Title
Optimization of the SWAT+ model to adequately predict different segments of a managed streamflow hydrograph

Journal
Hydrological Sciences Journal, 69(9)

ISSN
0262-6667

Authors
Tigabu, Tibebe B
Visser, Ate
Kadir, Tariq
et al.

Publication Date
2024-07-03

DOI
10.1080/02626667.2024.2364714

Optimization of the SWATPlus Model to Adequately Predict Different Segments of a Managed Streamflow Hydrograph

Tibebe B. Tigabu[1], Ate Visser[2], Tariq Kadir[3], Shalamu Abudu[3], Philip Cameron-Smith[2], Helen E. Dahlke[1]

[1]Department of Land, Air, and Water Resources, University of California, Davis, Davis, CA, USA, [2]Lawrence Livermore National Laboratory, Livermore, CA, USA, [3]California Department of Water Resources, CA, USA

Corresponding author: Tibebe Tigabu, Department of Land, Air, and Water Resources, University of California Davis, 1 Shields Ave, Davis, CA 95616, USA: E-mail: tbtigabu@ucdavis.edu

Optimization of the SWATPlus Model to Adequately Predict Different Segments of a Managed Streamflow Hydrograph

ABSTRACT: Complete representation of rainfall-runoff responses in complex, large watersheds using a single-objective parameterization approach in watershed models is often unachievable. In this study we present a calibration approach for the SWAT+ model that independently fits model parameters for different flow segments of the hydrograph. The approach is demonstrated for the Feather River, California, USA using daily streamflow from the Lake Oroville reservoir outlet gage. Results show that when model parameters were independently fitted for different flow segments the KGE, NSE, PBIAS, and RSR values improved to 0.96, 0.99, -3.3, and 0.10, respectively, compared to 0.72, 0.66, -9.30, and 0.53, respectively, achieved under a multi-objective and full hydrograph (average hydrograph) calibration. The results highlight when considering the average hydrograph and flow duration curves, a more balanced representation of both poorly and well-performing segments is achieved, emphasizing the importance of segment-specific parameterization and multi-objective evaluation for accurately representing different flow conditions.

Keywords: Feather; managed streamflow, optimization; flow segment, multi-objective functions, SWAT+

1. Introduction

Climate change will impact water resources globally, especially in high elevation regions that serve as the water towers of the world (Immerzeel et al., 2010; 2020). Predicting the impact of climate change on the timing and magnitude of streamflow in a warmer, low-to-no-snow future (Barnett et al., 2005; Siirila-Woodburn et al., 2021) is critical to adapt water resources infrastructure (Hedden-Nicely, 2022) and secure safe and reliable sources of water for human consumption, agriculture, ecosystem health, and industry. Advancements in large-scale integrated hydrologic models are required to quantify the current and future water supply and water quality conditions (Bailey et al. 2023). Prediction of current and future streamflow is one of the most important tasks in water resources management. Hydrologists have been using data-driven and physically based hydrologic models to simulate streamflow and other hydrologic processes in catchments. There are several such models that are being used globally. Among others, the Soil and Water Assessment Tools (SWAT) developed by Arnold et al. (1998), the MIKE SHE, which is the European Hydrological System Model (Refsgaard, and Storm 1995), and the Agricultural Policy/Environmental Extender abbreviated as APEX (Gassman 2009) are a few examples of watershed-scale models used worldwide (Golmohammadi et al. 2014).

The SWAT model has been extensively used to predict the terrestrial hydrologic cycle of watersheds, to evaluate best management practices, to simulate environmental flow, and to investigate the impacts of climate and land use changes on hydrologic processes (e.g., Bailey et al. 2023, Wagner et al. 2022, Liu et al. 2021, Mahmoodi et al. 2021, Tigabu et al. 2020, Kannan et al. 2019, Aliye et al. 2020). Due to the growing interest of users and the availability of a strong user group for the SWAT model, SWAT model developers have been advancing the model structure over time. Most recently, Bieger et al. (2017) completely restructured and advanced the SWAT model into the SWAT+ version, which is the most prominent revision in its history (Wagner et al. 2022). SWAT+ incorporates new features, including landscape unit and stream connectivity, to better represent various scales of watersheds and a consolidated file structure (White et al. 2022). SWAT+ is also capable of simulating managed flows, impacted by dams and reservoir operations, via a set of reservoir operation rules (Wu et al. 2020). Despite the extensive progress on advancing the SWAT model structure and a growing number of studies on calibration and uncertainty analysis, a holistic calibration approach is still missing due to the heterogeneity and complexity of watershed hydrologic processes encountered throughout the world. Moreover, calibration of hydrologic models, including SWAT, is always challenging due to uncertainties that arise from model inputs, model structure, parameters, and outputs (Wu et al. 2021). The challenge is far higher in regions with extensive water management infrastructure for flood protection and water delivery, such as dams, surface water reservoirs, and hydropower operations, and water diversion as can be found in the State of California, USA.

California is known for having the most complex water delivery systems in the world (Avanzi 2018). According to Avanzi (2018), the water delivery systems in California are implemented through a network of reservoirs, aqueducts, and groundwater pumps that deliver water from the headwaters in the northern and eastern portions of the state to population centers and agricultural land in the western and southern parts of the state. These numerous water infrastructures, including a collection of canals, pipelines, reservoirs, and hydroelectric power facilities, deliver clean water to

38 million Californians, 3.2 million hectares of farmland, and businesses throughout the state (DWR 2023). Moreover, the high seasonal and interannual hydrologic availability and extreme weather events in the state are other factors that make the water delivery system most intricate and complex (Hanak and Lund 2012). The Feather River watershed is one of the most important watersheds in California that provides a third of all water distributed by the Metropolitan Water District of Southern California through Oroville reservoir and canals (Avanzi et al. 2018). The complex hydrology, intensive water use system, Mediterranean climate, and large elevation range (including the rain-snow transition) of the Feather River make water management very challenging. To accommodate these challenges, the California Department of Water Resources uses various models to track streamflow and water deliveries to the State Water project, the nation's largest state-owned water and power generator and user-financed water system (DWR 2023). Application of a physically based hydrologic model like SWAT+ is needed to sustain water supply services under such multifold problems.

For the purpose of water management, it is also essential to calibrate and validate the model to accurately represent the spatial and temporal heterogeneities of hydrologic processes (Mengistu et al. 2019). To achieve an adequate calibration and to reduce uncertainty in model simulations (Guse et al. 2020; Kannan et al. 2019), both calibration and validation periods should include wet, average, and dry years (Arnold et al. 2012).

Various calibration approaches have been applied to predict hydrologic processes in watersheds. The traditional approach involves calibrating the model based on streamflow data at a catchment outlet point (Daggupati et al. 2015). However, new approaches have been developed, such as seasonal clustering of daily streamflow for calibration (Lakshmi and Sudheer 2021; Tigabu et al. 2023), multi-metric calibration for different parts of hydrographs (Pfannerstill et al. 2014), and incorporating streamflow signatures, flow duration curves, and spatially distributed remote sensing data in the model calibration (Alemayehu et al. 2022; Westerberg et al. 2011). More recently, the multicriteria sequential calibration and uncertainty analysis (MS-CUA) method was developed to better optimize SWAT simulations and to provide balanced uncertainty analyses compared to other calibration approaches (Wu et al. 2021). Multi-variable calibration approaches have also been shown to improve the performance of SWAT, particularly in predicting snow-affected streamflow (Chen et al. 2023, Liu et al. 2021). Similarly, the utilization of streamflow signatures (flow duration curve) and remote sensing information in hydrologic model evaluations ~~has~~ have emerged as crucial calibration options. Notably, Dal Molin et al. (2023), Alemayehu et al. (2022), Pfannerstill et al. (2017), Donnelly et al. (2016), Shafii and Tolson (2015), and Pfannerstill et al. (2014) ~~have~~ all have considered flow duration curves (FDCs) as calibration objectives for model performance evaluation.

Donnelly et al. (2016) incorporated flow signatures in the evaluation ~~of~~ of the performance of a multi-basin model ~~performance~~ across various sites within the domain, utilizing several model performance metrics to better understand dominant catchment processes. Their study indicated that simulated flows based on the semi-distributed and process-based HYPE model (Lindström et al. 2010) successfully captured observations, with dominant temporal variability also represented when considering all flow signatures simultaneously. Dal Molin et al. (2023) investigated the efficacy of a streamflow signature-based model calibration in predicting streamflow for six catchments in the Thur basin, in northeastern Switzerland. Their findings demonstrated that signature-based calibration of precipitation-streamflow models adequately predicts streamflow for ungagged catchments. Similarly, Alemayehu et al. (2022) found that utilizing flow duration curves (FDCs) from historical records and recent remote sensing-based evapotranspiration data in model calibration enhances the efficiency of hydrologic models in simulating catchment hydrologic processes. This underscores the superior efficiency of FDC-based calibration approaches for the SWAT model when integrating streamflow signatures and remote sensing information. Consequently, the authors suggest employing historical FDC records and recent remote sensing-based evapotranspiration data in model calibration to optimize the efficiency of hydrologic models in simulating catchment hydrologic processes. Overall, hydrologic model calibration frameworks that prioritize FDCs and remote sensing information as calibration objectives are essential for overcoming limitations associated with conventional calibration approaches.

Despite numerous calibration and validation studies of the SWAT model, achieving satisfactory calibration for large, complex watersheds remains challenging, as a single parameter set often fails to capture the diverse signatures of the streamflow hydrograph. As highlighted by Westerberg et al. (2011), conventional calibration performance measures suffer from four main limitations: uncertainty in observed streamflow, variable sensitivity of model performance across different flow segments, the influence of input/output errors, and the inability to evaluate model performance when observed and simulated flow magnitudes do not overlap in time. Therefore, the current study aims to demonstrate a novel calibration approach that can better represent the distinct flow signatures of managed streamflow using the SWAT+ model. Here, we independently parameterized the high-flow, middle-flow, and low-flow segments of a managed streamflow hydrograph. We believe that testing this calibration exercise is crucial, given the limited calibration exercises tested for the SWAT+ model. This study does not utilize remote sensing evapotranspiration and soil moisture data in the calibration and validation process because the scope of the study is to introduce a new method of model calibration to reproduce observed streamflow.

The study is conducted using the Feather River watershed in California, USA as an example. SWAT+ has a wider flexibility to include manmade structures such as reservoirs, weirs and ponds which can help to simulate regulated flows in a watershed (Wu et al. 2020). Multiple parameter sets will be proposed to improve the efficiency of the SWAT+ model in reproducing each flow segment independently, aiming to identify the segment of the FDC that most significantly influences the overall predictive performance of the model for managed streamflow.

This study assesses the feasibility of these methods using long-term managed streamflow data from the Feather River at the Oroville gauging station, juxtaposed with SWAT+ simulated streamflow at the same location. Oroville, the second-largest reservoir in California, serves crucial roles in water supply storage, hydropower generation, and flood control along the Feather River (Nelson et al., 2016). Given its well-documented reservoir operation rules in the SWAT+ model and extensive historical streamflow records at the outlet point, our calibration approach focuses on regulated flow at the outlet of the reservoir. Moreover, although calibration and validation weren't practical due to insufficient records at the outlet points of upstream water bodies, we have integrated these water bodies into our modeling framework. Hence, the specific objectives of the study are as follows:

1) To identify the most influential parameters for simulating high flow, middle flow, and low flow segments of the managed streamflow hydrograph using the SWAT+ model
2) To propose multiple parameters sets that can improve the efficiency of the SWAT+ model in reproducing each flow segment independently.
3) To evaluate the practicability of the proposed calibration and parameterization methods using long-term managed streamflow data from the Feather River at the Oroville gauging station and SWAT+ simulated streamflow at the same location.

2. Materials and methods

2.1. Study area

The Feather River is a large tributary to the Sacramento River (71,432 km^2), the largest river in northern California. It serves as the primary source of surface water for the state, flowing into Oroville Reservoir (Huang et al., 2012; Koczot et al., 2004). The Upper Feather River Watershed, above Oroville Dam, is situated in the Sierra Nevada Mountains of California. Using a digital elevation model with a resolution of 30 m by 30 m, we delineated the watershed to Oroville Dam and found that it covers an area of 9,427 km2, with altitudes ranging from 256 m to 2826 m above sea level.

There are three reservoirs and five powerhouses upstream of the North Fork Feather River, as depicted in Figure 1. The most prominent natural reservoir, Lake Almanor, is a spring-fed lake that has been expanded by the construction of Canyon Dam (Avanzi et al., 2018). The other two reservoirs are Butt Valley, located on Butt Creek with a capacity of 49,891 acre-feet (0.062 km^3), and Lake Almanor with a capacity of 1.308 million acre-feet (1.613 km^3). The Feather River watershed is characterized by a Mediterranean climate with warm, dry summers and cool, wet winters (Koczot et al., 2012). The mean long-term total annual precipitation is 1078 mm, while the mean monthly maximum and minimum temperatures are 21 °C and 4.8 °C, respectively.

Most of the watershed area is at elevations where winter temperatures can fluctuate from below to above freezing, and slight temperature changes can affect snow formation and melting. This feature of the watershed leads to complex streamflow variability, making changes in streamflow dependent on both temperature and precipitation lapse rates since temperature affects snowmelt and precipitation form. The land cover of the watershed primarily consists of coniferous trees, with some areas of shrubs and grassland mainly located in the agricultural valleys (Koczot et al., 2004).

Insert Figure 1 here.

2.2. Data

For this study, daily gridded precipitation and maximum and minimum temperature data were obtained from the California Department of Water Resources (CDWR 2022) based on the 4 km resolution PRISM product (PRISM, 2022). Spatial data, including land use/cover (NLCD 2001 Land Cover), soil (SSURGO), and digital elevation models, were obtained from publicly available resources listed in Table 1.

Long-term streamflow records from 1953 to the present at the Oroville dam gauging station's outlet point were obtained from the United States Geological Survey (USGS) National Water Information System (NWIS, https://waterdata.usgs.gov/nwis/) to calibrate and validate the SWAT+ model. The spatial (Figure 2), climate, and hydrologic data were used to establish and calibrate the SWAT+ model. All data used to establish and calibrate the SWAT+ model and their basic characteristics are presented in Table 1.

Insert Table 1 here.

Insert Figure 2 here.

2.3. Hydrologic model setup

This study was conducted with the SWAT+ model (Bieger et al. 2017), which is a completely revised version of SWAT (Arnold et al. 1998; Arnold and Fohrer 2005). SWAT+ is capable of simulating spatially distributed water balance and nutrient cycles based on hydrological response units (HRUs). The SWAT+ model for the Upper Feather watershed was constructed based on the pre-existing files of the SWAT2009 obtained from the California State Department of Water Resources (DWR, 2022). We used the long-term release of QGIS (version 3.22.10) to delineate watersheds and the burn function was applied to enforce the existing stream networks in the delineation of the DEM-based stream network. Moreover, we used the DEM Inversion function to classify the watershed into either landscape or flood plain zones and the Add Lake function to append the existing reservoir system.

SWAT+ incorporates a set of rules governing reservoir operations for surface water reservoirs in the United States. These rules are utilized without modification to simulate release scenarios. The default release rules, outlined in a decision table specific to the Oroville reservoir (Table 2), encompass five conditions, seven alternatives, and five action options. To facilitate the model's utilization of the decision table, key parameters such as the conditional variable, condition limits (limit variable), limit operator, and limit constant must be clearly defined (Arnold et al., 2018). In Table 2, the release rate is determined as a function of both reservoir volume and storage volume (e-pv). The conditional variable, representing reservoir volume, and the limit variable, denoting storage volume in hectares-meters (ha-m), jointly dictate the selection of alternatives and corresponding action entries. For instance, the implementation of the first alternative is contingent upon the satisfaction of the following conditional statement: if reservoir volume (conditional variable) > e-pv (storage volume in ha-m) = -6.761(limit constant) and reservoir volume < e-pv = 0.356 and month < 6.892, then the action entry involves releasing the base volume for drawdown days (dyrt) for multiple-use flood (multiple_use_fl). Similarly, the determination of action types for the remaining alternatives follows suit, relying on conditional statements derived from the interplay of the conditional variable, limiting variable, limit operation, and limit constants as outlined in Table 2. Here, it worths to note that calibration of parameters connected to reservoir storage and release are not the scope of this paper. As we used the default parameterization for the reservoir, there might be uncertainty associated with releases from reservoir.

Insert Table 2 here.

The watershed delineation step resulted in 59 subbasins and 583 channels. Following the watershed delineation step, the hydrologic response units (HRUs) were created by combining the land use, soil, and slope classes. To capture topographic effects on watershed processes, we classified the DEM into five slope classes as 0–2% (flat to very gently sloping), 2–5% (gently sloping), 5–8% (sloping), 8–15% (strongly sloping), and >15% (moderately steep to very steep) based on the Food and Agricultural Organization (FAO) slope guidelines (Jahn et al. 2006). Next, the model input files were compiled using the SWAT+ editor version 2.0.4. This resulted in 59 subbasins, 583 channels, 1131 routing units, 95,241 HRUs, and 117 aquifers objects for the first groundwater layer. To minimize the computing time, only the dominant HRUs were considered for calibration. The Hargreaves equation was used for calculating evapotranspiration (Hargreaves and Samani 1985), the variable storage method was used for channel routing, and the soil moisture function was used to calculate the average daily runoff curve number (CN).

2.4. Calibration and parametrization

2.4.1. SWAT+ Parameters, Calibration Data, and Objective Functions

The SWAT+ model of the Upper Feather River watershed is described by 19 parameters in total. For all parameters, an initial value based on prior studies is available in SWAT+. To limit the number of parameters under study here, we focus on the 19 parameters listed in Table 3.

Insert Table 3 here.

The model was calibrated and validated against daily observations of managed streamflow below Oroville dam. Five years of data (2005 – 2009) were used as model warm-up period to define appropriate initial conditions and to attain equilibrium conditions for the model. The 2010 to 2020 period was used for calibration and the five years for validation using the first (1995-1999) as warm-up period.

In accordance with Pfannerstill et al. (2014), the FDC of daily streamflow ~~volume~~ data was categorized into very high flow (0-5%), high flow (5-20%), middle flow (20-80%), low flow (80-95%), and very low flow (95-100%) segments based on the exceedance probability of average daily streamflows~~daily streamflow volume magnitude~~. This segmentation approach, as advocated by Pfannerstill et al. (2014), allows for a process-based calibration, capturing dominant watershed processes manifested in various parts of the hydrograph. The threshold for high flows was 2,000 m^3/s, coinciding with the 80th flow percentile and the threshold of low flows was 37 m^3/s, coinciding with the 10th flow percentile.

Four different objective functions were used for the automatic calibration to daily streamflow data: NSE, Kling–Gupta efficiency (KGE), PBIAS and RSR. The Nash-Sutcliff Efficiency (NSE) is a single metric that captures timing and magnitude errors between the simulated and observed mean daily streamflow (Eq. 1) (Nash-Sutcliffe 1970). The KGE (Eq. 2) is a modified version of NSE (Gupta et al. 2009), which targets maximizing its value to one based on the decomposition of the mean squared error into the three factors including mean (β), variability (α), and dynamics (correlation coefficient r) (Eq. 3) (Gupta et al. 2009).

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Q_o - Q_s\right)^2}{\sum_{i=1}^{n} \left(Q_o - \acute{Q}_o\right)^2} \qquad \text{(Eq.1)}$$

where, ‘n’ refers to the number of observations, ‘Q’ stands for streamflow, and subscripts ‘s’ and ‘o’ refer to simulated and observed, respectively.

$$KGE = 1 - ED \qquad \text{(Eq.2)}$$

$$ED = \sqrt{\left(r-1\right)^2 + ¿¿} \qquad \text{(Eq.3)}$$

where ED is the Euclidian distance from the ideal point and $\alpha = \frac{\delta_s}{\sigma_o}$, $\beta = \frac{\mu_s}{\mu_o}$, and

r = correlation coefficient.

The PBIAS (Eq. 4) (Gupta et al. 1999) is one of the objective functions related to error measure of simulations with reference to observation points. It is a popular method frequently used in determining whether a model simulation is overestimated or underestimated. PBIAS values can vary between $-\infty$ and $+\infty$, whereas its optimal value is zero.

$$PBIAS = 100 * ¿ \qquad \text{(Eq. 4)}$$

where, ‘n’, ‘Q’, ‘s’, and ‘m’ stand for the number of sampling points, streamflow, simulated, and observed, respectively.

The standardized root mean square error (RSR) is another error evaluation criterion that represents the ratio between the root mean square error (RMSE) to the standard deviation of the observations (Eq. 5). It is useful to understand the variation between the observed data and simulated data.

$$RSR = \frac{\sqrt{\sum (Q_m - Q_s)^2}}{\sqrt{\sum (Q_m - \acute{Q}_m)^2}} \qquad \text{(Eq. 5)}$$

where, ‘Q’, $\acute{Q}_m$ ‘s’, and ‘m’, streamflow, observed mean, simulated, and observed, respectively.

Both the NSE and KGE are measures of goodness-of-fit and need to be maximized. PBIAS and RSR are error coefficients and need to be minimized towards zero.

2.4.2. Sensitivity analyses

To identify the most sensitive model parameters, we adjusted the default value of each parameter (one-by-one) following appropriate change methods (Abbaspour et al. 2015) and compared the effect of these parameter changes on streamflow and water yield prediction with the default parameter set. One parameter at a time was considered and changes were applied within the SWAT+ editor interface. Several parameter value changes were applied for each parameter to quantify sensitivity and determine the maximum and minimum values for the automatic calibration. Finally, we identified the model parameters which showed the largest effect on streamflow and basin water balance components (Table 3) and further optimized their values via automatic calibration.

2.4.3. Manual and automatic calibration

We performed the model calibration using the SWATplusR package (Schürz 2019) and a multiple flow segment and multi-objective calibration approaches using performance metrics and signature metrics (Pfannerstill et al. 2014a; Pfannerstill et al. 2014b; Haas et al. 2016). The SWATplusR can automatically initiate multiple SWAT+ simulations with varying calibration parameter values. Moreover, it enables us to manage changes in model parameters, simulation periods and time steps and to store the simulated output variables.

The sensitive parameters shown in Table 3 control different aspects of the hydrological functioning of the Upper Feather River watershed. The sensitive parameters identified during the manual calibration were classified based on their relation to the model's spatial entities, such as HRUs, aquifers, soils, snow areas, and basin. These parameters were calibrated in four stages. In the first stage, we conducted 500 iterations (first iteration) using wide ranges of parameter values generated through Latin Hypercube sampling (LHS) method in R with the FME package (Soetaert and Petzoldt, 2010) for 19 parameters. The results helped determine new ranges for the subsequent phase (Appendix 1). In the second stage, we conducted 1000 iterations (second iteration) focusing solely on calibrating snowmelt, snowfall, precipitation lapse rate, and temperature lapse rate parameters independently. This was done for two reasons. First, these parameters influence water contribution to the system, introducing a level of uncertainty that may arise from the model's input data and introduce uncertainty from model (Abbaspour et al., 2017). Therefore, pre-parameter fitting was conducted to address this concern. Second, the decision to calibrate these parameters separately was influenced by the recognition that employing many model parameters during automatic calibration could lead to equifinality or difficulty in determining the optimal model parameterization from the available set (Casado-Rodríguez and del Jesus, 2022). Optimal parameter values were identified by comparing the performance of the snowfall, snowmelt, lapse rate parameters to the annual basin average precipitation. At this stage, our reference to calibrate the snowmelt parameters was the annual basin average precipitation.

In the third stage, keeping calibrated snow, temperature, and precipitation parameters constant, we performed 1000 iterations for the remaining 12 parameters using parameter sets generated through LHS method in the R FME package (Soetaert and Petzoldt, 2010). We chose not to employ the SUFI-2 calibration algorithm due to its tendency to produce only a single local solution.

The automated calibration was performed following a methodology outlined by Guse et al. (2020). This approach allowed exploration of potential model performances within the specified parameter space. Simulated streamflow by SWAT+ with these parameter sets was compared against observed managed streamflow, leading to the selection of new parameter value ranges for the fourth iteration, comprising 500 parameter sets (Appendix 1). Finally, in the fourth iteration, parameters were optimized for high, middle, and low flow segments using the last 500 parameter sets and multiple model evaluation statistics such as NSE, KGE, RSR, and PBIAS for managed streamflow prediction (Table 4).

To identify the optimal set of model parameters suitable for the high, middle, and low flow segments, ensuring acceptable statistical indices for all objective functions, the hydroGOF package in R (Zambrano-Bigiarini 2020) was employed. Throughout this process, the historically observed streamflow hydrograph and flow duration curves were used to select suitable set of parameters for the different flow segments. These parameter sets are calibrated to best fit the flows of the hydrograph within those segments, while other parts of the hydrograph might be less accurately simulated.

Insert Table 4 here.

Statistical indices for the specified objective functions were computed, and iterations with NSE values greater than 0.5 from the last 500 iterations were systematically selected. Subsequently, we scrutinized parameter sets that met the criteria of a KGE value greater than 0.5, PBIAS within the range of -25% to 25%, and RSR averageless than 1.

Ultimately, we derived a parameter sets that reasonably satisfied all objective functions for the average hydrograph and across the different segments of the hydrograph, including high flow, middle flow, and low flow (Appendix 2). Figure 3 below summarizes the sequential modeling approach adopted in this study. Here we want to emphasize that the simulated hydrograph is not stitched together from different flow segments. Each parameter set is used to simulate an entire hydrograph but with portions of the hydrograph either simulated very well or less accurate. The user can then choose which parameter set to run the model with, depending on the desired application (e.g. if model is used to simulate low flows or high flows or all flows).

Insert Figure 3 here.

3. Results

3.1. Model parameterization

Figure 4 illustrates the impact on water yield responses as a default model parameter value is adjusted to a new value. For instance, reducing the default the CN2 value by a value of 15 and increasing the awc value by 50% significantly decreases and increases the monthly basin water yield, respectively. Likewise, modifying other default parameter values results in either an increase or decrease in basin water yield (Figure 4, Appendix 3). By systematically adjusting one parameter at a time and analysing the simulated water balance components, it was possible to refine the maximum and minimum boundaries for each parameter (refer to Appendix 3). Among the 19 sensitivity model parameters that we identified during the preliminary sensitivity analyses, CN2, latq_co, snomeltmax, plaps, and temperature lapse rate tlaps were identified to be the most influential ones on streamflow (Figure 4).

Insert Figure 4 here.

Although all 19 parameters were identified as being sensitive during the preliminary (manual) calibration, some of them did not show clear patterns (Figure 5) in the automatic calibration because the relative sensitivity of each parameter is affected by the values of other parameters and the number of runs (Abbaspour et al. 2017).

Insert Figure 5 here.

3.2. Model Performance

The calibrated SWAT+ model performance was evaluated using multi-metric evaluation techniques that included multi-objective functions and focused model calibration on different segments of the streamflow hydrograph. Calibrating the model in simulating the average hydrograph within the calibration period (2010-2020) resulted in acceptable model performances. Based on Moriasi et al. (2007), we found several parameters sets that achieved adequate model performance ($0.5 \leq \text{NSE} \leq 0.65$) when the NSE was considered as the main objective function in simulating the daily streamflow. Building on Moriasi et al. (2007), our assessment of model performances also incorporated the criteria outlined in Towner et al. (2019), with the KGE serving as the chosen objective function. As per Towner et al. (2019), hydrologic simulations concluding with a KGE value falling within the range of 0.75 to 0.5 ($0.75 \geq \text{KGE} \geq 0.5$) are classified as intermediate, while simulations yielding a KGE value of 0.75 or higher are deemed as good. About 10% of simulations from a total of 500 model runs resulted in satisfactory or above satisfactory model performances when we considered the average hydrograph. However, both high and low flow segments of the hydrographs were consistently underestimated when using a single parameter set to capture the average hydrograph. Middle flows were relatively well captured by the model and achieved the best NSE (Figures 6 & 7). Similarly, optimization of the KGE to the average hydrograph resulted in satisfactory or more than satisfactory model performance with approximately 13% of the parameter sets in the range of adequate model performance. Optimizing for PBIAS resulted in nearly 12 % of the model simulations with an above satisfactory model performance. About 4% of the runs achieved a PBIAS value between -10% and 10% which can be interpreted as having achieved a very good model performance according to Moriasi et al. (2007). About 6% of the model runs resulted in a satisfactory model performance in all four metrics (NSE, KGE, PBIAS, and RSR) for the average hydrograph.

Insert Figure 6 here.

Insert Figure 7 here.

We further examined the best calibrated models for the average hydrograph based on NSE and KGE metrics (Table 4, left side). The parameter set with best overall NSE (0.64) shows the lowest NSE for the lowest flows (-2.73) and highest flows (0.45), but a very high NSE for the middle flows (0.98). Similarly, the parameter set with the best overall KGE (0.74) shows the lowest KGE for the lowest flows (0.23), low flows (0.51) and highest flows (0.54), but a very high KGE for the middle flows (0.91). This suggests that overall model calibration and model performance were largely influenced by the middle flow segment and that the very low flow and very high flow segments were largely omitted during the calibration when the model was calibrated using the average hydrograph (Table 4). Conversely, the best model performance metrics under PBIAS optimization (-0.1) were influenced by high flow segment (-2.9) whereas the low and lowest flow segments had higher absolute PBIAS values (35.8 and 52.6, respectively).

Insert Table 4 here.

According to Knoben et al. (2019), -0.41 is a benchmark value for KGE to decide whether a simulation is acceptable or not, whereas its equivalent value for NSE is zero based on Moriasi et al. (2007). Because the NSE and KGE include different metrics in the objective function, parameter sets that achieve high NSE values do not also automatically achieve high KGE values. For example, in our simulation a KGE value of 0.62 corresponded to the best NSE value (0.64) indicating that having lower KGE values does not mean poor model performance. Based on Moriasi et al. (2007) and Knoben et al. (2019) recommendations, the optimization of KGE objective has shown to yield more acceptable model runs compared to the optimization of the NSE objective. About 40% of the parameter sets resulted in good model performances, whereas only 10% of the parameter sets led to satisfactory model performance in the case of the NSE optimizations. Because the KGE is the modified version of the NSE that includes correlation, variability bias, and mean bias factors (Knoben et al. 2019), the KGE optimization results found in this study are more robust to different segments of the hydrograph. In contrast, the mean streamflow value and coefficient of variation are controlling factors in the NSE optimization that may result in disparity between observed and simulated hydrologic variables in a highly complex watershed with seasonal streamflow dynamics. Because a single objective function cannot comprehensively address these issues of model calibrations, using multi-objective functions and multi-segment calibration in this study yielded a deeper understanding of the model performance. We found that the use of multiple objective functions in evaluating the performance of the SWAT+ model increases the reliability of the calibration and finally resulted in calibrated model that was able to capture the seasonal and inter-annual variations in streamflow that are characteristic for the Feather River watershed.

Calibration and parametrization of the model for different segments of the streamflow hydrograph resulted in significantly improved model performances (Table 4, right side). For example, the NSE value increased from 0.72 to 0.99 for the high flow segment when we calibrated the model parameter sets for the high flow segment independently. Similarly, all other model performance metrics were improved when applying separate parameterization for each of the flow segment.

Figure 8 illustrates the impact of various parameter sets on the predictive efficiency of the SWAT+ model in replicating the high, middle, and low flow segments of the observed streamflow. Calibration of model parameters specifically for high and low flow segments led to enhanced curve fitting in the hydrograph for both high flow and low flow segments, compared to the calibration based on the average hydrograph. However, the middle flow segment showed no significant changes, as it was optimally represented during the calibration based on the average hydrograph (average).

Insert Figure 8 here.

The calibrated model, we achieved by joining the parametrization of multi-objective functions and flow segments, was more than satisfactory for the validation period (Figure 9) than the model that was calibrated to the average hydrograph as per the calibration guideline proposed by Moriasi et al. (2007). The performance metrics improved to 0.62, 0.67, -3.6, and 0.61, for NSE, KGE, PBIAS, and RSR, respectively. However, we found that the overall performance of the model during the calibration period (2010-2020) was higher than during the validation period (1995-2009), This lower model performance for the validation period (Table 5) might be due to changes in precipitation and temperature that occurred during that period since California entered a much drier climate regime in early 2000 (Pierce et al. 2016).

Insert Figure 9 here.

Insert Table 5 here.

3.3. Correlations between parameters and objective functions

Different objective functions were used to evaluate the goodness-of fit of a calibrated model in simulating streamflow. In this study, we evaluated the connection of the multi-objective functions to the model parameters. The results show that optimal parameter values resulted depending on which objective function was chosen. For instance, values of parameters which resulted in the best NSE did not necessarily result in a satisfactory KGE, PBIAS, and RSR at the same time. On one hand, the best NSE solution (0.64) was achieved when the default CN2 value was increased by an absolute value of 6.29. On the other hand, the best KGE solution (0.74) was achieved by increasing the default CN2 by an absolute value of 2.57. Similarly, parameter values which lead to optimal solutions for PBIAS differed from parameter values that resulted in optimal solutions for RSR. A 20% increase on the default value of available water content was required to get the best optimal solution for PBIAS, which is two times greater than the increase required to achieve an optimum solution for RSR. Although there were differences in parameter selections to obtain optimal solutions for NSE, KGE, and PBIAS, the parameter sets that led to best NSE solution also resulted in a good RSR. This means direct optimization of parameter sets for NSE resulted also in an optimum solution for RSR. Figure 10 below demonstrates the relationship between the NSE, RSR, KGE, and PBIAS values that were computed under different model runs and parameter sets. In maximizing the NSE values, the RSR values are minimized on the 1:1 line and the optimum NSE value is achieved at the smallest RSR value. Likewise, higher KGE values are associated with lower RSR values although there is more scatter at lower RSR and higher KGE values and the smallest RSR value does not result in the optimal KGE value. Thus, considering one objective function alone may result in statistically acceptable calibration and validation results, which can affect other objective functions or some part of hydrologic processes. In this study, we found parameter sets that optimized PBIAS to its best solution (value = -0.1) in calibrating the SWAT+. However, further inspection of the observed and simulated streamflow hydrographs as well as the statistical solutions for the different hydrograph components showed significantly high deviation for the peak and low flows (Figure 7 & 8). The parameter set resulting in the highest NSE had an acceptable (absolute) PBIAS value. Because the bias is included in the KGE objective function, the optimal KGE parameter set has an even smaller PBIAS. Overall, we found that the parameter set with highest KGE value also showed good results for all other objective functions. (Tables 4 & 5).

Insert Figure 10 here.

3.4. Objective functions and their connection to the streamflow hydrograph

NSE is one of the most used objective functions in calibrating hydrologic models. The NSE is calculated based on the observed mean and standard deviation and the best calibration solution found in this study was 0.64 with values of 0.62 and -9.3, for KGE and PBIAS, respectively. If model parameters were optimized using KGE as the objective function, the best calibration solution resulted in a KGE of 0.74 with corresponding values of 0.57 and -2.5 for NSE and PBIAS, respectively. Statistical solutions differed when the model was calibrated based on different objective functions and hydrograph segments which indicates that the improvement of one calibration criteria was achieved at the expenses of other calibration criteria and flow segments (Table 4). The optimum parameter set based on KGE and PBIAS reduced the NSE values, highlighting the potential differences in the model parameter selections and the part of hydrographs represented by each of the objective functions. Moreover, NSE and KGE metrics did not correlate linearly and the NSE metric did not relate well to the PBIAS and KGE metrics. However, for a perfect model simulation both KGE and NSE should have a value of 1. When using NSE as the main objective function the model performance is evaluated based on a benchmark value which is the mean value of the observation, while the Euclidean distance from the point of ideal model performance is considered in the case of KGE performance (Eq. 2).

3.5. Comparison of simulated percentiles under different objective functions

For each objective function and calibrated model parameter set the simulated streamflow was compared to the observed streamflow for both the calibration and validation periods using threshold values based on observed flow percentiles (Figure 11 & Figure 12). Variations in the fitted model parameters and objective functions led to significant differences in flow percentiles. For instance, the simulated flow volume value for the 5th percentile flow that corresponding to the flow exceedance probability of 95% ranged from 0 to 26 m^3/s, while the observed value was 18 m^3/s. When considering NSE as the primary objective function, the simulation output based on parameter sets fitted for the low flow segment provided the best representation of the observed value for the 5th percentile. In this case, the simulated value was overestimated by only 2%. Likewise, the parameter sets fitted to the average hydrograph underestimated flows by 7%. Conversely, the simulated flow rates using parameter sets fitted for the middle flow (streamflow volume corresponding to the flow exceedance probability between ***20% – 80%***), and high flows (flow exceedance probability between 5% - 20) segments resulted in underestimations of the flows by 18%, 34%, respectively.

Insert Figure 11 here.

Insert Figure 12 here.

Regarding the flow volume at the 80% flow exceedance probability, the model overestimated this value by 39%, 13%, and 1% when using parameter sets fitted to the high, middle, and low flow segments, respectively. The parameter values fitted for the extremely high and low flow segments led to an underestimation and overestimation of the volume corresponding to the 80% flow exceedance probability by 93% and 35%, respectively. Simulations using fitted parameters for the average hydrograph and middle flows segment resulted in a better representation of the median flow. The observed median flow was underestimated by about 2% and 0.3% when using models calibrated to the average hydrograph and middle flow segment, respectively. Similarly, the third quantile (a flow volume corresponding to the 25% flow exceedance probability) was better estimated by the model calibrated to the average hydrograph, whereas the flow volume at the 5% exceedance probability was best estimated by the high flow model with only a 0.7% underestimation. The middle flow and average flow models overestimated the flow volume at the 5% exceedance probability by 5% and 4%, respectively. These results are consistent with the corresponding model performance values (see Table 4 & Table 5).

Furthermore, the model parameters that led to optimal values of NSE, KGE, and PBIAS for the average hydrograph estimated different flow volumes with various uncertainty levels. A model fitted to the optimal value of NSE resulted in overestimations of the low, high, and very high flow volumes, and underestimation at the 50%, 25%, and 5% exceedance probabilities. In contrast, a model calibrated to the optimal solution of KGE resulted in an underestimation of the low flow (80 to 100% exceedance probabilities) and overestimation of the median flow. A model calibrated to the optimal value of PBIAS led to an overestimation for all flow volumes corresponding to the exceedance probabilities 5 to 100% and underestimation of flow volumes at 0 and 5% exceedance probabilities.

Overall, there were clear differences among the calibrated models for the optimal solutions of different objective functions and flow segments in mimicking the observed flow volumes. Thus, a single objective function and parameter sets might not appropriately represent the hydrologic processes of various flow stages. This could be due to the complex water abstractions, heterogeneity of land use/cover, high seasonal differences, and high variability of model parameters to adequately capture the dynamic hydrologic processes over different seasons. Consequently, it is worth calibrating the SWAT+ model separately for different flow segments using multi-objective functions to achieve a more accurate representation of streamflow and flow volumes.

4. Discussion

Although the new SWAT+ was used to simulate managed streamflow in a complex hydrologic system, characterized by high seasonal differences and frequent extreme flows, calibrating a comprehensive SWAT+ model for Feather River watershed posed challenges.

In our study, we independently parameterized the SWAT+ model for low flow, middle flow, and high flow segments of the hydrograph, leading to significant improvements in model performance. These improvements were achieved by appropriately representing sensitive model parameters. Among the 19 sensitive parameters, CN2 stood out as one of the most influential, causing overestimation or underestimation of flow segments. Its value was adjusted differently for each part of the flow segments.

Initially, when simulating the entire streamflow hydrograph using a calibrated SWAT+ model, underestimation occurred for high and middle flows, while overestimation was observed for low flow segments compared to observed streamflow. To mitigate this, we increased the CN2 value by 13.5, reducing the underestimation from 10.2% to 1.4% for the high flow segment. Similarly, increasing the CN2 value by 9.8 improved the simulation of middle flows and enhanced the goodness of fit for the hydrograph, which was previously underestimated. In contrast, for the low flow segment, which was overestimated during the average simulation, we decreased the CN2 value by 4.6 to achieve an adequate curve fitting and improved model performance. Likewise, the values of other parameters tailored to fit the average hydrographs were adjusted to new values that led to a satisfactory simulation of various flow segments when employing multi-objective functions and parameterization.

Various studies worldwide have demonstrated that updating model parameter values and employing multi-objective functions for different flow segments can enhance the predictive capabilities of hydrologic models. Tegegne et al. (2019) calibrated a SWAT model for two hydro-geographically distinct catchments, one well managed and the other one poorly managed and found that using different CN2 values for different flow components improved the model's ability to predict various streamflow stages. Pfannerstill et al. (2014) reported that calibrating the WAT model based on multiple flow segments improves its performance across different parts of the streamflow hydrograph, including the overall flow hydrograph and very low flows.

While combining flow segments for model calibration can yield a plausible model, the hydrograph and flow duration curves can incorporate both poorly and well-performing segments. Consequently, the overall performance of the model may be more influenced by either the good or poor performing segments. In our study, we obtained credible model performance for the entire streamflow calibration, at least considering the employed objective function. Based on Moriasi et al. (2007) and Towner et al. (2019), the statistical values for the objective functions fell within an acceptable range. However, significant deviations were observed for extreme high and low flows compared to the observed streamflow (Fig. 6 and Fig. 7). The observed discrepancy in the high flow segment may be ascribed to a limitation of the curve number method. The SCS-curve number method used in the SWAT model, as emphasized by Nie et al. (2011), does not consider the duration and intensity of rainfall. Moreover, the intricate water delivery cascading reservoirs system and the complex climate of California further contribute to the discrepancy in the high flow segment. Therefore, independent parameterization for different flow segments can contribute to constructing plausible models, increasing confidence in the use of calibrated models for various purposes. The improvements in the predictive power of the SWAT+ model achieved through calibration for different flow segments were also evident in the percentile flow estimates.

Furthermore, the independently calibrated SWAT+ model can predict future streamflow for high, low, and middle flows by incorporating projected precipitation and temperature data based on various emission scenarios. The differences observed in parameter selection and model performance across different hydrograph segments indicate that employing multiple parameter sets can significantly enhance the accuracy and reliability of predictions for high and low flows in the future. Consequently, the integration of a multi-parameter and multi-segment calibration in SWAT+ model will provide valuable insights into the most probable future conditions within the study catchment. Specially, the calibrated SWAT+ model will be utilized to explore the potential incidences of peak flows and low-flow events. This investigation will involve applying parameter sets tailored for high and low flow segments, considering future climate conditions as outlined by SSP and RCP scenarios. Implementation of multi-segment parameterization and calibration approaches becomes crucial in regions characterized by extreme climate conditions, such as California. These approaches are instrumental in simulating extremes, including droughts and floods, under diverse climate change scenarios, such as the Shared Socioeconomic Pathway (SSP) and the Representative Concentration Pathway (RCPs).

When considering the connections between objective functions and flow segments, optimizing the NSE and RSR resulted in similar improvements of different parts of the flow segments. Maximizing NSE for the entire streamflow led to underestimation of high and very low flow segments, while overestimating middle flows. Similarly, maximizing KGE significantly underestimated the low flow segment. In contrast, optimizing PBIAS resulted in an overestimation of the low flow segment. By calibrating the model using a combination of all objective functions, the middle flow segment was better simulated, with the lowest standardized root mean square error and highest NSE value. Moreover, all statistical values for the other flow segments were acceptable, indicating that combining multiple objective functions improves the overall model performance. However, each individual objective function exhibited slight deterioration compared to the optimal values obtained through individual calibration.

This finding aligns with Guta et al. (2009), and Garcia et al. (2017) who also reported improved simulation of seasonal and annual mean streamflow in TOPMODEL when using combined objective functions.

Comparing the parameter values fitted to the optimal solutions of NSE, KGE, and PBIAS (Table 4) with those fitted to the high, middle, and low segments, slight differences in the Pearson Correlation Coefficient (r) were observed. In both the NSE and KGE cases, parameter values fitted to the middle segment exhibited the highest correlation coefficient (r=0.97), while the low flow parameter values showed the highest correlation coefficient when the PBIAS (r=0.91) was used as objective function. When comparing the parameter correlation between NSE and KGE, NSE and PBIAS, and KGE and PBIAS, NSE and KGE demonstrated the highest correlation (r=0.96), followed by NSE and PBIAS (r=0.95), indicating that NSE and KGE have a similar effect on different parts of the flow segments

5. Conclusion

This study aimed to enhance the simulation of managed streamflow in the Feather River Watershed, Sierra Nevada, California, United States, by exploring multi-objective functions and multi-segment calibration approaches for the SWAT+ model. The study investigated how model parameters varied concerning different objective functions and distinct streamflow segments of the hydrograph (e.g. high, middle, and low flow segments). Model evaluation criteria, including NSE, KGE, PBIAS, and RSR, were employed to assess the simulated managed streamflow. The study's findings lead to the following conclusions:

1) Parameterizing hydrologic models based on different flow segments (e.g. high flow, middle flow, and low flow) of the hydrograph improved the prediction of streamflow compared to using a model calibrated to the average hydrograph.
2) When a single model calibration against the entire streamflow hydrograph was used the middle flow segment of the hydrograph was simulated best while low flow and high flow segments were underestimated in the simulation. Independently fitting parameters for different flow segments significantly enhanced the model's performance for each segment. This suggests that proposing multiple sets of model parameters can increase confidence in constructing a reliable model for managed streamflow predictions.
3) Optimizing model parameters for a single objective function to its optimal value may lead to a deterioration in another objective function. However, considering the hydrograph and flow duration curves can encompass both poorly and well-performing segments. Consequently, the overall performance of the model may be more influenced by either the good or poor-performing segment. Therefore, it is valuable to account for multi-objective functions simultaneously to obtain a credible model that can balance the trade-off between different objective functions.

In general, this study emphasizes the significance of independent flow segment calibration using multi-objective functions to accurately represent flow conditions during wet, average flow, and dry periods. Consequently, parameterizing hydrologic models based on different flow conditions is crucial for constructing reliable hydrologic models. As a result, this study provides valuable insights for water managers and researchers in effectively managing water resources during high and low flow seasons.

Acknowledgements

We would like to acknowledge the California Department of Water Resources for providing the PRISM climate data sources and Lawrence Livermore National Laboratory for the funding support under Project 22-SI-008.

Disclosure statement

The authors declare that they have no relevant financial or non-financial conflict of interests to disclose.

Funding Sources

Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project 22-SI-008.

References

Abbaspour, K.C., Vaghefi, S.A. and Srinivasan, R., 2017. A guideline for successful calibration and uncertainty analysis for soil and water assessment: a review of papers from the 2016 international SWAT conference. Water, 10(1), p.6.

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S.R.I.N.I.V.A.S.A.N.B., Srinivasan, R., Yang, H. and Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of hydrology, 524, pp.733-752.

Alemayehu, T., Gupta, H.V., van Griensven, A. and Bauwens, W., 2022. On the Calibration of Spatially Distributed Hydrologic Models for Poorly Gauged Basins: Exploiting Information from Streamflow Signatures and Remote Sensing-Based Evapotranspiration Data. Water, 14(8), p.1252.

Aliye, M.A., Aga, A.O., Tadesse, T. and Yohannes, P., 2020. Evaluating the performance of HEC-HMS and SWAT hydrological models in simulating the rainfall-runoff process for data scarce Region of Ethiopian Rift Valley Lake Basin.

Open Journal of Modern Hydrology, 10(04), p.105.

Al-Kaisi, M.M. and Lal, R., 2017. Conservation agriculture systems to mitigate climate variability effects on soil health. In Soil health and intensification of agroecosytems (pp. 79-107). Academic Press.

Arnold, J.G., Bieger, K., White, M.J., Srinivasan, R., Dunbar, J.A. and Allen, P.M., 2018. Use of decision tables to simulate management in SWAT+. Water, 10(6), p.713.

Arnold, J.G. and Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling. Hydrological Processes: An International Journal, 19(3), pp.563-572.

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development 1. JAWRA Journal of the American Water Resources Association, 34(1), pp.73-89.

Bailey, R.T., Abbas, S., Arnold, J., White, M., Gao, J. and Čerkasova, N., 2023. Augmenting the National agroecosystem model with physically based spatially distributed groundwater modeling. Environmental Modelling & Software, 160, p.105589.

Barnett, T.P., Adam, J.C. and Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438(7066), pp.303-309.

Bieger, K., Arnold, J.G., Rathjens, H., White, M.J., Bosch, D.D., Allen, P.M., Volk, M. and Srinivasan, R., 2017. Introduction to SWAT+, a completely restructured version of the soil and water assessment tool. JAWRA Journal of the American

Water Resources Association, 53(1), pp.115-130.

Daggupati, P., Yen, H., White, M.J., Srinivasan, R., Arnold, J.G., Keitzer, C.S. and Sowa, S.P., 2015. Impact of model development, calibration, and validation decisions on hydrological simulations in West Lake Erie Basin. Hydrological processes, 29(26), pp.5307-5320.

Dal Molin, M., Kavetski, D., Albert, C. and Fenicia, F., 2023. Exploring Signature-Based Model Calibration for Streamflow Prediction in Ungauged Basins. Water Resources Research, 59(7), p.e2022WR031929.

Donnelly, C., Andersson, J.C. and Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrological Sciences Journal, 61(2), pp.255-273.

DWR (California Department of Water Resources) 2023. The California State Water Project. https://water.ca.gov/programs/state-water-project#

Garcia, F., Folton, N. and Oudin, L., 2017. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrological sciences journal, 62(7), pp.1149-1166.

Gassman, P., Williams, J., Wang, X., Saleh, A., Osei, E., Hauck, L., Izaurralde, R.C. and Flowers, J., 2009. The Agricultural Policy Environmental Extender (APEX) model: An emerging tool for landscape and watershed environmental analyses.

Golmohammadi, G., Prasher, S., Madani, A. and Rudra, R., 2014. Evaluating three hydrological distributed watershed models: MIKE-SHE, APEX, SWAT. Hydrology, 1(1), pp.20-39.

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), pp.80-91.

Guse, B., Pfannerstill, M., Fohrer, N. and Gupta, H., 2020. Improving information extraction from simulated discharge using sensitivity-weighted performance criteria. Water Resources Research, 56(9), p.e2019WR025605.

Hanak, E. and Lund, J.R., 2012. Adapting California's water management to climate change. Climatic change, 111, pp.17-44.

Hargreaves, G.H. and Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Applied engineering in agriculture, 1(2), pp.96-99.

Hedden-Nicely, D.R., 2022. Climate change and the future of western US water governance. Nature Climate Change, 12(2), pp.108-110.

Homer, C., C. Huang, L. Yang, B. Wylie and M. Coan, 2004. Development of a 2001 national land cover database for the United States. Photogrammetric Engineering and Remote Sensing Vol.70,No.7,pp 829-840 or online at www.mrlc.gov/publications

Immerzeel, W.W., Lutz, A.F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B.J., Elmore, A.C. and Emmer, A., 2020. Importance and vulnerability of the world's water towers. Nature, 577(7790), pp.364-369.

Immerzeel, W.W., Van Beek, L.P. and Bierkens, M.F., 2010. Climate change will affect the Asian water towers. science, 328(5984), pp.1382-1385.

Jahn, R. Blume, H. P. Asio, V. B. Spaargaren, O. Schad, P. 2006 Guidelines for Soil

Description, Vol. 1. Food and Agriculture Organization of the United Nations, Rome, Italy.

Kannan, N., Santhi, C., White, M.J., Mehan, S., Arnold, J.G. and Gassman, P.W., 2019. Some challenges in hydrologic model calibration for large-scale studies: A case study of SWAT model application to Mississippi-Atchafalaya River Basin. Hydrology, 6(1), p.17.

Knoben, W.J., Freer, J.E. and Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences, 23(10), pp.4323-4331.

Koczot, K.M., Jeton, A.E., McGurk, B.J. and Dettinger, M.D., 2004. Precipitation-runoff processes in the Feather River Basin, northeastern California, with prospects for streamflow predictability, water years 1971–97. US Geological Survey Scientific Investigations Rep, 5202, p.82.

Koczot, K.M., Markstrom, S.L. and Hay, L.E., 2012. Watershed Scale Response to Climate Change—Feather River Basin, California (No. 2011-3125). US Geological Survey.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology research, 41(3-4), pp.295-319.

Liu, Z., Herman, J.D., Huang, G., Kadir, T. and Dahlke, H.E., 2021. Identifying climate change impacts on surface water supply in the southern Central Valley, California. Science of the Total Environment, 759, p.143429.

Mahmoodi, N., Kiesel, J., Wagner, P.D. and Fohrer, N., 2021. Spatially distributed impacts of climate change and groundwater demand on the water resources in a wadi system. Hydrology and Earth System Sciences, 25(9), pp.5065-5081.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L.,Harmel, R. D.&Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50 (3), 885–900.

Nelson, T., Hui, R., Lund, J. and Medellín–Azuara, J., 2016. Reservoir Operating Rule Optimization for California's Sacramento Valley. San Francisco Estuary and Watershed Science, 14(1).

Nie, W. Yuan, Y. Kepner, W. Nash, M. S. Jackson, M. Erickson, C. 2011 Assessing impacts of landuse and landcover changes on hydrology for the upper San Pedro watershed. Journal of Hydrology 407 (1–4), 105–114.

Pfannerstill, M., Guse, B. and Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. Journal of Hydrology, 510, pp.447-458.

Pierce, D.W., Cayan, D.R. and Dehann, L., 2016. Creating climate projections to support the 4th California climate assessment. University of California at San Diego, Scripps Institution of Oceanography: La Jolla, CA, USA.

PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu, data created 4 February 2014, accessed 28 July 2022. ttps://prism.oregonstate.edu/terms/

Refsgaard, J.C. and Storm, B., 1995. MIKE SHE.[in:] Singh VP (ed.), Computer Models of Watershed Hydrology. Water Resources Publication, Colorado, pp.809-

847.

Siirila-Woodburn, E.R., Rhoades, A.M., Hatchett, B.J., Huning, L.S., Szinai, J., Tague, C., Nico, P.S., Feldman, D.R., Jones, A.D., Collins, W.D. and Kaatz, L., 2021. A low-to-no snow future and its impacts on water resources in the western United States. Nature Reviews Earth & Environment, 2(11), pp.800-819.

Singh, V.P., 1995. Computer models of watershed hydrology. Water Resources Publications.

Tegegne, G., Kim, Y.O., Seo, S.B. and Kim, Y., 2019. Hydrological modelling uncertainty analysis for different flow percentiles: a case study in two hydro-geographically different watersheds. Hydrological Sciences Journal, 64(4), pp.473-489.

Tigabu, T.B., Wagner, P.D., Narasimhan, B. and Fohrer, N., 2023. Pitfalls in hydrologic model calibration in a data scarce environment with a strong seasonality: experience from the Adyar catchment, India. Environmental Earth Sciences, 82(15), p.367.

Tigabu, T.B., Wagner, P.D., Hörmann, G. and Fohrer, N., 2020. Modeling the spatio-temporal flow dynamics of groundwater-surface water interactions of the Lake Tana Basin, Upper Blue Nile, Ethiopia. Hydrology Research, 51(6), pp.1537-1559.

Towner, J., Cloke, H.L., Zsoter, E., Flamig, Z., Hoch, J.M., Bazo, J., Coughlan de Perez, E. and Stephens, E.M., 2019. Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. Hydrology and Earth System Sciences, 23(7), pp.3057-3080.

Trnka, M., Vizina, A., Hanel, M., Balek, J., Fischer, M., Hlavinka, P., Semerádová, D., Štěpánek, P., Zahradníček, P., Skalák, P. and Eitzinger, J., 2022. Increasing available water capacity as a factor for increasing drought resilience or potential conflict over water resources under present and future climate conditions. Agricultural Water Management, 264, p.107460.

Wagner, P.D., Bieger, K., Arnold, J.G. and Fohrer, N., 2022. Representation of hydrological processes in a rural lowland catchment in Northern Germany using SWAT and SWAT+. Hydrological Processes, 36(5), p.e14589.

Westerberg, I.K., Guerrero, J.L., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J.E. and Xu, C.Y., 2011. Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences, 15(7), pp.2205-2227.

Whittaker, G., Confesor, R.E.M.E.G.I.O., Di Luzio, M. and Arnold, J.G., 2010. Detection of overparameterization and overfitting in an automatic calibration of SWAT. Transactions of the ASABE, 53(5), pp.1487-1499.

White, M.J., Arnold, J.G., Bieger, K., Allen, P.M., Gao, J., Čerkasova, N., Gambone, M., Park, S., Bosch, D.D., Yen, H. and Osorio, J.M., 2022. Development of a Field Scale SWAT+ Modeling Framework for the Contiguous US. JAWRA Journal of the American Water Resources Association.

Wu, H., Chen, B., Ye, X., Guo, H., Meng, X. and Zhang, B., 2021. An improved calibration and uncertainty analysis approach using a multicriteria sequential algorithm for hydrological modeling. Scientific Reports, 11(1), p.16954.

Wu, J., Yen, H., Arnold, J.G., Yang, Y.E., Cai, X., White, M.J., Santhi, C., Miao, C. and Srinivasan, R., 2020. Development of reservoir operation functions in SWAT+ for national environmental assessments. Journal of Hydrology, 583, p.12455

Tables

Table 1: Spatial and hydrometeorological data sources and availability used this study.

Data type	Variable	Temporal/spatial resolution	Availability	Data source
Spatial data	Land use and land cover	2001/ 30 m x 30 m	Accessed July 2022	NLCD 2001 (Homer et al. 2004) https://www.mrlc.gov/data/nlcd-2001-land-cover-conus
	Soil	1:50000	Accessed July 2022	https://gdg.sc.egov.usda.gov
	Digital elevation model (DEM)	30 m x 30 m	Accessed July 2022	http://earthexplorer.usgs.gov/
Climate data	Precipitation and temperature (max/min)	Daily, 4 km x4 km	1915-2022	PRISM Obtained from California Department of Water Resources (DWR 2022)
Hydrology	Streamflow	At the outlet point of Oroville Lake	1953-2022	https://waterdata.usgs.gov/nwis/ (USGS 2022)

Table 2: Default decision rule for Oroville reservoir release as available in SWAT+ model (Beiger et al. 2017) on multipurpose use.

Xname	conds	alts	acts									
Oroville	5	7	5									
var	**obj**	**obj_num**	**lim_var**	**lim_op**	**lim_const**	**alt1**	**alt2**	**alt3**	**alt4**	**alt5**	**alt6**	**alt7**

vol	res	0	e-pv	=	-6.761	>	>	>	-	-	-	-
vol	res	0	e-pv	=	0.356	<	<	<	>	>	>	-
vol	res	0	e-pv	=	1.02	-	-	-	<	<	<	>
month	null	0	null	-	6.892	<	-	>	<	-	>	-
month	null	0	null	-	8.043	-	>	<	-	>	<	-

act_typ	obj	obj_num	name	option	const	const2	fp	outcome						
release	res	0	multiple_use_fl	dyrt	233.16170	0.13998	con1	y	y	n	n	n	n	n
release	res	0	multiple_use_nf	dyrt	402.9186	0.31844	con1	n	n	y	n	n	n	n
release	res	0	sfl_cont+mu_fl	dyrt	53.09800	0.31844	con2	n	n	n	y	y	n	n
release	res	0	sfl_cont+mu_nf	dyrt	153.4471	2.34415	con2	n	n	n	n	n	y	n
release	res	0	efc_cont	dyrt	0.69298	5.11667	con3	n	n	n	n	n	n	y

Conditions (conds); alternatives (alts); actions (acts): variable (var);* limit variable (lim_var); limit operator (lim_op); limit constant (lim_const); *action type (act_typ),* const*ant (*const*); file* pointer (fp); storage volume in ha-m(e-pv); day rate (dyrt). Multiple use flood (multiple_use_fl); multiple use non-flood (multiple_use_nf)*; seasonal* flood control multiple use flood (sfl_cont+mu_fl); seasonal flood control + multiple usenon-flood (sfl_cont+mu_nf); emergency flood control (efc_cont).

Table 3: List of sensitive model parameters identified during the manual calibration step that were subsequently calibrated during the automatic calibration steps for the high flow, middle flow, and low flow segments based on multi criteria evaluation statistics. *Parameters* highlighted in *bold were calibrated separately.*

				Calibrated parameter values for different segments of streamflow hydrographs and objective functions											
	Value ranges			NSE				KGE				PBIAS			
Parameter	Min	Max	Change method	Average	High	Middle	Low	Average	High	Middle	Low	Average	High	Middle	Low
snomelt_tmp	-5	5	replace	2.90	0.42	0.11	2.32	1.25	3.80	0.11	2.32	0.11	3.16	3.80	0.12
snofall_tmp	-5	5	replace	1.07	2.38	-2.20	-2.93	0.10	1.35	-2.20	-2.93	-2.65	-1.94	1.35	-2.54
snomelt_max	0	10	replace	3.15	2.19	6.83	4.57	5.99	4.01	6.83	4.57	3.10	3.77	4.01	6.64
snomelt_min	0	10	replace	1.99	0.85	4.73	3.38	3.98	1.82	4.73	3.38	1.82	3.48	1.82	4.24
snomelt_lag	0	1	replace	0.47	0.79	0.72	0.42	0.48	0.43	0.72	0.42	0.64	0.62	0.43	0.44
plaps	0	200	replace	19.65	26.42	22.92	26.07	17.10	24.60	22.92	26.07	31.33	30.56	24.60	29.03
tlaps	-10	10	replace	4.00	-2.42	7.93	-9.69	6.01	-0.30	7.93	-9.69	7.88	-0.23	-0.30	-6.06
cn2	-20	20	relative change	6.29	13.53	9.76	-4.64	2.58	4.84	9.76	-4.64	14.91	0.58	4.84	1.40
esco	0	1	replace	0.28	0.15	0.01	0.56	0.42	0.25	0.01	0.56	0.45	0.54	0.25	0.52
epco	0	1	replace	0.79	0.71	0.36	0.44	0.49	0.72	0.36	0.44	0.68	0.34	0.72	0.43
lat_ttime	0	180	replace	48.97	44.76	69.37	70.36	40.54	31.17	69.37	70.36	46.14	8.93	31.17	43.39
perco	0	1	replace	0.18	0.14	0.21	0.20	0.17	0.14	0.21	0.20	0.20	0.16	0.14	0.17
awc	0	1	relative change	0.08	0.09	0.44	0.20	0.32	0.33	0.44	0.20	0.20	0.46	0.33	0.34
alpha	0	1	replace	0.30	0.39	0.11	0.08	0.16	0.54	0.11	0.08	0.48	0.65	0.54	0.38

flo_min	0	50	replace	15.24	14.65	17.58	31.79	19.65	28.06	17.58	31.79	14.53	22.85	28.06	21.67
revap	0	1	replace	0.07	0.05	0.06	0.04	0.03	0.05	0.06	0.04	0.09	0.03	0.05	0.03
rchg_dp	0	1	replace	0.29	0.25	0.42	0.60	0.21	0.25	0.42	0.60	0.42	0.74	0.25	0.47
revap_min	0	500	replace	18.97	14.80	10.15	17.87	7.36	15.37	10.15	17.87	14.07	24.34	15.37	22.87
latq_co	0	1	replace	0.10	0.16	0.16	0.39	0.06	0.14	0.16	0.39	0.43	0.18	0.14	0.51

*Long name for model parameters:: cn2~ Condition II curve number, lat_ttime ~exponential of the lateral flow travel time, esco~soil evaporation compensation factor, epco~ plant water uptake compensation factor, snomelt_tmp ~ snowmelt temperature, snofall_tmp~ snow fall temperature, snomelt_max~maximum snowmelt temperature, snomelt_min~ minimum snowmelt temperature, snomelt_lag~ perco~ percolation coefficient; awc ~ available water capacity of soil layeralpha~ Condition II curve number, revap~threshold depth of water in shallow aquifer required to allow revap to occur, rchg_dp~recharge to deep aquifer (the fraction of root zone percolation that reaches the deep aquifer, revap_min~water table depth for revap to occur water table depth for revap to occur, flo_min~ plaps~ precipitation lapse rate: mm per km of elevation difference, tlaps~temperature lapse rate: deg C per km of elevation difference, latq_co~Plant ET curve number coefficient, replace ~ absolute value, relative change ~ add/subtract the value to the default one.. The change methods relative indicate add/subtract the value on the default and replace indicates replace the default value by the absolute value of the new number.

Table 4: Optimized objective function values as calibrated based on the average hydrograph, very high, high, middle, low, very low flow segments of the hydrographs. Values of interest are highlighted in bold.

Parameters adjusted based on the average hydrograph					Parameters adjusted based on flow segments			
Model parametrization based on best NSE								
Hydrograph segment	NSE	PBIAS	KGE	RSR	NSE	PBIAS	KGE	RSR
Average	**0.64**	-9.30	0.62	0.59				
Very high flow	0.45	-28.80	0.35	0.83	**0.88**	0.74	0.70	0.35
High flow	0.72	-10.20	0.79	0.63	**0.99**	0.20	0.97	0.10
Middle flow	**0.98**	**7.00**	**0.91**	**0.14**	**0.99**	-0.50	0.98	0.06
Low flow	0.56	0.60	0.36	0.83	**0.99**	-0.10	0.98	0.05
Very low flow	-2.73	-21.10	0.50	0.76	**0.96**	-1.50	0.91	0.19
Model parametrization based on best KGE								
Average	0.57	-2.5	**0.74**	0.65				
Very high flow	0.77	-7.9	0.54	0.48	0.88	-3.8	**0.88**	0.35
High flow	0.76	9.0	0.79	0.48	0.97	-3.3	**0.96**	0.18
Middle flow	**0.98**	**-4.8**	**0.91**	**0.14**	0.99	-0.5	**0.98**	0.06
Low flow	-3.63	-39.7	0.51	2.15	0.99	-0.1	**0.98**	0.05
Very low flow	-29.22	-61.7	0.23	5.48	0.84	-3.2	**0.96**	0.34
Parametrization based on best PBIAS								
Average	0.57	**-0.10**	0.65	0.66				
Very high flow	0.56	-18.40	0.39	0.66	0.82	**0.20**	0.60	0.41
High flow	**0.97**	**-2.90**	**0.95**	**0.18**	0.79	**-0.20**	0.56	0.46
Middle flow	0.93	13.60	0.86	0.25	0.96	**0.20**	0.94	0.08
Low flow	-2.79	35.80	0.51	1.94	0.95	**0.83**	0.00	0.22
Very low flow	-20.89	52.60	0.38	4.67	0.94	**-1.00**	0.81	0.25

Table 5: Optimized objective function values as validated based on the average, high, middle, low flow segments of the hydrographs. Values of interest are highlighted in bold.

Model parametrization based on best NSE					**Model parametrization based on best KGE**			
	NSE	**PBIAS**	**KGE**	**RSR**	**NSE**	**PBIAS**	**KGE**	**RSR**
Average	**0.62**	-3.60	0.67	0.61	0.52	2.50	**0.74**	0.69
High flow	**0.92**	-3.20	0.77	0.28	0.25	17.50	**0.76**	0.86
Middle	**0.97**	7.50	0.90	0.18	0.95	-0.20	**0.78**	**0.22**
Low	**-2.63**	-23.10	-0.01	1.90	-26.03	-67.40	**0.22**	5.20
Model parametrization based on best PBIAS					**Parameters optimized to the best of NSE for high flow segment**			
Average	0.50	**-18.60**	0.52	0.70	0.54	3.10	0.67	0.68
High	0.19	**-17.90**	0.71	0.90	**0.92**	**5.20**	**0.87**	**0.29**
Mid	0.92	**-11.70**	0.80	0.28	0.95	11.30	0.89	0.22
Low	-8.41	**-38.70**	0.16	3.07	-0.29	-2.80	-0.11	1.14
Parameters optimized to the best of NSE for middle flow					**Parameters optimized to the best of NSE for low flow**			
Average	**0.51**	2.90	0.68	0.70	**0.41**	-23.30	0.34	0.77
high	**0.84**	8.00	0.78	0.40	**-1.85**	-33.20	0.60	1.69
mid	**0.93**	**8.30**	**0.68**	**0.26**	**0.82**	-6.90	-0.77	0.42
low	**-1.96**	-21.30	0.38	1.72	**0.41**	**23.25**	**0.34**	**0.77**

Figures

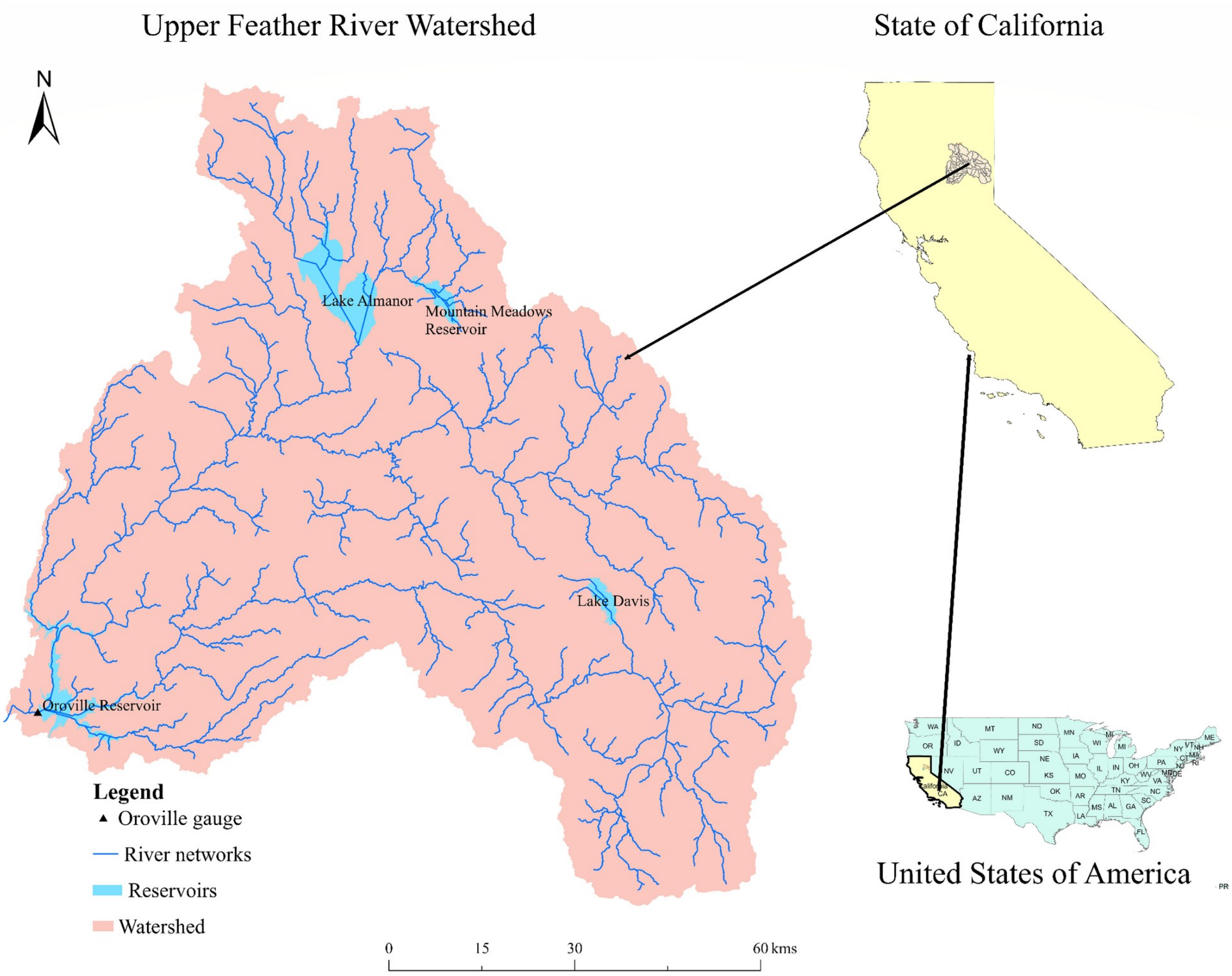


Figure 1: Location map of the Upper Feather River Watershed within Sierra Nevada Mountains, California, USA; reservoirs, river networks, and boundary of modeling Watershed.

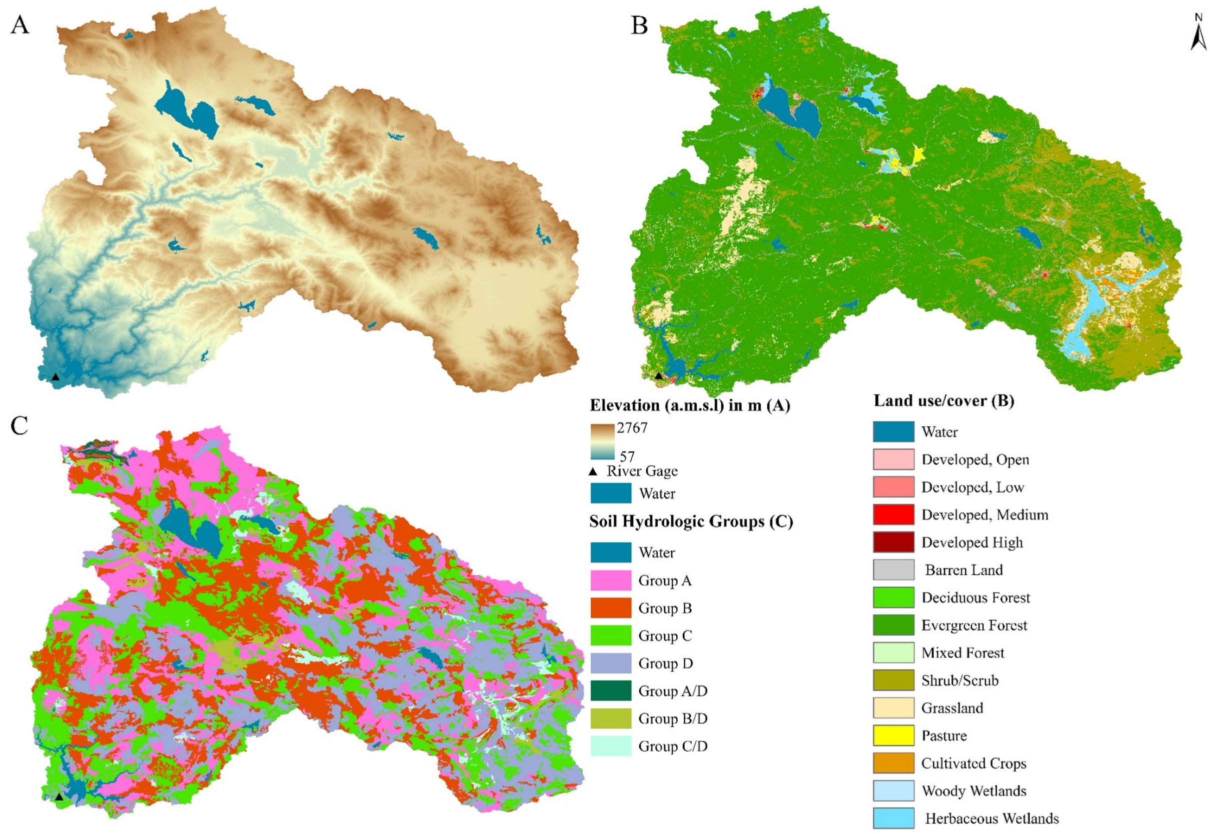


Figure 2: Spatial SWAT+ model input data including Digital Elevation Model (DEM) showing topography (A), land use/land cover, and the soil hydrologic groups within the Upper Feather River watershed, in the Sierra Nevada Mountains, California, USA.

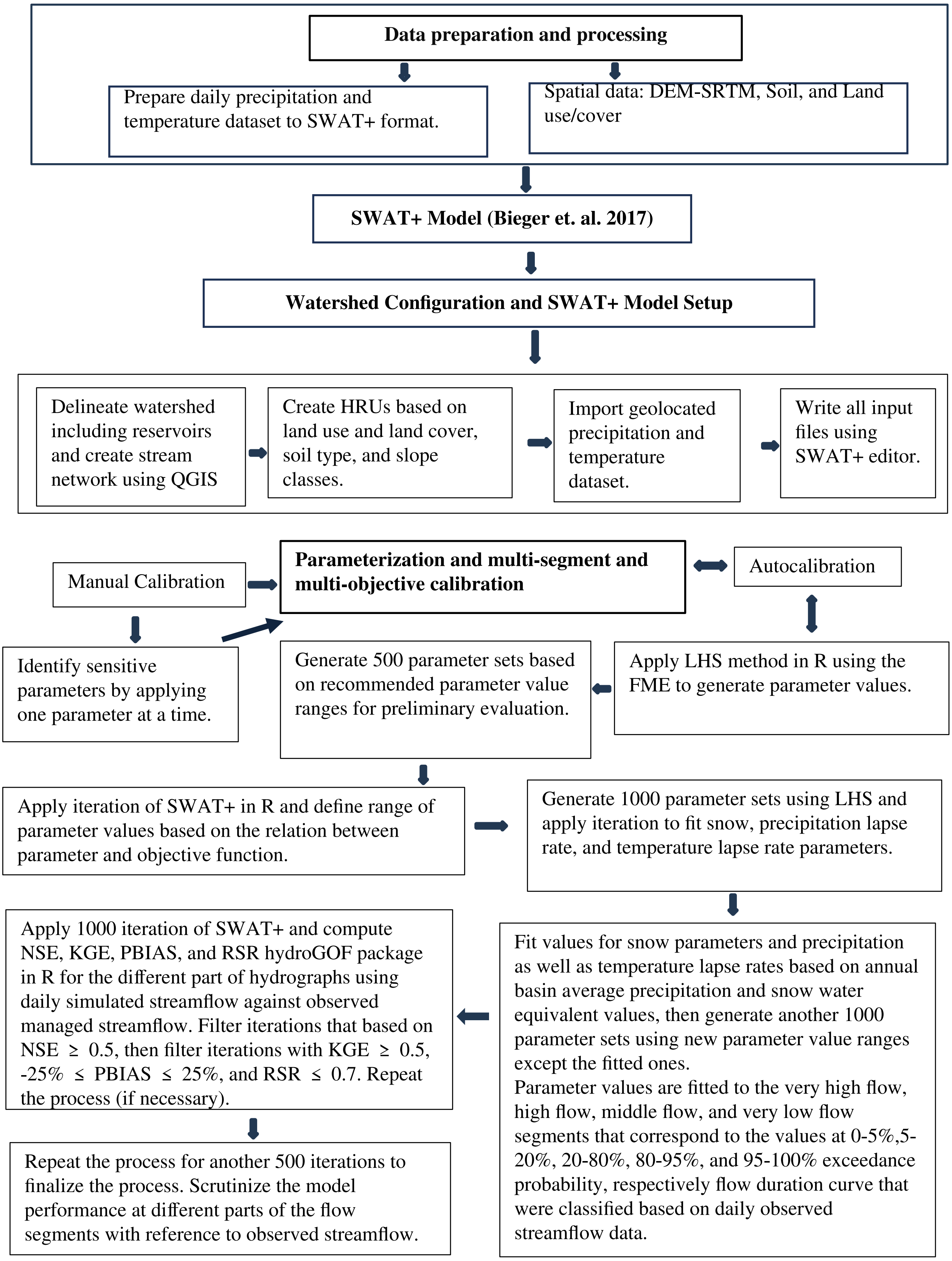

Data preparation and processing
Prepare daily precipitation and temperature dataset to SWAT+ format.
Spatial data: DEM-SRTM, Soil, and Land use/cover
SWAT+ Model (Bieger et. al. 2017)
Watershed Configuration and SWAT+ Model Setup
Delineate watershed including reservoirs and create stream network using QGIS
Create HRUs based on land use and land cover, soil type, and slope classes.
Import geolocated precipitation and temperature dataset.
Write all input files using SWAT+ editor.
Manual Calibration
Parameterization and multi-segment and multi-objective calibration
Autocalibration
Identify sensitive parameters by applying one parameter at a time.
Generate 500 parameter sets based on recommended parameter value ranges for preliminary evaluation.
Apply LHS method in R using the FME to generate parameter values.
Apply iteration of SWAT+ in R and define range of parameter values based on the relation between parameter and objective function.
Generate 1000 parameter sets using LHS and apply iteration to fit snow, precipitation lapse rate, and temperature lapse rate parameters.
Apply 1000 iteration of SWAT+ and compute NSE, KGE, PBIAS, and RSR hydroGOF package in R for the different part of hydrographs using daily simulated streamflow against observed managed streamflow. Filter iterations that based on NSE ≥ 0.5, then filter iterations with KGE ≥ 0.5, -25% ≤ PBIAS ≤ 25%, and RSR ≤ 0.7. Repeat the process (if necessary).
Fit values for snow parameters and precipitation as well as temperature lapse rates based on annual basin average precipitation and snow water equivalent values, then generate another 1000 parameter sets using new parameter value ranges except the fitted ones.
Parameter values are fitted to the very high flow, high flow, middle flow, and very low flow segments that correspond to the values at 0-5%,5-20%, 20-80%, 80-95%, and 95-100% exceedance probability, respectively flow duration curve that were classified based on daily observed streamflow data.
Repeat the process for another 500 iterations to finalize the process. Scrutinize the model performance at different parts of the flow segments with reference to observed streamflow.

Figure 3: This flowchart outlines the steps involved in the modeling process, including data preparation, watershed configuration, manual calibration, and multi-segment multi-objective function parameterization and calibration. It emphasizes the iterative nature of the calibration process to enhance model performance and confidence.

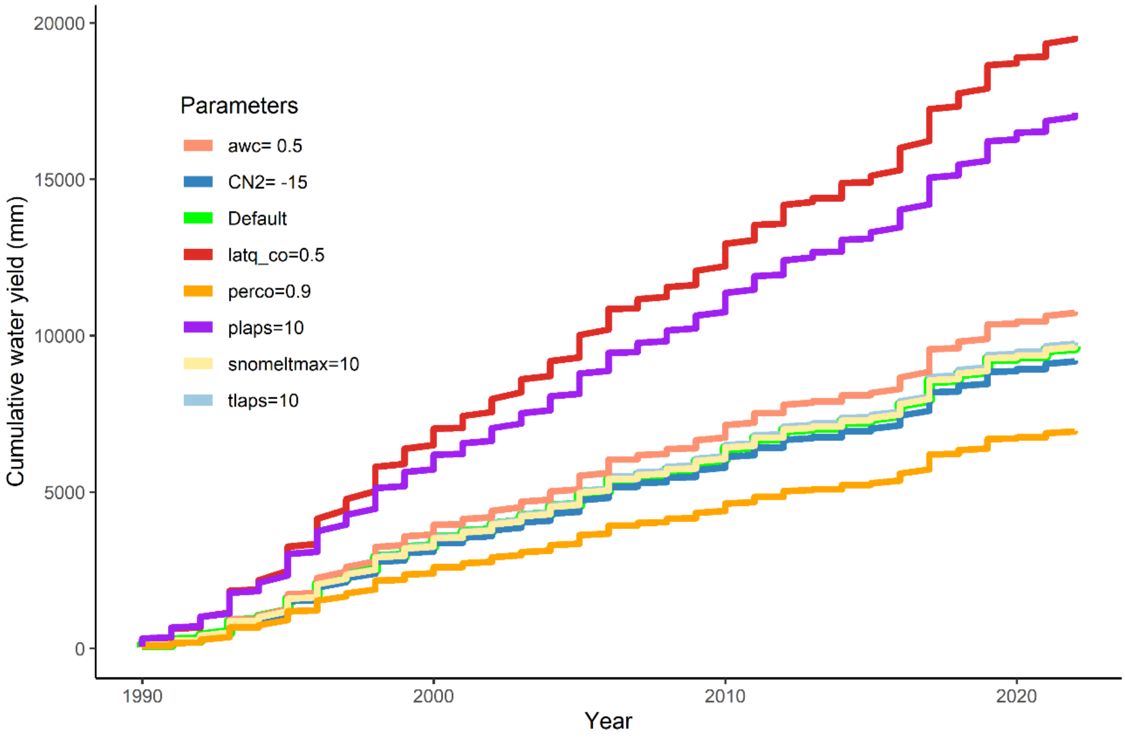


Figure *4*: Example graph showing the effects of different SWAT+ parameters on the simulated commutative monthly water yield values as compared to the default simulation. Each graph was generated from a monthly simulation output based on one parameter change at a time.

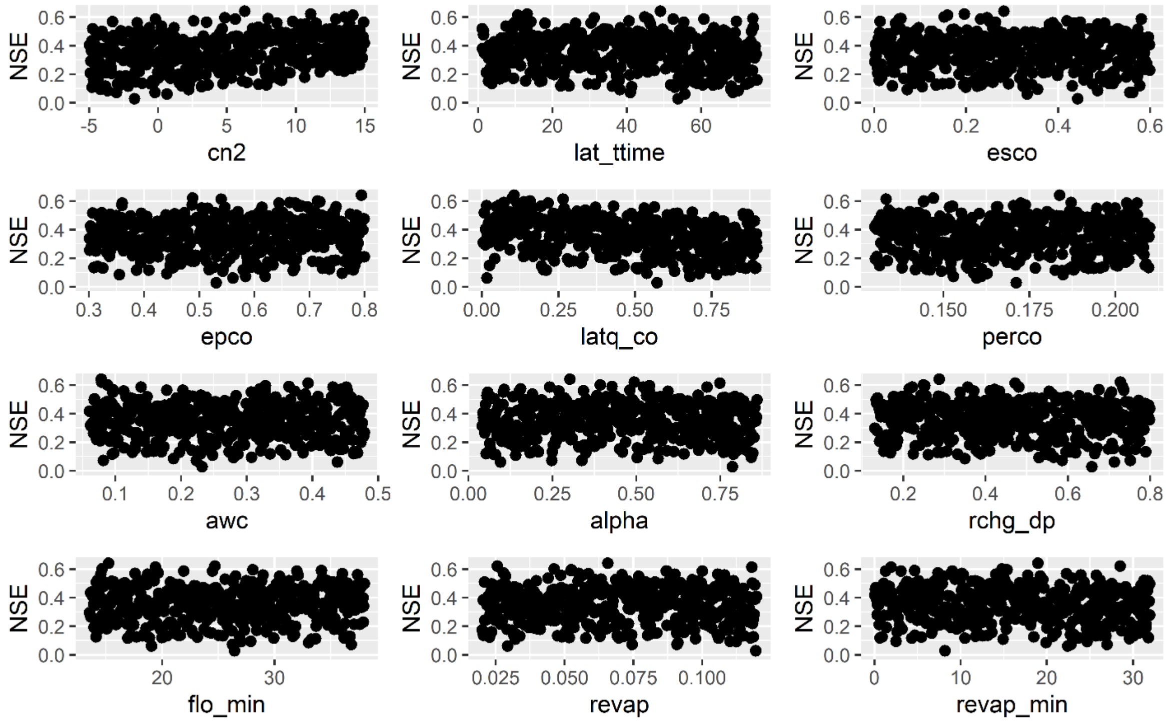


Figure 5: Scatter plots showing the association between NSE (as computed using simulated streamflow against observed managed streamflow records) and values of different parameters.

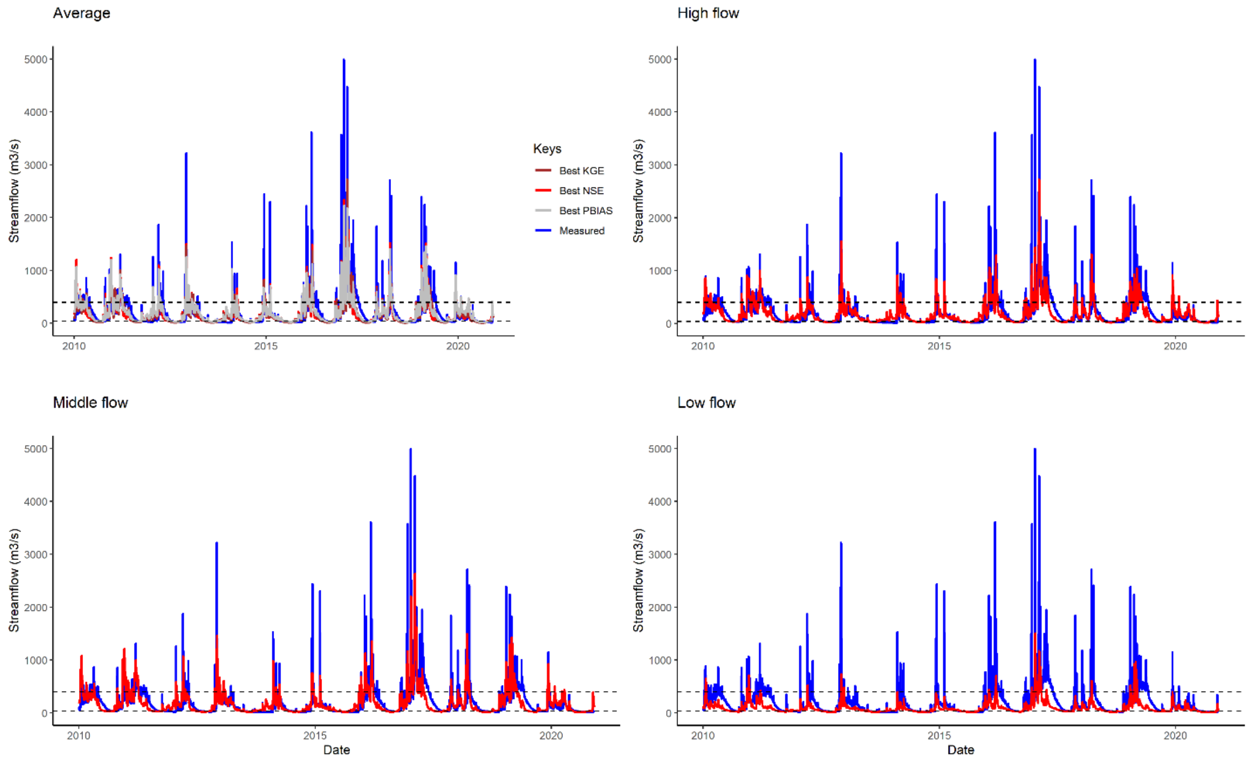


Figure 6: Hydrographs showing the comparison of daily observed streamflow data versus simulated streamflow under different calibration options that optimized model parameters based on the average hydrograph, or individual high flow, middle flow, and low flow segments that were independently optimized.

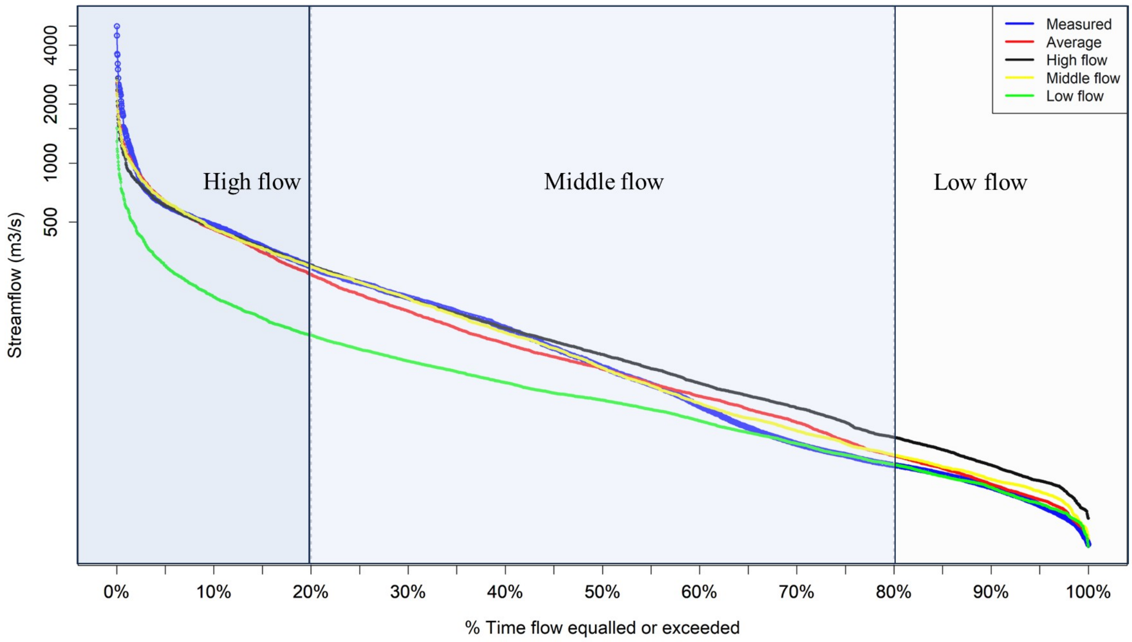


Figure 7: Flow duration curves of daily observed streamflow data versus simulated streamflow under different calibration options such as all flows (Average), high flow, middle flow, and low flow segments considered independently, and multi-objective functions are optimized.

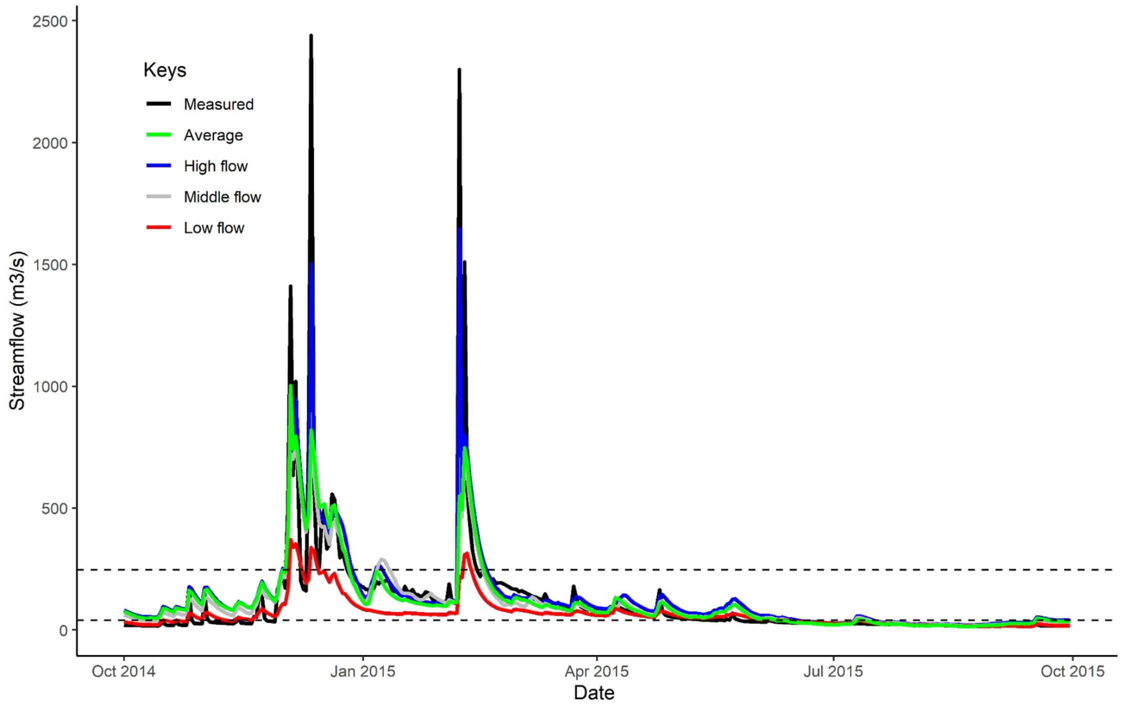


Figure 8: Hydrographs showing the comparison of daily observed streamflow data in water year 2014 versus streamflow simulated with parameter sets that are optimized to best capture either all flows (Average), or the high flow, middle flow, and low flow segments. The dashed lines indicate the streamflow thresholds that were used to optimize the high flow (> 240 m^3/s), middle flow (40 to 240 m^3/s), and low flow (<40 m^3/s) segments.

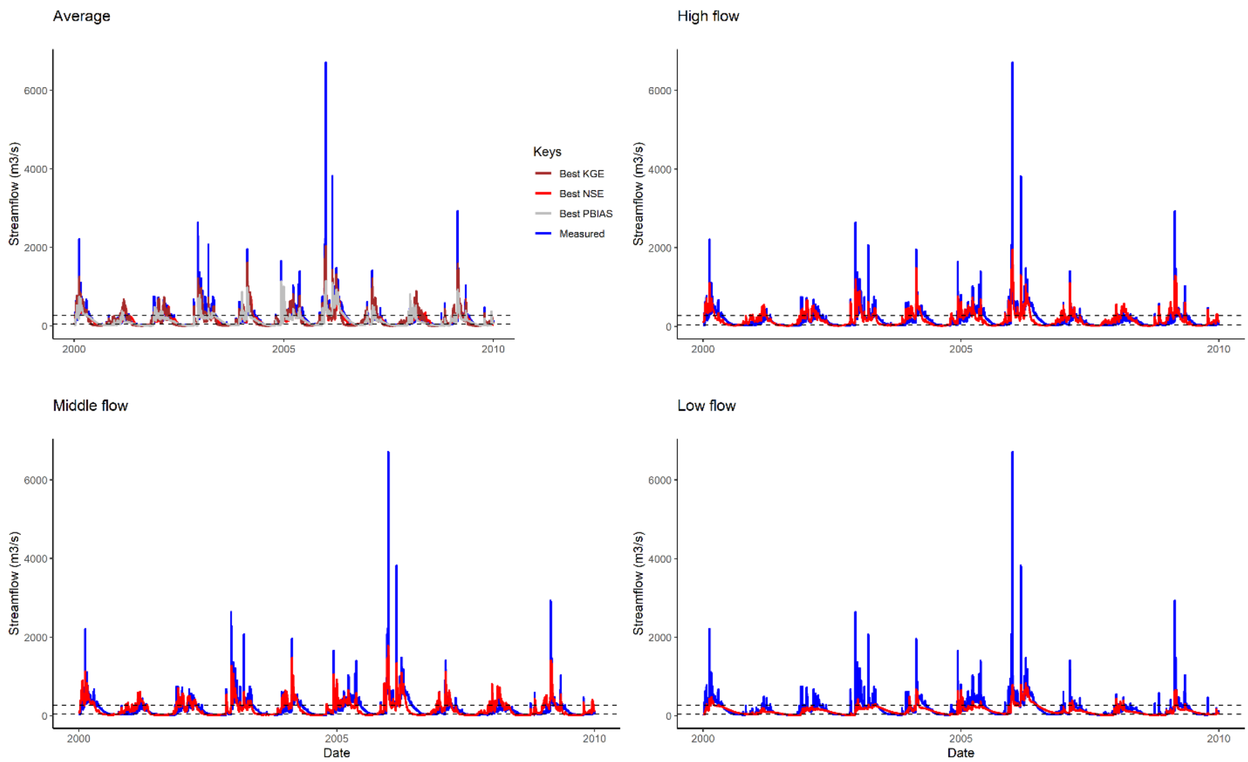


Figure 9: Hydrographs showing the comparison of daily observed streamflow data versus simulated streamflow for the validation period (2000-2009) which resulted from combining multi-objective functions (NSE, KGE, and PBIAS) with the calibration of different flow segments based on for the average hydrograph or individual high flow, middle flow, and low flow segments that were independently optimized.

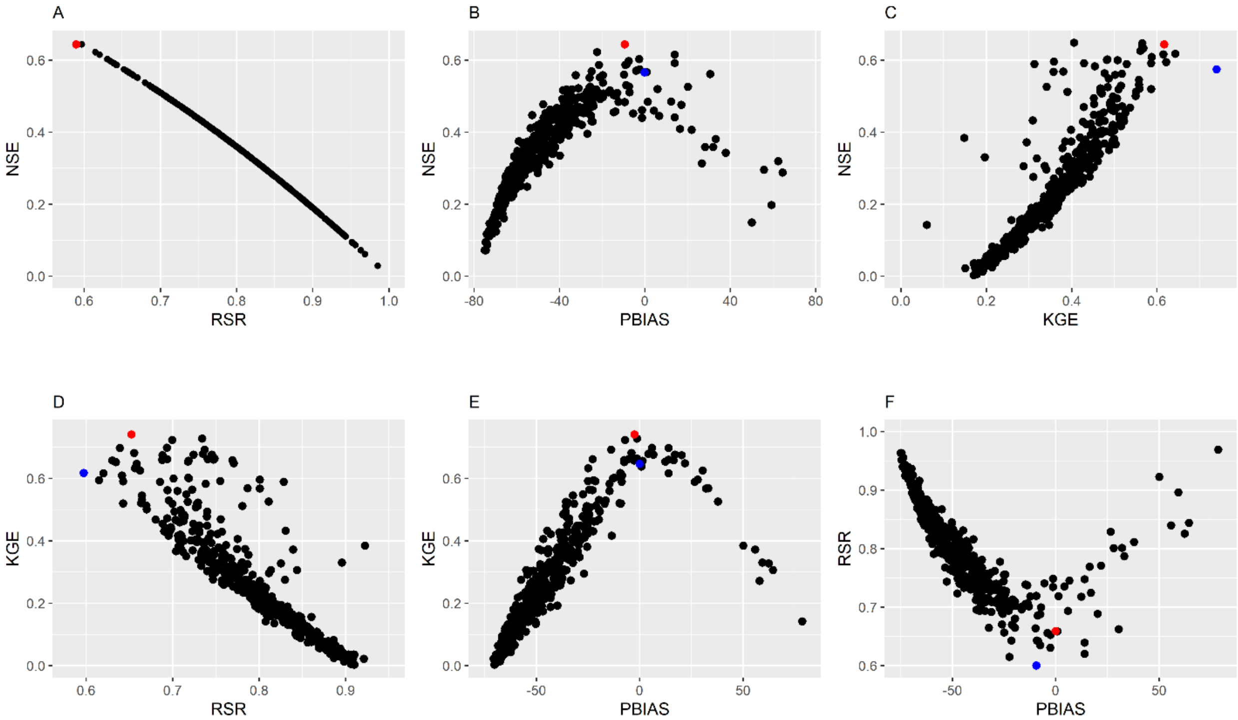


Figure 10: Scatterplot showing the correlation of different objective functions. The red points represent the best optimized values for each pair of objective functions with the target objective function shown on the vertical axis and the blue points represented the best values achieved based on the target objective function shown on the horizontal axis.

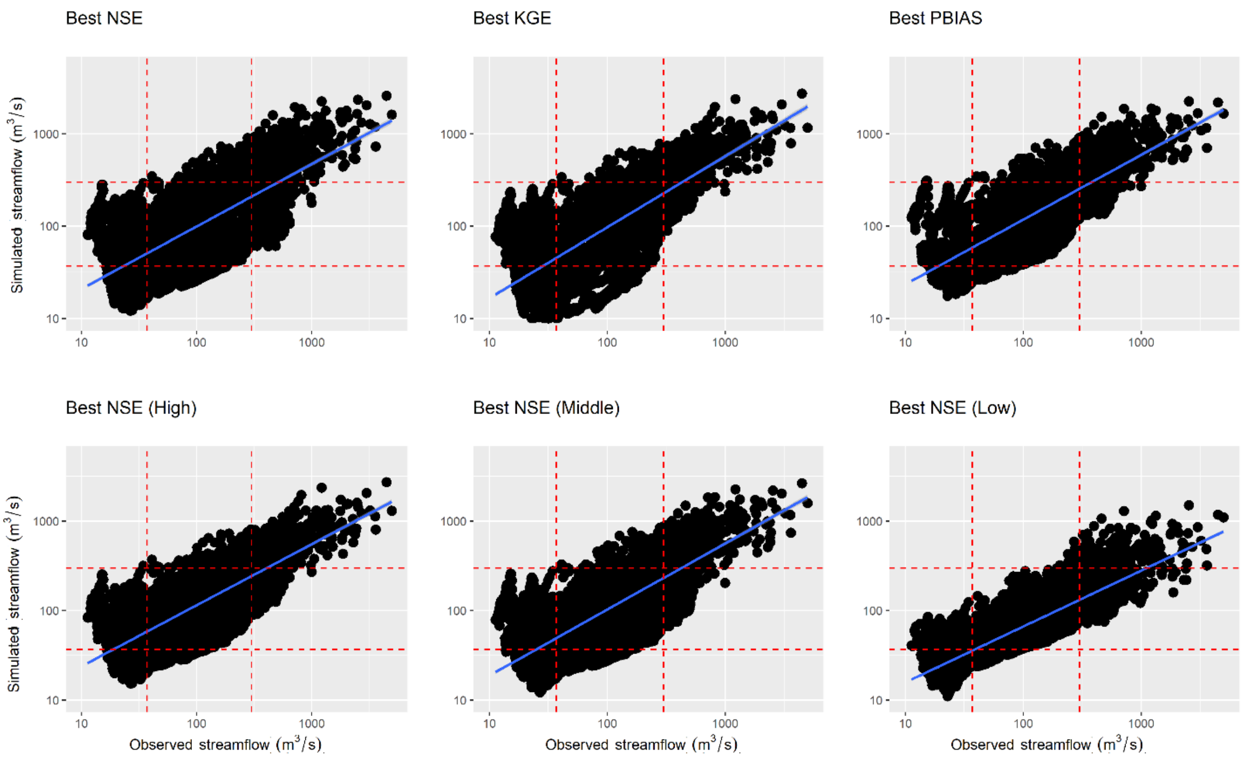


Figure 11: Scatterplot in the log10 scale showing the observed managed streamflow at Oroville gauging station against SWAT+ simulated streamflow at the same gauging station parameter sets optimized to the best of NSE, KGE, PBIAS, High flow, Middle flow, and Low flow for the water year period between 2010-2020 (calibration period). The red broken lines indicate threshold values of low and high flows.

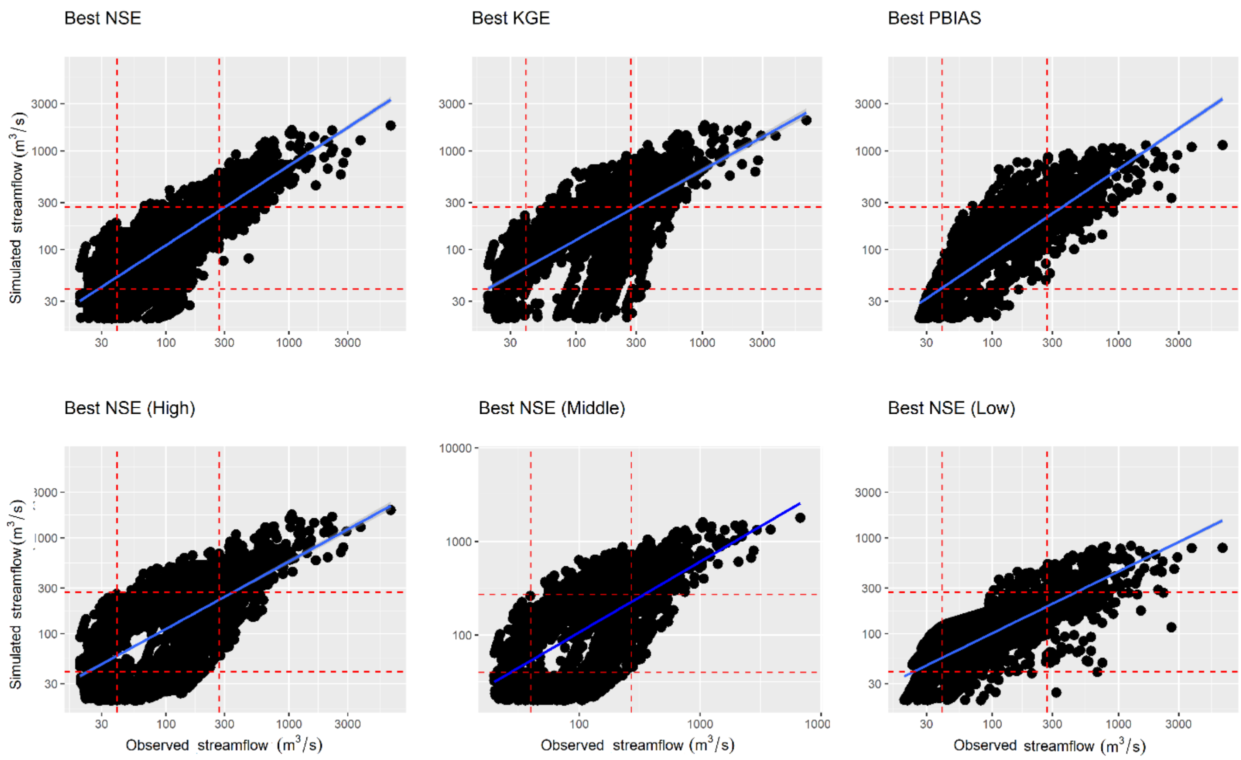


Figure 12: Scatterplot in the log10 scale showing the observed managed streamflow at Oroville gauging station against SWAT+ simulated streamflow at the same gauging station parameter sets optimized to the best of NSE, KGE, PBIAS, High flow, Middle flow, and Low flow for the water year period between 2000-2009 (validation period).

Appendices

Appendix 1: Parameter value ranges that were used for LHS resulted in 500,1000, 1000, and 500 parameter sets during the four phases of automatic calibration. The first phase of the 500 iterations were carried out to determine the minimum and maximum value ranges for 19 sensitive parameters, the next phase of 1000 iterations were carried out to calibrate snowmelt, snowfall, temperature, and precipitation lapse rate parameters, whereas the next 1000 and 500 iterations were used to calibrate other parameters.

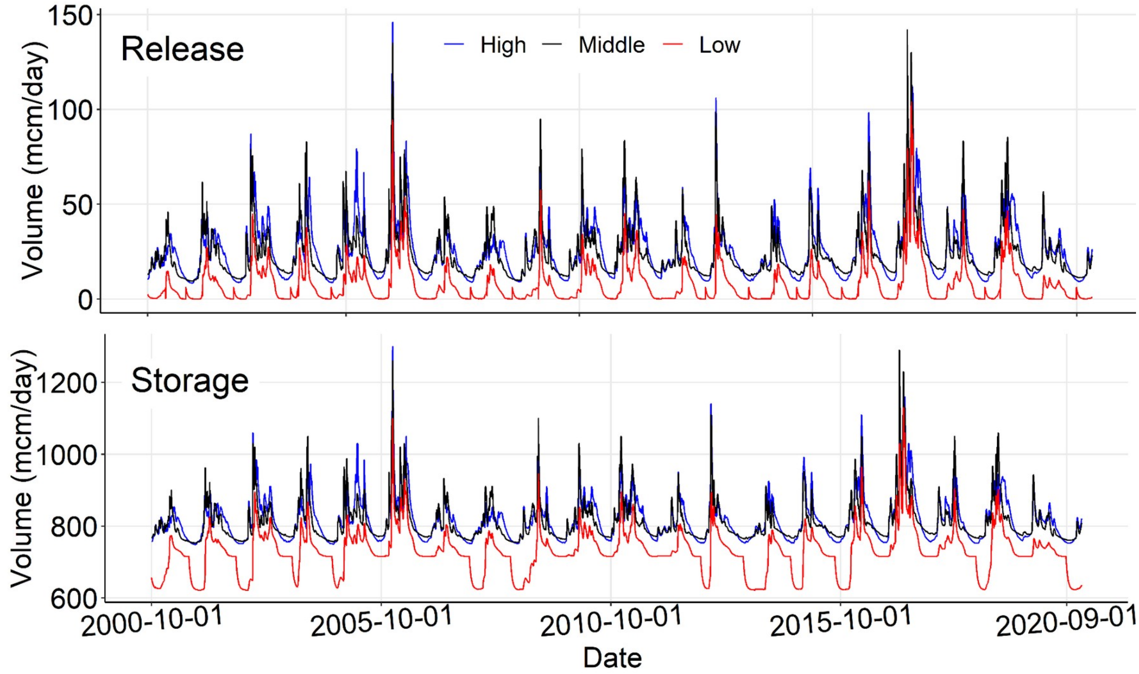
		Allowable range		First 500 runs (preliminary phase)		Second 1000 runs		Third 1000 runs		Fourth 500 runs (Final phase)		
Parameter	Default	Absolute min	Absolute max	Min	Max	Min	Max	Min	Max	Min	Max	change method
snomelt_tmp	*1*	*-5*	*5*	*-5.00*	*5.00*	*-5.00*	*5.00*					*replace*
snofall_tmp	*0.5*	*-5*	*5*	*-5.00*	*5.00*	*-5.00*	*3.75*					*replace*
snomelt_max	*4.5*	*0*	*10*	*0.00*	*10.00*	*-5.00*	*9.00*					*replace*
snomelt_min	*4.5*	*0*	*10*	*0.00*	*10.00*	*0.00*	*9.00*	*Calibrated independently in previous step*				*replace*
snomelt_lag	*1*	*1*	*1*	*0.00*	*1.00*	*0.00*	*1.00*					*replace*
plaps	*0*	*0*	*200*	*0.00*	*200.00*	*0.00*	*75.00*					*replace*
tlaps	*-10*	*-10*	*10*	*-10.00*	*10.00*	*-10.00*	*10.00*					*replace*
cn2	variable [35,98]	35.00	95.00	-20.00	20.00	default	default	-15.00	15.00	-5.00	15.00	relative
esco	0.95	0.00	1.00	0.00	1.00	0.95	0.95	0.00	0.60	0.00	0.60	replace
epco	1.00	0.00	1.00	0.01	1.00	1.00	1.00	0.30	0.80	0.30	0.80	replace
lat_ttime	0.00	0.50	180.00	0.00	180.00	0.00	0.00	1.00	75.00	1.00	75.00	replace
awc	variable [0,1]	0.01	1.00	0.01	0.50	default	default	0.06	0.48	0.06	0.48	relative change
perco	0.90	0.00	1.00	0.00	1.00	0.90	0.90	0.13	0.21	0.13	0.21	replace
alpha	0.05	0.00	1.00	0.00	1.00	0.05	0.05	0.04	0.86	0.04	0.86	replace
flo_min	3.00	0.00	50.00	0.00	50.00	3.00	3.00	13.50	40.00	13.50	38.00	replace
revap	0.02	0.02	0.20	0.02	0.20	0.02	0.02	0.02	0.12	0.02	0.12	replace
rchg_dp	0.05	0.00	1.00	0.00	1.00	0.05	0.0.5	0.13	0.80	0.13	0.80	replace
revap_min	5.00	0.00	500.00	0.00	500.00	5.00	5.00	0.00	32.00	0.00	32.00	replace
latq_co	0.01	0.00	1.00	0.00	1.00	0.01	0.01	0.00	0.90	0.00	0.90	relative

Appendix 3: Displays an optimization matrix designed to facilitate the integration of multi-parameters and multi-segment calibration, resulting in final parameter sets for running the model for the intended purpose. Each run encompasses multiple sets of parameter values and statistical indices for both the average and different flow segments. For brevity, this table excludes RSR and PBIAS computations for the average and other flow segments, considering space constraints.

						Flow segments separated based on historical streamflow threshold values									
		Average hydrograph				NSE for the different segments					KGE for the different segments				
run	Objective function	NSE	PBIAS	KGE	RSR	Very high	High	Mid	Low	Very low	Very high	High	Mid	Low	Very low
121	NSE	**0.64**	-9.30	0.62	0.60	0.45	0.72	0.98	0.56	-2.73	0.35	0.79	0.91	0.36	0.50
56	PBIAS	0.57	**-0.10**	0.65	0.66	0.56	0.97	0.94	-2.79	-20.89	0.39	0.95	0.86	0.51	0.38
454	KGE	0.57	-2.50	**0.74**	0.65	0.77	0.76	0.98	-3.63	-29.22	0.54	0.79	0.91	0.51	0.23
121	RSR	0.64	-9.30	0.62	**0.60**	0.45	0.72	0.98	0.56	-2.73	0.35	0.79	0.91	0.36	0.50
408	Very_high_NSE	0.31	-44.70	0.43	0.83	**0.88**	-2.64	-1.44	-24.53	-78.71	0.70	0.46	-0.04	-0.18	NA
340	High_NSE	0.60	-2.60	0.66	0.63	0.58	**0.99**	0.97	-4.18	-9.30	0.41	0.96	0.89	0.09	0.45
136	Mid_NSE	0.57	-4.00	0.68	0.66	0.66	0.98	**0.99**	0.42	-1.32	0.45	0.94	0.98	0.74	0.54
194	Low_NSE	0.31	-49.60	0.17	0.83	-0.29	-6.44	-0.05	**0.99**	0.82	0.10	0.26	0.21	0.98	0.94
303	Very_low_NSE	0.42	-29.50	0.39	0.76	0.25	-1.73	0.66	0.73	**0.96**	0.28	0.58	0.51	0.53	0.91
408	Very_high_KGE	0.31	-44.70	0.43	0.83	0.88	-2.64	-1.44	-24.53	-78.71	**0.70**	0.46	-0.04	-0.18	NA
102	High_KGE	0.60	-7.50	0.65	0.63	0.57	0.97	0.99	0.32	-5.86	0.40	**0.97**	0.94	0.29	0.70
136	Mid_KGE	0.57	-4.00	0.68	0.66	0.66	0.98	1.00	0.42	-1.32	0.45	0.94	**0.98**	0.74	0.54
194	Low_KGE	0.31	-49.60	0.17	0.83	-0.29	-6.44	-0.05	1.00	0.82	0.10	0.26	0.21	**0.98**	0.94
281	Very_low_KGE	0.46	-32.7.	0.37	0.73	0.10	-2.16	0.62	0.84	0.89	0.22	0.51	0.50	0.66	0.96

Appendix 3: Displays how changing the default value of a parameter affects water yield. The comparisons is between the simulation using the default parameters and simulations done by changing the value of a single parameter at a time.

Parameter	Default value	Minimum	Maximum	Change method	Changed value	% change on the default para value	% change in water yield
CN2	variable	35	98	relative	-15.00	variable	-14.29
tlaps	6.5	0	10	replace	10.00	53.85	1.22
latq_co	0.01	0	1	replace	0.50	98.00	102.86
awc	variable	0	1	relative	0.50	variable	11.43
snomelt_max	4.5	0	10	replace	10.00	122.22	0.12
perco	1	0	1	replace	0.90	10.00	27.96
plaps	0	-25	25	replace	10.00	infinity	77.55



Appendix 4: Daily volume of released and stored water from/in the Oroville Reservoir as simulated using calibrated SWAT+ models for high, middle, and low flows simulation. The release and storage patterns reflect the historical wet and dry years. As the calibration work did not focus on release and storage volumes specifically, the uncertainty levels could be higher than expected.