Title
A sequential Bayesian approach for hydrologic model selection and prediction

Journal
Water Resources Research, 45(12)

ISSN
0043-1397

Authors
Hsu, Kuo-lin
Moradkhani, Hamid
Sorooshian, Soroosh

Publication Date
2009-12-01

DOI
10.1029/2008wr006824

A sequential Bayesian approach for hydrologic model selection and prediction

Kuo-lin Hsu,[1] Hamid Moradkhani,[2] and Soroosh Sorooshian[1]

Received 10 January 2008; revised 28 July 2008; accepted 7 October 2008; published 22 January 2009.

[1] When a single model is used for hydrologic prediction, it must be capable of estimating system behavior accurately at all times. Multiple-model approaches integrate several model behaviors and, when effective, they can provide better estimates than that of any single model alone. This paper discusses a sequential model fusion strategy that uses the Bayes rule. This approach calculates each model's transient posterior distribution at each time when a new observation is available and merges all model estimates on the basis of each model's posterior probability. This paper demonstrates the feasibility of this approach through case studies that fuse three hydrologic models, auto regressive with exogenous inputs, Sacramento soil moisture accounting, and artificial neural network models, to predict daily watershed streamflow.

Citation: Hsu, K.-L., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resour. Res.*, *45*, W00B12, doi:10.1029/2008WR006824.

1. Introduction

[2] Engineers often use hydrologic models to describe a range of issues related to water resources planning and management, such as watershed runoff generation, reservoir systems operation, groundwater development and protection, and water distribution systems [*Chow et al.*, 1988; *Frevert and Singh*, 2002]. In the past few decades, the rapid development of computer capability and extensive data from ground observations and remote sensors have permitted the development of various model types, from "black box" to physically based models, and from "lumped" to distributed models, to solve real-world hydrologic problems. However, users select models primarily on the basis of subjective preference. A hydrologic model is a simplified presentation of a real-world processes and it may accurately describe certain processes but overlook others. Many studies have shown that predictions based on a single model tend to underestimate total predictive uncertainty [*Barnston et al.*, 2003; *Krishnamurti et al.*, 1999]. Others have concluded that effective combination of multiple models may provide more skillful predictions than a single model does [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Georgakakos et al.*, 2004].

[3] In the literature, many methods have been used to merge multiple data sources in hydrologic modeling. Arithmetic mean (AM) is a simple and widely used method; another popular method is weighted averaging (WA), which calculates each model's optimal weighting parameters from the inverse of the error variance of each model's estimates [*Daley*, 1991]. The WA method has been applied to many hydrologic and atmospheric science problems, such as merging precipitation estimates from multiple sources (gauges, radar, satellite, and numerical weather prediction models) [*Huffman et al.*, 1995; *Xie and Arkin*, 1996]. Other than AM and WA, some nonlinear and rule-based data fusion methods, such as fuzzy rule–based modeling and artificial neural networks (ANN), have been proposed to integrate several runoff models [*Abrahart and See*, 2002; *Coulibaly et al.*, 2005; *See and Abrahart*, 2001; *Shamseldin*, 1997; *Shamseldin et al.*, 1997; *Xiong et al.*, 2001]. *Shamseldin et al.* [1997] applied ANN to combine five rainfall-runoff models, examining the resulting model on 11 watersheds and demonstrating that the ANN and WA merging methods outperform AM. *Xiong et al.* [2001] proposed a Takagi-Sugeno type of fuzzy rule–based model to combine several rainfall-runoff model outputs. This method separates flow into several ranges represented by linguistic rules that are weighted and combined to generate output according to their membership function. Their results show that the fuzzy rule–based model is as effective as the WA and ANN methods. Similarly, [*See and Abrahart*, 2001] tested ANN models as a data fusion tool to combine streamflow estimates from four models, including ANN, fuzzy logic, autoregressive moving average (ARMA), and persistence forecasts, and demonstrated that the ANN fusion methods outperform the AM methods. *Xiong et al.* [2001], *Shamseldin* [1997], and *Coulibaly et al.* [2005] also used WA to combine several streamflow forecasts from a nearest-neighbor method (NNM), ANN, and a conceptual hydrologic model, and showed that combined multimodel predictions are more reliable than estimates from a single model. Recently, the Bayesian model average (BMA) technique has been proposed for estimating multimodel prediction and improving prediction uncertainty [*Ajami et al.*, 2007; *Raftery et al.*, 1997; *Vrugt and Robinson*, 2007]. For each model, the BMA method assigns an a priori distribution and finds a combined likelihood. The posterior distribution of each model's weighting parameter (level of participation) is cal-

[1] Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, Irvine, California, USA.
[2] Department of Civil and Environmental Engineering, Portland State University, Portland, Oregon, USA.

0043-1397/09/2008WR006824

culated using the expectation maximum (EM) or Markov chain Monte Carlo (MCMC) methods [*Gelman*, 2004; *Robert and Casella*, 2004]. A review of data fusion methods in the integration of hydrologic models can be found in work by *See* [2008].

[4] The multimodel combination strategies discussed above rely on finding a set of time-invariant weighting parameters that are assumed to be normally distributed. However, various hydrologic modeling approaches have shown that some models perform better for a certain portion of the hydrograph (e.g., wet seasons or high-flow periods), while others perform better for other portions of the hydrograph (e.g., dry seasons or low-flow scenarios). Although assigning fixed weights to models may improve a combined model's performance over that of a single model, fixed parameters do not provide the flexibility to assign higher weights to models that perform better for a particular simulation period. To resolve time variations that occur when merging model parameters, *Ajami et al.* [2007] assigned model weights on the basis of several flow categories, which could be determined by calendar month or flow range. In their approach the combined model selects the weighting parameters from the fixed pretrained parameter sets.

[5] This study used a sequential updating approach to blend multiple models. The proposed approach, the Bayesian combined estimator, is a probabilistically based predictor consisting of adjustable weights, in the form of individual models' posterior probabilities that are applied to model estimates at each prediction time and are updated sequentially. The multimodel estimate can be selected on the basis of each model's posterior probability, which may change over time and is estimated sequentially, resulting in improved prediction accuracy [*Petridis and Kehagias*, 1998; *Petridis et al.*, 2001].

[6] This paper's scope is as follows: section 2 describes Bayes rule and explains how to find a model's posterior probability from observations and its a priori distribution. Section 3 discusses streamflow prediction using multiple models, presenting the criteria necessary to evaluate model performance and the basic models that are used to simulate catchment rainfall-runoff processes. Section 4 provides case studies for combining several models, including auto regressive with exogenous (ARX) inputs, Sacramento soil moisture accounting (SAC-SMA), and ANN for predicting watershed runoff. Conclusions and future directions are provided in section 5.

2. Bayesian Multiple-Model Approach

[7] Bayesian inference updates the probability that a hypothesis may be true by referring to evidence or other supportive information. The posterior distribution of the quantity of interest is calculated from a set of observations and their a priori distribution. Consider that k models, $\boldsymbol{M} = \{M_1, \ldots, M_k\}$ are used to predict a quantity of interest, and data set $D = \{d_1, d_2, \ldots, d_t\}$ is used for model calibration. Our objective is to find the most plausible model, M_p, given the available observations D. The posterior distribution of a model, M_j, is calculated from Bayes' rule as

$$p(M_j|D) = \frac{p(D|M_j)p(M_j)}{p(D)}, \tag{1}$$

where $p(D|M_j)$ is the likelihood function of the model M_j, $p(M_j)$ is the a priori probability of the model M_j, $j = 1, 2, \ldots, k$, which satisfies $\sum_{j=1}^{k} p(M_j) = 1.0$, and $p(D) = \sum_{j=1}^{k} p(D|M_j)p(M_j)$ is the total probability of model j given by a set of data D. Using the models' posterior probability, we can calculate the estimate for the combined model. Let y denote the quantity of interest after combining k model outputs. The posterior probability of y given the multimodel combination is written as

$$p(y|M_1, M_2, \ldots, M_k, D) = \sum_{j=1}^{k} p(y|M_j, D)p(M_j|D). \tag{2}$$

The representation above is the marginal or integrated likelihood estimates of all models, M_j, $j = 1\ldots, k$. For the given set of models, the combined estimates rely on specifying the prior distribution of models $p(M_j)$ and the likelihood of each model, $p(D|M_j)$. In the absence of knowledge of model uncertainty, a uniform a priori is usually considered: that is, $p(M_j) = 1/k$, $j = 1,\ldots, k$. With uniform a priori, the model's posterior probability is essentially decided by the likelihood function. If the likelihood estimate of model M_j is assumed to be normally distributed with an error standard deviation of σ_j, it can be written as

$$p_j(D|M_j) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t(M_j)\right)^2}{2\sigma_j^2}\right), \tag{3}$$

where y_t is observation at time step t, n is the size of data; $\hat{y}_t(M_j)$ is the jth model estimate at time t; and $\varepsilon_t^j = y_t - \hat{y}_t(M_j)$ is the difference between the actual observation and jth model estimate at time t.

[8] Equation (3) is used to assign a higher probability to a lower square error. By substituting (3) into (1), the posterior probability of jth model can be calculated as

$$\begin{aligned} p(M_j|D) &= \frac{p(D|M_j)p(M_j)}{\sum_{i=1}^{k} p(D|M_i)p(M_i)} \\ &= \frac{\frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t(M_j)\right)^2}{2\sigma_j^2}\right) \cdot p(M_j)}{\sum_{i=1}^{k} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t(M_i)\right)^2}{2\sigma_i^2}\right) \cdot p(M_i)}. \end{aligned} \tag{4}$$

As shown in equation (2), all models are used to calculate the probability of the combined model's output. Each model is selected according to its posterior probability.

[9] In equation (4), the posterior probability of the jth model is estimated by processing a batch of observations (D) simultaneously. The limitation of batch calculation is that each model's posterior probability is fixed for the whole simulation period. However, if certain models give better predictions than others for a certain portion of the process, using them may generate a better solution than using a fixed weighting factor. Therefore, implementing the model sequentially provides a flexible framework by which to update models' posterior probability using newly available observations.

[10] Similar to equation (4), Bayes rule can be expressed in recursive form [*Moradkhani et al.*, 2005; *Petridis et al.*, 2001]. Let y_t be the observation of predictive variable at time t and $p_{t-1}(M_j)$ be the model prior distribution of the jth model at time t, given the measurement up to $t-1$, then $p_{t-1}(M_j)$ can be written as

$$p_{t-1}(M_j) = p(M_j|d_1, d_2, \ldots d_{t-1}). \quad (5)$$

The model posterior distribution in recursive form is represented as

$$p_t(M_j) = \frac{p(d_t|M_j) \cdot p_{t-1}(M_j)}{\sum_{i=1}^{k} p(d_t|M_i) \cdot p_{t-1}(M_i)}. \quad (6)$$

Assuming the likelihood estimate of model M_j at time t being a normal distribution with zero mean, it can be expressed as

$$p_j(d_t|M_j) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(y_t - \hat{y}_t(M_j))^2}{2\sigma_j^2}\right). \quad (7)$$

Therefore, the posterior probability of the jth model is calculated as

$$p_t(M_j) = \frac{\frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(y_t - \hat{y}_t(M_j))^2}{2\sigma_j^2}\right) \cdot p_{t-1}(M_j)}{\sum_{i=1}^{k} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(y_t - \hat{y}_t(M_i))^2}{2\sigma_i^2}\right) \cdot p_{t-1}(M_i)}. \quad (8)$$

As shown above, the posterior probability of the jth model at time t is determined by several factors: (1) The temporal error of the jth model output (i.e., $\varepsilon_t^j = y_t - \hat{y}_t(M_j)$) (when the temporal error of model estimate is high, the posterior probability of the jth model ($p_t(M_j)$) decreases) and (2) the a priori probability of the model (i.e., $p_{t-1}(M_j)$) (the model's posterior probability is positive proportional to its a priori probability).

[11] The model's posterior probability is calculated using the sequential Bayes's rule (equation (8)) and is recursively adjusted on the basis of available observations. The combined model estimate is

$$E(y_t|M_1, M_2, \ldots, M_k, D) = \sum_{j=1}^{k} E(y_t|M_j, D)p_t(M_j|D), \quad (9)$$

where $p_t(M_j|D)$ is the probability of the jth model; $E(y_t|M_j, D)$ is the expected value of the model M_j estimate, and $E(y_t|M_1, M_2, \ldots, M_k, D)$ is the expected value of the combined model estimate.

3. Fusion of Multiple Models for Streamflow Forecasting

3.1. Rainfall-Runoff Models

[12] The process of transforming rainfall to runoff in a watershed is very complex, and is influenced by factors such as rainfall distribution, soil characteristics, physiographic properties, and groundwater storage. Rain falling over a watershed may travel through many alternative paths to become flow in a river. If the rainfall intensity is strong enough, part of the rainfall may give rise to overland flow that travels quickly to the river. Other portions, however, may be detained on the surface or may infiltrate into the ground, taking substantially longer to reach the river. This complex, nonlinear process is not easily described by a simple model [*Chow et al.*, 1988; *Linsley et al.*, 1982]. Many studies have addressed rainfall-runoff (R-R) process modeling, and can provide a summary of model representation, identification, and calibration [*Beven*, 2001; *Duan*, 2003; *Frevert and Singh*, 2002].

[13] Although many models have been applied to simulate the R-R process, the process's complex nature makes it difficult for any one model to represent the process well at all times. This study explores ways to combine multiple models to improve the prediction of hydrologic processes. Several models, including SAC-SMA, ARX, and ANN models, were tested. A brief description of these models is provided below.

3.1.1. SAC-SMA Model

[14] The SAC-SMA model is a conceptual multistorage streamflow simulation model, developed and maintained by the U.S. National Weather Service [*Burnash et al.*, 1973]. The inputs to the model include mean basin precipitation at the current time interval, $r(t)$, and mean basin evapotranspiration. A description of the SAC-SMA model and the way it is calibrated by the shuffled complex evolution (SCE-UA) algorithm [*Duan et al.*, 1992] can be found in [*Sorooshian et al.*, 1993]. Discussions related to conceptual rainfall-runoff model identification and use can be found in work by *Boyle et al.* [2000, 2001], *Gupta et al.* [1998], and *Sorooshian et al.* [1993].

3.1.2. ARX Model

[15] The ARX model is a linear dynamic function that has been used extensively for predicting streamflow using observed rainfall and runoff sequences [*Wood and Szöllösi-Nagy*, 1980]. A model of ARX(n_1, n_2) is represented as

$$y_{t+1} = \sum_{i=0}^{n_1} a_i y_{t-i} + \sum_{j=0}^{n_2} b_j r_{t-j} + \varepsilon_{t+1}, \quad (10)$$

where a_i and b_j are parameters, and $y(t)$ and $r(t)$ are the observed streamflow and rainfall sequences, respectively. The time unit t is 1 day, and ε_{t+1} is the error of streamflow estimation. The case study uses three previous time intervals of rainfall and streamflow observations as the inputs to the model (i.e., $n_1 = n_2 = 2$).

3.1.3. Self-Organizing Linear Output Model

[16] ANN models have been found capable of modeling nonlinear systems, and have been applied to solve many hydrologic problems, such as runoff prediction and precipitation estimation from remote sensing measurements [*Abramowitz et al.*, 2006; *Hong et al.*, 2005; *Hsu et al.*, 1995; *Hsu et al.*, 1999; *Moradkhani et al.*, 2004]. Additional ANN applications have been discussed in several review articles [*Govindaraju and Rao*, 2000; *Maier and Dandy*, 2000]. The case study applies a special ANN model, named self-organizing linear output (SOLO), to the R-R process [*Hsu et al.*, 2002]. The SOLO model consists of a classification layer based on a self-organizing feature map (SOM) and a group of piecewise linear principal component regression functions to fit different portions of

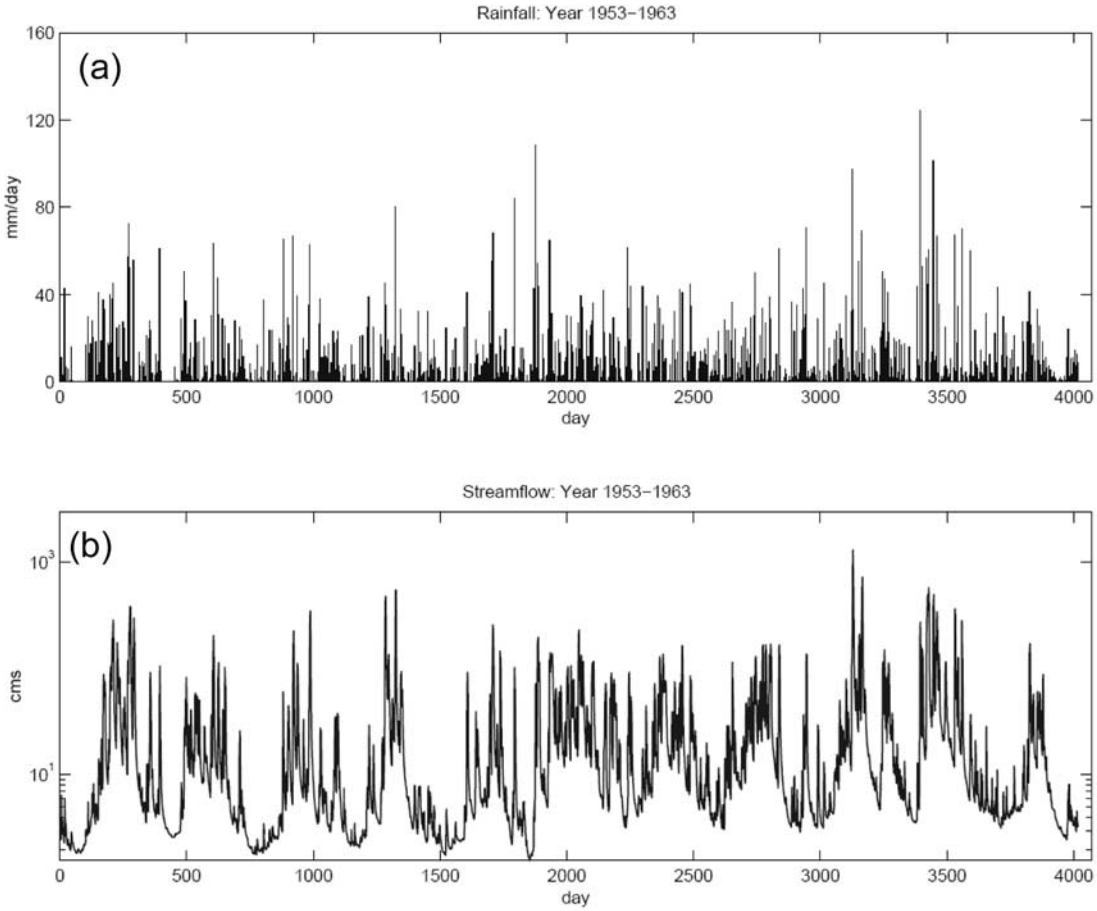

Figure 1. Daily rainfall-runoff time series used in the model calibration: (a) rainfall and (b) streamflow.

the hydrologic process. A previous study [*Hsu et al.*, 2002] has shown that SOLO is very efficient and effective in mapping the rainfall-runoff process. In the case study of this article, input variables include $x_t = [y(t), y(t-1), y(t-2), r(t), r(t-1), r(-2t)]$, which is consistent with the variables set in the ARX model. The size of the SOM network was set to 15×15. A detailed description of the SOLO architecture and model training can be found in work by *Hsu et al.* [2002].

3.2. Study Area and Data Coverage

[17] The selection of test basin with reliable data set is very important to the case study. During the test period, the watershed should cover a long enough period of data record to show a variety of flow patterns during the wet and dry periods. The Leaf River basin, consisting of the above data and hydrologic characteristics, is selected for this study. From the literature, Leaf River basin has been investigated by several researchers using conceptual hydrological models (such as SAC-SMA) and ANN models [*Boyle et al.*, 2000; *Duan et al.*, 1992; *Hsu et al.*, 2002; *Sorooshian et al.*, 1993; *Yapo et al.*, 1998]. The drainage area of the watershed is around 1948 km^2 and the data set includes daily streamflow, precipitation, and potential evapotranspiration estimates. For more detail description of the basin and data, readers can refer to the publications from *Brazil* [1988] and *Sorooshian and Gupta* [1983]. In this study, 36 years (1953∼1988) of data were selected for model calibration and evaluation. With respect to the data record length used in calibrating the model, *Yapo et al.* [1996] have suggested that more than 8 years of data are required to obtain stable parameters. The case study used the first 11 years of data for model calibration (see Figure 1), while the other 25 years of data were used to evaluate performance. Model comparisons are based on their ability to accurately predict streamflow 1 day ahead.

3.3. Fusion of Multiple Models

[18] In the multimodel approach, perhaps the most important (and rather difficult) step is model adaptation: the process of selecting a model in real time, when the data become available sequentially. Here we explore four options for combining models. Each model's importance in model integration is determined by a positive weight (probability), where the summation of all model weights is expected to be a unity. As expressed in equation (9), the higher a model's probability, the higher would be its degree of participation in calculating the combined estimate. If a model's weighting factor ($p_t(M_j)$) is equal to zero, the model makes no contribution to the combined estimation; likewise, if a model has a weighting factor of 1, it contributes fully in generating the combined estimate. This study considered several model combination strategies, as listed below.

[19] 1. The arithmetic mean (AM) approach treats all models equally. The probability assigned to each model is $(p_t(M_j|D) = 1/k$ for all models, where k is the number of models being used in the model combination. Our study used three basic models, where $p_t(M_j|D)$ is set to 1/3, for all models: $j \in \{1,2,3\}$.

[20] 2. In the weighted average (WA) approach the probability of each model used in the combination is calculated on the basis of the model's performance in the calibration phase. This probability is assigned according to the model's error variance as

$$p(M_j|D) = \left(1/\sigma^2_{\varepsilon j}\right) / \left(\sum_{i=1}^{k} 1/\sigma^2_{\varepsilon i}\right), \tag{11}$$

where $\sigma^2_{\varepsilon j}$ is the error variance of jth model, which is calculated after model calibration. In our case, the calibration data contains 11 years of the daily rainfall-runoff time series. This probability is a fixed value that does not change over time.

[21] 3. The sequential Bayesian model combination's (SBC) posterior probability is calculated on the basis of the Bayes rule of equation (8), while the combined model estimate is listed in equation (9). The model's posterior probability is adjusted recursively on the basis of newly available observations. Consider that only three models were included in both model calibration and validation, the models' initial priori probability is set to: $p_{t=0}(M_j) = 1/3$, for all models j. In the forecasting mode, because observations at time interval $t + 1$ are not yet available, instead of using the a posteriori probabilities of model estimates at time $t + 1$, the posterior probability at time t is used.

[22] 4. The sequential maximum a posteriori probability model selection (SMAP) setting selects only one model at each time interval. The candidate model, j^*, is selected according to the maximum a posteriori probability, i.e., $j^* = \arg \underset{j=1\cdots k}{\text{Max}}\{p_{t+1}(M_j)\}$. This setting enables us to select one model with the best performance at each time interval. The combined model estimate at time $t + 1$ is simply assigned as $\hat{y}_{t+1} = \hat{y}_{t+1}(M_{j^*})$. Again, because observations at forecasting time interval $t + 1$ were not yet available, the posterior probability at time t is used to select the best model at time $t + 1$. When the time interval progresses to $t + 1$, an observation is obtained and each model's posterior probability at $t + 1$ is updated.

3.4. Evaluation Statistics

[23] Five statistics frequently used in model evaluation were selected. They are bias (BIAS), root-mean-square error (RMSE), correlation coefficient (CC), Nash-Sutcliffe efficiency coefficient (NSE), and mean absolute error for flow > 200 cubic meters per second daily (cmsd) (MAE-Q200). The first two statistics, BIAS and RMSE are related to the mean and variance of residual, while CC and NSE are nondimensional coefficients. All the above four criteria evaluate the statistics for the whole calibration and validation time series of model estimates and observations. The MAE-Q200, on the other hand, is a threshold, which only evaluate model performance for observed flow greater than 200 cmsd. *Dawson et al.* [2007] provides a list of generalized metrics for model evaluation. More information about the generalized evaluation matrices can be found from HydroTest Web page, http://www.hydrotest.org.uk. Further description of the evaluation statistics are listed below.

[24] 1. The bias estimates (BIAS) statistics gives the mean of residual. The best value is zero; positive value means model overestimates, while negative value infers model underestimated:

$$BIAS = \sum_{t=1}^{n} (\hat{y}_t - y_t)/n. \tag{12}$$

[25] 2. The root-mean-square error (RMSE) statistics calculate the variance of the residual. The RMSE is always positive; the best value is zero; the higher the value, the poor the model performance:

$$RMSE = \sqrt{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2/n - 1}. \tag{13}$$

[26] 3. The correlation coefficient (CC): This is a dimensionless index which evaluates the linear relationship between the observed and model estimated flows. The value is in [−1 1]. The best value is 1.0:

$$CORR = \sum_{t=1}^{n} (y_t - y_m)(\hat{y}_t - \hat{y}_m) \Big/ \sqrt{\sum_{t=1}^{n} (y_t - y_m)^2 \sum_{t=1}^{n} (\hat{y}_t - \hat{y}_m)^2}. \tag{14}$$

[27] 4. The coefficient of Nash-Sutcliffe efficiency (NSE) index mainly refers model behavior to the mean value of the reference data. The value 1.0 is the best, while zero means that the model performs no better than the mean value of reference data:

$$NSE = 1 - \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 / \sum_{t=1}^{n} (y_t - y_m)^2. \tag{15}$$

[28] 5. The mean absolute error for flow $y_t >= 200$ cmsd(MAE(Q >= 200)) index calculates the absolute error of observed flow crossover a threshold. The threshold value can be assigned to different value to emphasize on different flow ranges (for example, ThQ >= 200 cmsd for high flow, and ThQ <= 10 cmsd for low flow). In this study, we select flow greater than threshold thdQ = 200 cmsd. This index is positive and the best value is zero:

$$MAE(Q >= 200) = \frac{1}{n_s} \sum_{\forall t, y_t \geq 200} |y_t - \hat{y}_t|. \tag{16}$$

In above equations, y_t is observed data at time t, $\hat{y}_t$ is the model estimate at time t, y_m is the mean value of observed data, and $\hat{y}_m$ is the mean value of multimodel estimation. For the sample size, n is the size of data, and n_s is the data sample size for $y_t >= 200$ cmsd.

4. Case Studies

[29] Two cases are provided. The first case includes the combination of three ARX models being trained from the low-, medium-, and high-flow hydrographs, and compares the result with the ARX, SAC-SMA and SOLO model estimates trained from full data range. The second case

Table 1. Evaluation Statistics for Single and Combined Models in Both the Calibration and Validation Data Periods[a]

	CAL					VAL				
Statistics	NSE	RMSE	CORR	BIAS	MAE (Q > 200)	NSE	RMSE	CORR	BIAS	MAE (Q > 200)
ARX-L	0.855	23.77	0.930	0.001	96.66	0.818	27.74	0.912	−0.01	91.23
ARX-M	0.839	24.99	0.948	0.117	123.33	0.822	27.43	0.934	1.574	115.92
ARX-H	0.914	18.31	0.956	−0.869	59.18	0.894	21.13	0.947	−0.787	57.42
AM	0.909	18.81	0.955	−0.250	83.29	0.887	21.79	0.943	0.259	78.19
WA	0.917	17.99	0.958	−0.378	76.95	0.896	20.90	0.947	0.025	72.32
SBC	0.955	13.17	0.979	0.342	46.67	0.942	15.64	0.971	0.526	47.87
SMAP	0.952	13.65	0.977	0.321	46.75	0.938	16.14	0.969	0.554	48.56
ARX	0.923	17.24	0.961	−0.105	66.69	0.900	20.53	0.949	0.036	65.66
SAC-SMA	0.918	17.79	0.959	−2.292	64.93	0.910	19.43	0.960	−5.444	56.37
SOLO	0.960	12.36	0.980	−0.075	44.44	0.931	17.01	0.965	−0.037	54.14

[a]Single models are ARX-L, ARX-M, ARX-H, ARX, SAC-SMA, SAC-SMA, and SOLO, and combined models are AM, WA, SBC, and SMAP.

demonstrates the combination of ARX, SAC-SMA, and ANN models in R-R modeling.

4.1. Case Study 1: Fusion ARX Models

[30] Three ARX linear models served as the basic models to simulate rainfall-runoff in the Leaf River Basin. These three models share the same ARX structure (i.e., $n_1 = n_2 = 2$), while the model parameters are calibrated separately from flows covered in three different ranges: low flow, medium flow, and high flow. Low flow indicates a daily flow in the range of 0–10 cm, while medium flow indicates a range 10–50 cm, and high flow indicates ranges higher than 50 cm. These three models, denoted as ARX-L, ARX-M, and ARX-H, are calibrated from objective functions set to fitting the models to low-flow, medium-flow, and high-flow observations, respectively.

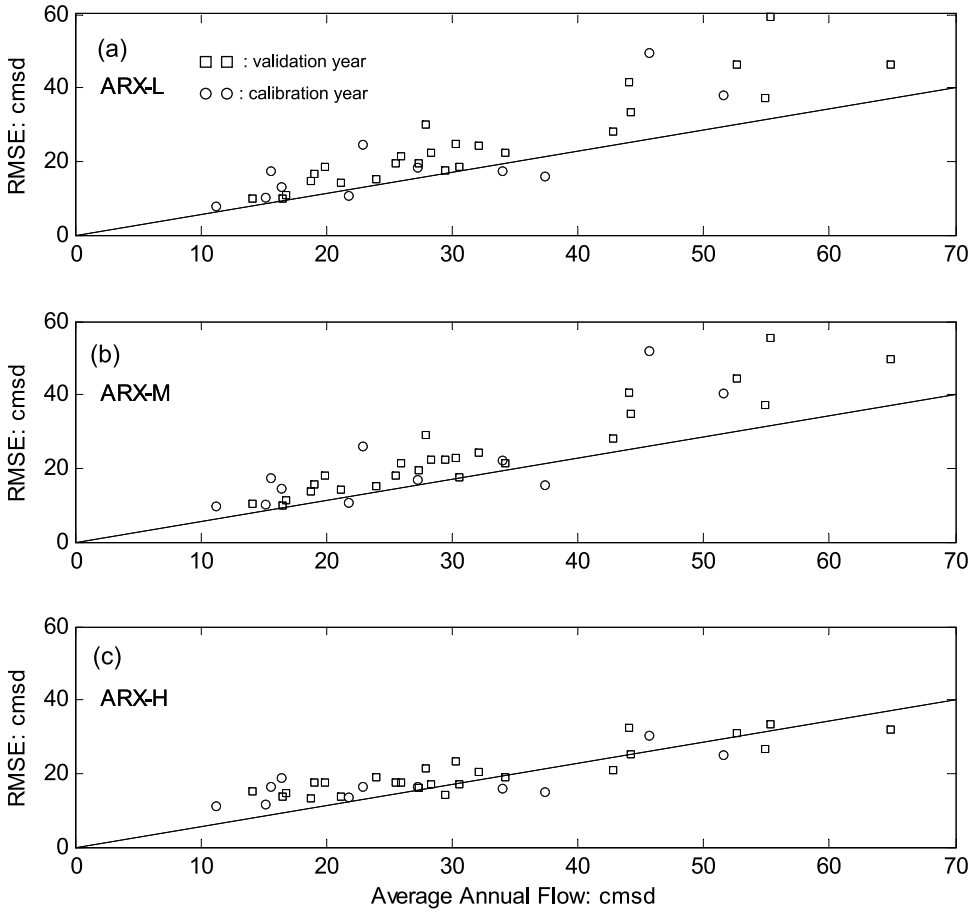

Figure 2. The root-mean-square error (RMSE) for testing models versus the average annual streamflow (cmsd) for each water year. Models are (a) ARX-L, (b) ARX-M, and (c) ARX-H. The circles represent the data covered in the calibration period, while the squares represent the data covered in the validation period.

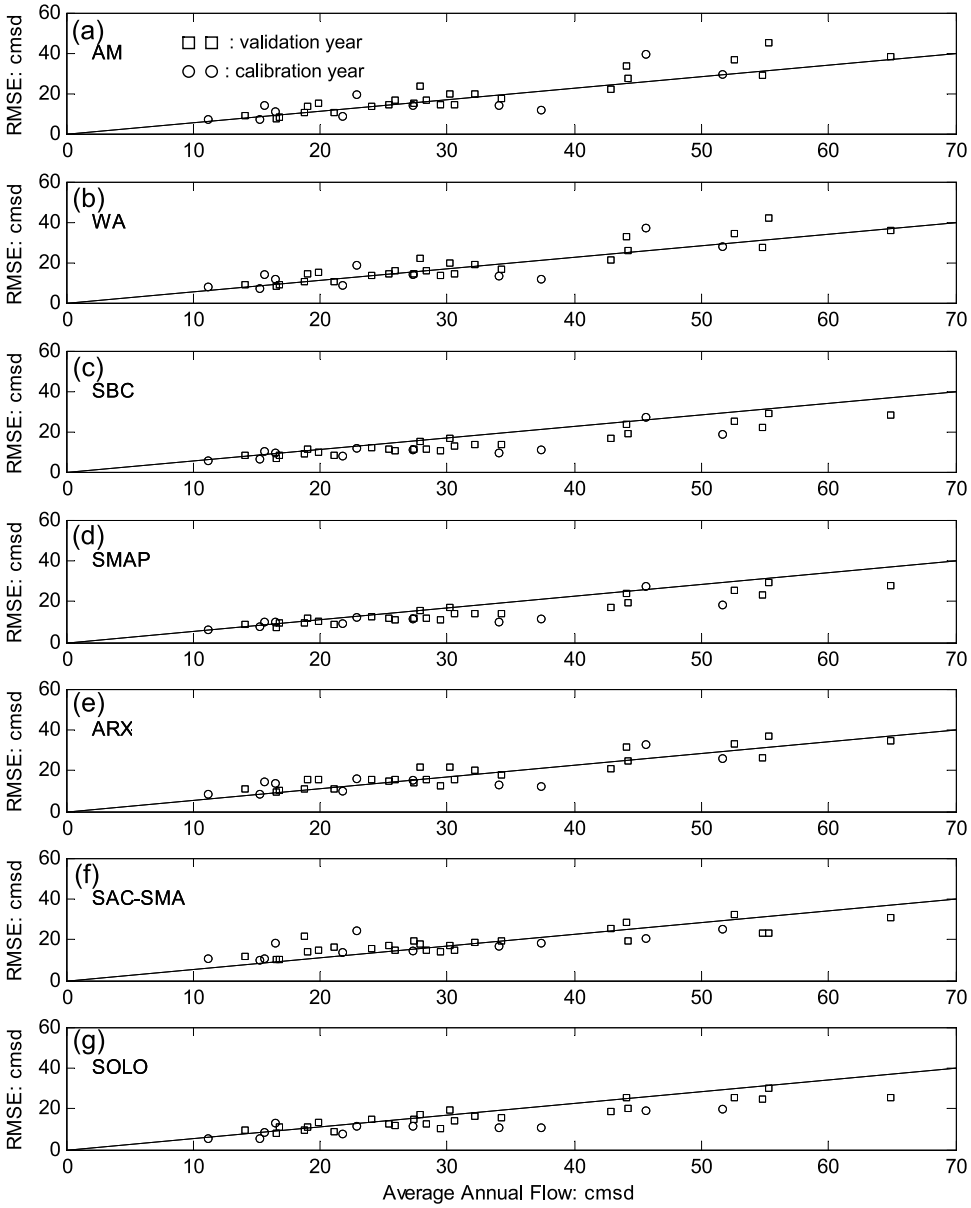

Figure 3. Similar to Figure 2 but for models (a) AM, (b) WA, (c) SBC, (d) SMAP, (e)ARX, (f) SAC-SMA, and (g) SOLO.

[31] For each specified flow range, the optimal model parameters, $\vec{\theta} = \{a_i, b_j\}_{i,j\in\{0,1,2\}}$ are found by minimizing the root-mean-square error (RMSE) of the streamflow residuals in the specified flow ranges, as depicted below:

$$ARX - L : \vec{\theta}^*_{ARX-L} = \arg\min_{\theta} F(\theta) = \min\left[\sum_t (y_t - \hat{y}_t)^2\right];$$

$$y_t \in q_L \in [0,\ 10]\text{cmsd},$$

$$ARX - M : \vec{\theta}^*_{ARX-M} = \arg\min_{\theta} F(\theta) = \min\left[\sum_t (y_t - \hat{y}_t)^2\right];$$

$$y_t \in q_M \in [10,\ 50]\text{cmsd},$$

$$ARX - H : \vec{\theta}^*_{ARX-H} = \arg\min_{\theta} F(\theta) = \min\left[\sum_t (y_t - \hat{y}_t)^2\right];$$

$$y_t \in q_H > 50\,\text{cmsd},$$

where $\vec{\theta}^*_i$ is the optimal parameters of model i; q_L, q_M, q_H are streamflows in the ranges designated as low, medium and high; and y_t and $\hat{y}_t$ are the observations and model estimates at time t.

[32] The case study applied the recursive Bayes rule to combine three linear ARX models for 1-day-ahead streamflow prediction. As described above, three ARX models (ARX-L, ARX-M, and ARX-H) were calibrated using 11 years of data with objective functions set for flows in three ranges (low flow: 0–10 cmsd; medium flow: 10–50 cmsd; and high flow: > 50 cmsd) After the models were calibrated, they were validated on the remaining 25 years of data. Because these linear models were calibrated in three flow ranges, it was expected that the ARX-L model would perform well in low-flow predictions, and similarly that the ARX-M and ARX-H models would forecast better in the medium- and high-flow ranges.

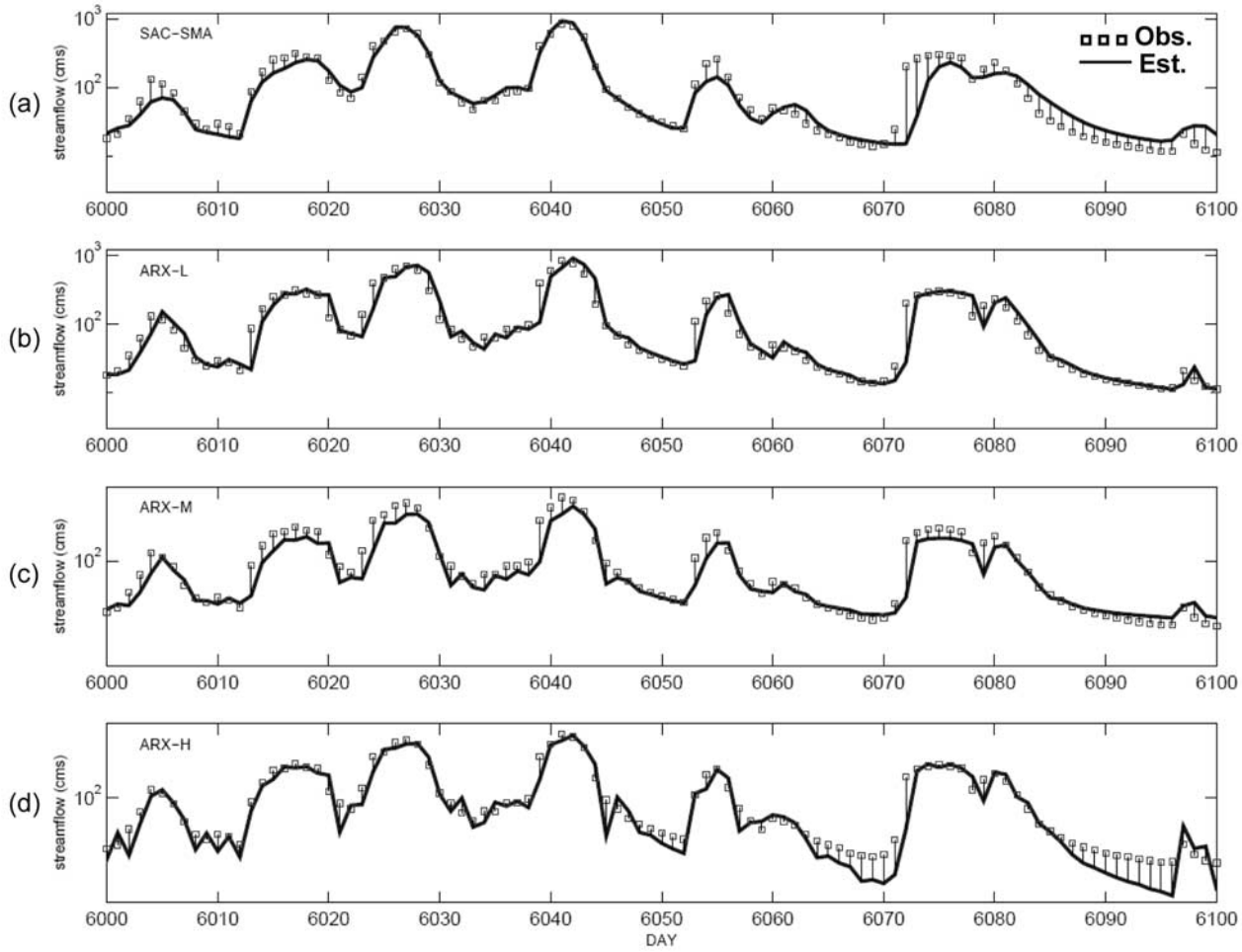

Figure 4. A close view of simulation flow time series from (a) SAC-SMA, (b) ARX-L, (c) ARX-M, and (d) ARX-H models within the 100-day simulation period. Observations are marked with squares, and model estimates are shown by solid lines.

[33] Although each model is expected to work well for flow prediction in its calibrated range, the quality of its estimates for ranges other than the calibrated one may not be consistent. The case studies used the recursive Bayes rule to calculate each model's posterior probability, which in turn was adaptively adjusted at each time interval when new observations became available. The expected value of the combined ARX prediction was calculated by $E(y_t|M_1, M_2, M_3, D) = \sum_{j=1}^{3} y_t(M_j)p_t(M_j|D)$. When $j = 1$, the model estimate was taken from the ARX-L model, whereas $j = 2$ and $j = 3$, the model estimates were taken from ARX-M and ARX-H, respectively. The posterior probability of the jth model was calculated on the basis of Bayes rule (see equation (8)).

[34] The initial a priori probability of a model is set to a uniform probability distribution for all models, i.e., $p_{t=0}(M_j) = 1/3$ for all models ($j = 1...3$). Gaussian likelihood functions are assumed to have a zero mean error and a variance (σ_j^2) calculated from the model residual over 11 years of calibration data (see equations (7) and (8)). To avoid the convergence of model's probability to zero, we set a threshold (h) for the posteriori. If the posterior probability of a model $p_t(M_j)$ is degenerated to zero, the model would not be able to move out of zero probability in a later time step. In this study, we set the threshold (h) equal to 0.01 in all models, if $p_t(M_j) < 0.01$.

[35] Ten models listed in Table 1 were evaluated. The study includes three basic ARX models (ARX-L, ARX-M, and ARX-H) and four merged models (AM, WA, SBC, and SMAP), which were combined in various forms from those three ARX models. Also, the ARX, SAC-SMA, and SOLO models are provided for the comparison.

[36] When combining models, the probability of models, $p_t(M_i)$, is considered to be fixed or varying in time. The combined weights of AM and WA were fixed at all times for all three ARX models. The AM assigns constant weights as $p(M_i) = \{1/3, 1/3, 1/3\}$ to {ARX-L, ARX-M, ARX-H} models, and the WA, on the other hand, assigns $p(M_i) = \{0.2785, 0.2520, 0.4695\}$ to those three ARX models. These probabilities were calculated from the error variations of the calibrated models (equation (11)), where the high-flow model (ARX-H) takes a higher probability, as its error variance ($\sigma^2_{\varepsilon_{ARX-H}} = 18.31^2$) is lower than that of the low-flow ($\sigma^2_{\varepsilon_{ARX-L}} = 23.77^2$) and medium-flow ($\sigma^2_{\varepsilon_{ARX-M}} = 24.99^2$) models.

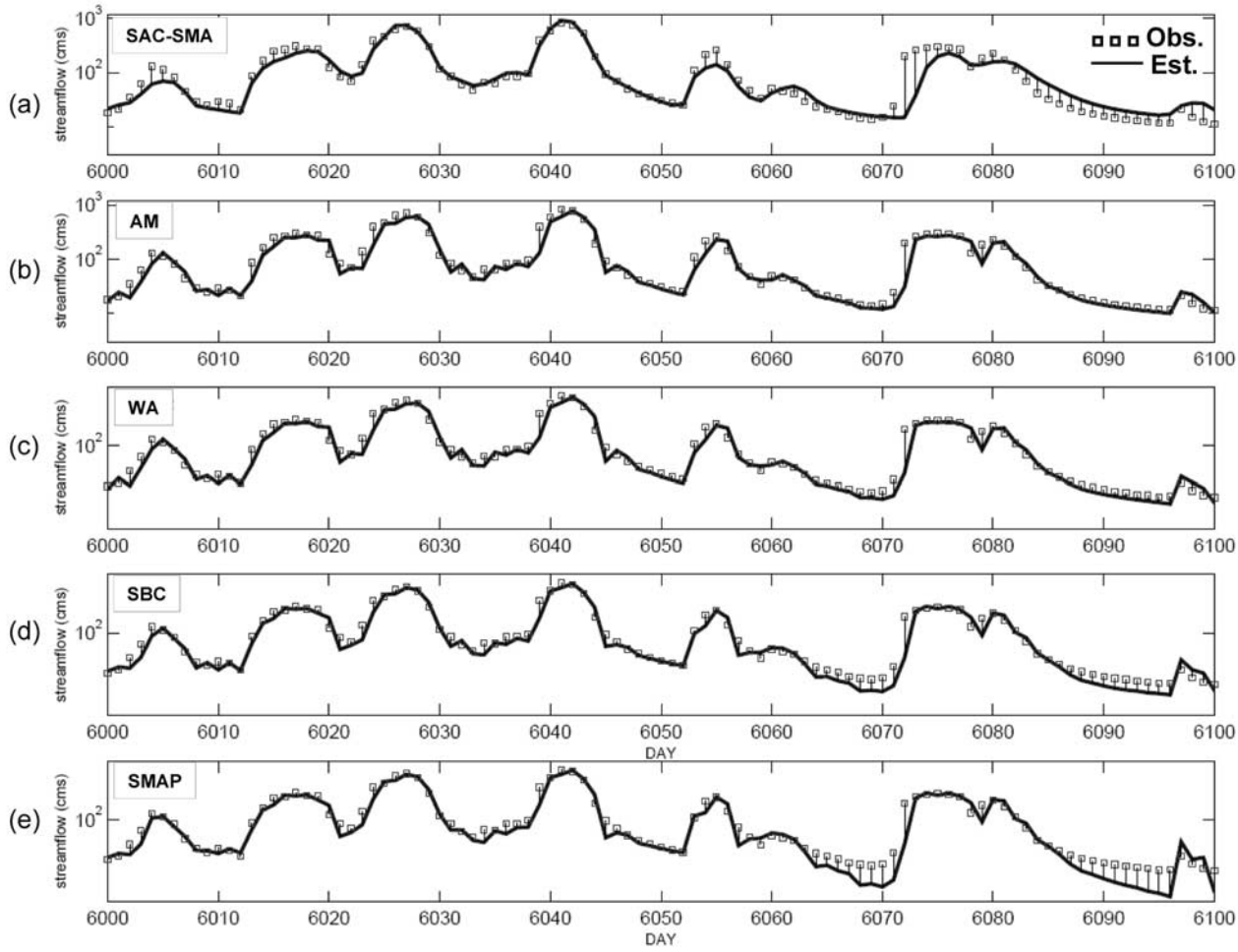

Figure 5. Similar to Figure 4 but for model estimates from (a) SAC-SMA, (b) AM, (c) WA, (d) SBC, and (e) SMAP models.

[37] Sequential calculations were used to assign variable posterior probabilities to the combined models SBC and SMAP. The SBC calculates each model's posterior probability recursively. When the observed flow is in the low-flow range, the likelihood function favors the ARX-L model and augments its posterior probability; therefore, the merged prediction is highly weighted toward the low-flow ARX-L model. Likewise, when the observed flow hydrograph is in the high-flow range, ARX-H provides a better basic model prediction, and the SBC assigns a higher likelihood to the ARX-H, thereby increasing its posterior probability. As higher probabilities are shifted toward models capable of providing better prediction, the combined model's overall performance improves. The SMAP selects only one model with the highest posterior probability to generate a prediction at each time step.

[38] Table 1 summarizes the performance of the individual models and various multimodels in the calibration and validation years. It shows that the both ARX (fully range calibration) and SAC-SMA models outperform all three ARX models (ARX-L, ARX-M, ARX-H), with higher NSE and CORR and lower RMSE estimates. The BIAS, however, SAC-SMA is somewhat higher in both calibration and validation periods. For the combined models with fixed probability at all times (i.e., AM and WA), the performance is very close to that of the ARX and SAC-SMA in both the calibration and testing periods. The SOLO model, on the other hand, outperforms all the other models in the evaluation. SOLO can be considered as a version of multimodel system using a set of 15 $\times$ 15 ARX models in its forecasting. In addition, data preprocessing using principal component regression also helps to find stable the regression parameters. And therefore, it is not surprising that the SOLO model can outperform the other models.

[39] Five performance measures for models {SBC, SMAP, ARX, SAC-SMA, and SOLO} were calculated in the validation period, as NSE = {0.942, 0.938, 0.900, 0.910, 0.931}, RMSE = {15.64, 16.14, 20.53, 19.43, 17.01 cmsd}, CORR = {0.971, 0.969, 0.949, 0.960, 0.965}, BIAS = {0.526, 0.554, 0.036, −5.444, −0.037 cmsd}, and MAE(Q >= 200) = {47.87, 48.56, 65.66, 56.37, 54.14 cmsd}. These statistics demonstrate that both SBC and SMAP outperform ARX and SAC-SMA in all evaluation categories. In addition, the SBC performs the best of all the combination schemes except the SOLO model. From the MAE(Q >= 200), both SBC and SMAP are better than SOLO. Figures 2 and 3 display the RMSE versus the average annual streamflow for ARX-L, ARX-M, and ARX-H models on each water year. The circles represent the calibration years, while the squares represent the validation years. Figures 2 and 3

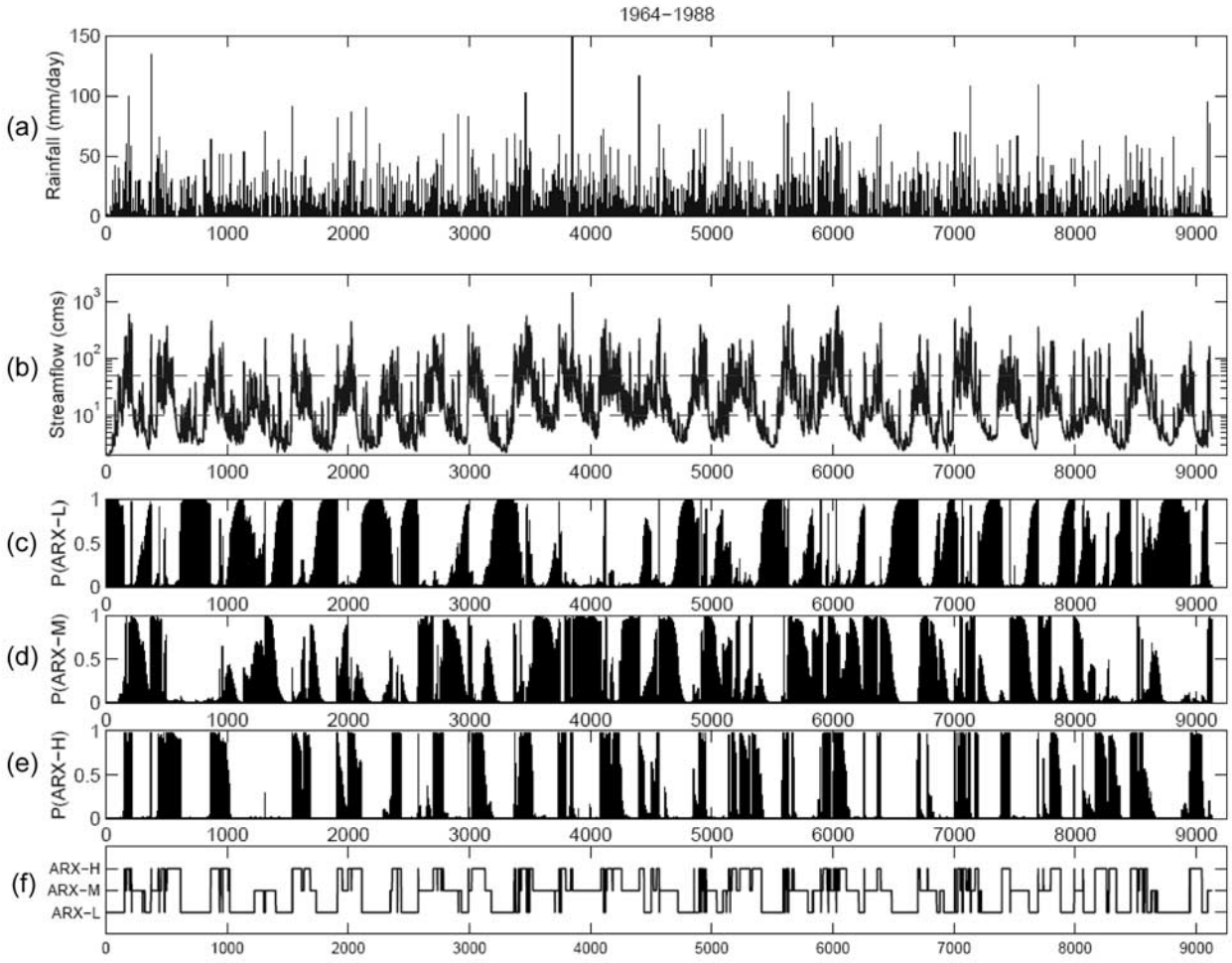

Figure 6. Sequential Bayesian simulation and model selection in the validation period (1964–1988): (a) daily rainfall time series, (b) daily streamflow time series, (c) posterior probability of the ARX-L model, (d) posterior probability of the ARX-M model, (e) posterior probability of the ARX-H model, and (f) the model identified as having the maximum a posteriori probability.

show that error variance increases with a year's "wetness" (i.e., the annual RMSE increases with annual streamflow).

[40] An arbitrary line is added to each graph from (0, 0) to (70, 40) to provide a basis for simple visual comparison. The SAC-SMA model (see Figure 3) performed well on high-flow (wetter) years but not as well on low-flow (drier) years. SOLO model, however, shows consistently well in both high- and low-flow years. The annual RMSE plot of ARX-H is similar to the SAC-SMA plot. SAC-SMA and ARX-H share very similar evaluation statistics (see Table 1), although in general, SAC-SMA performs better. The ARX-L and ARX-M models, with most annual RMSEs above the reference line, do not perform well in high-flow years. Comparing the SAC-SMA and ARX with the combined models (see Figure 3) in terms of daily RMSE and annual streamflow, the fixed weight models (AM and WA) perform better in low-flow years, but still have considerably higher RMSEs in high-flow years. The combined models with adjustable weight (SBC and SMAP) show better statistics than SAC-SMA and ARX in both low- and high-flow years. The performance of SOLO model is similar to the SBC and SMAP, but substantial better than AM and WA.

[41] The daily time series of estimates shows that the SAC-SMA model missed a few low flows, but in general fit the observed flow very well and showed generated flows in a more or less log linear relationship for the flows in the recession period. On the other hand, the ARX-L estimates fit very well with observations in the low-flow periods; ARX-M missed several low flows, but fit better in both medium and high flows; and the ARX-H fit very well over the high-flow period, but missed both medium and low flows substantially. Figure 4 displays the simulation within a 100-day period. Both the SAC-SMA and ARX-H models fit well in high-flow periods, but the ARX-L is expected to fit better than the others in the low-flow ranges.

[42] From the validation results from the multimodels (AM, WA, SBC, and SMAP), it shows that the combined models with fixed weights (i.e., AM and WA) tend to give estimates between the models' highest and lowest values. Although they improved the overall statistics from the ARX-L, ARX-M, and ARX-H models (see Table 1), they tended to overestimate medium and low flows and under-

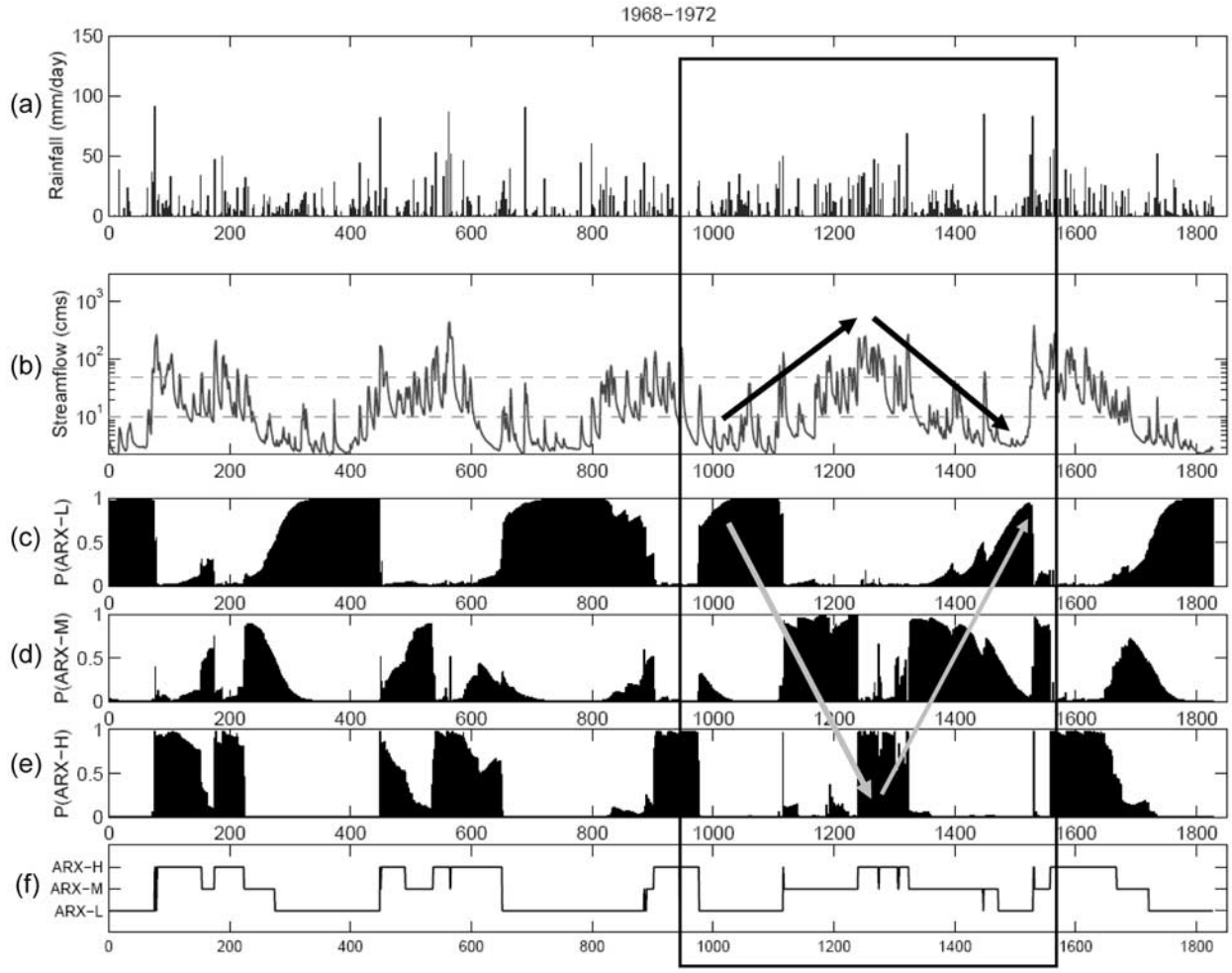

Figure 7. A close view of multimodel simulation and selection in the 5-year simulation period: (a) daily rainfall time series, (b) daily streamflow time series, (c) posterior probability of the ARX-L model, (d) posterior probability of the ARX-M model, (e) posterior probability of the ARX-H model, and (f) the model identified as having the maximum a posteriori probability.

estimate high flows. However, combined models with adjustable weights based on the posterior probability of ARX models showed the ability to give estimates extending to a much wider range of values than those of fixed-weight combined models. Compared with AM and WA, the SBC and SMAP models substantially improved estimates in the middle-flow section. Although the SBC and especially the SMAP models showed several spikes in the estimates for low-flow days, their overall performance, as shown in Table 1, was significantly better than that of the others. For better visualization of model estimates in different combinations, 100 days of simulated streamflow are shown in a logarithmic scale in Figure 5. The AM and WA showed a high residual, while the SCA-SMA performed well in high-flow ranges, but missed some medium to low flows. The SBC and SMAP, on the other hand, performed consistently throughout the flow ranges.

Table 2. Evaluation Statistics for Single Models and Combined Models in Both Calibration and Validation Data Periods[a]

	CAL					VAL				
Statistics	NSE	RMSE	CORR	BIAS	MAE ($Q \geq 200$)	NSE	RMSE	CORR	BIAS	MAE ($Q \geq 200$)
ARX	0.923	17.24	0.961	−0.105	66.69	0.900	20.53	0.949	0.036	65.66
SAC-SMA	0.918	17.79	0.959	−2.292	64.93	0.910	19.43	0.960	−5.444	56.37
SOLO	0.960	12.36	0.980	−0.075	44.44	0.931	17.01	0.965	−0.037	54.14
WA	0.961	12.28	0.981	−0.824	44.64	0.947	14.84	0.974	−1.815	45.49
SBC	0.975	9.735	0.988	−0.012	31.23	0.968	11.54	0.984	−0.775	32.55
SMAP	0.973	10.25	0.986	−0.005	31.94	0.964	12.32	0.982	−0.781	34.64

[a]Single models are ARX, SAC-SMA, and SOLO, and combined models are WA, SBC, and SMAP.

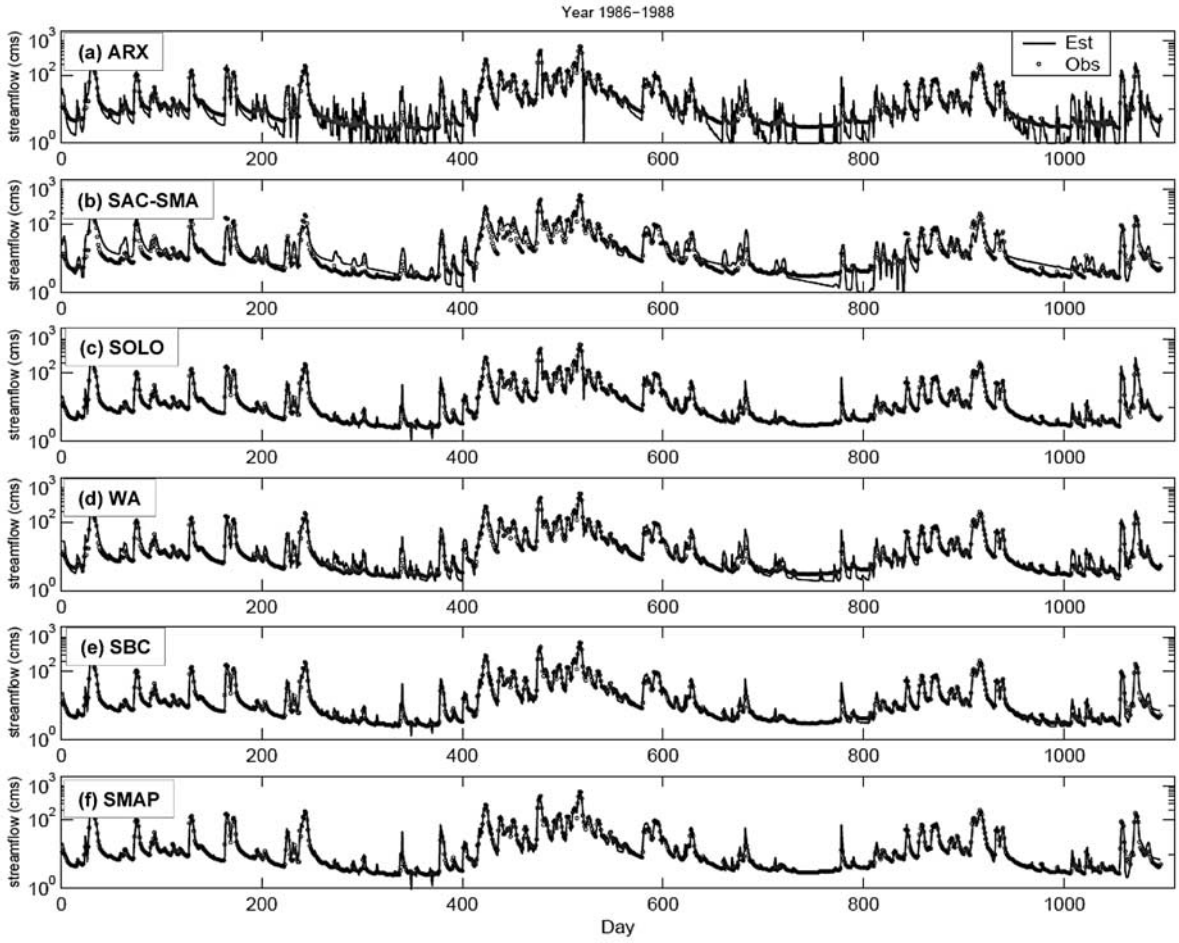

Figure 8. Simulation time series of testing models (1986–1988).

[43] The SBC requires the posterior probability data from the basic ARX models at each time interval, while the SMAP merging strategy picks up the estimate from the ARX model with the highest probability. Figure 6 displays the sequential evolution of model combination and selection based on the ARX models' posterior probability. Figures 6a and 6b show the rainfall and runoff time series. Figures 6c–6e describe the probability of the ARX-L, ARX-M, and ARX-H models at each time interval. SBC merges the three ARX model estimates on the basis of their probability. Figure 6f shows the ARX model with the highest probability at each time interval. Finally, the SMAP merging model essentially selects the ARX estimate with the highest probability.

[44] A close look at the probability evolution of ARX models through the first 5-year validation period is listed in Figure 7. During this time period, from day 1100 to day 1450 in Figures 7b–7e, streamflow increases from low flow (day 1100) to high flow (around day 1300) and then decreases to low flow (around day 1450). The ARX models' probability is consistent with the choice to assign higher probability to the ARX-L model for the low-flow period, to the ARX-M model for the medium flow, and to the ARX-H for the high flow, and then reverses the order that assigned high probability to the ARX-L and low probability to the ARX-H and ARX-M during the recession period. The model selection pattern is repeated over the 5-year period.

[45] The model with the highest probability of all the ARX models is plotted in Figure 7f, which shows how SMAP selects a model from all the ARX models. The selected models (ARX-L and ARX-H) correspond very well to the flow status (low and high flows) at the specified time period.

4.2. Case Study 2: Fusion of ARX, SAC-SMA, and SOLO Estimates

[46] Case study 2 describes the combination of estimates from three different R-R models (ARX, SAC-SMA, and ANN). The ARX is a linear system model, SAC-SMA is a conceptual hydrologic model, and SOLO is a neural network model. Similar to case study 1, all these models are calibrated from 11 years of data and then validated using 25 years data. All these models were trained for 1-day-ahead streamflow forecasting.

[47] Three combined models (WA, SBC, and SMAP) were calibrated and validated in the same data period. Table 2 summarizes the performance of all these models. In the basic model group, SOLO outperforms the others in all evaluation criteria. Evaluation statistics for ARX and SAC-SMA are close, and ARX performs slightly better in the calibration period, while SAC-SMA shows better performance in the validation period.

[48] The combined models (WA, SBC, and SMAP) showed better statistics in almost all evaluation criteria

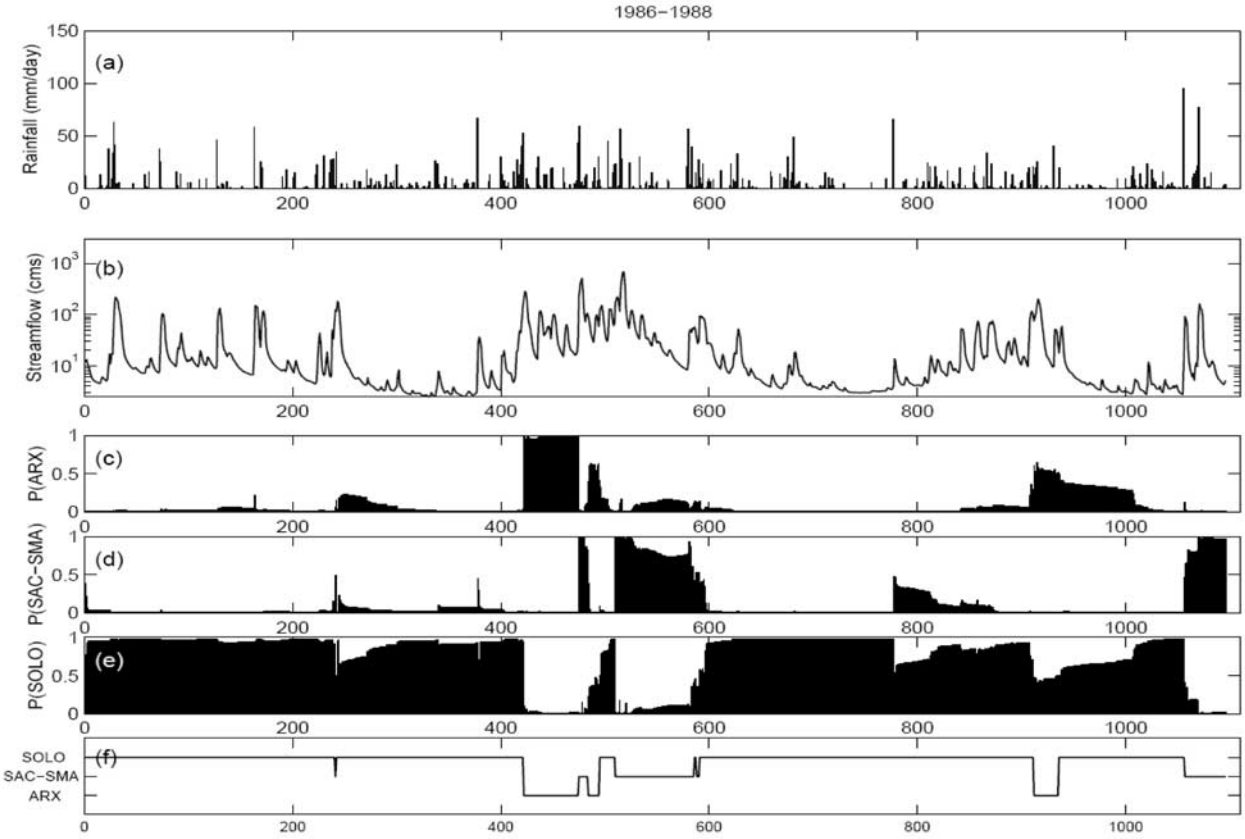

Figure 9. (a) Rainfall and (b) runoff time series (1986–1988); posterior probability of (c) ARX, (d) SAC-SMA, and (e) SOLO models; and (f) the model with the maximum probability.

except the BIAS index (see Table 2), in which the ARX and SOLO models give the lowest values. For the other indexes, NSE, RMSE, and CORR, the combined models give substantially better results than ARX, SAC-SMA, and SOLO. This performance was consistent in both calibration and validation events.

[49] For the combined models in the calibration period, SBC and SMAP significantly outperformed WA; the RMSE was 9.73 and 10.25 cmsd for SBC and SMAP, but was 12.28 cmsd for WA. Similarly, in the validation period, the RMSE for SBC and SMAP was 11.54 and 12.32 cmsd, but was only 14.48 cmsd for WA. For the other statistics, such as NSE, CORR, BIAS, and MAE($Q >= 200$), it is clear that SBC and SMAP are better than WA. As previously stated, the model weights for SBC and SMAP were adjusted at each time interval, while WA is a fixed-weight scheme whose weights were calculated from the calibration data. In general, a significant gain in accuracy from sequential adjustment was observed.

[50] The daily time series of the ARX, SAC-SMA, and SOLO models during the validation years (1986–1988) are listed in Figures 8a–8c. The ARX model fit well in the high-flow period, but registered rather spiky for the predictions in both medium- and low-flow periods. The SAC-SMA model fit the observed flow well, registering as somewhat log linear in the flow's recession period. The SOLO model seems to fit well with the observations for most periods. The time series plots of the combined models (WA, SBC, and SMAP) in the same validation period are plotted in Figures 8d and 8e. The model combined with fixed weights (WA) showed good fitting over the high-flow period, but either overestimated or underestimated values in part of the medium- and low-flow periods. Unlike the ARX estimates, the averaging of multiple estimates caused the WA time series to be smooth throughout the validation period. The flow for SBC and SMAP models fit generally well at all ranges (see Figures 8e and 8f). As shown in the evaluation statistics in Table 2, SBC and SMAP outperformed the other models in all evaluation criteria. Table 2 shows that SBC's performance is better than SMAP's, although the graphic in Figure 8 suggests that they are very close.

[51] Figure 9 shows the rainfall (Figure 9a) and the runoff (Figure 9b) time series; the posterior probability of ARX (Figure 9c), SAC-SMA (Figure 9d), and SOLO (Figure 9e) models; and the model with the highest probability (Figure 9f). The models' posterior probability plot (see Figures 9c–9e) makes it clear that of all the combination models, the SOLO model covers most of the time periods studied (see Figure 9f), while the others (ARX and SAC-SMA) cover only a limited time during the high-flow and recession periods. The low-flow period is covered mainly by the SOLO estimates. Again, as shown in Figure 9f, if only one model was selected at each time interval, the SOLO model was selected for more than 80% of the simulation periods, while the ARX and SAC-SMA models were selected for 8% and 12%, respectively.

[52] The improvement from the SBC and SMAP is mainly from the continuing adjustment of model probability through new available observations as well as the model's priori probability. As there is no perfect model, each model may estimate better for certain flow ranges. Providing a suitable framework being capable of selecting effective models for generating prediction is critical. Bayes' rule provides a basic framework for sequential model section. Although from case studies, multimodel fusion approaches (e.g., SBC and SMAP) output perform those fixed probability approaches (e.g., AM and WA) consistently, for the operational hydrology, potential issues may be raised from the observation which is not available in real time. When the time latency of observations increases, the model tends to give prediction without appropriate information and therefore the forecasting skill can be downgraded substantially. In the case studies, we only consider for one time step ahead ($t + 1$) prediction of streamflow from multimodels, the prediction uncertainty may arise when the forecasting time step extended. As the prediction extended for three time steps (e.g., y_{t+3}), for example, one key uncertainty is from to the precipitation forcing (i.e., r_{t+1} and r_{t+2}) prediction being not reliable at time $t + 1$ and $t + 2$. Meanwhile because the observed streamflow is not available for the next two time steps, all the models need to either give their forecasts from the time delayed observations to up to current time, or use model predictions at time $t + 1$ and $t + 2$, (i.e., $\hat{y}_{t+1}$ and $\hat{y}_{t+2}$). In the study, we introduced using sequential modeling techniques in fusion several model estimations. The capability of multiple time step prediction using the sequential Bayesian model selection approach will be explored in the continuing investigation.

5. Conclusions

[53] This paper has discussed several multiple-model estimates for watershed runoff prediction. The proposed multiple-model approaches (SBC and SMAP) use the sequential Bayesian rule to calculate and update individual models' probability. A weighting factor for model combination then was assigned according to the probability assigned to each model. Case studies demonstrated that the multiple-model strategy is effective for improving streamflow forecasting. Two case studies were demonstrated using (1) three basic linear ARX models, and (2) three more sophisticated models (ARX, SAC-SMA, and SOLO). The evaluation shows that although using fixed weights for model combinations (e.g., AM and WA) may improve overall prediction performance, they may not be optimal for every time interval. Setting the weighting factor so that it is sequentially adjustable may improve the estimates for most portions of the hydrograph. The SBC and SMAP approaches enabled us to combine model estimates adaptively according to their conditional probability, calculated up to the most nearest observations available. The evaluation of SBC and SMAP showed that their performance was better than those of the individual models and the fixed weighting factor fusion methods, such as AM and WA.

[54] The SBC and SMAP model fusion approaches can be generalized to combine estimates from all kinds of models, such as physically based, conceptual, and data-driven models (e.g., SAC-SMA, ARX, artificial neural networks and fuzzy rule–based models). Experiments including other types of models will be explored in future studies.

[55] **Acknowledgments.** Partial financial support of this study was made available through research grants from NASA NEWS (grant NNX06AF93G), NASA EOS (grant NA56Gp185), NOAA CPPA (grant NA04OAR4310086 and NA070AR4310203), and NSF SAHRA (grant EAR 9876800). The authors also would like to thank Diane Hohnbaum for her editing of this manuscript.

References

Abrahart, R. J., and L. See (2002), Multi-model data fusion for river flow forecasting: An evaluation of six alternative methods based on two contrasting catchments, *Hydrol. Earth Syst. Sci.*, *6*(4), 655–670.

Abramowitz, G., H. Gupta, A. Pitman, Y. P. Wang, R. Leuning, H. Cleugh, and K.-L. Hsu (2006), Neural error regression diagnosis (NERD): A tool for model bias identification and prognostic data assimilation, *J. Hydrometeorol.*, *7*(1), 160–177, doi:10.1175/JHM479.1.

Ajami, N. K., Q. Y. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, *43*, W01403, doi:10.1029/2005WR004745.

Barnston, A. G., S. J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak (2003), Multimodel ensembling in seasonal climate forecasting at IRI, *Bull. Am. Meteorol. Soc.*, *84*(12), 1783–1796, doi:10.1175/BAMS-84-12-1783.

Beven, K. J. (2001), *Rainfall-Runoff Modelling: The Primer*, 360 pp., John Wiley, Chichester, U. K.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*(12), 3663–3674, doi:10.1029/2000WR900207.

Boyle, D. P., H. V. Gupta, S. Sorooshian, V. Koren, Z. Y. Zhang, and M. Smith (2001), Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resour. Res.*, *37*(11), 2749–2759, doi:10.1029/2000WR000207.

Brazil, L. E. (1988), Multilevel calibration strategy for complex hydrologic simulation models, Ph.D. dissertation, 217 pp. Colo. State Univ., Fort Collins.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system; conceptual modeling for digital computers, 204 pp., State of Calif, Dep. of Water Resour., Sacramento, Calif.

Chow, V. T., D. R. Maidment, and L. W. Mays (1988), *Applied Hydrology*, 572 pp., McGraw-Hill, New York.

Coulibaly, P., M. Hache, V. Fortin, and B. Bobee (2005), Improving daily reservoir inflow forecasts with model combination, *J. Hydrol. Eng.*, *10*(2), 91–99, doi:10.1061/(ASCE)1084-0699(2005)10:2(91).

Daley, R. (1991), *Atmospheric Data Analysis*, 457 pp. Cambridge Univ. Press, Cambridge, U. K.

Dawson, C. W., R. J. Abrahart, and L. M. See (2007), HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, *Environ. Modell. Software*, *22*(7), 1034–1052, doi:10.1016/j.envsoft.2006.06.008.

Duan, Q. (2003), *Calibration of Watershed Models, Water Sci. Appl.*, vol. 6, 345 pp. AGU, Washington, D. C.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.

Duan, Q., N. K. Ajami, X. G. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Adv. Water Resour.*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Frevert, D. K., and V. P. Singh (2002), *Mathematical Models of Small Watershed Hydrology and Applications*, 950 pp. Water Resour. Publ., Highlands Ranch, Colo.

Gelman, A. (2004), *Bayesian Data Analysis*, 2nd ed., 668 pp. Chapman and Hall, Boca Raton, Fla.

Georgakakos, K. P., D. J. Seo, H. Gupta, J. Schaake, and M. B. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, *298*(1–4), 222–241, doi:10.1016/j.jhydrol.2004.03.037.

Govindaraju, R. S., and A. R. Rao (2000), *Artificial Neural Networks in Hydrology*, 329 pp. Kluwer Acad., Dordrecht, Netherlands.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable mea-

sures of information, *Water Resour. Res.*, *34*(4), 751–763, doi:10.1029/97WR03495.

Hong, Y., K. L. Hsu, S. Sorooshian, and X. G. Gao (2005), Self-organizing nonlinear output (SONO): A neural network suitable for cloud patch-based rainfall estimation at small scales, *Water Resour. Res.*, *41*, W03008, doi:10.1029/2004WR003142.

Hsu, K. L., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, *Water Resour. Res.*, *31*(10), 2517–2530, doi:10.1029/95WR01955.

Hsu, K. L., H. V. Gupta, X. G. Gao, and S. Sorooshian (1999), Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation, *Water Resour. Res.*, *35*(5), 1605–1618, doi:10.1029/1999WR900032.

Hsu, K. L., H. V. Gupta, X. G. Gao, S. Sorooshian, and B. Imam (2002), Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, *Water Resour. Res.*, *38*(12), 1302, doi:10.1029/2001WR000795.

Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn (1995), Global precipitation estimates based on a technique for combining satellite-based estimates, rain-gauge analysis, and NWP model precipitation information, *J. Clim.*, *8*(5), 1284–1295, doi:10.1175/1520-0442(1995)008<1284:GPEBOA>2.0.CO;2.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (1999), Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, *285*(5433), 1548–1550, doi:10.1126/science.285.5433.1548.

Linsley, R. K., J. L. H. Paulhus, and M. A. Kohler (1982), *Hydrology for Engineers*, 3rd ed., 508 pp. McGraw-Hill, New York.

Maier, H. R., and G. C. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications, *Environ. Modell. Software*, *15*(1), 101–124, doi:10.1016/S1364-8152(99)00007-9.

Moradkhani, H., K. Hsu, H. V. Gupta, and S. Sorooshian (2004), Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, *J. Hydrol.*, *295*(1–4), 246–262, doi:10.1016/j.jhydrol.2004.03.027.

Moradkhani, H., K. L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Petridis, V., and A. Kehagias (1998), *Predictive Modular Neural Networks: Applications to Time Series*, 314 pp., Kluwer Acad., Boston, Mass.

Petridis, V., A. Kehagias, L. Petrou, A. Bakirtzis, S. Kiartzis, H. Panagiotou, and N. Maslaris (2001), A Bayesian multiple models combination method for time series prediction, *J. Intell. Robot. Syst.*, *31*(1–3), 69–89, doi:10.1023/A:1012061814242.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997), Bayesian model averaging for linear regression models, *J. Am. Stat. Assoc.*, *92*(437), 179–191, doi:10.2307/2291462.

Robert, C. P., and G. Casella (2004), *Monte Carlo Statistical Methods*, 2nd ed., 645 pp., Springer, New York.

See, L. (2008), Data fusion methods for integrating data-driven hydrological models, *Stud. Comput. Intell.*, *79*, 1–18, doi:10.1007/978-3-540-75384-1_1.

See, L., and R. J. Abrahart (2001), Multi-model data fusion for hydrological forecasting, *Comput. Geosci.*, *27*(8), 987–994, doi:10.1016/S0098-3004(00)00136-9.

Shamseldin, A. Y. (1997), Application of a neural network technique to rainfall-runoff modelling, *J. Hydrol.*, *199*(3–4), 272–294, doi:10.1016/S0022-1694(96)03330-6.

Shamseldin, A. Y., K. M. O. Connor, and G. C. Liang (1997), Methods for combining the outputs of different rainfall-runoff models, *J. Hydrol.*, *197*(1–4), 203–229, doi:10.1016/S0022-1694(96)03259-3.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268, doi:10.1029/WR019i001p00260.

Sorooshian, S., Q. Y. Duan, and V. K. Gupta (1983), Calibration of rainfall-runoff models: Application of global optimization to the Sacramento soil moisture accounting model, *Water Resour. Res.*, *29*(4), 1185–1194, doi:10.1029/92WR02617.

Vrugt, J. A., and B. A. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, *43*, W01411, doi:10.1029/2005WR004838.

Wood, E. F., and A. Szöllösi-Nagy (Eds.) (1980), *Real-Time Forecasting/Control of Water Resource Systems: Selected Papers From an IIASA Workshop, October 18–21, 1976*, 1st ed., 330 pp., Pergamon, Oxford, U. K.

Xie, P. P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*(4), 840–858, doi:10.1175/1520-0442(1996)009<0840:AOGMPU>2.0.CO;2.

Xiong, L. H., A. Y. Shamseldin, and K. M. O'Connor (2001), A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system, *J. Hydrol.*, *245*(1–4), 196–217, doi:10.1016/S0022-1694(01)00349-3.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, *181*(1–4), 23–48, doi:10.1016/0022-1694(95)02918-4.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1–4), 83–97, doi:10.1016/S0022-1694(97)00107-8.

K.-L. Hsu and S. Sorooshian, Center for Hydrometeorology and Remote Sensing, Department of Civil and Environmental Engineering, University of California, E-4130 Engineering Gateway, Irvine, CA 92697-2175, USA. (kuolinh@uci.edu)

H. Moradkhani, Department of Civil and Environmental Engineering, Portland State University, 1930 S.W. 4th Avenue, Suite 200, Portland, OR 97201, USA.