Check for updates

ASCE

Simulating Hydropower Discharge using Multiple Decision Tree Methods and a Dynamical Model Merging Technique

Tiantian Yang[1]; Xiaomang Liu[2]; Lingling Wang[3]; Peng Bai[4]; and Jingjing Li[5]

Abstract: Hydropower release decision making relies on multisource information, such as climate conditions, downstream water quality, inflow and storage, regulation and engineering constraints, and so on. The decision tree (DT) method is one of the commonly used techniques to simulate reservoir operation and release strategies because of its simplicity and effectiveness. However, the performances and simulation accuracy vary among different DT models due to many structures and splitting rules associated with each DT model. In this study, we propose a dynamic merge technique (DMerge), which adopts a concept from particle swarm optimization, to postprocess outputs from different DT models with the purpose of increasing the simulation accuracy and producing a model ensemble with dynamically changing weights throughout the validation phase. A case study of Shasta Lake in northern California is presented, where the daily hydropower releases are predicted and compared using the DMerge, AdaBoost DT, random forest, and extremely randomized trees methods. Results show that the DMerge method has the best statistics compared to other popular DT algorithms. Furthermore, scenario tests were carried out to analyze the sensitivity to model inputs (i.e., hydrological condition, reservoir storage and regulation, climate phenomenon indices, and water quality) with respect to explaining the variability of hydropower releases. According to the results, we found that the hydropower releases are a complex decision-making process and water quality and climate conditions could play an even more significant role than both hydrological forcing and system states in our case study. The proposed DMerge method is a robust and efficient technique in solving water-energy prediction and simulation problems, and it is suitable for joint use with other data-driven approaches. **DOI: 10.1061/(ASCE)WR.1943-5452.0001146.**
© 2019 American Society of Civil Engineers.

Author keywords: Hydropower; Environmental impacts; Reservoir operation; Decision tree; Climate change; Artificial intelligence.

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

Introduction

Hydropower is an important renewable and clean energy resource produced mostly within reservoir systems or by run-of-river dams. In California, hydroelectric generation provides about 15% of all electricity generation (CDWR 2017). Hydropower is a flexible energy source to provide power during peak loads and extra spinning reserves (Chang et al. 2013; Kaygusuz 2004; Li et al. 2015; Tarroja et al. 2014), which is an extra power generating capacity being available by increasing the power output of assets that are already connected to the electric grid. Hydropower is also more stable than solar and wind power, because it is not vulnerable to day-night shifts and abrupt changes in weather conditions (CDWR 2017; Yang 2015; Yang et al. 2017a). In many developing countries, such as China, hydropower resources have been under rapid exploitation in recent decades (Cheng et al. 2017; Ji et al. 2015; Li et al. 2014, 2015). Examples include the recently built Xiluodu hydropower dam, Ertan hydropower dam, and so on, and existing constructions of hydropower facilities along the Jinsha River, Yalong River, and Yangtze River upstream in southwest China (Zhang et al. 2018). Hydropower production depends on many factors, which include (1) local hydrological conditions and available water, such as precipitation, evaporation, and upstream inflows to reservoirs; (2) reservoir storage levels (i.e., the head difference between forebay water level and tailwater), engineering constraints, and system operating rules; (3) large-scale climatic conditions that influence both demand and supply of water and electrical power; and (4) water quality parameters for the benefits of the aquatic ecosystem and survival of fish species downstream from hydropower facilities, such as water temperature, turbidity, and dissolved oxygen (DOE 2014a, b). Modeling hydropower production and turbine water releases from reservoirs is challenging due to the complexity and variety of the decision variables associated with the generation and human decision-making processes (Conklin et al. 2007; Kaygusuz 2004; Madani 2011). How to better simulate hydropower releases and how to include as many relevant decision variables as possible in the modeling framework are two research questions that are of interest to both scientists and operators.

In order to mathematically model the reservoir decision-making process, one challenge associated with the physical and mass-balance model is the lack of capability of utilizing ancillary

[1]Assistant Professor, Key Laboratory of Soil and Water Loss Process and Control on the Loess Plateau of Ministry of Water Resources, Yellow River Institute of Hydraulic Research and Univ. of Oklahoma, 202 W. Boyd St., Room 427, Norman, OK 73019. Email: Tiantian.Yang@ou.edu

[2]Associate Professor, Key Laboratory of Water Cycle and Related Land Surface Process, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 1 Datun Rd., Chaoyang Qu, Beijing 100101, China (corresponding author). Email: hydroliu@163.com

[3]Professor, Key Laboratory of Soil and Water Loss Process and Control on the Loess Plateau of Ministry of Water Resources, Yellow River Institute of Hydraulic Research, 12 Chengbei Rd., ZiJingShan ShangQuan, Jinshui Qu, Zhengzhou, Henan 450003, China. Email: llwang@yrihr.com.cn

[4]Assistant Professor, Key Laboratory of Water Cycle and Related Land Surface Process, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 1 Datun Rd., Chaoyang Qu, Beijing 100101, China. Email: baip@igsnrr.ac.cn

[5]Assistant Professor, Dept. of Geosciences and Environment, California State Univ. Los Angeles, 5151 State University Dr., Los Angeles, CA 90032. ORCID: https://orcid.org/0000-0002-4237-7389. Email: jli104@calstatela.edu

Note. This manuscript was submitted on November 15, 2018; approved on May 22, 2019; published online on November 30, 2019. Discussion period open until April 30, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

information to predict or simulate hydropower release schedules. For example, under certain circumstances, large-scale climatic phenomenon (e.g., atmospheric river events) are driving the local hydrology and inflow amount to reservoirs over the western US. Large-scale climatic conditions, such as Pacific decadal oscillation (PDO) (Mantua and Hare 2002) and El Niño southern oscillation (ENSO) (Cane 2005), are known to impact precipitation variability and thus water availability for our study areas. These factors impact hydropower production potential by changing local dry-wet conditions, extreme precipitation events, timing of spring snowmelt, and demand for electrical power (Barlow et al. 2001; Cook et al. 2007; Dettinger et al. 2011; Namias and Cayan 1981; Redmond and Koch 1991). Furthermore, regulations and constraints on water quality may also be as influential as other decision variables for discharge simulation, and physical-based reservoir simulation models lack the ability to simultaneously use different types of decision variables and mimic the fuzzy reservoir release strategies. From a statistical point of view, the sensitivity evaluation of model inputs is also useful to help operators to extract key information from a complex set of data records and to identify the important variables that possess the high predictability of reservoir release patterns.

Hydropower scheduling decision making (i.e., the amount of turbine releases at a certain time) is a complex process because it depends on multiple decision factors, which are beyond what is required for reservoir outflow simulation. Most of the factors do not have a direct physical relationship with the hydropower generation governing law (i.e., Production $= \rho g Q \Delta h$, where ρ = density of water; g = gravity constant; Q = turbine water release rate; and Δh = water static head difference between forebay and tailwater elevations). For example, reservoir storage level and Δh together decide how much gravity potential is needed for hydropower production. System constraints, flood control, and water supply operating rules regulate how much water can be used for hydropower production, including the design elements of the hydropower plant itself.

To better simulate hydropower releases and take advantage of the multisource information available to operators, artificial intelligence (AI) and data-mining (DM) techniques have gained much popularity in the fields of reservoir operation and system decision making during the last decade. The AI and DM tools mainly focus on deriving a simulation model for reservoir release, evaluating the existing operation rules, or predicting future flows as a regression model, rather than optimizing the system yield (storage or discharge) (Ashaary et al. 2015; Castelletti et al. 2010; Chang et al. 2016; Cheng et al. 2015; Galelli and Castelletti 2013; Kumar et al. 2013a; Shamim et al. 2016; Yang et al. 2017a). Among different forms of AI and DM methods, the decision tree (DT) method is one of the most popular techniques that is commonly used in reservoir discharge simulation (Zagona et al. 2001; Bessler et al. 2003; Cheng et al. 2008; Sattari et al. 2012; Tsai et al. 2012; Wei 2012; Zhang et al. 2015; Yang et al. 2016; Kim et al. 2019). The reason is that the actual reservoir discharge operation is a multistaged decision-making process, which shares a similar logic used in DT methods, i.e., the true or false conditional Boolean logic. For example, under normal operations, dam operators typically use current storage level and rule curves (i.e., the relationship between discharge and storage) to decide whether to release a certain amount of water from reservoirs (Raso et al. 2014; Zagona et al. 2001). Dam operators also take certain control actions based on whether incoming flows from upper streams exceed a certain level that could result in flooding. Theoretically, those control actions mentioned above can be well interpreted by properly designed if-then logic and corresponding thresholds using storage levels and incoming streamflow as decision variables (Schwanenberg et al. 2014, 2015).

In the literature, there are a number of studies that focus on applying DT methods to reservoir discharge simulation. For example, Wei (2012) employed two standard DT algorithms, namely, the classification and regression trees (CART) and a C5.0 decision tree algorithm, for a reservoir discharge simulation problem during typhoon events over the Taiwan area. Sattari et al. (2012) tested the if-then logic from the DT method with multiple reservoir release data sets in Iran and concluded that the decision tree techniques are suitable for determining reservoir operating rules for irrigation. Yang et al. (2016) carried out a large simulation of reservoir discharge using a CART algorithm and a random forest (RF) algorithm, and they achieved a high simulation accuracy when mimicking reservoir releases. However, when applying DT methods to the hydropower release simulation, different models are not always in agreement with each other with respect to the overall performance along the prediction horizon, which is a common phenomenon for many statistical and AI models. Different DT methods, for example, the AdaBoost tree methods (Banfield et al. 2007; Bauer and Kohavi 1999; Breiman 1996; Dietterich 2000a, b), RF algorithm (Breiman 2001), and extremely randomized trees algorithm (Geurts et al. 2006, 2007; Marée et al. 2007), use various mechanisms of adding randomness and constructing ensemble candidates (i.e., whether the candidate trees are required to have complex or simple structures). As a result, different DT models will perform differently on the same problem or data set. A single DT model's performance could also vary during the test period since the predictors used in the test data set are treated as independent data points. Therefore, to select a single DT model that consistently performs well during the test period is a difficult task due to the fact that model performances are not stable during the entire test period and the model structures are not consistent compared among different DT models. The research question, therefore, becomes how to improve the prediction accuracy utilizing the outputs of different DT models and the observation data that gradually become available and up to date as time goes on.

To address the above research and application question, in this study, we propose to use data-driven approaches to fuse different types of information required by hydropower release decision making and combine the outputs from many many decision tree models for better accuracy. Traditionally, an ensemble of the outputs from many models is constructed by assigning equal weights to different models' outputs, also termed the simple model average (SMA) (Hagedorn et al. 2005). In this paper, we introduce a nonequal averaging technique to postprocess different DT model outputs, termed the dynamic merging technique (DMerge). The DMerge approach identifies the changes in participating models' performances through time and significantly reduces the biases in the final ensemble. Different from the SMA method, the philosophy used in the DMerge approach is to combine two sources of information to determine the model ensemble weights at the next time step when the observations are unknown. The first source of information is the model performances at the current time step (assuming observation data become available and the previous time step prediction becomes historical). The second source of information is the general historical model performances in all previous time steps. This idea is inspired by the heuristic searching algorithm in the field of optimization, i.e., the mechanics used in particle swarm optimization (PSO) (Eberhart and Kennedy 1995; Kennedy 2011; Kennedy et al. 2001). In PSO algorithms, the direction of a particle for the next time step is determined based on the current best and historical best locations of the moving particles. In the DMerge approach, we use this similar idea to weight

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

model candidates for the next time step based on the current and historical best models' performances. The advantage of the DMerge technique is that it can efficiently use any observation data, which continuously become available through the prediction horizon (test period), and it dynamically incorporates this information to enhance the performance of the final ensemble.

Specifically, our study's goals are to (1) simulate hydropower releases based on a large variety of decision variables besides hydrological information and state variables and (2) demonstrate the superiority of the proposed DMerge method with regards to the accuracy of deriving more accurate hydropower release simulations compared to any single decision tree method. To achieve these goals, a case study is carried out to simulate the hydropower releases from Shasta Lake located in northern California, which is the headwater reservoir for California's Central Valley Project. The model inputs include 26 decision variables: 3 hydrological variables (i.e., precipitation, inflow, evaporation); 2 state and regulation variables (i.e., current storage volume and the conservative pool elevation above storage elevation); 18 climate phenomenon indicators [i.e., ENSO, PDO, and arctic oscillation (AO)]; and 3 water quality indicators (i.e., the water temperature, dissolved oxygen, and water turbidity measured downstream from the reservoir). The results from the DMerge method are compared against several popular decision tree methods and the standard SMA method. Both statistical and trajectory comparisons are carried out for our case study. The results show that the proposed DMerge approach has improved performance over the SMA method, as well as any single DT model. A sensitivity analysis is also carried out to test the importance of different input variables in explaining the variability of hydropower releases.

The specific objectives and contributions of this study are (1) introducing a new predictive approach, termed DMerge, which is capable of using updated observation data to select the best ensemble members, and capable of producing a more consistent and reliable prediction than any single DT model; (2) comparing different DT methods in support of hydropower simulation using multiple information sources; and (3) evaluating the importance of different model inputs, particularly water quality and climate information, with respect to the predictability of hydropower releases. Though this study only focuses on using DT methods in hydropower scheduling problems, the proposed DMerge could be universally applicable to combine other AI and DM approaches.

In the rest of the paper, the "Methodology" section summarizes the methodologies of individual decision trees and the proposed DMerge technique. The "Study Site, Data, and Model Settings" section introduces the study site, data, and model setting. The results are shown in the "Results" section. The "Discussion" section provides discussions and limitations based on our experiments. Major findings and our conclusion are summarized in the "Conclusion" section.

Methodology

AdaBoost Decision Tree

Boosting algorithms have been introduced as a class of algorithms that convert a weak learning predictor with performance close to random guessing into a strong predictor (Freund and Schapire 1996). In other words, boosting techniques combine multiple predictors of the same type and weight the models' contribution according to their performance to achieve better accuracy (Drucker 1997; Freund and Schapire 1997). Among the boosting algorithms, the AdaBoost algorithm has gained a great deal of attention in the field of computer science. The AdaBoost algorithm adapts the contribution weighting of each predictor using the error term, which is produced by each weak learning algorithm (Freund and Schapire 1997; Wu et al. 2004). The AdaBoost algorithm also applies weights to training examples that emphasize the hard-to-predict points in the input population (Dietterich 2000a, b). At the beginning of the model training process, the training points have equal weights. During the model training process, the AdaBoost algorithm dynamically updates the training weights according to the performance of the base learning algorithms. The joint use of the AdaBoost algorithm and DT methods, i.e., AdaBoost trees (Dietterich 2000a, b; Drucker 1997; Freund 1995), has already been applied in various fields, e.g., economics (Alfaro Cortés et al. 2007; Kim and Upneja 2014), biology (Che et al. 2011), remote sensing (Briem et al. 2002; Chan and Paelinckx 2008), and physics (Roe et al. 2005). However, application of the AdaBoost algorithm in the fields of hydrology and water resources management is rarely reported.

Random Forest

The random forest (RF) algorithm (Breiman 2001) is a nonparametric, white-box ensemble tree method based on the standard CART algorithm (Breiman et al. 1984) and bagging tree algorithm (Breiman 1996; Tao et al. 2018). The RF algorithm builds an ensemble with ensemble members consisting of a collection of tree-structured predictors. These tree-structured predictors depend on independent and randomly sampled vectors with the same probability distribution for all the trees (Breiman 2001). The general procedure used in the RF algorithm starts at a random selection of predictors and a split of a bootstrap data set using a user-defined splitting criterion, such as the root-mean-square error, mean-square error, or Gini diversity index to maximize the homogeneity (or minimize the summed error rates) of splitting nodes. A similar splitting procedure is performed for each of the partitioned subsets until a user-defined stopping criterion is met. This procedure is repeated using randomly selected predictors and multiple ensemble candidate trees are built using different combinations of predictors. According to Liaw and Wiener (2002), a RF model includes the following key features: (1) all of the data are bootstrapped, (2) for each bootstrapped sample, an unpruned CART model is built using randomly selected decision variables as the best split criterion, and (3) the prediction is obtained by aggregating the prediction of all the trees built on different bootstrapped data. There have been numerous successful applications of the RF for reservoir operation (Bai et al. 2016; He et al. 2016; Kumar et al. 2013a, b; Li et al. 2010; Wei and Hsu 2009), and its performances have been proven to be superior to other DT methods, such as CART, ID3 (Quinlan 1986), ID4.5, and bagging tree (Quinlan 1996; Tao et al. 2018).

Extremely Randomized Trees

Introducing randomness in both the predictors and training data selection (bootstrap) allows the tree-growing process to enhance the model performances as shown in the RF algorithm (Breiman 2001). In recent years, a new way of adding randomness to DT methods has been developed by Geurts et al. (2006, 2007). They entitled the new algorithm *extremely randomized trees* (Extra-Trees). The Extra-Trees algorithm is similar to RF but differs in the following two aspects: (1) the Extra-Trees algorithm uses all of the training data during the tree-growing process instead of constructing different subsets of training data sets for each ensemble candidate and (2) Extra-Trees employs both randomly selected predictors and predictor values when splitting the training data,

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

whereas the RF methods find the best split (i.e., select the optimal predictor and predictor values by minimizing a classification function) among random subsets of variables. After a number of test splits on randomly selected predictor and predictor values, the split with the best skill scores is used for the next iteration (Ernst et al. 2005). Geurts et al. (2006) demonstrated that the Extra-Trees algorithm results are able to significantly decrease the final ensemble tree variance, but slightly increase the bias. Furthermore, Geurts et al. (2006) have also shown that Extra-Trees is able to eliminate variance with acceptable biases after properly adjusting the randomness level in choosing predictors and predictor values in the model setting.

Dynamic Merging Method

Inspired by the concept of particle swarm optimization (Eberhart and Kennedy 1995; Kennedy 2011; Shi 2001), a dynamic merging (DMerge) technique is proposed and demonstrated in Fig. 1 and Eq. (2). The core concept of the DMerge method is to use nonequal weights to dynamically create a single model averaging result with two specific models: one is the current best-performing model at time step t, and another is the historical best-performing model over the horizon from 0 to (t–1). As shown in Eq. (2), using the DMerge method, the prediction value for time step t consists of two parts. The first part is the prediction value from the current best-performing model at time step (t–1), which is weighted by a user-defined coefficient $C1$. The current best-performing model is defined as the model with the lowest bias between the model prediction and the observation at time step (t–1). The second part is defined as the prediction value from the historical best-performing model over the horizon from 0 to (t–1) weighted by a coefficient $C2$ [Eq. (2)]. Note that the historical best-performing model [the second part in Eq. (2)] is not necessarily identical to the current best-performing model [the first part in Eq. (2)]. For example, in Fig. 1, at time step $t = 0$, both the current and historical best models are Model 2, and the DMerge prediction will be identical to the Model 2 result at the next time step ($t = 1$). At time step $t = 1$ we predict a value at $t = 2$; the DMerge method is identical to simple model averaging as no historical information has been recorded. At time step $t = 2$, while trying to predict the value at time step $t = 3$, Model 3 becomes both the historical best and current best model, and DMerge is identical to the Model 3 prediction result at the next time step ($t = 3$). At time step $t = 3$ we want to obtain the prediction value for time step $t = 4$; the current best-performing model becomes Model 1, and both Model 1 and Model 3 results at the next time step ($t = 4$) are used to construct the DMerge result. A similar process continues at each prediction step by re-evaluating which model(s) are the current best and historical best-performing model(s) and constructing a weighted prediction value for the next time step. A generalized mathematical expression is given in the following Eqs. (1) and (2) for SMA and DMerge, respectively:

$$\mathrm{SMA}_{t=i} = \frac{\sum_1^K M_{j_{t=i}}}{K}, \quad \text{for } i = 1, 2, \ldots N \tag{1}$$

$$\mathrm{DMerge}_{t=i} = \begin{cases} \dfrac{\sum_1^K M_{j_{t=i}}}{K}, & \text{for } i = 1 \\ C1 \cdot M_{\text{Current Best}_{t=i-1}} + C2 \cdot M_{\text{Historical Best}_{t=1,2,\ldots,i-1}} & \text{for } i = 2, 3, \ldots, N, \text{and } C1 + C2 = 1 \end{cases} \tag{2}$$

where K = number of ensemble members; N = total time step for the prediction horizon; $M_{j_{t=i}}$ = prediction value obtained from the jth participating model for time step i; $C1$ and $C2$ = weighting factors for current and historical best-performing models, respectively. The constraint is $C1 + C2 = 1$.

The presented DMerge approach in Fig. 1 and Eq. (2) is based on some existing theories and assumptions used in the field of swarm intelligence (SI) (Eberhart and Kennedy 1995; Yang et al. 2018), in which the core concept and theory are to move particles toward a direction regulated by two locations: one is associated with the best fitness value so far during the search, and another is the current best location with the highest fitness value of a given objective function (Kennedy et al. 2001; Kennedy 2011). In the SI, it is also assumed that all searching particles have no information of where the global optimal is, and in our case, the observation for the next time step ($t = T + 1$) always remains unknown. According to PSO as invented by Kennedy et al. (2001), Kennedy (2011), Eberhart and Kennedy (1995), and Shi (2001), the strategy of SI, i.e., moving the existing particles toward a direction that is correlated to the locations of current and historical best particles (i.e., the individual with the best fitness), is proven as an effective approach to find global optimal solutions for complex optimization problems. In other words, the direction of a particle for the next time step is determined by both (1) the current best location among all participating members and (2) the historical best location found during the evolution of the population. It is worth mentioning that at the initial prediction time step ($t = 0$) the historical best is not available because no prediction has been made before time step 0 (Fig. 1). Therefore, the DMerge method will have a value identical to that of the SMA method. As time goes on ($t = 1, 2, 3 \ldots, N - 1$, N), the DMerge method is able to learn which predictive model has consistently good performance over all prediction time steps and uses this model to weight against the current best model.

Study Site, Data, and Model Settings

Study Site

Shasta Lake in northern California is selected as our case study (Fig. 2). The Shasta Lake dam is the ninth tallest in the United States and the biggest hydropower dam in California. The major functionalities of the Shasta Lake dam are hydropower, irrigation,

J. Water Resour. Plann. Manage., 2020, 146(2): 04019072

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

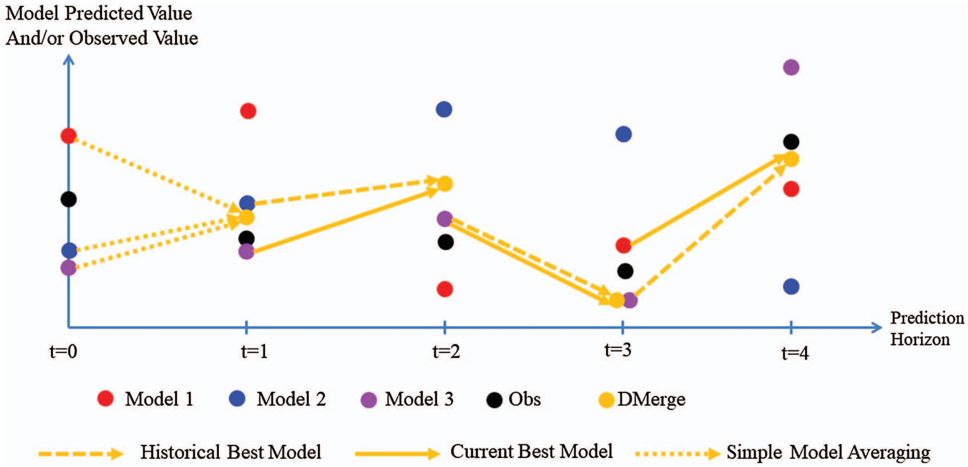

Fig. 1. (Color) Concept of the DMerge approach.

and water supply for the Central Valley Project in California. The capacity of the lake is around 5.6×10^9 m^3 at full pool level, and the lake has an elevation of 325 m above sea level. The total hydropower generation capability for the Shasta Dam is about 710 MW, which is more than 30% of the total generation capacity of the California Central Valley Project. Besides the important role of hydropower generation, the Shasta Dam also provides the flows for fish migration and spawning in the Sacramento River. There are many fish species living in Shasta Lake, such as Chinook salmon and steelhead trout. To protect those fish populations, strict regulations and hydropower releasing rules have been enforced by federal endangered species act (ESA) from 1989 (EPA 2013b). California state agencies have carried out water quality monitoring at hydropower facilities at the Shasta Dam to monitor the water temperature, dissolved oxygen, and water turbidity that are required by law to stay within the tolerance limits for spawning salmonids (DOE 2014b; EPA 2013b; FERC 2011; Kimmell and Veil 2009; Veil et al. 1993). The selection of the study site is based on the data availability from the publicly accessible source in California. In the California Data Exchange Center (CDEC), Shasta Lake has the most complete set of data (Table 1) as well as the longest data records as compared to many other reservoirs in California.

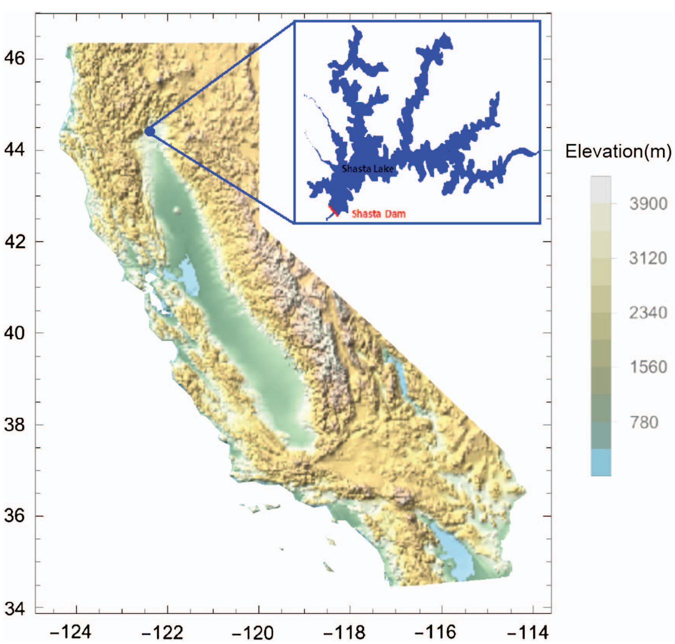

Fig. 2. (Color) Location of the Shasta Dam located in northern California. (The map of California was made with Wolfram Mathematica using DEM files courtesy of webGIS.)

Data

There are 26 different types of data used for model training and validation. The selection of input variables is based on some findings in an earlier study (Yang et al. 2017a), as well as the correlation coefficient between raw model inputs and hydropower releases. More justifications and discussion are provided later in the "Input Variable Cross-Correlation" and "Limitations and Future Work" sections of this paper. Daily hydrological data including precipitation, lake evaporation, and inflow amounts were retrieved from CDEC. CDEC is the official data portal for the state and federal water resource operating agencies in California. The daily observation on lake storage volume and the conservation pool level for the Shasta Lake were also obtained form CDEC. The duration of observation data is five years from January 1, 2010, to December 31, 2015. The early data before 2010 were not used in this study due to significant missing and unreliable records.

Also, 17 climate indices were retrieved from the National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA-ESRL) that represent different climate and atmospheric activities, including teleconnections, atmosphere, ENSO, and the variations of the Pacific and Atlantic sea surface temperatures. Table 1 lists detailed information about the data used as model inputs. The model outputs are the daily hydropower releases.

The raw data obtained are in different temporal resolutions. To obtain daily hydropower releases, the model inputs were either aggregated or disaggregated into a daily time step so that the model holds a constant input-output data resolution. For instance, hourly data were accumulated into daily, while climate information with a monthly time step resulted in daily inputs with the same values throughout a month. Similar ways of data handling are also presented by Naeini et al. (2018) and Zhang et al. (2018).

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

Table 1. Information of selected model input data

No.	Group	Short name	Resolution	Description	Source/station
1	Hydrological condition	Inflow	Daily	Inflow amount	CDEC/SHA
2		Precipitation	Daily	Point accumulated precipitation	CDEC/SHA
3		Evaporation	Daily	Computed accumulated lake evaporation	CDEC/SHA
4	State and regulation	Storage	Daily	Measured storage volume	CDEC/SHA
5		Regulation	Daily	Calculated water level above reservoir conservative pool	CDEC/SHA
6	Water quality	Water temperature	Hourly	Hourly measured water temperature in degrees	CDEC/SHD
7		Water dissolved oxygen	Hourly	Hourly measured dissolved oxygen in ml/L or ppm	CDEC/SHD
8		Water turbidity	Hourly	Hourly measured dissolved oxygen in nephelometric turbidity unit (NTU)	CDEC/SHD
9	Climate index	PNA	Monthly	Pacific North American index	NOAA Climate Prediction Center (CPC)
10		WP	Monthly	Western Pacific index	NOAA Climate Prediction Center (CPC)
11		NAO	Monthly	North Atlantic oscillation	NOAA Climate Prediction Center (CPC)
12		SOI	Monthly	Southern oscillation index	NOAA Climate Prediction Center (CPC)
13		WHWP	Monthly	Western hemisphere warm pool	Wang and Enfield (2001)
14		ONI	Monthly	Oceanic Nino index	NOAA Climate Prediction Center (CPC)
15		MEI	Monthly	Multivariate ENSO index	Wolter and Timlin (1998)
16		Nino1 + 2	Monthly	Extreme Eastern Tropical Pacific SST (0–10S, 90W–80W)	CPC
17		Nino 3	Monthly	Eastern Tropical Pacific SST (5N–5S, 150W–90W)	CPC
18		Nino 34	Monthly	East Central Tropical Pacific SST (5N–5S, 170–120W)	CPC
19		Nino 4	Monthly	Central Tropical Pacific SST (5N–5S, 160E–150W)	CPC
20		PDO	Monthly	Pacific decadal oscillation	Zhang et al. (1997)
21		TNI	Monthly	Indices of El Niño evolution	Trenberth and Stepaniak (2001)
22		AO	Monthly	First leading mode from the EOF analysis of monthly mean height anomalies	CPC
23		QBO	Monthly	Quasi-biennial oscillation	NCEP/NCAR Reanalysis
24		CENSO	Monthly	Bivariate ENSO time series	Standardized SOI and standardized Nino3.4 SST Time series
25		EPO	Monthly	East Pacific/North Pacific oscillation	NOAA Climate Prediction Center (CPC)
26		Seasonality	Monthly	Month of a year	Calendar

Model Settings

In this study, we use the period of 2010–2013 as calibration and the period of 2014–2015 as validation as shown in Fig. 3. The CDEC data portal contains longer records of data back to 2007 for Shasta Lake. However, many of the data listed in Table 2 have significant missing and unreliable records. To test the model performance, a short period of data would reveal the models' capability in capturing the relationships among multiple variables. In real operation and practical uses of the proposed statistical models, it is suggested to use a complete data set and the longest possible data records to train a statistical model so that model performance can be maximized.

To compare the sensitivity of model inputs, three different test scenarios were designed, as shown in Fig. 3. The influence of various categories of data on explaining the behaviors of model outputs during the validation period is investigated. Scenario 1 uses all the decision variables listed in Table 1 as model inputs. Scenario 2 intentionally removes the water quality decision variables from the model inputs. The purpose is to examine the performance impact of the water quality input data. Finally, under Scenario 3, the climate indices are further excluded from the model inputs and only hydrological, state, and regulation variables are used as model inputs. The purpose of Scenario 3 is to test the impacts of climate indices on hydropower releases like the examination of water quality impacts under Scenario 2.

Under different scenarios, the decision tree models (Fig. 3) have identical hyperparameter settings: the maximum tree depth is set to 50, number of estimators for ensemble trees is set to 1,000 (which ensures a sufficient number of learners being built), the minimum number of data points in each leaf is set to 2, the maximum number of features is set to the square root of the number of decision variables, and the root-mean-square error is used as a splitting criterion for the AdaBoost tree, the RF algorithm, and the ERT algorithm. According to many default settings used with PSO algorithms, the coefficient for the current best member [$C1$ in Eq. (2)] in the

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

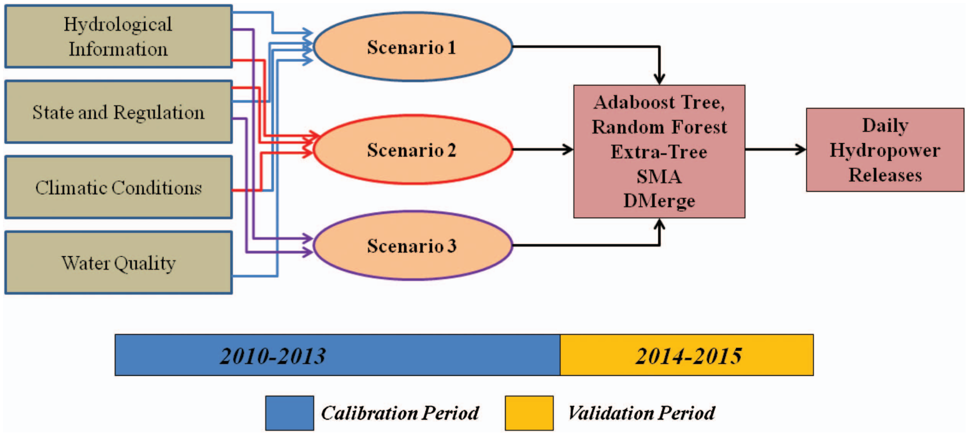

Fig. 3. (Color) Designed scenarios and modeling structure.

Table 2. The statistics between simulated and observed discharges using AdaBoost, RF, Extra-Trees, SMA, and DMerge under different scenarios

	CORR		RMSE (m^3/s)		NSE		KGE	
Models/statistics	Cal	Val	Cal	Val	Cal	Val	Cal	Val
				Scenario 1 (all inputs)				
AdaBoost tree	**0.998**	0.951	**235.905**	668.021	**0.995**	0.872	0.940	0.902
Random forest	0.985	0.904	614.384	829.026	0.969	0.803	0.916	0.863
Extra-Trees	0.962	0.893	981.440	860.191	0.920	0.787	0.806	0.779
SMA	0.988	0.940	558.293	673.987	0.974	0.870	0.914	0.884
DMerge	0.993	**0.959**	436.571	**551.169**	0.984	**0.913**	**0.963**	**0.934**
				Scenario 2 (hydrological, state and regulation, and climate inputs)				
AdaBoost tree	**0.996**	0.911	**299.659**	870.453	**0.993**	0.782	**0.944**	0.874
Random forest	0.973	0.904	812.221	809.635	0.945	0.812	0.861	0.831
Extra-Trees	0.941	0.863	1,196.472	954.575	0.881	0.738	0.763	0.732
SMA	0.981	0.914	704.317	779.693	0.959	0.825	0.885	0.852
DMerge	0.988	**0.939**	555.099	**650.794**	0.974	**0.878**	0.917	**0.895**
				Scenario 3 (hydrological and state and regulation inputs only)				
AdaBoost tree	**0.988**	0.831	**529.992**	1,285.775	**0.977**	0.525	0.894	0.765
Random forest	0.954	0.836	1,091.366	1,192.326	0.901	0.592	0.940	0.798
Extra-Trees	0.874	0.786	1,774.608	1,333.839	0.737	0.489	0.867	0.743
SMA	0.964	0.826	1,031.162	1,238.712	0.911	0.560	0.954	0.785
DMerge	0.978	**0.848**	822.976	**1,153.355**	0.943	**0.618**	**0.963**	**0.809**

Note: The bolded numbers are the best statistics among all methods in calibration and validation periods, respectively.

DMerge method is set to 0.7, and the coefficient for the historical best [$C2$ in Eq. (2)] is set to 0.3. The coefficients are determined by some initial tests by authors, in which the performances are stable and optimal in the training sets. More discussion of the systematic derivation of the coefficients will be provided in the "Limitations and Future Work" section.

Results

Model Performances

Four types of statistical measures were selected in this study: the correlation coefficient (CORR), the root-mean-square error (RMSE), the Nash-Sutcliffe model efficiency coefficient (NSE) (Nash and Sutcliffe 1970), and the Kling-Gupta efficiency (KGE) from Gupta et al. (2009), as shown in the following Eqs. (3)–(6), respectively:

$$\mathrm{CORR} = \frac{\sum_{i=0}^{n}((Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim},i})(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs},i}))}{\sqrt{\sum_{i=0}^{n}(Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim},i})^2 \sum_{i=0}^{n}(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs},i})^2}} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=0}^{n}(Q_{\mathrm{sim},i} - Q_{\mathrm{obs},i})^2/n} \tag{4}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=0}^{n}(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i})^2}{\sum_{i=0}^{n}(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs},i})^2} \tag{5}$$

$$\mathrm{KGE} = 1 - \sqrt{(1 - \mathrm{CORR})^2 + (1 - \sigma_s/\sigma_o)^2 + (1 - \mu_s/\mu_o)^2} \tag{6}$$

where $Q_{\mathrm{obs},i}$ and $Q_{\mathrm{sim},i}$ = observed and simulated hydropower discharges at time step t, respectively; $\bar{Q}_{\mathrm{obs},i}$ and $\bar{Q}_{\mathrm{sim},i}$ = means of the observed and simulated values, respectively; n = total number of data points; μ_s and σ_s = mean and standard deviation of the simulated discharges, respectively; and μ_o and σ_o = mean and standard deviation of the observed hydropower releases, respectively. The KGE values range from negative infinity to 1 and a KGE value equal to 1 suggests an ideal case in which simulation matches observation (Gupta et al. 2009). According to Moriasi et al. (2007), for monthly

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

simulation a KGE value or a NSE value larger than 0.50 can be considered satisfactory. In Table 2, the CORR, RMSE, NSE, and KGE values for both calibration and validation periods are calculated and compared using the simulation results obtained from the DT models, the SMA method, and the proposed DMerge method under different simulation scenarios. In Table 2, the bolded values indicate the best statistics under each scenario and period.

According to Table 2, the DMerge method produces promising results with the best CORR, NSE, KGE, and RMSE values when compared to other algorithms during the validation periods. The NSE and KGE values from DMerge are consistently higher than 0.5 under all scenarios. The AdaBoost tree algorithm demonstrates the highest CORR, NSE, and KGE and the lowest RMSE as compared to the RF and Extra-Trees algorithms during the model calibration period, whereas the performances during the validation period are significantly worse than the DMerge. As compared among SMA, AdaBoost tree, RF, and Extra-Trees algorithm, the SMA method is slightly better than other individual models as the averaging of different model outputs cancels the biases for most of the time steps, particularly for the cases where some models are overestimating and others are underestimating. It is worth mentioning that the DMerge method is able to generate even better statistics than each individual method and the SMA method, and therefore, it is believed to be the best-performing model in all validation periods (Table 2).

Under Scenario 1, the best statistical measures are produced by the DMerge method, which are 0.959, 0.913, and 0.934 during the validation period for CORR, NSE, and KGE, respectively. The worst-performing model is the Extra-Trees algorithm, of which the CORR, NSE, and KGE values are 0.893, 0.787, and 0.779 for CORR, NSE, and KGE during the validation period, respectively. The SMA method is able to produce better statistical measures as compared to the Extra-Trees method and is comparable to random forest results. The calculated CORR, NSE, and KGE by the SMA method are 0.940, 0.870, and 0.884, respectively. Under Scenarios 2 and 3, similar results are observed with respect to the ranks of model performances. The ranking of model performances can be summarized as DMerge > AdaBoost tree > SMA > RF > Extra-Trees algorithm.

According to Table 2, the model performances shown tend to decrease as fewer model inputs are used. For example, the CORR value of the DMerge method decreases from 0.959 (Scenario 1) to 0.939 (Scenario 2), and finally to 0.848 (Scenario 3). The NSE value of the DMerge method decreases from 0.913 (Scenario 1) to 0.878 (Scenario 2), and further to 0.618 (Scenario 3). The KGE value of the DMerge method decreases from 0.934 (Scenario 1) to 0.895 (Scenario 2), and further to 0.809 (Scenario 3). Similar decreasing patterns are shown in the calculated statistics (Table 2) from the AdaBoost tree, RF, and Extra-Trees algorithms and the SMA method. The model performance rankings are the same under Scenarios 1 and 2 (DMerge > AdaBoost tree > SMA > RF > Extra-Trees algorithms), while under Scenario 3 the AdaBoost tree, RF, and Extra-Trees algorithms show similar results and the model ranking varies slightly with respect to different statistical measures.

Fig. 4 shows the comparison between simulated and observed daily hydropower discharge under different scenarios during the validation period (January 1, 2014, to December 31, 2015). In this figure, the observations are shown as black dots, and the red, blue, pink, yellow, and green lines represent the simulated discharges by AdaBoost, RF, Extra-Trees, SMA, and the proposed DMerge method, respectively. According to Fig. 4(a), all algorithms were able to produce reasonable simulations with a good match to observation. However, during February and March of 2015 in Fig. 4(a), only the DMerge and AdaBoost algorithms were able to capture the sudden hydropower reduction, while other models overestimated the daily hydropower releases during this period. Another supporting case is shown in Fig. 4(a) during the period between April and May of 2015, where the high hydropower releases from Shasta Lake were significantly underestimated by the RF, Extra-Trees, and SMA methods.

Note that in Fig. 4(a), starting around June 2015, the AdaBoost tree algorithm started to overestimate the hydropower releases, while the proposed DMerge method could identify that the AdaBoost method was no longer a current best algorithm and the results from the other two methods become relatively better than those derived by the AdaBoost method. Similar cases also occurred in the winter of 2015 [Fig. 4(a)]. According to Fig. 4(a), during the winter of 2015, RF and Extra-Trees algorithm began to overestimate the hydropower releases, while the DMerge and AdaBoost tree methods demonstrated a good match with respect to observations. Those above-mentioned cases in Fig. 4(a) are direct evidence that the DMerge approach can produce better predictions by dynamically identifying the best-performing model and merging the corresponding model prediction sets. Another interesting phenomenon is shown during the period of June to October of 2015 [Fig. 4(b)], in which the AdaBoost tree algorithm significantly overestimated the hydropower releases. However, the DMerge method was able to capture the variation of hydropower releases and retains similar predictive performances to RF and the Extra-Trees algorithms, which

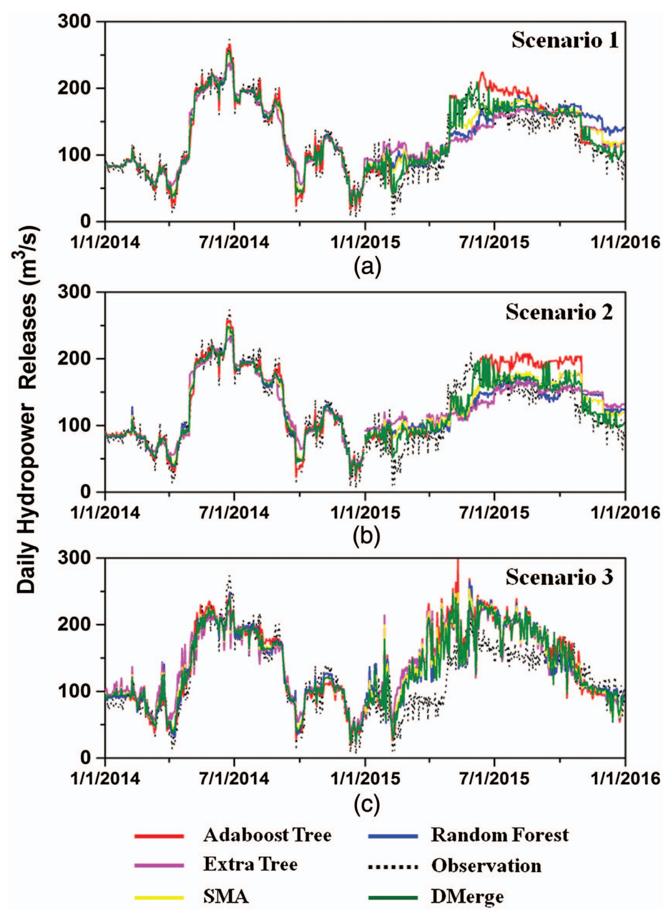

Fig. 4. (Color) Comparison of the observation (black dots) and the prediction and using AdaBoost (red), RF (blue), Extra-Trees (pink), SMA (yellow), and DMerge (green), under (a) Scenario 1; (b) Scenario 2; and (c) Scenario 3 during the validation period (January 1, 2014, to December 31, 2015).

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

are performing more satisfactorily than the AdaBoost tree algorithm during this particular prediction period. Similarly, in Fig. 4(b), after October 2015, RF and Extra-Trees algorithms began to overestimate the hydropower releases, and the AdaBoost tree algorithm became a better prediction model. The proposed DMerge method identified the performance changes of different candidate models and retains the best-performing models during October to December 2015. Under Scenario 3, all models were unable to provide a reasonably good prediction after April 2015 because only hydrological and state and regulation variables were employed as model inputs. The reduced model inputs resulted in the deterioration of model predictive performances. More in-depth discussions about the usefulness of information and variables are provided in the next section.

Decision Variable Contribution

In this section, we analyze the sensitivity to model inputs. The goal is to provide a mathematical evaluation of the decision variables' contribution to the derivation of the simulation results as shown in Fig. 4. To evaluate the contributions of model inputs, we use the reduction of the percentage of mean-square error (MSE) for each decision variable in the tree-growing process. According to Breiman (1996, 2001), Hancock et al. (2005), and Liaw and Wiener (2002), the percentage of MSE reduction is a standard indicator to measure the functioning of each split in the tree-growing process. For each split, the differences in MSE with and without the split are first achieved. Then, the MSE reduction is summed up by each decision variable for all the splits in the tree. Last, the summed MSE reduction for each decision variable is normalized in a manner that a higher percentage of MSE reduction indicates a more efficient split using such model input. The sum of percentages of MSE reduction for all decision variables will always equal 1. A lower MSE reduction percentage of a decision variable would indicate that the decision variable is less important compared to others. Theoretically, a zero percentage of MSE reduction for a decision variable suggests that this decision variable is not used in the tree-growing process and has nearly no predictability of the target variable (i.e., hydropower releases in our case study).

The calculated percentages of MSE reduction for all decision variables are shown in Fig. 5. According to Fig. 5, the impact of hydropower generation on the ecosystem is not neglectable in reservoir decision making. In addition, in all cases (Scenarios 1–3), the storage and inflow are two of the most important decision variables relevant to the patterns of hydropower releases. This is because of the facts that most of reservoirs in the United States are controlled by rule curves, which essentially describe the relationship between storage and allowable releases, and the reservoir inflow is the main source to replenish the reservoir storage as compared to other water supplies to the reservoir.

Input Variable Cross-Correlation

It is highly possible that the employed model inputs are collinear and correlated. In order to examine the correlation among different inputs, as well as the relationship between raw model inputs and hydropower releases, we conducted the following experiment. In Fig. 6, we map out the correlation coefficients of each model input against one another [Fig. 6(a)] and also calculated the raw correlation coefficients between model inputs against the target variable [Fig. 6(b)]. The diagonal line in Fig. 6(a) indicates correlation coefficients values of 1, i.e., the self-correlation of each model input with respect to itself. The numbers on the x- and y-axes correspond to the number of model inputs listed in Table 2. The correlation coefficients in Fig. 6 are derived using all of the data sets (2010–2015). The authors have examined the correlation coefficient values within different splits of data in calibration (2010–2013), validation (2014–2015), and the entire set (2010–2015), in which the correlation coefficients are consistent with minimal variations among different periods.

Discussion

Model Performances and Suitability

The evaluation metric obtained from different models (Table 2) shows that all methods are able to reach satisfactory simulation results for most of the cases except for Scenario 3. Table 2 also shows that the AdaBoost algorithm tends to outperform the RF and Extra-Trees algorithm when there are extended predictors used for model training (e.g., Scenarios 1 and 2). The model performances of AdaBoost, RF, and the Extra-Trees algorithm are similar to one another under Scenario 3. Nevertheless, the newly developed DMerge method is able to increase the model prediction accuracy under each scenario consistently, and the improvement of statistics is significant according to results in Table 2. With respect to the NSE and KGE values during the validation period, the improvement allowed by the DMerge method is about 15%–30%.

The DMerge method has shown advantages because of its self-learning mechanism, which iteratively evaluates the best-performing models and weights the current best-performing model against the model that outperforms others for the maximum number of historical time steps. The results in Figs. 4(a and b) along with our analysis in the results section indicate that the DMerge method is able to capture dynamic changes of model performance and adjust the ensemble prediction. Therefore, it provides better results than the SMA method or any single model. With regards to the SMA approach, the model weights remain constant over time, which inevitably produces an ensemble that lies between the best prediction and the worst prediction at each time step as shown in Figs. 4(a and b). The proposed DMerge method is able to give improved statistics over the SMA method in both calibration and validation periods. For example, for the DMerge method, the NSE and KGE values under Scenario 1 (validation period) are 0.931 and 0.934, respectively. The NSE and KGE values produced by DMerge under Scenario 2 (validation period) are 0.878 and 0.895, respectively. For the SMA method, the NSE and KGE values under Scenario 1 (validation period) are 0.870 and 0.884, respectively. The NSE and KGE values produced by the SMA method under Scenario 2 (validation period) are 0.825 and 0.852, respectively. As compared to the above-summarized NSE and KGE values from both DMerge and SMA, the DMerge method will produce consistently better predictions than the SMA method, which is employed as the baseline reference.

In addition, the model performance with ensemble tree methods, especially the extremely randomized trees method, suffers from high dimensional data sets, since the bias increases as the randomization level increases. The proposed DMerge method is not a randomization process; rather, it is a model-averaging tool to reduce biases associated with an individual model. Considering a situation when one single model, i.e., extremely randomized trees, processes a greater amount of bias than others, the DMerge method identifies that this model is not the current best model, and the model with the lowest bias is used to develop the hybrid prediction for next time step. The logic is explained in the methodology section. The increases of model bias are likely to happen when the dimensionality of the data set becomes larger. Some standard computer

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

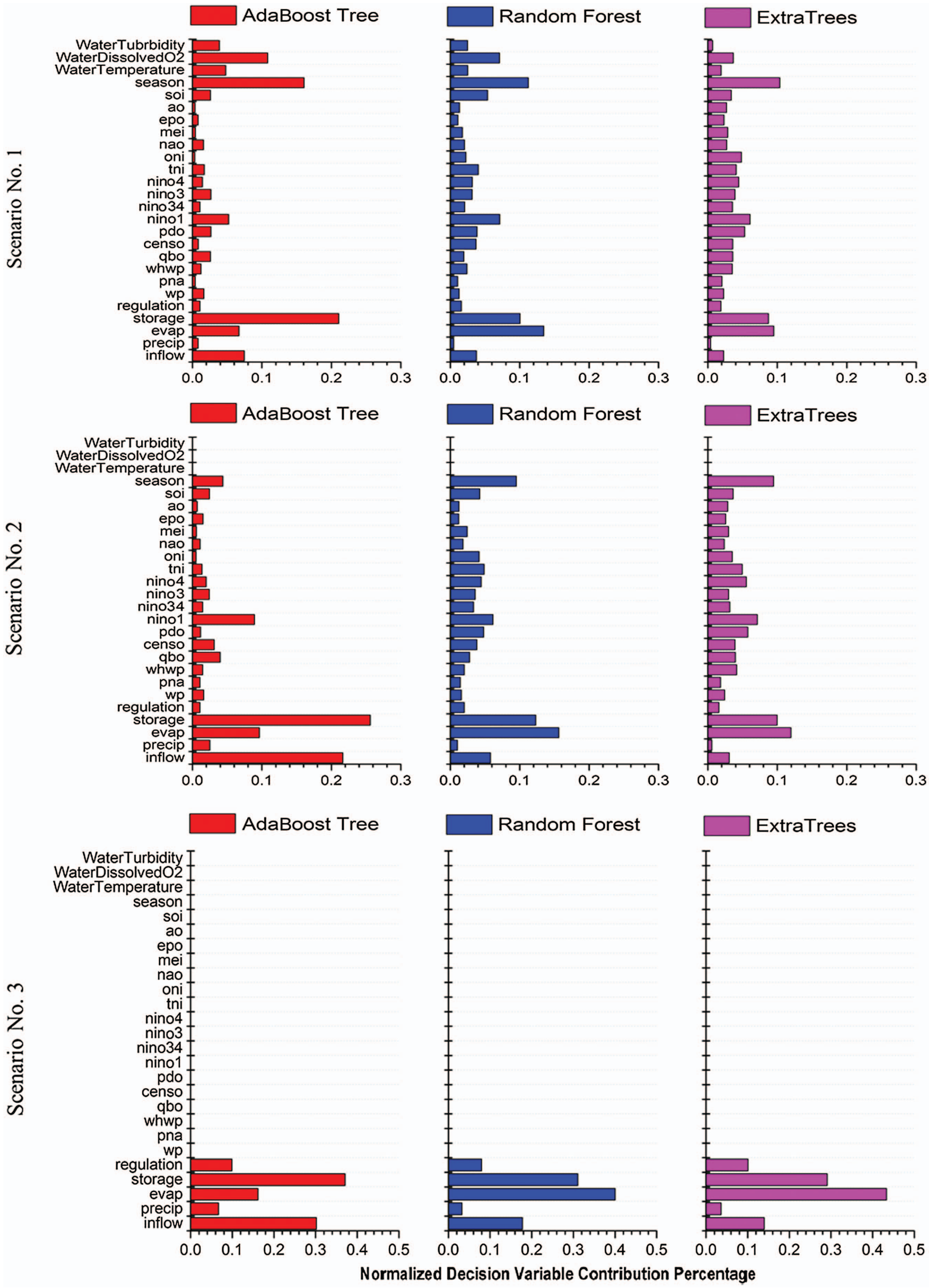

Fig. 5. (Color) Normalized MSE reduction percentages for all decision variables under different scenarios.

J. Water Resour. Plann. Manage., 2020, 146(2): 04019072

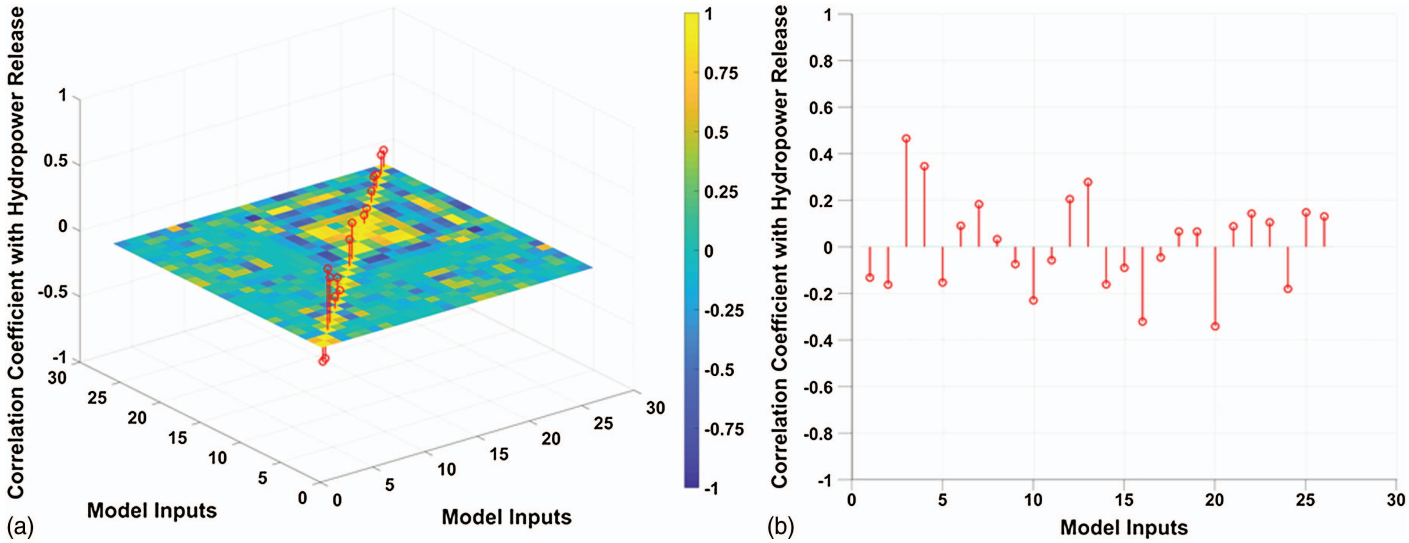

Fig. 6. (Color) (a) Correlation coefficient among model inputs; and (b) correlation coefficient between each model input against hydropower release.

science benchmark data sets with high dimensionality could be further tested, for example, the Machine Learning Repository from UC Irvine, which is heavily used for artificial neural network algorithm development (Yang et al. 2017b).

It is also worth mentioning that the proposed DMerge method can be used for other types of predictive models, such as artificial neural networks, support vector machines, or any linear regression models. The only assumption for DMerge is that users will have different options of data-driven models for time-dependent prediction problems and different predictive models may exhibit different performances over time. As new data gradually become available, the DMerge method can provide a means of injecting updated information into the step-by-step prediction process.

Impacts of Different Decision Variables

According to Fig. 5, Nino 1 is identified as the most important climate predictor among climate information for the hydropower releases for Shasta Lake. This finding is similar to the conclusion from Gutiérrez and Dracup (2001), in which reservoir discharges in the Columbia River basin (western US) were highly correlated to the El Niño indicators. Garen (1993) claimed that many climate indicators, including El Niño and PDO, were useful information for predicting the water supply conditions in the Western US. Similar conclusions were drawn from the study by Montoya et al. (2014) and Pagano and Garen (2003), in which the large-scale climate indicators were found to be associated with April 1 snow water equivalent in the Sierra Nevada of California. In another study, Lü et al. (2011) identified that Nino1, 2, and Nino3.4 were representative of the streamflow variation of the Yellow River basin of China. In addition, many previous studies pointed out the connections between the El Niño phenomenon and streamflow in the United States (Cayan et al. 1999; Cayan and Webb 1993; Dettinger et al. 2011; Redmond and Koch 1991; Miao et al. 2010). However, the means to mathematically quantify the impacts of a specific climate index on hydropower release management have rarely been reported. In this study, we tackle this issue by using the reduction of MSE to evaluate the relative impacts of climate indices on the hydropower releases from Shasta Lake as shown in Fig. 5.

It is worth mentioning that, besides the El Niño indices, seasonality and reservoir inflows also demonstrate strong correlations with hydropower releases (Fig. 6). In California, the water supply conditions strongly vary from season to season. Winter precipitation is the primary driving force of the hydrological cycle. However, as a certain portion of winter precipitation is stored as snowpack along the Sierra Nevada, the impacts of precipitation on hydropower releases is limited (Fig. 5). Different from precipitation, the reservoir inflow, especially during runoff season (April 1–August 31 each year), can play an important role in explaining the variation of hydropower releases, accounting for about 15%–30% for the best-performing DT model under Scenarios 2 and 3. The reason is that when reservoir inflows are expected to be high, reservoirs are facing the risk of potential flooding. As a result, water in the storage flood control pool or the conservation pool needs to be released either from a spillway or hydropower turbines. In addition, this study uses the daily inflow and storage volume as predictors, while monthly inflow accumulation and storage could be better predictors for monthly hydropower approximation. As the current simulation is based on the daily scale, the variations exhibited at daily time steps could provide better predictability.

As we compare the statistical metrics among different simulation scenarios (Table 2), another interesting finding is that the statistical performances of all models become worse as the number of model inputs decreases from Scenario 1 to Scenario 3. After purposely removing water quality indices from Scenario 1, the statistics all decrease to some extent as shown in Table 2. Similarly, the deteriorations of statistical measures are also observed under Scenario 3 when the climate indices from the set of model inputs under Scenario 2 are intentionally excluded. This indicates that both water quality and climate conditions are useful information to determine hydropower releases of the Shasta Dam. The high impacts of water quality indicators on hydropower releases are due to ecosystem constraints downstream of the hydropower turbines. After the water goes through the hydropower turbines, the velocity potential of water increases due to the decrease of gravity potentials and increased heat of the released water (DOE 2014b). The heated water has less capacity to hold dissolved oxygen, which can create dead zones in water bodies (DOE 2014a; EPA 2013a, b). All the effects have detrimental impacts on the ecosystem and living conditions of spawning fish. According to Fig. 5, in general, the dissolved oxygen has a higher sensitivity to hydropower releases, followed by the water temperature and water turbidity indicators. Therefore, including those water quality variables in a

J. Water Resour. Plann. Manage., 2020, 146(2): 04019072

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

multiobjective decision support model is essential, because significant deterioration of all statistical measures are observed after removing those variables from the modeling approach employed in this paper. The scenario experiments carried out in this paper (Table 2 and Fig. 4) suggest that water quality information should be considered in the hydropower decision-making process. Otherwise, the simulation accuracy will be compromised. Similarly, the inclusion of climate information also increases the simulation accuracy of hydropower release as compared to the statistics between Scenario 2 and Scenario 1. Furthermore, as demonstrated in this paper, the data-driven models (DT methods) are able to incorporate a large number as well as various types of decision variables into a hydropower modeling process. This unique feature of a data-driven model provides the flexibility of adding additional ancillary information to various modeling problems and allows decision makers and operators to carry out sensitivity analyses on different types of information that are tailored to reservoir operations.

Reservoir Operating Rules

In California, most of the reservoirs are strictly controlled and operated by the operation manuals designed by the US Army Corps of Engineers (USACE) (USACE 2016). These operation manuals are also called rating/rule curves, which describe the empirically preferred reservoir storage level and discharge relationships that satisfy essential engineering constraints and flood control requirements (Louks and Sigvaldason 1981). In California, the reservoir discharges, including hydropower releases, must obey these reservoir rule curves, and only in special cases can the discharge deviate from the rule curves (Kelly 1986; Yang et al. 2015). The primary purpose of these rule curves is to regulate reservoir discharges so that the risk of flooding and the possibility of dam seepage can be controlled within certain allowable ranges. As shown in Fig. 5, the storage variable has consistently high importance with respect to explaining the variation of hydropower releases, which is in line with the concept of reservoir rule curves. According to the best-performing DT model under each scenario (Fig. 5), the contribution of storage level on hydropower releases ranges from more than 20%–50% as compared to other decision variables.

In California, the snow water equivalent (SWE) content, or snow depth information, also has high importance for reservoir operation and storage planning. In the current model input set, we only included seasonality as one of the inputs. The use of SWE content needs further investigation in terms of the delay effects of snow melting in April and solid precipitation accumulation in mountainous areas. The impact of SWE on hydropower is suggested to be considered over a longer period time in a monthly or yearly time step in which the delay effects of snow melting could be better captured along the seasonal variation of hydropower releases. Daily hydropower already contains a high level of randomness and noise due to the complex decision-making process constrained by multiple factors.

Selections of Model Inputs and Sensitivity Analysis

In this study, we use 26 decision variables (Table 1), which includes hydrological information, climate indices, and reservoir operational data, as well as some water quality indicators related to hydropower generation. According to Fig. 6(a), it is obvious that some model inputs are correlated as shown with bright yellow (positive correlation) or dark blue (negative correlation) cells. The high correlation values are most likely to appear for Model inputs 15–24 [Fig. 6(a)], which contain the El Niño indices, CENSO, PDO, AO, and other indices (Table 1). Most of the correlated climate indices are related to direct measures of SST, El Niño, and ENSO. Therefore, high correlations among the raw model inputs are unavoidable. Though some model inputs are correlated or collinear, we also found in Fig. 6(b) that the correlation coefficients between individual raw model inputs and hydropower releases are relatively low, ranging from −0.4 to 0.4. According to Fig. 6(b), among the Model inputs 15–24, the variables with the highest correlation coefficient values are 16 (Nino1) and 20 (CENSO) with correlation coefficient values about −0.4 to −0.3. This indicates that a single climate index may not lead to confident predictability of reservoir releases and the uses of multiple climate indices along with other ancillary information together could potentially result in better predictability of reservoir operation.

From Fig. 6(a), we also observe that many model inputs have some level of correlations with other predictors in the same group. Though the raw hydrological information and reservoir storage (Model inputs 1–5) have relatively higher correlation coefficients than other individual predictors; noise and the equifinality issue may exist. Many raw model inputs contain seasonality and exhibit high correlation against the last model input. Further study is suggested to extract eigenvalues of multidimensional data instead of using raw inputs. Principal component analysis could be used to create noncorrelated model input sets (Chu et al. 2014; Naeini et al. 2018). Data preprocessing and dimensionality reduction techniques could potentially promote models' performance. However, the physical interpretation of how raw model inputs influence the target variable could be compromised.

In addition, this study selects the model inputs based on some prior studies on climate indices influencing reservoir operation (Kim et al. 2019; Yang et al. 2017a). The inclusion of water quality information in hydropower release simulations is rarely investigated in the literature, partially because many presented case studies do not have ecosystem management or fish species protection concerns along with hydropower modeling. However, in Fig. 6(b), it is obvious that the correlation coefficient of water turbidity, temperature, and dissolved oxygen against hydropower releases are not negligible (about 0.2–0.3), which are comparable to many other raw model inputs. Hydropower production will affect the downstream water temperature, turbidity, and dissolved oxygen, all of which are unfavorable or even detrimental to fish spawning and aquatic wildlife survival (Conklin and Young 2008; Conklin et al. 2007; DOE 2014b). Power plants (including hydropower plants) that withdraw water and then release it back into the environment at an elevated temperature must comply with temperature limits under the National Pollutant Discharge Elimination System (NPDES) program (Veil et al. 1993) authorized by Section 316(a) of the US Federal Clean Water Act (CWA), although water temperature is not included in the US Environmental Protection Agency (EPA) list of priority pollutants (EPA 2013a, b). At higher temperatures of intake water, hydropower plants may reduce electricity production to meet the discharge temperature limit or risk paying fines (Kimmell and Veil 2009). In California, strict state regulations and laws protecting fish species and ecosystems that regulate the allowable levels of water quality for hydropower dams have been established for years (CDWR 2013, 2015; Conklin and Young 2008; Conklin et al. 2007). Those factors mentioned above are all important for modeling hydropower decision making. However, water quality and ecosystem concerns belong to a different category of information besides hydrology and climate forcing. There are neither direct links among those variables nor well-understood physical processes that relate those decision variables to hydropower releases. This study tries to connect the neglected water quality with hydropower simulation in a statistical modeling

J. Water Resour. Plann. Manage., 2020, 146(2): 04019072

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

framework in support of an integrated water resources management for the Shasta Lake operation.

Limitations and Future Work

In the presented experiments, the model input data are not lagged, which means the hydropower releases at any time step depend on the corresponding predictors at the same current time step. It is inevitable that some model inputs could have a certain level of autocorrelation. However, as the temporal resolution of target values (hydropower releases) is at a daily scale, the uses of lagged model inputs could be arguable. For example, the water storage is a result of the water balance of the total reservoir inflow, outflow (including hydropower releases), and previous time-step water storage. The lagged water storage (state variable) is partially dependent on the hydropower releases in the previous time step. A well-trained regression model would identify the linear relationship between the target variable and its dependent decision variable, instead of a set of independent decision variables. In our case, the DT models capture the hydropower releases that have a certain level of autocorrelation at a daily scale with respect to the lagged storage volume. Instead of lagging reservoir storage, we also recommend that climate conditions be intentionally lagged for several steps as those large-scale teleconnections and ENSO phase changes could happen at the temporal scale of months to years. However, using the lagged climate indices should be carried out with a longer data set than the current study, because some climate indices, such as PDO, have annual time scale duration when changing phases. The impacts of those climate indices become more significant as long-term observation of reservoir operation data and water quality data become available to use in the employed modeling framework in this paper. Nevertheless, with the current experiment settings, results show that hydropower releases from the Shasta Dam can still be well simulated (Fig. 4) and satisfactory statistics are achievable (Table 2). Furthermore, as demonstrated in Fig. 6, many climate indices have similar contribution percentages and cross-correlation may exist within those model inputs. Similar conclusions are also drawn from a prior study (Yang et al. 2017a). However, the cross-correlation among decision variables does not undermine the findings of this study. All DT methods employed in this study successfully and consistently identify that the Nino 1 index is the most useful indicator from the climate variable category, and this finding is also in line with many previous studies as mentioned above. However, this phenomenon can also be explained by the fact that the reservoir inflows correlate well with the variation of El Niño indices for a short period of time at daily temporal scales.

Besides the lag effects of climate indices, water quality predictors are also not lagged in this study. The current simulations are conducted on the daily scale, while the water quality indicators are observed at an hourly scale. Based on some personal communications with some water managers and dam operators, when ecosystem and fish species activities coincide with hydropower generation, water turbidity, temperature, and dissolved gas at the outlet of a hydropower facility should be kept at certain desired levels, especially during fish migration and spawning seasons. This indicates that though a coarse resolution (daily) is used for water quality variables in our case study, it is not detrimental to estimate the hydropower at a daily scale. The lag effects are more influential when upscaling from daily to monthly as evidenced by some prior studies (Raman and Chandramouli 1996; Hejazi and Cai 2011). If the resolution of simulation is monthly, it is suggested to construct prior time step inputs together with predictors at the current time step.

Another limitation is the setting of coefficients for current and historical best ensemble members. In the presented experiments, the values of coefficients for the current and historical best members are 0.7 and 0.3, respectively. These coefficient values are based on many default settings in the uses of the PSO algorithm in the field of optimization. Furthermore, the authors have tested some other combinations of coefficients and found that values of 0.7 and 0.3 for current and historical best members in Eq. (2), respectively, are acceptable. In order to systematically determine the best coefficients, another layer of optimization and sensitivity tests is needed, which is rather beyond the proof-of-concept goal of this study. Also, the performances of different candidate models can also affect the optimal combination of coefficients in the proposed DMerge algorithm. Future work could adopt a brutal search method or a heuristic optimization scheme to determine the weights for different problems using other regression models.

Further study is also suggested to determine the optimal lag times on each climate index for water resources planning and investigate the predictability of grouped model inputs or dimensionality-reduced model inputs for hydropower. It will also be useful to relate the physical explanations of certain release events to the variability of certain climate indices. For example, the mechanism of fast horizontal water vapor transfers from tropical regions to northern California (e.g., the atmospheric river events) deserves further investigation and should be conducted in other research. In addition, the soft data (expert knowledge) can also be quantified and incorporated into the proposed data-driven models to further enhance models' predictive performances for hydropower discharge simulation. The data-driven approaches, as demonstrated in this paper, have the flexibility of adding ancillary information or removing nonuseful predictors, which should be emphasized in further studies of reservoir operation and water resources management.

Conclusion

In this study, a newly developed DMerge technique is proposed to postprocess multiple DT models' outputs and provide a more accurate ensemble prediction. The performance of the DMerge method was compared with the AdaBoost tree, RF, and Extra-Trees algorithms with respect to multiple statistics on a hydropower simulation problem. Experiments were carried out under different scenarios to simulate the hydropower discharges from the Shasta Dam using multiple types of information. A total of 26 types of decision variables were selected as model inputs, including hydrological information, climate phenomenon indices, reservoir state and regulation, and water quality indicators. The model inputs sensitivities were evaluated with respect to their predictability for hydropower releases. This paper provides a promising way to combine different AI and DM model outputs and demonstrates that improvements in the prediction accuracy of hydropower releases are achievable. Furthermore, the approaches employed in this paper are able to provide a flexible modeling capability for decision makers to utilize various types of information for water resources planning. The following findings and conclusions are summarized based on the presented experiments:

1. A DMerge method appears to be a superior model ensemble algorithm as compared to the SMA method, as well as individual DT models, including the AdaBoost, RF, and Extra-Trees algorithms in our study case. The DMerge method is inspired by the concept used in particle swarm optimization that both historical and most current information are weighted to approximate further status. The advantage of the DMerge method lies in its structure of self-learning and bias-removing mechanism

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

in producing future time step prediction, in which the historical best-performing model is weighted against the current best-performing models in a dynamical ensemble process.
2. Our case study over Shasta Lake in northern California suggests that the proposed DMerge method could produce better statistical measures than other DT methods and the SMA method, which indicates a stronger ability to simulate decision making in hydropower releases. Furthermore, in our case study, climate condition is identified as one of the most important factors for hydropower releases. Nino 1 and seasonality are identified as representative indicators within the climate category for explaining the variation of hydropower releases from Shasta Lake.
3. In our hydropower simulation study, the AdaBoost tree method is superior to the RF and the Extra-Trees algorithms, especially when the number of predictors is sufficient (Scenario 1). Under Scenario 3, which has a limited number of model inputs as compared to other scenarios, the AdaBoost tree, RF, and the Extra-Trees algorithms have similar results as compared to one another. The proposed DMerge and SMA methods are able to increase the prediction accuracy under all experiment scenarios, and the DMerge approach is able to produce better results than the commonly used SMA method.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository or online from the California Data Exchange Center (http://cdec.water.ca.gov/index.html) and the NOAA Earth System Research Laboratory (http://www.esrl.noaa.gov/psd/data/climateindices/list/). The digital elevation model raster files are obtained from http://www.webgis.com/srtm3.html for the State of California. Information about the *Journal*'s data-sharing policy can be found here: https://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0001263.

Acknowledgments

The financial support for this study is from the DOE (Prime Award No. DE-IA0000018), CEC (Award No. 300-15-005), the Open Foundation of Key Laboratory of Soil and Water Loss Process and Control on the Loess Plateau of Ministry of Water Resources, and a NASA Minority University Research and Education Project (MUREP) Institutional Research Opportunity grant (NNX15AQ06A).

References

Alfaro Cortés, E., M. Gámez Martínez, and N. García Rubio. 2007. "Multi-class corporate failure prediction by Adaboost.M1." *Int. Adv. Econ. Res.* 13 (3): 301–312. https://doi.org/10.1007/s11294-007-9090-2.

Ashaary, N. A., W. H. W. Ishak, and K. R. Ku-Mahamud. 2015. "Forecasting model for the change of reservoir water level stage based on temporal pattern of reservoir water level." In *Proc., 5th Int. Conf. on Computing and Informatics*, 692–697. Kedah, Malaysia: Univ. Utara Malaysia.

Bai, P., X. Liu, T. Yang, K. Liang, and C. Liu. 2016. "Evaluation of streamflow simulation results of land surface models in GLDAS on the Tibetan plateau." *J. Geophys. Res.: Atmos.* 121 (20): 12–180. https://doi.org/10.1002/2016JD025501.

Banfield, R. E., L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer. 2007. "A comparison of decision tree ensemble creation techniques." *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (1): 173–180. https://doi.org/10.1109/TPAMI.2007.250609.

Barlow, M., S. Nigam, and E. H. Berbery. 2001. "ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow." *J. Clim.* 14 (9): 2105–2128. https://doi.org/10.1175/1520-0442(2001)014<2105:EPDVAU>2.0.CO;2.

Bauer, E., and R. Kohavi. 1999. "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants." *Mach. Learn.* 36 (1): 105–139. https://doi.org/10.1023/A:1007515423169.

Bessler, F. T., D. A. Savic, and G. A. Walters. 2003. "Water reservoir control with data mining." *J. Water Resour. Plann. Manage.* 129 (1): 26–34. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:1(26).

Breiman, L. 1996. "Bagging predictors." *Mach. Learn.* 24 (2): 123–140. https://doi.org/10.1007/BF00058655.

Breiman, L. 2001. "Random forests." *Mach. Learn.* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. *Classification and regression trees*. Boca Raton, FL: CRC Press.

Briem, G. J., J. A. Benediktsson, and J. R. Sveinsson. 2002. "Multiple classifiers applied to multisource remote sensing data." *IEEE Trans. Geosci. Remote Sens.* 40 (10): 2291–2299. https://doi.org/10.1109/TGRS.2002.802476.

Cane, M. A. 2005. "The evolution of El Niño, past and future." *Earth Planet. Sci. Lett.* 230 (3–4): 227–240. https://doi.org/10.1016/j.epsl.2004.12.003.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa. 2010. "Tree-based reinforcement learning for optimal water reservoir operation." *Water Resour. Res.* 46 (9): 1–19. https://doi.org/10.1029/2009WR008898.

Cayan, D. R., K. T. Redmond, and L. G. Riddle. 1999. "ENSO and hydrologic extremes in the western United States." *J. Clim.* 12 (9): 2881–2893. https://doi.org/10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2.

Cayan, D. R., and R. H. Webb. 1993. *El Niño/southern oscillation and streamflow in the western United States*. Cambridge, UK: Cambridge University Press.

CDWR (California Department of Water Resources). 2013. *California Water Plan Update 2013*. Sacramento, CA: CDWR.

CDWR (California Department of Water Resources). 2015. *California Water Plan Update 2018*. Sacramento, CA: CDWR.

CDWR (California Department of Water Resources). 2017. *Connecting the dots between water, energy, food, and ecosystems issues for integrated water management in a changing climate*. Sacramento, CA: CDWR.

Chan, J. C.-W., and D. Paelinckx. 2008. "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery." *Remote Sens. Environ.* 112 (6): 2999–3011. https://doi.org/10.1016/j.rse.2008.02.011.

Chang, F. J., Y. C. Wang, and W. P. Tsai. 2016. "Modelling intelligent water resources allocation for multi-users." *Water Resour. Manage.* 30 (4): 1395–1413. https://doi.org/10.1007/s11269-016-1229-6.

Chang, M. K., J. D. Eichman, F. Mueller, and S. Samuelsen. 2013. "Buffering intermittent renewable power with hydroelectric generation: A case study in California." *Appl. Energy* 112 (Dec): 1–11. https://doi.org/10.1016/j.apenergy.2013.04.092.

Che, D., Q. Liu, K. Rasheed, and X. Tao. 2011. In *Software tools and algorithms for biological systems*, edited by H. R. Arabnia, and Q.-N. Tran, 191–199. New York: Springer.

Cheng, C., L. Yan, A. Mirchi, and K. Madani. 2017. "China's booming hydropower: Systems modeling challenges and opportunities." *J. Water Resour. Plann. Manage.* 143 (1): 02516002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000723.

Cheng, C. C., N. S. Hsu, and C. C. Wei. 2008. "Decision-tree analysis on optimal release of reservoir storage under typhoon warnings." *Nat. Hazards* 44 (1): 65–84. https://doi.org/10.1007/s11069-007-9142-1.

Cheng, C. T., W. J. Niu, Z. K. Feng, J. J. Shen, and K. W. Chau. 2015. "Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization." *Water* 7 (12): 4232–4246. https://doi.org/10.3390/w7084232.

Chu, W., T. Yang, and X. Gao. 2014. "Comment on "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

(ZS) and high-performance computing" by Eric Laloy and Jasper A. Vrugt." *Water Resour. Res.* 50 (3): 2775–2780. https://doi.org/10.1002/2012WR013341.
Conklin, D. E., and P. S. Young. 2008. *Ecological evaluation of hydropower pulsed releases on California stream system*. Davis, CA: University of California, Davis.
Conklin, D. E., P. S. Young, University of California, Davis, California Energy, Commission, and Public Interest Energy, Research. 2007. *Hydropower-related pulsed flow impacts on stream fishes, amphibians, and macroinvertebrates: PIER consultant report*. Davis, CA: University of California, Davis.
Cook, E. R., R. Seager, M. A. Cane, and D. W. Stahle. 2007. "North American drought: Reconstructions, causes, and consequences." *Earth Sci. Rev.* 81 (1): 93–134. https://doi.org/10.1016/j.earscirev.2006.12.002.
Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan. 2011. "Atmospheric rivers, floods and the water resources of California." *Water* 3 (2): 445–478. https://doi.org/10.3390/w3020445.
Dietterich, T. G. 2000a. "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization." *Mach. Learn.* 40 (2): 139–157. https://doi.org/10.1023/A:1007607513941.
Dietterich, T. G. 2000b. "Multiple classifier systems." In *Proc., First international workshop, MCS*, 1–15. Berlin: Springer.
DOE. 2014a. *Environment baseline. Vol. 4 of Energy-water nexus*. Washington, DC: DOE.
DOE. 2014b. *The water-energy nexus: Challenges and opportunities*. Washington, DC: DOE.
Drucker, H. 1997. *Improving regressors using boosting techniques*. La Jolla, CA: International Conference on Machine Learning.
Eberhart, R. C., and J. Kennedy. 1995. *A new optimizer using particle swarm theory.* New York: IEEE.
EPA. 2013a. *Priority pollutants*. Washington, DC: EPA.
EPA. 2013b. *Proposed effluent guidelines for the steam electric power generating category.* Washington, DC: EPA.
Ernst, D., P. Geurts, and L. Wehenkel. 2005. "Tree-based batch mode reinforcement learning." *J. Mach. Learn. Res.* 6 (Apr): 503–556.
FERC. 2011. *Order granting petition*. Washington, DC: Iberdrola Renewables, Inc., PacifiCorp, NextEra Energy Resources, LLC, Invenergy Wind North America LLC, and Horizon Wind Energy LLC v. Bonneville Power Administration.
Freund, Y. 1995. "Boosting a weak learning algorithm by majority." *Inf. Comput.* 121 (2): 256–285. https://doi.org/10.1006/inco.1995.1136.
Freund, Y., and R. E. Schapire. 1996. "Experiments with a new boosting algorithm." In *Proc., 13th Int. Conf. on International Conference on Machine Learning (ICML'96)*. Bari, Italy: International Conference on Machine Learning.
Freund, Y., and R. E. Schapire. 1997. "A decision-theoretic generalization of on-line learning and an application to boosting." *J. Comput. Syst. Sci.* 55 (1): 119–139. https://doi.org/10.1006/jcss.1997.1504.
Galelli, S., and A. Castelletti. 2013. "Assessing the predictive capability of randomized tree-based ensembles in streamflow modeling." *Hydrol. Earth Syst. Sci.* 17 (7): 2669–2684. https://doi.org/10.5194/hess-17-2669-2013.
Garen, D. C. 1993. "Revised surface-water supply index for western United States." *J. Water Resour. Plann. Manage.* 119 (4): 437–454. https://doi.org/10.1061/(ASCE)0733-9496(1993)119:4(437).
Geurts, P., D. Ernst, and L. Wehenkel. 2006. "Extremely randomized trees." *Mach. Learn.* 63 (1): 3–42. https://doi.org/10.1007/s10994-006-6226-1.
Geurts, P., N. Touleimat, M. Dutreix, and F. d'Alché-Buc. 2007. "Inferring biological networks with output kernel trees." *BMC Bioinf.* 8 (2): S4. https://doi.org/10.1186/1471-2105-8-S2-S4.
Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez. 2009. "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling." *J. Hydrol.* 377 (1): 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.
Gutiérrez, F., and J. A. Dracup. 2001. "An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño–Southern Oscillation indicators." *J. Hydrol.* 246 (1–4): 181–196. https://doi.org/10.1016/S0022-1694(01)00373-0.
Hagedorn, R., F. J. Doblas-Reyes, and T. Palmer. 2005. "The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept." *Tellus A* 57 (3): 219–233.
Hancock, T., R. Put, D. Coomans, Y. Vander Heyden, and Y. Everingham. 2005. "A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies." *Chemom. Intell. Lab. Syst.* 76 (2): 185–196. https://doi.org/10.1016/j.chemolab.2004.11.001.
He, X., N. W. Chaney, M. Schleiss, and J. Sheffield. 2016. "Spatial downscaling of precipitation using adaptable random forests." *Water Resour. Res.* 52 (10): 8217–8237. https://doi.org/10.1002/2016WR019034.
Hejazi, M. I., and X. M. Cai. 2011. "Building more realistic reservoir optimization models using data mining: A case study of Shelbyville Reservoir." *Adv. Water Resour.* 34 (6): 701–717. https://doi.org/10.1016/j.advwatres.2011.03.001.
Ji, C., Z. Jiang, P. Sun, Y. Zhang, and L. Wang. 2015. "Research and application of multidimensional dynamic programming in cascade reservoirs based on multilayer nested structure." *J. Water Resour. Plann. Manage.* 141 (7): 04014090. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000489.
Kaygusuz, K. 2004. "Hydropower and the world's energy future." *Energy Sources* 26 (3): 215–224. https://doi.org/10.1080/00908310490256572.
Kelly, K. F. 1986. *Reservoir operation during drought. Case studies*. Davis: DTIC Document.
Kennedy, J. 2011. *Encyclopedia of machine learning*. New York: Springer.
Kennedy, J., J. F. Kennedy, R. C. Eberhart, and Y. Shi. 2001. *Swarm intelligence*. Burlington: Morgan Kaufmann.
Kim, S. Y., and A. Upneja. 2014. "Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models." *Econ. Modell.* 36 (Jan): 354–362. https://doi.org/10.1016/j.econmod.2013.10.005.
Kim, T., J. Y. Shin, H. Kim, S. Kim, and J. H. Heo. 2019. "The use of large-scale climate indices in monthly reservoir inflow forecasting and its application on time series and artificial intelligence models." *Water* 11 (2): 374. https://doi.org/10.3390/w11020374.
Kimmell, T., and J. Veil. 2009. *Impact of drought on US steam electric power plant cooling water intakes and related water resource management issues*. Lemont, IL: Argonne National Laboratory.
Kumar, A. R. S., M. K. Goyal, C. S. P. Ojha, R. D. Singh, and P. K. Swamee. 2013a. "Application of artificial neural network, fuzzy logic and decision tree algorithms for modelling of streamflow at Kasol in India." *Water Sci. Technol.* 68 (12): 2521–2526. https://doi.org/10.2166/wst.2013.491.
Kumar, A. R. S., M. K. Goyal, C. S. P. Ojha, R. D. Singh, P. K. Swamee, and R. K. Nema. 2013b. "Application of ANN, fuzzy logic and decision tree algorithms for the development of reservoir operating rules." *Water Resour. Manage.* 27 (3): 911–925. https://doi.org/10.1007/s11269-012-0225-8.
Li, C., J. Zhou, S. Ouyang, X. Ding, and L. Chen. 2014. "Improved decomposition–coordination and discrete differential dynamic programming for optimization of large-scale hydropower system." *Energy Convers. Manage.* 84 (Aug): 363–373. https://doi.org/10.1016/j.enconman.2014.04.065.
Li, X., M. Paster, and J. Stubbins. 2015. "The dynamics of electricity grid operation with increasing renewables and the path toward maximum renewable deployment." *Renewable Sustainable Energy Rev.* 47 (Jul): 1007–1015. https://doi.org/10.1016/j.rser.2015.03.039.
Li, X.-Z., L.-Z. Xu, and Y.-G. Chen. 2010. "Implicit stochastic optimization with data mining for reservoir system operation." In *Proc., 2010 Int. Conf. on Machine Learning and Cybernetics*, 2410–2415. New York: IEEE.
Liaw, A., and M. Wiener. 2002. "Classification and regression by randomForest." *R News* 2 (3): 18–22.
Louks, D. P., and O. T. Sigvaldason. 1981. *Multiple reservoir operation in North America*. Reston, VA: ASCE.
Lü, A., S. Jia, W. Zhu, H. Yan, S. Duan, and Z. Yao. 2011. "El Niño-Southern Oscillation and water resources in the headwaters region of the Yellow River: links and potential for forecasting." *Hydrol. Earth*

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

Syst. Sci. 15 (4): 1273–1281. https://doi.org/10.5194/hess-15-1273-2011.

Madani, K. 2011. "Hydropower licensing and climate change: Insights from cooperative game theory." *Adv. Water Resour.* 34 (2): 174–183. https://doi.org/10.1016/j.advwatres.2010.10.003.

Mantua, N. J., and S. R. Hare. 2002. "The Pacific decadal oscillation." *J. Oceanogr.* 58 (1): 35–44. https://doi.org/10.1023/A:1015820616384.

Marée, R., P. Geurts, and L. Wehenkel. 2007. "Random subwindows and extremely randomized trees for image classification in cell biology." *BMC Cell Biol.* 8 (1): S2. https://doi.org/10.1186/1471-2121-8-S1-S2.

Miao, C. Y., J. R. Ni, and A. G. L. Borthwick. 2010. "Recent changes of water discharge and sediment load in the Yellow River basin, China." *Prog. Phys. Geogr.* 34 (4): 541–561. https://doi.org/10.1177/0309133310369434.

Montoya, E., J. Dozier, and W. Meiring. 2014. "Biases of April 1 snow water equivalent records in the Sierra Nevada and their associations with large-scale climate indices." *Geophys. Res. Lett.* 41 (16): 5912–5918. https://doi.org/10.1002/2014GL060588.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." *Trans. ASABE* 50 (3): 885–900. https://doi.org/10.13031/2013.23153.

Naeini, M. R., T. Yang, M. Sadegh, A. AghaKouchak, K. Hsu, and S. Sorooshian. 2018. "Shuffled complex-self adaptive hybrid evolution (SC-SAHEL) optimization framework." *Environ. Modell. Software* 104 (Jun): 215–235. https://doi.org/10.1016/j.envsoft.2018.03.019.

Namias, J., and D. R. Cayan. 1981. "Large-scale air-sea interactions and short-period climatic fluctuations." *Science* 214 (4523): 869–876. https://doi.org/10.1126/science.214.4523.869.

Nash, J. E., and J. V. Sutcliffe. 1970. "River flow forecasting through conceptual models part I—A discussion of principles." *J. Hydrol.* 10 (3): 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Pagano, T., and D. Garen. 2003. "Use of climate information in official western US water supply forecasts." In *Proc., World Water and Environmental Resources Congress 2003*. Reston, VA: ASCE.

Quinlan, J. R. 1986. "Induction of decision trees." *Mach. Learn.* 1 (1): 81–106.

Quinlan, J. R. 1996. *Bagging, boosting, and C4*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Raman, H., and V. Chandramouli. 1996. "Deriving a general operating policy for reservoirs using neural network." *J. Water Resour. Plann. Manage.* 122 (5): 342–347. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:5(342).

Raso, L., D. Schwanenberg, N. C. van de Giesen, and P. J. van Overloop. 2014. "Short-term optimal operation of water systems using ensemble forecasts." *Adv. Water Resour.* 71 (Sept): 200–208. https://doi.org/10.1016/j.advwatres.2014.06.009.

Redmond, K. T., and R. W. Koch. 1991. "Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices." *Water Resour. Res.* 27 (9): 2381–2399. https://doi.org/10.1029/91WR00690.

Roe, B. P., H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor. 2005. "Boosted decision trees as an alternative to artificial neural networks for particle identification." *Nucl. Instrum. Methods Phys. Res., Sect. A* 543 (2–3): 577–584. https://doi.org/10.1016/j.nima.2004.12.018.

Sattari, M. T., H. Apaydin, F. Ozturk, and N. Baykal. 2012. "Application of a data mining approach to derive operating rules for the Eleviyan irrigation reservoir." *Lake Reservoir Manage.* 28 (2): 142–152. https://doi.org/10.1080/07438141.2012.678927.

Schwanenberg, D., F. M. Fan, S. Naumann, J. I. Kuwajima, R. A. Montero, and A. A. Dos Reis. 2015. "Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty." *Water Resour. Manage.* 29 (5): 1635–1651. https://doi.org/10.1007/s11269-014-0899-1.

Schwanenberg, D., M. Xu, T. Ochterbeck, C. Allen, and D. Karimanzira. 2014. "Short-term management of hydropower assets of the Federal Columbia River power system." *J. Appl. Water Eng. Res.* 2 (1): 25–32. https://doi.org/10.1080/23249676.2014.912952.

Shamim, M. A., M. Hassan, S. Ahmad, and M. Zeeshan. 2016. "A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels." *KSCE J. Civ. Eng.* 20 (2): 971–977. https://doi.org/10.1007/s12205-015-0298-z.

Shi, Y. 2001. "Particle swarm optimization: Developments, applications and resources." In *Proc., 2001 Congress on Evolutionary Computation*, 81–86. New York: IEEE.

Tao, Y., T. Yang, M. Faridzad, L. Jiang, X. He, and X. Zhang. 2018. "Non-stationary bias correction of monthly CMIP5 temperature projections over China using a residual-based bagging tree model." *Int. J. Climatol.* 38 (1): 467–482. https://doi.org/10.1002/joc.5188.

Tarroja, B., A. AghaKouchak, R. Sobhani, D. Feldman, S. Jiang, and S. Samuelsen. 2014. "Evaluating options for balancing the water–electricity nexus in California: Part 2—Greenhouse gas and renewable energy utilization impacts." *Sci. Total Environ.* 497–498 (Nov): 711–724. https://doi.org/10.1016/j.scitotenv.2014.06.071.

Trenberth, K. E., and D. P. Stepaniak. 2001. "Indices of El Niño evolution." *J. Clim.* 14 (8): 1697–1701. https://doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2.

Tsai, C. C., M. C. Lu, and C. C. Wei. 2012. "Decision tree-based classifier combined with neural-based predictor for water-stage forecasts in a river basin during typhoons: A case study in Taiwan." *Environ. Eng. Sci.* 29 (2): 108–116. https://doi.org/10.1089/ees.2011.0210.

USACE (United States Army Corps of Engineers). 2016. *Engineering and design water control management*. Washington, DC: USACE.

Veil, J., J. VanKuiken, S. Folga, and J. Gillette. 1993. *Impact on the steam electric power industry of deleting Section 316 (a) of the Clean Water Act: Energy and environmental impacts*. Washington, DC: DOE.

Wang, C., and D. B. Enfield. 2001. "The tropical Western Hemisphere warm pool." *Geophys. Res. Lett.* 28 (8): 1635–1638. https://doi.org/10.1029/2000GL011763.

Wei, C. C. 2012. "Discretized and continuous target fields for the reservoir release rules during floods." *Water Resour. Manage.* 26 (12): 3457–3477. https://doi.org/10.1007/s11269-012-0085-2.

Wei, C. C., and N. S. Hsu. 2009. "Optimal tree-based release rules for real-time flood control operations on a multipurpose multireservoir system." *J. Hydrol.* 365 (3–4): 213–224. https://doi.org/10.1016/j.jhydrol.2008.11.038.

Wolter, K., and M. S. Timlin. 1998. "Measuring the strength of ENSO events: How does 1997/98 rank?" *Weather* 53 (9): 315–324. https://doi.org/10.1002/j.1477-8696.1998.tb06408.x.

Wu, B., H. Ai, C. Huang, and S. Lao. 2004. "Fast rotation invariant multi-view face detection based on real adaboost." In *Proc., Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 79–84. New York: IEEE.

Yang, T. 2015. *A framework to provide optimal management strategies for California's reservoirs in achieving sustainable water supply and high hydropower productivity.* Irvine, CA: University of California.

Yang, T., A. A. Asanjan, M. Faridzad, N. Hayatbini, X. Gao, and S. Sorooshian. 2017b. "An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis." *Inf. Sci.* 418–419 (Dec): 302–316. https://doi.org/10.1016/j.ins.2017.08.003.

Yang, T., A. A. Asanjan, E. Welles, X. Gao, S. Sorooshian, and X. Liu. 2017a. "Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information." *Water Resour. Res.* 53 (4): 2786–2812. https://doi.org/10.1002/2017WR020482.

Yang, T., X. Gao, S. L. Sellars, and S. Sorooshian. 2015. "Improving the multi-objective evolutionary optimization algorithm for hydropower reservoir operations in the California Oroville–Thermalito complex." *Environ. Modell. Software* 69 (Jul): 262–279. https://doi.org/10.1016/j.envsoft.2014.11.016.

Yang, T., X. Gao, S. Sorooshian, and X. Li. 2016. "Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme." *Water Resour. Res.* 52 (3): 1626–1651. https://doi.org/10.1002/2015WR017394.

Yang, T., K. Hsu, Q. Duan, S. Sorooshian, and C. Wang. 2018. "Method to estimate optimal parameters." In *Handbook of hydrometeorological*

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.

ensemble forecasting, edited by Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. Cloke, and J. Schaake. Berlin: Springer.

Zagona, E. A., T. J. Fulp, R. Shane, T. Magee, and H. M. Goranflo. 2001. "Riverware: A generalized tool for complex reservoir system modeling." *J. Am. Water Resour. Assoc.* 37 (4): 913–929. https://doi.org/10.1111/j.1752-1688.2001.tb05522.x.

Zhang, D., J. Lin, Q. Peng, D. Wang, T. Yang, S. Sorooshian, X. Liu, and J. Zhuang. 2018. "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm." *J. Hydrol.* 565 (Oct): 720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050.

Zhang, Y., J. M. Wallace, and D. S. Battisti. 1997. "ENSO-like interdecadal variability: 1900-93." *J. Clim.* 10 (5): 1004–1020. https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zhang, Z. B., S. H. Zhang, S. M. Geng, Y. Z. Jiang, H. Li, and D. W. Zhang. 2015. "Application of decision trees to the determination of the year-end level of a carryover storage reservoir based on the iterative dichotomizer 3." *Int. J. Electr. Power Energy Syst.* 64 (Jan): 375–383. https://doi.org/10.1016/j.ijepes.2014.06.073.

Downloaded from ascelibrary.org by Calif State Univ-Los Angeles on 01/23/20. Copyright ASCE. For personal use only; all rights reserved.