AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2013WR013855

Key Points:
- Reducing climatic uncertainty plays a critical role in developing streamflow forecasts
- Hydrologic models forced with multimodel climate forecasts perform better
- Systematic uncertainty reduction results in improved streamflow forecasts

Correspondence to:
A. Sankarasubramanian,
sankar_arumugam@ncsu.edu

Citation:
Singh, H., and A. Sankarasubramanian (2014), Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models, *Water Resour. Res.*, *50*, 1288–1307, doi:10.1002/2013WR013855.

Received 20 MAR 2013
Accepted 13 JAN 2014
Accepted article online 17 JAN 2014
Published online 15 FEB 2014

Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models

Harminder Singh[1] and A. Sankarasubramanian[1]

[1]Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, North Carolina, USA

Abstract Recent studies show that multimodel combinations improve hydroclimatic predictions by reducing model uncertainty. Given that climate forecasts are available from multiple climate models, which could be ingested with multiple watershed models, what is the best strategy to reduce the uncertainty in streamflow forecasts? To address this question, we consider three possible strategies: (1) reduce the input uncertainty first by combining climate models and then use the multimodel climate forecasts with multiple watershed models (MM-P), (2) ingest the individual climate forecasts (without multimodel combination) with various watershed models and then combine the streamflow predictions that arise from all possible combinations of climate and watershed models (MM-Q), (3) combine the streamflow forecasts obtained from multiple watershed models based on strategy (1) to develop a single streamflow prediction that reduces uncertainty in both climate forecasts and watershed models (MM-PQ). For this purpose, we consider synthetic schemes that generate streamflow and climate forecasts, for comparing the performance of three strategies with the true streamflow generated by a given hydrologic model. Results from the synthetic study show that reducing input uncertainty first (MM-P) by combining climate forecasts results in reduced error in predicting the true streamflow compared to the error of multimodel streamflow forecasts obtained by combining streamflow forecasts from all-possible combination of individual climate model with various hydrologic models (MM-Q). Since the true hydrologic model structure is unknown, it is desirable to consider MM-PQ as an alternate choice that reduces both input uncertainty and hydrologic model uncertainty. Application on two watersheds in NC also indicates that reducing the input uncertainty first is critical before reducing the hydrologic model uncertainty.

1. Introduction

Skillful seasonal streamflow forecasts are useful to planners and water managers for quantifying water availability and for allocation. Seasonal streamflow forecasts are typically obtained either by developing statistical relationships between the climate forecasts and initial land-surface conditions (e.g., snow water equivalent) and the observed streamflow [*Sankarasubramanian et al*., 2008] or by downscaling the precipitation and temperature forecasts into a land-surface model's (LSM) grid scale, so that the downscaled forcings could be ingested into the LSM [*Luo et al*., 2008; *Mahanama et al*., 2011; *Sinha and Sankarasubramanian*, 2013]. It is natural to expect that the skill of streamflow forecasts developed based on the latter approach would depend on the skill of downscaled climate forecasts as well as based on the ability of the LSM in predicting the observed streamflow. Since the skill of the climate forecasts from general circulation models (GCMs) varies substantially across the models [*Devineni and Sankarasubramanian*, 2010], research institutes, and operational agencies have been issuing multimodel climate forecasts [*Goddard et al*., 2003] for reducing model uncertainty. Recently, *Koster et al*. [2011] compared the performance of different LSMs under climatological forcings for developing multimodel runoff forecasts. Thus, the availability of climate forecasts from multiple GCMs and various watershed models provides an opportunity to reduce the uncertainty in seasonal streamflow forecasts.

Recent studies on climate forecasts have shown that combining multiple models reduces the uncertainty in climate forecasts [*Rajagopalan et al*., 2002; *Devineni and Sankarasubramanian*, 2010]. *Weigel et al*. [2008] clearly demonstrated using a synthetic model setup that multimodel forecasts outperform the best single model only if single models exhibit overconfidence in prediction. Recently, *Devineni et al*. [2010] improved

the skill in predicting winter precipitation and temperature over the United States by optimally combining multiple GCMs using an algorithm that assesses the models' skills conditioned on El Nino Southern Oscillation (ENSO) state. Thus, it has become a common practice that real-time climate forecasts are issued based on the net assessment whose inputs are primarily obtained by combining multiple climate models [*Goddard et al*., 2003; *Barnston et al*., 2003].

Similarly, on comparing the performance of multiple watershed models, studies have demonstrated the utility of combining hydrologic models in order to improve streamflow predictions [*Georgakakos et al*., 2004; *Ajami et al*., 2007; *Marshall et al*., 2005, 2006; *Oudin et al*., 2006; *Najafi et al*., 2011; *Li and Sankarasubramanian*, 2012; *Parrish et al*., 2012]. Since each hydrological model has uncertainty due to its inherent process representation, it is advantageous to combine the strengths of various hydrologic models for improving streamflow predictions [*Vrugt et al*., 2006; *Duan et al*., 2007; *Marshall et al*., 2007a, 2007b; *Vrugt and Robinson*, 2007; *Chowdhury and Sharma*, 2009]. It has also been shown that multimodel streamflow forecasts outperform individual model forecasts even if the individual model forecasts from statistical models' are well calibrated [*Regonda et al*., 2005; *Devineni et al*., 2008; *Wang et al*., 2012]. Recently, *Li and Sankarasubramanian* [2012] demonstrated using a synthetic study that multimodel streamflow predictions perform better than individual model predictions as measurement uncertainty and model structural uncertainty increases. *While these multimodel combination studies have focused exclusively on reducing the uncertainty on climate forecasts or on the hydrologic models, there is no systematic approach available on how to reduce the combined uncertainty in seasonal streamflow forecasts arising from both climate forecasts and hydrological models*.

The main objective of this study is to identify a systematic approach for reducing the uncertainty in seasonal streamflow forecasts developed using a hydrological model. Given that the multimodel climate forecasts perform better than individual model GCMs, which is the right approach that systematically reduces uncertainty in streamflow forecasts? To address this question, we consider three possible strategies:

1. *Reduce the input uncertainty first by combining climate models and then use the multimodel climate forecasts with multiple watershed models (MM-P)*.

2. *Ingest individual GCMs' climate forecasts (without multimodel combination) with various watershed models and then combine the streamflow predictions that arise from all possible combinations of climate and watershed models (MM-Q) and*

3. *Combine the streamflow forecasts obtained from multiple watershed models based on strategy (1) to develop a single streamflow prediction that first reduces uncertainty in climate forecasts and then subsequently in the watershed models (MM-PQ)*.

For simplicity, we refer to these three uncertainty reduction strategies as *MM-P*, *MM-Q*, and *MM-PQ*, respectively. Evaluating the above three strategies can provide a clear understanding on how uncertainty in streamflow forecasts could be reduced given the plethora of GCMs and hydrologic models. We evaluate the above strategies to predict a known "true flow" under a synthetic generation scheme as well as by evaluating their ability to predict the observed streamflow for a watershed in NC.

2. Experimental Design

This section describes the overall experimental design that is involved in developing synthetic precipitation and streamflow forecasts for evaluating the combination strategies. The primary idea behind evaluating the strategies using synthetic generation scheme is that given the true climate forcings and true streamflow, the performance of the three forecasting strategies could be compared with the true flow. For this purpose, we consider the observed seasonal precipitation and observed potential evapotranspiration as the true climatic forcings with true streamflow from a given watershed model being obtained using the methodology discussed in section 2.1. Observed monthly precipitation, streamflow, and potential evapotranspiration are obtained for 35 years for the Tar River Site from the national hydroclimatic database available for the continental United States [*Vogel and Sankarasubramanian*, 1999; *Sankarasubramanian and Vogel*, 2002]. Using the observed precipitation, precipitation forecast with a specified skill (i.e., correlation) and streamflow with a specified measurement error are obtained. Figure 1 shows the overall experimental design for evaluating the performance of a given streamflow forecast develop strategy.

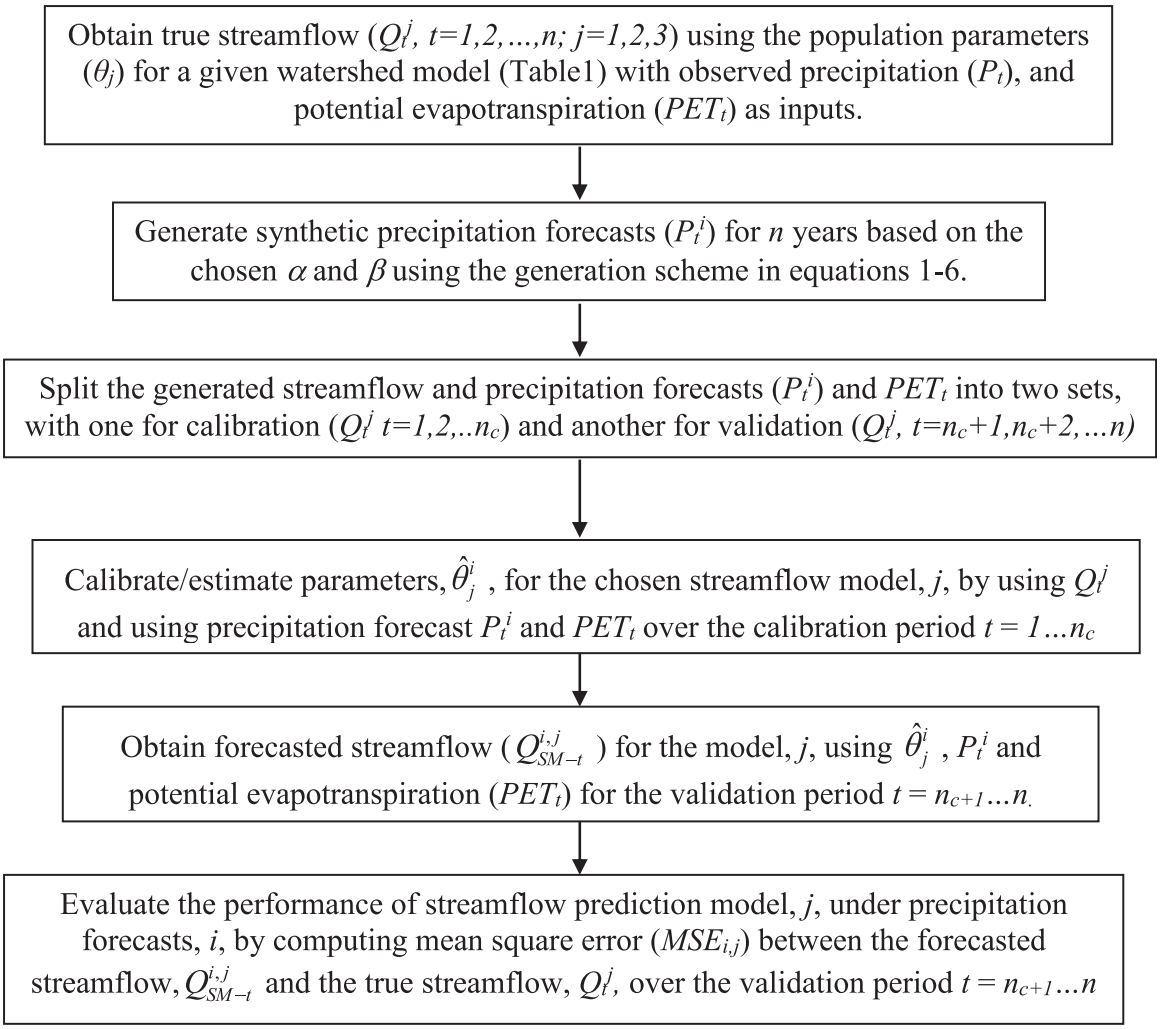

Figure 1. Experimental design for evaluating the performance of streamflow forecasts developed from a precipitation forecasting scheme and the calibrated watershed model.

2.1. True Streamflow Generation From a Given Hydrological Model

Under this subsection, we present the formulation for obtaining model-based true streamflows for evaluating the proposed strategies. To obtain model-based true streamflow, we consider two stochastic streamflow models and one conceptual water balance model. The streamflow generation scheme is similar to that of the scheme employed by *Li and Sankarasubramanian* [2012] for understanding why multimodel streamflow estimates perform better than the individual model estimates. Table 1 provides the parameters of linear, log-linear and "abcd" model. The true streamflow is generated using one of these watershed models based on the observed winter (January–March) precipitation and potential evapotranspiration available for the Tarboro site. We use the observed winter (January–March) precipitation (P_t), streamflow (Q_t), and potential evapotranspiration (PET_t) for the Tar River at Tarboro to estimate the population parameters (Table 1) by minimizing the sum of squares of errors between the observed streamflow and the model predicted

Table 1. Summary of Candidate Streamflow Generation Models Used in the Synthetic Study as Well as for Application

Models	Population Model	Population Parameters	Measurement Error
Linear	$Q_t^1=\hat{a}_1+\hat{b}_1*P_t+\hat{c}_1*PET_t+\varepsilon_t^1$	$a_1=41.15$, $b_1=0.58$, $c_1=-0.77$	$\varepsilon_t^1 \sim N(0, f*\hat{\sigma}_\varepsilon^1)$
Nonlinear	$Y_t=\hat{a}_2+\hat{b}_2*\log(P_t)+\hat{c}_1*\log(PET_t)+\varepsilon_t^2 Q_t^2=10^{Y_t}$	$a_2=0.99$, $b_2=1.08$, $c_2=-0.80$	$\varepsilon_t^2 \sim N(0, f*\hat{\sigma}_\varepsilon^2)$
"abcd"	Conceptual water balance model with G_{t-1}, S_{t-1}, PET_t, P_t as inputs and Q_t^3 as output [*Sankarasubramanian and Vogel*, 2002]	$a=0$, $b=148.60$, $c=0.48$, $d=0$	$\varepsilon_t^3 \sim N(0, f*\hat{\sigma}_\varepsilon^3)$

streamflow. The "*abcd*" model is calibrated using the seasonal input data and the parameters of the seasonal model are considered as "*true*" parameters for the synthetic study.

Using the population parameters specified in Table 1, we generate true seasonal streamflow, Q_t^j, where $j = 1, 2, 3$ denoting linear, log-linear and "abcd" models, respectively, from each watershed model, *j*, using the observed winter seasonal precipitation and PET. It is important to note that the true flows also contain a measurement/output error (ε_t^j), which follows Gaussian noise with zero mean and standard deviation, *f** $\hat{\sigma}_\varepsilon^j$, where "*f*" denotes a factor that control the residual standard deviation $(\hat{\sigma}_\varepsilon^j)$ for model "*j*." The residual standard deviation $(\hat{\sigma}_\varepsilon^j)$ for a given model is estimated based on the residuals between the observed winter streamflow and the model-estimated flow. For $f = 0$, the generated flow does not have any measurement error resulting in flows being exactly as that of the model estimates for the Tar River at Tarboro. The case of "$f = 0$" is specifically considered to find out how different multimodel combination strategies perform given the true flow are available for calibration. The measurement error is added explicitly to the model-estimated flow to generate many realizations of true flows. The "true" flows from the true models allow us to compare between single model streamflow predictions and various multimodel streamflow predictions developed using the precipitation forecasts.

2.2. Single Model Precipitation Forecasts Generation

Synthetic precipitation forecasts, $P_{m,t}^i$, where $m = 1, 2, \ldots, M$ denote the ensemble members and $i = 1, 2, \ldots, N$ represent the number of climate models issuing forecasts, are generated for the season, *t*, based on the toy model suggested by *Weigel et al.* [2008]. The skill of the synthetic precipitation forecasts is controlled by two parameters, α and β, with α denoting the correlation between the forecasted precipitation and the true precipitation and β indicating the overconfidence of the forecasted precipitation. We consider the observed winter (January–March) precipitation (P_t) at Tar River at Tarboro over the period 1952–1986 as the true precipitation with winter climatology being represented by the mean (μ_P) and standard deviation (σ_P). *Weigel et al.* [2008] considered synthetic climate forecasts with zero mean and one standard deviation. To apply the *Weigel et al.* [2008] generation scheme for the Tar River at Tarboro data, we standardize the observed precipitation to obtain x_t:

$$x_t = \frac{P_t - \mu_P}{\sigma_P} \tag{1}$$

$$\varepsilon_\beta^i \sim N(0, \beta^i) \tag{2}$$

$$\varepsilon_m^i \sim N\left(0, \sqrt{1-(\alpha^i)^2-(\beta^i)^2}\right) \tag{3}$$

Two noise terms ε_β and ε_m follow Gaussian distribution with zero mean and the respective standard deviations as specified in equations (2) and (3). The noise term, ε_β, specifies the overconfidence of the forecast, which forces all the members of ensemble to be far away from the true precipitation (P_t). It is important to note that the noise term ε_β which denotes the overconfidence of the forecasts is fixed for a given year. Thus, the parameter β relocates the ensemble mean for generating the overconfidence forecasts using Gaussian noise term with zero mean and standard deviation β. For $\beta = 0$, the term $\varepsilon_\beta = 0$ in equation (4) resulting in the generated forecast being well dispersed around the observed precipitation x_t. The parameter α ranges between 0 and 1 and controls the correlation between true precipitation and the synthetic forecasts. Thus, the noise term, ε_m, controls the correlation between the true precipitation (*P*) and the issued forecast over the period 1952–1986. For further details on the analytical derivation of how α and β control the correlation and the forecast overconfidence, see *Weigel et al.* [2008]:

$$X_{t,m}^i = \alpha^i \cdot x_t + \varepsilon_{\beta,t}^i + \varepsilon_{t,m}^i \tag{4}$$

$$\overline{X}_t^i = \sum_{m=1}^{M} X_{t,m} \tag{5}$$

$$P_t^i = \overline{X}_t^i \cdot \sigma_P + \mu_P \tag{6}$$

The two noise terms ε_β and ε_m are added to the standardized precipitation after adjusting it based on the skill parameter α (equation (4)). The members of the ensemble, $X_{t,m}^i$, are then averaged to obtain the ensemble

mean X_t^i (equation (5)) and are transformed back using the observed winter climatology of precipitation (equation (6)). It is important to note that the generated precipitation preserves the observed climatology, μ_P and σ_P, for any given value of α and β. We assume the precipitation forecast ensemble for each winter season, t, to constitute 100 members ($M = 100$) following Gaussian distribution. The conditional mean of the forecasted precipitation is equal to the observed precipitation if $\alpha = 1$ and $\beta = 0$. Thus, by adjusting these two parameters, α and β, the center and the spread of the precipitation forecast ensemble are controlled. For instance, a precipitation forecast with $\alpha = 0.9$ and $\beta = 0$ represents a well-dispersed highly skillful forecast, while $\alpha = 0.5$ and $\beta = 0.85$ represents an overconfident moderately skillful forecast. By assuming different α and β, we synthetically generate climate forecasts having different skills for the Tarboro site. In the next section, we discuss how streamflow forecasts could be developed using the precipitation forecasts for a given skill.

2.3. Single Model Streamflow Forecasts Development

To begin with, using the procedure detailed in section 2.1, we first obtain the "true" streamflow, Q_t^j, where j denotes the streamflow model, by using the observed precipitation and PET for the 35 year period using the population parameters (Table 1) and the chosen value of f, which basically controls the measurement error in the "true" streamflow. Based on the details given in section 2.2, we then generate 35 years of synthetic climate forecasts based on the chosen α and β. The "true" streamflow, Q_t^j, is split into two sets with the first 20 years ($t = 1, 2, \ldots, n_c$; n_c denotes the number of years of calibration) of flow being used for calibration and the remaining 15 years ($t = n_c + 1, n_c + 2, \ldots, n$; n denotes the number of total years of evaluation) for validation. Following that, we estimate the model parameters, $\hat{\theta}_j^i$, using Q_t^j, precipitation forecasts (P_t^i), where i denotes a precipitation forecasting model, and PET_t by minimizing the sum of squares of errors between Q_t^j and the calibrated flow $\hat{Q}_t^j$ over the 20 year calibration period. The calibrated parameters, $\hat{\theta}_j^i$, are subsequently used with precipitation forecasts (P_t^i) and PET_t to estimate the forecasted streamflow, $Q_{SM-t}^{i,j}$, by the individual model j for the validation period. It is important to note that the forecasted streamflow by a given model, j, could vary depending on the skill of precipitation forecasts, P_t^i, which is determined by α and β equations (1)–(6).

We also consider a conceptual watershed model, *abcd*, for estimating the single model forecasts of streamflow. The "*abcd*" model originally suggested by *Thomas* [1981] has been employed by various monthly and annual water balance studies [*Vogel and Sankarasubramanian*, 2000; *Sankarasubramanian and Vogel*, 2002a]. For details, see *Sankarasubramanian and Vogel* [2002b]. For forecasting the winter streamflow using "*abcd*" model, apart from P_t^i and PET_t, the model also requires initial soil moisture and groundwater states, S_{t-1} and G_{t-1}, over the calibration and validation period. We upfront develop these estimates, S_{t-1} and G_{t-1}, over the entire 35 year period by simulating the "*abcd*" model at seasonal time scale using observed seasonal precipitation and potential evapotranspiration and the model parameters (shown in Table 1). Thus, for forecasting each year winter streamflow, we use the simulated initial soil moisture and groundwater states from the fall season along with P_t^i and PET_t for performing calibration and validation. This ensures the forecasted streamflow using "*abcd*" model does not use any information available during the winter forecasting season. Given the single model streamflow forecasts, we combine them next to develop multimodel streamflow forecasts.

2.4. Multimodel Precipitation and Streamflow Forecasts Development

Both GCMs and watershed models invariably contain errors due to different sources that include quality of input data, initial states, uncertainty in parameter estimation and the actual model structure itself [*Feyen et al.*, 2001; *Devineni and Sankarasubramanian*, 2010]. Recently, multimodel combinations have emerged as a way to reduce model uncertainty by combining multiple models to obtain improved predictions [*Georgakakos et al.*, 2004; *Ajami et al.*, 2007; *Devineni et al.*, 2008; *Li and Sankarasubramanian*, 2012]. Multimodel predictions are also able to capture the strength of single models resulting in improved predictability [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Devineni et al.*, 2008].

For combining multiple models, there are various methods using weighted averages, including simple or weighted average of single model predictions [*Georgakakos et al.*, 2004; *Shamseldin et al.*, 1997; *Xiong et al.*, 2001]. Other studies have explored statistical techniques such as multiple linear regression [*Krishnamurthi et al.*, 1999] and Bayesian model averaging [*Duan et al.*, 2007] for multimodel combinations. In this study,

multimodel combinations are obtained from single models by using weights which are obtained based on the performance of the single model over the calibration period [*Li and Sankarasubramanian*, 2012]:

$$MSE_i = \sum_{t=1}^{n_c} (P_t - P_t^i)^2 \tag{7}$$

$$W_i = \frac{MSE_i^{-1}}{\sum_{i=1}^{n_m} MSE_i^{-1}} \tag{8}$$

$$P_{MM-t} = \sum_{i=1}^{n_m} P_t^i * W_i \qquad t = n_c+1,\ n_c+2,\ \ldots, n \tag{9}$$

Given the precipitation forecasts, P_t^i, over the calibration period (20 years), we compute the skill of the synthetic precipitation forecast by computing the mean square error between the forecasted mean, P_t^i, and the true precipitation, P_t, available for the Tar River at Tarboro using equation (7). Given that we have n_m climate forecasts from the synthetic scheme in section 2.2, we obtain weights for individual precipitation forecasting scheme by giving higher weights for the best-performing model (equation (8)). Using the weights, W_i, obtained for each model, we obtain multimodel precipitation forecasts, P_{MM-t}, over the validation period (equation (9)). One could obtain further improvements in multimodel precipitation forecasts by pursuing optimal model combination [*Rajagopalan et al.*, 2002] or optimal model combination conditioned on the predictor state [*Devineni and Sankarasubramanian*, 2010]. Those approaches are not pursued here, since the focus of this article is to compare the three strategies proposed in section 1. Application of such methods will only result in further improvements in multimodel predictions:

$$MSE_{i,j} = \sum_{t=1}^{n_c} (Q_t^j - Q_{SM-t}^{i,j})^2 \tag{10}$$

$$W_{i,j} = \frac{MSE_{i,j}^{-1}}{\sum_{\forall i,j} MSE_{i,j}^{-1}} \tag{11}$$

$$Q_{MM-Q-t} = \sum_{\forall i,j} Q_{SM-t}^{i,j} * W_{i,j} \qquad t = n_c+1,\ n_c+2,\ \ldots, n \tag{12}$$

Similar to developing multimodel precipitation forecasts, we also combine streamflow forecasts, $Q_{SM-t}^{i,j}$ $(i = 1, 2, \ldots, n_m; j = 1, 2, 3)$, developed from individual hydrological models using different precipitation forecasts. To begin with, we first compute mean square error between the single model prediction, $Q_{SM-t}^{i,j}$ and the true flow for the model, Q_t^j, over the calibration period ($t = 1, 2, \ldots, n_c$). This results in a total of n_m*3 MSE estimates from different combinations of precipitation forecasts forced with various streamflow prediction models. These MSE estimates (equation (10)) are converted into weights (equation (11)) for each forecast by giving higher weights for the forecasts that perform the best. Given the weights, $W_{i,j}$, the individual model predictions are converted into multimodel forecasts using equation (12) over the validation period. This basically gives the multimodel forecasts corresponding to the second strategy *MM-Q*.

2.5. Performance Evaluation of Single Model Forecasts and Multimodel Forecasts

To evaluate the three strategies proposed in section 1 for reducing the uncertainty in streamflow predictions, we repeat the experimental design, outlined in Figure 1, 100 times to get 100 mean square errors (MSE) ($MSE_{i,j}$) for a given precipitation forecasting scheme (i) and the candidate watershed model (j). Synthetic precipitation forecasts with different skills could be generated using equations (1)–(6) for the 35 year period with each year forecast being represented with 100 realizations. The ensemble mean is used as the forecast input for the streamflow prediction models. The first model is the single model setup in which the individual precipitation forecasts (P_t^i) are forced with one of the streamflow prediction models (Table 1) to

Table 2. Summary of Streamflow Forecasts Developed From Different Climatic and Hydrologic Model Combinations for the Synthetic Study

Model Indices-Schemes	Brief Description
$Q^{i,j}_{SM}$	Streamflow is obtained by forcing the individual streamflow model "*j*" with a single climate model input "*i*"
Q^{j}_{MM-P}	Streamflow is obtained by forcing the individual watershed model "*j*" with the multimodel, P_{MM} as the input
$Q_{MM\text{-}Q}$	Streamflow is obtained by combining all $Q^{i,j}_{SM}$
$Q_{MM\text{-}PQ}$	Streamflow is obtained by combining all Q^{j}_{MM-P}

obtain the modeled streamflow $(Q^{i,j}_{SM-t})$. Thus, in the individual model streamflow predictions denoted as $Q^{i,j}_{SM-t}$ (Table 2), there is no combination of precipitation or streamflow forecasts from multiple models.

The first strategy, *MM-P*, reduces the input uncertainty first by combining the synthetic precipitation forecasts to develop multimodel precipitation forecasts (equation (9)), P_{MM-t}, which are then used as an input with the streamflow prediction model (Table 1), *j*, to obtain streamflow (Q^{j}_{MM-P-t}). The second strategy, *MM-Q*, reduces the uncertainty in streamflow prediction by first ingesting the individual precipitation forecasts (without multimodel combination) with the individual watershed models and then combines the individual streamflow $(Q^{i,j}_{SM-t})$ to obtain a multimodel streamflow (equation (12)) which is denoted by Q_{MM-Q-t}. The third strategy, *MM-PQ*, combines all Q^{j}_{MM-P-t} obtained from the strategy *MM-P* to obtain $Q_{MM-PQ-t}$. The multimodel modeled streamflow, Q^{j}_{MM-P-t} from the *MM-P*, are combined by evaluating their performance against true streamflow (similar to equations (10)–(12) to give a single streamflow time series denoted by $Q_{MM-PQ-t}$.

In summary, we have streamflow predictions from single models, $Q^{i,j}_{SM}$ and three multimodel combinations Q^{j}_{MM-P}, Q_{MM-Q}, and Q_{MM-PQ} as shown in Table 2. The performance of these streamflow predictions is evaluated using mean square error, which is computed based on the streamflow predictions and the true model streamflow Q^{j}_{t} during the validation period $t = n_{c+1}, \ldots, n$. For each precipitation forecasts *i* and hydrologic model index *j*, the evaluation of streamflow predictions $Q^{i,j}_{SM}$ gives a single value of mean square error. The first multimodel combination, *MM-P* (denoted by Q^{j}_{MM-P}) yields mean square error for each hydrologic model index, *j*, under a given realization. The second and third multimodel combinations denoted by Q_{MM-Q} and Q_{MM-PQ} each yield one mean square error under a given realization. For the results, the experimental design outlined in Figure 1 is repeated 1000 times for all single and multimodel combinations to obtain 1000 realizations of mean square error values for evaluating different strategies.

3. Results and Analyses

In this section, we compare the performance of single models and the three multimodel strategies based on the estimated mean square error from 1000 realizations. Section 3.1 presents results from three precipitation forecasts forced with the linear streamflow prediction model (Table 1). Under section 3.2, we consider three candidate streamflow prediction models (Table 1) with the true streamflow being generated from one of the hydrologic models. Thus, under this case, we explicitly consider uncertainties across climate models and hydrologic models by analyzing the proposed three strategies, *MM-P*, *MM-Q*, and *MM-PQ*, for reducing the uncertainty in streamflow forecasts based on MSEs from 1000 realizations. This helps us to identify the right strategy that will reduce uncertainty in streamflow forecasts by considering both input (precipitation) uncertainty and output (streamflow) uncertainty.

3.1. Uncertainty Reduction in Streamflow Forecasts With Known Hydrologic Model

As discussed earlier, we consider the linear streamflow prediction model (Table 1) to be the true model as well as the candidate watershed model with the measurement error being equal to zero ($f = 0$ in Table 1). Since we employ the same stochastic streamflow generation model for both true streamflow generation and for estimation, the uncertainty in streamflow forecasts primarily arises from precipitation forecasts. For this purpose, we generate three precipitation forecasts with different skills by varying α and β. Selected values of α and β are very realistic to the values exhibited by GCMs [*Wang et al.* 2013] as well as to the values suggested by *Weigel et al.* [2008]. *Weigel et al.* [2008] show the minimum number of models (around 3–5) required to result in improved multimodel performance. Each precipitation forecast is used with the candidate model—linear streamflow prediction model—to estimate three single model streamflows denoted by Q^{i}_{SM}. The first multimodel streamflow, Q_{MM-P}, is derived first by combining the three precipitation forecasts and then the combined precipitation forecasts, P_{MM}, is forced with the linear model. The second multimodel

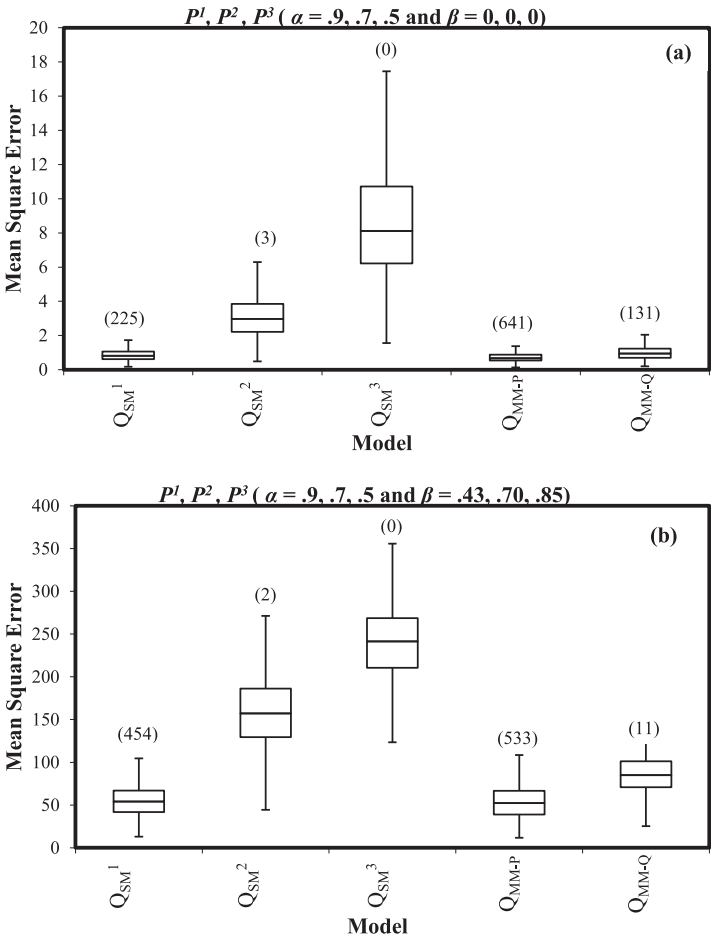

Figure 2. Box-plots of MSEs in estimating true streamflow under known hydrologic model structure being forced by (a) well-dispersed ($\beta = 0$) precipitation forecasts having different skills (α) and (b) overdispersed ($\beta \neq 0$) precipitation forecasts having different skills. Numbers above the box-plots show the number of times each forecasting model resulted with the lowest MSE out of the 1000 realizations.

combination streamflow, Q_{MM-Q}, is derived by combining the streamflow from all the single model (Q^i_{SM}; $i = 1, 2, 3$) forecasts.

We first (Figure 2) consider precipitation forecasts with varying skills by adjusting values of α and also by allowing the forecasts to be well dispersed ($\beta = 0$) (Figure 2a) or overconfident (Figure 2b) by choosing nonzero values of β. Nonzero β results in overconfidence forecasts as it shifts each year forecast further away from the observation. It is important to note that in evaluating the precipitation forecasting schemes, we compare their ability to predict the true streamflow obtained with the true climate forcings (i.e., observed precipitation and potential evapotranspiration). It is obvious from Figure 2a, skillful precipitation forecasts have lesser MSEs compared to the streamflow estimated from precipitation forecasts having poor skill. Comparing Figures 2a and 2b, MSEs produced by a well-dispersed precipitation forecast have a lower MSE values compared to the overconfident precipitation forecasts. This is primarily due to the ensemble spread of the precipitation forecasts becoming narrower under increased values of β.

The single models perform better as the skill (α) of the precipitation increases. We also see that the two multimodel streamflow forecasts (*MM-P* and *MM-Q*) perform better than single model streamflow forecasts, $Q^2_{SM}$and Q^3_{SM}, due to the reduction in climate model uncertainty. Since the skill of the precipitation forecasts varies due to α, we have used different weights to combine the precipitation forecasts (equations (7)–(9)) and streamflow forecasts (equations (10)–(12)) in the multimodel combination. By giving higher weights for the best-performing model under the calibration period, multimodel precipitation forecasts and multimodel streamflow forecasts perform better than the best single model (model with the highest value of α in Figure 2) forecasts.

Figure 2 also shows the number of times each model resulted with the lowest MSE value out of the 1000 realizations. From the numbers above the box-plot, we clearly see that *MM-P* performs even better than the best-performing individual model forecasts (Q^1_{SM}). This shows that the multimodel combination Q_{MM-P} (reducing input uncertainty first) performs better than the best single model and Q_{MM-Q}. This implies that for systematic uncertainty reduction in developing streamflow forecasts, we first need to reduce input (precipitation) uncertainty using multimodel combination, since the reduced uncertainty in the inputs to the watershed model results in better accuracy in predicting the true streamflow under known hydrologic model structure. We also infer that this is true with the precipitation forecasts being well dispersed

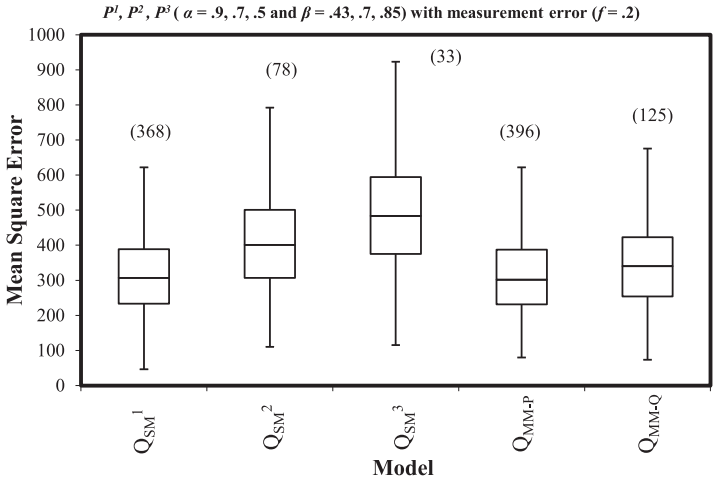

Figure 3. Box-plots of MSEs in estimating true streamflow under known linear model structure with measurement error ($f = 0.2$) under precipitation forecasts that are overdispersed ($\beta \neq 0$) precipitation forecasts having different skills. Numbers above the box-plots show the number of times each forecasting model resulted with the lowest MSE out of the realizations.

(Figure 2a) or overconfident (Figure 2b). Furthermore, though multimodel streamflow forecasts, *MM-Q*, performs poorly in comparison to *MM-P* and the best single model (Q^2_{SM}), it provides a much better performance in comparison to the other two individual ($Q^2_{SM}$and Q^3_{SM}) models.

Figure 3 considers the measurement error by selecting the term $f = 0.2$ in the synthetic streamflow generation and analyzes the performance of different streamflow forecasting models with precipitation forecasts being overconfident with varying β. Increasing f beyond 0.2 resulted in generation of higher percentage of negative flows and also produced true flows without adhering to any of the candidate model structures. We are not considering well-dispersed forecasts ($\beta = 0$) anymore since most of the precipitation forecasts from climate models are often overconfident [*Devineni and Sankarasubramanian*, 2010]. Hence, it is meaningful to consider that case for detailed analyses. Further, *Weigel et al*. [2008] clearly showed that if the climate forecasts are well dispersed, though it is unrealistic, then multimodel climate forecasts cannot outperform the best-performing single model forecasts. Hence, we limit our analyses to a more realistic scenario with precipitation forecasts having β not equal to zero.

From Figure 3, as the forecasts are overconfident, we see that *MM-P* performs better than *MM-Q* and the best-performing single model streamflow forecasts based on the spread of the MSE as well as based on the number of times each scheme had lowest MSE. Increased β values result with the ensemble spread being narrower and the ensemble mean of the precipitation forecasts being away from the observed values. Forcing such precipitation forecasts into watershed models result with inaccurate streamflow predictions. However, combining more number of precipitation forecasts increases the multimodel ensemble spread, thereby forcing the multimodel mean approach toward the true precipitation values. These findings are completely in line with *Weigel et al*. [2008] who showed that multimodel combinations result in better predictions as the dispersion factor, β, increases. Thus, forcing the watershed model with multimodel precipitation forecasts result in better streamflow forecasts as seen in the case of $Q_{MM\text{-}P}$. However, the performance of $Q_{MM\text{-}Q}$ is not as good as $Q_{MM\text{-}P}$ (Figure 3), since the former does not reduce the uncertainty in precipitation forecast before forcing it in the watershed model. The analyses presented in Figures 2 and 3 clearly shows that under known hydrologic model structure, the performance of multimodel, *MM-P*, is better than that of best-performing single model streamflow forecasting scheme. However, in reality, the forecaster does not know the true hydrologic model structure. Thus, it is prudent to consider multiple hydrologic models for forecasting streamflow. The next section considers multiple hydrologic models with the true streamflow being generated by one of the candidate models.

3.2. Sources of Model Uncertainty—Climate Models and Hydrologic Models

In the previous section, we considered the model uncertainty only from the climate models alone. Here we consider model uncertainties arising from both climate models as well as from the hydrologic model. For this purpose, we consider all three models (Table 1) to be candidate models with the true streamflow being generated by either linear model or "*abcd*" watershed model. Similar to the previous section, we consider three precipitation forecasts having different skills. Each precipitation forecast is forced with three streamflow prediction models to develop nine streamflow forecasts, $Q^{i,j}_{SM}$ where $i = 1, 2, 3$ and $j = 1, 2, 3$. We also combine the three precipitation forecasts using equations (7)–(9) and use the combined precipitation forecast with each candidate model to develop three multimodel streamflows (Q^j_{MM-P} with $j = 1, 2, 3$).

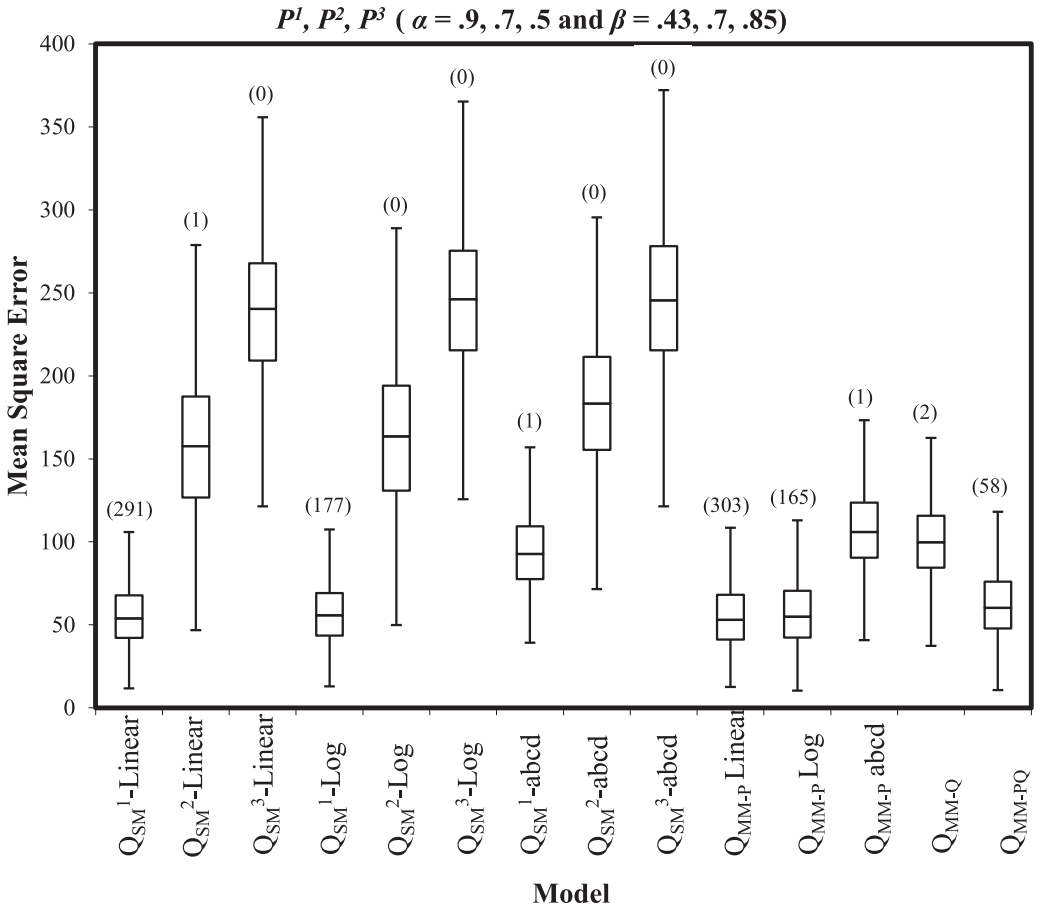

Figure 4. Box-plots of MSEs for 1000 realization for both single model streamflow forecasts and for three multimodel combination strategies (*QMM-P*, *QMM-Q*, and *QMM-PQ*) with the true streamflow arising from linear model with no measurement error ($f = 0$) with precipitation forecasts having different skill (α) and being overconfident (β).

Streamflow forecasts from nine single models $Q_{SM}^{i,j}$ are combined using equations (10–12) to develop multimodel forecasts (Q_{MM-Q}). Similarly, streamflow forecasts obtained from three multimodel precipitation forecasts, Q_{MM-P}^{j}, are combined separately (using equations (10–12)) to develop multimodel streamflow forecasts, Q_{MM-PQ}, which reduces first the input uncertainty followed by hydrologic model uncertainty. Thus, in this section, we present MSEs from 14 models that include nine single models and five multimodels. All the multimodel combinations on precipitation/streamflow are obtained using weights dependent upon the skill of the forecast during the calibration period.

Figure 4 presents the box-plots of MSEs for 14 models with the linear model as the true model having no output error ($f = 0$). From Figure 4, we can see that Q_{MM-P}^{Linear} and $Q_{SM}^{1-Linear}$ are the two best-performing models, which is to be expected, since the true flow arises from the linear model. We can also see that Q_{SM}^{abcd} and Q_{MM-P}^{abcd} perform the worst across all the models. This is partly because the true streamflow is generated from a linear and "*abcd*" model is a nonlinear water balance model. The multimodel streamflow forecasts, Q_{MM-P}, that reduces first the input uncertainty performs better than their respective streamflow forecasts obtained with individual model precipitation forecasts alone. For example, in the case of the linear candidate model, Q_{MM-P}^{Linear} performs better than the best single model, $Q_{SM}^{1,Linear}$. Similarly, log-linear and "*abcd*" candidate model forced with multimodel precipitation forecast perform better than their respective counterpart (Q_{MM-P}) in the single model. Since we are using varying weights during combination, we assign higher weight to the precipitation forecast that exhibit good skill. Multimodel streamflow forecasts, Q_{MM-Q}, which reduces the hydrologic model uncertainty by combining the streamflow over all the nine single models, performs poorer than the multimodel combinations, Q_{MM-P} (except the abcd model) and Q_{MM-PQ}. This is to be expected since Q_{MM-PQ} estimates streamflow by first reducing input uncertainty and then the hydrologic model uncertainty by combining the streamflow from the multimodel combination Q_{MM-P} instead of

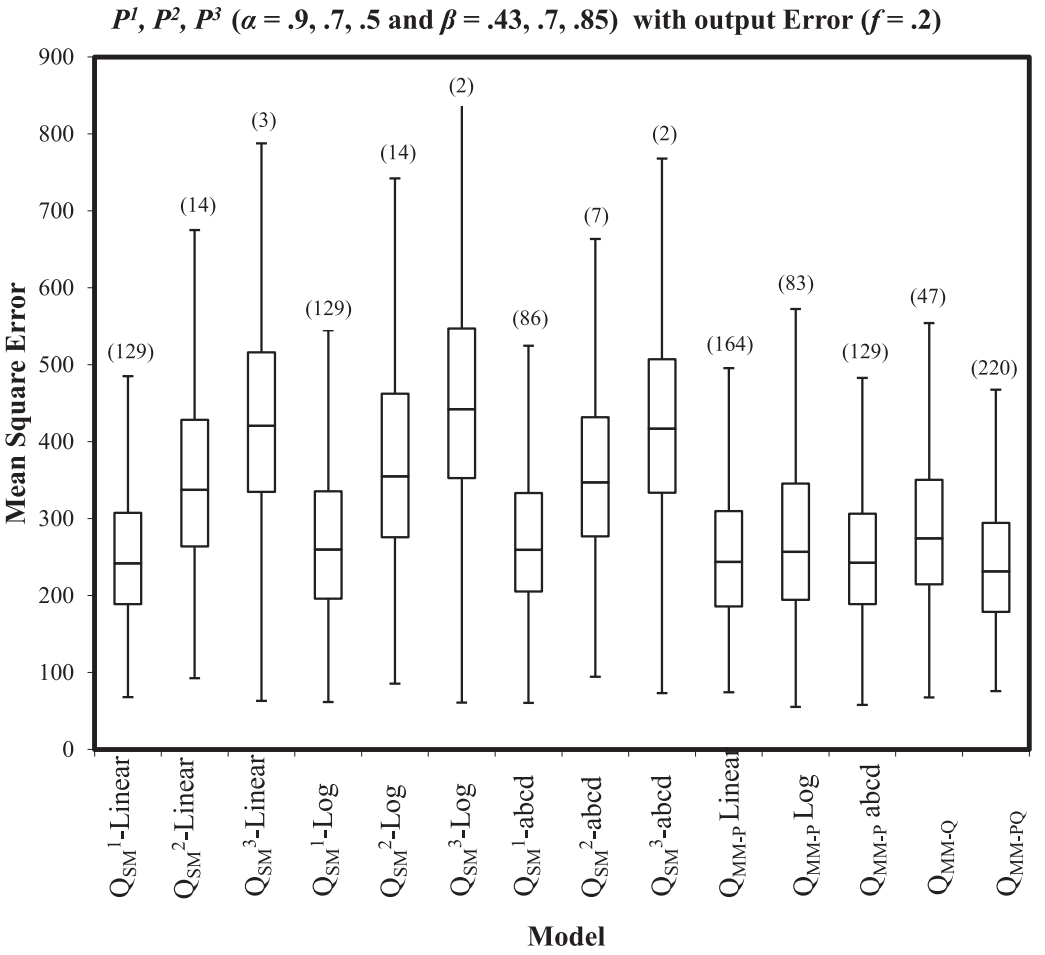

Figure 5. Same as Figure 4 but with streamflow measurement error ($f = 0.2$).

combining all the nine streamflow forecasts at once. Though Q_{MM-P}^{Linear} performs better than Q_{MM-PQ} in terms of the numbers of times it had the lowest MSEs, the spread of MSE based on the box-plot for both schemes is very similar. In real-world applications, we do not know the true model. Under such situations, it is desirable to consider Q_{MM-PQ}, since it systematically reduces the uncertainty by first reducing the uncertainty in the forcings, which is subsequently followed by reduction in hydrologic model uncertainty.

Figure 5 shows the box-plots of mean square error of all models with true streamflow generated by a linear model with measurement error ($f = 0.2$). From Figure 5, we can primarily see the MSE have increased even for the best linear model, since we have added noise to the true model streamflow. But the relative performance of the single and multimodel combinations has remained the same. Forcing the candidate watershed models with multimodel precipitation forecasts, *MM-P*, improve the performance of streamflow forecasts obtained from individual precipitation forecasts. However, the performance of $Q_{SM}^{1,Linear}$ and Q_{MM-P}^{Linear} is relatively worse compared to Figure 4, since the true flow from the linear model is corrupted with the measurement error ($f = 0.2$). Comparing the performance of multimodel combination strategies, we understand that *MM-PQ*, perform better than *MM-P* and *MM-Q* under the presence of measurement error. This could be understood both from the spread of the box-plot as well as from the number of times, *MM-PQ*, streamflow forecasting strategy performed better to the other schemes. This is different from Figure 4, which showed that under no measurement error, the strategy *MM-P* performs better than *MM-Q*. Thus, to reduce the uncertainty in streamflow prediction with the presence of measurement error, it is important to first reduce the input uncertainty, which needs to be followed with reduction in hydrologic model uncertainty.

Figure 6 shows the performance of the 14 models with the true flow arising from the "*abcd*" model with measurement error ($f = 0.2$). Under this case, the performance of single "abcd" models, Q_{SM}^{abcd}, is better than

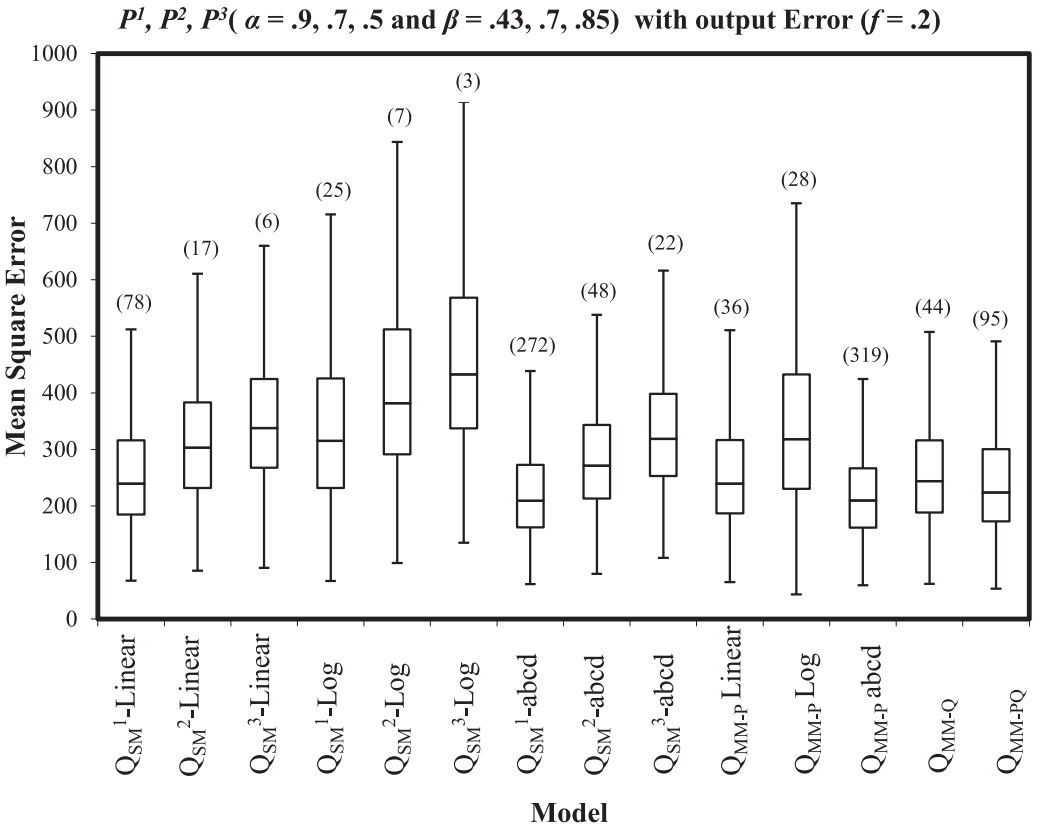

Figure 6. Same as Figure 5, but true flows arise from "abcd" model with output error ($f = 0.2$).

the linear or log-linear models which is to be expected since the true model is "*abcd*." We also see that the multimodel combination, Q^{abcd}_{MM-P}, performs the best among the multimodels which reduce input uncertainty first before reducing output uncertainty. The performance of Q_{MM-PQ} is also very similar to *MM-P* developed using "abcd" model. Though the median of Q_{MM-PQ} is slightly higher than that of Q^{abcd}_{MM-P}, the spread of the box-plot is very similar. Based on the number of times each model resulted with the lowest MSE out of the 1000 realizations, we understand that Q^{abcd}_{MM-P} outperforms Q_{MM-PQ}. But, upon comparing the MSE between the two schemes under each realization, we found the difference between MSEs is very small. In reality, since we do not know the true hydrologic model structure, it would be desirable to consider Q_{MM-PQ} since it reduces uncertainty in the forcings as well as the uncertainty in the hydrologic model structure. Further, the multimodel combination schemes given in equations (7)–(9) (for precipitation forecasts) and equations (10)–(12) (for streamflow forecasts) do not consider optimal combination as pursued by *Li and Sankarasubramanian* [2012] or *Rajagopalan et al.* [2002]. Multimodel combination schemes provided in section 2.4 only provides model averaging by weighing the individual models' based on past performance. Application of optimal combination methods [*Rajagopalan et al*., 2000] for multimodel forecasts will only result in improved performance, since those methods provide higher weights for the best-performing model. Thus, from Figure 6, it is again clear that reducing the input certainty first results in improved prediction if the underlying hydrologic model structure is known. However, in reality, we do not know the underlying hydrologic structure. Given similar performance of Q^{abcd}_{MM-P} and Q_{MM-PQ} based on the spread of the box-plot, it seems logical to consider Q_{MM-PQ} as a viable strategy in developing streamflow forecasts for a given basin.

To recapitulate the findings under this section, we have evaluated the proposed three multimodel combination strategies for uncertainty reduction in streamflow forecasts development based on two scenarios: (a) reducing only uncertainty in climate forecasts (Figures 2 and 3) and (b) reducing uncertainty under both climate forecasts and hydrological models (Figures 4 and 6). When considering no hydrological model

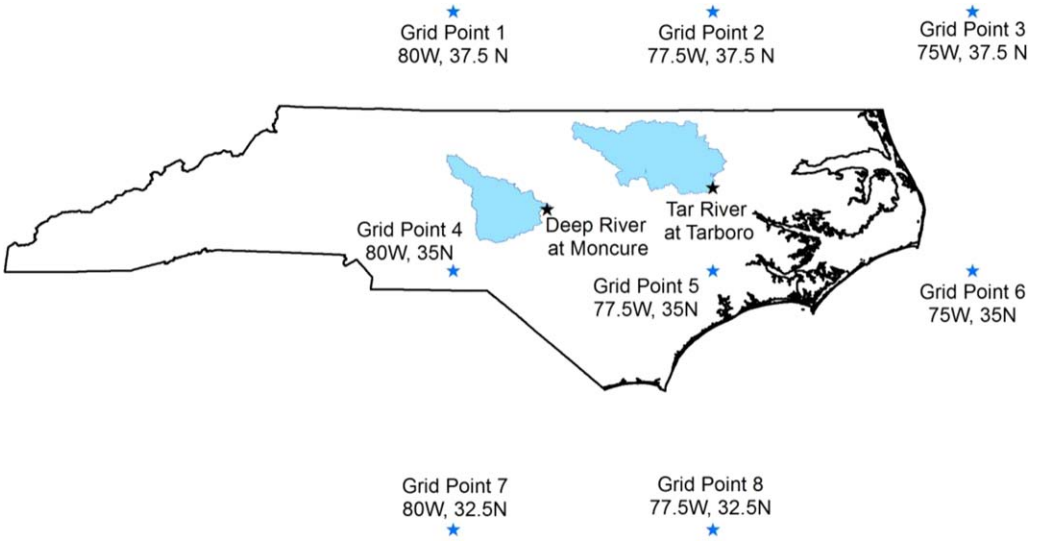

Figure 7. Location of the Tar River and Deep River in NC along with the latitude and longitude of the various grid points of precipitation forecasts used for the analysis.

uncertainty, reducing input uncertainty first by using *MM-P* strategy performs better than reducing output by using *MM-Q* regardless of whether the precipitation forecasts is well dispersed or overconfident. We can attribute this to the fact that reduced uncertainty in the inputs of the watershed model results in better accuracy in predicting the true streamflow arising from the same model. While we only consider the linear model to be the population and candidate model in section 3.1, the same conclusions can be reached if one uses other model, log or "*abcd*" model, as population and candidate model. In section 3.2, we consider multiple candidate hydrologic models focusing on uncertainty reduction under both climate forecasts and hydrological models, which allows us to evaluate the streamflow predictions from all three proposed strategies *MM-P*, *MM-Q*, and *MM-PQ*. We find that regardless of the candidate models employed, it is important to first reduce the input uncertainty, which needs to be followed with reduction in hydrologic model uncertainty in order to reduce the uncertainty in streamflow forecasts given that we do not know the true underlying hydrologic model structure. Both strategies *MM-P* and *MM-PQ* perform better in comparison to *MM-Q*. Comparing the performance of *MM-P* and *MM-PQ*, we understand that *MM-P* performs better for the most skillful precipitation forecasts. Pursuing optimal multimodel combination strategies as suggested by *Li and Sankarasubramanian* [2012] and *Rajgopalan et al*. [2002] will result only in further improvement in MM-PQ, since it will give higher weights for the best-performing model under MM-P. To confirm these findings in real settings, we intend to evaluate performance of three multimodel combination strategies in predicting the observed streamflow at Tar River at Tarboro using five coupled GCMs.

4. Application

In this section, we investigate how the proposed multimodel combination strategies perform in predicting the observed streamflow in two river basins, Tar River at Tarboro and Deep River at Moncure, in North Carolina (Figure 7). The observed winter seasonal streamflow and potential evapotranspiration are obtained for the Tar River at Tarboro site (02083500) and for the Deep River at Moncure (02102000) from the HCDN database of *Vogel and Sankarasubramanian* [2005]. Winter season is the most critical for filling up the reservoirs in the Southeastern US, since most of the winter precipitation is in the form of rainfall resulting in significant runoff. Further, the skill of climate forecasts is also significant during that season since the hydroclimatology of the region is significantly influenced by the ENSO. Given the basin is a rainfall-runoff dominated basin, this is a critical season for reducing the uncertainty in seasonal climate forecasts. Precipitation forecasts for the two basins are obtained from five coupled GCMs (Table 3) developed as part of the ENSEMBLES project [*Palmer et al*., 2004]. The precipitation forecasts from eight grid points over the domain (longitude: 80°W to 75°W; latitude: 32.5°N to 37.5°N) available at a monthly time step from 1981 to 1999 are considered for developing winter streamflow forecasts. The monthly precipitation forecasts issued in every November

Table 3. List of Coupled GCMs From Which the Seasonal Precipitation Forecasts Were Obtained for Forcing the Hydrologic Models (Source: DEMETER)

Ocean Model	Atmospheric Model	Institution	References
HOPE	IFS CY31R1	ECMWF	*Balmaseda et al.* [2008]
HadGEM2-O	HadGEM2-A	UKMO	*Collins et al.* [2008]
OPA8.2	ARPEGE4.6	MF	*Daget et al.* [2009]
MPI-OMI	ECHAM5	IFM-GEOMAR	*Keenlyside et al.* [2005]
OPA8.2	ECHAM5	CMCC-INGV	*Alessandri et al.* [2010]

from each GCM are averaged to obtain the seasonal (December–February) precipitation forecasts for each year for further analysis.

The precipitation from the five GCMs (Table 3) ($i = 1, \ldots, 5$) is used with the candidate models ($j = 1, \ldots, 3$) to obtain the single model streamflow $Q_{SM}^{i,j}$. There are 15 single models which are combined to obtain the multimodel streamflow denoted by Q_{MM-Q}. The multimodel precipitation data is based on the algorithm developed by *Devineni and Sankarasubramanian* [2010] which evaluates the skill of GCMs conditioned on the forecasted Nino3.4. The primary reason to consider ENSO as the conditioning variable is due to its influence in modulating the winter precipitation over the Southern United States [*Quan et al.*, 2006; *Devineni and Sankarasubramanian*, 2010]. For instance, if a GCM that performs well during a particular El-Nino condition (i.e., positive anomalous SST condition over the tropical Pacific), higher weights are given to that GCM as opposed to the other GCMs. For additional details on the algorithm, please see *Devineni and Sankarasubramanian* [2010]. The multimodel streamflow Q_{MM-P}^{j} is obtained by using the multimodel precipitation forecasts with the candidate streamflow model j. The final multimodel combination Q_{MM-PQ} is also obtained by combining the three streamflow forecasts, Q_{MM-P}^{j}, forced with multimodel precipitation forecasts.

The winter precipitation forecasts from the above five GCMs (Table 3) and the multimodel algorithm are statistically downscaled using principal component regression to obtain the seasonal average precipitation over the two watersheds. Since the precipitation forecasts from the eight grid points (Figure 7) of each GCM are spatially correlated, we employ principal component regression (PCR) to obtain downscaled time series of precipitation over the two river basins. Given that we employ PCR for downscaling the precipitation, any biases in GCM forecasts will be obviously removed at the basin-level downscaled values. The downscaled multimodel precipitation forecast performs better than the downscaled precipitation forecasts from individual CGCMs in capturing the variability with the observed precipitation (Table 4).

In order to evaluate the performance of various multimodel methods, we perform split-sample validation with the first 20 years of data (1961–1980) being used to calibrate the various hydrological models and the last 19 years (1981–1999) being used for validation. The parameters of each streamflow forecasting model in Table 1 obtained during the calibration period are used with the observed potential evapotranspiration and the single/multimodel precipitation forecasts (in Table 4) to obtain the winter streamflow forecasts from a given scheme.

Table 4. Correlation Between the Observed Precipitation and the Downscaled Precipitation for the Winter Season for Each GCM as Well as With the Downscaled Multimodel Precipitation Forecasts of *Devineni and Sankarasubramanian* [2010][a]

	Downscaled Precipitation Forecasts	
Models	Tar River	Deep River
CMCC	0.47	0.52
ECMWF	0.60	0.62
FRANCE	0.47	0.37
GEOMAR	0.52	0.56
UKMO	0.45	0.48
Multimodel	0.62	0.68

[a]Correlation above 0.33 is considered to be statistically significant.

Figures 8 and 9 show the observed winter streamflow and the forecasted streamflow by three multimodel schemes, $Q_{MM\text{-}P}$ (*linear*), $Q_{MM\text{-}Q}$, and $Q_{MM\text{-}PQ}$, for the Tar River at Tarboro and the Deep River at Moncure, respectively, under both the calibration (Figures 8a and 9a) and the validation (Figures 8b and 9b) periods. Based on Figures 8 and 9, we clearly see the best-performing individual model ends up improving its performance upon forcing with the multimodel climate forecasts. The performance of the models based on MSE and correlation under the validation period are summarized in Figures 10 and 11, respectively, for both basins. The performance of each scheme is summarized using the mean square error (Figure 10), which is obtained between the observed streamflow and the forecasted streamflow over the validation period. Model weights for each streamflow

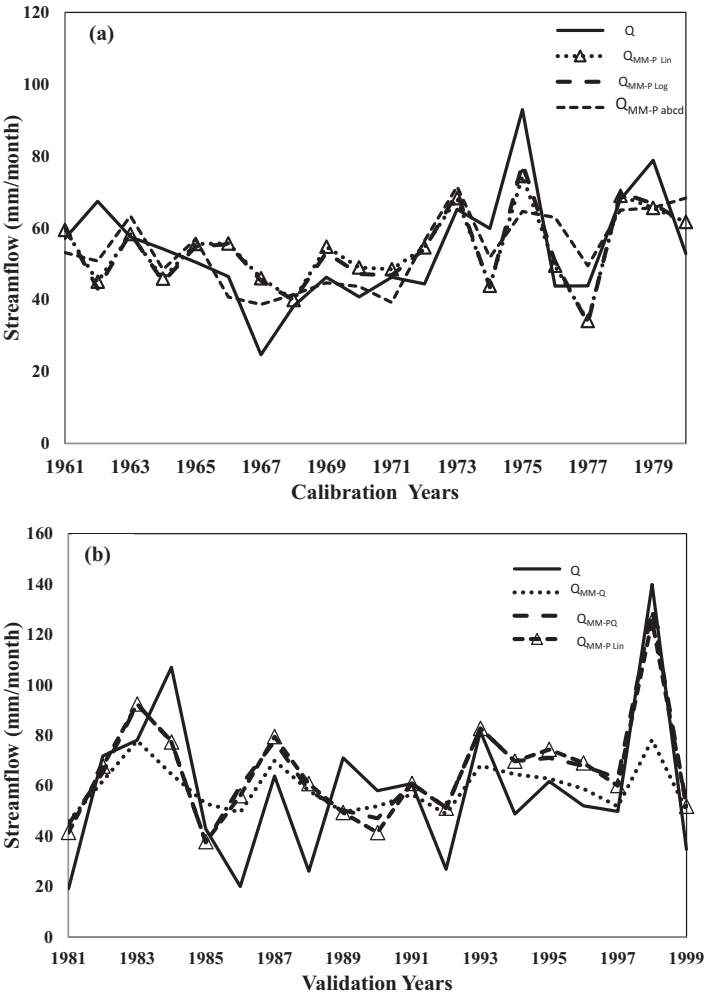

Figure 8. Performance of the proposed three multimodel schemes in forecasting seasonal streamflow for Tar River at Tarboro over the (a) calibration period (1961–1980) and (b) validation period (1981–1999).

forecasting model under strategies, *MM-Q* and *MM-PQ*, are summarized in Table 5.

As reflected by the mean square error (Figure 10), *MM-PQ* better predicts the high flow year in comparison to the performance of *MM-Q*. We can also see that linear model forced with the multimodel precipitation forecasts perform better than the *MM-Q* in predicting the observed streamflow. The performance of *MM-Q* is poor in predicting streamflow during wet years. Thus, from the synthetic study as well as from the application, out of the three multimodel strategies considered in the study, the two-step combination ($Q_{MM\text{-}PQ}$) performs the best in comparison to the rest of the multimodel ($Q_{MM\text{-}P}$ and $Q_{MM\text{-}Q}$) strategies. Comparing the correlation across different schemes (Figure 11), we infer that combination strategies, $Q_{MM\text{-}Q}$ and $Q_{MM\text{-}PQ}$, perform very similar to the strategy $Q_{MM\text{-}P}$. This indicates the multimodel precipitation forecasts developed by *Devineni and Sankarasubramanian* [2010] primarily provide information in improving the variability, whereas the strategy, $Q_{MM\text{-}PQ}$ overall reduces the conditional-bias in prediction resulting in reduced MSE (Figure 10).

Results presented in Figures 10 and 11 are pretty consistent with the findings from the synthetic setup. From Figure 10, it is clear that employing multimodel combination that reduces input uncertainty, $Q_{MM\text{-}P}$, performs better than their respective counterparts of single model forecasts and also performs better than the multimodel streamflow forecasts $Q_{MM\text{-}Q}$, which is obtained by the combination of 15 individual model streamflow forecasts. Comparing Tables 4 and 5, we clearly see that the best-performing individual climate forecasting model, ECMWF, receives higher weight than the rest of the model under *MM-Q* under each streamflow forecasting model. Similarly, the best-performing streamflow forecasting model, log-linear model for the Tar River basin and linear model for the Deep river basin, receives higher weight (Table 5) than the rest of the streamflow forecasting models under *MM-PQ*. Thus, we infer that reducing the input uncertainty (climate models) through multimodel combinations is more critical than reducing hydrological model uncertainty. The overall best performance is obtained by reducing input uncertainty followed by reducing hydrologic model uncertainty as indicated by the model *MM-PQ*. Thus, application of different multimodel strategies for predicting streamflow for the two watersheds confirms the findings from the synthetic study that two-step uncertainty reduction—reducing the uncertainty first with the climate forcings followed by the uncertainty in streamflow prediction arising from different watershed models—results in

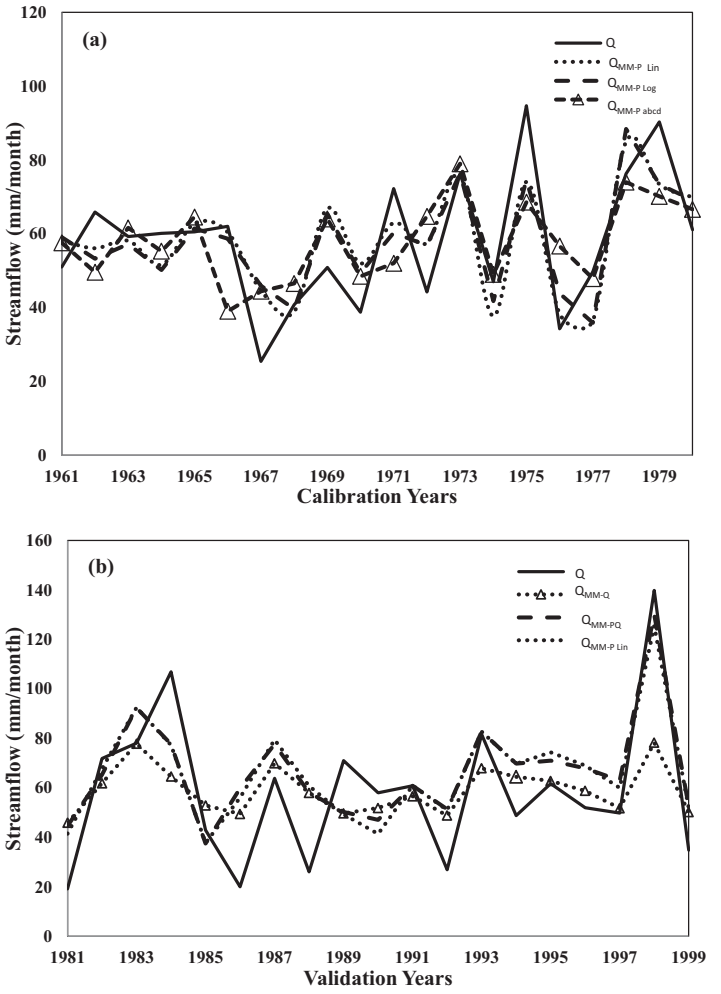

Figure 9. Performance of the proposed three multimodel schemes in forecasting seasonal streamflow for Deep River at Moncure over the (a) calibration period (1961–1980) and (b) validation period (1981–1999).

better performance than the multimodel combination of single model streamflow forecasts, which are obtained by forcing individual precipitation forecasts with candidate watershed models.

5. Discussion

Several sources of uncertainties influence streamflow forecasts development that includes climate forecasts [*Devineni and Sankarasubramanian*, 2010], selected downscaling model [*Yuan and Wood*, 2012], initial conditions [*Wood and Lettenmaier*, 2008; *Shukla et al*., 2012], hydrologic models [*Bohn et al*., 2010] and their calibration [*Ajami et al*., 2007; *Shi et al*., 2008], and streamflow measurement errors [*Li and Sankarasubramanian*, 2012]. For rainfall-runoff dominated basins, the primary source of uncertainty in seasonal streamflow forecasting relies on the climatic inputs—precipitation and temperature—followed by hydrologic model uncertainty [*Li et al*., 2009; *Sinha and Sankarasubramanian*, 2013]. On the other hand, basins dominated by snow melt, uncertainty on initial conditions could play an important role in seasonal streamflow forecasting [*Mahanama et al*., 2012]. To address these challenges, studies have considered Bayesian model averaging either on the climatic inputs [*Luo and Wood*, 2008], or on the hydrologic models [*Ajami et al*., 2007; *Marshall et al*., 2005, 2006; *Oudin et al*., 2006; *Najafi et al*., 2011]. Hence, most of the above studies focused only on uncertainty reduction in climate forecasts or on the hydrological models, but not on both at the same time by considering multiple climate models and hydrologic models for streamflow forecasts development. Further, in this context, we considered simple multimodel combination procedure that is conditioned on the overall model performance (equations (9) and (12)). We did not intend to consider advanced multimodel combination algorithms (e.g., *Devineni and Sankarasubramanian* [2010], for climate models or *Najafi et al*. [2011], for hydrologic models) since application of such methodologies uniformly across all the three strategies is expected to result in improved performance for each strategy. However, the relative improvements across the three strategies due to various multimodel combination algorithms is worthy of further investigation.

This study evaluates three different strategies, *MM-P*, *MM-Q*, and *MM-PQ*, for reducing the uncertainty in streamflow forecasts utilizing multiple climate forecasts and multiple hydrologic models. The skill in forecasting the streamflow is evaluated based on MSE that quantify the accuracy in prediction. But, under synthetic experiments, the relative performance of three different strategies is strictly discussed in terms of the spread of the MSE (based on the box-plots in Figures 2–6) for a given "true" streamflow trace. It is important

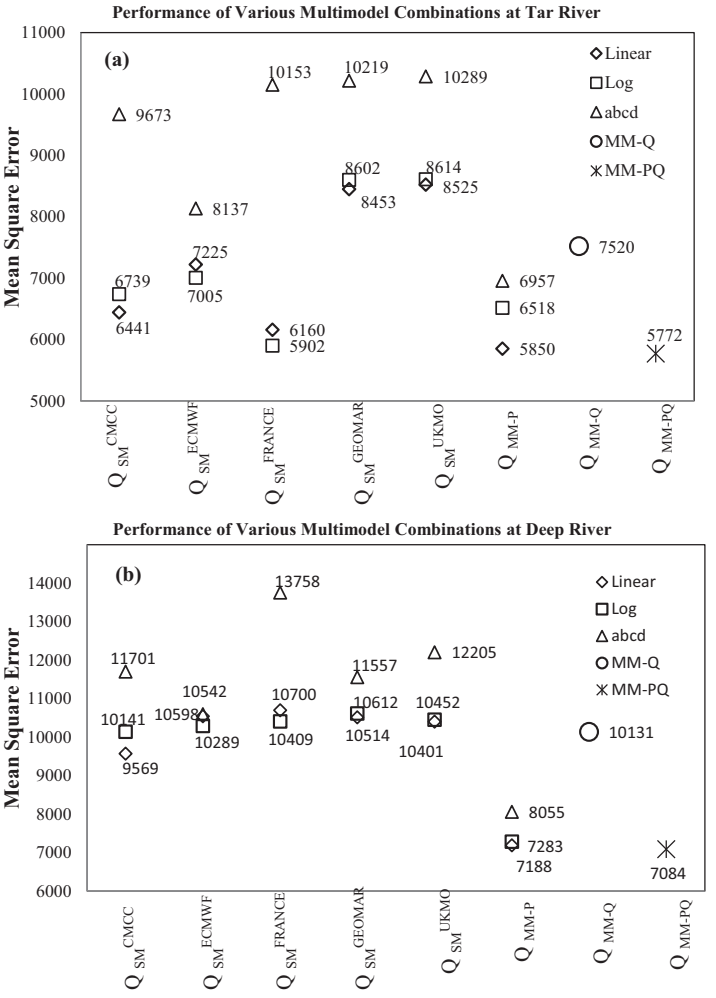

Figure 10. MSE of streamflow forecasts obtained from individual GCMs and multimodel combination schemes in predicting the observed streamflow for (a) Tar River at Tarboro and (b) Deep River at Moncure over the validation period (1981–1999).

to note that the skill in climate forecast is usually low, but the uncertainty across climate models (i.e., the range of prediction) is quite high. On the other hand, uncertainty in hydrologic models is relatively lower, but their accuracy/skill in predicting streamflow is high provided the inputs are proper. This study demonstrates through synthetic experiments that systematic uncertainty reduction (based on the spread of box-plots) in streamflow forecasts could be achieved by first reducing uncertainty across the climate models followed by reducing uncertainty across hydrologic models.

By applying the above three strategies for two watersheds in NC using retrospective climate forecasts from ENSEMBLES experiment with three streamflow forecasting models, we also show significant improvements in accuracy in predicting the observed streamflow by ingesting multimodel climate forecasts with three streamflow forecasting models and then combining the streamflow forecasts based on the multimodel climate forecasts to obtain single streamflow prediction. Thus, there is potential in extending the recommended strategy of *MM-PQ* from this study in a Bayesian model combination context [e.g., *Rajagopalan et al*., 2002; *Luo and Wood*, 2008; *Najafi et al*., 2011], so that uncertainty estimates (e.g., posterior distribution of MSE) in streamflow prediction accuracy could be developed even for application. We intend to consider this as part of our future study.

6. Concluding Remarks

Given that climate forecasts from multiple climate models are available, which could be ingested with multiple watershed models, the study outlined three possible strategies for reducing the uncertainty in streamflow prediction for developing multimodel streamflow forecasts. A systematic analysis is performed by evaluating the performance of three strategies under a synthetic streamflow setup with true streamflow being known as well as by evaluating their in predicting the observed winter streamflow for two river basins, Tar River at Tarboro and Deep River at Moncure, in NC. Based on the above analysis, we conclude the following:

1. Multimodel streamflow forecasts obtained by reducing input (climate forecast) uncertainty performs better than the streamflow predictions obtained from single models even if the single model having the true model's structure (Figures 2 and 3). This primarily stems from the uncertainty reduction on the forcings.

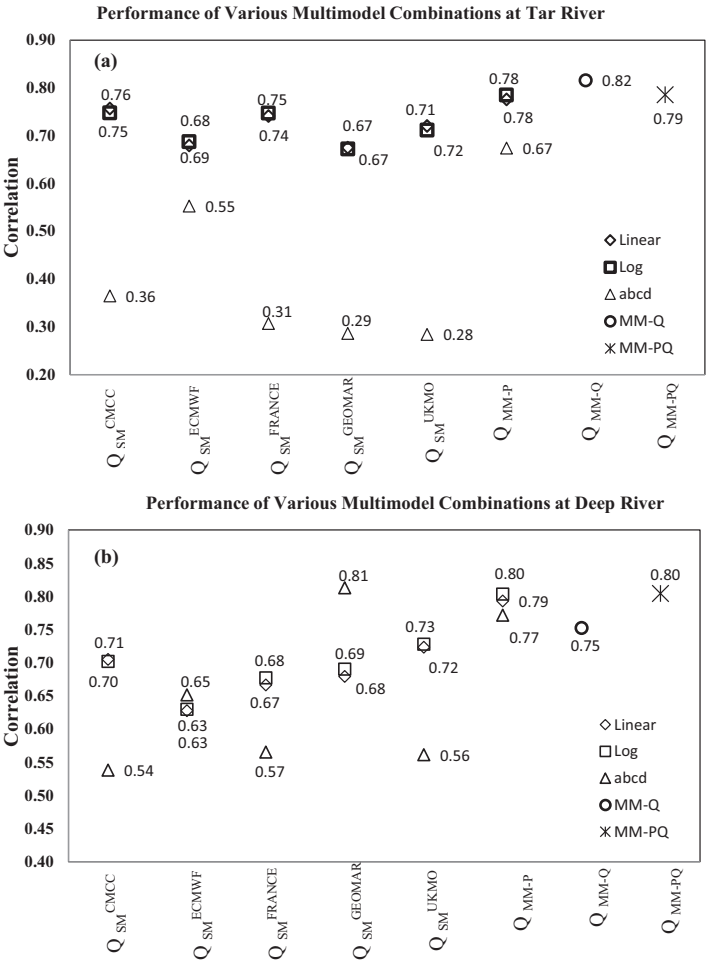

Figure 11. Correlation of streamflow forecasts obtained from individual GCMs and multimodel combination schemes in predicting the observed streamflow for (a) Tar River at Tarboro and (b) Deep River at Moncure over the validation period (1981–1999).

2. Reducing the input (climate forecast) uncertainty through multimodel combinations is more critical than reducing uncertainty in the hydrologic model structure (Figures 4–6). Thus, the multimodel combination strategy, *MM-P*, consistently performs better than *MM-Q*. This primarily arises due to the enormous role of precipitation explaining the variability in streamflow indicating the need for good precipitation forecast for basins in the rainfall-runoff regime.

3. Comparing the performance of the three outlined multimodel strategies, reducing input (climate forecast) uncertainty followed by reducing hydrologic model uncertainty (*MM-PQ*) provides better results than multimodel streamflow forecasts obtained by combining the streamflow forecasts developed using all single model streamflow forecasts forced with individual precipitation forecasts (*MM-Q*). Though *MM-PQ* performs better than *MM-P* under few selected

Table 5. Model Weights Used in the Multimodel Combination Schemes *MM-Q* and *MM-PQ*

		Weights for Each Model in Scheme MM-Q	
Streamflow Forecasting Model	Climate Forecasts	Tar River	Deep River
Linear	CMCC	0.063	0.069
	ECMWF	0.071	0.098
	FRANCE	0.065	0.065
	GEOMAR	0.055	0.065
	UKMO	0.057	0.069
Log	CMCC	0.062	0.066
	ECMWF	0.071	0.101
	FRANCE	0.066	0.064
	GEOMAR	0.055	0.063
	UKMO	0.057	0.067
abcd	CMCC	0.072	0.041
	ECMWF	0.092	0.082
	FRANCE	0.070	0.047
	GEOMAR	0.064	0.046
	UKMO	0.081	0.058
		Weights for Each Model in Scheme MM-PQ	
Linear	MM-P	0.337	0.393
Log	MM-P	0.351	0.368
"abcd"	MM-P	0.312	0.239

cases (Figure 5) and under application, the spread of MSE from the box-plot is very similar for both strategies. Given that we do not know the true hydrologic model structure, it is desirable to reduce the uncertainty in hydrological model structure by combining the streamflow forecasts available from *MM-P*.

Given that we do not know the true model structure, it is imperative to reduce model uncertainties in both sources—climate model and hydrologic model—for improving the streamflow predictions. The study provides a definitive outline on the development of multimodel streamflow forecasts using climate forecasts available from multiple GCMs by forcing with the different hydrologic models. Though we have demonstrated the results through synthetic study and by application for one basin, it would be interesting to evaluate these strategies with retrospective climate forecasts to understand how the skill of individual and multimodel streamflow forecasts over a large region. Our future effort will focus in evaluating this over the Southeast United States and incorporate this in the experimental inflow and storage forecasts (http://www.nc-climate.ncsu.edu/inflowforecast) which is currently being developed for water supply systems in NC.

Acknowledgments
This study was supported by the North Carolina Water Resources Research Institute. The writers also would like to thank the three anonymous reviewers whose valuable comments led to significant improvements in the manuscript.

References

Ajami, N. K., Q. 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.

Alessandri, A., A. Borrelli, S. Masina, A. Cherchi, S. Gualdi, A. Navarra, P. Di Pietro, and A. F. Carril (2010), The INGV-CMCC seasonal prediction system: Improved ocean initial conditions, *Mon. Weather Rev.*, *138*, 2930–2952.

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*, 1783–1796, doi:10.1175/BAMS-84-12-1783.

Bohn, T. J., M. Y. Sonessa, and D. P. Lettenmaier (2010), Seasonal hydrologic forecasting: Do multi-model ensemble averages always yield improvements in forecast skill?, *J. Hydrometeorol.*, *11*(6), 1357–1371, doi:10.1175/2010JHM1267.1.

Chowdhury, S., and A. Sharma (2009), Long-range Nin∼ o-3.4 predictions using pairwise dynamic combinations of multiple models, *J. Clim.*, *22*, 793–805.

Devineni, N., and A. Sankarasubramanian (2010), Improved categorical winter precipitation forecasts through multimodel combinations of coupled GCMs, *Geophys. Res. Lett.*, *37*, L24704, doi:10.1029/2010GL044989.

Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, *44*, W09404, doi:10.1029/2006WR005855.

Duan, Q., N. K. Ajami, X. 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.

Feyen, L., K. J. Beven, F. De Smedt, and J. Freer (2001), Stochastic capture zones delineated within the generalized likelihood uncertainty estimation methodology: Conditioning on head observations, *Water Resour. Res.*, *37*, 625–638.

Georgakakos, K. P., D.-J. Seo, H. V. Gupta, J. Schaake, and M. B. Butts (2004), Characterising streamflow simulation uncertainty through multimodel ensembles, *J. Hydrol.*, *298*(1–4), 222–241.

Goddard, L., A. G. Barnston, and S. J. Mason (2003), Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997–2001, *Bull. Am. Meteorol. Soc.*, *84*, 1761–1781.

Koster, R. D., et al. (2011), The second phase of the global land-atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill, *J. Hydrometeorol.*, *12*(5), 805–822.

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*, 1548–1550.

Li, W., and A. Sankarasubramanian (2012), Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination, *Water Resour. Res.*, *48*, W12516, doi:10.1029/2011WR011380.

Li, H., L. Luo, E. F. Wood, and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, *114*, D04114, doi:10.1029/2008JD010969.

Luo, L., and E. F. Wood (2008), Use of Bayesian merging techniques in a multi-model seasonal hydrologic ensemble prediction system for the Eastern U.S., *J. Hydrometeorol.*, *5*, 866–884.

Mahanama, S. P. P., B. Livneh, R. D. Koster, D. Lettenmaier, and R. H. Reichle (2012), Soil moisture, snow, and seasonal streamflow forecasts in the United States, *J. Hydrometeorol.*, *13*, 189–203, doi:10.1175/JHM-D-11-046.1.

Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, *41*, W10422, doi:10.1029/2004WR003719.

Marshall, L., A. Sharma, and D. Nott (2006), Modeling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, *42*, W11409, doi:10.1029/2005WR004613.

Marshall, L., A. Sharma, D. Nott (2007a), A single model ensemble versus a dynamic modeling platform: Semi-distributed rainfall runoff modeling in a Hierarchical Mixtures of Experts framework, *Geophys. Res. Lett.*, *34*, L01404, doi:10.1029/2006GL028054.

Marshall, L., A. Sharma, and D. Nott (2007b), Towards dynamic catchment modelling: A Bayesian hierarchical mixtures of experts framework, *Hydrol. Processes*, *21*(7), 847–861, doi:10.1002/hyp.6294.

Najafi, M. R., H. Moradkhani and I. W. Jung (2011), Assessing the uncertainties of hydrologic model selection in climate change impact studies, *Hydrol. Processes*, *25*, 2814–2826, doi:10.1002/hyp.8043.

Palmer, T. N., et al. (2004), Development of a European multimodel ensemble system for Seasonal-to-interannual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*, 853–872, doi:10.1175/BAMS-85-6-853.

Parrish, M., H. Moradkhani, and C. M. DeChant (2012), Towards reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resour. Res.*, *48*, W03519, doi:10.1029/2011WR011116.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *J. Hydrol.*, *320*, 62–83, doi:10.1016/j.jhydrol.2005.07.016.

Overgaard, J., D. Rosbjerg, and M. B. Butts (2006), Land-surface modeling in hydrological perspective—A review, *Biogeosciences*, *3*, 229–241..

Quan, X., et al., (2006), Diagnosing sources of US seasonal forecast skill, *J. Clim.*, *19*, 3279–3293, doi:10.1175/JCLI3789.1.
Rajagopalan, B., U. Lall, and S. E. Zebiak (2002), Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles, *Mon. Weather Rev.*, *130*, 1792–1811.
Regonda, S. K., B. Rajagopalan, M. Clark, and J. Pitlick (2005), Seasonal cycle shifts in hydroclimatology over the western United States, *J. Clim.*, *18*, 372–384.
Sankarasubramanian, A., and R. M. Vogel (2002a), Annual hydroclimatology of the United States, *Water Resour. Res.*, *38*(6), 1083, doi: 10.1029/2001WR000619.
Sankarasubramanian, A., and R. M. Vogel (2002b), Comment on the paper: Basin hydrologic response relations to distributed physiographic descriptors and climate, *J. Hydrol.*, *263*, 257–261.
Sankarasubramanian, A., U. Lall, and S. Espuneva (2008), Role of retrospective forecasts of GCM forced with persisted SST anomalies in operational streamflow forecasts development, *J. Hydrometeorol.*, *9*, 212–227.
Shamseldin, A. Y., K. M. O'Connor, and G. C. Liang (1997), Methods for combining the outputs of different rainfall-runoff model, *J. Hydrol.*, *197*, 203–229.
Shi, X., A. W. Wood, and D. P. Lettenmaier (2008), How essential is hydrologic model calibration to seasonal streamflow forecasting?, *J. Hydrometeorol.*, *9*(6), 1350–1363.
Sinha, T., and A. Sankarasubramanian (2013), Role of climate forecasts and initial land-surface conditions in developing operational streamflow and soil moisture forecasts in a rainfall-runoff regime: Skill assessment, *Hydrol. Earth Syst. Sci.*, *17*, 721–733, doi:10.5194/hess-17–721-2013.
Thomas, H. A. (1981), Improved methods for national water assessment, *Rep. contract WR 15249270*, U.S. Water Resour. Counc., Washington, D. C.
Vogel, R. M., and A. Sankarasubramanian (1999), On the validation of a watershed model, poster presentation at IUGG, Univ. of Buckingham.
Vogel, R. M., and A. Sankarasubramanian (2000), Scaling properties of annual streamflow in the continental United States, *J. Hydrol. Sci.*, *45*, 465–476.
Vogel, R. M., and A. Sankarasubramanian (2005), *Monthly Climate Data for Selected USGS HCDN Sites, 1951–1990*, Oak Ridge Natl. Lab. Distrib. Active Arch. Cent., Oak Ridge, Tenn., doi:10.3334/ORNLDAAC/810. [Available at http://www.daac.ornl.gov.].
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.
Vrugt, J. A., B. Ó. Nualláin, B. A. Robinson, W. Bouten, S. C. Dekker, and P. M. A. Sloot (2006), Application of parallel computing to stochastic parameter estimation in environmental models, *Comput. Geosci.*, *32*(8), 1139–1155, doi:10.1016/j.cageo.2005.10.015.
Wang, H., A. Sankarasubramanian, and R. S. Ranjithan (2013), Integration of climate and weather information for improving 15-day ahead accumulated precipitation forecasts, *J. Hydrometeorol.*, *14*(1), 186–202, doi:10.1175/JHM-D-11-0128.1.
Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Q. J. R. Meteorol. Soc.*, *134*, 241–260, doi:10.1002/qj.210.
Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, *35*, L14401, doi:10.1029/2008GL034648.
Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier (2002), Long-range experimental hydrologic forecasting for the Eastern United States, *J. Geophys. Res.*, *107*, 4429, doi:10.1029/2001JD000659.
Xiong, L. H., A. Y. Shamseldin, and K. M., O'Connor (2001), A nonlinear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system, *J. Hydrol.*, *245*(1–4), 196–217.
Xu, C. Y., and V. P. Singh (2004), Review on regional water resources assessment models under stationary and changing climate, *Water Resour. Manage.*, *18*, 591–612.
Yuan, X., and E. F. Wood (2012), Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast, *Water Resour. Res.*, *48*, W12519, doi:10.1029/2012WR012256.