Modeling Earth Systems and Environment
https://doi.org/10.1007/s40808-018-0483-4

ORIGINAL ARTICLE

Sensitivity and uncertainty analysis for streamflow prediction using multiple optimization algorithms and objective functions: San Joaquin Watershed, California

Manashi Paul[1] · **Masoud Negahban-Azar**[1]

Received: 5 April 2018 / Accepted: 7 June 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract
Uncertainty analysis prior to the model calibration is key to the effective implementation of the hydrologic models. The major application of sensitivity analysis is to indicate the uncertainties in the input parameters of the model, which affects the model performance. There are different optimization algorithms developed and applied in the hydrologic model, which can be performed with different objective functions to calibrate and quantify the uncertainties in the system. The purpose of this study was to evaluate the model calibration performance and sensitivity of parameters using three optimization algorithms and five objective functions for predicting monthly streamflow. Sequential Uncertainty Fitting (SUFI-2), Generalized Likelihood Uncertainty Estimation (GLUE), and Parameter Solution (ParaSol) were used to calibrate the monthly streamflow for the semi-arid San Joaquin Watershed in California by using Soil and Water Assessment Tool (SWAT) model. The best performance metrics (R^2, NSE, PBIAS, P-factor, and R-factor) were obtained by SUFI-2 while using NSE as the objective function. The coefficient of determination (R^2), Nash–Sutcliffe Efficiency (NSE), the percentage of bias (PBIAS), Kling-Gupta efficiency (KGE) and Ratio of the standard deviation of observations to root mean square error (RSR) were used as an objective function to assess the monthly calibration performance. KGE was found to be a suitable objective function to calibrate this complex and snowmelt-dominated watershed. The findings from this study will serve as a guideline for hydro-ecological researchers to achieve further watershed management goals.

Keywords Agricultural watershed · Hydrological modelling · SWAT · SWAT-CUP · Model performance criteria

Introduction

Hydrologic cycle has close interactions with the surface and subsurface processes by integration of climate, land use and land cover (LULC) and ocean systems (Kumar et al. 2017; Paul et al. 2017; Zhang et al. 2017). To identify the potential impacts of LULC changes, soil degradation, and climate changes on the ecosystem, it is necessary to study the hydrological parameters, such as surface runoff, soil moisture, evapotranspiration (ET), groundwater, streamflow etc. (Kumar et al. 2017; Paul 2016; Talib and Randhir 2017). Assessment of hydrology has been a long-standing research topic in studying agricultural management, flood forecasting and inundation mapping, soil degradation, nutrient losses, and biodiversity conservation practices (Morton and Olson 2014; Paul 2016; Paul et al. 2017; Rajib et al. 2016; Schilling et al. 2014).

Hydrologic models are effective tools to understand and simulate the hydrologic processes to evaluate the impact of climate and LULC changes, to investigate water quality, and to plan the water resources management (Paul 2016; Paul et al. 2017; Shao et al. 2017; Wang et al. 2016). However, the successful application of distributed hydrologic models depends on proper calibration–validation and uncertainty analysis to capture the existing complex environmental condition (Abbaspour et al. 2015a; Kouchi et al. 2017). Calibration is performed by the appropriate selection of the model input parameters regarding suitable ranges to simulate the hydrological process accurately which is assessed by comparing model outputs for a given set of observed data (streamflow, sediment, etc.) (Arnold et al. 2012; Kouchi

✉ Masoud Negahban-Azar
mnazar@umd.edu

[1] Department of Environmental Science and Technology, University of Maryland, 1430 An. Sci. Bldg., College Park, MD 20740, USA

Published online: 13 June 2018

et al. 2017). The accurate calibration of the hydrological models is a challenging task due to the uncertainties in hydrological modeling. According to several studies, the main sources of uncertainties are the input errors due to inaccurate and interpolated measurements; inaccuracy due to over-simplification of the model; errors in model structure or hypothesis and algorithms; inaccuracies of observation used to calibrate the model; and errors in parameterization and output ambiguity (Abbaspour et al. 2008; Khoi and Thom 2015; Kouchi et al. 2017; Singh et al. 2013; Xue et al. 2013). Therefore, sensitivity analysis is crucial to quantify the uncertainty of the system, determine the effect of input parameters on the outputs, the integral knowledge of data, and optimize the design of a system.

Although model uncertainty is difficult to quantify, using a suitable calibration method can manage to control these large uncertainties (Khoi and Thom 2015; Kouchi et al. 2017; Rostamian et al. 2008; Wu and Chen 2015a). A number of optimization algorithms have been developed in the literature and applied in hydrologic models to calibrate the model, quantify the uncertainty of the system, and rank the influence of various parameters on the system. For instance, the Sequential Uncertainty Fitting procedure (SUFI-2) (Abbaspour et al. 2004), Generalized Likelihood Uncertainty Estimation method (GLUE) (Beven and Binley 1992), Bayesian Inference (Box and Tiao 2011), Parameter Solution (ParaSol) (van Griensven and Meixner 2006), Particle Swarm Optimization (PSO) (Eberhart and Kennedy 1995; Kennedy and Eberhart 1995) etc. have been developed and applied to many hydrological studies. Several studies, evaluate the uncertainty of the model systems and input parameters using different optimization techniques. For example, (Khoi and Thom 2015) applied four uncertainty techniques (SUFI-2, GLUE, ParaSol and PSO) to evaluate the parameter uncertainty analysis of streamflow simulation at the Srepok River watershed in Vietnam, they reported that the SUFI-2 method has the advantages to provide more reasonable simulated results compared to other three techniques. Yesuf et al. (2016) have investigated SUFI-2 and GLUE techniques for the Maybar Watershed in Ethiopia, and found that both techniques were able to produce reasonable simulated results for uncertainty analysis, calibration, and validation of the hydrologic model (Yesuf et al. 2016).

The optimization algorithms differ based on the assessment strategies, determination of the set of parameter ranges, and examination of the desired threshold for a particular objective function (Kouchi et al. 2017). Model simulation performance is evaluated graphically (to compare how well the simulated values fit with the observed data), and statistically (referred as goodness-of-fit criteria, and performance or efficiency criteria). The efficiency criteria can also be used as objective function during the calibration process to help identify an optimal parameter that means parameter sets are adjusted according to a specific search scheme to optimize certain calibration criteria (objective functions) (Madsen 2003; Muleta 2011). Different objective functions rely on different features of the variable that is targeted for calibration, especially when several sites and/or different variables are being calibrated. For Example, Nash–Sutcliffe efficiency tends to rely on the peaks (Nash and Sutcliffe 1970), while mean absolute error relies more on average deviations, or least-square errors aim to fit the hydrograph for high flows. Therefore, the relevance and choice of an objective function is an important decision to calibrate a watershed model because an inappropriate objective function can give a good output in statistical terms while it is conflicting from the reality (Abbaspour 2013; Molina-Navarro et al. 2016; Muleta 2011). For instance, Garcia et al. (2017) identified the best objective functions to calibrate the parameter set and estimate the robustness and sensitivity of the rainfall-runoff model in 691 French watersheds. Molina-Navarro et al. (2017) assessed the impact of the objective function for multi-site and multi-variable calibration using a hydrologic model in the Odense catchment in Denmark. Muleta (2011) examined sensitivity of hydrologic model performances to the objective function during automated calibrations in the Little River Experimental Watershed (LREW) in Georgia.

Looking at these recent studies, it is evident that the capability of using different optimization algorithms in relation to different objective functions needs to be verified in different regions. Despite this importance, there is no study that exclusively focuses on the semi-arid region like central valley of California. In this study, the San Joaquin watershed located in the central valley of California was selected, where water management activities are intense. San Joaquin is an agricultural watershed, where its hydrology is affected by the several reservoirs and dams that are operated for the extensive agricultural irrigation. A hydrologic model will be an the basis of developing the strategies for sustainable water resources management for such a complex hydrology to watershed managers, agricultural producers, and policy makers. However, before the development of the water resource management, it is necessary to identify and quantify the large uncertainties subjected to the distributed hydrological modelling using different optimization algorithms and objective functions within the same hydrologic modeling framework. The aim of this study is to (1) evaluate hydrologic model performances and calibration results using three different optimization algorithms (SUFI-2, GLUE, and ParaSol), and (2) evaluate the impact of five objective functions (R^2, NSE, PBIAS, KGE and RSR) on the monthly streamflow simulations. To achieve these objective, SWAT-CUP (SWAT Calibration and Uncertainty Programs) (Abbaspour et al. 2008) was used for model calibration–validation and sensitivity analysis coupled with the distributed hydrological model the Soil and Water Assessment Tool

(SWAT) (Arnold et al. 2012) in the San Joaquin Watershed. SWAT is one of the most widely used distributed models, and it has been applied worldwide for hydrologic and water quality simulations. In recent years, many studies compared the performance of SWAT models under different optimization algorithms coupled with SWAT-CUP to calibrate the streamflow and uncertainty analysis (Kumar et al. 2017; Molina-Navarro et al. 2017; Uniyal et al. 2015; Yesuf et al. 2016; Zhang et al. 2015), that establish its applicability and scientific acceptance under many different circumstances.

Methodology

Study area

The San Joaquin watershed is located in central California and flows through the highly agricultural region of the Sacramento-San Joaquin Delta (Fig. 1). The watershed has total drainage area of 15357.7 km^2 approximately (USGS Hydrologic Unit Code 18040001). The land use is primarily dominated by agricultural land (37.4%). The remaining area consists of grass/pasture (29.7%), shrubland (7.9%), fallow/idle cropland (7.8%), urban (7.3%), forest (6.8%), and water (3.1%) (Table 1). Based on the USDA National Agricultural Statistics Service-Cropland Data Layer (NASS-CDL)

Table 1 Characteristics of land use and land covers in San Joaquin watershed

	Area (km^2)	Area (Acre)	% Watershed area
Agricultural land	5744.7	1,419,548	37.4
Grass/pasture	4560.9	1,127,017	29.7
Shrubland	1214	300014.4	7.9
Fallow and Barren	1192.3	294629.3	7.8
Urban	1128	278,742	7.3
Forest	1041.7	257,412	6.8
Water andWetland	475.9	117600.7	3.1

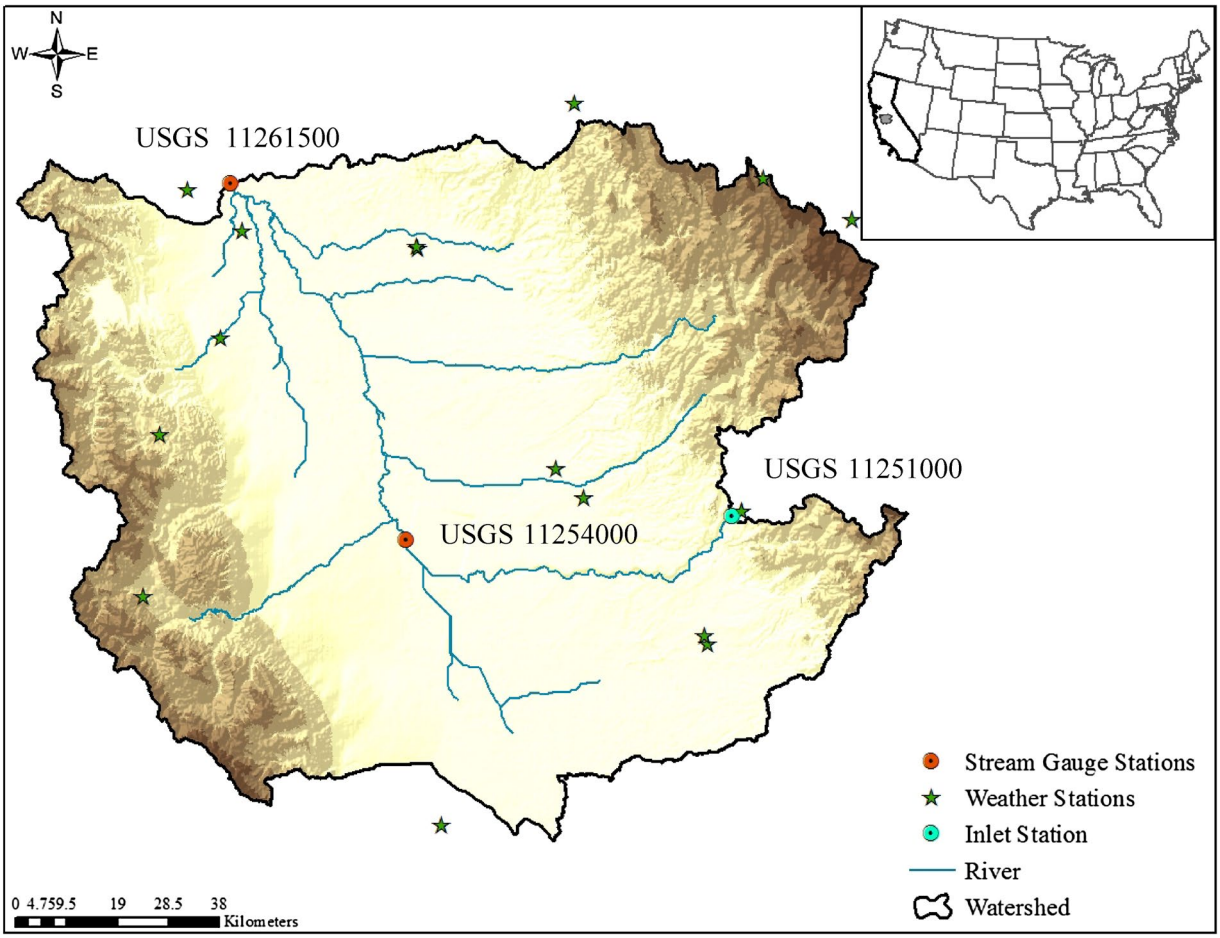

Fig. 1 Location of San Joaquin watershed in California, with selected weather stations and the United States Geological Survey's streamflow gauge stations at respective watershed outlets and inlets

2016, agricultural land is mainly dominated by almond (10.35%), followed by vineyard (8.2%), alfalfa (4.2%), winter wheat (2.6%), tomatoes (2.4%) and cotton (2.3%). The San Joaquin River originates from the high Sierra Nevada Mountains and flows through the northwest of the central valley before reaching the Sacramento-San Joaquin Delta. The Sacramento-San Joaquin Delta has an arid-to-semiarid climate, where average annual precipitation is 323 mm (12.5 inches), and the average annual temperature is 17.1 °C with minimum and maximum of 9.7 and 24.5 °C respectively (Service 2017). Soils in the San Joaquin Delta are composed of mainly ultisols in the high Sierra Nevada ecoregion and recent alluvial soil in the Central Valley ecoregion (Gronberg and Domagalski 1998). A United States Geological Survey (USGS) gauging station 11,261,500 was used as an outlet at Fremont in California. In addition, another USGS station at Mendota (11,254,000) was used in the upstream during calibration. One watershed inlet was defined at the USGS station 11,251,000 below the dam on the San Joaquin River at Friant, California (Fig. 1).

Hydrologic model

In this study, SWAT model (version 2012 with its ArcSWAT interface) was used to delineate the San Joaquin Watershed. SWAT is a physically based semi-distributed and time continuous hydrologic model (Arnold et al. 1998). SWAT is a watershed scale model, was developed to evaluate the impact of land management practices and climate on the water in large and complex watersheds over long periods of time. In SWAT, different water balance components, and water resources (e.g., blue and green waters) are calculated through explicit calculation at subbasin level. In this model, a watershed is divided into a number of subbasins and based on homogeneous soil types, land-use types, and slope classes categorized into hydrological response units (HRUs) that allow a high level of spatial detail simulation.

Model setup

The SWAT model requires input data include Digital Elevation Map (DEM), land use map, soil map, and weather/climate data as the main input data (Neitsch et al. 2011). A 30 m resolution DEM data derived from the USGS National Elevation Dataset (USGS-NED 2013) was used to delineate the watershed boundary. A 30 m resolution of land use data from the NASS-CDL 2016 and soil data from the State Soil Geographic (STATSGO) database were used. Daily precipitation and daily maximum and minimum temperature data for 15 years (2002–2016) were obtained from the National Climatic Data Center (NCDC) website for 18 weather stations (Fig. 1). The multiple HRU options was used to represent the soil and land use types where a single HRU represents a unique combination of soil type and land use. This watershed was discretized into 3902 HRUs in 73 subbasins. Surface runoff is determined using the modified Soil Conservation Service (SCS) Curve Number (CN) method (Arnold et al. 1998; Neitsch et al. 2011; Service 1972; Wu et al. 2012). The Penman–Monteith method (Monteith 1965) was used to estimate the potential evapotranspiration (PET). Channel routing is calculated using the Muskingum routing method or variable storage routing method (Arnold et al. 1998).

Calibration/uncertainty analysis programs

The SWAT-CUP enables sensitivity analysis, calibration, validation, and uncertainty analysis of SWAT models (Abbaspour 2013). The uncertainty analysis was done using three different techniques—Sequential Uncertainty Fitting (SUFI-2), Generalized Likelihood Uncertainty Estimation (GLUE), and Parameter Solution (ParaSol) which are implemented in the SWAT-CUP 2012.

SUFI-2

Based on the Bayesian framework, the sequential and fitting process is used in SUFI-2 for calibration, validation, and sensitivity and uncertainty analysis (Khoi and Thom 2015; Kouchi et al. 2017). In SUFI-2, all sources of parameter uncertainties (e.g., model input, model structure, and measured data) are accounted and described as uniform distributions (Abbaspour 2013). The model's goodness-of-fit and uncertainty are determined by two indices: P-factor and R-factor. Latin hypercube sampling is used to obtain the propagation of the uncertainty and known as P-factor. The P-factor is the percentage of observed data bracketed by the 95% prediction uncertainty (95PPU) (determined at the 2.5 and 97.5% levels of the cumulative distribution of output variables). The R-factor is quantifying the strength of a calibration/uncertainty analysis by the average thickness of the 95PPU band divided by the standard deviation of the observed data (Abbaspour 2013). The value of the P-factor ranges between 0 and 100% and R-factor ranges between 0 and infinity. Theoretically, where the simulation exactly corresponds to the observed data, the P-factor, and R-factor incline to be 1 and 0, respectively. The aim of the SUFI-2 is bracketing most of the observed data with the smallest possible uncertainty band that means good results with a relatively large P-factor and small R-factor. An objective function is defined before uncertainty analysis and assigned with a required stopping rule.

GLUE

The GLUE is relatively simple and widely used in hydrology. GLUE was first introduced by Beven and Freer (2001) to allow the equifinality of parameter sets during the estimation of model parameters in over parameterized models. The GLUE depends on the Monte Carlo method where a large number of simulations are performed with a randomly chosen parameter values selected from prior parameter distributions. In this method, a likelihood value is assigned to each set of parameter values to compare the predicted simulation and observed data and evaluate the simulated parameter combination into the real system. However, GLUE rejects the concept of a unique global optimum parameter set within some particular model structure used in the most calibration procedures in the hydrological modeling. GLUE is different from other optimization algorithms because of its the acceptability of different parameter sets which can produce fit model predictions with similarly good performance. The objective of GLUE is to identify a set of behavioral models within the universe of possible model or parameter combinations (Abbaspour 2013; Blasone et al. 2008; Kouchi et al. 2017). Similar to SUFI-2, all souces of uncertainty are also accounted in GLUE for parameters uncertainty (Beven and Freer 2001).

ParaSol

The ParaSol method combines the objective functions into the global optimization criterion and the modified Shuffle Complex (SCE-UA) algorithm is used for uncertainty analysis to find the optimum solutions (Abbaspour 2013; Duan et al. 1992; van Griensven and Meixner 2006). Similar to GLUE methodology, a threshold or criterion value is used to divide the performed simulations into 'good' and 'not good' simulations after the optimization of the modified SCE-UA. However, unlike GLUE, the threshold value can be defined by the χ^2-statistics where the selected simulations correspond to the confidence region (CR) or Bayesian statistics which could point out the high probability density (HPD) region for parameters or the model outputs (Abbaspour 2013; Wu and Chen 2015b). Through global SCE algorithm, the minimization of a single function can be done up to 16 parameters (Abbaspour 2013; Duan et al. 1992).

Objective functions

The coefficient of determination (R^2), Nash–Sutcliffe Efficiency (NSE), the percentage of bias (PBIAS), Kling-Gupta efficiency (KGE), and Ratio of standard deviation of observations to root mean square error (RSR) were used as objective functions to assess the agreement between simulated and observed streamflow hydrographs.

Coefficient of determination

Standard regression R^2 (Krause et al. 2005) is an indicator in which the model explains the total variance in the observed data (the squared ratio between the covariance and the multiplied standard deviations of the observed and predicted values.). R^2 describes the degree of collinearity between the observed and simulated values. Therefore, large R^2 value can be obtained with a poor model which consistently overestimates or underestimates the observations.

Nash-Sutcliffe efficiency

Dimensionless NSE (Nash and Sutcliffe 1970) is addressing the temporal dynamics and the most widely used statistics for hydrologic calibration. The values range from negative infinity to 1, where 1 shows a perfect model; zero implies that observed mean is as good as predicted model; and less than zero means observed mean is a better predictor than the model. NSE is sensitive to extreme values due to the squared differences between observed and simulated values (Krause et al. 2005).

Percentage of bias

PBIAS (Yapo et al. 1996) is robust and commonly used to determine how well the model simulates the average magnitudes for the output response of interest. According to Moriasi et al. (2007), error index PBIAS is one of the measures that should be included in the model performance reports. PBIAS measures the tendency of the simulated values to be larger or smaller than their observed counterparts. Positive PBIAS values indicate a tendency of the model simulations to overestimate and negative values indicate to underestimate the observations respectively.

Kling-Gupta efficiency

NSE uses the observed mean as baseline, which can lead to overestimation of model skill for highly seasonal variables(e.g., runoff in snowmelt-dominated basins). To overcome this problem associated with NSE, Gupta et al. (2009) proposed an alternative performance indicator KGE, based on the equal weighting of three sub-components: linear correlation (r), bias ratio (β) and variability (α), between simulated and observed discharge (Eq. 4).

Ratio of standard deviation of observations to RMSE

RSR standardizes the root mean square error (RMSE) using standard deviation of the observations. Moriasi et al. (2007) developed RSR based on the recommendation by Singh et al. (2005). Moriasi et al. (2007) mentioned that RSR providing

a normalized value to compare model performances across watersheds or various constituents.

The corresponding performance efficiency criteria for these five objective functions were established according to a recent review of Moriasi et al. (2015) and Thiemig et al. (2013) (Table 2). The formulations of these five objective functions are as follows:

$$R^2 = \frac{[\sum_i (Y^{obs,i} - Y^{obs}_{mean})(Y^{sim,i} - Y^{sim}_{mean})]^2}{\sum_i \sqrt{(Y^{obs,i} - Y^{obs}_{mean})^2} \sum_i \sqrt{(Y^{sim,i} - Y^{sim}_{mean})^2}} \quad (1)$$

$$NSE = 1 - \frac{\sum_i (Y^{obs} - Y^{sim})^2}{\sum_i (Y^{obs} - Y^{obs}_{mean})^2} \quad (2)$$

$$PBIAS = \frac{\sum_{i=1}^{n} (Y^{obs} - Y^{sim})}{\sum_{i=1}^{n} Y^{obs}} x\ 100 \quad (3)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2};\ \alpha = \frac{\sigma_{sim}}{\sigma_{obs}};\ \beta = \frac{Y^{obs}_{mean}}{Y^{obs}_{mean}} \quad (4)$$

$$RSR = \frac{RMSE}{STDEV} = \frac{\sqrt{\sum_{i=1}^{n} (Y^{obs} - Y^{sim})^2}}{\sqrt{\sum_{i=1}^{n} (Y^{obs,i} - Y^{obs}_{mean})^2}} \quad (5)$$

Where Y^{obs} is the observed data, Y^{sim} is the simulated output, and Y^{obs}_{mean} is the mean of observed data, Y^{sim}_{mean} is the mean of simulated output, r is the linear regression coefficient between simulated output and observed data, where σ_{sim} and σ_{obs} are the standard deviation of simulated output and observed data.

Calibration, validation and sensitivity analysis

The calibration protocol presented by Abbaspour et al. (2015b) was followed to calibrate the SWAT model. The SWAT model was calibrated for monthly streamflow from 2009 to 2016 with one-year warm-up period (2008) and validated from 2002 to 2007. The following paragraphs describe the calibration procedure used in this study in a step-by-step process.

SWAT model parameters

SWAT model contains over 200 hydrological parameters, and clearly all of them may not significantly contribute to the output. Therefore, it is necessary to identify the most sensitive input parameters and their ranges for streamflow simulation. In this study, initially, 32 parameters and their initial value ranges were selected based on the literature review on and nearby Central Valley watersheds (Burke and Ficklin 2017; Chen et al. 2017; Luo et al. 2008).

Local sensitivity

Local sensitivity process was taken, where a single parameter was allowed to change in the input parameters and other parameters kept constant. It is also called as one-at-a-time analysis since it is an indicator only for the addressed point estimates instead of the entire distribution.

Global sensitivity

The second approach to sensitivity analysis is the global sensitivity analysis, where a global set of samples are used to explore the design space. SWAT-CUP uses t-stat (high absolute values suggest more sensitivity) and p-value (values close to zero suggest a high level of significance) to identify the relative significance of individual parameters. For the global sensitivity process, 1000 numbers of iterations were selected to identify the most sensitive input parameters. Total of 18 sensitive parameters were identified by Hypercube One-at-a-time (LH-OAT) and global sensitivity analysis using the SUFI-2 (Table 3). These parameters were used for the streamflow simulations at two stream gauge stations (Fig. 1). Table 4 shows that the sensitive parameters yielded by the three optimization techniques (SUFI-2, GLUE, and ParaSol) are the same. However, there are variations in the ranking of the sensitive parameters. This variation in the sensitivity ranking of the parameters is attributed

Table 2 Performance evaluation criteria for flow measures for watershed scale models (adapted from Kouchi et al. (2017); Moriasi et al. (2015); Thiemig et al. (2013))

Measure	Temporal scale	Very good	Good	Satisfactory	Not satisfactory
R^2	D-M-A[a]	$R^2 > 0.85$	$0.0.75 < R^2 \leq 0.85$	$0.60 < R^2 \leq 0.75$	$R^2 \leq 0.6$
NSE	D-M-A	NSE > 0.80	0.70 < NSE ≤ 0.80	0.50 < NSE ≤ 0.70	NSE ≤ 0.50
PBAIS (%)	D-M-A	PBIAS < ± 5	± 5 ≤ PBIAS < ± 10	± 10 ≤ PBIAS < ± 15	PBIAS ≥ ± 15
KGE	M	0.9 ≤ KGE ≤ 1	0.75 ≤ KGE < 0.9	0.5 ≤ KGE < 0.75	KGE < 0.5
RSR	M	0 ≤ RSR ≤ 0.5	0.5 < RSR ≤ 0.6	0.6 < RSR ≤ 0.7	RSR > 0.7

[a]D, M and A denoted daily, monthly, and annual temporal scales, respectively

Table 3 Descriptions and initial ranges of the most sensitive parameters used for model calibration for San Joaquin watershed

Parameter	Definition[a]	Scale of input	Adjustment[b]	Initial range
Groundwater				
ALPHA_BF	Baseflow recession constant (days)	Watershed	1	0.01–1
GW_DELAY	Groundwater delay (days)	Watershed	2	1–500
GW_REVAP	Groundwater "revap" coefficient	Watershed	1	0.01–0.20
REVAPMN	Re-evaporation threshold (mm H_2O)	Watershed	1	0.01–500
GWQMN	Threshold groundwater depth for return flow (mm H_2O)	Watershed	1	0.01–5000
Soil water				
SOL_K	Soil saturated hydraulic conductivity (mm/hr)	HRU	3	− 15 to 15
SOL_AWC	Available soil water capacity (mm H_2O/mm soil)	HRU	3	− 15 to 15
Channel Flow				
CH_N(2)	Main channel Manning's *n*	Reach	1	0.01–0.15
CH_K(2)	Main channel hydraulic conductivity (mm/hr)	Reach	1	5–100
Surface Runoff				
CN2	Curve number for moisture condition II	HRU	3	− 0.3 to 0.1
EPCO	Plant uptake compensation factor	HRU	1	0.75–1
ESCO	Soil evaporation compensation factor	HRU	1	0.75–1
Lateral Flow				
HRU_SLP	Average slope steepness (m/m)	HRU	1	0–1
Snow				
SFTMP	Snowfall temperature (^oC)	Watershed	1	0–5
SMFMN	Melt factor for snow on December 21 (mm H_2O/ ^oC-day)	Watershed	1	0–10
SMFMX	Melt factor for snow on June 21 (mm H_2O/ ^oC-day)	Watershed	1	0–10
SMTMP	Snow melt base temperature (^oC)	Watershed	1	− 2 to 5
TIMP	Snow pack temperature lag factor	Watershed	1	0–1

[a]Source: Neitsch et al. 2001

[b]Type of change to be applied to the existing parameter value: '1' means the original value is to be replaced by a value from the range, '2' means a value from the range is added to the original value, '3' means the original value is multiplied by the adjustment factor (1 + given value within the range)

Table 4 Sensitivity of the model parameters for monthly streamflow simulation generated by the three optimization algorithms

Parameters	SUFI-2			GLUE			ParaSol		
	Ranking	t-stat	P-value	Ranking	t-stat	P-value	Ranking	t-stat	P-value
r__CN2.mgt	1	− 26.82	0	1	− 39.90	0	1	22.71	0
r__SOL_K().sol	2	− 14.08	0	2	− 34.00	0	2	21.63	0
r__SOL_awc().sol	3	13.10	0	3	25.11	0	3	− 16.88	0
v__ESCO.hru	4	− 5.23	0	6	− 4.35	0	13	0.13	0.90
v__ALPHA_BF.gw	5	− 3.56	0	5	− 5.61	0	9	2.15	0.03
v__HRU_SLP.hru	6	− 2.68	0.01	4	− 16.06	0	4	8.90	0
v__REVAPMN.gw	7	2.26	0.02	10	0.84	0.40	6	− 3.45	0
v__GWQMN.gw	8	2.03	0.04	7	3.84	0	8	− 3.35	0
v__GW_REVAP.gw	9	0.99	0.32	9	1.90	0.06	5	− 4.46	0
v__CH_K2.rte	10	− 0.52	0.60	11	− 0.36	0.72	7	3.39	0
v__EPCO.hru	11	0.43	0.67	13	0.11	0.91	11	0.58	0.56
a__GW_DELAY.gw	12	0.29	0.77	8	2.77	0.01	10	0.82	0.41
v__CH_N2.rte	13	− 0.26	0.79	12	− 0.29	0.77	12	− 0.26	0.80

due to the difference in the sampling techniques used for selecting the random samples. Similar results were found by Uniyal et al. (2015) and revealed that any of the techniques (GLUE, SUFI-2, ParaSol, etc.) could be used for the sensitivity analysis.

Rainfall and snowmelt

Rainfall and snowmelt both are driving variables in the watershed hydrology. Therefore, it is better to abstain to calibrate simultaneously with other model parameters (Kouchi et al. 2017). To avoid this identifiability problems with other parameters, the snow parameters and their values were fixed initially in the model. Then, rest of the parameters were used to calibrate the model.

Optimization algorithms

To compare the three optimization algorithms, similar conditions were used regarding calibration parameters and their initial ranges, and statistical criteria. In SUFI-2, the sample size for one iteration could be set in the range of 500–1000 simulations (Abbaspour 2013; Wu and Chen 2015a; Xue et al. 2013), however, 500 is recommended by Abbaspour (2013). In this study, three iterations with 500 simulations in each iteration (total 1500 simulations) were conducted for uncertainty analysis with a preset threshold value NSE = 0.5. Initially, like SUFI-2, same 1500 simulation runs were used for GLUE and ParaSol to compare the sensitivity of calibration performance of the model. However, fewer runs could not produce satisfactory results regarding P-factor and R-factor (results not shown here) while using GLUE and ParaSol as optimization algorithms. Moreover, different studies and literature reviews recommended that GLUE and ParaSol need higher simulations (Abbaspour 2013; Khoi and Thom 2015; Wu and Chen 2015). In the GLUE, generally a wide physically meaningful ranges are used for each parameter to cover more possible behavioral solution. This requires one iteration with a large number of simulation runs (maximum 10,000). Yang et al. (2008) performed the GLUE with 1000, 5000, 10,000 and 20,000 runs and best performance found for a sample size of 10,000 runs. In this study, the parameters range has been selected based on the literature review (mentioned above), therefore, a small simulation runs of 5000 was used with same initial parameter ranges used in SUFI-2. To keep the consistency with SUFI-2, the NSE = 0.5 was selected as the objective function to screen the behavioral and non-behavioral simulations for GLUE. ParaSol optimization technique also requires a larger number of simulations (> 5000). However, similar to GLUE a smaller sample size of 3000 runs and same initial parameter ranges used to conduct uncertainty analysis.

Evaluating objective functions

To evaluate the effect of different objective functions on the streamflow simulation, same initial parameter ranges were used (Table 3). Total of five objective functions were used separately (one objective function per calibration run) to optimize the parameter ranges to calibrate the monthly streamflow. In each iteration, the behavior threshold of each objective function was set based on Table 2 (as example, for NSE it was 0.5) to get the satisfactory results. Finally, the performance of each calibration results was compared using the five efficiency criteria (Table 2).

Results and discussion

Performance sensitivity to optimization algorithms

SUFI-2

The results from the third iteration were used for uncertainty analysis. SUFI-2 found 384 behavioral simulations in last 500 simulations. The 95PPU for the simulated monthly discharge after the third iteration shown in Fig. 2. It is found that 70% measurements at the Fremont station and 60% measurements at the Mendota station were bracketed by the 95PPU during the calibration period and 78 and 62% during the validation period. The relative width of 95% probability band (R-factor) was near 1 during both calibration and validation period (Table 5). These results indicated that SUFI-2 was capable of capturing the observations during the calibration and validation periods. In the Fremont station, the values of the performance measures of the best simulation were within the criteria suggested by (Moriasi et al. 2015), except PBAIS during the calibration period (Table 5). However, Mendota station did not have satisfactory results during both calibration and validation periods, except R^2 (Table 5). At the outlet (Fremont station), for the best simulation the values of R^2, NSE, and PBAIS were 0.91, 0.84, and −40.45% respectively during the calibration period; and 0.88, 0.84, and −7.33% respectively during the validation period.

GLUE

Like SUFI-2, the same threshold value (NSE = 0.5) was used in the GLUE. However, GLUE achieved only 285 behavioral simulations out of 5000 simulation runs. Figure 3 shows the 95PPU plot with 2.5 and 97.5% of the accumulated distribution of prediction uncertainty from the behavioral simulations. The small P-factor (21–30%) indicated that the GLUE optimization algorithm was not able to predict the reasonable observations at both Fremont and Mendota Stations (Table 5). In Fig. 3 it is also

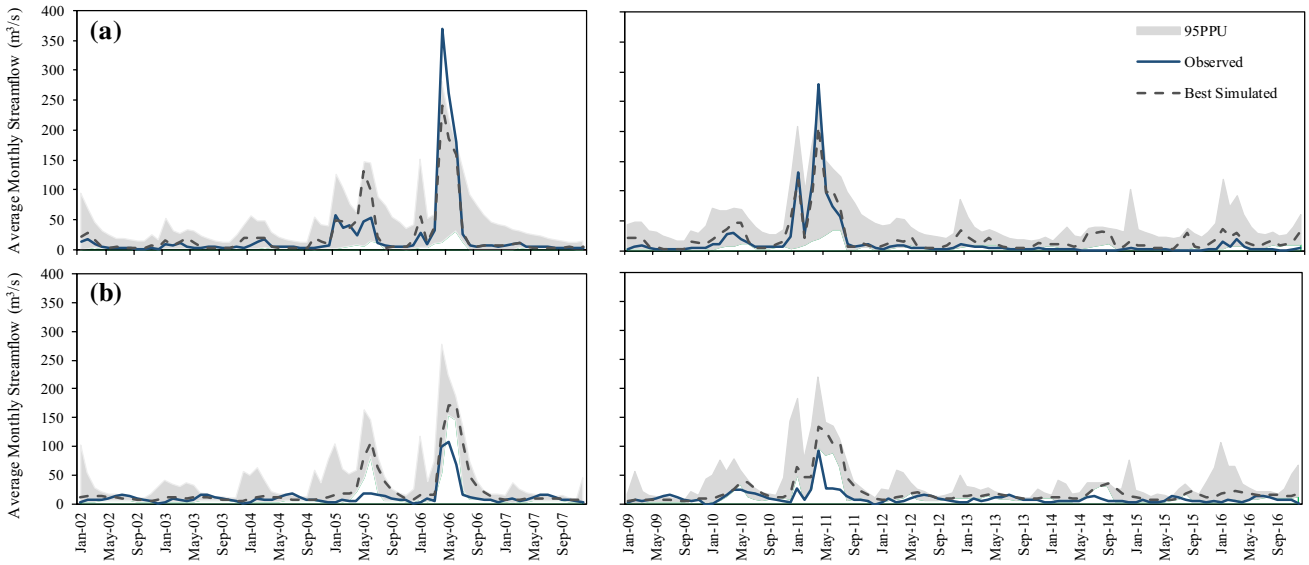

Fig. 2 The best simulated and observed monthly streamflow with 95PPU for calibration (2009–2016) and validation (2002–2007) periods at **a** Fremont station and **b** Mendota station by using the SUFI-2

Table 5 Performance of the three optimization algorithms for the calibration and validation periods in San Joaquin watershed

Optimization techniques	Stations		R^2	NSE	PBAIS	P-factor	R-factor
SUFI-2 500 + 500 + 500 runs 384 behavioral simulations	Fremont	Cal	0.91	0.84	− 40.45	0.70	1.02
		Val	0.88	0.84	− 7.33	0.78	0.98
	Mendota	Cal	0.76	− 1.97	− 55.54	0.60	0.85
		Val	0.81	− 0.96	− 32.13	0.62	0.87
GLUE 5000 runs 285 behavioral simulations	Fremont	Cal	0.89	0.83	− 40.01	0.29	0.74
		Val	0.88	0.84	− 4.17	0.30	0.71
	Mendota	Cal	0.75	− 2.00	− 43.02	0.23	0.45
		Val	0.81	− 0.95	− 66.31	0.21	0.43
ParaSol 3000 runs 2002 behavioral simulations	Fremont	Cal	0.90	0.83	− 47.44	0.61	0.33
		Val	0.88	0.87	− 14.87	0.58	0.32
	Mendota	Cal	0.75	− 2.04	− 61.62	0.17	0.24
		Val	0.80	− 1.22	− 45.26	0.16	0.22

shown that the 95PPU region from GLUE was narrower (R-factor = 0.74 and 0.45 at Fremont and Mendota Station respectively) than the SUFI-2 (R-factor = 1.02 and 0.85) during the calibration period. However, similar to SUFI-2, calibration and validation results were "very good" is terms of R^2 and NSE at the outlet (Tables 2, 5). At the Fremont Station, for the best simulation R^2, NSE, and PBAIS were found to be 0.89, 0.83, and − 40.41% respectively during the calibration period; and 0.88, 0.84, and − 4.17% respectively during the validation period. However, at the Mendota stations performance criteria did not meet except R^2 (Table 5).

ParaSol

ParaSol algorithm was also applied to compare the sensitivity performance with other two optimization algorithms. Unlike GLUE, the ParaSol achieved 2002 behavioral simulations in 3000 simulation runs. The statistical summary of behavioral simulation results is presented in Table 5 and the hydrograph of the observed and best-simulated streamflow with 95PPU in Fig. 4. Figure 4 showed that ParaSol algorithms obtained a very narrow uncertainty region and only 61–63% observed streamflows at the Fremont station, and 17–16% observed streamflows at

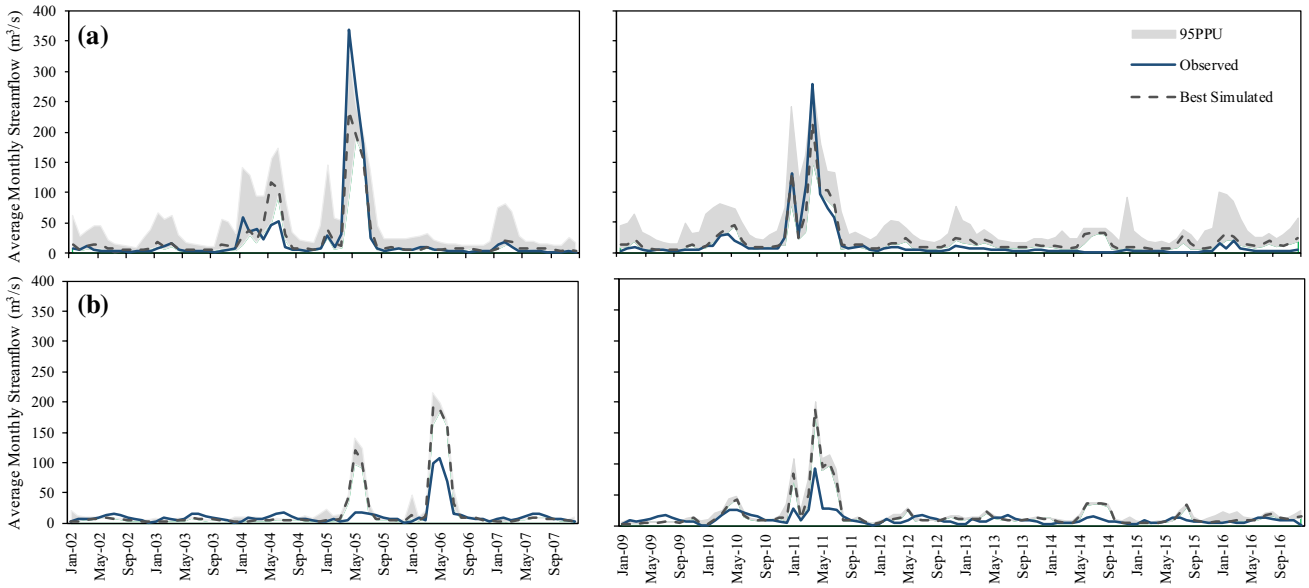

Fig. 3 The best simulated and observed monthly streamflow with 95PPU for calibration (2009–2016) and validation (2002–2007) periods at **a** Fremont station and **b** Mendota station by using the GLUE

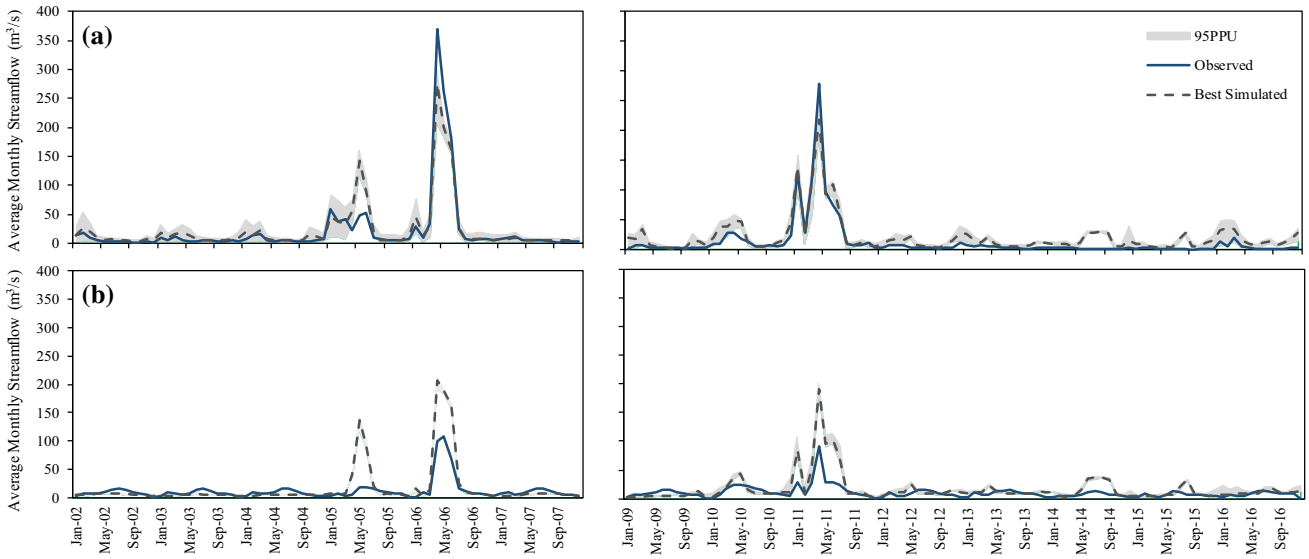

Fig. 4 The best simulated and observed monthly streamflow with 95PPU for calibration (2009–2016) and validation (2002–2007) periods at **a** Fremont station and **b** Mendota station by using the ParaSol

the Mendota station were covered by the 95PPU (Table 5; Fig. 4). At the outlet (Fremont station) for the best simulation the values of R^2, NSE, and PBAIS were found to be 0.90, 0.83, and −47.44% respectively during the calibration period; and 0.88, 0.87, and −14.87% respectively during the validation period (Table 5).

Comparison among three optimization algorithms

The comparison among the SUFI-2, GLUE, and ParaSol were conducted in three aspects: the model performance, the model prediction uncertainty, and the model computational efficiency.

The model performance Amongst three optimization algorithms, there are quite small differences in model performances in the streamflow simulation during both calibration and validation periods (Table 5). Using all the three algorithms, the R^2 and NSE values were ranked as "very good" at the watershed outlet during both calibration and validation periods (Tables 2, 5). However, using all three algorithms, the calibrated model always overestimated the streamflows (-ve PBAIS) to capture the peak flow in 2011. Model performance at Mendota station was not satisfactory based on performance criteria (Table 2), except R^2. In the previous study, Chen et al. (2017) also found similar results for the Mendota station using the SUFI-2 algorithm. This may be attributed to insufficient and uneven spatial distribution of the weather stations and due to the influence of the reservoir in the upstream (Fig. 1). Another possible reason for the mismatch could be the caused by the intense human activities in the upper reaches of the watershed including irrigation channels or small hydropower dams (Fig. 1). In this study, the model performance during the validation period revealed the model capability to encompass the variation of observed streamflows in magnitudes (Figs. 2, 3, 4) and according to the statistical performances sometimes showed better than calibration periods (Table 5). This demonstrates the ability of the model to reproduce the discharge in the San Joaquin Watershed, which can serve as a base model for management and future scenarios analysis.

The model prediction uncertainty The SUFI-2 achieved satisfactory simulations of the streamflow, and indicating a reasonable uncertainty in the calibration and validation results. The SUFI-2 algorithm yielded the similar R^2 (0.91) and NSE (0.84) from the best simulation compared to other two algorithms (GLUE and ParaSol), and generated more balanced prediction uncertainty ranges (R-factor 0.85 to 1.02) with the best coverage of measurement (P-factor 0.60 to 0.78) at the same time (Table 5). However, GLUE and ParaSol barely showed the improvement on P-factor (ranges from 0.16 to 0.63) and R-factor (ranges from 0.22 to 0.74) comparing the SUFI-2. The R-factors for the streamflow of SUFI-2 showed a better performance than GLUE and followed by ParaSol (Table 5). This revealed that the prediction uncertainty range from the SUFI-2 algorithm was wider than that from the GLUE and ParaSol. The ParaSol algorithms generated very narrow prediction uncertainty bands (R-factor 0.22 to 0.33) which were not distinct from the best prediction. This may have resulted from a violation of the statistical assumption of independent and normally distributed residuals. Overall, in this study, the R-factor values found quite small for GLUE and ParaSol that generated the small bands of the 95PPU, and thus small number of observed streamflows were bracketed by the 95PPU (small P-factors) (Figs. 3, 4). Small values of P and R factor in Fremont and Mendota station, indicated that GLUE and ParaSol were not successful in capturing the uncertainty (95PPU, R-factor, P-factor) based on the defined conditions (i.e., initial parameter ranges, number of simulation runs, and behavioral threshold value).

The model computational efficiency The last aspect of the comparative analysis was the model computational efficiency. In this study, an intensive computation was applied for GLUE (5000 simulation runs) and ParaSol (3000 simulation runs) compared to SUFI-2 (total 1500 simulation runs in three iterations). The SUFI-2 algorithm succeeded to get 384 behavioral solutions in last 500 simulations, while GLUE and ParaSol found 285 and 2002 behavioral simulations in 5000 and 3000 simulations, respectively. Although GLUE and ParaSol used a larger number of simulations, the P-factor and R-factor of SUFI-2 showed a better performance than GLUE and ParaSol. Therefore, SUFI-2 was easy to implement compared to other algorithms because the high efficient Latin Hypercube (LH) sampling method can reduce the sampling sizes within a certain space (Khoi and Thom 2015; Wu and Chen 2015a). In the ParaSol, another high efficient sampling method SCE-UA was applied to localize the global optimum of the parameter ranges (Wu and Chen 2015a; Yang et al. 2008; Zhang et al. 2015). The GLUE required large intensive computations (5000 runs) due to use of relatively simple Monte Carlo sampling algorithm, although the application of the GLUE was easier than the other two methods on the sensitivity analysis and global optimization calculation. Therefore, GLUE has the low computational efficiency for high dimensional and complex models, which require more computational resources and time for estimating the uncertainty. The correlation matrix among the best simulation of the streamflows from three optimization algorithms also concluded that the GLUE was less efficient for uncertainty analysis than the SUFI-2, and ParaSol (Table 6).

This study has been done for the semi-arid to the arid climate in the Central Valley of California. These findings were similar to the study conducted by Kouchi et al. (2017), where the SUFI-2, GLUE, and PSO were used to assess the uncertainty estimates for the Karkheh River Basin and Salman Dam Basin (Iran) located in the semi-arid and arid

Table 6 The correlation matrix among the best simulated streamflow obtained by three optimization algorithms

Sensitivity Techniques	SUFI-2	GLUE	ParaSol
SUFI-2	1	0.97	0.99
GLUE		1	0.96
ParaSol			1

regions respectively and indicated the advantages of using the SUFI-2. Chen et al. (2017) also applied the SUFI-2 algorithm for the uncertainty analysis and the monthly streamflows calibration in the San Joaquin watershed. In general, it can be concluded that the SUFI-2 technique is the promising technique in the calibration and uncertainty analysis in the semi-arid to arid regions.

Sensitivity of model parameters to three optimization algorithms

Table 7 and Fig. 5 show the best estimates and 95% uncertainty ranges of all parameters resulting from the GLUE and ParaSol, and posterior parameter ranges resulting from SUFI-2. Each calibrated parameter range has large overlaps

Table 7 List of best estimates and the final parameter uncertainty ranges of the parameters based on all three optimization algorithms applied in San Joaquin watershed

Parameter	Initial rang	Uncertainty range and best parameter estimate[a]		
		SUFI-2	GLUE	ParaSol
ALPHA_BF	0.01-1	0.359 (0.1, 0.6)	0.511 (0.07, 1)	0.046 (0.01, 1)
GW_DELAY	1-500	207.25 (1, 250)	269.145 (15, 490)	263.00 (0, 500)
GW_REVAP	0.01–0.20	0.006 (0.01, 0.1)	0.108 (0.01, 0.2)	0.188 (0.01, 0.2)
REVAPMN	0.01–500	101.1 (0.01, 250)	81.342 (13.5, 487)	279.23 (0.01, 500)
GWQMN	0.01–5000	4886 (3000, 5000)	2502.745 (100, 4780)	2941.6 (0.01, 5000)
SOL_K	− 25 to 25	− 7.42 (− 10, 5)	− 0.212 (− 22.7, 22.29)	− 15.416 (− 25, 4.8)
SOL_AWC	− 25 to 25	5.975 (− 5, 10)	4.574 (− 19.454, 25)	2.562 (− 12.96, 25)
CH_N(2)	0.01–0.25	0.112 (0.07, 0.14)	0.135 (0, 0.25)	0.065 (0, 0.25)
CH_K(2)	5–100	5.3 (5, 50)	99.21 (4, 100)	2.506 (0, 89.12)
CN2	− 0.3 to 0.3	− 0.176 (− 0.35, − 0.01)	0.018 (− 0.28, 0.12)	− 0.142 (− 0.3, 0.1)
EPCO	0.01–1	0.534 (0.25, 0.75)	0.526 (0.08, 0.98)	0.154 (0, 1)
ESCO	0.01–1	0.453 (0.25, 0.75)	0.525 (0.08, 0.98)	0.522 (0, 1)
HRU_SLP	0–1	0.169 (0, 0.5)	0.334 (0.01, 0.75)	0.190 (0, 1)

[a]c (a, b) for each parameter means: c is the best parameter estimate, (a, b) is the 95% parameter uncertainty range except SUFI-2 (in SUFI-2, this interval denotes the final parameters distribution)

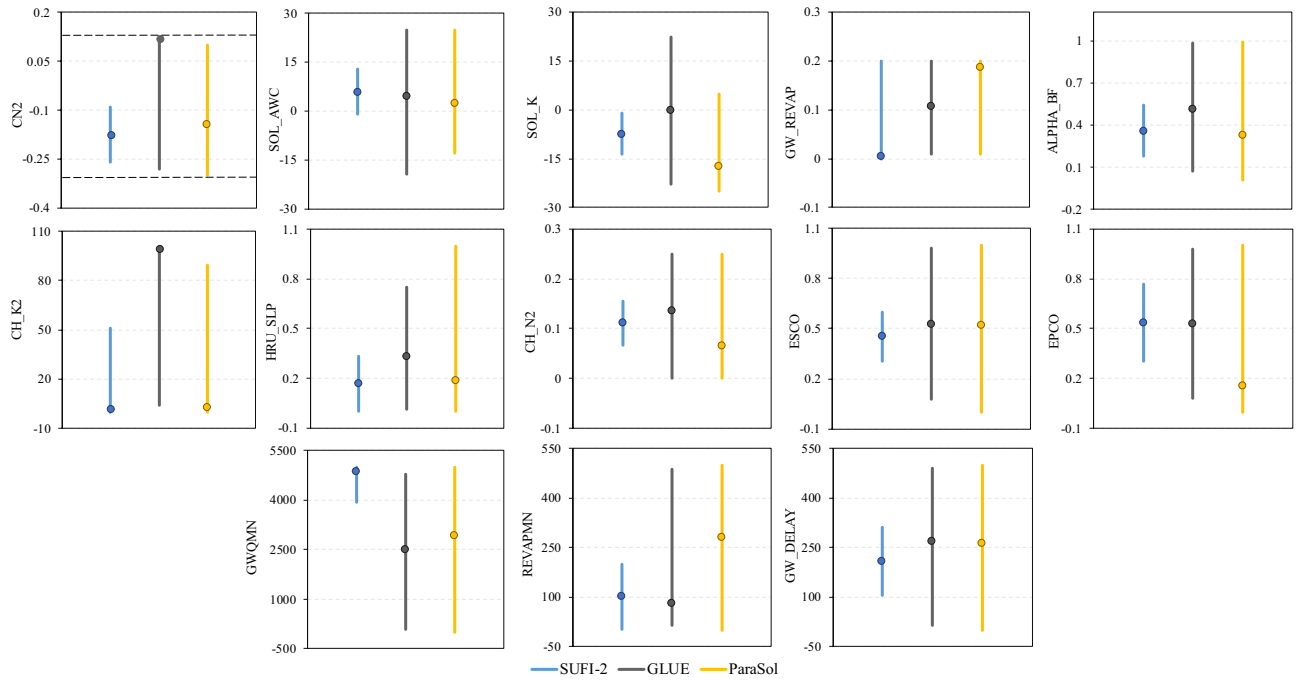

Fig. 5 Final parameter uncertainty ranges with best estimates (points in each line) of the calibrated parameters by three optimization algorithms in San Joaquin watershed

by all three algorithms, although SUFI-2 showed the narrower ranges compared to other two optimization algorithms. The reason behind this is, SUFI-2 creates a combination of all calibrated parameters values for each simulation, and after each iteration, the parameter ranges expressed to narrower distribution from the initial wider distribution.

Performance sensitivity to objective functions using SUFI-2

SUFI-2 allows to use different objective functions and to modify the threshold individually to optimize the calibration parameters. To identify the best objective function, five objective functions (R^2, NSE, PBAIS, KGE, and RSR) were evaluated to calibrate the model using SUFI-2. Table 8 and Fig. 6 shows the model performance by five objective functions using SUFI-2 in the San Joaquin Watershed.

Robustness of model performance

Figure 6 and Table 8 show that all five objective functions performed well at the outlet (Fremont station). Table 8 shows that R^2, NSE and PBIAS values remain almost consistent, while KGE and RSR values showed more variability by all objective functions. The calibration results yielded from different objective functions showed "very good" R^2 and NSE values, while PBAIS values were "unsatisfactory" (Table 8). Only "satisfactory" PBAIS (-14.93%) found at the Fremont station while PBAIS was used as an objective function in the iteration (Table 8). However, a small number of behavioral simulations (56) found compared to the other objective functions (474, 329, 251, and 323 for R^2, NSE, KGE, and PBAIS respectively). In addition, from the hydrograph, it is revealed that the best-simulated streamflow from each objective function (R^2, NSE, KGE, and RSR) captured the observed streamflow satisfactorily, while it seemed to have slightly underestimated for PBAIS (Fig. 6). Based on the comparative analysis between Fig. 6 and Table 8, the final calibration results suggest that using KGE as objective function might be the best option to obtain a good calibration in a complex watershed.

Sensitivity of model parameters

Unlike the different optimization algorithms, the parameters obtained by each objective function showed different ranges

Table 8 Calibration results for best simulated and observed monthly streamflow by five different objective functions using the SUFI-2 in San Joaquin watershed

Objective function in iterations	No. of behavioral simulations	Stations	R^2	NSE	PBIAS	KGE	RSR
R^2 (0.60)	474	Fremont	0.91	0.82	− 53.36	0.56	0.37
		Mendota	0.78	− 1.52	− 54.60	− 0.45	1.59
NSE (0.50)	384	Fremont	0.91	0.84	− 40.45	0.62	0.40
		Mendota	0.76	− 1.97	− 55.54	− 0.36	1.51
PBIAS (± 15)	56	Fremont	0.80	0.77	− 14.93	0.70	0.48
		Mendota	0.78	− 1.48	− 42.80	− 0.44	1.58
KGE (0.50)	251	Fremont	0.89	0.85	− 32.93	0.74	0.38
		Mendota	0.78	− 1.50	− 53.50	− 0.45	1.58
RSR (0.70)	323	Fremont	0.91	0.84	− 51.41	0.60	0.40
		Mendota	0.79	− 1.28	− 52.30	− 0.37	1.51

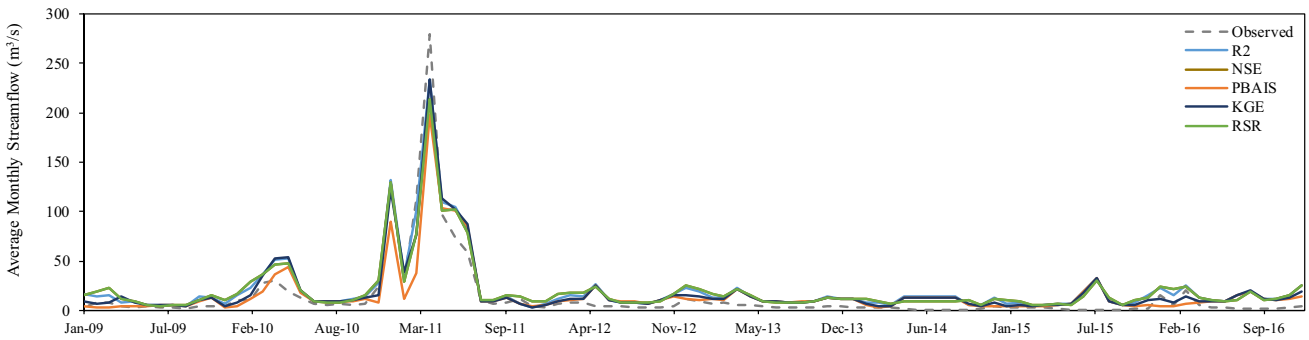

Fig. 6 Comparison of the monthly observed and best simulated streamflow obtained at Fremont station when R^2, NSE, PABIAS, KGE, and RSR used as objective function

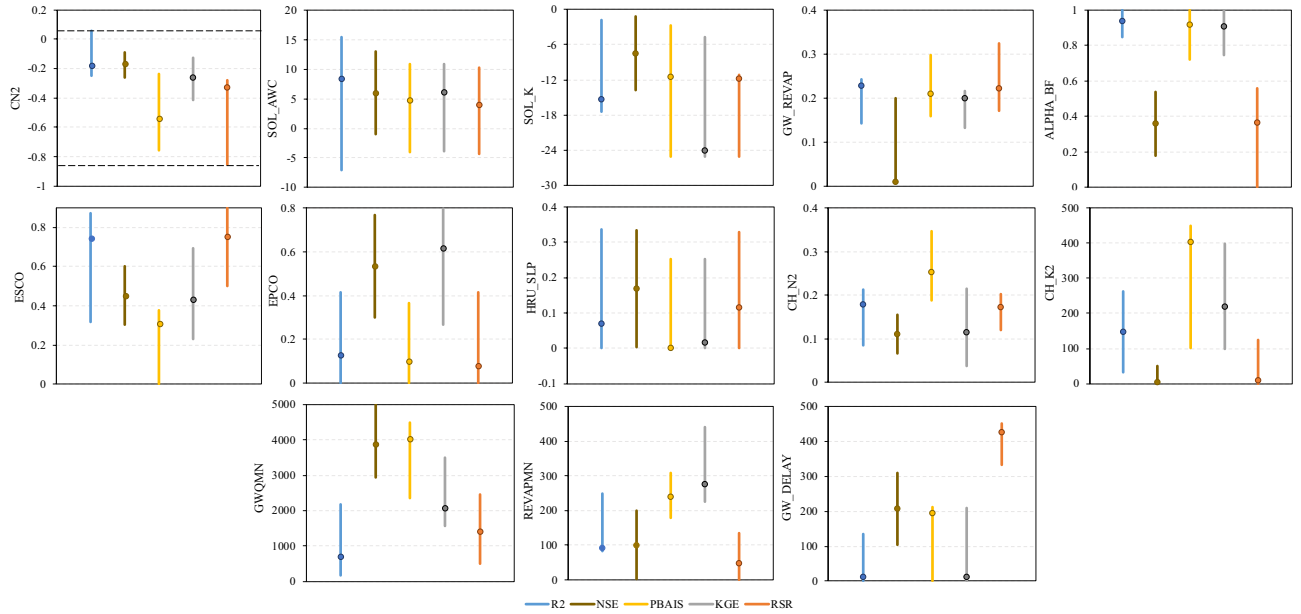

Fig. 7 Final parameter uncertainty ranges with best estimates (points in each line) of the calibrated parameters by five objective functions using SUFI-2 in San Joaquin watershed

Table 9 The correlation matrix among the best simulated streamflow obtained by five different objective functions using the SUFI-2

Objective function	R^2	NSE	PBIAS	KGE	RSR
R^2	1	0.98	0.91	0.95	0.98
NSE		1	0.94	0.94	0.99
PBIAS			1	0.98	0.94
KGE				1	0.98
RSR					1

(Fig. 7) since different objective functions solve different problems (Kouchi et al. 2017). Figure 7 also presents that the identified optimal values for the calibrated parameters were different to each other. However, any clear dominance was not detected among the objective functions with regard to producing optimum parameters values. These results explain the concept of parameter "non-uniqueness" and the concept of "conditionality" of the calibrated parameters, where an unconditional parameter range is defined as parameter range to calibrate the model (Kouchi et al. 2017). Therefore, the unconditional parameter range of CN2 for San Joaquin Watershed would be in the range indicated by the dashed line in Fig. 7. This also indicates a large parameter uncertainty associated with the choice of objective functions regarding parameter ranges.

However, the optimal parameter values were very different from each other while simulated streamflow was not different from each other (Table 9) and considered as "very good" when judged by 4 different performance criteria (R^2, NSE, KGE, and RSR) (Tables 2, 8). These results were consistent with the equifinality concept (Beven and Freer 2001; Muleta 2011), which illustrates that multiple sets of parameters can simulate different and acceptable representations of the watershed characteristics. Relative robustness of the performance criteria was also examined using the correlation matrix (Table 9). Table 9 shows the inter-correlation among the best-simulated streamflows determined from five different objective functions for the calibration period. Table 9 indicates that except PBIAS, the best-simulated streamflows from other four objective functions were well correlated ($r = 0.95$–0.99).

Importance of the study

Successful calibration of the distributed hydrologic models is important for future analysis for watershed or crop managements, and evaluation of impacts of climate and land use changes on the hydrology. Prior to calibration, it is critical to assess the model uncertainties. According to (Abbaspour 2013), reporting the model uncertainty is necessary otherwise the calibration will be "meaningless" and "misleading". The key application of sensitivity analysis is to indicate the uncertainties in the input parameters of the model. Another application of sensitivity analysis is in the utilization of models by managers and decision-makers, which helps to understand the uncertainties, and pros and cons with the limitations and scope of a hydrologic model.

It is also necessary to test the sensitivity of the hydrologic models to different optimization algorithms. However, most applications are only reporting a single optimization algorithm. One of the main reasons is some of the uncertainty analyses techniques are difficult to apply (e.g., the need for testing statistical assumptions). In addition, for a complex hydrologic model, another restriction is the number of simulation runs required for the uncertainty analysis, which needs high CPU speed and parallel computation technology.

This study provides an insight into a hydrologic model response to three different optimization techniques (SUFI-2, GLUE, and ParaSol) for streamflow simulation. The findings show that SUFI-2 is more suitable for the semi-arid San Joaquin watershed to estimate the parameter uncertainty of the streamflow. This study also revealed that all three techniques produced acceptable calibration results, however, with different parameter ranges. This is an important step toward the development of strategies for sustainable water resources management in such semi-arid region with heavy agricultural activities like Central Valley, California.

The second objective of this study was to find a suitable objective function for the streamflow simulation. In most of the studies, NSE used as the objective function, however, in this study it was clear that KGE was more efficient than NSE. One of the reasons could be since KGE was developed (Sect. "Kling-Gupta efficiency" and Eq. 4) to overcome the problem associated with NSE, where observed mean used as a baseline, which can lead to overestimation of model skill for highly seasonal variables (e.g., runoff in snowmelt-dominated watersheds). Therefore, in a snowmelt-dominated watershed like San Joaquin watershed (Lettenmaier and Gan 1990) KGE could be the more suitable to use as the objective function.

The results presented in this manuscript were developed for both low flow (Year 2015) and high flow (Year 2011) simulations, and proved to be suited for both low and high flow simulations. Hydrological models are generally used to simulate the streamflow at ungauged sites by transferring model parameters from gauged to ungauged subbasins. In this study, available streamflow was found for only 2 gauge stations- at the outlet (Fremont station) and at the upstream (Mendota station), despite of having a very large drainage area (15,357.7 km^2). The choice of the objective function used for gauged watersheds might influence the simulation of the regionalized models on ungauged sites. It is likely that the model parameter, transfered from gauged to ungauged watersheds, will carry much more uncertainty than the choice of the objective functions used in gauged watersheds. Therefore, it is recommended to test the sensitivity of the objective functions in smaller gauged watersheds.

Summary and conclusion

The objectives of this study were to evaluate the different optimization algorithms and multiple objective functions to simulate the monthly streamflow with the calibration of the parameter set of a distributed hydrologic model. The SWAT hydrologic model was developed for a large semi-arid watershed in Central Valley, California (San Joaquin watershed). First, three different optimization algorithms (SUFI-2, GLUE, and ParaSol) were evaluated for monthly streamflow simulations. The optimization algorithms were implemented in the SWAT-CUP 2012. The calibration performance and sensitivity of parameters of these algorithms were compared through evaluating the P-factor, R-factor, R^2, NSE, and PBIAS of the best simulation. Afterward, model calibration performance and sensitivity of parameters were evaluated by five objective functions (R^2, NSE, PBAIS, KGE, and RSR) using SUFI-2. Based on the results obtained from this study the following conclusion can be drawn:

1. By comparing the results from three optimization algorithms, the SUFI-2 performed better than the other two algorithms due to the good R^2 and NSE values of the best simulation results and the best prediction uncertainty ranges (P-factor), and the relative coverage of measurements (R-factor).
2. Different objective functions presented different range of the parameters with distinct optimal values while simulating similar streamflow with satisfactory performance criteria.
3. In case of conducting hydrological simulation for streamflow, the SUFI-2 algorithm with KGE as objective function coupled with SWAT model is preferred for the semi-arid and snowmelt-dominated watersheds like San Joaquin watershed.

The calibration and validation performance are not sensitive to the choice of optimization algorithm and objective function, but the obtained parameters are different. Therefore, using the calibrated optimal parameter sets achieved in this study, the local water resource managers and decision makers can obtain more confident prediction intervals for the streamflow simulations.

Acknowledgements This work was supported by the United States Department of Agriculture-National Institute of Food and Agriculture, Grant number 20166800725064, that established CONSERVE: A Center of Excellence at the Nexus of Sustainable Water Reuse, Food, and Health.

References

Abbaspour KC, Yang J, Reichert P, Vejdani M, Haghighat S, Srinivasan R (2008) SWAT calibrating and uncertainty programs–A User Manual. Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Zurich, Switzerland

Abbaspour KC (2013) SWAT-CUP 2012. SWAT Calibration and Uncertainty Program—A User Manual

Abbaspour KC, Johnson C, Van Genuchten MT (2004) Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. Vadose Zone J 3:1340–1352

Abbaspour K, Rouholahnejad E, Vaghefi S, Srinivasan R, Yang H, Kløve B (2015a) A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. J Hydrol 524:733–752

Abbaspour KC, Rouholahnejad E, Vaghefi S, Srinivasan R, Yang H, Kløve B (2015b) A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. J Hydrol 524:733–752. https://doi.org/10.1016/j.jhydrol.2015.03.027

Arnold JG, Srinivasan R, Muttiah RS, Williams JR (1998) Large area hydrologic modeling and assessment part I: model development. JAWRA 34:73–89

Arnold JG, Moriasi DN, Gassman PW, Abbaspour KC, White MJ, Srinivasan R, Santhi C, Harmel R, Van Griensven A, Van Liew MW (2012) SWAT: model use, calibration, and validation. Trans ASABE 55:1491–1508

Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6:279–298

Beven K, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29

Blasone R-S, Vrugt JA, Madsen H, Rosbjerg D, Robinson BA, Zyvoloski GA (2008) Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. Adv Water Resour 31:630–648

Box GE, Tiao GC (2011) Bayesian inference in statistical analysis. Wiley, Hoboken

Burke WD, Ficklin DL (2017) Future projections of streamflow magnitude and timing differ across coastal watersheds of the western United States. Int J Climatol 37:4493–4508

Chen H, Luo Y, Potter C, Moran PJ, Grieneisen ML, Zhang M (2017) Modeling pesticide diuron loading from the San Joaquin watershed into the Sacramento-San Joaquin Delta using SWAT. Water Res 121:374–385

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

Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. Micro Machine and Human Science, 1995 MHS'95. In: Proceedings of the Sixth International Symposium on. IEEE. pp 39–43

Garcia F, Folton N, Oudin L (2017) Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrol Sci J 62:1149–1166

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol 377:80–91

Kennedy J, Eberhart R (1995) Proceedings of IEEE international conference on neural networks. Perth, Australia

Khoi DN, Thom VT (2015) Parameter uncertainty analysis for simulating streamflow in a river catchment of Vietnam. Global Ecol Conserv 4:538–548

Kouchi DH, Esmaili K, Faridhosseini A, Sanaeinejad SH, Khalili D, Abbaspour KC (2017) Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms. Water 9:384

Krause P, Boyle D, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. Adv Geosci 5:89–97

Kumar N, Singh SK, Srivastava PK, Narsimlu B (2017) SWAT Model calibration and uncertainty analysis for streamflow prediction of the Tons River Basin, India, using Sequential Uncertainty Fitting (SUFI-2) algorithm. Model Earth Syst Environ 3:30

Lettenmaier DP, Gan TY (1990) Hydrologic sensitivities of the Sacramento-San Joaquin River Basin, California, to global warming. Water Resour Res 26:69–86

Luo Y, Zhang X, Liu X, Ficklin D, Zhang M (2008) Dynamic modeling of organophosphate pesticide load in surface water in the northern San Joaquin Valley watershed of California. Environ Pollut 156:1171–1181

Madsen H (2003) Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv Water Resour 26:205–216

Molina-Navarro E, Hallack-Alegria M, Martinez-Perez S, Ramirez-Hernandez J, Mungaray-Moctezuma A, Sastre-Merlin A (2016) Hydrological modeling and climate change impacts in an agricultural semiarid region. Case study: Guadalupe River basin, Mexico. Agric Water Manag 175:29–42. https://doi.org/10.1016/j.agwat.2015.10.029

Molina-Navarro E, Andersen HE, Nielsen A, Thodsen H, Trolle D (2017) The impact of the objective function in multi-site and multi-variable calibration of the SWAT model. Environ Model Softw 93:255–267

Monteith J (1965) Evaporation and environment. Symp Soc Exp Biol 19:4

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transa ASABE 50:885–900

Moriasi DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation criteria. Trans ASABE 58:1763–1785

Morton LW, Olson KR (2014) Addressing soil degradation and flood risk decision making in levee protected agricultural lands under increasingly variable climate conditions. J Environ Protect 5:1220

Muleta MK (2011) Model performance sensitivity to objective function during automated calibrations. J Hydrol Eng 17:756–767

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. J Hydrol 10:282–290

Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2011) Soil and water assessment tool theoretical documentation version 2009. Texas Water Resources Institute, Technical Report No. 406. Texas A&M University System. College Station, TX

Paul M (2016) Impacts of land use and climate changes on hydrological processes in South Dakota Watersheds

Paul M, Rajib MA, Ahiablame L (2017) Spatial and temporal evaluation of hydrological response to climate and land use change in three South Dakota watersheds. JAWRA 53:69–88

Rajib MA, Ahiablame L, Paul M (2016) Modeling the effects of future land use change on water quality under multiple scenarios: a case study of low-input agriculture with hay/pasture production. Sustain Water Qual Ecol 8:50–66. https://doi.org/10.1016/j.swaqe.2016.09.001

Rostamian R, Jaleh A, Afyuni M, Mousavi SF, Heidarpour M, Jalalian A, Abbaspour KC (2008) Application of a SWAT model for estimating runoff and sediment in two mountainous basins in central Iran. Hydrol Sci J 53:977–988

Schilling KE, Gassman PW, Kling CL, Campbell T, Jha MK, Wolter CF, Arnold JG (2014) The potential for agricultural land use change to reduce flood risk in a large watershed. Hydrol Process 28:3314–3325

Service USC (1972) Sect. 4: hydrology. In: National Engineering Handbook. SCS, USDA, USA.
Service YW (2017) U.S. Climate Data
Shao W, Cai J, Liu J, Luan Q, Mao X, Yang G, Wang J, Zhang H, Zhang J (2017) Impact of water scarcity on the Fenhe River Basin and Mitigation Strategies. Water 9:30
Singh J, Knapp HV, Arnold J, Demissie M (2005) Hydrological modeling of the Iroquois River watershed using HSPF and SWAT. JAWRA 41:343–360
Singh V, Bankar N, Salunkhe SS, Bera AK, Sharma J (2013) Hydrological stream flow modelling on Tungabhadra catchment: parameterization and uncertainty analysis using SWAT CUP. Curr Sci:1187–1199
Talib A, Randhir TO (2017) Climate change and land use impacts on hydrologic processes of watershed systems. J Water Clim Change 8:jwc2017064
Thiemig V, Rojas R, Zambrano-Bigiarini M, De Roo A (2013) Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. J Hydrol 499:324–338. https://doi.org/10.1016/j.jhydrol.2013.07.012
Uniyal B, Jha MK, Verma AK (2015) Parameter identification and uncertainty analysis for simulating streamflow in a river basin of Eastern India. Hydrol Process 29:3744–3766
USGS-NED (2013) National Elevation Dataset: United States Geological Survey National Map Viewer.. Available at: http://viewer.nationalmap.gov/viewer/ Accessed 10 March, 2013
van Griensven A, Meixner T (2006) Methods to quantify and identify the sources of uncertainty for river basin water quality models. Water Sci Technol 53:51–59
Wang R, Bowling LC, Cherkauer KA (2016) Estimation of the effects of climate variability on crop yield in the Midwest USA. Agric For Meteorol 216:141–156
Wu H, Chen B (2015a) Evaluating uncertainty estimates in distributed hydrological modeling for the Wenjing River watershed in China by GLUE, SUFI-2, and ParaSol methods. Ecol Eng 76:110–121
Wu H, Chen B (2015b) Evaluating uncertainty estimates in distributed hydrological modeling for the Wenjing River watershed in China by GLUE, SUFI-2, and ParaSol methods. Ecol Eng 76:110–121. https://doi.org/10.1016/j.ecoleng.2014.05.014
Wu Y, Liu S, Li Z (2012) Identifying potential areas for biofuel production and evaluating the environmental effects: a case study of the James River Basin in the Midwestern United States. GCB Bioenergy 4:875–888
Xue C, Chen B, Wu H (2013) Parameter uncertainty analysis of surface flow and sediment yield in the Huolin Basin, China. J Hydrol Eng 19:1224–1236
Yang J, Reichert P, Abbaspour K, Xia J, Yang H (2008) Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J Hydrol 358:1–23
Yapo PO, Gupta HV, Sorooshian S (1996) Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J Hydrol 181:23–48
Yesuf HM, Melesse AM, Zeleke G, Alamirew T (2016) Streamflow prediction uncertainty analysis and verification of SWAT model in a tropical watershed. Environ Earth Sci 75:806
Zhang J, Li Q, Guo B, Gong H (2015) The comparative study of multi-site uncertainty evaluation method based on SWAT model. Hydrol Process 29:2994–3009
Zhang L, Karthikeyan R, Bai Z, Wang J (2017) Spatial and temporal variability of temperature, precipitation, and streamflow in upper Sang-kan basin, China. Hydrol Process 31:279–295