

# Water Resources Research

## RESEARCH ARTICLE

10.1029/2019WR025228

### Key Points:

- This presents a methodology that uses remote sensing and distributed models to generate reservoir inflow forecasts in data-scarce regions
- This combines ensemble inflow forecasts with stochastic programming with recourse for hydropower optimization
- This studies the impact of uncertainty in precipitation forecasts and model parameters on hydropower production

### Correspondence to:

A. Koppa,  
akashkoppa@ucla.edu

### Citation:

Koppa, A., Gebremichael, M., Zambon, R. C., Yeh, W. W.-G., & Hopson, T. (2019). Seasonal Hydropower Planning for Data-Scarce Regions Using Multimodel Ensemble Forecasts, Remote Sensing Data, and Stochastic Programming. *Water Resources Research*, 55. <https://doi.org/10.1029/2019WR025228>

Received 6 MAY 2019

Accepted 12 OCT 2019

Accepted article online 24 OCT 2019

## Seasonal Hydropower Planning for Data-Scarce Regions Using Multimodel Ensemble Forecasts, Remote Sensing Data, and Stochastic Programming

**Akash Koppa<sup>1</sup> , Mekonnen Gebremichael<sup>1</sup>, Renato C. Zambon<sup>2</sup> , William W.-G. Yeh<sup>1</sup> , and Thomas M. Hopson<sup>3</sup> **

<sup>1</sup>Department of Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, CA, USA,

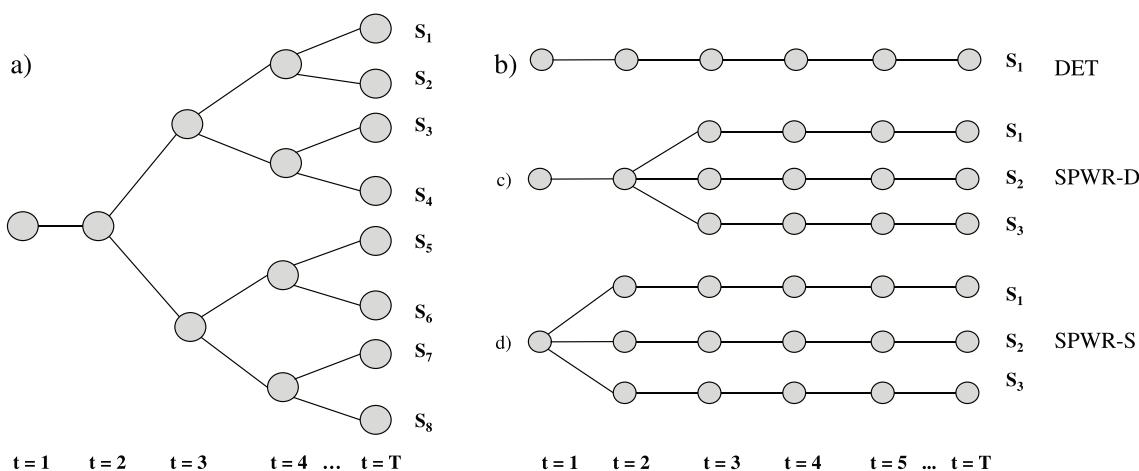
<sup>2</sup>Department of Hydraulic and Environmental Engineering, University of São Paulo, Brazil, <sup>3</sup>National Center for Atmospheric Research, Boulder, CO, USA

**Abstract** In data-scarce regions, seasonal hydropower planning is hindered by the unavailability of reliable long-term streamflow observations, which are required for the construction of inflow scenario trees. In this study, we develop a methodological framework to overcome the problem of streamflow data scarcity by combining precipitation forecasts from ensemble numerical weather prediction models, spatially distributed hydrologic models, and stochastic programming. We use evapotranspiration as a proxy for streamflow in generating reliable reservoir inflow forecasts. Using the framework, we compare three different formulations of inflow scenario structures and their applicability to data-scarce regions: (1) a single deterministic forecast, (2) a scenario fan with the first stage deterministic, and (3) a scenario fan with all stages stochastic. We apply the framework to a cascade of two reservoirs in the Omo-Gibe River basin in Ethiopia. Future reservoir inflows are generated using a 3-model 30-member ensemble seasonal precipitation forecast from the North American Multimodel Ensemble and the Noah-MP hydrologic model. We then perform deterministic and stochastic optimization for hydropower operation and planning. Comparing the results from the three different inflow scenario structures, we observe that the uncertainty in reservoir inflows is significant only for the dry stages of the planning horizon. In addition, we find that the impact of model parameter uncertainty on hydropower production is significant ( $0.14\text{--}0.18 \times 10^6$  MWh).

### 1. Introduction

In data-scarce regions, the lack of SF measurements hinders accurate prediction of water availability (Sivapalan et al., 2003) and hence reliable planning and management of reservoirs for purposes such as flood control, irrigation, and hydropower. For example, although sub-Saharan Africa has witnessed significant growth in installed hydropower capacity, the efficiency of the dams is low (Conway et al., 2015; Conway et al., 2017). In this study, we identify and address key challenges facing effective hydropower optimization in data-scarce regions. Specifically, the focus of the study is on hydropower planning at seasonal time scales. Conducted monthly, semiannually, or annually, seasonal hydropower planning is essential for generating optimal monthly release and storage decisions to fulfill a particular objective (e.g., maximization of hydropower production) or a set of objectives of a given reservoir system (Lund, 1996). For this purpose, multistage stochastic programming with recourse models are widely used (Yeh, 1985). Past studies have primarily focused on improving reservoir inflow generation and stochastic optimization techniques in data-rich catchments. In contrast, we combine global satellite remote sensing, ensemble seasonal precipitation forecasts, and stochastic optimization models to develop a framework that addresses critical issues hindering seasonal hydropower planning in data-scarce regions. Specifically, we identify the following challenges for seasonal hydropower planning in data-scarce regions and the approach we adopt to overcome each of them:

- 1. Generation of reservoir inflow scenarios**—In traditional seasonal hydropower planning, possible future reservoir inflows are derived from long-term SF records. These future scenarios are structured in the form of trees (Figure 1). In data-scarce regions, because of the lack of historical records, scenario trees cannot be constructed reliably. In this study, we develop a methodology to generate reservoir inflow scenarios in data-scarce regions by combining ensemble seasonal precipitation forecasts, a spatially distributed hydrologic model, and satellite-based estimates of precipitation and evapotranspiration.



**Figure 1.** A visual representation of different reservoir inflow scenario structures—(a) Scenario tree, (b) a single deterministic forecast, (c) first stage deterministic and the rest stochastic, and (d) all stages stochastic (SPWR-S).

2. **Consideration of different sources of uncertainty**—Seasonal hydropower planning using ensemble seasonal precipitation forecasts and hydrologic models introduces uncertainties in (a) precipitation forecasts, (b) estimates of initial hydrologic conditions (such as SWE and soil moisture), and (c) hydrologic model parameters. Quantifying these uncertainties is a major challenge in improving seasonal hydropower planning in data-scarce regions where SF data is unavailable. In this study, we explore the potential of satellite-based evapotranspiration as a proxy for SF in quantifying the uncertainties.
3. **Formulation of the hydropower optimization model**—The classical formulation of a stochastic programming with recourse model, used for seasonal hydropower planning, considers the first stage to be deterministic and the subsequent stages to be stochastic. In data-scarce catchments, this assumption may not be appropriate. In this study, we reformulate the stochastic programming with recourse model for hydropower planning to consider the uncertainty in seasonal reservoir inflow forecasts at the first/immediate stage.

In addressing these issues associated with seasonal hydropower planning in data-scarce regions, we answer the following research questions: (1) Can seasonal precipitation forecasts combined with remotely sensed estimates of precipitation and evapotranspiration generate reliable reservoir inflow scenarios in the absence of SF observations? (2) How do uncertainties in precipitation forecasts and model parameters impact seasonal reservoir inflow forecasts and hence hydropower production? (3) To what extent does incorporation of inflow uncertainty in the first/immediate stage of a stochastic programming with recourse model affect the release policy and hence hydropower production?

In the subsequent subsections, we present a detailed literature review of different methodologies for generating reservoir inflows at multiple time scales, the importance and approaches of incorporating different sources uncertainty, and different assumptions in formulating stochastic programming with recourse models.

### 1.1. Generation of Reservoir Inflow Scenarios

Scenario trees are generated from SF observations using different statistical methods such as (1) simulation-based methods which sample different scenarios from the distribution of the random variable (SF; Turgeon, 2005), (2) optimization techniques such as moment matching which match different moments of the sampled scenarios with observations (Høyland & Wallace, 2001), and (3) clustering algorithms such as neural gas (Martinetz et al., 1993) which sample directly from observations rather than a distribution. Xu et al. [2015] presents a detailed review of different scenario tree generation methods. In the absence of reliable SF measurements, these techniques are ineffective. Therefore, in data-scarce catchments, the combination of ensemble precipitation forecasts and hydrologic models has the potential to generate reliable inflow scenarios (Block et al., 2009). Several previous studies have tested the validity of combining hydrologic

models and ensemble precipitation forecasts for generating inflow forecasts for reservoir optimization at shorter time scales (Lee et al., 2008; Saavedra Valeriano et al., 2010; Wang et al., 2012). At seasonal time scales, which is the focus of our study, scenario trees are valuable as SF forecasts suffer from large uncertainty, due mainly to lack of skill of long-lead precipitation forecasts (Shukla & Lettenmaier, 2011). As a consequence, studies which quantify the value of long-lead forecasts on reservoir operations have focused on methods such as ensemble SF prediction (Alemu et al., 2011; Anghileri et al., 2016; Day, 1985; Wood & Lettenmaier, 2006) and regression-based methods (Pagano et al., 2004). All these methods require either long-term observations of either precipitation or hydrologic variables such as SF or snow (regression-based methods). However, the development of ensemble climate prediction systems has resulted in improved forecasts of precipitation at seasonal time scales (Lavers et al., 2009). Although the use of such climate forecasts for operational hydropower planning is not widespread (Block, 2011), several studies have tried to combine ensemble seasonal precipitation forecasts with hydrologic models to generate SF forecasts (Block et al., 2009; Yuan et al., 2011). We evaluate the potential of global seasonal precipitation forecasts and hydrologic models to generate inflow scenarios in data-scarce regions. However, this does not completely obviate the need for SF observations. SF measurements are still required for calibrating hydrologic models. In the subsequent subsection, we detail potential solutions for calibrating hydrologic models without SF measurements.

### 1.2. Consideration of Different Sources of Uncertainty

Multimodel ensembles (MME) are used widely for characterizing the uncertainty in precipitation forecasts, arising primarily from model physics and initial atmospheric conditions (Krishnamurti et al., 2015; Shrestha et al., 2015). For seasonal climate forecasts, a number of MME developed in the past decade such as DEMETER (Palmer et al., 2004), ENSEMBLES (Weisheimer et al., 2009), the North American Multimodel Ensemble (NMME; Kirtman et al., 2014), and the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFSv1 and CFSv2; Saha et al., 2006, Saha et al., 2014) have shown promise (Siegmund et al., 2015; Yuan et al., 2011). These forecasts can help account for precipitation forecast uncertainty in reservoir inflow forecasts and hence hydropower planning in data-scarce regions as they do not require ground-based measurements for their generation. In this study, we account for precipitation forecast uncertainty by using ensemble seasonal precipitation forecasts from NMME, which has been evaluated in different hydroclimatic regions (Becker et al., 2014; Becker & van den Dool, 2016; Cash et al., 2017).

Satellite-based remote sensing enables spatially and temporally continuous monitoring of hydrologic fluxes such as evapotranspiration (ET), soil moisture, total water storage, and snow water equivalent (SWE; Lettenmaier et al., 2015). In the absence of reliable SF records, the potential of such remote sensing-based data sets in characterizing model parameter and catchment initial condition uncertainty needs to be explored. Spatially distributed calibration of hydrologic models solely with evapotranspiration data has proven to be effective for simulating the water balance of irrigation-dominated catchments (Becker et al., 2019), for improving the representation of vegetation dynamics (Rajib et al., 2018), and for informing SF simulations (Immerzeel & Droogers, 2008; Koppa et al., 2019; López López et al., 2017; Wanders et al., 2014). Based on the findings of these studies, we use evapotranspiration as a proxy for SF in calibrating the hydrologic model and quantify model parameter uncertainty, using a Bayesian calibration methodology (Koppa et al., 2019).

Like model parameter uncertainty, remotely sensed soil moisture and SWE data sets can be used to characterize the uncertainty in the estimates of initial hydrologic conditions. Techniques such as the ensemble Kalman filter have found widespread use for soil moisture (Yatheendradas et al., 2012) and SWE (Alvarez-Garreton et al., 2015) data assimilation. In this study, we do not explicitly account for the uncertainty in initial hydrologic conditions. However, we assume that a well-calibrated model leads to good representation of the initial hydrologic conditions.

### 1.3. Formulation of the Hydropower Optimization Model

In many studies concerning stochastic programming with recourse models, the first stage is assumed to be deterministic. For example, Xu et al. (2014) consider the inflow forecast at the first stage to be sufficiently accurate in developing a Bayesian hydropower optimization model with a horizon of 5 and 10 days. In another study, precipitation forecasts from a deterministic Numerical Weather Prediction (NWP) model are used to generate SF forecasts for the first stage and second stage 30-member MME (Wang et al., 2012).

Even when the planning horizon is seasonal to interannual, the first stage (month) inflow is assumed to be deterministic, and the optimal release is updated continuously when more observations of the current month's inflow become available (Kim & Palmer, 1997). For example, Etkin et al. (2015) develop and test a stochastic decision support tool for seasonal multipurpose reservoir operation in which the immediate stage is deterministic. As reliable SF observations are not available in data-scarce regions to validate the reservoir inflow forecasts, the assumption of a deterministic first stage may not be appropriate. For example, Séguin et al. (2017) highlight the importance of stochastic scenarios on hydropower by comparing scenario trees of varying complexities: (1) full scenario tree with SPWR-D, (2) scenario tree with only the median scenario at all stages, and (3) scenario fan. They conclude that, for hydropower planning, stochastic scenarios (scenario tree or fans) are preferable over deterministic scenarios. In this study, we extend the comparisons presented in Séguin et al. (2017) by reformulating the stochastic programming with recourse model for hydropower planning to consider the uncertainty in seasonal reservoir inflow forecasts at the first/immediate stage. Specifically, we evaluate the impact of considering deterministic and stochastic first stages in reservoir inflows on the optimized released decisions and hence hydropower production.

#### 1.4. Organization of the Study

First, we present the methodological framework developed in this study for seasonal hydropower planning in data-scarce regions. Next, we detail the study area selected to test the developed framework in a data-scarce catchment (Appendix A presents the experiment design of the case study). Finally, we present and discuss the results of the application of the hydropower planning framework to a data-scarce catchment, followed by a summary of the main conclusions of the study and potential directions in which the study can be extended in the future.

## 2. Methodology

In contrast to traditional hydropower planning, we combine NWP-based MME precipitation forecasts, remote sensing-based estimates of precipitation and evapotranspiration, and spatially distributed hydrologic models to generate reservoir inflow scenarios in data-scarce regions. Specifically, we evaluate three different reservoir inflow scenario structures (Figure 1) and their impacts on release decisions and hydropower production—(1) a single deterministic forecast (DET), (2) a scenario fan, but with the first stage deterministic (SPWR-D), and (3) a scenario fan with all stages stochastic (SPWR-S). For this, we develop a methodological framework for seasonal hydropower planning in data-scarce regions consisting of three major components: (1) precipitation forecasting, (2) hydrologic forecasting, and (3) hydropower optimization. In Figure 2, we present a flowchart of the entire framework. In the subsequent subsections, we describe each of the three major components in detail.

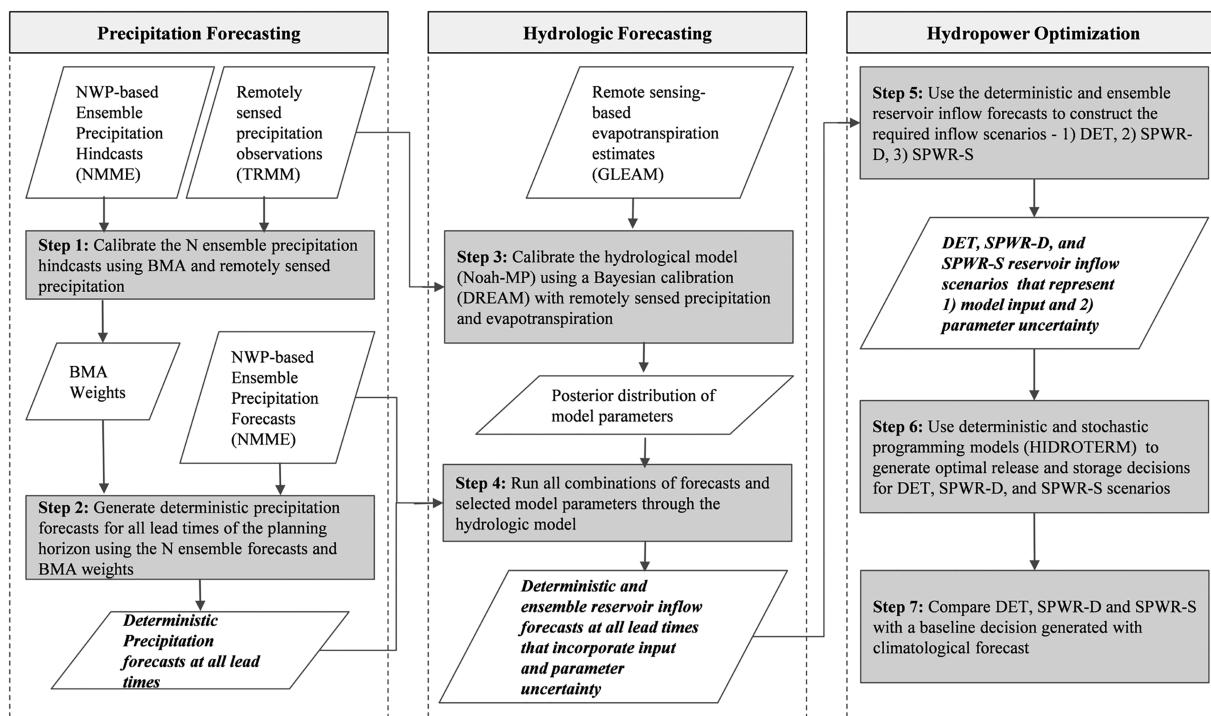
### 2.1. Precipitation Forecasting

The main objective of this component of the hydropower planning framework is to produce the required precipitation forecasts for the hydrologic forecasting component. For the SPWR-S scenario in which forecasts at all stages are stochastic, we use the raw MME forecasts. However, from Figure 1 it is evident that DET and SPWR-D scenarios require DET at the first stage. To generate these DET, we use Bayesian Model Averaging (BMA) to calibrate MME precipitation forecasts (Steps 1 and 2 in Figure 2).

BMA is a statistical approach to post-processing forecast ensembles generated from multiple statistical (Hoeting et al., 1999) or dynamical models (Raftery et al., 2005). BMA of a forecast ensemble results in a calibrated and sharp predictive probability density function (PDFs), represented as a weighted averaged of the PDFs of the ensemble members. Following Sloughter et al. (2007), the BMA predictive PDF can be mathematically represented as:

$$P(x|f_1, f_2, \dots, f_n) = \sum_{n=0}^{Ns} \omega_n c_n(x|f_n), \quad (1)$$

where  $\omega_n$  is the posterior probability of ensemble forecast member  $f_n$  being the best one, determined in the calibration or training period using observed or reference data of the hydrologic variable under consideration  $x$  (e.g., precipitation or evapotranspiration); and  $c_n(x|f_n)$  is the conditional PDF associated with the ensemble forecast  $f_n$  of the hydrologic quantity  $x$ . For variables such as temperature, the conditional PDF  $c_n(x|f_n)$  can



**Figure 2.** Flow chart for seasonal hydropower planning framework developed in this study. The precipitation forecasts, observational data sets, hydrologic model, and optimization algorithm used in the case study are mentioned in parenthesis. The precipitation forecasting, hydrologic forecasting, and hydropower optimization steps are repeated at each stage of the hydropower planning horizon

be assumed to be normally distributed (Raftery et al., 2005), but a gamma distribution is more appropriate for precipitation (Sloughter et al., 2007), evapotranspiration (Khanmohammadi et al., 2018), and SF (Vogel & Wilson, 1996).

In our study, we use a mixture of point mass at zero and a gamma distribution as the conditional PDF (Sloughter et al., 2007). We determine the BMA weights ( $\omega_n$ ) of ensemble precipitation members using satellite-based estimates of precipitation. To calibrate the ensemble precipitation forecasts (determine the BMA weights), we define a calibration period that precedes the hydropower planning horizon. The length of the calibration period is equal to the length of the planning horizon. In other words, if the planning horizon starts from month  $t$  and ends at month  $T$ , the BMA calibration period starts from month  $t - T$  and ends at month  $t - 1$ . We adopt a rolling scheme, wherein we recalculate the BMA weights at every stage of the stochastic programming with recourse model by moving the calibration period forward by 1 month. The DET for the DET and SPWR-D scenario structures then are generated as the weighted average of the ensemble precipitation forecast members under consideration. The deterministic and raw ensemble precipitation forecasts then are used as inputs into the hydrologic model to generate the deterministic and ensemble reservoir inflow forecasts (described below).

We present the details of the satellite-based precipitation data set and ensemble precipitation forecast data used in the case study in Appendix A.1.

## 2.2. Hydrologic Forecasting

The main objective of the hydrologic forecasting component of the framework is to generate reliable reservoir inflow forecasts. For this, we use the deterministic and ensemble precipitation forecasts produced in the previous step as input into a calibrated spatially distributed hydrologic model (Steps 3 and 4 in Figure 2). In the absence of SF measurements for the calibration of the hydrologic model, we use evapotranspiration as a proxy for streamflow (SF). The selection of ET as a proxy for SF is motivated by the findings of several previous studies concluding that incorporating ET, in the absence of SF observations, into calibration improves the accuracy of SF compared with an uncalibrated model (Immerzeel & Droogers, 2008; Koppa

et al., 2019; López López et al., 2017; Zink et al., 2018). In addition, to quantify the model parameter uncertainty, its impact on the inflow forecasts, and hence the hydropower, we adopt a formal Bayesian calibration approach to derive the posterior probability distribution of model parameters. Specifically, we utilize the Differential Evolution Adaptive Metropolis (DREAM) Markov Chain Monte Carlo Scheme (Vrugt et al., 2008; Vrugt et al., 2009).

The DREAM algorithm has been used extensively to quantify parameter uncertainty in hydrologic (Shafii et al., 2014) and hydrogeologic models (Laloy et al., 2013). DREAM is a multichain Markov Chain Monte Carlo simulation algorithm that automatically tunes the scale and orientation of the proposal distribution en route to the target distribution. It is designed for increasing the sampling efficiency of complex, high-dimensional parameter spaces, while maintaining detailed balance and ergodicity (Vrugt, 2016). In this study, we use the MT-DREAM (ZS) version of DREAM, which utilizes multitry sampling (MT), snooker updating, and sampling from an archive of past states to improve the rate of convergence and make use of parallel computing resources. Being a Bayesian algorithm, DREAM requires the selection of (1) a prior distribution of the selected parameters and (2) a likelihood function. The prior distribution represents the modeler's knowledge of the selected parameters. In this study, we assume a uniform prior for all the parameters selected for calibration. A uniform distribution is a noninformative prior, implying that we do not have any prior knowledge of the possible values of the selected parameters. A likelihood function is used to summarize the distance between the model simulations and the corresponding observations. In this study, we select the Laplacian likelihood based on the findings of Schoups and Vrugt (2010); residual errors in rainfall-runoff models are better represented by a Laplacian distribution than a Gaussian distribution. The uniform priors and the likelihood function are then used by the DREAM algorithm to derive posterior distributions of model parameters, which are used to quantify model parameter uncertainty.

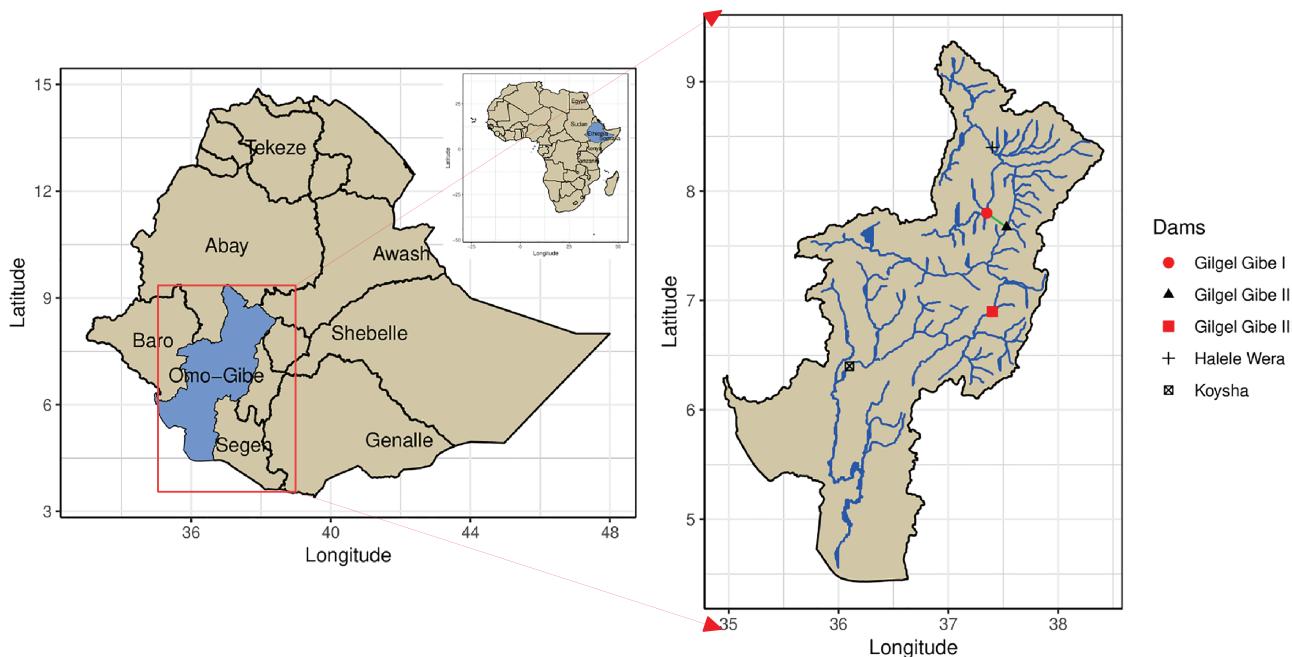
We describe the specific steps involved in the hydrologic forecasting component. First, we calibrate the hydrologic model using the DREAM algorithm to derive the posterior distribution of model errors. Next, we define a limit of acceptability or error threshold to distinguish between behavioral and nonbehavioral solutions (Beven, 2006). “Behavioral” solutions are model parameter sets, derived from calibration, that result in errors that are within an acceptable limit (error threshold) for a specific model output (ET in this study). We use the model parameter set with the least model error for generating inflow scenarios. The remaining behavioral model parameter sets then are used to quantify the uncertainty in ET and SF forecasts due to uncertainty in model parameters. We describe the setup, calibration, and validation of the specific hydrologic model used in our study, as well as the configuration of the DREAM calibration algorithm in the next (experiment design) section.

We present details of the satellite-based ET data, hydrologic model setup, calibration, and uncertainty quantification used in the case study in Appendices A.1. and A.2..

### 2.3. Hydropower Optimization

We develop a multistage deterministic and stochastic programming with recourse model to optimize hydropower production of a complex reservoir system at seasonal time scales. We consider a nonlinear power production function (a function of reservoir storage and release) at each individual power plant. The optimization model is developed for planning purposes; thus, monthly time periods are used (Yeh, 1985). First, we provide a generic example of a multistage stochastic programming with recourse model. Then, we describe the specific hydropower optimization scheme implemented in this study.

Consider a system of  $R$  reservoirs. Let the seasonal hydropower planning horizon be  $T$  months, which is divided into  $T$  stages (each stage is 1 month). Let the monthly time period be  $t, t = 1, 2, \dots, T$ . In a multistage stochastic programming with recourse model, at the beginning of each time period  $t$ , a reservoir inflow scenario tree or fan with  $n$  scenarios is constructed to represent possible future inflows into the reservoir. These scenario trees or fans are used as an input to an optimization model that minimizes or maximizes the expected value of a specified objective function. We adopt a “rolling horizon control” scheme in which the optimal release decisions are adopted only for the immediate stage ( $t$ ). Then, the planning horizon ( $T$ ) is “rolled” forward ( $t = t+1$ ), the reservoir inflows are reforecasted, and the model is reoptimized (recourse). The rolling horizon control is a general purpose control scheme that involves repeatedly solving a constrained optimization problem as well as using predictions of future inflows and constraints over a moving time horizon to choose the control action. This process is repeated until the end of the planning horizon



**Figure 3.** The Omo-Gibe River basin consisting of a cascade of five reservoirs, located in Ethiopia. The hydropower reservoirs used in this study (Gilgel Gibe I and Gilgel Gibe III) are highlighted in red. The green line represents a tunnel which connects Gibe I and Gibe II reservoirs.

( $t = T$ ). Specifically, the optimization procedure developed as part of the framework consists of three main steps (represented in Figure 2)—(1) construction of deterministic and stochastic reservoir inflow scenarios; (2) optimization of reservoir operation by a deterministic and multistage stochastic programming with recourse model, with a rolling horizon; and (3) comparison of the optimized results with a baseline policy generated by climatological forecasts.

First, the deterministic and ensemble reservoir inflow forecasts generated in the hydrologic forecasting component are used to construct the three scenario structures compared in this study—(1) DET, (2) SPWR-D, and (3) SPWR-S (Figure 1). For the construction of the completely deterministic or DET scenario, only the DETs generated from BMA of ensemble precipitation forecasts are used at all stages of the planning horizon ( $T$ ). As all stages are deterministic, the DET scenario fan consists of only one inflow scenario. For the SPWR-D scenario, DET (from BMA) is used for just the first stage of the planning horizon, and  $n$  ensemble forecasts (generated from raw ensemble precipitation forecasts) are used for the remaining  $T - 1$  stages (Figure 1). Therefore, SPWR-D scenario fan consists of  $n$  inflow scenarios. Finally, for the SPWR-S scenario, only ensemble forecasts are used at all stages resulting in a completely stochastic scenario fan, consisting of  $n$  scenarios. Note that the DET, SPWR-D, and SPWR-S scenarios are constructed (Step 5 in Figure 2) at the beginning of each iteration of the multistage stochastic programming with recourse model, i.e., at each stage  $t$  of the planning horizon  $T$ .

Next, the different scenarios are used as an input to a nonlinear programming (NLP) hydropower optimization model to generate optimal release and storage decisions (Step 6 in Figure 2). We choose NLP because it can accommodate the nonlinear power production functions without approximation. It has been shown that NLP models are most accurate in generating reliable release and storage policy for reservoir management and operation (Barros et al., 2003). An NLP model requires the formulation of an objective function and a set of constraints that accurately represent the hydropower system being optimized. We present the detailed mathematical formulation of the NLP model used in this study in Appendix A.3.

Finally, the hydropower produced from DET, SPWR-D, and SPWR-S scenarios is compared with a baseline policy (Step 7 in Figure 3). Generally, a baseline policy is the existing release policy adopted in the specific reservoir system. The hydropower produced from existing policy is compared with the hydropower

**Table 1**  
*Details of the Five Hydropower Reservoirs in the Omo-Gibe River Basin*

Reservoir name	Start of operation	Hydropower capacity	Maximum storage
Gilgel Gibe I	2004	210 MW	840 Mm <sup>3</sup>
Gilgel Gibe II	2010	420 MW	0.15 Mm <sup>3</sup> (Run-of-river)
Gilgel Gibe III	2015	1870 MW	13,700 Mm <sup>3</sup>
Koysa Dam	Planned	2160 MW	-

produced using the hydropower planning framework developed in this study. In the absence of existing release policy or hydropower generation information, a baseline generated using climatological forecast can be used (as is done in the case study presented here). This comparison can be done either at every stage of the planning horizon or at the end of the planning horizon.

### 3. Study Area and Time Period for Case Study

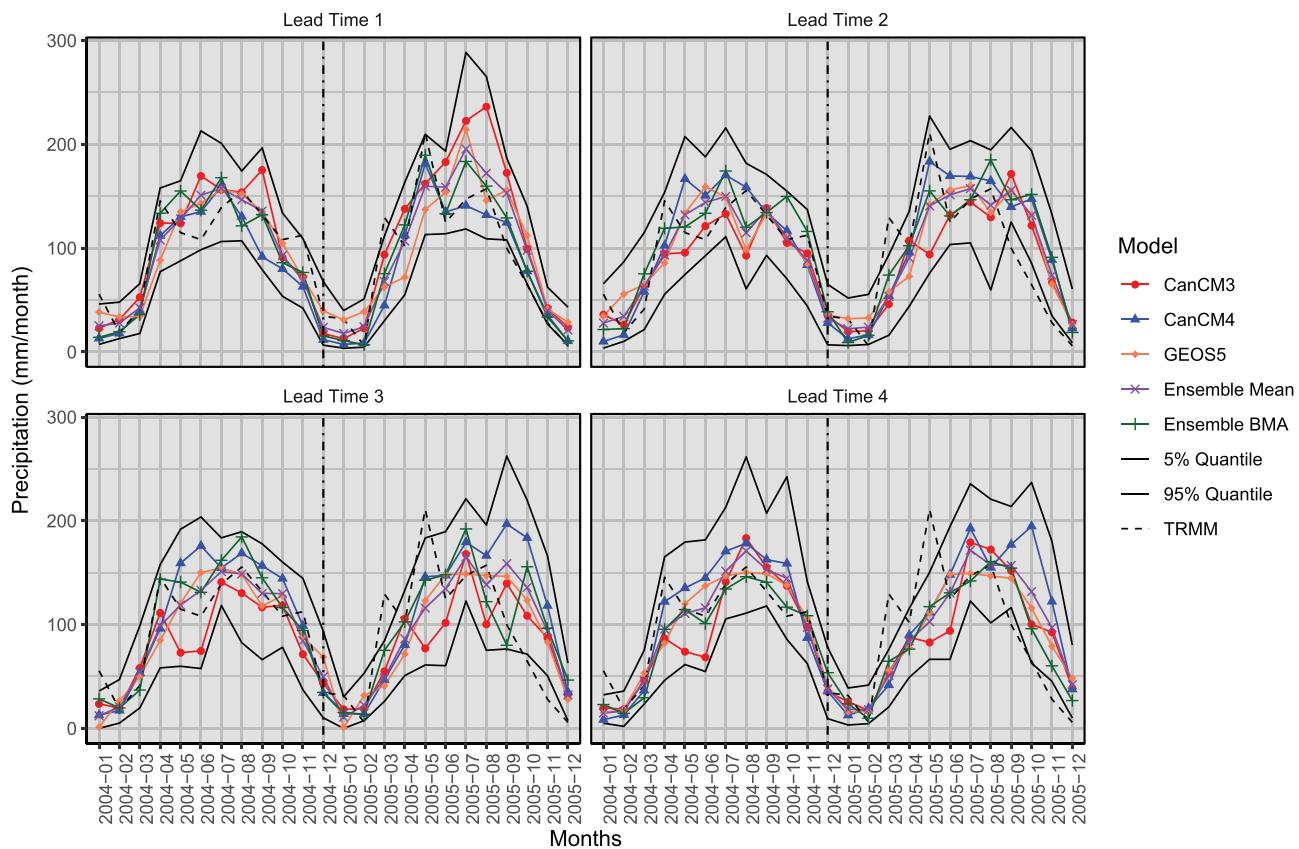
We test the seasonal hydropower planning framework developed in this study in the Omo-Gibe River basin in East Africa (Figure 3). Spanning an area of 79,000 km<sup>2</sup>, the river basin is spread across Ethiopia and Kenya, and the outlet of the basin is at Lake Turkana. The focus of our study is on the Ethiopian part of the Omo-Gibe River basin. The elevation ranges from about 700 m to 3,100 m. The average rainfall in the basin is about 1,150 mm with a humid north and arid south. The annual temperature varies between 17 °C and 29 °C (Chaemiso et al., 2016). In terms of hydropower capacity, the study area consists of three power plants in operation (Gilgel Gibe I, II, and III) and one planned hydropower plant (Koysa). Table 1 presents the details of the hydropower plants. To test the proposed framework, we use a cascade of two reservoirs, consisting of Gibe I and Gibe III. Inflows to the reservoirs are forecasted 8 months in advance, and the planning horizon is assumed to be 8 months. The system is optimized with an 8-month rolling horizon. To test the framework, we select February–September 2005 as the planning horizon. The primary reasons for selecting February–September 2005 are (1) the availability of precipitation forecasts (lead times of 9 months) and the availability of historical hydropower production data (8 months from February–September 2005). We calibrate the hydrologic model for the year 2004 using satellite-based estimates of ET. We present the detailed description of the experiment design for the case study in Appendix A.

### 4. Results and Discussion

First, we validate the seasonal precipitation hindcasts from NMME using the satellite-based TRMM data set. Next, we validate the calibrated Noah-MP hydrologic model with remote sensing-based ET estimates from GLEAM. We then present the seasonal hydropower planning results derived from HIDROTERM for the study region (Appendix A presents the details of the experiment design of the case study). Finally, we quantify the uncertainty in optimal reservoir power release decisions and hydropower arising due to uncertainty in model parameters.

#### 4.1. Validation of the NMME Precipitation Forecasts

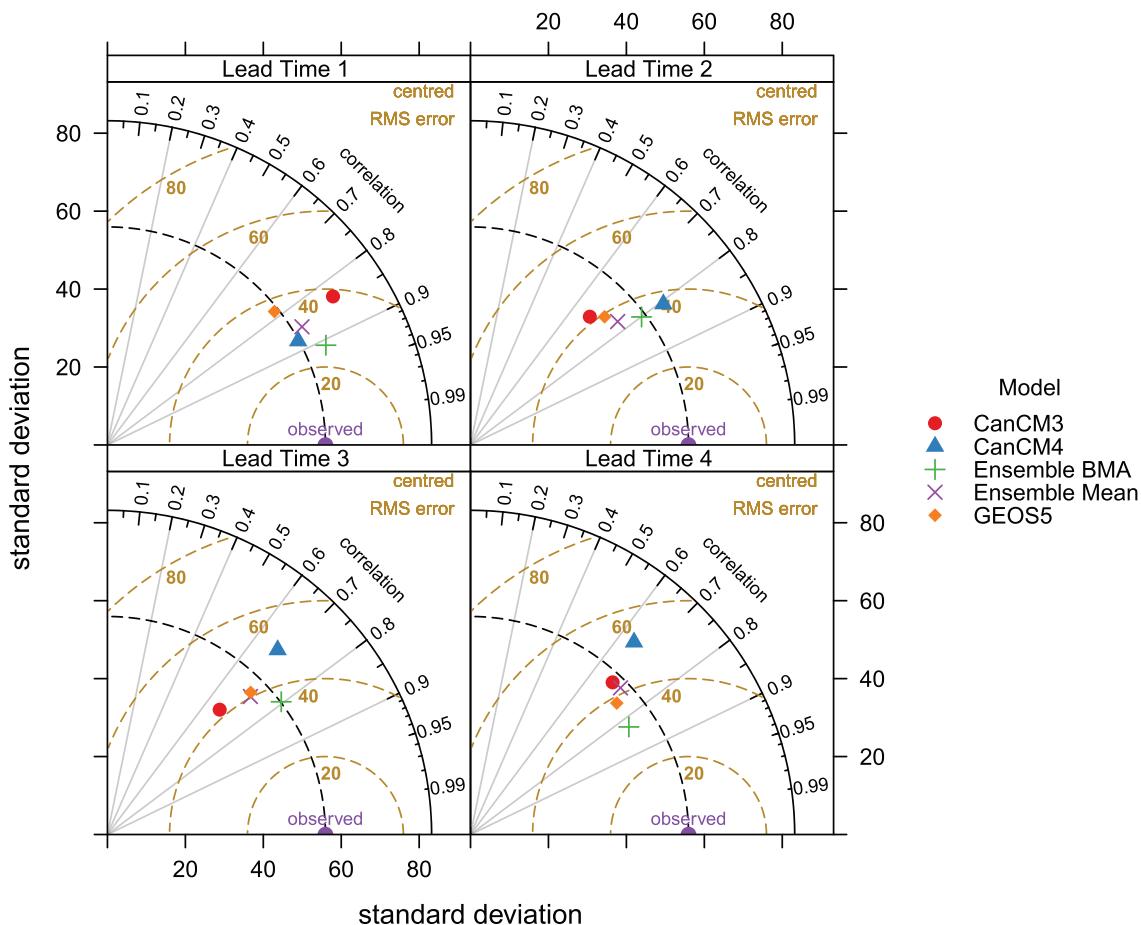
To validate the 30-member seasonal ensemble precipitation forecasts from NMME, we use the TRMM remotely sensed precipitation estimates. Figure 4 presents a time series comparison of different precipitation forecast models with the observed data. We also present Taylor diagrams (Taylor, 2001) to represent the root mean square error (RMSE), correlation with observations, and standard deviation of different forecast models (Figure 5). A Taylor diagram represents the standard deviation of the observed data and modeled results, radially, along the x and y axes (e.g., the standard deviation of the observed precipitation is approximately 55 mm). The correlation of the different models with the observed data is shown along the circumference of the Taylor diagram, increasing radially from the y-axis to the x-axis (e.g., the observed data is shown on the x-axis, representing a correlation of 1.0, and CanCM3 has a correlation of 0.82 at lead time of 1 month). Finally, the RMSE of models is represented as contours (dashed yellow lines in Figure 5), with RMSE values increasing from the x-axis to y-axis (e.g., the observed data is shown on the x-axis, representing an RMSE of 0.0, and CanCM3 has an RMSE of approximately 40 mm/month at lead time of 1 month). The closer the point representing a model (e.g., CanCM3) is to the observed data (purple point in Figure 5), the



**Figure 4.** Time series comparison of precipitation from different NMME models and observations (TRMM) for different lead times (1–4 months) and for the calibration (12 months of 2004) and validation (12 months of 2005) time periods. We present the mean of the 10 members from CanCM3, CanCM4, and GEOS-5. “Ensemble Mean” and “Ensemble BMA” are the simple mean and Bayesian Model Averaging of all 30 ensemble members. In addition, we present the 5% and 95% quantiles of all 30 ensemble members. The vertical line demarcates the calibration (2004) and validation (2005) time periods.

better the performance of the model. We see that the 30 ensemble members, represented by the 5% and 95% quantiles (gray lines in Figure 4), encompass the observed data (dashed black lines in Figure 4) up to lead times of 3 months (lead time 1–3). The months of March, May, October, and November of the year 2005 (validation time period) are the exceptions. Also, all the raw ensemble members underestimate the precipitation for the month of May 2005 and overestimate precipitation for the months of October and November of the same year. In addition, we see that the difference between the 5% and 95% quantiles increases with increase in lead time. For example, the mean difference between the 5% and 95% quantiles at lead time of 1 month is 78.5 mm and at lead time of 4 month is 95.1 mm. For comparison, the mean difference at 2-month and 3-month lead times are 91.1 mm and 93.4 mm, respectively. This is expected, as the uncertainty in the initial and boundary conditions driving the NWP models is higher at longer lead times. At lead times longer than 4 months (lead times 4–8), the raw ensemble members considerably overestimate the observed precipitation in the second half of the year 2005.

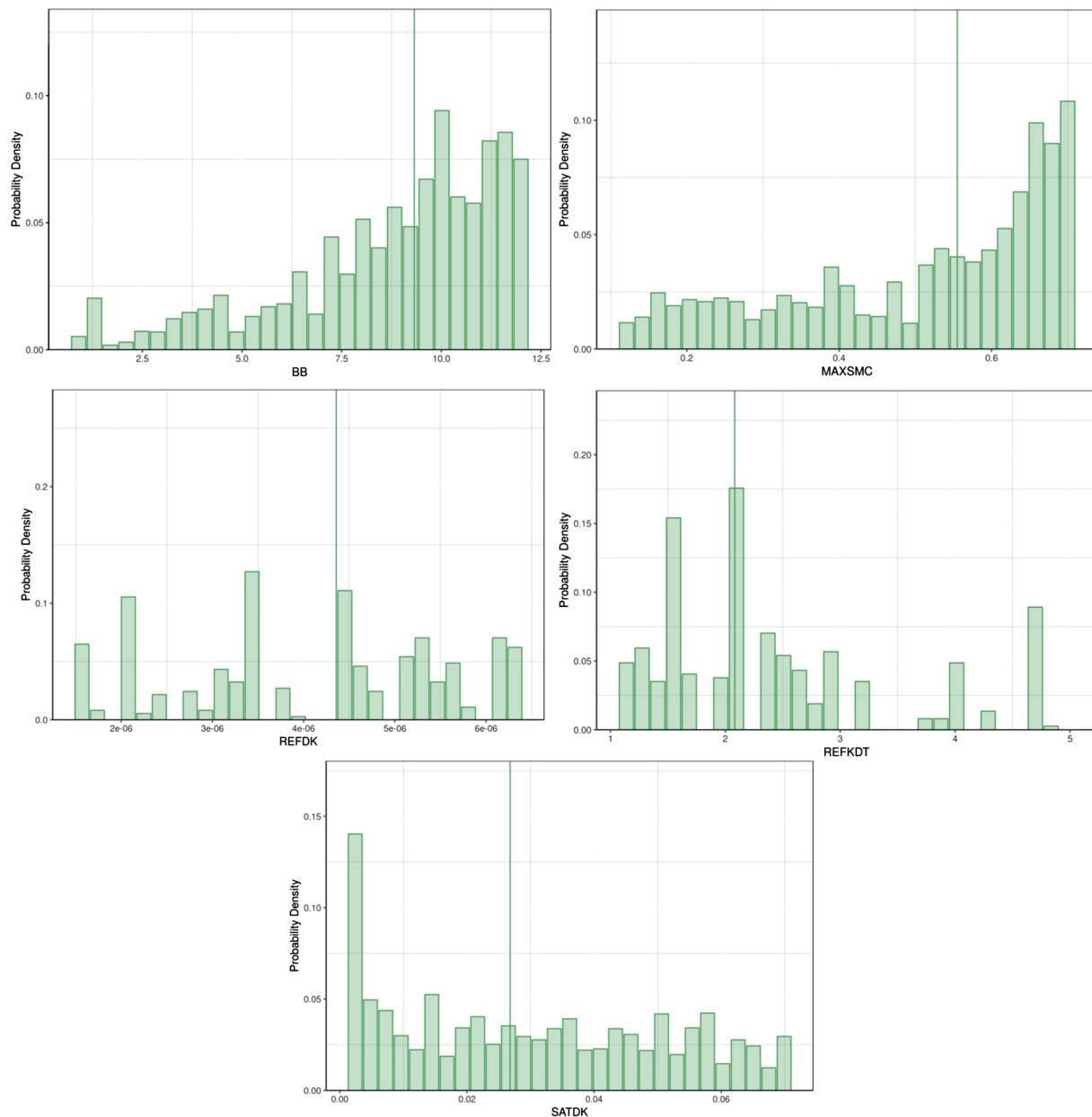
To understand the differences in performance among the three NMME models under consideration (CanCM3, CanCM4, and GEOS-5), we compare the average of the ten constituent members. The stochastic programming with recourse model implemented in this study uses forecasts that are updated at the start of each month, while the release decisions are implemented only for the immediate stage. Therefore, we focus on the performance of the forecasts at shorter lead times. We present the results of longer lead times (5–8 months) in Appendix B. At lead times of 1–3 months, we see that the CanCM4 model consistently outperforms the other two models (Figures 4 and 5). Both the CanCM3 and GEOS-5 models overestimate the precipitation in the second half of 2005 (validation time period). This is especially true for the lead time 1 month, where the CanCM4 model exhibits lower RMSE



**Figure 5.** Taylor diagrams of precipitation from the CanCM3, CanCM4, GEOS-5, Ensemble Mean, and BMA models determined for different lead times (1–4 months) and for the calibration (12 months of 2004) and validation (12 months of 2005) time periods. The CanCM3, CanCM4, and GEOS-5 models are calculated as the mean of the 10 ensemble members of the three models. “Ensemble Mean” and “Ensemble BMA” are the simple mean and Bayesian Model Averaging of all 30 ensemble members.

and higher correlation with the observed data (Figure 5). Additionally, at lead time 1 month, the CanCM4 model is able to capture the two peaks in the observed precipitation of the year 2005. Between CanCM3 and GEOS-5, GEOS-5 significantly underestimates the peaks in 2005. At longer lead times, CanCM4 model performs relatively poorly compared with the other models. As expected, the forecast accuracy of all the models deteriorates with increasing lead times.

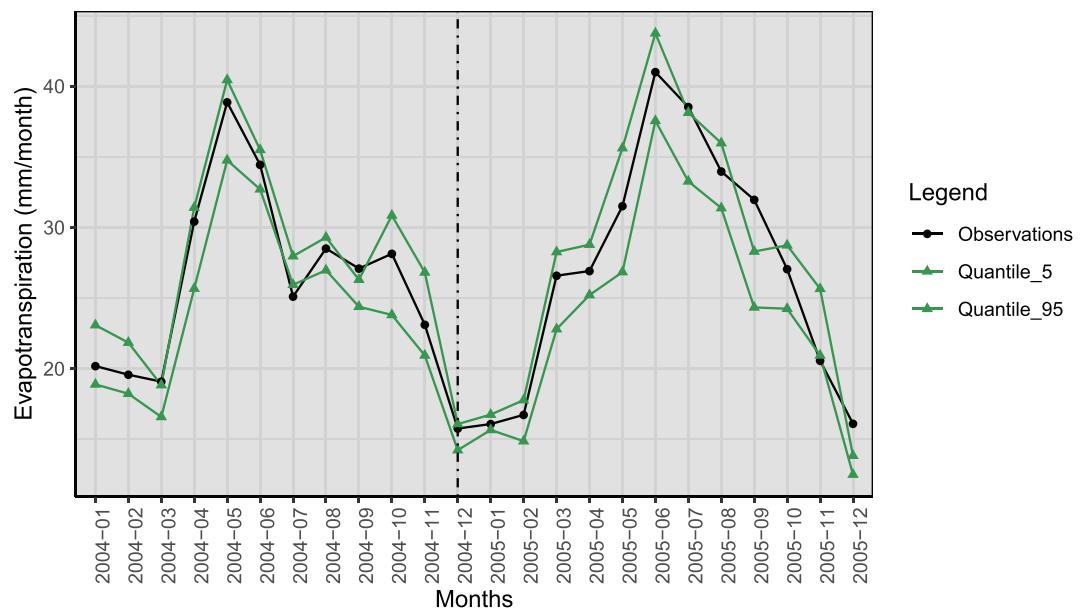
Finally, we compare the simple mean and Bayesian Model Averaging of all 30 ensemble members with the individual models. We see that the ensemble BMA outperforms all the other models at almost every lead time (Figures 4 and 5). At shorter lead times (1–3 months), the ensemble BMA is able to capture the quantity and the timing of the peaks and troughs seen in the observed precipitation (Figure 4). The low RMSE (approximately 25 mm/month) and high correlation coefficient (greater than 0.9) at lead time 1 month supports this conclusion. At shorter lead times, the CanCM4 model outperforms the ensemble mean in terms of both the magnitude of error and correlation with the observed data. This highlights the advantage of the BMA technique, which provides higher weights to better performing models (CanCM4 in this case). The ensemble mean, which weighs all the ensemble members equally, is biased by the lower performing CanCM3 and GEOS-5 models. The improvements are particularly greater in the second half of the year 2005, in which the overestimation of precipitation by CanCM3 and GEOS-5 is reduced by ensemble BMA and the ensemble mean. In addition, the variance of the ensemble BMA also seems to match the variance of the observed precipitation better than the individual members, except at lead time 1 month, where the CanCM4 standard deviation matches the observed value better.



**Figure 6.** Posterior probability density functions of the Noah-MP hydrologic model parameters considered for calibration using the MT-DREAM (ZS) algorithm and ET estimates from GLEAM. The green line represents the 50% quantile values for each of the parameters.

#### 4.2. Validation of the Noah-MP Hydrologic Model

As stated in the methodology section, we use evapotranspiration as a proxy for SF in calibrating the Noah-MP hydrologic model for the Omo-Gibe River basin. Accordingly, we validate the hydrologic model with remote sensing-based ET estimates from GLEAM. First, we present the posterior distribution of Noah-MP model parameters derived from the MT-DREAM (ZS) algorithm using a uniform prior and Laplacian likelihood (Figure 6). The primary reason for presenting the posterior distribution of model parameters is to analyze the impact of using ET as the target variable for calibration on the parameters selected for calibration. Compared to the prior distributions of calibrated parameters, which are assumed to be uniform, the posterior distributions may either be uniform or nonuniform. A uniform posterior distribution of a model parameter implies that ET data cannot help in calibrating that model parameter. In other words, evapotranspiration is



**Figure 7.** Time series comparison of evapotranspiration from ET-calibrated Noah-MP model (green) and observed ET estimates (black) from GLEAM for the calibration (12 months of 2004) and validation (12 months of 2005) time periods. The 5% and 95% quantiles from the behavioral solutions of Bayesian calibration are used to determine the uncertainty in modeled ET (green band). The vertical line demarcates the calibration (2004) and validation (2005) time periods.

not sensitive to that model parameter. On the other hand, a nonuniform posterior implies that ET data can inform the calibration of that model parameter. Figure 6 presents the posterior distributions of the five model parameters considered for calibration—BB, MAXSMC, REFDK, REFKDT, and SATDK. First, we analyze the shape (uniform or nonuniform) of the posterior distribution of each parameter to check if ET can inform the calibration of that model parameter. Next, we try to physically interpret the most probable values of each parameter.

In Figure 6, we see that the distributions of all the parameters, except REFDK, are clearly nonuniform. This implies that ET data can help in informing the calibration of all the parameters, except REFDK. The distributions of BB and MAXSMC, which are soil-related parameters, are skewed toward the higher values. Physically, we can identify the high values of BB (greater than 10.0) and MAXSMC (greater than 0.4) parameters combined with lower SATDK seen in Figure 6 with sandy clay, silty clay, and clay soil textures. Compared to the soil-related parameters, the distributions of the runoff-related parameters are more uniform. While the REFKDT parameter is more skewed toward the lower values, the REFDK parameter is more uniform. REFKDT is a parameter that controls the partitioning of total runoff into surface and subsurface components. Physically, a lower value of REFKDT implies that a larger portion of the total runoff is being partitioned into surface runoff. The more uniform distribution of REFDK parameter could imply that ET is not very sensitive to that parameter. The optimized parameter set from Bayesian calibration is used to construct the posterior distribution of simulated ET errors. Then, we define an error threshold of 50% quantile to select a set of behavioral (acceptable) parameter sets. The error threshold was subjectively selected by selecting a reasonable error value (approximately 10 mm/month) after comparing the values of different quantiles of the posterior distribution of ET errors. Twenty parameter sets from the behavioral solutions are used to quantify the uncertainty in inflow forecasts, and hence the hydropower, due to uncertainty in model parameters.

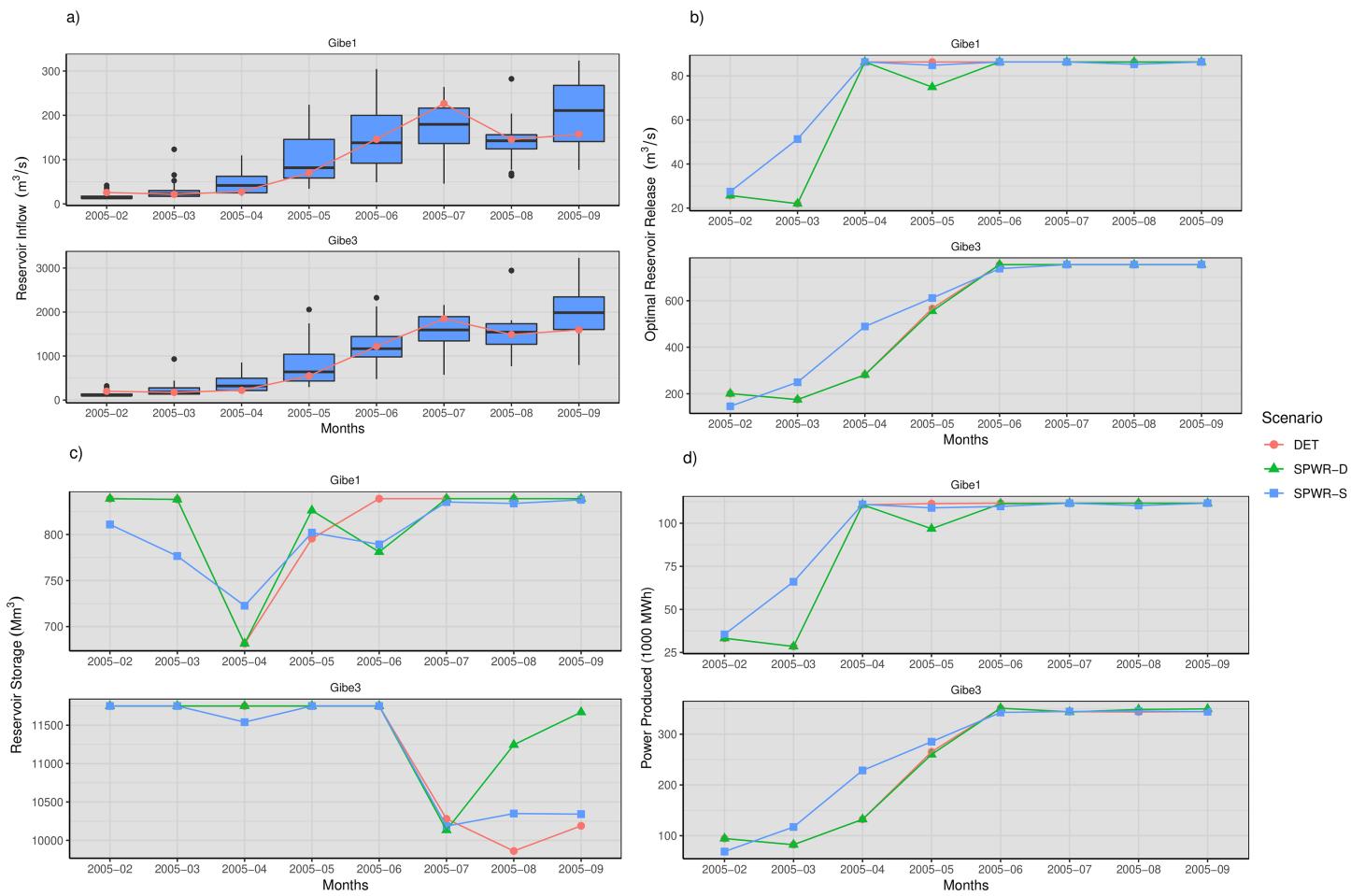
We validate the set of selected behavioral solutions with the GLEAM ET reference data set for the year 2005 using time series analysis (Figure 7) and relevant error metrics. A visual examination of Figure 7 reveals that the ET-calibrated Noah-MP model performs well in simulating the evapotranspiration across the Omo-Gibe River basin. In the calibration period (12 months of 2004), we see that the reference ET data (black line) is closer to the 95% quantile of the modeled ET for the months of April, May, and

June. This implies that the behavioral solutions consistently underestimate ET for the months with higher observed evapotranspiration. In the validation period (12 months of 2005), the reference data set corresponds to the median modeled ET values for the first 6 months. In this last 6 months of 2005, we see that the model consistently underestimates (except November 2005) the evapotranspiration across the study area. This consistent underestimation of ET, coupled with the fact that the NMME forecasts overestimate precipitation in the second half of 2005, can lead to overestimation of reservoir inflows. It is interesting to note that the model results show higher uncertainty for the month of October in both the calibration and validation time periods compared to other months. Additionally, the model is able to capture the quantity and timing of the peaks and troughs seen in the reference ET data set. A low mean RMSE of 3 mm/month and a high mean correlation coefficient of 0.97, determined for the validation time period, support this conclusion. We note that the primary assumption of our study is that ET can be used as a proxy for SF in data-scarce regions. Consequently, we assume that the high fidelity of the Noah-MP model in simulating ET across the Omo-Gibe River basin leads to accurate simulation of SF (based on the findings of Koppa et al., 2019) and hence the reservoir inflows required for seasonal hydropower planning.

#### 4.3. Seasonal Hydropower Planning in the Study Region

We select the best performing Noah-MP model parameter set from the 20 behavioral solutions to generate the reservoir inflow forecasts required to optimize the cascade of 2 reservoirs (Gibe I and Gibe III) in the study region. We construct the three scenario structures (DET, SPWR-D, and SPWR-S) for the months February–September 2005 by combining the NMME precipitation forecasts and the Noah-MP hydrologic model as detailed in the methodology section. The three inflow scenario structures then are used as inputs to a deterministic and stochastic programming with recourse model to generate the optimal power releases, reservoir storage variations, and the associated hydropower production (Figure 8). In a stochastic programming with recourse model, only the immediate stage release decisions are implemented. At the end of the immediate stage, scenario tree or fan structures are regenerated, and the system is reoptimized with a rolling horizon. Figure 8 shows the results obtained from the optimization model for the immediate stage. First, we compare the differences in reservoir inflows among the three scenarios. We then analyze the impact of these differences in the scenario structure on the optimized release decisions, storage variations, and hydropower production. Figure 8a presents the deterministic (from BMA) and stochastic (raw forecast ensembles) inflows used to construct the three scenario trees. We see that the first 3 months of the study period are relatively dry compared to the rest of the months. Also, the deterministic inflows into both Gibe I and Gibe III are consistently lower than the mean of the 30-member ensemble, with June and July being the exceptions among months with higher flows. Additionally, we also see that there is considerable uncertainty in the reservoir inflow arising from uncertainty in the input precipitation ensembles. To compare the uncertainty of inflows across different months and reservoirs, we calculate coefficient of variation and coefficient of range values (Table 2). The high values of measures of absolute and relative dispersion reflect the high uncertainty in the inflow values. The inflows in the month of March 2005 exhibit the largest uncertainty, with a coefficient of variation of 0.74 and 0.69 for the Gibe I and Gibe III reservoirs, respectively. The month of July 2005 has the least uncertainty, with a coefficient of variation of 0.31 and 0.24 for the Gibe I and Gibe III inflows, respectively.

In Figure 8b, we present the results of the optimized release decisions generated for the three inflow scenario structures: (1) DET, (2) SPWR-D, and (3) SPWR-S. We see that the optimized release decisions generated with the DET and SPWR-D inflow scenario structures are very similar to each other, except for the month of May 2005. We note here that the results correspond to the immediate stage of the stochastic programming with recourse model. Therefore, the similarity between the DET and SPWR-D release decisions may be due to the fact that the immediate stage of the SPWR-D model is deterministic and that the value derived from BMA of ensemble members is equal to the DET inflow scenario structure. We see that the optimized release decisions are consistently higher in the SPWR-S case compared with either the DET or SPWR-D scenario structures. This can be attributed to the fact that the inflows in the DET scenarios are consistently lower than the mean of the ensemble members (Figure 8a). It is interesting to note that the uncertainty in the inflows represented in the SPWR-S scenarios only impacts the dry periods (February–May 2005). In the wet months, the power releases reach the capacity of the power plants.



**Figure 8.** (a) Reservoir inflows, (b) optimal release decisions, (c) storage, and (d) power produced in the Gibe 1 and Gibe 3 reservoirs for the 8-month planning horizon (February–September 2005) and three scenario structures (DET, SPWR-D, and SPWR-S). Also shown are results for climatological reservoir inflows (SEASN). Note: the results correspond to the first stage of each iteration of the deterministic and stochastic programming with recourse model.

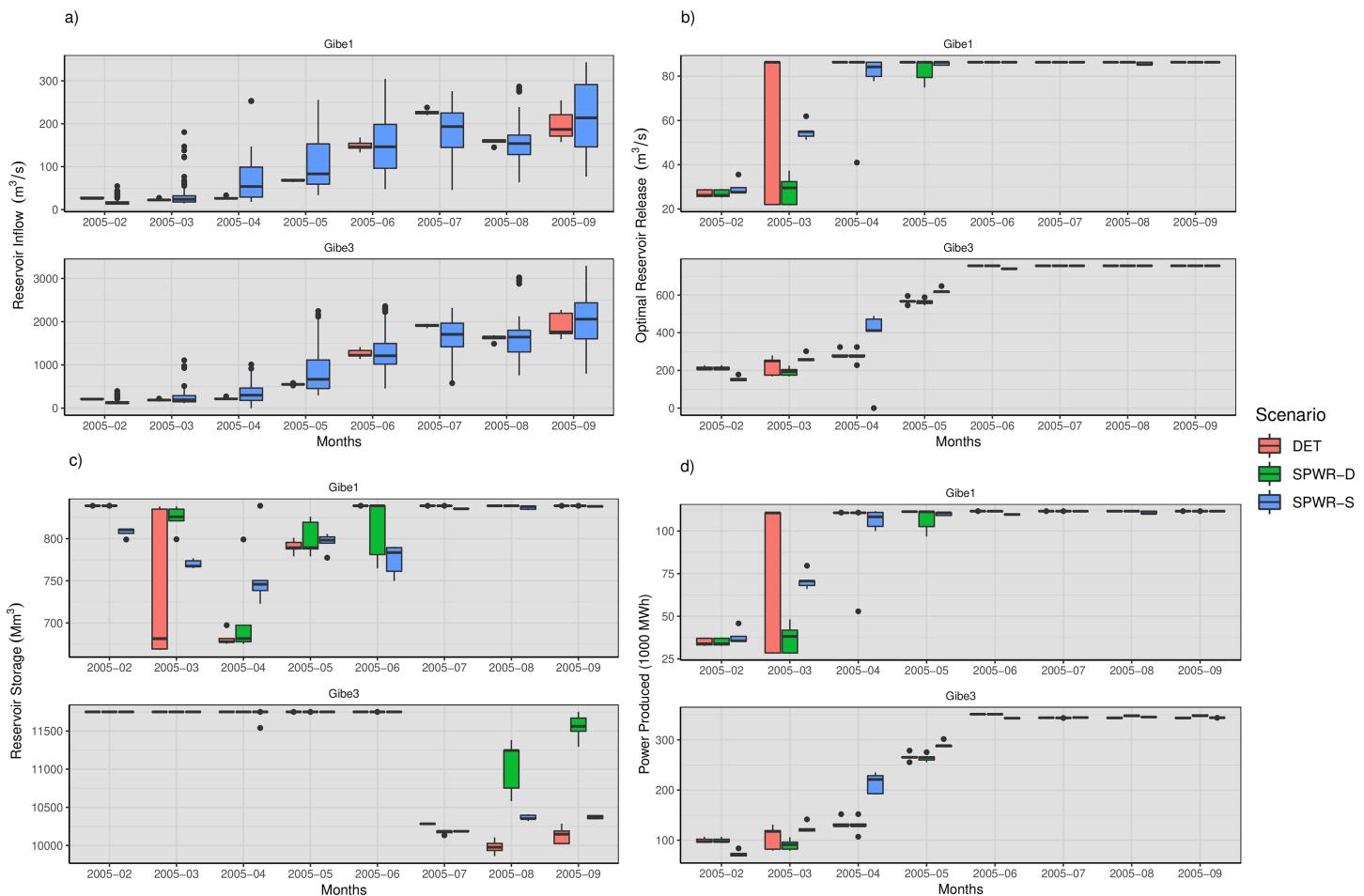
For reference we also present in Figure 8c the storage variations corresponding to the optimized release decisions. Figure 8d presents the optimized hydropower productions corresponding to the optimized release decisions for each of the three inflow scenario structures. The differences among the three inflow structures match the differences in optimal release decisions: (1) In the dry periods (February–May 2005), the SPWR-S inflow scenario structure leads to higher hydropower production compared with the DET and SPWR-D cases; (2) the differences among the scenario structures and the uncertainty in reservoir inflows do not have any impact in the wet months of the study period.

**Table 2**

Different Measures of Absolute and Relative Dispersion Determined for the Ensemble Inflows Into the Gibe I and Gibe III Reservoirs Generated Using the Best Performing Noah-MP Parameter Set.

Month	Range ( $m^3/s$ )	Standard deviation ( $m^3/s$ )	Coefficient of variation	Coefficient of range
February	(29.9, 226.1)	(7.7, 58.8)	(0.45, 0.43)	(0.56, 0.54)
March	(108.3, 815.8)	(21.6, 160.9)	(0.74, 0.69)	(0.79, 0.78)
April	(91.0, 681.4)	(25.8, 184.2)	(0.55, 0.49)	(0.71, 0.67)
May	(189.5, 1,760.4)	(59.4, 498.1)	(0.59, 0.60)	(0.73, 0.75)
June	(255.5, 1,847.1)	(64.5, 456.7)	(0.45, 0.37)	(0.72, 0.66)
July	(218.3, 1,583.8)	(54.0, 390.8)	(0.31, 0.24)	(0.71, 0.58)
August	(218.6, 2,172.9)	(53.1, 512.1)	(0.37, 0.34)	(0.63, 0.59)
September	(246.5, 2,430.8)	(81.4, 570.1)	(0.39, 0.29)	(0.62, 0.60)

Note. In the parentheses, values for Gibe I inflows are presented first, followed by values for Gibe III inflows.



**Figure 9.** Uncertainty in (a) reservoir inflows, (b) optimal release decisions, (c) storage, and (d) power produced in the Gibe 1 and Gibe 3 reservoirs for the 8-month planning horizon (February–September 2005) and three scenario structures (DET, SPWR-D, and SPWR-S), due to uncertainty in model parameters derived from Bayesian calibration. Note: the results correspond to the first stage of each iteration of the deterministic and stochastic programming with recourse model.

Finally, we compare the optimized hydropower values with the actual power produced at the Gibe I reservoir for the months of February–September 2005. Due to unavailability of historic hydropower production data, we do not present the same comparison for Gibe III reservoir. The actual power produced in Gibe I for the 8-month study period is  $0.53 \times 10^6$  MWh (sourced from Ethiopian Electric Power). In comparison, the optimized hydropower generated from the DET, SPWR-D, and SPWR-S scenario structures are  $0.73 \times 10^6$  MWh,  $0.71 \times 10^6$  MWh, and  $0.76 \times 10^6$  MWh, respectively. The higher power produced in the SPWR-S scenario structure may be attributed to the consistently higher values of reservoir inflow compared with the DET and SPWR-D scenarios. We note that reservoir evaporation has been ignored in our study, which, when included, may reduce the optimized hydropower values.

To avoid this bias, we compare the power production results of the three reservoir inflow scenarios with a baseline scenario generated with climatology. The climatological reservoir inflow forecast is generated using 10 years (January 2000 to December 2009) of observed precipitation data from TRMM and meteorological variables from the Global Data Assimilation System. The climatological inflow forecast is used as an input to the hydropower optimization model to generate optimized release policy and power production. For the 8-month planning horizon (February–September 2005), the hydropower generated from the climatological inflow forecasts in Gibe I is approximately  $0.65 \times 10^7$  MWh. In comparison, all the different scenarios (DET, SPWR-D, and SPWR-S) represent a significant improvement in terms of power production. In Gibe III, the power produced from DET, SPWR-D, and SPWR-S scenarios are  $1.96 \times 10^6$ ,  $1.96 \times 10^6$ , and  $2.1 \times 10^6$  MWh, respectively. The power produced from the climatological inflow forecast is approximately  $1.91 \times 10^6$  MWh.

**Table 3**

Different Measures of Absolute and Relative Dispersion Determined for the Ensemble Inflows Into the Gibe I and Gibe III Reservoirs Generated Using All Behavioral Noah-MP Parameter Sets

Month	Range (m <sup>3</sup> /s)	Standard Deviation (m <sup>3</sup> /s)	Coefficient of variation	Coefficient of range
February	(43.1, 301.8)	(8.2, 61.9)	(0.46, 0.43)	(0.65, 0.62)
March	(165.7, 990.8)	(24.9, 168.5)	(0.79, 0.68)	(0.85, 0.81)
April	(234.8, 1,009.0)	(85.3, 235.8)	(0.94, 0.74)	(0.87, 0.71)
May	(221.9, 1,949.9)	(60.9, 508.0)	(0.58, 0.60)	(0.77, 0.77)
June	(256.9, 1,904.9)	(62.6, 454.1)	(0.42, 0.35)	(0.73, 0.68)
July	(230.3, 1,740.2)	(53.4, 399.6)	(0.29, 0.24)	(0.72, 0.60)
August	(223.0, 2,264.7)	(50.2, 510.4)	(0.33, 0.32)	(0.63, 0.60)
September	(266.4, 2,494.6)	(80.5, 557.9)	(0.38, 0.28)	(0.63, 0.61)

Note. In the parentheses, values for Gibe I inflows are presented first, followed by values for Gibe III inflows.

$\times 10^6$  MWh. The reduced production in the case of climatological forecast can be attributed to the lower inflows during the wet months of the planning horizon (July–September 2005 in Figure 8a). However, we note that the climatological inflows during the dry season (February–May 2005) are consistently higher than the forecasts. The reduced inflows are reflected in the reduced optimized release and hence hydropower produced during the wet months from both Gibe I and Gibe III reservoirs.

In addition to incorporating uncertainty in reservoir inflow forecasts using stochastic programming, we quantify the impact of model parameter uncertainty on inflow forecasts, power release decisions, and hydropower production (Figure 9). For quantifying the impact of model parameter uncertainty, we use inflow forecasts generated from 20 behavioral model parameter sets and 30 ensemble precipitation forecasts (total of 600 inflow forecasts). Comparing Figures 8a and 9a, we see that the range and standard deviation of reservoir inflows have increased for all months in the study period. We present the specific values of range and standard deviation values in Table 3. However, we see little difference in the coefficient of variation and coefficient of range values, with the exception of April 2005. This indicates that model parameter uncertainty does not have a significant impact on the deterministic and stochastic inflow forecasts. We also present the corresponding release decisions (Figure 9b), storage variations (Figure 9c), and the optimized hydropower values (Figure 9d). Except for March 2005 for the DET and SPWR-D scenarios, the uncertainty in the release decisions and the corresponding hydropower production due to model parameter uncertainty is not very high. It is evident that the impact of model parameter uncertainty on release decisions is more pronounced during the dry months. In the wet months, model parameter uncertainty has no impact on the release decisions.

Finally, we compare the uncertainty in total hydropower production by the two reservoirs for the three inflow scenario structures. For the DET scenario, the combined hydropower production from Gibe I and Gibe III varies from  $2.68 \times 10^6$  to  $2.83 \times 10^6$  MWh. Similarly, for the SPWR-D scenario structure, the combined hydropower production varies from  $2.65 \times 10^6$  to  $2.75 \times 10^6$  MWh and from  $2.80 \times 10^6$  to  $2.91 \times 10^6$  MWh for the SPWR-S scenario structure. This shows that model parameter uncertainty can impact hydropower production by approximately  $0.14 \times 10^7$ – $0.18 \times 10^7$  MWh or 5–6% of the total power, which could be significant in seasonal planning of large hydropower systems.

## 5. Conclusions and Future Work

In this study, we developed a framework for seasonal hydropower planning in regions where reliable SF measurements are unavailable. Within this framework, we investigated the potential of combining seasonal precipitation forecasts from NWP models, ET-calibrated spatially distributed hydrologic models, and stochastic programming with recourse models for optimizing hydropower production in data-scarce regions. We compared three different inflow scenario structures that combine deterministic and stochastic reservoir inflow forecasts: (1) a single DET; (2) a scenario fan, but with the SPWR-D; and (3) a SPWR-S. In addition, we used a Bayesian calibration approach to quantify the impact of hydrologic model parameter uncertainty on the reservoir inflow forecasts, optimal release decisions, and hydropower production. We applied the framework to a cascade of two reservoirs in the Omo-Gibe River basin using NMME seasonal forecasts, the Noah-

MP hydrologic model, and a deterministic and stochastic programming with recourse model. We draw the following conclusions, addressing each of the three research questions stated in section 1.

Addressing research question 1: (a) The NWP-based 3-model, 30-member ensemble precipitation forecasts from NMME are accurate at short lead times (1–3 months). (b) BMA of the ensemble precipitation forecasts outperforms the ensemble mean as well as the individual models at all lead times. (c) The Noah-MP model calibrated with remote sensing-based ET estimates (GLEAM) performs well in simulating evapotranspiration in the calibration and validation time periods. Based on our previous findings (Koppa et al., 2019), we can reasonably conclude that the calibration of the Noah-MP hydrologic model with ET results in accurate simulation of SF. Thus, the combination of NWP-based ensemble precipitation forecasts and ET calibration hydrologic model has the potential to generate reliable reservoir inflow forecasts.

Addressing research question 2: (a) The ensemble seasonal inflow forecasts exhibit considerable uncertainty, with the deterministic inflow values consistently lower than the mean of the raw ensembles. However, parameter uncertainty does not have significant impact on the uncertainty of reservoir inflows. (b) The uncertainty in the inflow forecasts affect the optimized release decisions only in the dry months of the study period. This finding agrees with previous studies in which accurate SF forecasts improve system performance mainly in dry situations such as droughts (Turner et al., 2017). (c) In the wet period, all the ensemble reservoir inflows are very high, and the optimized power releases are governed by the capacity of the power plants. (d) Model parameter uncertainty exerts significant impact ( $0.14\text{--}0.18 \times 10^7$  MWh or 5–6% of the total power produced) on the optimized hydropower results.

Addressing research question 3: (a) In terms of hydropower production (combined Gibe I and Gibe III power production), the inflow scenario fan with SPWR-D leads to the most conservative estimate ( $2.67 \times 10^6$  MWh), followed by the completely deterministic scenario structure ( $2.69 \times 10^6$  MWh) and the SPWR-S ( $2.86 \times 10^6$  MWh). Thus, we do not see significant differences in hydropower production between the different scenario structures. However, for a large hydropower system, as opposed to the one presented here, the differences could be significant; (b) we see a considerable improvement in the optimized hydropower generated using the developed framework, compared to both the historic hydropower production in Gibe I ( $0.53 \times 10^7$  MWh) and the total optimized hydropower generated from climatology in Gibe I and Gibe III ( $2.56 \times 10^6$  MWh).

Currently, the NWP-based seasonal precipitation forecasts are available for 8–12 months. This hinders the applicability of the framework in seasonal hydropower planning, which requires forecasts with lead times greater than 12 months. Therefore, future work involves extending seasonal forecasts beyond lead times of 12 months. In our study, we assume that evapotranspiration is a reliable proxy for SF in calibrating and validating the hydrologic model. Although this is a valid assumption based on existing calibration literature, it would be interesting to analyze the impact of using other water balance components (such as soil moisture and total water storage) as proxies for SF. The framework developed in this study for seasonal hydropower planning provides the flexibility of testing different forecasts of seasonal precipitation, forecast post-processing methods of other BMA, different hydrologic models, and optimization algorithms. Finally, future work also involves application of the developed methodology to other regions with higher installed hydropower capacities.

## Appendix A: Experiment Design A

### A.1. Observational and Forecast Data

In the study area (Omo-Gibe River basin), we use monthly estimates of ET from the Global Land Evaporation Amsterdam Model (GLEAM; Martens et al., 2017) as observational data for calibrating the hydrologic model. Specifically, we use the GLEAM v3 ET data set, which assimilates remotely sensed soil moisture and vegetation optical depth from multiple satellites. The spatial resolution of the GLEAM data set is  $0.25^\circ \times 0.25^\circ$ . Similarly, the precipitation input for calibrating the hydrologic model in the Omo-Gibe River basin is the Tropical Rainfall Measuring Mission Multisatellite Precipitation Analysis data set (Huffman et al., 2007). Specifically, we utilize the real-time version (TMPA 3B42RT). The spatial resolution of the TMPA 3B42RT data set is  $0.25^\circ \times 0.25^\circ$ , and the temporal resolution is every 3 hours. We select

GLEAM ET based on the findings of Koppa and Gebremichael (2017) in which multiple satellite-based ET and precipitation data sets were ranked using a framework based on the Budyko hypothesis (Budyko, 1974).

For generating the seasonal reservoir inflow forecasts, we use the ensemble seasonal precipitation forecasts from the NMME (Kirtman et al., 2014). The NMME consists of 9 partner models with the number of ensemble members in each model varying from 6 to 28. The hydrologic model requires daily forecasts to run. Of the nine models, only three models provide the seasonal forecast at daily time step. Therefore, we select the three NMME models which provide daily forecast: (1) Goddard Earth Observation System Version 5 (GEOS-5; Borovikov et al., 2017), (2) Third Generation Canadian Coupled Global Climate Model (CanCM3; Merryfield et al., 2013), and (3) Fourth Generation Canadian Coupled Global Climate Model (CanCM4; Merryfield et al., 2013). The CanCM3 and CanCM4 models issue forecasts for lead times of up to 12 months. GEOS-5 forecasts are available for lead times of up to 9 months. With 10 ensemble members for each model, our study uses a total of 30 ensembles from 3 models. The hindcasts of NMME are available for the time period 1981–2010. The spatial resolution is  $1.0^\circ \times 1.0^\circ$ , and the temporal resolution is daily. For calculating the BMA weights we use the EBMAForecast package in R (Montgomery et al., 2016).

## A.2. Hydrologic Model: Setup, Calibration, and Model Parameter Uncertainty

To translate the NMME precipitation forecasts into reservoir inflow forecasts, we choose the Noah-MP (Multiparameterization) Land Surface Model (Niu et al., 2011), driven through NASA's Land Information System (Kumar et al., 2006). The Noah-MP model builds on the original Noah Land Surface Model by incorporating a dynamic groundwater model, improved representation of vegetation canopy, and snowpack. All the static input data sets required for running the Noah-MP model are sourced from NASA's Land Information System data portal (<https://portal.nccs.nasa.gov/lisdata>). The important static input data sets are the land cover map, sourced from the United States Geological Survey; the soil texture map from State Soil Geographic data set, sourced from the United States Department of Agriculture; and the elevation map from GTOPO30, sourced from the United States Geological Survey. Albedo, greenness fraction, and temperature are sourced from the NCEP reanalysis. The meteorological forcings required by the Noah-MP model include precipitation, air temperature, surface pressure, specific humidity, wind speed, and radiation. Barring TRMM precipitation, all meteorological forcings are derived from the Global Data Assimilation System, sourced from the Environmental Modeling Center of the NCEP (Derber et al., 1991). The spatial resolution of the data set is  $0.47^\circ \times 0.47^\circ$ . The Noah-MP model is set at a spatial resolution of  $5 \text{ km} \times 5 \text{ km}$ . The meteorological inputs, including the forecasts, are interpolated onto the model grid using bilinear interpolation. The model is spun-up for a period of 68 years by looping through the year 2003 until the groundwater and soil moisture storage reach equilibrium. The model time step is 3 hours. The number of soil layers in the model is four with thicknesses 10 cm, 30 cm, 60 cm, and 100 cm. Specific Noah-MP model physics options selected for different processes are detailed in Table A1.

The Noah-MP model contains 71 standard parameters (present in user-defined tables) and 139 hard-coded parameters (present in the model code). The Noah-MP model output has been found to be sensitive to about two thirds of the 71 standard parameters (Cuntz et al., 2016). As the study is a calibration experiment involving multiple calibration cases, we keep the dimension of the calibration problem manageable by selecting five of the most sensitive parameters from the Cuntz et al. (2016) study. The selected parameters are two surface runoff-related parameters (REFDK and REFKDT), the exponent in the Brooks-Corey equation, soil porosity, and hydraulic conductivity at saturation. Of the five parameters, BB, MAXSMC, and SATDK are related to soil texture. As there are 12 soil texture classes, the total number of parameters selected for calibration of the Noah-MP hydrologic model is 38 (Table A2 presents a detailed breakdown of the parameters with maximum and minimum values used for calibration). The minimum and maximum values of the parameter ranges are selected from literature (MAXSMC and SATDK ranges from Cai et al., 2014, BB and REFDK ranges from Cosby et al. (1984), and REFKDT range from Mendoza et al. (2015)). The minimum and maximum values are adjusted to improve the rate of convergence of the calibration algorithms.

To calibrate the hydrologic model and quantify the uncertainty in reservoir inflow forecasts due to uncertainty in model parameters, we use the MT-DREAM (ZS) algorithm. Specific configuration options and parameters of the MT-DREAM (ZS) algorithm used in this study are detailed in Table A3. In the absence of SF observations, we use satellite-based estimates of ET to calibrate the hydrologic model for the year 2004.

**Table A1**  
*Noah-MP Model Physics Options*

Model physics	Selected physics option
Vegetation model	Use table Leaf Area Index (4)
Canopy stomatal resistance	Ball-Berry (1) (Ball et al., 1987)
Soil moisture factor for stomatal resistance	Original Noah (1) (Chen et al., 1997)
Runoff and groundwater	TOPMODEL with groundwater (1) (Niu et al., 2007)
Surface layer drag coefficient	Original Noah (2) (Chen et al., 1997)
Frozen soil permeability	Linear effects, more permeable (1) (Niu & Yang, 2006)
Radiation transfer	Modified two-stream (1) (Yang & Friedl, 2003)
Snow surface albedo	CLASS (2) (Verseghe et al., 1991)
Rainfall and snowfall partitioning	Jordan Scheme (1) (Jordan, 1991)
Lower boundary of soil temperature	Original Noah (2) (Chen et al., 1997)
Snow and soil temperature time scheme	Semiimplicit (1)
Super-cooled liquid water	No iteration (Niu & Yang, 2006) (1)

*Note.* The number in the brackets represents the internal Noah-MP model code for the selected physics option.

Specifically, error residuals are determined at all of the  $5 \text{ km} \times 5 \text{ km}$  grid cells and time steps (monthly) over the entire Omo-Gibe River basin. On a workstation with 16 processors, MT-DREAM (ZS) required around 12,000 iterations to converge to a solution.

### A.3. Hydropower Optimization Model

In this study, the formulation of the hydropower optimization model is based on HIDROTERM, an NLP optimization model previously developed for planning the operation of the Brazilian hydrothermal system (Zambon et al., 2012). The model, originally deterministic, was modified to solve the multistage stochastic programming with recourse model for our study (Zambon et al., 2012).

The objective function is represented by

$$\min_{R', R''} ZH = \sum_{s=1}^{ns} \sum_{t=1}^{nt} \left\{ p_s \cdot dt_t \cdot \left( D_t - \sum_{i=1}^{ni} P_{i,s,t} \right)^2 \right\}, \quad (2)$$

where  $i$  = hydropower plant/reservoir index;  $ni$  = number of hydropower plants/reservoirs;  $dt_t$  = time period duration ( $10^6 \text{ s}$ );  $s$  = scenario index;  $ns$  = number of scenarios;  $t$  = time period index;  $nt$  = number of time periods;  $p_s$  = probability associated with each scenario;  $P_{i,s,t}$  = power production (MW);  $D_t$  = objective demand, usually the total demand minus the fixed generation, though it can be defined arbitrarily by the user (e.g., as the maximum installed power capacity MW); and  $ZH$  = model objective ( $10^6 \text{ s.MW}^2$ ). Power release  $R'_{i,s,t}$  and nonpower release  $R''_{i,s,t}$  are the decision variables.

The model minimizes the expected value of the quadratic departures from the demand so that the hydropower production will follow the specified demand variations. The model is subject to the following set of constraints:

- Continuity equation for the storage reservoirs, including evaporation losses:

**Table A2**

*Details of Noah-MP Parameters Used for Calibration*

Parameter	Description	Sensitive variable	Total parameters	Units	Minimum	Maximum
REFDK	Surface runoff parameter	SF	1	m/s	1.4e-06	6.5e-06
REFKDT	Surface runoff parameter	SF	1	No units	1.0	5.0
BB1-BB12	Exponent in the Brooks Corey Equation	SF, ET	12	No units	0.5	12.0
MAXSMC1-MAXSMC12	Soil porosity	SF, ET, SM	12	No units	0.1	0.7
SATDK1-SATDK12	Saturated hydraulic conductivity	SF, SM	12	m/s	2.0e-06	7.0e-02

*Note.* Soil texture classes for BB, MAXSMC, and SATDK (from 1 to 12): Sand, loamy sand, sandy loam, silt loam, silt, loam, sandy clay loam, silt clay loam, clay loam, sandy clay, silty clay and clay.

**Table A3**  
*MT-DREAM (ZS) and AMALGAM Configuration*

DREAM Option	Specified Option
Number of generations	600
Number of Markov chains	3
Number of forward model parameters	38
Number of crossover values	3
Number of Multitries	4
Number of chain pairs proposal	1
Prior distribution	Uniform
Likelihood function	Laplacian likelihood

*Note.* All other MT-DREAM (ZS) parameters are set to default values.

$$S_{i,s,t} = S_{i,s,t-1} + dt_t \left( \sum_{m_i=1}^{nm_i} R_{m_i,s,t} + I_{i,s,t} - R_{i,s,t} - UC_{i,t} - DT_{i,t} \right) - E_{i,s,t}. \quad (3)$$

Total release is the summation of the power and nonpower releases:

$$R_{i,s,t} = R'_{i,s,t} + R''_{i,s,t}. \quad (4)$$

Evaporation loss from a storage reservoir can be estimated by

$$E_{i,s,t} = \frac{A_{i,s,t} + A_{i,s,t-1}}{2} \cdot IE_{i,t}. \quad (5)$$

Note that in the application of the model to the Omo-Gibe River basin, evaporation loss is ignored.

2. The maximum and minimum storage constraints as well as the maximum and minimum release constraints can be imposed as follows:

$$S_{i,t}^{\min} \leq S_{i,s,t} \leq S_{i,t}^{\max}, \quad (6)$$

$$R_{i,s,t} \geq R_{i,t}^{\min}, \quad (7)$$

$$R''_{i,s,t} \geq R_{i,t}^{\min}. \quad (8)$$

3. Power production is limited by generator capacities:

$$P_{i,s,t} \leq P_i^{\max} \cdot ID_{i,t}. \quad (9)$$

4. Ending storage requirement:

$$S_{i,s,t=nt} \geq S_{i,\text{final}}. \quad (10)$$

5. Water head is the difference from reservoir forebay and tailrace water level:

$$Hg_{i,s,t} = \frac{H_{i,s,t} - H_{i,s,t-1}}{2} - HT_{i,s,t}. \quad (11)$$

6. Power production is the function of water head and power release:

$$P_{i,s,t} = \varepsilon_i \cdot (Hg_{i,s,t} - \Delta H_{i,s,t}) \cdot R'_{i,s,t}. \quad (12)$$

7. Reservoir forebay water level, reservoir forebay area, and tailrace water level are estimated by polynomials:

$$H_{i,s,t} = a_{0i} + a_{1i} \cdot S_{i,s,t} + a_{2i} \cdot S_{i,s,t}^2 + a_{3i} \cdot S_{i,s,t}^3 + a_{4i} \cdot S_{i,s,t}^4, \quad (13)$$

$$A_{i,s,t} = c_{0i} + c_{1i} \cdot H_{i,s,t} + c_{2i} \cdot H_{i,s,t}^2 + c_{3i} \cdot H_{i,s,t}^3 + c_{4i} \cdot H_{i,s,t}^4, \quad (14)$$

$$HT_{i,s,t} = b_{0i} + b_{1i} \cdot R_{i,s,t} + b_{2i} \cdot R_{i,s,t}^2 + b_{3i} \cdot R_{i,s,t}^3 + b_{4i} \cdot R_{i,s,t}^4. \quad (15)$$

8. Turbine flow maximum limit:

$$R'_{i,s,t}^{\max} = d_{0i} + d_{1i} \cdot Hg_{i,s,t} + d_{2i} \cdot Hg_{i,s,t}^2 + d_{3i} \cdot Hg_{i,s,t}^3 + d_{4i} \cdot Hg_{i,s,t}^4, \quad (16)$$

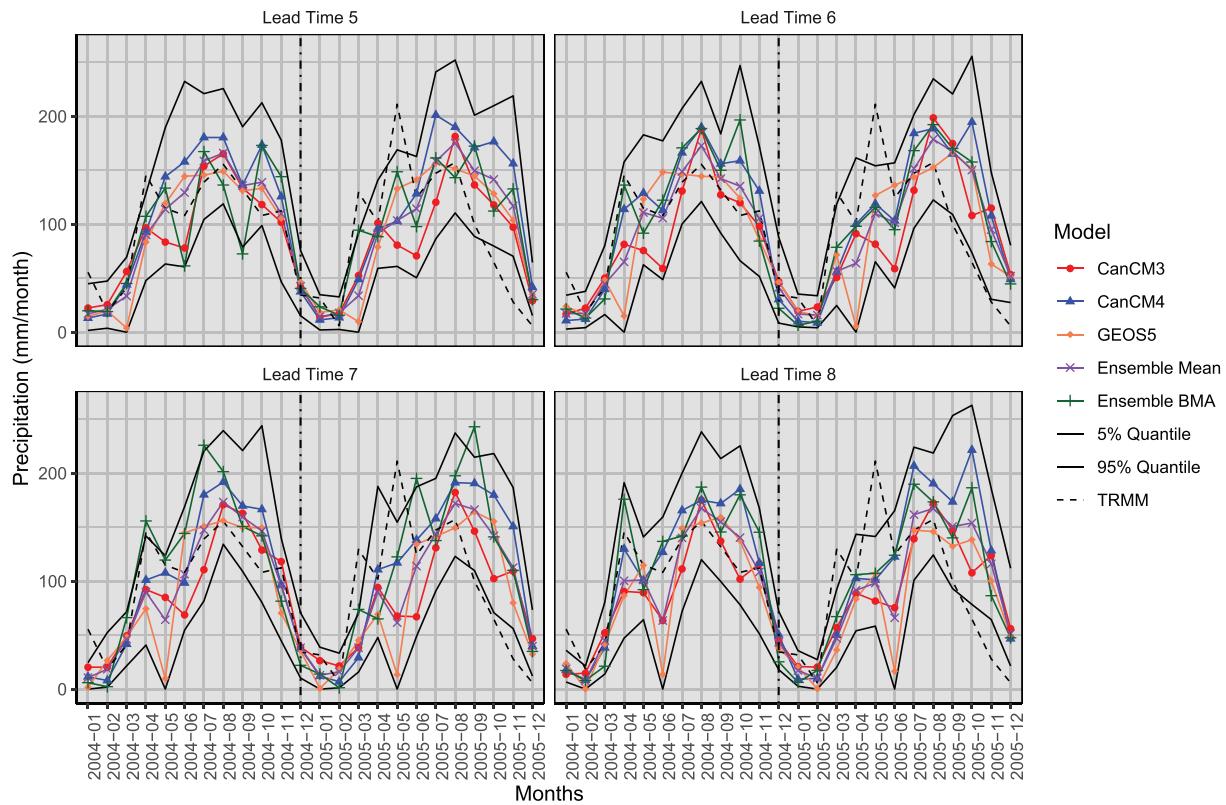
where  $m_i$  = index for each of the hydropower plants/reservoirs upstream of reservoir  $i$ ;  $nm_i$  = number of hydropower plants/reservoirs upstream of reservoir  $i$ ;  $A_{i,s,t}$  = reservoir forebay area ( $\text{km}^2$ );  $DT_{i,t}$  = water

transfer, positive indicating out-of-the-basin and negative indicating into-the-basin ( $\text{m}^3/\text{s}$ );  $E_{i,s,t}$  = evaporation loss ( $10^6 \text{ m}^3$ );  $Hg_{i,s,t}$  = water head (m);  $\Delta H_{i,s,t}$  = head loss (m);  $\varepsilon_i$  = specific productivity ( $\text{MW}/(\text{m}^3/\text{s})/\text{m}$ );  $H_{i,s,t}$  = reservoir forebay water level (m);  $HT_{i,s,t}$  = reservoir tailrace water level (m);  $ID_{i,t}$  = availability index due to maintenance and system expansion ( $0 < ID_{i,t} \leq 1$ ) (-);  $IE_{i,t}$  = evaporation rate during time period ,(m);  $I_{i,s,t}$  = lateral inflow ( $\text{m}^3/\text{s}$ );  $P_{i,s,t}$  = power production (MW);  $P_{i,t}^{\max}$  = maximum power production (MW);  $R''_{i,s,t}$  = spill or nonpower release ( $\text{m}^3/\text{s}$ );  $R'_{i,s,t}$  = power release ( $\text{m}^3/\text{s}$ );  $R_{i,s,t}$  (and  $R_{mi,s,t}$ ) = total release ( $\text{m}^3/\text{s}$ );  $R_{i,t}^{\min}$  = minimum total release ( $\text{m}^3/\text{s}$ );  $R_{i,t}^{\min'}$  = minimum nonpower release ( $\text{m}^3/\text{s}$ );  $S_{i,s,t}$  = storage ( $10^6 \text{ m}^3$ );  $S_{i,t}^{\max}$  = maximum storage ( $10^6 \text{ m}^3$ );  $S_{i,t}^{\min}$  = minimum storage ( $10^6 \text{ m}^3$ );  $S_{i,\text{final}}$  = final reservoir storage (assumed as half the max storage);  $UC_{i,t}$  = consumptive use ( $\text{m}^3/\text{s}$ );  $a_{0b}, a_{1b}, a_{2b}, a_{3b}, a_{4b}$  = storage-level fourth order polynomial coefficients;  $b_{0b}, b_{1b}, b_{2b}, b_{3b}, b_{4b}$  = release-tailrace level fourth order polynomial coefficients;  $c_{0b}, c_{1b}, c_{2b}, c_{3b}, c_{4b}$  = level-area forth order polynomial coefficients; and  $d_{0b}, d_{1b}, d_{2b}, d_{3b}, d_{4b}$  = head-maximum power release fourth order polynomial coefficients.

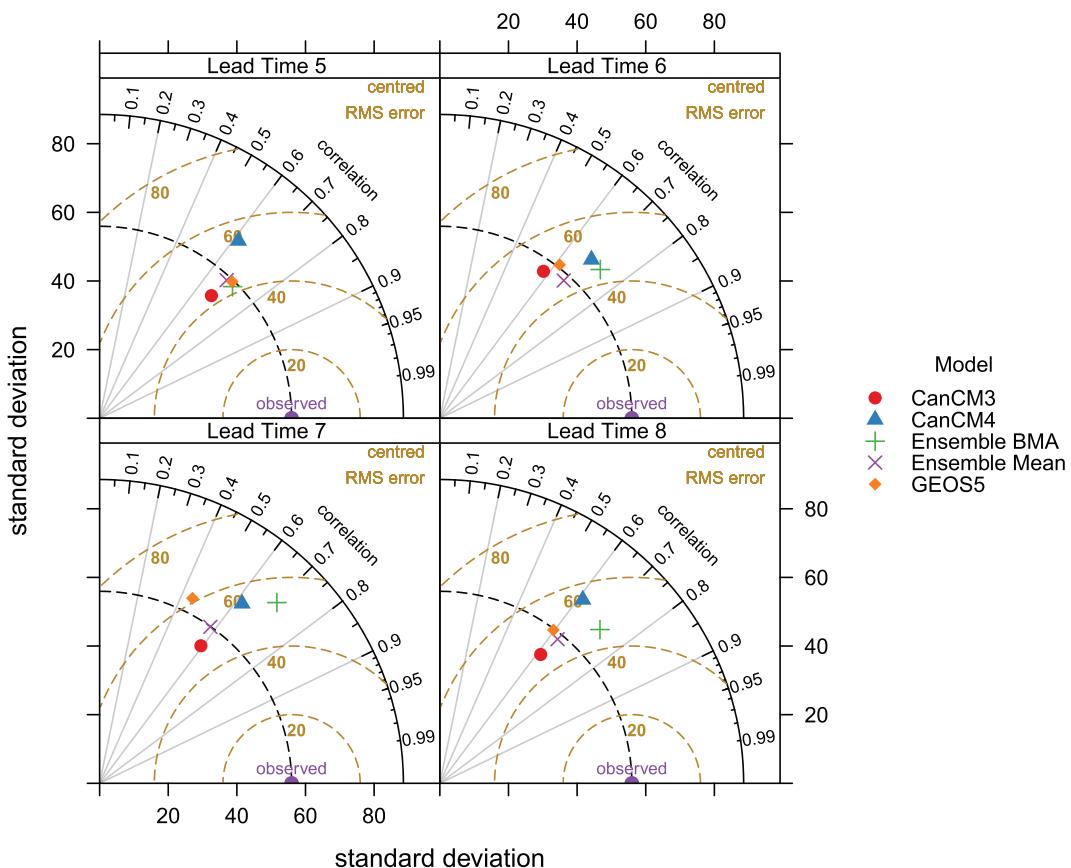
The deterministic equivalent of the stochastic model is solved by NLP using the General Algebraic Modeling System package (, 2018). The DET has 32 decision variables and 320 constraints. The SPWR-D scenario has 844 decision variables and 8,720 constraints. The SPWR-S scenario has 960 decision variables and 9,600 constraints. The execution times of each optimization run for the DET, SPWR-D, and SPWR-S scenarios are on an average 0.28 seconds, 1.65 seconds, and 2.01 seconds, respectively.

## Appendix B: Validation of NMME Precipitation Forecasts

We present the time series comparison between the three NMME models (30-member ensemble) and the TRMM remote sensing data for lead times of 5 months to 8 months in Figure B1. The corresponding Taylor diagrams are presented in Figure B2.



**Figure B1.** Time series comparison of precipitation from different NMME models and observations (TRMM) for different lead times (5–8 months) and for the calibration (12 months of 2004) and validation (12 months of 2005) time periods. We present the mean of the 10 members from CanCM3, CanCM4, and GEOS-5. “Ensemble Mean” and “Ensemble BMA” are the simple mean and Bayesian Model Averaging of all 30 ensemble members. In addition, we present the 5% and 95% quantiles of all 30 ensemble members. The vertical line demarcates the calibration (2004) and validation (2005) time periods.



**Figure B2.** Taylor diagrams of precipitation from the CanCM3, CanCM4, GEOS-5, Ensemble Mean, and BMA models determined for different lead times (5–8 months) and for the calibration (12 months of 2004) and validation (12 months of 2005) time periods. The CanCM3, CanCM4, and GEOS-5 models are calculated as the mean of the 10 ensemble members of the three models. “Ensemble Mean” and “Ensemble BMA” are the simple mean and Bayesian Model Averaging of all 30 ensemble members.

### Appendix C: List of Abbreviations and Acronyms C

A list of important abbreviations and acronyms used in the study is presented in Table C1.

**Table C1**  
*A glossary of Important Abbreviations and Acronyms Used in the Study*

Acronym	Description
BB	Exponent in the Brooks-Corey equation (Noah-MP model parameter)
BMA	Bayesian Model Averaging
CanCM	Canadian Coupled Global Climate Model
DET	Single deterministic forecast of reservoir inflow (scenario structure)
DREAM	Differential Evolution Adaptive Metropolis
GDAS	Global Data Assimilation System
GEOS-5	Goddard Earth Observation System Version 5
GLEAM	Global Land Evaporation Amsterdam Model
MAXSMC	Soil Porosity (Noah-MP model parameter)
MME	Multimodel Ensembles
NWP	Numerical Weather Prediction
REFDK and REFKDT	Surface runoff parameters (Noah-MP model parameters)
SATDK	Hydraulic conductivity at saturation (model parameter)
SPWR-D	Inflow scenario fan with first stage deterministic (scenario structure)
SPWR-S	Inflow scenario fan with all stages stochastic (scenario structure)
TMPA	Tropical Rainfall Measuring Mission
TRMM	TRMM Multisatellite Precipitation Analysis

## Acknowledgments

We acknowledge funding support from the NASA Applied Sciences Water Resources Application Grant NNX15AC33G. We thank Jasper Vrugt for help with the DREAM algorithm. We thank Solomon Demissie for providing historic power production data for Gilgel Gibe I. We thank three anonymous reviewers for their constructive comments which have improved the article. The NMME seasonal precipitation forecasts have been sourced from the NMME website (<https://www.cpc.ncep.noaa.gov/products/NMME/data.html>). The GLEAM ET data set has been sourced from the GLEAM website ([www.gleam.eu](http://www.gleam.eu)). All hydrologic model configuration files are openly available at the UCLA website (<https://doi.org/10.6084/m9.figshare.10120835.v3>).

## References

- Alemu, E. T., Palmer, R. N., Polebitski, A., & Meaker, B. (2011). Decision support system for optimizing reservoir operations using ensemble streamflow predictions. *Journal of Water Resources Planning and Management*, 137(1), 72–82. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000088](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000088)
- Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., Su, C.-H., & Robertson, D. R. (2015). Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. *Water Resources Research*, 52, 5357–5375. <https://doi.org/10.1002/2015WR018429>
- Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, 52, 4209–4225. <https://doi.org/10.1002/2015WR017864>
- Ball, J. T., Woodrow, I. E., & Berry, J. A. (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In *Progress in Photosynthesis Research* (pp. 221–224). Netherlands, Dordrecht: Springer. [https://doi.org/10.1007/978-94-017-0519-6\\_48](https://doi.org/10.1007/978-94-017-0519-6_48)
- Barros, M. T., Tsai, F. T., Yang, S. L., Lopes, J. E., & Yeh, W. W. (2003). Optimization of large-scale hydropower system operations. *Journal of Water Resources Planning and Management*, 129(3), 178–188. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2003\)129:3\(178](https://doi.org/10.1061/(ASCE)0733-9496(2003)129:3(178)
- Becker, E., & van den Dool, H. (2016). Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment. *Journal of Climate*, 29(8), 3015–3026. <https://doi.org/10.1175/JCLI-D-14-00862.1>
- Becker, E., van den Dool, H., & Zhang, Q. (2014). Predictability and forecast skill in NMME. *Journal of Climate*, 27(15), 5891–5906. <https://doi.org/10.1175/JCLI-D-13-00597.1>
- Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., & Schüth, C. (2019). Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived ET data. *Journal of Hydrology*, 575s7. <https://doi.org/10.1016/j.jhydrol.2019.123944>
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, 320(1-2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>
- Block, P. (2011). Tailoring seasonal climate forecasts for hydropower operations. *Hydrology and Earth System Sciences*, 15(4), 1355–1368. <https://doi.org/10.5194/hess-15-1355-2011>
- Block, P. J., Souza Filho, F. A., Sun, L., & Kwon, H. (2009). A streamflow forecasting framework using multiple climate and hydrological models. *JAWRA Journal of the American Water Resources Association*, 45, 828–843. <https://doi.org/10.1111/j.1752-1688.2009.00327.x>
- Borovikov, A., Cullather, R., Kovach, R., Marshak, J., Vernieres, G., Vikhlaev, Y., et al. (2017). GEOS-5 seasonal forecast system. *Climate Dynamics*, 1–27. <https://doi.org/10.1007/s00382-017-3835-2>
- Budyko, M. (1974). *Climate and life, International Geophysics Series*. Cambridge, Massachusetts, United States: Academic Press.
- Cai, X., Yang, Z.-L., David, C. H., Niu, G.-Y., & Rodell, M. (2014). Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. *Journal of Geophysical Research: Atmospheres*, 119, 23–38. <https://doi.org/10.1002/2013JD020792>
- Cash, B. A., Manganello, J. V., & Kinter, J. L. (2017). Evaluation of NMME temperature and precipitation bias and forecast skill for South Asia. *Climate Dynamics*, 1–18. <https://doi.org/10.1007/s00382-017-3841-4>
- Chaeviso, S. E., Abebe, A., & Pingale, S. M. (2016). Assessment of the impact of climate change on surface hydrological processes using SWAT: A case study of Omo-Gibe River basin, Ethiopia. *Modeling Earth Systems and Environment*, 2(4), 1–15. <https://doi.org/10.1007/s40808-016-0257-9>
- Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-Layer Meteorology*, 85(3), 391–421. <https://doi.org/10.1023/a:1000531001463>
- Conway, D., Dalin, C., Landman, W. A., & Osborn, T. J. (2017). Hydropower plants in Eastern and Southern Africa increase risk of concurrent climate-related electricity supply disruption. *Nature Energy*, 2(12), 946–953. <https://doi.org/10.1038/s41560-017-0037-4>
- Conway, D., van Garderen, E. A., Deryng, D., Dorling, S., Krueger, T., Landman, W., et al. (2015). Climate and Southern Africa's water-energy-food nexus. *Nature Climate Change*, 5, 837. <https://doi.org/10.1038/nclimate2735>
- Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres*, 121, 10,676–10,700. <https://doi.org/10.1002/2016jd025097>
- Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20(6), 682–690. <https://doi.org/10.1029/wr020i006p00682>
- Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, 111(2), 157–170. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)
- Derber, J. C., Parrish, D. F., & Lord, S. J. (1991). The new global operational analysis system at the national meteorological center. *Weather and Forecasting*, 6(4), 538–547. [https://doi.org/10.1175/1520-0434\(1991\)0062.0.co;2](https://doi.org/10.1175/1520-0434(1991)0062.0.co;2)
- Etkin, D., Kirshen, P., Watkins, D., Roncoli, C., Sanon, M., Some, L., et al. (2015). Stochastic programming for improved multiuse reservoir operation in Burkina Faso, West Africa. *Journal of Water Resources Planning and Management*, 141(3), s. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000396](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000396)
- GAMS. (2018). General Algebraic Modeling System. <https://www.gams.com>
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model Averaging: A tutorial. *Statistical science*, 382–401. <https://doi.org/10.1214/ss/1009212519>
- Hoyland, K., & Wallace, S. W. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2), 295–307. <https://doi.org/10.1287/mnsc.47.2.295.9834>
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8, 38–55. <https://doi.org/10.1175/JHM560.1>
- Immerzeel, W., & Droogers, P. (2008). Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology*, 349(3), 411–424. <https://doi.org/10.1016/j.jhydrol.2007.11.017>
- Jordan, R. (1991). A One-Dimensional Temperature Model for a Snow Cover: Techni- cal Documentation for SNTHERM.89. Cold Regions Research and Engineering Lab, Hanover NH Tech. Rep.
- Khanmohammadi, N., Rezaie, H., Montaseri, M., & Behmanesh, J. (2018). Regional probability distribution of the annual reference evapotranspiration and its effective parameters in Iran. *Theoretical and Applied Climatology*, 134(1-2), 411–422. <https://doi.org/10.1007/s00704-017-2283-6>

- Kim, Y.-O., & Palmer, R. N. (1997). Value of seasonal flow forecasts in Bayesian stochastic programming. *Journal of Water Resources Planning and Management*, 123(6), 327–335. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1997\)123:6\(327\)](https://doi.org/10.1061/(ASCE)0733-9496(1997)123:6(327))
- Kirtman, B. P., Min, D., Infantini, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- Koppa, A., & Gebremichael, M. (2017). A Framework for validation of remotely sensed precipitation and evapotranspiration based on the Budyko hypothesis. *Water Resources Research*, 53, 8487–8499. <https://doi.org/10.1002/2017wr020593>
- Koppa, A., Gebremichael, M., & Yeh, W. W.-G. (2019). Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. *Advances in Water Resources*, 130, 129–146. <https://doi.org/10.1016/j.advwatres.2019.06.005>
- Krishnamurti, T. N., Kumar, V., Simon, A., Bhardwaj, A., Ghosh, T., & Ross, R. (2015). A review of multimodel super ensemble forecasting for weather, seasonal climate, and hurricanes. *Reviews of Geophysics*, 54, 336–377. <https://doi.org/10.1002/2015RG000513>
- Kumar, S., Peters-Lidard, C., Tian, Y., Houser, P., Geiger, J., Olden, S., et al. (2006). Land information system: an interoperable framework for high resolution land surface modeling. *Environmental Modelling and Software*, 21(10), 1402–1415. <https://doi.org/10.1016/j.envsoft.2005.07.004>
- Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., & Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov Chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49, 2664–2682. <https://doi.org/10.1002/wrcr.20226>
- Lavers, D., Luo, L., & Wood, E. F. (2009). A multiple model assessment of seasonal climate forecast skill for applications. *Geophysical Research Letters*, 36, L23711. <https://doi.org/10.1029/2009GL041365>
- Lee, Y., Kim, S. K., & Ko, I. H. (2008). Multistage stochastic linear programming model for daily coordinated multi-reservoir operation. *Journal of Hydroinformatics*, 10(1), 23–41. <https://doi.org/10.2166/hydro.2008.007>
- Lettenmaier, D., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51, 7309–7342. <https://doi.org/10.1002/2015WR017616>
- López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*, 21(6), 3125–3144. <https://doi.org/10.5194/hess-21-3125-2017>
- Lund, J. R. (1996). Developing seasonal and long-term reservoir system operation plans using HEC-PRM. No. HEC-RD-40, Hydrologic Engineering Center Davis CA.
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., et al. (2017). GLEAM v3: Satellite-based land evaporation and rootzone soil moisture. *Geoscientific Model Development*, 10(5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- Martinet, T. M., Berkovich, S. G., & Schulten, K. J. (1993). Neural-gas network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks*, 4(4), 558–569. <https://doi.org/10.1109/72.238311>
- Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., et al. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, 16(2), 762–780. <https://doi.org/10.1175/jhm-d-14-0104.1>
- Merryfield, W. J., Lee, W., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization. *Monthly Weather Review*, 141, 2910–2945. <https://doi.org/10.1175/MWR-D-12-00216.1>
- Montgomery, J. M., Hollenbach, F.M., & Ward, M.D. (2016). EBMAforecast: Ensemble BMA Forecasting. R package version 0.52. <https://CRAN.R-project.org/package=EBMAforecast>
- Niu, G.-Y., & Yang, Z.-L. (2006). Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *Journal of Hydrometeorology*, 7(5), 937–952. <https://doi.org/10.1175/jhm538.1>
- Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. *Journal of Geophysical Research*, 112, D07103. <https://doi.org/10.1029/2006jd007522>
- Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, 116, D12109. <https://doi.org/10.1029/2010jd015139>
- Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of official western US seasonal water supply outlooks, 1922–2002. *Journal of Hydrometeorology*, 5(5), 896–909. [https://doi.org/10.1175/1525-7541\(2004\)005<0896:EOOWUS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2)
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., et al. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMIETER). *Bulletin of the American Meteorological Society*, 85(6), 853–872. <https://doi.org/10.1175/BAMS-85-6-853>
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian Model Averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174. <https://doi.org/10.1175/MWR2906.1>
- Rajib, A., Evenson, G. R., Golden, H. E., & Lane, C. R. (2018). Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. *Journal of hydrology*, 567, 668–683. <https://doi.org/10.1016/j.jhydrol.2018.10.024>
- Saavedra Valeriano, O. C., Koike, T., Yang, K., Graf, T., Li, X., Wang, L., & Han, X. (2010). Decision support for dam release during floods using a distributed biosphere hydrological model driven by quantitative precipitation forecasts. *Water Resources Research*, 46, W10544. <https://doi.org/10.1029/2010WR009502>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, 27(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., et al. (2006). The NCEP climate forecast system. *Journal of Climate*, 19(15), 3483–3517. <https://doi.org/10.1175/JCLI3812.1>
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46, W10531. <https://doi.org/10.1029/2009wr008933>
- Séguin, S., Audet, C., & Côté, P. (2017). Scenario-Tree Modeling for Stochastic Short-Term Hydropower Operations Planning. *Journal of Water Resources Planning and Management*, 143(12), 04017073. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000854](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000854)

- Shafii, M., Tolson, B., & Matott, L. S. (2014). Uncertainty-based multi-criteria calibration of rainfall-runoff models: A comparative study. *Stochastic Environmental Research and Risk Assessment*, 28(6), 1493–1510. <https://doi.org/10.1007/s00477-014-0855-x>
- Shrestha, D. L., Robertson, D. E., Bennett, J. C., & Wang, Q. J. (2015). Improving precipitation forecasts by generating ensembles through postprocessing. *Monthly Weather Review*, 143(9), 3642–3663. <https://doi.org/10.1175/MWR-D-14-00329.1>
- Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences*, 15(11), 3529–3538. <https://doi.org/10.5194/hess-15-3529-2011>
- Siegmund, J., Bliefernicht, J., Laux, P., & Kunstmann, H. (2015). Toward a seasonal precipitation prediction system for West Africa: Performance of CFSv2 and high-resolution dynamical downscaling. *Journal of Geophysical Research: Atmospheres*, 120, 7316–7339. <https://doi.org/10.1002/2014JD022692>
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6), 857–880. <https://doi.org/10.1623/hysj.48.6.857.51421>
- Sloughter, J. M., Raftery, A. E., Gneiting, T., & Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian Model Averaging. *Monthly Weather Review*, 135, 3209–3220. <https://doi.org/10.1175/MWR3441.1>
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Turgeon, A. (2005). Solving a stochastic reservoir management problem with multilag autocorrelated inflows. *Water Resources Research*, 41, W12414. <https://doi.org/10.1029/2004WR003846>
- Turner, S. W. D., Bennett, J. C., Robertson, D. E., & Galelli, S. (2017). Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrology and Earth System Science*, 21, 4841–4859. <https://doi.org/10.5194/hess-21-4841-2017>
- Verseghy, D. L., McFarlane, N. A., & Lazare, M. (1991). Class-A Canadian land surface scheme for GCMS, II. Vegetation model and coupled runs. *International Journal of Climatology*, 13(4), 347–370. <https://doi.org/10.1002/joc.3370130402>
- Vogel, R. M., & Wilson, I. (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *Journal of hydrologic Engineering*, 1(2), 69–76. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1996\)1:2\(69\)](https://doi.org/10.1061/(ASCE)1084-0699(1996)1:2(69))
- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software*, 75, 273–316. <https://doi.org/10.1016/j.envsoft.2015.08.013>
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov Chain Monte Carlo simulation. *Water Resources Research*, 44, W00B09. <https://doi.org/10.1029/2007WR006720>
- Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273–290. <https://doi.org/10.1515/IJNSNS.2009.10.3.273>
- Wanders, N., Bierkens, M. F. P., de Jong, S. M., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, 50, 6874–6891. <https://doi.org/10.1002/2013WR014639>
- Wang, F., Wang, L., Zhou, H., Saavedra Valeriano, O. C., Koike, T., & Li, W. (2012). Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. *Water Resources Research*, 48. <https://doi.org/10.1029/2011WR011366>
- Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Deque, M., et al. (2009). ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions skill and progress beyond DEMETER in forecasting tropical pacific SSTs. *Geophysical Research Letters*, 36, L21711. <https://doi.org/10.1029/2009GL040896>
- Wood, A. W., & Lettenmaier, D. P. (2006). A test bed for new seasonal hydrologic forecasting approaches in the Western United States. *Bulletin of the American Meteorological Society*, 87(12), 1699–1712. <https://doi.org/10.1175/BAMS-87-12-1699>
- Xu, W., Zhang, C., Peng, Y., Fu, G., & Zhou, H. (2014). A two stage Bayesian stochastic optimization model for cascaded hydropower systems considering varying uncertainty of flow forecasts. *Water Resources Research*, 50, 9267–9286. <https://doi.org/10.1002/2013WR015181>
- Yang, R., & Friedl, M. A. (2003). Modeling the effects of three-dimensional vegetation structure on surface radiation and energy balance in boreal forests. *Journal of Geophysical Research*, 108(D16) 8615. <https://doi.org/10.1029/2002jd003109>
- Yatheendradas, S., Lidard, C. D. P., Koren, V., Cosgrove, B. A., De Goncalves, L. G. G., Smith, M., et al. (2012). Distributed assimilation of satellite-based snow extent for improving simulated streamflow in mountainous, dense forests: An example over the DMIP2 western basins. *Water Resources Research*, 48, W09557. <https://doi.org/10.1029/2011WR011347>
- Yeh, W. W.-G. (1985). Reservoir management and operations models: A state-of-the-art re-view. *Water Resources Research*, 21(12), 1797–1818. <https://doi.org/10.1029/WR021i012p01797>
- Yuan, X., Wood, E. F., Luo, L., & Pan, M. (2011). A first look at Climate Forecast System Version 2 (CFSv2) for hydrological seasonal prediction. *Geophysical Research Letters*, 38, L13402. <https://doi.org/10.1029/2011GL047792>
- Zambon, R. C., Barros, M. T. L., Barbosa, P. S. F., Francato, A. L., Lopes, J. E. G., & Yeh, W. W.-G. (2012). Two-stage stochastic optimization of large-scale hydrothermal system. In *World Environmental and Water Resources Congress 2012* (pp. 2472–2481). Reston, VA: American Society of Civil Engineers.
- Zambon, R. C., Barros, M. T. L., Lopes, J. E. G., Barbosa, P. S. F., Francato, A. L., & Yeh, W. W.-G. (2012). Optimization of large-scale hydrothermal system operation. *Journal of Water Resources Planning and Management*, 138(2), 135–143. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000149s](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000149s)
- Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. *Water Resources Research*, 54, 2976–2998. <https://doi.org/10.1002/2017WR021346>