

Improving Model Performance Using Dynamic Evaluation and Proper Objective Function

Misgana Muleta

Abstract

Models have become important decision making aids. Model evaluation (i.e., global sensitivity analysis, calibration and uncertainty analysis), is crucial to improve their prediction accuracy and reduce the likelihood of making decisions that could lead to undesirable policy outcomes. The conventional approach assumes that model parameters are insensitive to season irrespective of the temporal variability of input forcings such as rainfall. This assumption could significantly compromise model performance for low flow seasons and/or high flow seasons depending on the calibration method pursued. This study will demonstrate the advantage of dynamic (seasonal) model evaluation in improving performance compared to the traditional approach. In addition, the impact of the goodness-of-fit criteria (e.g., mean of sum of square of residuals, Nash-Sutcliffe efficiency criteria, volume based efficiency criteria, etc) used as an objective function during automatic calibration on model performance has been examined. Objective functions that would improve the accuracy of simulating high flows as well as low flows were identified. The added values of using multiobjective calibration, over the more widely used single objective calibration, has also been explored. The Little River Experimental Watershed, one of the U.S. Department of Agriculture's experimental watersheds, has been used to illustrate the approaches tested in the study. Soil and Water Assessment Tool is the watershed simulation model used for the work. Results show that the season based model calibration approach significantly improved model performance, and calibration is sensitive to the efficiency measure used as object function. As such, multiple efficiency criteria should be used to report model performance as no single efficiency measure performed consistently well in describing goodness of model results. Another important finding is that parameter values that are significantly divergent from their "true" values may lead to model performance that may be considered near perfect even when judged using multiple efficiency measures underlining the challenge of parameter identifiability.

Introduction

Computer models are routinely used for planning and management of water resources and address the challenges faced by local and global stressors such as climate change, land use change and population increase. However, because models are simplifications of the real world, accuracy of model predictions cannot be taken for granted. Models must be calibrated before they are used as decision making aids to ensure that simulation results are scientifically sound and defensible (U.S. EPA, 2002). Sensitivity analysis (SA), a technique used to identify the relative significance of model inputs, parameters and/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 significant portion of the output uncertainty, and 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. 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. 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. Using global SA, several studies including Tang et al. (2007), Wagener et al. (2003) and van Werkhoven et al. (2008) have demonstrated sensitivity of influential model parameters to season for the watersheds they studied. However, most model evaluation procedures 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 as well as 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 to the conventional approach of assuming seasonal invariability of dominant parameters and their optimal values.

How well a model simulation fits the observed data is evaluated either visually (i.e., graphical comparison of model simulations with observations) or using one or more statistical measures commonly referred to as efficiency criteria or goodness-of-fit criteria. Efficiency criteria are derived from the residual (error) between the simulated and observed output. Many such measures have been used in water resources modeling (Nash and Sutcliffe, 1970; Legates and McCabe, 1999; Krause et al., 2005; Moriasi et al., 2007; Gupta et al., 2009). The Nash-Sutcliffe efficiency (*NSE*) criteria originally proposed by Nash and Sutcliffe (1970) and the root of mean of square of errors (*RMSE*) between the simulated and observed hydrologic variable are the most commonly used efficiency measures (Gupta et al., 2009). In automatic calibration studies, in addition to evaluating how well model simulations fit observed data, one or more efficiency criteria are used as objective function(s) during optimization to help identify parameter sets that result in model simulations that closely match observations. The studies done in the past on *NSE* and other goodness-of-fit criteria examined how well the measure(s) describe model performance. However, no study has investigated sensitivity of model performance to the goodness-of-fit criteria used as objective function during automatic calibration. As such, this study examines the effect of the efficiency criteria used as objective function during automatic calibration on model performance; identifies objective functions that are reasonably sensitive to both low flows and high flows, and ought to be used for single objective automatic calibration attempts and

investigate an efficiency measure that best describes model performance and needs to be used to report goodness of modeling results.

A widely used watershed simulation model known as Soil and Water Assessment Tool (SWAT) has been applied to the headwaters of the Little River Experimental Watershed, one of the United States Department of Agriculture (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 using the principle parameters identified by each global SA methods. The most widely used efficiency criteria have been examined with regard to their effectiveness as objective function during the calibration attempt, and performance of the calibrated models have been evaluated using a number of goodness-of-fit criteria. Performance of the calibration results has been verified using the traditional split-sampling approach as well as 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.

Watershed Simulation Model and Study Watershed

Soil and Water Assessment Tool (SWAT) (Arnold et al., 1999), the simulation model used for this study, is one of the most widely-used and well supported 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, potentially, 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 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 ground water flow. SWAT operates within ESRI's ArcGIS 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, SWAT version 2005 (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 technical theory of the hydrologic processes modeled by SWAT, the reader is referred to Neitsch et al. (2005).

As shown in Figure 1 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. The LREW has been selected because it is heavily gauged for rainfall as well as streamflow (Bosch et al, 2007), and because data are readily accessible online (<ftp://www.tiftonars.org/>) from the Southeast Watershed Research Laboratory (SEWRL). 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 about 31% row crop agriculture, 10% pasture, 50% forest, and 7% urban area (Bosch et al. 2006).

Only the upper 116 km² of the LREW has been used for this study to minimize computational demand of the model, and also because the headwater subwatersheds have denser streamflow and rainfall gauges. Twelve precipitation gauges and five streamflow gauges (see Figure 1) with long-term data (i.e., 1967-2006) are available for the headwaters from the SEWRL. Climate data 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>) as the climate data available from SEWRL starts only from 2004. The geographic data used to setup SWAT model including topography, land use, stream networks, and rainfall and streamflow gauging locations have been obtained from the SEWRL. SSURGO soil map has been obtained from the Natural Resources Conservation Service (NRCS) soil data mart (<http://soildatamart.nrcs.usda.gov/>).

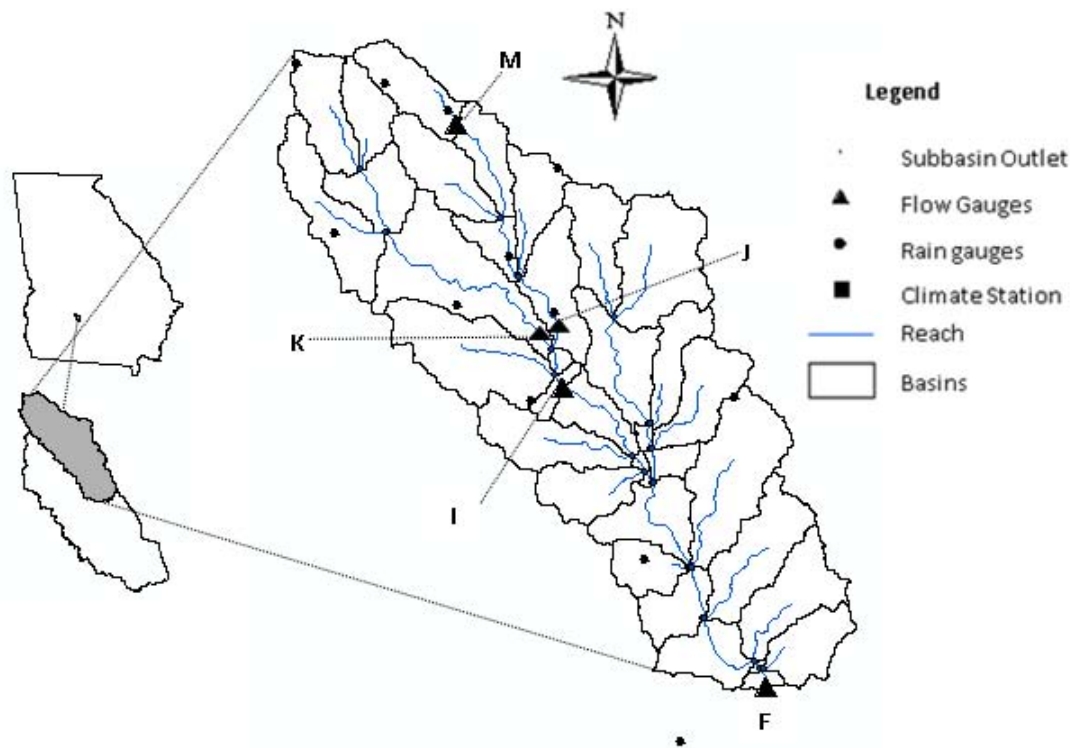


Figure 1. Location Map of the Study Area and the Gauging Stations

Global Sensitivity Analysis and Calibration Methods

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 in to: the sum of the linear (first-order) terms due 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. 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 non-additive models.

Dynamically Dimension Search (DDS) (Tolson and Shoemaker 2007), used for calibration in the study, 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 as 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.

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 mainly from the SEWRL. Missing values for the precipitation data have been filled 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 to 2006 data, mean daily precipitation of the twelve rain gauges in the study watershed varied from 3.18 mm to 3.45 mm. The minimum and the maximum daily rainfall correlation factors among the twelve rain gauges were 0.77 and 0.98, respectively. These results indicate 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 Soil Survey Geographic (SSURGO) soil data used in this study provides the highest resolution soil map for a county-wide soil database in the United States. 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 km² study watershed was delineated and subdivided into 37 subbasins and 96 HRUs using ArcSWAT (Winchell et al. 2008) as shown in Figure 1.

Seasonal Model Evaluation

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). Dominant model structure and parameters may depend on forcings and antecedent conditions (Tang et al. 2007). According to Tang et al. (2007) and van Werkhoven et al. (2008), forcings, mainly 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 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. Runoff coefficient, as used here, is defined as the ratio of areally averaged total monthly rainfall to the total monthly runoff measured at the outlet of the watershed. Areally averaged monthly rainfall totals, monthly runoff totals and monthly runoff coefficients obtained for the watershed are given in Table 1.

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

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. (2008), dynamic parameter sensitivity may not be described based on 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 runoff coefficient greater than 0.1 (i.e., December to April); 2) months with runoff coefficient less than 0.1 (i.e., June to October); 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 where 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 advantage of the dynamic (i.e., seasonally varying) model evaluation technique attempted in this study.

The streamflow data collected at gauge *F* (outlet of the study watershed as shown in Figure 1) was used for the SA as well as calibration. One year data (i.e., 1999) was used as a warm-up period to diffuse the effect of antecedent conditions, and four year data (i.e., 2000 - 2003) was used for the sensitivity analysis as well as calibration. Performance of the calibration attempt was verified using the traditional split sampling approach (i.e., the 2004-2006 data at the calibration site was used for verification) as well as by 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) was used as output for the SA and as objective function for the calibration attempts. Root Mean Square Error (*RMSE*), Percent Bias (*% Bias*), 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/under estimate the observations. Results of the dynamic model calibration and the conventional calibration attempts are summarized in Table 2. The results clearly show superiority of season based model calibration in improving model performance.

Table 2. Efficiency Criteria Values Obtained Using Seasonal and Conventional Calibrations at Gauges F and J

Gage	Period	RMSE (m3/sec)		NSE		% Bias		Mean Annual Streamflow (mm)		
		Seas.	Conv.	Seas.	Conv.	Seas.	Conv.	Observed	Seas.	Conv.
F	Cal.	1.12	1.91	0.67	0.41	-3.3	55.3	248.4	240.0	385.7
	Ver.	2.36	2.57	0.41	0.45	-4.8	47.2	282.6	269.0	416.0
J	Cal.	0.19	0.40	0.79	0.38	12.3	85.1	237.2	266.5	438.8
	Ver.	0.43	0.47	0.37	0.49	-36.3	6.1	144.5	91.9	153.2

Sensitivity of Model performance to Efficiency Criteria

Based on results of the seasonal sensitivity analysis study summarized in Table 2, dynamic model calibration was pursued to examine sensitivity of model performance to the efficiency criteria used during model calibration. Single objective automatic calibration was performed using nine commonly used different efficiency measures as objective function, one efficiency measure per calibration run. Performance of each calibration result was then tested using eleven different efficiency criteria. Streamflow data from gage *F* was used for the calibration. One year data (i.e., 1999 data) was used as a warm-up period to diffuse the effect of antecedent conditions, and four year data (i.e., 2000 - 2003) was used for calibration. Performance of the calibration exercise was verified using the traditional split sampling approach (i.e., the 2004- 2006 data at the calibration site (gage *F*) was used for verification) as well as by analyzing capability of the calibrated model to simulate accurate streamflow at internal gauges not used for calibration (i.e., gages *I*, *J*, *K*, and *M*) for 2000-2006 data.

Two case studies were considered for the research. In the first case, the actual data (i.e., soil, land use, topography, climate and streamflow) available for the watershed were used to build SWAT model. Then, DDS was used to calibrate the model using 9 different objective functions. For each objective function, three separate calibration attempts were made: one for the low flow season only, one for the wet season only, and one using the conventional calibration approach (i.e., both seasons combined) where model parameters are assumed season insensitive. For each calibration run, 3000 SWAT simulations were used for DDS. This means that a total of 27 calibration attempts (i.e., three calibration scenarios for each of the 9 objective functions), where each calibration requires 3000 SWAT simulations, were made for the first case study. In the second case, instead of using the actual streamflow available for the watershed, the streamflow simulated by the calibrated SWAT during the first case study with one of the objective functions was used as observed data, and then calibrations were repeated with each one of the 9 efficiency measures used as objective function one at a time. This second approach would eliminate uncertainties due to model structure and input data from the modeling process as the streamflow simulated by the calibrated model is considered as “observed data”, and the same input data that produced the “observed streamflow” is being used to recalibrate the same model. Furthermore, true values of the 12 sensitive parameters are known and that would help to identify the objective function that produces “optimal” parameters that are the closest to the true values. Therefore, the second case study would help further elucidate relative effectiveness of the 9 efficiency measures as objective function to minimize parametric uncertainty. To minimize the computational demand, only dynamic model evaluation approach was considered for the case study implying that total of 18 calibration attempts were made for the second case. In addition, for both dry season and wet season calibration runs, only 2000 SWAT simulations per calibration attempt were used for this hypothetical scenario.

For both case studies, all twelve sensitive parameters were assumed to follow uniform distribution as done in Muleta and Nicklow (2005), and the lower and upper bounds recommended in Neitsch et al. (2005) were used for majority of the parameters. 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 and/or other watershed behaviors. During the model calibration, the baseline values assigned to each spatially varying parameter were altered by multiplying the base lines by the sampled multipliers, or by adding the sampled values to the baseline. This way, the parameters would be scaled up or down while preserving their spatial variability. Results of the hypothetical scenario are summarized in Table 3.

Conclusions

The major conclusions of the study are 1) the season based model calibration approach has significantly improved model performance and the identified optimal parameter values showed significant sensitivity to season; 2) automatic model calibration is sensitive to the efficiency measure used as object function; 3) multiple efficiency criteria should be used to report model performance as no single efficiency measure performed consistently well in describing goodness of model results; 4) relative to all the efficiency measures tested in the study, *MNSE*D was identified as the most robust with regard to its effectiveness as objective function during automatic model calibration for both low flow seasons as well as high flow season. This shows that *MNSE*D may be a better choice for use as objective function for

single objective calibration applications; 5) parameter values that are significantly divergent from their true values may lead to model performance that may be considered near perfect even when judged using multiple efficiency measures.

Table 3. Efficiency Criteria Values Obtained At Gauge F When Various Efficiency Measures Were Used As Objective Function

Period	Criteria	MAE	RMSE	NSE	NSED	MNSE	MNSESD	LN_NSE	LN_NSED	VE	PBIAS	RSR
Cal.	MAE	0.22	0.57	0.89	0.90	0.78	0.80	0.88	0.88	0.76	-1.79	11.56
	RMSE	0.16	0.28	0.97	0.97	0.84	0.86	0.81	0.80	0.83	3.46	5.61
	NSE	0.30	0.53	0.91	0.91	0.72	0.74	0.79	0.78	0.67	10.59	10.74
	NSED	0.24	0.41	0.95	0.95	0.77	0.80	0.74	0.76	0.74	-5.68	8.19
	MNSE	0.12	0.25	0.98	0.98	0.88	0.90	0.97	0.97	0.87	-0.81	5.02
	MNSESD	0.08	0.22	0.98	0.98	0.92	0.93	1.00	1.00	0.92	-2.14	4.45
	LN_NSE	0.25	0.71	0.84	0.84	0.74	0.77	0.99	0.98	0.72	-7.02	14.44
	LN_NSED	0.16	0.37	0.95	0.95	0.84	0.86	0.97	0.97	0.82	-5.04	7.51
	VE	0.25	0.59	0.86	0.87	0.75	0.77	0.95	0.94	0.72	-5.87	11.97
Ver.	MAE	0.39	1.80	0.70	0.71	0.67	0.71	0.85	0.92	0.62	-3.33	21.85
	RMSE	0.18	0.36	0.98	0.98	0.85	0.87	0.84	0.89	0.82	0.94	4.40
	NSE	0.42	1.61	0.75	0.76	0.68	0.70	0.81	0.87	0.59	5.14	19.46
	NSED	0.33	1.57	0.76	0.77	0.74	0.78	0.78	0.87	0.68	-4.16	19.02
	MNSE	0.13	0.38	0.97	0.97	0.89	0.91	0.97	0.98	0.88	-0.50	4.64
	MNSESD	0.08	0.30	0.98	0.99	0.94	0.94	1.00	1.00	0.92	-1.91	3.60
	LN_NSE	0.27	0.76	0.90	0.91	0.77	0.80	0.96	0.97	0.74	-5.22	9.20
	LN_NSED	0.19	0.54	0.96	0.96	0.85	0.87	0.88	0.90	0.82	-4.37	6.51
	VE	0.30	0.71	0.90	0.91	0.75	0.78	0.90	0.91	0.71	-7.85	8.59

References

- Arnold, J.G., Williams, J.R., Srinivasan, R., King, K.W.(1999). "SWAT: Soil and Water Assessment Tool." U.S. Dept. of Agriculture, Agricultural Research Service, Temple, TX.
- Bosch, D. D., Sheridan, J. M., Lowrance, R. R, Hubbard, R. K., Strickland, T. C., Feyereisen, G. W., Sullivan, D. G.(2007). "Little River Experimental Watershed database." *Water Resour. Res.* 43, W09470, doi:10.1029/2006WR005844.
- Bosch, D.D., Sullivan, D.G., Sheridan, J. (2006). "Hydrologic impacts of land-use changes in coastal plain watersheds." *Transactions of the ASABE*, 49(3), 423-432.
- 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." *Transactions of the ASABE*, 50(4), 1211-1250.
- 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.*, 377 (1-2), 80-91.
- Krause, P., Boyle, D.P., and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, 5, 89-97.
- Legates, D.R. and G.J. McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model evaluation, *Water Resour. Res.* **35**, 233-241.

- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations", *Transactions of the ASABE*, 50(3), 885–900.
- Muleta, M.K., Nicklow, J.W. (2005). "Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model." *J. Hydrol.* 306, 1-19.
- Nash, J.E., Sutcliffe, J.V.(1970). "River flow forecasting through conceptual models. Part I-A discussion of principles." *J. Hydrol.* 125, 277-291.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R.(2005). "Soil and Water Assessment Tool-Version 2005-Theoretical Documentation", Temple, Texas, USA.
- Saltelli, A. (2000). "What is sensitivity analysis?" In: Saltelli, A., Chan, K., and Scott, E. M. (Eds.), *Sensitivity analysis*. Wiley, NY.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J. , Gatelli, D. , Saisana, M., Tarantola, S. (2008). "Global Sensitivity Analysis: The Primer." Wiley, New York.
- Sheshukov, A., Daggupati, P., Lee, M. C., Douglas-Mankin, K.(2009). "ArcMap Tool for Pre-processing SSURGO Soil Database for ArcSWAT." *Proceedings of the 5th International SWAT Conference*, Boulder, CO., Aug. 5-7.
- Sobol', I.M. (1993). "Sensitivity estimates for non-linear mathematical models", *Math. Modeling Comput. Experiment*, 1(4), 407-414.
- Tang, Y., Reed, P., van Werkhoven, K., Wagener, T. (2007). "Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis." *Water Resour. Res.*, 43, W06415, doi:10.1029/2006WR005813.
- Tolson, B. A., Shoemaker, C. A.(2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration." *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.
- U. S. EPA (2002). Guidance for quality assurance project plans for modeling. *EPA QA/G-5M. Report EPA/240/R-02/007*, Washington, D.C.
- van Werkhoven, K., Wagener, T., Reed, P., Tang, Y. (2008). "Rainfall characteristics define the value of streamflow observations for distributed watershed model identification.", *Geophys. Res. Lett.*, 35, L11403, doi:10.1029/2008GL034162.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., Gupta, H. V. (2003). "Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis." *Hydrol. Process.* 17(2), 455– 476.
- White, E.D., Feyereisen, G.W., Veith, T.L., Bosch, D.D. (2009). "Improving Daily Water Yield Estimates in the Little River Watershed: SWAT Adjustments", *Transactions of the ASABE*, Vol. 52(1): 69-79
- Winchell, M., Srinivasan, R., Di Luzio, M., Arnold, J. (2008). "ArcSWAT 2.1 Interface for SWAT 2005: User's Guide", Blackland Research Center, Temple, Texas.