The General Linear Model Based Hydrologic Post-processor for Ensemble Streamflow Predictions – Experimental Results Using MOPEX Datasets

D. Jiang, Q. Duan, Beijing Normal University, China
John Schaake, Consultant, USA

Introduction:

In an ensemble streamflow prediction (ESP) system, the reliability and accuracy of hydrologic predictions are negatively affected by a number of uncertainties. Uncertainty in the hydrologic simulations (i.e., the raw ensemble) is one of them. Here we employ a "General Linear Model (GLM)" to post-process hydrologic simulations from MOPEX data archive. We conducted a number of experiments using the GLM postprocessor according the instructions given by the workshop organizer. The GLM Post-Processor is show to be able to remove the mean bias when applied to hydrologic model simulations and produce reliable ensemble predictions.

The General Linear Model Methodology

Let Z1 be the observations for the forecast period, Z2 the predictor vector, which contains simulated and observed streamflow for the analysis period, and simulated streamflow for the forecast period :

$$Z_1 = \begin{bmatrix} q_{obs,1} \\ \vdots \\ q_{obs,n_f} \\ \vdots \\ q_{obs,N_f} \end{bmatrix} \qquad Z_2 = \begin{bmatrix} q_{sim,n_f} \\ q_{obs,n_f} \\ q_{sim,n_f} \end{bmatrix} = \begin{bmatrix} q_{sim,1} \\ \vdots \\ q_{sim,N_f} \\ q_{obs,-N_a} \\ \vdots \\ q_{obs,-1} \\ q_{sim,-N_a} \\ \vdots \\ q_{sim,-1} \end{bmatrix}$$

Further, let Z1,2 be the predictand given the predictor vector, i.e., Z1,2 = Z1 | Z2. The GLMPP model can be expressed as:

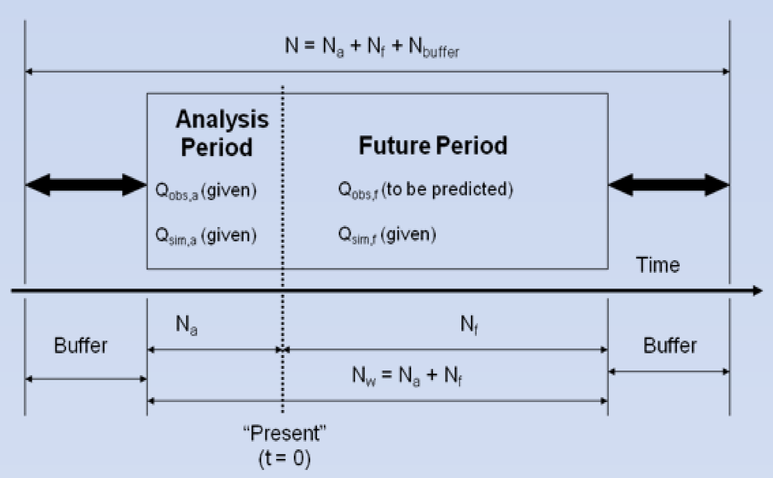
$$Z_{1,2} = A \cdot Z_2 + B \cdot E \qquad E=N(0,1)$$

Denote $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$. The covariance matrix for Z is: $\Sigma = \begin{bmatrix} \Sigma_{11}\Sigma_{12} \\ \Sigma_{21}\Sigma_{22} \end{bmatrix}$

So Z1,2 can be solved analytically (Valencia and Schaake, 1973) with

$$A = \Sigma_{12} \cdot \Sigma_{22}^{-1} \qquad BB^T = \Sigma_{11} - \Sigma_{12} \cdot \Sigma_{22}^{-1} \cdot \Sigma_{21}$$

A schematic of the data window



Because there are typically only a few historical years of data available to calibrate the GLMPP, we use a buffer period to increase the number of samples and help to calibrate the GLM's parameters.

GLM Properties:

(1) it preserves the "skill" of the raw ensemble forecast;

(2) it removes mean bias;

(3) it produces an ensemble of members that represent in an "equally-likely" sense the observed hydrograph that is being predicted;

(4) it preserves temporal scale dependency relationships in streamflow hydrographs and the uncertainty in the predictions;

(5) it produces ensemble predictions of future streamflow events that have very nearly the same climatology as the corresponding streamflow observations.

Data :

The data used to test the GLM are **daily** data produced by the Model Parameter Estimation Experiment (MOPEX) Tucson Workshop.
The MOPEX data set contains streamflow discharge simulations and corresponding observations from **12** Southeastern US basins generated by **6** different hydrological models. The simulation period covers **36** years, starting on **January 1, 1962** and ending on **December 31,1997.**

In the experiments, we used **two different data sets** to test the GLM. One set of simulations were done using a priori model parameter estimates (**apr**), while the other set with calibrated model parameters (**cal**).

Preliminary Results:

In this study, we designed various experiments under different conditions, including the variation of analysis period , forecast period, buffer period, calibrate period and verification period.

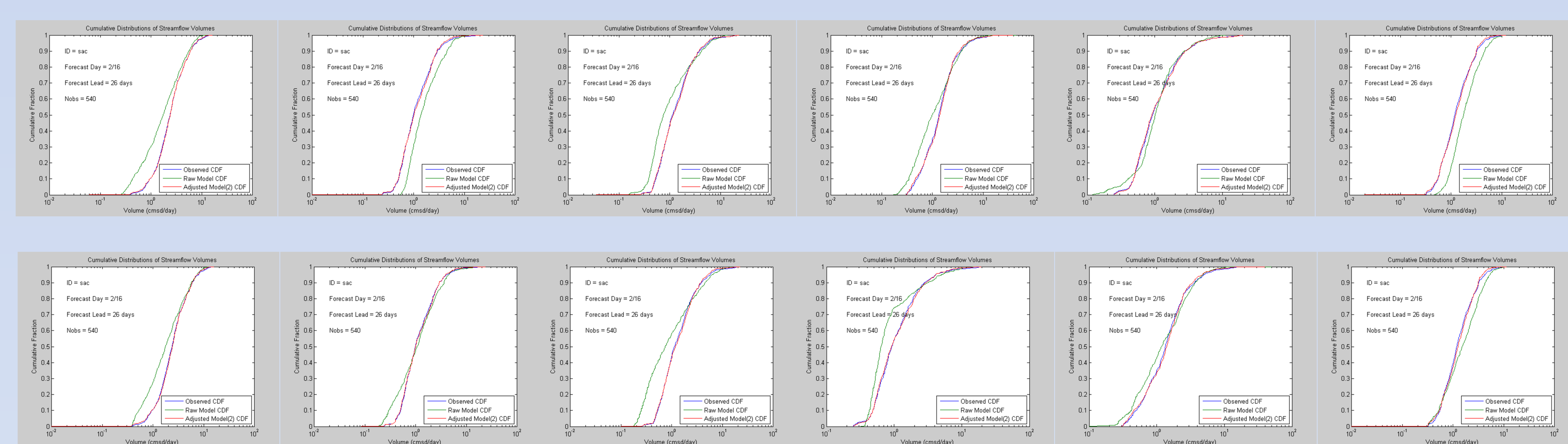
Most results show that the adjusted ensemble is much better than the raw ensemble and the GLM can reduce the mean bias in the streamflow simulations in both the calibration and verification periods.
And most of the results using the cal data set are better than those using the apr data set.
The figures below are partial results of the experiments.

(In each experiment, the upper figures are produced by using the **apr** data set; the lower figures are produced by using the **cal** data set.)

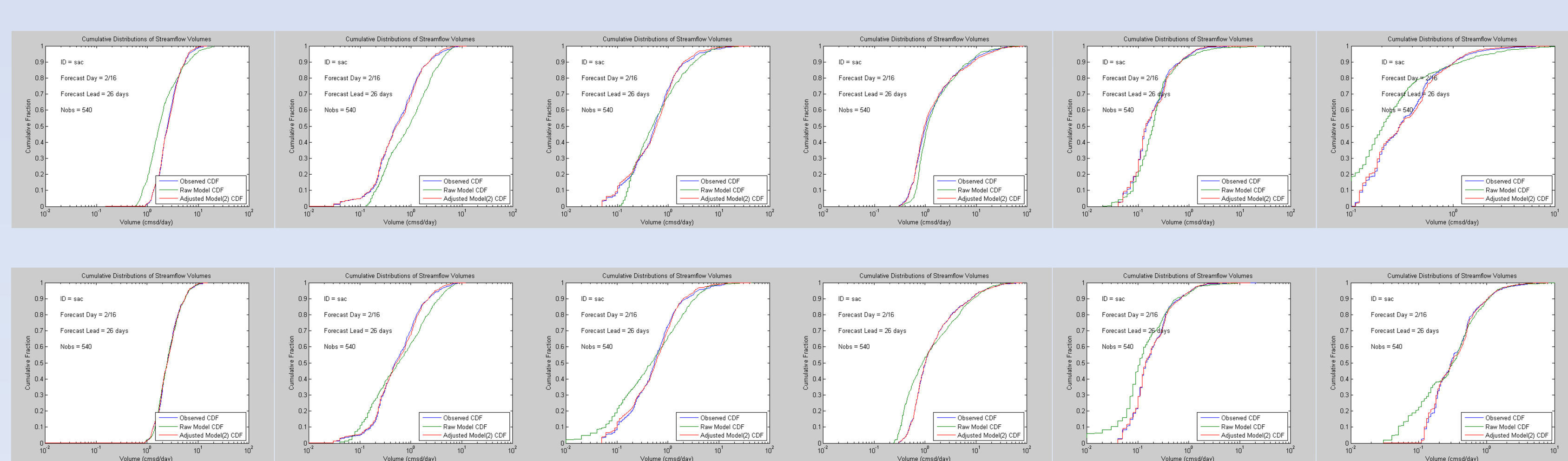
Case 1:

Forecast Day: Feb 16 na:30 days nf:30 days buffer:15 days
Model : SAC Basins: all
Length of period to calibrate the GLM parameters: years 1-36
Length of verification period: 0 years

3054500 1068500 1643000 168000 3179000 3364000



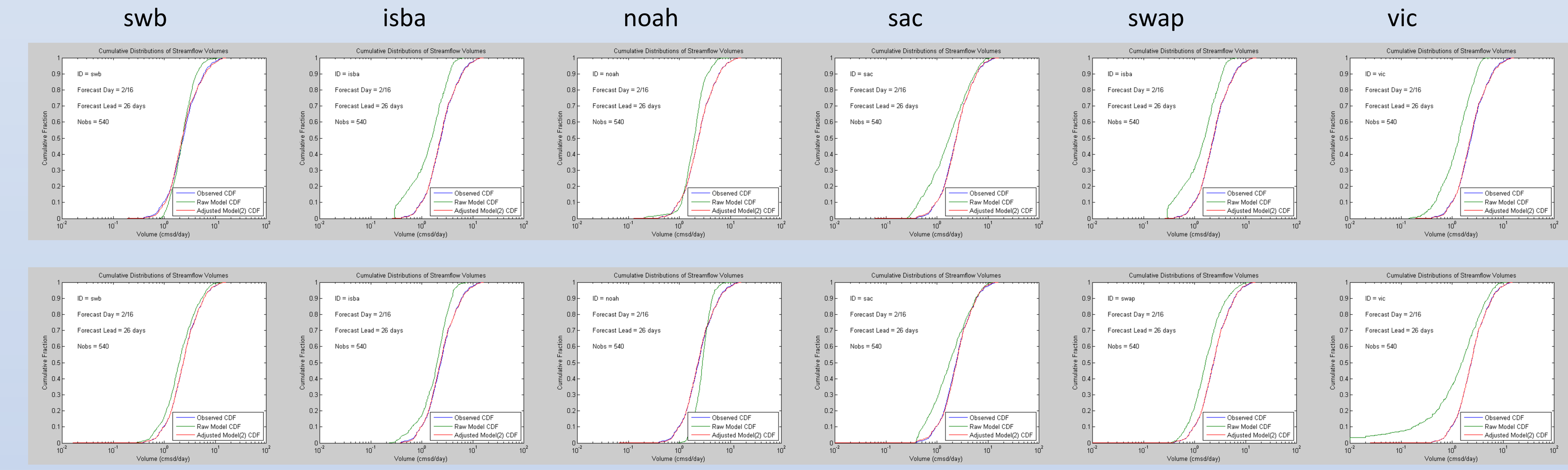
3451500 5455500 7186000 7378500 8167500 8172000



Case 2:

Forecast Day: Feb 16 na:30 days nf:30 days buffer:15 days
Models : all Basins: 3054500
Length of period to calibrate the GLM parameters: 36 years
Length of verification period: 36 years

swb isba noah sac swap vic



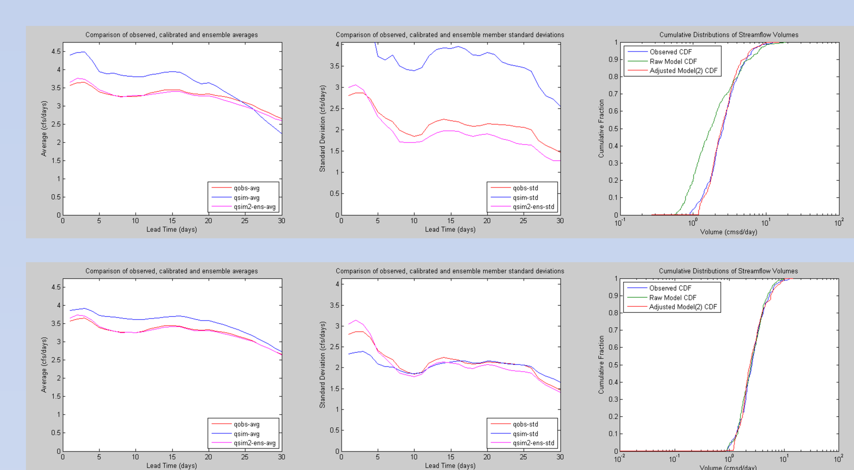
Case 3:

Forecast Day: Feb 16 na:30 days nf:30 days buffer:15 days
Model : SAC Basins: 3451500

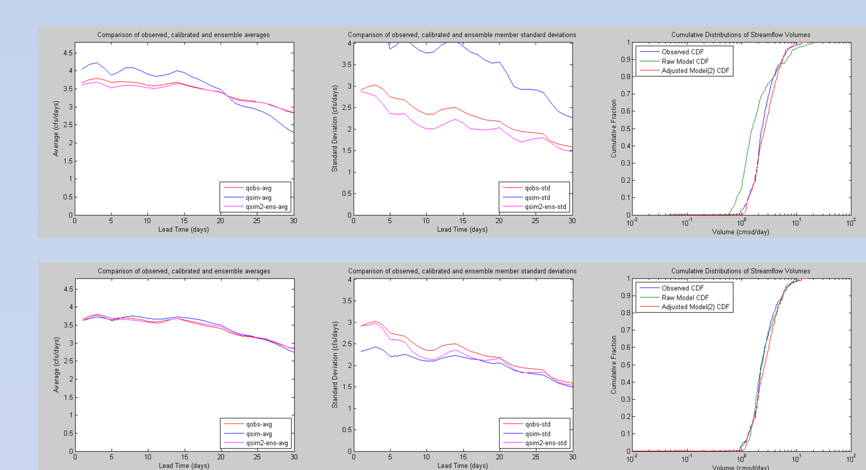
Length of calibrate period: years 1-18
Length of verification period: years 19-36

Length of calibrate period: odd years
Length of verification period: even years

Average Standard Deviation CDF



Average Standard Deviation CDF



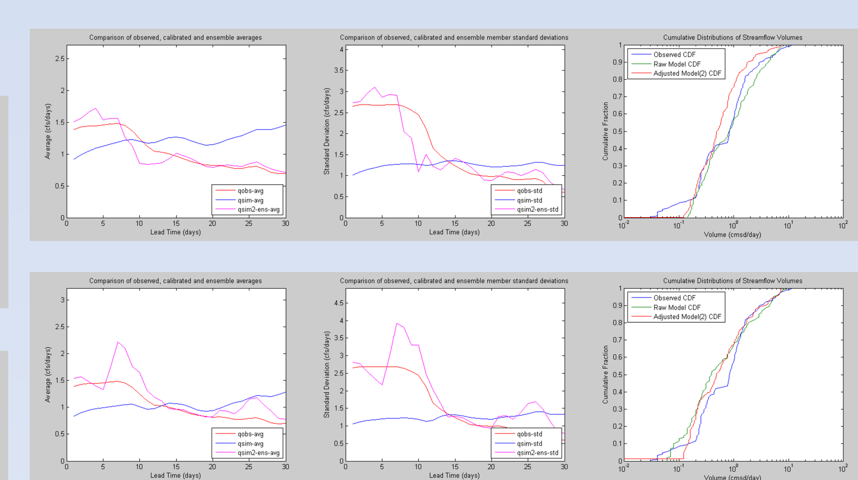
Case 4:

Forecast Day: Feb 16 na:30 days nf:30 days buffer:15 days
Model : SAC Basins: 5455500

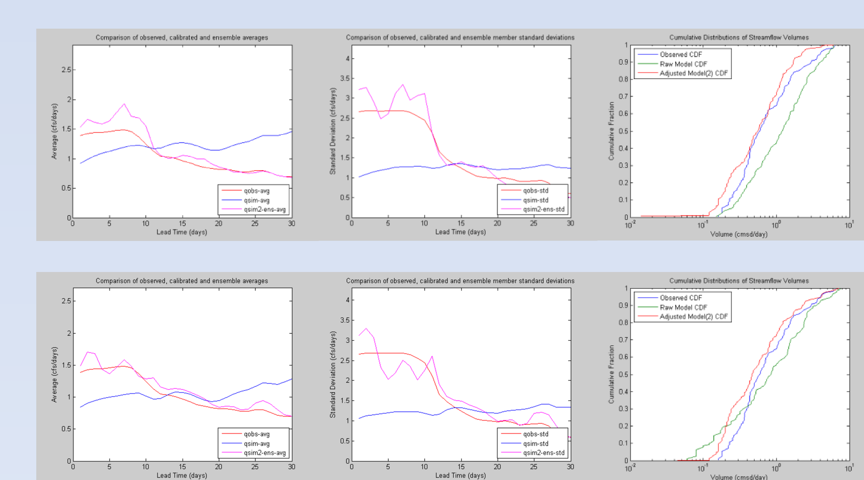
Length of calibrate period: years 1-9
Length of verification period: years 10-18

Length of calibrate period: years 1-9
Length of verification period: years 19-27

Average Standard Deviation CDF



Average Standard Deviation CDF



Contact: Qingyun Duan, qyduan@bnu.edu.cn