

# Improving Model Performance Using Season-Based Evaluation

Misgana K. Muleta

---

**Abstract:** Computer models have become vital decision-making tools in many areas of science and engineering including water resources. However, models should be properly evaluated before use to improve the likelihood of making sound decisions based on their results. The model evaluation technique practiced today in hydrology assumes that model parameters are season insensitive and attempts to identify “optimal” values that would describe watershed behavior during dry and wet seasons. This assumption could compromise accuracy of model predictions. This study demonstrates performance improvement that would be achieved when a season-based model evaluation approach is pursued. A global sensitivity analysis (SA) model has been used to investigate seasonal sensitivity of streamflow parameters of a watershed simulation model on the headwaters of the Little River Watershed, one of the United States Department of Agriculture’s experimental watersheds. Two separate analyses have been performed: the conventional approach in which model parameters are assumed to be season insensitive; and a season-based evaluation in which the influential parameters may vary for months with a low runoff coefficient and months with a high runoff coefficient. The sensitivity analysis helped to identify dominant model and watershed behaviors for the conventional annual approach and for the wet and dry seasons. The SA results show that the influential parameters exhibited modest seasonal sensitivity for the experimental watershed. Model calibration was then performed by using the dynamically dimensioned search (DDS) algorithm for the conventional and season-based approaches using the principal parameters identified by the global SA model. Performance of the calibration attempts have been verified with the traditional split-sampling technique and also by assessing effectiveness of the model in predicting internal watershed behaviors through comparison of simulated streamflow with observations at multiple internal sites not used for calibration. Several efficiency measures have been used to test goodness of the model simulations. The season-based model evaluation technique showed superior performance compared with the traditional method of assuming constant model parameters across seasons. The watershed simulation model exhibited reasonable accuracy in simulating streamflow at the internal sites and for the verification periods when parameter values are allowed to vary from dry to wet season. The “optimal” parameter values identified by the calibration attempts showed significant seasonal sensitivity.

## Introduction

Watershed simulation models have become powerful decision-making aids. Before their use, however, models must be carefully evaluated to ensure that predictions are scientifically sound and defensible (U.S. EPA 2002). Model evaluation refers to procedures such as sensitivity analysis, calibration, and uncertainty analysis that are undertaken to improve model accuracy and to quantify the uncertainty associated with model predictions (Matott et al. 2009). Sensitivity analysis (SA), a technique used to identify the relative significance of model inputs, parameters, or structures on output uncertainty, is an essential model evaluation procedure (Saltelli et al. 2008). Sensitivity analysis helps to understand model behaviors and its consistency with the watershed dynamics

exhibited from observations. SA is commonly used to identify (1) the most influential model parameters (inputs that are not readily measurable and must be estimated) that need to be calibrated; (2) model inputs that describe a significant portion of the output uncertainty that, if measured more accurately, has the greatest potential to reduce output uncertainty; (3) dominant model structures (i.e., model assumptions, abstractions, and methods/theories) that may be more applicable to the watershed and would help reduce output uncertainty.

Many methods are available to perform SA but can be broadly classified as local and global methods (Saltelli 2000). In local SA, the response of the output is investigated around a fixed point in the input space. As the analysis is done around a local point, the entire parameter range cannot be explored. As such, when the perturbation moves away from the local point used during the SA, results become less descriptive of the actual input-output response surface. Also, the more nonlinear the relationship between the input and output variables, which is typical in hydrologic models, the less reliable it is to employ local techniques (Helton 1993). Unlike the local techniques, global SA methods explore the entire range of input factors, thus improving the accuracy of describing the actual input-output relationship (Saltelli et al. 1999).

Following Saltelli's (1999) review of various SA methods and their relative weaknesses and strengths, application of global SA methods has been steadily rising in the area of water resources modeling. Muleta and Nicklow (2005) developed a quantitative global SA method that uses sampling-based stepwise regression analysis and demonstrated its capability to identify parameters to be calibrated by using a widely used distributed watershed model known as soil and water assessment tool (SWAT) (Arnold et al. 1999). van Griensven et al. (2006) combined the one-factor-at-a-time (OAT) method (Morris 1991) with Latin hypercube sampling and developed a qualitative global SA model that has been integrated with SWAT. Tang et al. (2006) compared the performance of four local and global SA methods using a lumped conceptual hydrologic model and showed robustness of the Sobol' (1993) method. On the basis of comparison results, Tang et al. (2007) applied the Sobol' (1993) method on a spatially distributed conceptual hydrologic model and investigated parameter sensitivity at various temporal (i.e., annual, monthly, and rainfall event) and spatial (grid-to-grid) patterns. Likewise, Tang and Zhuang (2009) developed a global SA and Bayesian inference framework and applied it to a monthly time step process-based biogeochemistry model and demonstrated seasonal variability of the principle parameters. Global SA methods have also been used to understand model behaviors across different hydroclimatic regions (van Werkhoven et al. 2008a), to simplify problem complexity for multiobjective calibration (van Werkhoven et al. 2009) and for dynamic identifiability analysis (Abebe et al. 2010).

With global SA, several studies including Tang et al. (2007), Wagener et al. (2003), and van Werkhoven et al. (2008b) have demonstrated sensitivity of influential model parameters to season for the watersheds they studied. From the perspective of diagnostic analysis study, Li et al. (2010) and Tian et al. (2010) have also shown sensitivity of dominant model behaviors to season. However, most model evaluation-procedure practices in hydrology today assume temporal invariability of the dominant parameters and their respective "optimal" values. This assumption could compromise capability of the model to effectively extract information from the observed data and to develop more accurate model that can simulate acceptable watershed responses during dry and wet seasons of a year. For example, White et al. (2009) obtained slight improvement in model performance by allowing seasonal variability of a single parameter during model calibration. This study investigates the advantage of conducting season-based global sensitivity analysis and automatic calibration in improving accuracy of model simulations compared with the conventional approach of assuming seasonal invariability of dominant parameters and their optimal values.

A widely used watershed simulation model known as the soil and water assessment tool (SWAT) has been applied to the headwaters of the Little River Experimental Watershed, one of the USDA Agricultural Research Service (ARS) experimental watersheds. A global SA method known as Sobol' (Sobol' 1993) has been used to investigate sensitivity of SWAT's streamflow parameters at three time periods: annual, months with low runoff coefficient, and months with high runoff coefficient in an attempt to identify the dominant model and watershed behaviors during wet and dry seasons. Then, the dynamically dimensioned search (DDS) (Tolson and Shoemaker 2007) algorithm has been used to calibrate SWAT for the three time periods using the principle parameters identified for each time period. Performance of the calibration results was verified by using the traditional split-sampling approach and by assessing effectiveness of the model in predicting internal watershed behaviors through comparison of simulated streamflow with observations at multiple internal sites

not used for model calibration. Overall, the objectives of this study are to explore seasonal sensitivity of the dominant model parameters and to investigate whether season-based model calibration would improve model performance compared with the traditional calibration approach. Seasonal sensitivity of the "optimal" parameter values would also be explored.

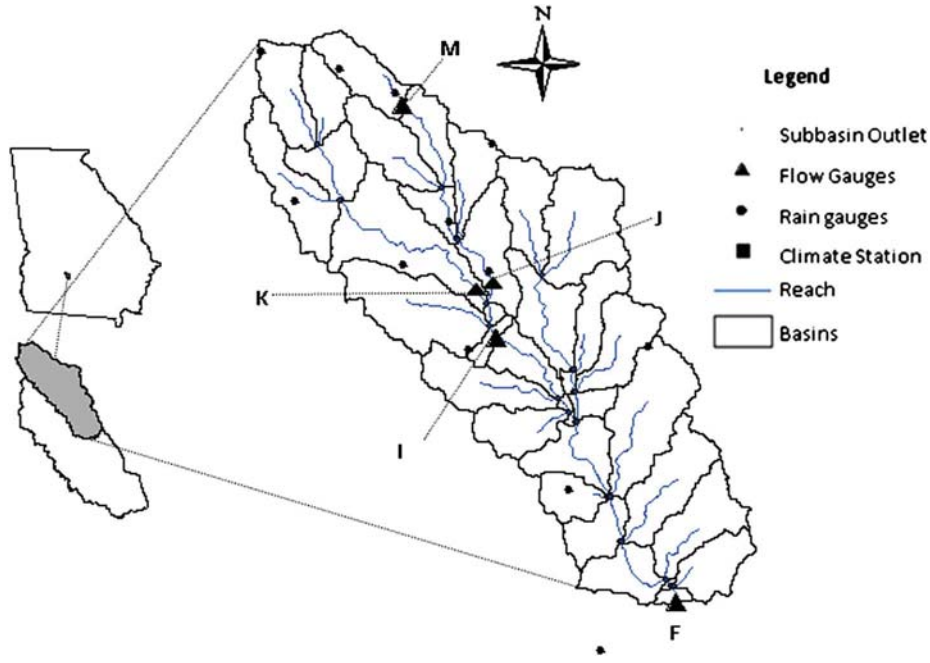
## Materials

### *Watershed Simulation Model*

Over the last three decades, numerous hydrologic and water quality simulation models have evolved. Of these, SWAT (Arnold et al. 1999) is one of the most widely used watershed simulation models in use today (Gassman et al. 2007). SWAT is a physically based, spatially distributed model that uses information regarding climate, topography, soil properties, land cover, and human activities, such as land management practices, to simulate numerous physical processes, including surface runoff, groundwater flow, streamflow, sediment concentration, pesticides, nutrients such as nitrogen and phosphorous, pathogens, and bacteria. Spatially, the model subdivides a watershed into subwatersheds and further partitions subwatersheds into hydrologic response units (HRUs) based on land cover, soil, and overland slope diversity in the subwatershed. Major hydrologic processes modeled by SWAT include snowpack and snow melt, surface runoff, potential evapotranspiration estimated by the Penman-Monteith, Hargreaves, or Priestley method; percolation, simulated by a combination of a layered routing technique with a crack-flow model; lateral subsurface flow or interflow, simulated by a kinematic storage model; and groundwater flow. Sediment yield from each subbasin is computed with the modified universal soil loss equation, which applies runoff as an erosive factor. The transport of nutrients, including phosphorous and nitrogen, is also simulated. SWAT models plant growth and can estimate crop yield. SWAT also allows modeling of detailed land management operations and best management practices (BMPs), including ponds, wetlands, filter strips, and swales. SWAT operates within Environmental Systems Research Institute (ESRI)'s ArcGIS (Winchell et al. 2008) platform greatly simplifying the preparation of model inputs and visualization of outputs. SWAT has been extensively used in the United States and Europe (Gassman et al. 2007). In this study, SWAT2005 (Neitsch et al. 2005) has been used to solve the governing hydrologic equations, and to determine streamflow outputs at desired locations throughout the demonstration watershed. For detailed theory of the hydrologic processes modeled by SWAT, the reader is referred to Neitsch et al. (2005).

### *Study Watershed and Data*

Headwaters of the Little River Experimental Watershed (LREW), one of the USDA-ARS's experimental watersheds, located in Georgia, United States, has been used to demonstrate the research objectives outlined in the introduction (see Fig. 1). The LREW has been selected because it is heavily gauged for rainfall and streamflow (Bosch et al. 2007) and because data are readily accessible online (<ftp://www.tiftonars.org/>) from the Southeast Watershed Research Laboratory (SEWRL). Drainage area of the LREW is approximately 334 km<sup>2</sup> and the watershed is located in the headwaters of the Upper Suwannee River Basin. The watershed consists primarily of low-gradient streams and is located mainly on sandy soils underlain by limestones that form locally confined aquifers. Land use within the watershed is made up of approximately 31% row crop agriculture, 10% pasture, 50% forest, and 7% urban area (Bosch et al. 2006).



**Fig. 1.** Location map of the study area and the gauging stations

Only the upper 116 km<sup>2</sup> of the LREW has been used for this study to minimize computational demand of the model and also because the headwater subwatersheds have more dense streamflow and rainfall gauges. Twelve precipitation gauges and five streamflow gauges (Fig. 1) with long-term daily data (i.e., 1967–2006) are available for the headwaters from the SEWRL. This study used daily precipitation, climate, and streamflow data from 1999–2006. Daily maximum and minimum temperature for a station near the watershed has been obtained from the U.S. Historical Climatology Network (<http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html>) because the climate data available from SEWRL starts only from 2004. The geographic data used to set up the SWAT model including topography, land use, and stream networks have also been obtained from the SEWRL. The Soil Survey Geographic (SSURGO) soil map has been obtained from the Natural Resources Conservation Service (NRCS) soil data mart (<http://soildatamart.nrcs.usda.gov/>).

### Global Sensitivity Analysis Method

Sobol's method (Sobol' 1993), the global sensitivity analysis method used for this study, is a variance-based SA approach that decomposes total variance of the output ( $y$ ) into the contribution of the individual model parameters ( $x_i$ ). Variance of the output can be decomposed into: the sum of the linear (first-order) terms owing to individual parameters ( $x_i$ ); the sum of two-way interactions (i.e., the effect of parameters  $x_i$  and  $x_j$  that cannot be explained by the sum of the individual effects of  $x_i$  and  $x_j$ ); plus sums of higher-order interactions (Campolongo and Saltelli 1997). As such, the method can determine the first-order (main-effect) as well as the total sensitivity indices for each parameter, accounting for higher-order interaction effects between the parameters. In addition, the method is model independent in that, unlike regression and correlation analysis-based techniques, it works for nonlinear and nonadditive models.

Assuming that the parameters are independent, decomposition of the total output variance  $V$  can be described as

$$V = \sum_{i=1}^N V_i + \sum_{i < j \leq N} V_{ij} + \dots + V_{12\dots N} \quad (1)$$

where  $N$  = total number of parameters;  $V_i$  = variance of the model output contributed by parameter  $x_i$ ;  $V_{ij}$  = portion of the output variance explained by the interaction of parameters  $x_i$  and  $x_j$ ; and so on. The first-order sensitivity index ( $S_i$ ) for parameter  $x_i$  is estimated as (Hall et al. 2005)

$$S_i = \frac{V_i}{V} \quad (2)$$

and the total effect  $S_{Ti}$  is determined as (Tang et al. 2007)

$$S_{Ti} = 1 - \frac{V_{\sim i}}{V} \quad (3)$$

where  $V_{\sim i}$  denotes the variance from all parameters except from  $x_i$ ;  $S_{Ti}$  represents the average variance that would remain as long as  $x_i$  remains unknown and is an indicator of the parameter interactions within the model. Parameters with small first-order indices but high total sensitivity indices affect the model output mainly through interactions (Hall et al. 2005).

A Monte Carlo technique is commonly used to determine the sensitivity indices described in Eqs. (2) and (3). The function  $f(x_1, x_2, \dots, x_N)$  is defined in the  $N$ -dimensional unit cube as

$$N^N = (X | 0 \leq x_1 \leq 1; 0 \leq x_2 \leq 1, \dots, 0 \leq x_N \leq 1) \quad (4)$$

and the function is then decomposed into summands of increasing dimension as (Hall et al. 2005)

$$f(x_1, x_2, \dots, x_N) = f_0 + \sum_{i=1}^N f_i(x_i) + \sum_{1 \leq i < j \leq N} f_{ij}(x_i, x_j) + \dots + f_{1,2,\dots,N}(x_1, x_2, \dots, x_N) \quad (5)$$

Then, the following numerical schemes can be used to estimate  $f_0$ ,  $V$ ,  $V_i$ , and  $V_{\sim i}$  and determine the sensitivity indices described in Eqs. (2) and (3) using  $M$  Monte Carlo realizations (Hall et al. 2005)



$$\hat{f}_0 = \frac{1}{M} \sum_{w=1}^M f(x_w) \quad (6)$$

$$\hat{V} = \frac{1}{M} \sum_{w=1}^M f^2(x_w) - \hat{f}_0^2 \quad (7)$$

$$\hat{V}_i = \frac{1}{M} \sum_{w=1}^M f(x_{(\sim i)w}^{(a)}, x_{iW}^{(a)}) f(x_{(\sim i)w}^{(b)}, x_{iW}^{(a)}) - \hat{f}_0^2 \quad (8)$$

$$\hat{V}_{\sim i} = \frac{1}{M} \sum_{w=1}^M f(x_{(\sim i)w}^{(a)}, x_{iW}^{(a)}) f(x_{(\sim i)w}^{(a)}, x_{iW}^{(b)}) - \hat{f}_0^2 \quad (9)$$

where  $x_w$  = sampled Monte Carlo realization within the unit hypercube;  $x_{(\sim i)w}$  refers to all  $M - 1$  samples except  $x_{(iw)}$ . Superscripts (a) and (b) indicate that two sampling data matrices, each with  $M \times N$  dimension, are being used for  $x$ . To calculate high-order indices, Sobol' (1993) requires  $M \times (2N + 1)$  model simulations and can be computationally demanding, especially for distributed hydrologic models where  $N$  can be large. Saltelli (2002) introduced an efficient technique that calculates the first-order, the total-order and the  $(N - 2)$ -order indices with  $M \times (N + 2)$  model simulations. In this study, the Sobol' method implemented in SimLab 2.2 (<http://simlab.jrc.ec.europa.eu/>) has been used by integrating SimLab 2.2 with SWAT. SimLab 2.2 requires  $M \times (2N + 2)$  model runs to compute the full set of total- and first-order indices for the Sobol' method.

### Dynamically Dimensioned Search

Dynamically dimension search (DDS) (Tolson and Shoemaker 2007) has been developed to improve computational efficiency of calibrating spatially distributed watershed simulation models. DDS is a simple, single-objective, heuristic search method that starts by globally searching the feasible region and incrementally localizes the search space as the number of simulation approaches the maximum allowable number of simulations (the only stopping criteria used by the algorithm). Progress from global to local search is achieved by probabilistically reducing the number of model parameters modified from their best value obtained thus far. New potential solutions are created by perturbing the current parameter values of the randomly selected model parameters only. The perturbation magnitudes are randomly sampled from a normal distribution with a mean of zero. The best solution identified thus far is maintained and never updated with a solution with an inferior value of the objective function. One beauty of the DDS is that it requires no algorithmic parameter tweaking because the only parameters to set are the maximum number of model evaluations and the scalar neighborhood size perturbation parameter ( $r$ ) that defines the random perturbation size standard deviation as a fraction of the decision variable range. The recommended value of 0.2 (Tolson and Shoemaker 2007) has been used for  $r$  in this study.

Tolson and Shoemaker (2007) compared DDS with the shuffled complex evolution (SCE) for several optimization test functions, as well as real and synthetic SWAT2000 calibration formulations, and showed DDS to be more efficient and effective than SCE. In two of their test cases, DDS required only 15–20% of the number of model evaluations used by SCE to find equally good values of the objective function. Muleta (2010) also conducted preliminary comparison of DDS, SCE, Genetic Algorithms (GAs), and Parameter Estimation (PEST) for automatic calibration of SWAT2005 and showed robustness of DDS in terms of efficiency and effectiveness. In this study, the source code of the DDS algorithm described in

Tolson and Shoemaker (2007) has been obtained from the first author and has been integrated with SWAT to calibrate streamflow for the study watershed by using the influential SWAT parameters identified by the global SA method described previously.

## Methodology

### Data Preprocessing and Watershed Delineation

The data required by SWAT2005, the watershed simulation model used in this study, have been obtained for the headwaters of the LREW primarily from the SEWRL. Missing values for the precipitation data have been filled by using areally averaged precipitation determined from gauges with available data for that particular day. Areal average precipitation was used because of homogeneity of precipitation in the study area. Based on the 1968–2006 data, mean daily precipitation of the 12 rain gauges in the study watershed varied from 3.18 to 3.45 mm. The minimum and the maximum daily rainfall correlation factors among the 12 rain gauges were 0.77 and 0.98, respectively. These results indicate a reasonably homogeneous spatial rainfall pattern across headwaters of the LREW. Precipitation and other climate data were then formatted in the way that is readable by ArcSWAT, the ArcGIS interface that prepares SWAT2005 inputs and parameters from climate and watershed data (Winchell et al. 2008).

The SSURGO soil data used in this study provide the highest resolution soil map for a countywide soil database in the United States. Several studies have shown that SSURGO improves accuracy of SWAT's streamflow estimates compared with the State Soil Polygon (STATSGO) soils (Wang and Melesse 2006; Geza and McCray 2008). Because SSURGO soils cannot be directly used by ArcSWAT, SWATioTools (Sheshukov et al. 2009), an ArcMap GIS extension tool that processes the SSURGO soils into the format that is readable by ArcSWAT, has been used to preprocess the SSURGO soils. The land cover image used for the study was for year 2003 and was also preprocessed to synchronize the names used in the original map with SWAT's land cover types. Once the climate, the land use, and the soil data were preprocessed, the 116 km<sup>2</sup> study watershed was delineated and subdivided into 37 subbasins and 96 HRUs using ArcSWAT (Winchell et al. 2008) as shown in Fig. 1. From the perspective of streamflow modeling, subdividing the watershed into 37 subwatersheds and 96 HRUs may be unnecessary because several studies including Muleta et al. (2007) and FitzHugh and Mackay (2000) have demonstrated that using a large number of subwatersheds and HRUs would not improve accuracy of SWAT's streamflow simulation. This study is, however, part of an ongoing research that would involve modeling of nonpoint source (NPS) pollutants such as sediment and nutrients whose simulation accuracy is sensitive to the number of subwatersheds and HRUs (Muleta et al. 2007). As such, detailed delineation has been used considering the anticipated NPS pollution study.

### Seasonal Sensitivity Analysis and Calibration

Through dynamic identifiability analysis, several studies have shown that model and watershed behaviors may react differently to the same model parameter during various periods of a year (Wagener et al. 2003; Tang et al. 2007; Abebe et al. 2010). Dominant model structure and parameters may depend on forcings and antecedent conditions (Tang et al. 2007). Various model diagnostic analysis studies have also shown sensitivity of dominant model behaviors to various seasons of a year (Li et al. 2010; Tian et al. 2010). According to Tang et al. (2007) and van Werkhoven et al. (2008b), forcings, primarily rainfall, is responsible for dynamic

sensitivity of the model they used on their demonstration watershed. For the headwaters of LREW, however, careful review of the observed rainfall and runoff data showed strong seasonality of the rainfall-runoff relationship that cannot be described using rainfall alone. As a result, seasonality of the watershed's rainfall-runoff behavior was described in this study using monthly runoff coefficients determined from 39 years (i.e., 1968–2006) of rainfall and runoff data. The runoff coefficient as used in this study is defined as the ratio of the total monthly runoff measured at the outlet of the watershed to areally averaged total monthly rainfall. Areal averaged monthly rainfall totals, monthly runoff totals, and monthly runoff coefficients obtained for the watershed are given in Table 1.

Table 1 reveals interesting information regarding rainfall-runoff characteristics of the watershed. Except for March, the highest monthly rainfall totals were recorded for the watershed in June, July, and September. However, monthly runoff coefficients of these three months (i.e., June, July, and September) are among the lowest. This indicates that unlike the finding of Tang et al. (2007) and van Werkhoven et al. (2008b), dynamic parameter sensitivity may not be described on the basis of rainfall alone for the watershed used in this study. To test seasonal sensitivity of SWAT2005 streamflow parameters and also to test the improvement in model accuracy that may be achieved by calibrating SWAT2005 for separate seasons, both SA and calibration runs were performed on the following three time periods: (1) months with a runoff coefficient greater than 0.1 (i.e., December to April); (2) months with a runoff coefficient less than 0.1 (i.e., June to October); and (3) all months combined irrespective of their runoff coefficients, which is typical of the model evaluation methods practiced today. For the season-based evaluation, November and May were used as transition months in which model parameters were changed linearly from their respective dry season values to wet season values and vice versa, respectively. The conventional model evaluation approach has been used as baseline to compare the advantage of the dynamic (i.e., seasonally varying) model evaluation technique attempted in this study. Unlike the moving window approach used to define the time window upon which dynamic identifiability analysis is performed (Wagener et al. 2003; Abebe et al. 2010), the season-based analysis pursued in this work is operationally more practical because it requires only two distinct time periods (i.e., dry months and wet months as defined by monthly runoff coefficients). The global SA method was used to identify the key model parameters for each time period. Then, DDS was used to calibrate SWAT2005

**Table 1.** Monthly Runoff Coefficients Calculated for the Study Area

Month	Monthly rainfall total (mm)	Monthly runoff total (mm)	Runoff coefficient
January	248.6	57.7	0.23
February	258.3	63.4	0.25
March	698.4	76.0	0.11
April	286.8	43.4	0.15
May	186.9	18.6	0.10
June	437.4	15.2	0.03
July	473.8	15.0	0.03
August	320.2	13.8	0.04
September	455.7	10.3	0.02
October	272.6	7.2	0.03
November	279.2	11.5	0.04
December	255.9	25.3	0.10

by using the top 10 influential parameters identified for each time period.

The streamflow data collected at Gauge F (outlet of the study watershed as shown in Fig. 1) was used for the SA and calibration. One-year data (i.e., 1999) were used as a warm-up period to diffuse the effect of antecedent conditions, and four-year data (i.e., 2000–2003) were used for the sensitivity analysis and calibration. Performance of the calibration attempt was verified by using the traditional split-sampling approach (i.e., the 2004–2006 data at the calibration site were used for verification) in addition to assessing the capability of the calibrated model to simulate streamflow with reasonable accuracy at the internal gauges not used for calibration (i.e., Gauges I, J, K, and M). The Nash-Sutcliffe efficiency (NSE) criterion (Nash and Sutcliffe 1970) described in Eq. (10) was used as output for the SA and as objective function for the calibration attempts. Root mean square error [RMSE; Eq. (11)], percent bias [% Bias; Eq. (12)], and agreement of the observed and simulated mean annual streamflow have been used as additional criteria to compare goodness of the calibrated model predictions. Moriasi et al. (2007) recommended percent bias as one of the measures that should be included in model performance reports. Percent bias describes whether model simulations over or underestimate the observations

$$\text{NSE} = 1 - \frac{\sum_{i=1}^N (Y_i - O_i)^2}{\sum_{i=1}^N (O_i - O_{\text{mean}})^2} \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - O_i)^2} \quad (11)$$

$$\% \text{BIAS} = 100 * \frac{\sum_{i=1}^N (O_i - Y_i)}{\sum_{i=1}^N O_i} \quad (12)$$

where  $Y$  = model simulated output;  $O$  = observed hydrologic variable;  $O_{\text{mean}}$  = mean of the observations, which the NSE uses as a benchmark against which performance of the hydrologic model is compared; and  $N$  = total number of observations. In this study, NSE is used as objective function because of its popularity in the hydrologic literature. However, in spite of its popularity, NSE has many well-documented limitations in describing goodness of model performance and when used as objective function during model calibrations (ASCE 1993; Moriasi et al. 2007; Gupta et al. 2009). An ongoing study is investigating sensitivity of model performance to the goodness-of-fit criteria used as objective function during model calibration.

## Results and Discussion

### Sensitivity Analysis

Sensitivity analysis was performed for the study watershed using the Sobol' method for the three time periods described formerly. A total of 20 SWAT2005 streamflow parameters were considered for the SA. All 20 parameters were assumed to follow uniform distribution as done by Muleta and Nicklow (2005), and the lower and upper bounds recommended in Neitsch et al. (2005) were used for majority of the parameters. A list of the parameters and their ranges are provided in Table 2. Some of these model parameters (e.g., NRCS's curve number, CN2) vary from HRU to HRU, from subbasin to subbasin, or from reach to reach, depending on soil,

land cover, slope or other watershed behaviors. During the sensitivity analysis and model calibration, the baseline values assigned to each spatially varying parameter were altered by multiplying the baseline value by the sampled multipliers or by adding the sampled values to the baseline value as shown in Table 2. This way, parameters would be scaled up or down while preserving their spatial variability. Once the input sample is generated, SWAT2005 was executed to simulate streamflow at Gauge F. The simulated and the observed streamflow at site F were used to determine NSE.

One of the subjective decisions that need to be made while conducting SA is the number of input-output samples to use. To analyze sensitivity of results to the number of input-output samples and also to investigate the minimum number of samples needed by the global SA method to generate steady ranking of the influential parameters, three different input-output sample sizes were investigated for Sobol' for each of the three time periods described previously. As described in the global SA method section, if a parameter range is divided into  $M$  intervals and given that  $N$ , the total number of model parameters is 20 for this study, the Sobol' method requires  $M \times (2N + 2)$  model simulations. Input-output sample sizes of  $M = 16, 32,$  and  $64$  were considered in this study, implying that sample sizes of 672, 1,344, and 2,688 were tested for Sobol'. The results showed no significant difference in parameter rankings for sample sizes of 1,344 and 2,688, indicating that  $M = 32$  may be used for this study. However, all the SA results reported in this study are obtained by using  $M = 64$ .

SA results are provided in Tables 3, which presents the parameter rankings obtained using  $M = 64$  for the dry season (June to October), the wet season (December to April), and when both seasons are combined irrespective of the runoff coefficient. The parameter rankings did not exhibit significant shift among the conventional annual approach and the wet season scenarios, especially for the top nine sensitive parameters. The largest shift observed among the annual and the wet season scenarios was for Sol\_K (soil

hydraulic conductivity), which dropped from 10 for the annual scenario to 14 for the wet season scenario. The remaining top 10 parameters either maintained their rankings or were shifted up or down only by one step. This insensitivity in rankings among the wet season and the annual scenarios may be attributed to the fact that the annual scenario result is biased toward the wet season streamflow (e.g., peak flows) because the NSE criteria was used as output for the SA. To evaluate this possible reason, an ongoing study is investigating sensitivity of model evaluation results to the goodness-of-fit criteria used as output or objective function.

Table 3 shows that the wet season and the dry season scenarios showed noticeable shift in parameter rankings. Significant shifts were obtained for Esco (soil evaporation compensation factor), which dropped from 4th under the wet season scenario to 10th for the dry season scenario; Ch\_N2 (Manning's roughness coefficient for main channels), which dropped from 7th under the wet season scenario to 11th for the dry season case; and Sol\_K that moved up from 14th for the wet season scenario to 7th for the dry season scenario. These results provide some insight, in most cases well-known processes that could be presumed a priori, regarding the relative importance of the hydrologic processes represented by the respective parameters during the wet and dry seasons. Dropping of Esco in ranking during dry season indicates that soil evaporation plays less role toward estimating streamflow during dry months because soil moisture content of the soil is generally lower compared with the wet season, when soil moisture content is relatively high and is available for evaporation. Likewise, a drop in the importance of Ch\_N2 during dry months shows that channel flow routing is not a very important process during the dry season because of insufficient flow in the channel for the routing parameters to make significant change in the simulated streamflow. Most of the rain recorded during the dry season may be lost to infiltration (as shown by the importance of Sol\_K during dry months), transpiration, and depression losses and would produce less streamflow

**Table 2.** Model Parameters and Ranges Used for the Sensitivity Analysis and Calibration

Name	Description	Range of values	
		Minimum	Maximum
Alpha_Bf	Base flow alpha factor (days)	0.1	1
Biomix	Biological mixing efficiency	0	1
Blai	Leaf area index for crop	0	1
Canmx	Maximum canopy storage index	0	10
Ch_K2	Effective hydraulic conductivity in main channel alluvium (mm/h)	0	150
Ch_N2	Manning's $n$ for the main channels	0.0	0.5
Cn2 <sup>a</sup>	SCS runoff curve number for moisture condition II	25%	25%
Epc0	Plant evaporation compensation factor	0	1
Esco	Soil evaporation compensation factor	0	1
GW_Delay <sup>b</sup>	Groundwater delay (days)	10	10
GW_Revap <sup>b</sup>	Groundwater revap coefficient	0.036	0.036
Gwqmn <sup>b</sup>	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	5000	5000
Revapmn <sup>b</sup>	Threshold depth of water in the shallow aquifer for "revap" to occur (mm)	500	500
Slope <sup>a</sup>	Average slope steepness (m/m)	50%	50%
Slsbbsn <sup>a</sup>	Average slope length (m)	50%	50%
Sol_Alb <sup>a</sup>	Soil albedo	50%	50%
Sol_Awc <sup>a</sup>	Available water capacity of the soil layer (mm/mm soil)	50%	50%
Sol_K <sup>a</sup>	Soil hydraulic conductivity (mm/h)	50%	50%
Sol_Z <sup>a</sup>	Soil depth	50%	50%
Surlag	Surface runoff lag time (days)	0	10

<sup>a</sup>Indicates spatially distributed parameters in which baseline values are scaled by a multiplier sampled from the range.

<sup>b</sup>Indicates parameters in which baseline values are scaled by addition/subtraction of a value sampled from the range.



**Table 3.** Ranks of SWAT2005 Streamflow Parameters Obtained Using Sobol' with 2,688 Samples

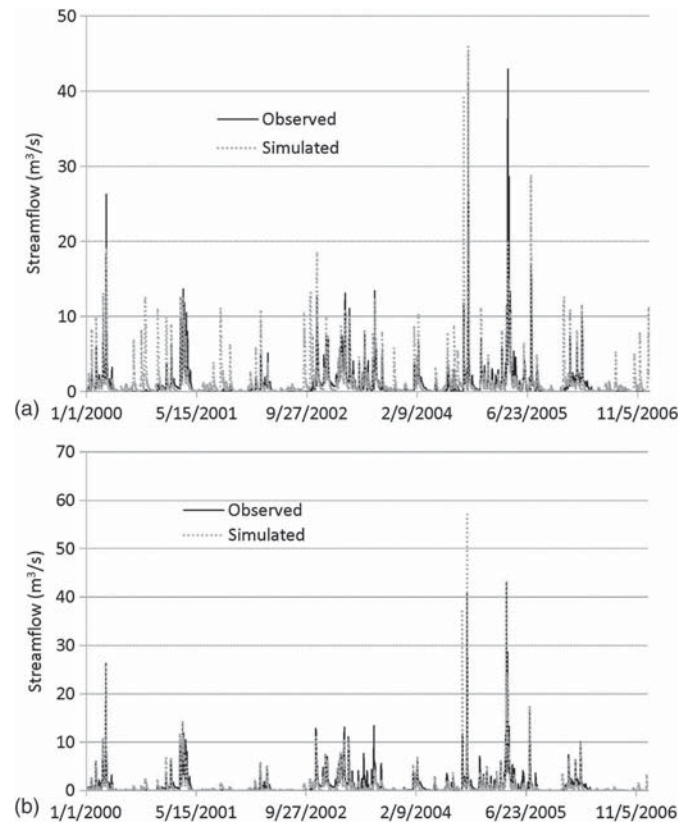
Parameters	All months combined	Months with $C > 0.1$	Months with $C < 0.1$
Cn2	1	1	1
Gwqmn	2	2	3
Alpha_Bf	3	3	2
SoI_Z	4	5	6
Esco	5	4	10
Slope	6	6	5
Ch_K2	7	8	4
Ch_N2	8	7	11
Surlag	9	9	8
SoI_K	10	14	7
Epc0	11	13	14
Canmx	12	12	9
SoI_Awc	13	10	12
Blai	14	11	13
Biomix	15	16	18
SoI_AlB	16	15	16
Slsbbsn	17	17	17
GW_Delay	18	18	15
GW_Revap	19	19	20
Revapmn	20	20	19

Note: Rank 1 is assigned to the most influential parameter and rank 20 is assigned to the parameter that has the least influence on the output.  $C$  is the runoff coefficient.

to be routed in the channel. The result is consistent with the streamflow measurements that show very little (if any) flows at the watershed outlet (Gauge F) during dry months as shown in Fig. 2. Overall, the fact that the SA results agreed with well-known processes that can be assumed a priori provides confidence in application of the season-based SA approach pursued in this study to detect the capability of various hydrologic models to properly describe observed watershed behaviors.

### Calibration

Automatic calibration was then carried out by using the top 10 crucial parameters identified by the SA for the respective time periods. As an alternative to setting the less-sensitive parameters to nominal values, preliminary calibration runs were performed for the conventional calibration in addition to the season-based calibration scenarios. The less-sensitive parameters were assigned the values obtained from the preliminary calibration for the respective scenarios. Maximum iterations of 500 and 4,000 were used for DDS during the preliminary and the main calibration runs, respectively. Results of the calibration exercise are provided in Tables 4 and 5, and in Figs. 2–4. Table 4 shows values of the RMSE, NSE, % Bias, and mean annual streamflow for the conventional and seasonal calibration approaches. Results are provided for the calibration period (i.e., 2000–2003) and the verification period (i.e., 2004–2006) for the calibration site (Gauge F) in addition to the internal gauges (i.e., I, J, K, and M). Fig. 1 may be referred to for the location of the gauges in the watershed. Observation of the conventional calibration results from Table 4 and Figs. 2–4 reveals that SWAT simulated streamflow for the watershed satisfactorily but not very well. However, these results are comparable with past studies that applied SWAT to the Little River Watershed or its sub-basins (van Liew et al. 2007; White et al. 2009).



**Fig. 2.** Comparison of observed and simulated daily streamflow at site F when calibration was done with (a) seasonally invariant influential model parameters; (b) seasonally varying influential model parameters

Table 4 clearly shows superiority of the season-based model calibration over the traditional approach of lumping all seasons together. As demonstrated with multiple goodness-of-fit measures, SWAT2005 simulated streamflow significantly better when model parameters are allowed to vary between wet and dry season instead of assuming seasonal insensitivity. In addition, significant improvement in streamflow simulation accuracy has been achieved at the internal gauges and for the verification period, except for the verification period of Gauge J, when the conventional calibration performed better than the season-based calibration. According to Moriasi et al. (2007), streamflow simulations are generally considered satisfactory if  $NSE > 0.5$  and % Bias is within  $\pm 25\%$ . According to this criterion, the season-based evaluation produced satisfactory results at most gauges for both calibration and verification periods.

Similar conclusions could be drawn regarding superiority of the season-based model evaluation from observation of Figs. 2–4. Fig. 2 compares observed and simulated streamflow for the calibration and verification periods at Gauge F when calibration is performed with conventional techniques as in Fig. 2(a) and when seasonal calibration is performed [Fig. 2(b)]. Fig. 3(a) and 3(b) show observed and simulated streamflow at site K (an internal gauge) when calibration is performed by lumping seasons together [Fig. 3(a)] and when seasonal evaluation is pursued [Fig. 3(b)]. Fig. 4 uses a scatter plot to illustrate the advantage of the season-based calibration for Gauge F. Finally, Table 5 shows the optimal parameter values obtained by using the seasonal and the traditional calibration approaches. Even though all the 20 parameters are listed in Table 5, “optimal” values for the 10 less-sensitive parameters were identified with the preliminary calibration attempt as the

**Table 4.** Efficiency Criteria Values Obtained Using Seasonal and Conventional Calibrations

Gauge	Period	RMSE (m <sup>3</sup> /s)		NSE		Bias (%)		Mean annual streamflow (mm)		
		Seasonal	Conventional	Seasonal	Conventional	Seasonal	Conventional	Observed	Seasonal	Conventional
F	C	1.12	1.91	0.67	0.41	3.3	55.3	248.4	240.0	385.7
	V	2.36	2.57	0.41	0.45	4.8	47.2	282.6	269.0	416.0
I	C	0.45	0.84	0.75	0.42	9.3	48.4	242.5	219.9	360.0
	V	0.89	1.16	0.56	0.46	23.8	21.4	165.2	125.9	200.5
J	C	0.19	0.40	0.79	0.38	12.3	85.1	237.2	266.5	438.8
	V	0.43	0.47	0.37	0.49	36.3	6.1	144.5	91.9	153.2
K	C	0.17	0.34	0.71	0.36	13.4	85.4	151.7	172.1	281.3
	V	0.37	0.52	0.61	0.44	4.2	58.8	55.0	57.3	87.4
M	C	0.06	0.08	0.38	0.26	129.5	216.0	131.0	220.2	373.4
	V	0.11	0.15	0.40	0.28	82.2	186.5	33.4	60.4	95.7

Note: The columns labeled “Seasonal” and “Conventional” provide results of season-based and conventional calibrations, respectively; C = calibration and V = verification.

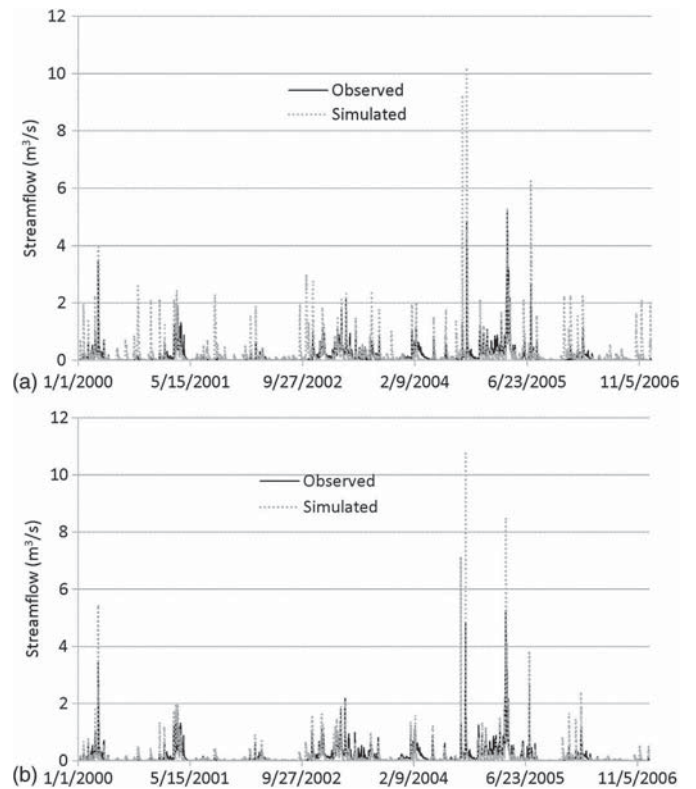
**Table 5.** Optimal Parameter Values Obtained Using Seasonal and Conventional Calibrations

Parameter	Conventional	Seasonal calibration	
		Wet season	Dry season
Alpha_Bf	0.98	1.00	0.72
Biomix	0.02	0.91	0.92
Blai	0.99	0.97	0.45
Canmx	8.61	0.15	9.83
Ch_K2	141.00	87.56	57.32
Ch_N2	0.04	0.08	0.34
Cn2	3.75	11.88	23.80
Epc0	0.52	0.13	0.52
Esco	0.91	0.94	0.02
Gw_Delay	8.71	8.77	5.89
Gw_Revap	0.03	0.01	0.01
Gwqmn	1881.99	170.17	740.67
Revapmn	444.54	474.13	122.11
Slope <sup>a</sup>	15.68	13.86	13.86
Slsbbsn <sup>a</sup>	23.02	45.90	45.90
SoI_Alb <sup>a</sup>	12.42	11.95	11.95
SoI_Awc <sup>a</sup>	22.86	22.90	22.90
SoI_K <sup>a</sup>	3.13	36.18	36.18
SoI_Z <sup>a</sup>	41.51	48.57	48.57
Surlag	1.65	1.67	4.86

<sup>a</sup>The parameter was assumed to be seasonally invariant.

primary calibration runs were executed using the top 10 influential parameters for the respective calibration scenarios.

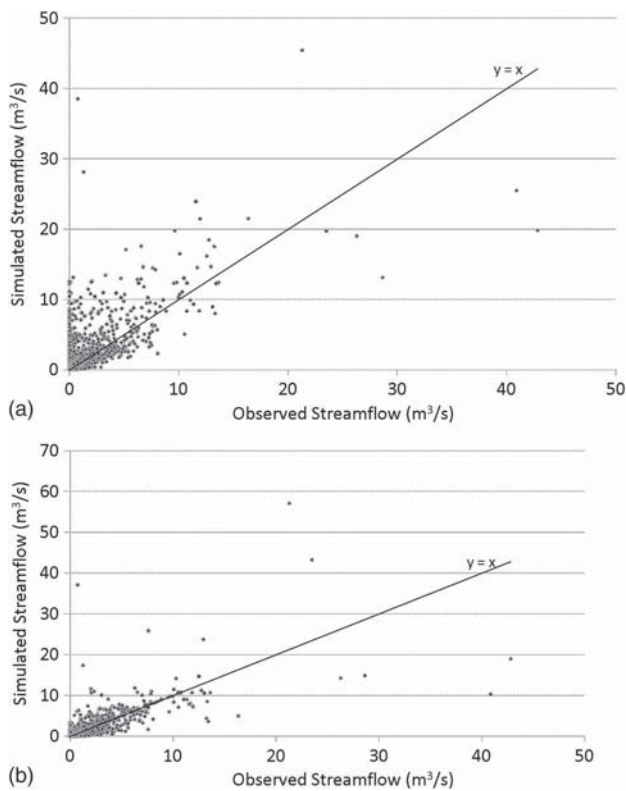
Table 5 shows that for most of the influential parameters, the identified optimal values exhibit considerable sensitivity to season. Close observation of optimal values for the three cases (the conventional, wet season, and the dry season) reveals some insights regarding important physical processes in the watershed during dry and wet seasons. As an example, the optimal values obtained for the most sensitive parameter (i.e., CN2) during the wet season ( 11.88%) and dry season ( 23.8%) indicate that the baseline CN2 values assigned to each HRUs in the watershed need to be reduced by approximately 12 and 24% during the wet and dry seasons, respectively. This agrees with the well-known knowledge that the runoff coefficient would be higher during wet seasons than during dry seasons resulting from antecedent moisture conditions. On

**Fig. 3.** Comparison of observed and simulated daily streamflow at site K when calibration was done with (a) seasonally invariant influential model parameters; (b) seasonally varying influential model parameters

the contrary, the optimal CN2 value obtained for the conventional scenario suggests that the parameter needs to be increased by 3.75%. Increasing CN2 by 3.75% clearly leads to overprediction of streamflow and is confirmed by the results given in Table 4.

The optimal values obtained for the two influential groundwater flow parameters (i.e., Gwqmn and Alpha\_Bf) show that groundwater flow contribution is lower during dry season. A lower Alpha\_Bf indicates faster groundwater flow recession, implying that groundwater contribution to streamflow (if any) would last for a shorter duration in the dry season. A higher Gwqmn indicates the need to have a higher water table level in the shallow aquifer before groundwater starts to contribute to streamflow. Therefore,





**Fig. 4.** Scatter plot comparison of observed and simulated daily streamflow at site *F* when calibration was done with (a) seasonally invariant influential model parameters; (b) seasonally varying influential model parameters

optimal values of both parameters indicate that conditions that favor groundwater contribution to streamflow are more stringent during the dry season. Results of the soil evaporation compensation factor (*Esco*) indicate that evaporation can be extracted from deeper soil levels during the dry season compared with how deep it needs to go during the wet season. Additionally, the surface runoff lag time (*Surlag*) results show that lag time would be longer during dry seasons, which is consistent with knowledge that can be presumed a priori.

The identified seasonality of the influential parameters and their optimal values could be attributed to one or both of the following reasons: (1) The parameter may represent a physical process that exhibits significant seasonality. For example, parameters such as *CN* and Manning's roughness would vary from season to season depending on land cover, especially for agricultural lands and grass lands. Likewise, groundwater parameters such as *Gwqmn* and *Alpha\_Bf* may exhibit seasonality depending on the water table level and soil moisture conditions as described in the previous paragraphs. (2) The input data used for the study and/or structure of SWAT may not adequately describe hydrologic processes of the Little River Watershed, especially when the conventional model evaluation approach is pursued. A detailed diagnostic analysis study could help identify the specific cause of the seasonality exhibited by each parameter. The primary objective of this study was to investigate the advantage of a season-based model evaluation in improving model performance. As such, no attempt has been made to determine the specific contribution of uncertainties owing to model structure, model parameter, or input data to the parametric seasonality exhibited for the Little River Watershed.

## Conclusions

The study compares performance of the traditional model evaluation (SA and calibration) method with a season-based model evaluation approach. A global SA method known as Sobol' (Sobol' 1993; Saltelli 2002) has been applied by using SWAT2005 on the headwaters of the Little River Experimental Watershed, one of the USDA-ARS experimental watersheds. The global SA method has been used to investigate sensitivity of SWAT2005 streamflow parameters at three time periods—annual, months with a low runoff coefficient, and months with a high runoff coefficient—in an attempt to understand the dominant model and watershed behaviors during wet and dry seasons. The DDS algorithm (Tolson and Shoemaker 2007) has been used to calibrate SWAT for the three time periods by using the principle parameters identified by each global SA method. Performance of the calibration results have been verified with the traditional split-sampling approach and by analyzing effectiveness of the model in predicting internal watershed behaviors by comparing simulated streamflow with observations at multiple internal sites that were not used to calibrate the model.

The major conclusions are (1) the season-based SA helped to understand important hydrologic processes during the dry and wet season; (2) compared with the conventional model calibration technique, the season-based model calibration approach pursued in this study has significantly improved model performance for both calibration and verification periods at the calibration site and at multiple internal gauges that were not used for calibration; (3) the optimal parameter values identified by the season-based calibration technique showed significant sensitivity to season; (4) the traditional model evaluation approach of aggregating model parameters across seasons would compromise model performance compared with dynamic (i.e., season-based) model evaluation approach; and (5) similar studies need to be undertaken to confirm these results across various hydroclimatic regions.

## Acknowledgments

The source code for DDS (Tolson and Shoemaker 2007) was obtained from the first author of the code. Comments provided by three anonymous reviewers have helped improve the quality and the readability of the article.

## References

- Abebe, N. A., Ogden, F. L., and Pradhan, N. R. (2010). "Sensitivity and uncertainty analysis of the conceptual HBV rainfall—runoff model: Implications for parameter estimation." *J. Hydrol. (Amsterdam)*, 389(3–4), 301–310.
- Arnold, J. G., Williams, J. R., Srinivasan, R., and King, K. W. (1999). "SWAT: Soil and water assessment tool." *U.S. Dept. of Agriculture, Agricultural Research Service, Temple, TX*.
- ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, and Drainage Division. (1993). "Criteria for evaluation of watershed models." *J. Irrig. Drain. Eng.*, 119(3), 429–442.
- Bosch, D. D., et al. (2007). "Little river experimental watershed database." *Water Resour. Res.*, 43, W09470.
- Bosch, D. D., Sullivan, D. G., and Sheridan, J. M. (2006). "Hydrologic impacts of land-use changes in coastal plain watersheds." *Trans. ASABE*, 49(2), 423–432.
- Campolongo, F., and Saltelli, A. (1997). "Sensitivity analysis of an environmental model: An application of different analysis methods." *Reliab. Eng. Syst. Saf.*, 57(1), 49–69.

- FitzHugh, T. W., and Mackay, D. S. (2000). "Impacts of input parameter spatial aggregation an agricultural nonpoint source pollution model." *J. Hydrol. (Amsterdam)*, 236(1–2), 35–53.
- Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). "The soil and water assessment tool: Historical development, applications, and future research directions." *Trans. ASABE*, 50(4), 1211–1250.
- Geza, M., and McCray, J. E. (2008). "Effects of soil data resolution on SWAT model stream flow and water quality predictions." *J Environ. Manage*, 88(3), 393–406.
- 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 modeling." *J. Hydrol. (Amsterdam)*, 377(1–2), 80–91.
- Hall, J., Tarantola, S., Bates, P. D., and Horritt, M. S. (2005). "Distributed sensitivity analysis of flood inundation model calibration." *J. Hydraul. Eng.*, 131(2), 117–126.
- Helton, J. C. (1993). "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal." *Reliab. Eng. Syst. Saf.*, 42(2–3), 327–367.
- Li, H., Sivapalan, M., and Tian, F. (2010). "Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2 basins: The Blue River and the Illinois River." *J. Hydrol. (Amsterdam)*, in press.
- Matott, L. S., Babendreier, J. E., and Purucker, S. T. (2009). "Evaluating uncertainty in integrated environmental models: A review of concepts and tools." *Water Resour. Res.*, 45(6), W06421.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." *Trans. ASABE*, 50(3), 885–900.
- Morris, M. D. (1991). "Factorial sampling plans for preliminary computational experiments." *Technometrics*, 33(2), 161–174.
- Muleta, M. K. (2010). "Comparison of model evaluation methods to develop a comprehensive watershed simulation model." *Proc., 2010 Conf. of the Environmental and Water Resources Institute*, ASCE, Reston, VA.
- Muleta, M. K., and Nicklow, J. W. (2005). "Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model." *J. Hydrol. (Amsterdam)*, 306(1–4), 127–145.
- Muleta, M. K., Nicklow, J. W., and Bekele, E. G. (2007). "Sensitivity of a distributed watershed simulation model to spatial scale." *J. Hydrol. Eng.*, 12(2), 163–172.
- Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part I—A discussion of principles." *J. Hydrol. (Amsterdam)*, 10(3), 282–290.
- Neitsch, S. L., Arnold, J. G., Kiniry, J. R., and Williams, J. R. (2005). *Soil and water assessment tool-version 2005—Theoretical documentation*, Grassland, Soil and Water Research Laboratory, Temple, TX.
- Saltelli, A. (1999). "Sensitivity analysis: Could better methods be used?" *J. Geophys. Res.*, 104(D3), 3789–3793.
- Saltelli, A. (2000). "What is sensitivity analysis?" *Sensitivity analysis*, A. Saltelli, K. Chan and E. M. Scott, eds., Wiley, New York.
- Saltelli, A. (2002). "Making best use of model evaluations to compute sensitivity indices." *Comput. Phys. Commun.*, 145(2), 280–297.
- Saltelli, A., et al. (2008). *Global sensitivity analysis: The primer*, Wiley, New York.
- Saltelli, A., Tarantola, S., and Chan, K. P. S. (1999). "A quantitative model-independent method for global sensitivity analysis of model output." *Technometrics*, 41(1), 39–56.
- Sheshukov, A., Daggupati, P., Lee, M. C., and Douglas-Mankin, K. (2009). "ArcMap tool for pre-processing SSURGO soil database for ArcSWAT." *Proc., 5th Int. SWAT Conf.*, Texas A&M Univ., College Station, TX.
- Sobol', I. M. (1993). "Sensitivity estimates for non-linear mathematical models." *Math. Modelling Comput. Experiment*, 1(4), 407–414.
- Tang, J., and Zhuang, Q. (2009). "A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based terrestrial ecosystem model." *J. Geophys. Res.*, 114, D15303.
- Tang, Y., Reed, P., van Werkhoven, K., and Wagener, T. (2007). "Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis." *Water Resour. Res.*, 43(6), W06415.
- Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K. (2006). "Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation." *Hydrol. Earth Syst. Sci. Discuss.*, 3(6), 3333–3395.
- Tian, F., Li, H., and Sivapalan, M. (2010). "Model diagnostic analysis of seasonal switching of runoff generation mechanisms in the Blue River basin, Oklahoma." *J. Hydrol. (Amsterdam)*, in press.
- Tolson, B. A., and Shoemaker, C. A. (2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration." *Water Resour. Res.*, 43, W01413.
- U.S. EPA. (2002). "Guidance for quality assurance project plans for modeling." *EPA QA/G-5M. Rep. EPA/240/R-02/007*, Washington, DC.
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R. (2006). "A global sensitivity analysis tool for the parameters of multi-variable catchment models." *J. Hydrol. (Amsterdam)*, 324(1–4), 10–23.
- Van Liew, M. W., Veith, T. L., Bosch, D. D., and Arnold, J. G. (2007). "Suitability of SWAT for the conservation effects assessment project: Comparison on USDA agricultural research service watersheds." *J. Hydrologic Engng.*, 12(2), 173–189.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2008a). "Characterization of watershed model behavior across a hydroclimatic gradient." *Water Resour. Res.*, 44, W01429.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2008b). "Rainfall characteristics define the value of streamflow observations for distributed watershed model identification." *Geophys. Res. Lett.*, 35, L11403.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2009). "Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models." *Adv. Water Resour.*, 32(8), 1154–1169.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V. (2003). "Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis." *Hydrol. Processes*, 17(2), 455–476.
- Wang, X., and Melesse, A. M. (2006). "Effects of STATSGO and SSURGO as inputs on SWAT model's snowmelt simulation." *J. Am. Water Resour. Assoc.*, 42(5), 1217–1236.
- Winchell, M., Srinivasan, R., Di Luzio, M., and Arnold, J. (2008). *ArcSWAT 2.1 interface for SWAT 2005: User's guide*, Blackland Research Center, Temple, TX.
- White, E. D., Feyereisen, G. W., Veith, T. L., and Bosch, D. D. (2009). "Improving daily water yield estimates in the Little River Watershed: SWAT adjustments." *Trans. ASABE*, 52(1), 69–79.