now
- Supports flexible conditional verification
- Computes several key metrics

Status ...

- Released to several RFCs for experimental forecasting
- Enhancement underway (e.g., sampling uncertainty)
- Fully documented and freely available (but w/o support for non-RFC users)
Modified box plot of ensemble forecast errors against observed value.

Huntingdon.Huntingdon.Temperature.Cond_obs at lead hour 6 without date and value conditions.
Real-time (Prognostic) Verification
The aim is to estimate:

\[ F_{X|Y}(x|y,S) \]: conditional cdf of observed given ensemble forecast, based on historical analogs to live forecast;

\[ S = \text{additional conditioning information for analogs (e.g., QPF for conditioning flow forecasts)} \]

\[ X = \text{observed variable} \]

\[ Y = \{Z_1, \ldots, Z_m\} \text{ ensemble forecast made of } m \text{ members} \]

**Approach to estimating** \( F_{X|Y}(x|y,S) \):

- **Indicator regression (transformed X and Y)**
- **Expectation of an indicator variable of** \( X \) **at threshold** \( b_k \) **is equal to**

\[
\text{Prob}[X \leq b_k]
\]

\[
\text{Prob}[X \leq b_k|Y \leq b_k] = E[I(x,b_k)|I(y,b_k)=1]
\]
Real-time Verification: Informal Example

Temperature (°F)

- Live forecast (L)
- Analog forecasts (H): $\mu_H = \mu_L \pm 1.0°C$
- Analog observations

Forecast lead day

April 16-18, 2008
NCAR DTC Verification Workshop, Boulder CO
Two classes of verification at NWS

- **Diagnostic**: simpler metrics/tools (EVS)
- **Real-time**: new concept, being tested

Outstanding science issues

- Communicating ensemble/probabilistic verification information to non-experts
- Addressing sampling uncertainty in metrics
- Verification of multi-scale variables (e.g. streamflow)
- Non-stationarity (e.g. climate change)
- Timing and shape errors of hydrograph (potential approach: MODE?)
Collaborations

Thorpepx-Hydro project

- NCEP-OHD joint project
- Verification of meteorological and hydrologic ensembles

Other key collaborators

- University of Iowa (Allen Bradley)
- Iowa State University (Kristie Franz)
- University of California, Irvine (Soroosh Sorooshian et al.)
- Participants of HEPEX