Article type : Technical Paper

Assessing Parameter Uncertainty of a Semi-Distributed Hydrology Model for a Shallow Aquifer Dominated Environmental System

S. Samadi, D. L. Tufford, and G. J. Carbone

Research Assistant Professor (**Samadi**), Department of Civil and Environmental Engineering, Research Associate Professor (**Tufford**), Department of Biological Sciences, and Professor (**Carbone**), Department of Geography, University of South Carolina, 300 Main Street, Columbia, South Carolina 29208 (E-Mail/Samadi: samadi@cec.sc.edu).

Abstract: This paper examines the performance of a semi-distributed hydrology model (i.e., Soil and Water Assessment Tool; SWAT) using Sequential Uncertainty FItting (SUFI-2), Generalized Likelihood Uncertainty Estimation (GLUE), Parameter Solution (ParaSol), and Particle Swarm Optimization (PSO). We applied SWAT to the Waccamaw watershed, a shallow aquifer dominated Coastal Plain watershed in the southeastern United States. The model was calibrated (2003-2005) and validated (2006-2007) at two US Geological Survey (USGS) gaging stations using significant parameters related to surface hydrology, hydrogeology, hydraulics, and physical properties. SWAT performed best during intervals with wet and normal antecedent conditions with varying sensitivity to effluent channel shape and characteristics. In addition, the calibration of all algorithms depended mostly on Manning's *n*-value for the tributary channels as the surface friction resistance factor to generate runoff. SUFI-2 and PSO simulated the same

relative probability distribution tails to those observed at an upstream outlet, while all methods (except ParaSol) exhibited longer tails at a downstream outlet. The ParaSol model exhibited large skewness suggesting a global search algorithm was less capable of characterizing parameter uncertainty. Our findings provide insights regarding parameter sensitivity and uncertainty as well as modeling diagnostic analysis that can improve hydrologic theory and prediction in complex watersheds. ***Editor's note****: This paper is part of the featured series on SWAT Applications for Emerging Hydrologic and Water Quality Challenges. See the February 2017 issue for the introduction and background to the series.*

(**Key Terms:** SWAT; Parameter Uncertainty; Shallow Aquifer; Modeling Diagnostic Analysis.)

INTRODUCTION

Rivers in the US Atlantic Coastal Plain flow through broad, densely vegetated floodplains with low topographic gradient, and are strongly influenced by groundwater dynamics of a shallow water table. Hydrologic simulation models must capture these interactions and to estimate flow accurately that aid sustainable water resources management in this rapidly growing region. Thus, some researchers have applied the Soil & Water Assessment Tool (SWAT) model to the forested wetland watersheds of the southeastern US (SEUS) Coastal Plain (Wu and Xu, 2006; Feyereisen et al., 2007; Lam et al., 2010; Amayta and Jha, 2011; Joseph and Guillaume, 2013; Samadi et al., 2014; Samadi, 2016).

However, watershed modeling in the Coastal Plain landscape has inherently high levels of uncertainty associated with stochasticity, model structure, input parameters, initial conditions and measurement errors. Structural uncertainty comes from over-simplification of hydrological processes in the conceptual structure of a hydrology model, including the effects of wetlands or reservoirs (e.g., Yang *et al*., 2008), surface/groundwater interaction (e.g., Tian *et al*., 2012), or infrastructure (e.g., Yang *et al*., 2008). Input parameter uncertainty derives from spatial interpolation of model inputs (e.g., Baffaut et al., 2015a) or parameter non-uniqueness. Initial conditions and measurement errors are caused by in situ observations that can be collectively considered as data errors.

In recent decades, various mathematical methods have been developed to treat uncertainty and calibrate hydrology models. Methods to represent model parameter, state, and prediction uncertainty include Bayesian approaches (Beven and Binley, 1992; Kuczera and Parent, 1998; Thiemann et al., 2001; Vrugt et al., 2003b, 2009a; Abbaspour et al., 2007; Tonkin and Doherty, 2009; Moore et al., 2010; Schoups and Vrugt, 2010; Laloy and Vrugt, 2012; Pourreza-Bilondi and Samadi, 2016; Pourreza-Bilondi et al., 2016), set-theoretic (Klepper et al., 1991; Vrugt et al., 2001), sequential data assimilation (Madsen et al., 2003; Vrugt et al., 2005; Moradkhani et al., 2005), stochastic optimization techniques (Duan et al., 1992; Eberhart and Kennedy, 1995), and multi-model averaging methods (Georgekakos et al., 2004; Ajami et al., 2007; Vrugt and Robinson, 2007). These methods differ in mathematical rigor, underlying assumptions about the residual error distribution, and how explicitly those assumptions are expressed in the modeling procedure.

Among these, optimization methods and Bayesian algorithms are well known techniques available to research communities. Optimization is an appropriate method for solving continuous and discrete problems. Bayesian methods enable calibration of complex multivariate distributions by casting them as the invariant distribution of a Markov chain (MC: McMillan and Clark, 2009). The Particle swarm optimization (PSO; Eberhart and Kennedy, 1995) and the Parameter Solutions (ParaSol; modified algorithm of Duan et al., 1992) procedures are two optimization and uncertainty analysis techniques used in this study. The generalized likelihood uncertainty estimation (GLUE; Beven and Binley, 1992) and the Sequential Uncertainty FItting algorithm (SUFI-2; Abbaspour et al., 2004) are two well-known (informal) Bayesian models that propose a pragmatic approach to uncertainty estimation (Schoups and Vrugt, 2010). Both Bayesian and optimization algorithms have found widespread applications in hydrology because of their flexibility and suitability for applying in a number of simple to complex research problems.

Despite the importance of addressing uncertainty in complex hydrological systems, few studies have compared various uncertainty algorithms in the coastal plain watershed, and these are either limited to applications of simple uncertainty model (e.g., Amatya and Jha 2011) or one Bayesian algorithm (e.g., Joseph and Guillaume, 2013; Samadi, 2016). In particular, no relevant studies

have used a range of uncertainty and optimization experiments to simulate hydrological variables in the Coastal Plain watersheds with extensive alluvial and non-alluvial forested wetlands and riparian storage.

This study is thus a first attempt to link the SWAT semi-distributed hydrology model with various optimization and uncertainty algorithms (i.e., PSO and ParaSol, GLUE and SUFI-2) in a coastal plain watershed of SEUS. The first objective of the study is to identify and quantify the uncertainty associated with daily streamflow simulation using various techniques during wet, moderate, and dry hydrological conditions. The second objective is to investigate: (i) how parameter uncertainty affects model performance? (ii) how the sensitivity of SWAT parameters varies in different uncertainty algorithms? and (iii) what is likely the role of shallow water table in runoff generation? Here, we aim to better understand the interaction between surface and subsurface flow and the way in which shallow aquifer shapes and controls Coastal Plain simulation. This is a substantial challenge as this type of simulation is tempered by the heterogeneity in physical characteristics of watershed (soil moisture, evapotranspiration, etc.), complexities in storage capacity due to extensive alluvial and non-alluvial forested wetlands and riparian storage, and difficulties in computing subsurface contribution to the river system (see Shirmohammadi et al., 1986; Amatya and Jha, 2011). In the pursuit of model parameter uncertainty, we proposed several diagnostic analyses to test model identification and information content of simulation, and to ensure that models are 'working for the right reasons' (e.g., Kirchner, 2006).

The SWAT Calibration and Uncertainty Programs (SWAT-CUP; version 2012) framework was used to infer these techniques and to assess parameter sensitivity. Using multiple uncertainty algorithms, we present an integrated picture of the state of the art of hydrological simulation for a complex landscape from a parameter uncertainty perspective. We evaluate the models in terms of parameter uncertainty and their ability to simulate daily observed streamflow. Accurate hydrogeological parameterization of a Coastal Plain watershed and consequent improvement in explaining model sensitivity and parameter uncertainty, and diagnostic analysis are also explored.

METHODS

Study Area

The Waccamaw River watershed (hydrologic unit code 03040206, area=311, 685 ha) is on the lower Coastal Plain of eastern North and South Carolina (Figure 1). The watershed has little topographic gradient (99% is < 5% slope), wide floodplains, and complex groundwater characteristics due to poorly drained soils, a shallow water table, and extensive wetlands. Elevation ranges from 6 – 46 m above mean sea level. The watershed is in a humid subtropical climate with hot summers and mild winters. Precipitation in the basin falls almost exclusively as rainfall, with an annual average of 1309 mm during the study period (2003-2007).

The downstream reaches of the Waccamaw River are tidally influenced, so this modeling study focused on the non-tidal portion of the river. Streamflow data from two US Geological Survey (USGS) gaging stations, at Freeland (34°05'42N, 78°32'54W) and Longs (33°54'45N, 78°42'55W), were used as subwatershed outlets (Figure 1). Daily precipitation, minimum and maximum temperature, wind speed, and solar radiation were obtained from climate stations at Loris, Whiteville7, and Longwood, all located in North Carolina (Figure 1). Data from these stations were interpolated using Thiessen polygon and Inverse Distance Weighting (IDW) methods to capture the spatial continuity of rainfall fields in the study area.

Waccamaw land use information was obtained from USGS National Land Cover Data portal on 13 Sep 2012 (http://viewer.nationalmap.gov/viewer/). Forested wetlands were the dominant land use, occupying approximately 28% of the watershed (Table 1). Agricultural uses were 26% and developed uses (residential, commercial, and industrial) were 5%. Much of the developed land is widely distributed in the watershed.

Five soil series (i.e., Meggett, Croatan, Rains, Woodington and Norfolk) account for 78% of the watershed area and four land cover classes (i.e., WETF, FRSE, AGRR, and RNGB) cover 86% of the area (see Table 2). Meggett soils are found in the broad floodplains along lower Coastal Plain streams and rivers. Woodington, Rains, and Norfolk soils are found in interfluvial areas. All five series are hydric soils. Croatan soils are often found in non-alluvial wetlands. They are exclusively in the eastern section of the watershed, which is part of the Green Swamp (Riggs et al., 2000).

Approximately 90.5% of the soils are one of four series, all of which are either hydrologic group B, D, or B/D. Only 9.5% of the soils are hydrologic group A; there is no group C soils (Table 2). Hydrologic group D soils (poorly drained) are adjacent to the main channel and hydrologic group B and B/D soils (mostly of the Rains series in the downstream reaches) adjacent to those.

SWAT Model

SWAT is a watershed modeling program developed by the US Department of Agriculture (USDA)–Agricultural Research Service (Arnold *et al*., 1998) to simulate hydrology and water quality at various scales. It was developed to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large complex watersheds with varying soils, land use, and management conditions (Neitsch et al., 2004). ARCSWAT 2009 (Aeronautical Reconnaissance Coverage SWAT; version2009.10.1) was used for this research. The SWAT system is embedded within a geographic information system (GIS) that integrates various spatial environmental data including soil, land cover, climate, and topographic features. Information on major soil types, soil hydrologic groups, and soil layer properties was obtained from USDA STATSGO database (http://websoilsurvey.sc.egov.usda.gov/App/HomePage.html). This simulation was initially performed within ARCSWAT; then the outputs were linked to different uncertainty models. We also ran SWAT manually when it was needed for instance, when we used different modules for evapotranspiration estimation methods, river routing model, model optimization, etc. User can change the SWAT output files (e.g., .rch) and rerun the SWAT executable file.

ARCSWAT subdivided the Waccamaw River watershed into 28 sub watersheds (see Figure 1) and 2020 Hydrologic Response Units (HRUs) connected by a stream network. The HRUs vary in terms of land cover, forest-covered area, cultivation, and hydrologic behavior. HRU definition was established at a 0% threshold for slope, soil, and land use layers to avoid missing small land use units.

The hydrological cycle simulated in SWAT is based on the water balance equation found in Arnold et al., (1998). This research used the Priestly-Taylor (P-T; Priestly and Taylor, 1972)

evapotranspiration method because this radiation-based method produced better results in a forested wetland ecosystem with wet and humid surfaces (Samadi, 2016). The Muskingum method (Schroeter and Epp, 1988) was used for channel flow routing module for this study. SWAT estimates the surface runoff volume from each HRU using either the Soil Conservation Service (SCS) curve number (CN) method (USDA-SCS, 1972) or a new CN method that includes a one-parameter soil moisture depletion curve (see Kannan et al., 2008) to account for antecedent soil moisture conditions (CN ranges 25 to 98 in the SWAT model). We used the improved one-parameter depletion coefficient for adjusting the CN based on plant ET. This new approach is more suitable for watersheds with low storage soils and dense vegetation (see Kannan et al., 2008).

In this research, SWAT was calibrated using 18 parameters related to surface runoff, soil properties, river hydraulics and groundwater parameters (see sensitive parameter section for the list of parameters). In addition, global sensitivity analysis was implemented to perform the sensitivity analysis. This was conducted by allowing all parameters to be changed from their absolute values during the simulation period (see Samadi and Meadows, 2017 for further information).

SWAT was calibrated using different starting points. Each simulation period was shifted by 1 year, such that subsequent periods have 2 years of data in common. Overall, five different calibration periods were considered (i.e., 2003-2005, 2004-2006, 2005-2007, 2003-2006, and 2004-2007) and parameter sensitivity was evaluated for each data set. Model sensitivity did not vary significantly during those subsequent periods, suggesting that 3 years of daily streamflow data contain enough information about parameter estimation for the watershed. Since no significant variations in parameter estimates between different calibration periods were anticipated, this study used 2003-2005 and 2006-2007 as calibration and validation periods, respectively. We used 2000-2002 as warm-up period as well.

Figure 2 illustrates the SWAT modeling steps and the link between SWAT and the uncertainty algorithm at the watershed scale. Inputs for SWAT come from observed meteorological records (rainfall, temperature, etc.) and physical variables (land use, elevation) in the basin. These

variables drive empirically-derived estimates of baseflow and the more processed-based SWAT model. Simulated hydrological response (i.e., modeled streamflow) was calibrated using optimized parameters and compared to measured streamflow. The fit between the simulated and observed values was quantified with a variety of objective functions. Note that while uncertainties in rainfall-runoff prediction come from variety of sources (i.e., model structure, parameters, initial conditions, and measurement error, Vrugt et al., 2003; Liu and Gupta, 2007), this study focuses on parameter uncertainty associated with daily streamflow modeling.

SWAT-CUP Program

The SWAT Calibration and Uncertainty Procedures program is designed to aid model calibration by integrating several calibration/uncertainty analysis approaches using a single interface. Each SWAT-CUP calibration algorithm runs until an objective function is satisfied (in this study Nash-Sutcliffe Efficiency (NSE); Nash and Sutcliffe, 1970). Users select the candidate parameters considering dominant hydrological processes based on knowledge of a basin or in reference to those suggested in the literature. The four algorithms in this research are briefly described below. More descriptions can be found in Abbaspour (2013).

SUFI-2: The SUFI-2 technique, as a Bayesian framework, starts by assuming a large parameter uncertainty within a physically meaningful range, so that the measured data initially fall within 95PPU (95% prediction uncertainty band; 95PPU). The uncertainty band is narrowed during each iteration while monitoring the p- and r-factors (defined in Model Performance Evaluation section). In each iteration, SUFI-2 updates the previous parameter ranges by calculating the sensitivity matrix (equivalent to Jacobian) and the equivalent of a Hessian matrix (Magnus and Neudecker, 1988), followed by the covariance matrix. Parameters are then updated in such a way that new ranges are always smaller than previous ranges and are centered on the best simulation (Abbaspour *et al.*, 2004 and 2007). In this algorithm, uncertainty of input parameters is depicted as uniform distributions, while model output uncertainty is quantified at the 2.5 % and 97.5 % levels of the cumulative distribution function (CDF) of the output variables obtained through the Latin Hypercube Sampling method (McKay et al., 1979). Thus, parameter uncertainty reflects all potential sources: the conceptual model, model inputs, and model parameters. SUFI-2 first

defines the goal function $g(\theta)$ with meaningful parameter ranges $[\theta_{abs\min,}\theta_{abs\max}]$. Then, a Latin Hypercube sampling is carried out in the initial hypercube $[\theta_{abs\min,}\theta_{abs\max}]$ to evaluate the corresponding goal function and to perform the sensitivity analysis (based on sensitivity matrix and the parameter covariance matrix). The 95PPU of a sensitive parameter is then computed to find the best solutions (i.e., parameters that produce the optimal goal function). The 95PPU and two indices (i.e., p-factor and r-factor; defined in Model Performance Evaluation section) are then calculated. If those two factors have satisfactory values, the parameter range travels from $[\theta_{abs\min,}\theta_{abs\max}]$ to $[\theta_{\min,}\theta_{\max}]$ as the posterior parameter distribution. Based on Bayes' theorem, the probability density of the posterior parameter distribution will be driven from the prior density function to calculate the likelihood function of the model (see Yang et al., 2008). SUFI-2 (or any other algorithms) uses the same parameter sets for sensitivity analysis, uncertainty assessment and calibration. Three iterations, with 500 runs in each iteration, were conducted for calibration using the SUFI-2 method (and for the rest of algorithms except GLUE). After each iteration, the parameter ranges were updated as the prior distribution for the next iteration.

GLUE: The generalized likelihood uncertainty estimation methodology of Beven and Binley (1992), inspired by the Hornberger and Spear (1981) method of sensitivity analysis, was one of the first attempts to represent predictive uncertainty in hydrology. This method maps the uncertainty in the modeling onto the parameter space and operates within the context of Monte Carlo analysis coupled with Bayesian estimation and propagation of uncertainty (Blasone et al., 2008). This approach, based on the concept of equifinality (Beven and Binley, 1992), argues that many parameter sets can typically provide acceptable (or ‘‘behavioral’’) simulations of an environmental system. GLUE randomly samples a large number of parameter sets from the prior distribution and each set is classified as either behavioral or non-behavioral through a comparison of the “likelihood measure” with the given threshold value. Finally, the prediction uncertainty is described as prediction quintile from the cumulative distribution realized from the weighted behavioral parameter sets (Abbaspour, 2013). Five GLUE simulation runs were performed with sample sizes of 2,000, 6,000, 8,000, 12,000 and 20,000. Different threshold values (i.e., 0.5, 0.60, 0.65, and 0.70) were also chosen. The result of 20,000 runs with the threshold value of 0.60 was used to summarize the 95 PPU in this study.

ParaSol: The ParaSol method aggregates objective functions (OF's) into a global optimization criterion (GOC) and minimizes these OF's or GOC using the Shuffle Complex (SCE-UA) algorithm (Abbaspour, 2013). The SCE algorithm is a global search algorithm that combines the direct search method of the simplex procedure with the concept of a controlled random search of Nelder and Mead (1965), a systematic evolution of points in the direction of global improvement, competitive evolution (Holland, 1975), and the concept of complex shuffling (Abbaspour, 2007). The SCE algorithm minimizes a single function for up to 16 parameters (Duan et al., 1992) and selects an initial "population" by random-sampling throughout parameter space to optimize each model parameter. After optimization of the modified SCE-UA, the simulations are divided into "good" and "not good" simulations using a threshold value (similar to the GLUE methodology). Next, the prediction uncertainty is constructed from "good" simulations based on an objective function (here NSE).

PSO: Particle swarm optimization is a population-based stochastic optimization technique and shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). Compared to GA, PSO is easier to implement and there are fewer parameters to adjust (Abbaspour, 2013). In each iteration, each particle is updated by following two "best" values (Abbaspour, 2013). The first one is the pbest solution, the best solution so far (fitness). Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population (Eberhart and Kennedy, 1995). Initially, the upper and lower bounds of sensitive parameters are specified in the PSO algorithm. The values for the parameters are generated randomly within the bounds for each particle. These parameters are then fed into the SWAT model. Next, the fitness function is evaluated based on a defined objective function. The fitness evaluation of the particle is compared with the pbest value. If the current value is better than pbest, pbest is set equal to the current value in dimensional space. The fitness evaluation is then compared with the overall prior best value. If the current value performs better than gbest (i.e., global best, the best values of each individual), gbest is reset to the current value of the particle's array index. The velocity and position of the particle are then changed according to the current value. Next, the particle navigates a potential solution toward pbest and gbest

through stochastic mechanism. More information on PSO algorithm can be found in Eberhart and Kennedy, (1995).

Model Evaluation and Diagnostics

The basic criteria used to evaluate model performance are: NSE, Mean Square Error (MSE), p-factor, and r-factor. The NSE (Equation 1) is used frequently to evaluate model performance (Nash and Sutcliffe, 1970).

$$NSE = 1 - \frac{\sum_{t=1}^{N}(O_i - S_i)^2}{\sum_{t=1}^{N}\left(O_i - \bar{O}\right)^2} \quad \text{(Equation 1)}$$

Where $\bar{O}$, O_i and S_i are mean of observed data, daily observed and simulated values, respectively. NSE is sensitive to extreme values like peak flow. NSE values between 0.0 and 1.0 are generally considered acceptable, with 1.0 being a perfect match between the model and measured data (NSE>0.65 considers good and very good), the values between 0.50 and 0.65 are considered satisfactory and NSE<0.50 is unsatisfactory (see Moriasi et al., 2007). Values less than 0.0 indicate the mean of the observed values are better predictors than the model (e.g., Krause et al., 2005; Moriasi et al., 2007, 2015). Limitations of the NSE are oversensitivity to extreme values because differences are squared (Harmel and Smith, 2007; Willmott et al., 2015) and sensitivity to hydrograph shape (Moussa, 2008).

The MSE (Equation 2) is defined as:

$$MSE = \sum_{j=1,n}\left[x_{i,measured} - x_{i,simulated}\right]^2 \quad \text{(Equation 2)}$$

Where *n* is the number of pairs of measured (*x*-measured) and simulated (*x*-simulated) variables and *j* represents the rank (Abbaspour, 2013). The MSE is highly influenced by the magnitude of the difference between observed and simulated values. It is useful for assessing the relative magnitude of the bias among models (Abbaspour, 2013).

The p-factor is the percentage of data bracketed by a 95PPU (maximum value is 100%), and an r-factor (or d-factor; Equations 3 and 4), is the average width of the uncertainty band divided by the standard deviation of the corresponding measured variable (minimum value is zero; Abbaspour et al., 2004, 2007). Theoretically, the value for p-factor ranges between 0 and 100%, while the r-factor ranges between 0 and infinity. A p-factor of 1 and r-factor of zero represents a simulation that corresponds exactly with measured data.

$$R-factor = \frac{\overline{d_x}}{\sigma_x} \quad \text{(Equation 3)}$$

$$\overline{d_x} = \frac{1}{k}\sum_{i=1}^{k}(X_U - X_L) \quad \text{(Equation 4)}$$

Where $\overline{d_x}$ is the average distance between the upper and the lower 95PPU, X_U and X_L represent the upper and the lower boundaries of the 95PPU, while σx is the standard deviation of the measured data.

From these indices, we selected NSE as the likelihood function since it is widely used as an objective measure in hydrologic simulations (Beven and Freer, 2001; Yang et al., 2008). The cutoff threshold to separate "behavioral" from "non-behavioral" parameter sets is another subjective choice within an uncertainty algorithm. The selection of the threshold value is an entirely arbitrary choice that affects prediction uncertainty (e.g., Montanari, 2005; Mantovan and Todini, 2006). Small cutoff thresholds lead to larger "behavioral" simulations and larger uncertainty bands whereas greater threshold values decrease the numbers of "behavioral" models and reduce 95PPU (Blasone et al., 2008; Sellami et al., 2013). Based on our trial-error approaches, when we set SWAT's objective function on NSE values, and considered cutoff thresholds in a range of 0.55 (SUFI-2, PSO) to 0.60 (GLUE and ParaSol), the model simulated low flow, peak flow, and hydrograph recession curves more accurately.

We also compared the output from the four uncertainty algorithms in terms of parameter sensitivity by checking the p-value (values close to zero suggest high level of significance) and the t-stat (a measure of sensitivity, larger absolute values are more sensitive) of calibration results. In addition, model diagnostic analysis including violin plots, marginal posterior

distribution, and autocorrelation plots of residual errors, were used to explore error significance in simulation.

A violin plot is a boxplot combined with kernel density plots, added on each side of the boxplot, to exhibit the probability distribution of the data set. A marginal posterior distribution specifies the posterior distribution for a subset of the model parameters independent from other parameters. In probability theory, posterior density aims to display how the probability of (optimized) sensitive parameter ranges can vary as a function of the integral of the parameter's density over the area under the density function (the lower and upper values of the optimized parameter). It is calculated by dividing the range of optimized parameters obtained by different uncertainty algorithms, into a number of discrete "bins" of equal width. The degree of bias in parameter estimation can be also checked by inspecting the densities of parameter distribution. This approach was applied to high sensitive parameters obtained by four uncertainty algorithms.

In the context of model diagnostic analysis, checking for autocorrelation is typically a sufficient test of randomness since the residual errors from poor fitting models tend to display non-subtle randomness (i.e. deterministic randomness that lacks a pattern). Autocorrelation of residuals reveals information not accounted for in the (uncertainty) model; indicating heteroscedasticity and nonnormality in the modeling (see Sorooshian and Dracup, 1980). These diagnostic approaches provided reliable insights into the behavior and magnitude of simulation error over time.

RESULTS

Flow Calibration and Validation Using the SUFI-2 Algorithm

SWAT performed well when parameters were fitted with the SUFI-2 algorithm, achieving NSE values greater than 0.71 during the calibration and validation periods (Table3; Figures 3 & 4). The p-factor and r-factor indicated that there was some uncertainty in simulating low flows at both Freeland and Longs outlets. When parameters were fitted with SUFI-2, SWAT consistently overestimated streamflow during the summer periods (low flow), which suggests that characteristics of base flow are not well represented by the model. It is also possible that ET was

underestimated during the growing season or that spatial and temporal variability of precipitation was not appropriately characterized in the model.

The 95PPU computed by SUFI-2 bracketed most observations during 2004, 2005, and 2006, whereas it slightly overestimated the magnitude and shape of recession curves in 2003 and 2007 (Figures 3 and 4). This indicates high rates of uncertainty in the recession calculation within SWAT as seen elsewhere (e.g., Yang et al., 2008). To further understand hydrologic processes in the basin, discharges during wet (2006), and dry (2007) years were validated individually. The wet year p-factor was over 90% of the observation while the dry year exhibited lower p-factor (51%; results not shown here). Our results demonstrated that precipitation in the wet period primarily contributes to surface flow and lateral flow while shallow aquifer is the primary contributor to river discharge during dry periods (e.g., Samadi and Meadows, 2017). Because the knowledge of SWAT parameters for simulating shallow subsurface flow is insufficient, the uncertainty of model parameters in simulated results was larger in dry years than in wet years.

Calibration and Validation Using the GLUE Algorithm

The Monte Carlo random sampler demonstrated that many parameter solutions with similarly good values of the NSE (>0.60) were within the posterior ranges. The NSE is 0.64 and 0.87 for calibration and 0.81 and 0.87 for validation at the Freeland and Longs outlets, respectively. These performances indicate that when the GLUE algorithm is used to calibrate SWAT parameters, there is good agreement between measured and simulated flows in both calibration and validation periods (Table 3; Figures 5 and 6) as most of the surface flow observations are bracketed by the 95PPU (>72%). The relatively larger r-factor value for the GLUE model compared to other methods indicates larger uncertainties may remain in simulation and GLUE overpredicted 95PPU as reported by Setegn et al. (2010) and Sellami et al. (2013).

Calibration and Validation Using the ParaSol Algorithm

ParaSol results indicated that most of the calibrated parameters influenced daily streamflow calibration. The sensitive parameters yielded by ParaSol were entirely different from other algorithms and thus showed significant variation in the ranking of sensitive parameters. While the method presented a good NSE, it did not bracket the observed flow satisfactorily, especially

during wet periods. For instance, only 15% of measurements at the Freeland gage and 25% at the Longs gage were bracketed by 95PPU during the calibration period (see Figures 7 and 8). Although, ParaSol approximated the global maximum of NSE values, it produced uncertainty band that was too narrow because of inaccurate assumptions about independently and normally distributed errors made by the model (see Yang et al., 2008). Thus, ParaSol estimated only a small portion (15%-25% percent) of the parameter uncertainty. ParaSol failed to summarize and derive the satisfactory prediction uncertainty; however, the best simulation matches observations well. This is related in part to error associated with the model structure, measurement and initial condition, which results in an underprediction of 95PPU.

Calibration and Validation Using the PSO Algorithm

PSO results indicated that most of the surface flow observations were bracketed by the 95PPU i.e., 87% and 79% of the observations during the calibration period in the upstream and downstream outlets. Periods of overestimation indicated that there were still mismatches, especially in the recession calculation (Figures 9 and 10). The NSE values were high for both calibration and validation periods at upstream and downstream portions (see Table 3). PSO produced large r-factor values (large uncertainty band) for both the calibration (>1.39) and validation (>1.23) periods. This might be attributed to the fact that PSO disregards parameter uncertainty associated with other sources such as the conceptual model and input data that are explicitly accounted by SUFI-2.

Comparing the Results of Four Uncertainty Algorithms

The model performance criteria used in this research were acceptable for all algorithms with SUFI-2 having the best overall fit and ParaSol the worst (Table 3). Some of the performance values were higher during validation than calibration. This may be because the validation interval was generally wetter than the calibration period so the relatively poor performance of the model during low flows would have less effect on model performance.

SUFI-2, PSO, and GLUE produced the closest simulation of annual flows in most years. However, to some extent, some of peak flow values were not well simulated, which may be due to repeated precipitation events that leave the surficial aquifer close to saturated in the watershed.

Another important reason is that SWAT shows high sensitivity to both surface and subsurface runoff (see Sensitive Parameters section), although the timing of the peak is well reproduced at both outlets. In addition, the daily rainfall inputs and daily time step calculations are not sufficient during large storm events that occur at the end of a day since watershed response occurs a day later than simulated. Additionally, measured streamflow data have uncertainties, especially after major rainfall events when there is large or lengthy floodplain residency and standard rating curves do not apply (e.g., Herschy, 2008). Another reason may be attributed to the complex interactions among the water table, soils, floodplain, and stream channel that are highly variable and dependent upon antecedent conditions in the coastal plain watershed.

At the Freeland outlet, the tendency was for all algorithms to underestimate total flow. Overestimation was more common at the Longs outlet. Simulations with all methods (except ParaSol) produced the closest fit at the Longs outlet compared to the Freeland outlet, indicating that uncertainty and errors associated with streamflow simulation have greater influence in the upstream portion of the watershed.

Model performance was unsatisfactory in some intervals of low flow, but in most cases the 95% probability bands were not seriously affected and most simulated flows were positioned in the 95% PPU band. Low flow estimates are mainly related to the deficiency of groundwater module in SWAT, which appears to underestimate Coastal Plain groundwater response. More importantly, shallow water tables may severely restrict the amount of water recharging to the river system. These shallow aquifers along with poor natural drainage created an excessive soil-water condition that is difficult for most hydrology models to understand and capture their interaction. Computing this dynamics is much more difficult in riparian zone and wetlands areas, and reflects the fact that SWAT is less capable in computing saturated areas and soil storage capacity.

In addition, water-holding capacities of soils are low in the Coastal Plain watersheds. These collectively cause soil moisture deficiencies and create more difficulties in capturing subsurface processes and simulating low flow magnitude over time. Previous field based studies (Shirmohammadi et al., 1986) in the Coastal Plain also indicated that during dry periods the

water table in the riparian zone continued to be lowered by transpiration from the riparian forest. As the water table continues to drop, it may reach a level too low for the model to calculate its magnitude especially if rainfall does not occur. Eventually, this may lead to underestimation of low flow dynamics.

Model performance during low flow intervals also may be influenced by the high spatial and temporal variability of precipitation during the wet season and by the sensitivity of the model results to initial conditions at the beginning of the dry season as advocated by Amatya and Jha, (2011). Other potential reasons include not considering the soil condition in the new curve number method and inadequate numerical implementation of Darcy's equation in the model. SWAT keeps track of the amount of water recharging to the shallow aquifer system (delayed by GW_DELAY). Once it reaches the water table, SWAT releases that water to the river system at a rate determined by ALPHA_BF. Therefore, this insufficient knowledge may increase error and uncertainty in low flow simulation. In addition, SWAT showed more sensitivity to soil properties during high hydraulic periods when the interaction between surface and groundwater is high particularly in the forested wetland portion of the watershed.

Modeling Performance Analysis

A comparative analysis was conducted in two aspects: parameter sensitivities and model diagnostic analysis.

Sensitive Parameters

The sensitivity ranking of the parameters varied among the four algorithms (Table 4). Parameters related to channel hydraulics like Manning's *n*-value for the tributary channels and curve number value showed more sensitivity to streamflow generation in all algorithms. Waccamaw streamflow depends mostly on Manning's *n*-value; it is a major contributor to the uncertainty in streamflow prediction at both upstream and downstream outlets. Manning's *n*-value depends on the surface friction resistance, form resistance, wave resistance, and resistance due to flow unsteadiness. There are some artificial channels and channelized streams in the upper basin that may complicate calculating a Manning's *n*-value during flow simulation. Due to these complexities, the exact value of Manning's *n*-value is often uncertain. Further, the shallow

aquifer system that is well connected to the river system makes definition of the Manning's *n*-value especially challenging for any hydrology model.

Posterior density responses of the model parameters were visually investigated and the marginal posterior densities were constructed for sensitive parameters with high to moderate sensitivity (Figures 11 and 12). The x-axis in each graph is fixed to the prior range of each individual parameter to facilitate comparison of sensitive parameters. The y-axis is represented the relative likelihood for sensitive parameter to take on a given value.

Comparing the parameter posterior distribution indicated that parameter values in the SUFI-2 model tended to be uniformly distributed, and optimal parameter values were spread across the entire posterior range. In specific, parameter ranges from the GLUE method were wider than other models. This may be attributed to the fact that GLUE considers parameter correlations, whereas, for instance, SUFI-2 does not. In the SUFI-2 method, all parameter sets from samples are set as behavioral parameters that contribute to the final uncertainties. In the GLUE method, parameter sets are viewed as behavioral parameter sets when their likelihood values are higher than the threshold value; the likelihood values below the threshold value is considered as nonbehavioral parameter sets and removed from further analysis. Only the behavioral parameter sets would contribute to the final uncertainty ranges. This may lead to more reasonable uncertainty ranges by the GLUE method, but in our research, uncertainty bands widened in the GLUE simulation.

The posteriors distributions of some aggregate parameters in the GLUE model have obvious peak areas (e.g., CH_N2, CN2, OV_N, and SLSUBBSN; see Figure 11 and Table 4) that have significant influence on the results of model simulation. The correlations between most parameters in the GLUE model (the correlation matrix of the posterior distribution) are small except the correlations of CH_N2, CN2, OV_N, CH_K2, and SLSUBBSN (results not shown here). This indicates that the correlations between those sensitive parameters along with hydraulic conductivity in tributary channel (CH_K2) cannot be ignored in the GLUE simulation.

Saturated hydraulic conductivity (SOL_K), Manning's *n*-value for the main channel, and runoff curve number were three of the most sensitive parameters in the SUFI-2 analysis (see Table 4 and Figure 12). All are characteristics of upland and wetland sections of the watershed, so they can have several values depending on land use and soil type. Average slope length (SLSUBBSN) was one of the most sensitive parameters in all models except SUFI-2. This is the distance that sheet flow forms and dominates the surface runoff process then concentrates at the origin of the rill. It appears that those methods showed more sensitivity to land use, topography and soil types because SLSUBBBSN largely depends on these physical characteristics.

Manning's *n*-value for overland flow was one of the most sensitive parameters in PSO, GLUE and ParaSol models. This parameter controls water transmission in different land use categories (wetland, forest, pasture, etc.). Groundwater properties (GW_REVAP) and the soil evaporation compensation factor (ESCO) also showed sensitivity in ParaSol. This may be attributed to inadequate evapotranspiration and shallow aquifer parameterization in the SWAT model and the complexities in soil and groundwater system as discussed above. Table 4 presents the parameters chosen to calibrate streamflow and their ranks, and best parameter values estimated by SUFI-2.

Based on SUFI-2 result, calibrated values of the CN ranged from 50 (Norfolk soil in evergreen forest) to 89 (Croatan soil in agricultural row crops; Table 5). As anticipated, higher CN values are associated with hydrologic group D and lower with group A (Table 2). For each soil series, the highest CN is associated with agricultural row crops, the cover class that is most likely to have the lowest infiltration rate of the four land cover classes. These results are aligned with expectations based on their landscape position and characteristics (see Tables 1 and 2 for soil and land use information).

Calibrated values of the hydraulic conductivity range from moderately high (Meggett) to very high (Croatan and Norfolk; Table 5). For all soils the surface layer has higher conductivity than intermediate layers; Croatan and Norfolk have a lower layer with very high conductivity. The calibrated hydraulic conductivities are indicative of the dominance of sand and relatively small amount of clay in these soils. Loamy textures also are prevalent.

Based on these simulation results, it appears that additional refinement of our parameterization is required to represent surface flow generated from interfluvial uplands where most precipitation infiltrates prior to reaching the stream. Model performance also would benefit from additional work on the physical processes that occur in surficial aquifers within the alluvial floodplain storage along streams.

Modeling Diagnostic Analysis

The violin plots of the simulated streamflow are illustrated for all methods (Figures 13 and 14). These plots were created using both calibration and validation outputs generated by the models. The shape of the density mass function (or probability density function (PDF)) for SUFI-2 and PSO were very close to observations while ParaSol and GLUE, respectively, underestimated and overestimated the observed PDF. In addition, SUFI-2, PSO and GLUE exhibited approximately similar median values compared to observation. More dispersion and skewness were noted in the ParaSol estimated streamflow. The probability of streamflow values (longer density) is highest in the GLUE and the PSO models, correspondingly in upstream and downstream portions. The 25th and 75th percentile (thick black lines in Figures 13 and 14) range of PDF distribution is effectively narrowed by all models. SUFI-2 and PSO showed the same relative tails to observed values at the upstream outlet while all methods (except ParaSol) exhibited longer tails at downstream gage. Further, SUFI-2, PSO and ParaSol underestimated the 5th and 95th percentile ranges of streamflow values (thin black line in Figures 13 and 14) while GLUE provided better estimates at Freeland. SUFI-2 provided a close estimate of the 5th and 95th percentile ranges at Longs.

Close inspections of the modeling residuals reveal that the errors are autocorrelated at the first lag in all methods (Figure 15). Autocorrelation is higher at Longs outlet than Freeland outlet. The residuals show that the model captures the patterns in the data during high lag at Freeland outlet and there is a large amount of autocorrelation left in the Longs residual. This suggests that the model can be slightly improved especially for Longs outlet. Note that the autocorrelation results are quite similar for SUFI-2, PSO and GLUE at both outlets. Since most of these errors are correlated, it seems that residuals have heteroscedasticity and nonnormality, especially at the downstream outlet. This indicates that error variance changes with the magnitude of streamflow

(heteroscedasticity) and skewness increases in modeling over time (nonnormality or inhomogeneity) as explained by Sorooshian and Dracup, (1980); Duan et al., (1992); and Schoups and Vrugt, (2010). This reflects the fact that residual errors increases as linear and/or nonlinear function of mean/median streamflow, which may cause fluctuation in the simulation.

Summary and Discussion

This paper compares different uncertainty algorithms used with the SWAT semi-distributed hydrology model. SWAT was used to simulate daily streamflow during wet, moderate, and dry hydrological conditions at two streamflow gages in the southeastern US. Model analyses include assessment of model performance and diagnosis of model output during simulation period.

Parameter uncertainty analysis revealed that all algorithms performed best during intervals with wet and normal antecedent conditions with varying sensitivity to channel shape and characteristics, soil and groundwater properties. Further, the calibration of all algorithms depended mostly on Manning's *n*-value for the tributary channels as the surface friction resistance factor to generate runoff. This along with curve number parameter showed more sensitivity to daily streamflow simulation.

SUFI-2 and PSO showed the same relative distribution tails compared to observation at the upstream outlet while all methods (except ParaSol) exhibited longer tails at the downstream outlet. In addition, large skewness and dispersion was defined by the ParaSol model (r-factor<0.24 and p-factor<25%), suggesting that a global search algorithm is less capable of computing parameter uncertainty of a low gradient Coastal Plain watershed.

The p-factor and r-factor computed using SUFI-2 bracketed more than 90% and 75% of the observed streamflow during the calibration and validation periods at upstream and downstream USGS gages in the watershed. Furthermore, the model predicted most peak flows with the least uncertainty and error. The model consistently overestimated low flows but performed relatively well during high flows, especially during periods with wet antecedent conditions. Specifically, SUFI-2 showed that the dry years had slightly larger prediction uncertainties than the wet years. This was because groundwater dominated the surface flow during dry periods, thus SWAT

inadequately simulated below ground processes. Good model performance during wet periods was related, in part, to higher daily discharge rates, and, thus, wider uncertainty bounds (e.g., Yang et al., 2008; Coxon et al., 2014).

The differences in model performance for low flow events in upstream and downstream portions are related at least in part to the impact of wetland and saturated areas in modeling performance. The downstream gage that had lower performance (with larger portions of wetlands), used surface runoff parameterizations in which saturated areas are controlled by water storage in the shallow aquifer. Variability in saturated areas in this outlet has large control on model parameterization. This approach was not accounted for by SWAT due to its deficiency in generating saturated areas and saturation excess runoff (e.g., Lyon et al., 2004; Easton et al., 2008; Schneiderman et al., 2007). However, Baffaut et al., (2015b) proposed adjustment of the percolation routine to better simulate saturation close to the surface, this approach was not explored in this research. These are reasonable interpretation of the model results; however, there are no measured data to assess if defined saturated areas by the model are realistic for the Waccamaw watershed. In this context, it is worth noting that no model has a mechanism for generating infiltration-excess runoff, and the saturated area parameterization may compensate for model weakness (e.g. Clark et al., 2008) particularly for the coastal plain with a typically wet surface.

Diagnosis of model errors shows that the groundwater response to low flow events is quite consistent in that most algorithms either overestimate or underestimate low flow events. However, the consistency in model errors may also arise because errors in input data may affect different modules and processes of a watershed model in similar ways. Moreover, different uncertainty models may not bracket the measurement errors because the SWAT model, like other similar hydrology models, has similar weaknesses such as no mechanism for generating saturated areas and saturation excess runoff, and no vegetation submodel (the model domain extending from the top of the vegetation to the base of active groundwater i.e., the Earth's Critical Zone' see Anderson et al., 2008). In addition, errors in discharge data resulting from uncertainty in the stage-discharge rating curve might affect the simulation results especially

during repeated rainfall events when the measured values were extrapolated outside the range of the established rating curve.

There has been a continued debate within the hydrological community regarding to the choice of the most suitable approach for uncertainty assessment to capture different sources of error and uncertainty (Schoups and Vrugt, 2010; Laloy and Vrugt, 2012; Beven et al., 2012; Pourreza-Bilondi and Samadi, 2016; Pourreza-Bilondi et al., 2016). As with all methods of uncertainty analysis, our methods have their own limitations and other approaches could have been adopted, such as the use of multi-objective algorithm (Vrugt and Robinson, 2007; Sadeghi-Tabas et al., 2016 a, b), and MCMC algorithm based on GL (generalized log-likelihood) function (Schoups and Vrugt, 2010; Laloy and Vrugt, 2012). However, the need for robust simulation given the complexities and difficulties in the coastal plain modeling meant that this study provided the basis for future model development and hypothesis testing. Here, the resulting simulation for each algorithm and the sensitivity and diagnostic analyses act as a guide to the dominant processes operating in the shallow aquifer dominated environmental system and additional analysis including field based research may be needed to conclude this definitively, alongside this type of analysis over many more coastal plain watersheds.

Further work is also needed to evaluate model performance with respect to multiple criteria, including assessment of model performance based on flow duration curves (FDC), assessment of model performance in floodplain portion and water table processes, and assessment of model performance with respect to ‘‘diagnostic signatures’’ that are extracted from the data to explain different hydrological processes in the watershed (Gupta et al., 2008). More incisive diagnostic testing that does not depend solely on discharge data (e.g., McMillan et al., 2012) is needed, as is in-depth analysis of different parts of the hydrographs (low flow, peak flow, and recession curve). Future research will extend this methodology by using GL-based MCMC Bayesian method as well as data assimilation (DA) algorithm to many more coastal plain watersheds across the Southeast and advance our ability to simulate watershed response, benchmark watershed processes, and make skillful prediction in ungagged catchments.

ACKNOWLEDGEMENTS

We express our gratitude to NOAA–National Oceanic and Atmospheric Administration for financial supporting of this research (Grant # NA060AR4310007) to Carolinas Integrated Sciences & Assessments (CISA). We would also thank to all of the institutions (mentioned throughout the paper) for giving free access to their software and data. We thank three anonymous reviewers for their thoughtful and constructive comments on the manuscript.

LITERATURE CITED

Abbaspour, K.C., C.A. Johnson, and M.T. van Genuchten, 2004. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal* 3(4):1340–1352. DOI: 10.1016/j.proenv.2015.07.047

Abbaspour, K.C., 2013. SWAT-CUP 2012: SWAT Calibration and Uncertainty Programs – A User Manual. Swiss Federal Institute of Aquatic Science and Technology, Eawag, Duebendorf, Switzerland.

Abbaspour, K.C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan, 2007. Modelling hydrology and water quality in the pre-Alpine/Alpine Thur watershed using SWAT. *Journal of Hydrology* 333:413–430. DOI: 10.1016/j.jhydrol.2006.09.014

Amatya, K.M. and M.K. Jha, 2011. Evaluating the SWAT model for a low-gradient forested watershed in Coastal South Carolina. *Transactions of the American Society of Agricultural and Biological Engineers* 54(6):2151-2163. DOI: 10.13031/2013.40671

Anderson, S.P., R.C. Bales, and C.J. Duffy, 2008. Critical zone observatories: Building a network to advance interdisciplinary study of Earth surface processes. *Mineral. Magazine*, 72(1):7–10. DOI: 10.1180/minmag.2008.072.1.7

Arnold, J.G., R. Srinivasan, R.S. Muttiah, J.R. Williams, 1998. Large area hydrologic modeling and assessment. Part I: model development. *Journal of the American Water Resources Association* 34(1):1-89. DOI: 10.1111/j.1752-1688.1998.tb05961.x

Baffaut, C., S.M. Dabney, M.D. Smolen, M.A. Youssef, J.V. Bonta, M.L. Chu, J.A. Guzman, V. Shedekar, M.K. Jha, and J.G. Arnold, 2015a. Hydrologic and water quality modeling: spatial and temporal considerations. *Transactions of the American Society of Agricultural and Biological Engineers* 58(6):1661–1680. DOI: 10.13031/trans.58.10714

Baffaut, C., E.J. Sadler, F. Ghidey, and S.H. Anderson, 2015b. Long-Term Agroecosystem Research in the Central Mississippi River Basin: SWAT Simulation of Flow and Water Quality in the Goodwater

Creek Experimental Watershed. *Journal of Environmental Quality* 44: 84-96. DOI: 10.2134/jeq2014.02.0068

Beven, K.J. and A.M. Binley, 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6(3):279–98. DOI: 10.1002/hyp.3360060305

Beven, K. and J. Freer, 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249:11-29. DOI: 10.1016/S0022-1694(01)00421-8

Beven, K., P. Smith, I. Westerberg, and J. Freer, 2012. Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al. *Water Resources Research* 48: W11801. DOI: 10.1029/2012WR012282

Blasone, R.S., J.A. Vrugt, H. Madsen, D. Rosbjerg, B.A. Robinson, and G.A. Zyvoloski, 2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources* 31(4):630–648. DOI: 10.1016/j.advwatres.2007.12.003

Coxon, G., J. Freer, T. Wagener, N.A. Odoni, and M. Clark, 2014. Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. *Hydrological Processes* 28(25):6135–6150. DOI: 10.1002/hyp.10096

Clark, M.P., A.G. Slater, D.E. Rupp, R.A. Woods, J.A. Vrugt, H.V. Gupta, T. Wagener, and L.E. Hay, 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research* 44, W00B02, DOI: 10.1029/2007WR006735

Duan, Q., S. Sorooshian, and V. Gupta, 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resources Research* 28(4):1015–1031. DOI: 10.1029/91WR02985

Eberhart, R.C. and J. Kennedy, 1995. A new optimizer using particle swarm theory. Proceedings of the sixth international symposium on micro machine and human science pp. 39-43. IEEE service center, Piscataway, NJ, Nagoya, Japan. DOI: 10.1109/MHS.1995.494215

Easton, Z.M., D.R. Fuka, M.T. Walter, D.M. Cowan, E.M. Schneiderman, and T.S. Steenhuis, 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *Journal of Hydrology* 348:279-291. DOI: 10.1016/j.jhydrol.2007.10.008

Feyereisen, G.W., R. Lowrance, T.C. Strickland, J.M. Sheridan, R.K. Hubbard, and D.D. Bosch, 2007. Long-Term Water Chemistry Database, Little River Experimental Watershed, Southeast Coastal Plain, United States. *Water Resources Research* 43(9):W09474. DOI: 10.1029/2006WR005835

Gupta, H.V., T. Wagener, and Y. Liu, 2008. Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes* 22:3802–3813. DOI: 10.1002/hyp.6989

Hargreaves, G.H. and Z.A. Samani, 1982. Estimation of potential evapotranspiration. *Journal of Irrigation and Drainage Division* 108(3):223-230

Harmel, R.D. and P.K. Smith, 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology* 337(3-4):326-336. DOI: 10.1016/j.jhydrol.2007.01.043

Herschy, R.W., 2008. Streamflow Measurement, Third Edition. CRC Press, Boca Raton, FL. ISBN: 9780415413428

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI. ISBN: 9780262082136

Hornberger, G.M. and R.C. Spear, 1981. An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management* 12:7–18.

Joseph, J.F. and J.H.A. Guillaume, 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environmental Modelling & Software* 46:292–298. DOI: 10.1016/j.envsoft.2013.03.012

Kannan, N., C. Santhi, J.R. Williams, and J.G. Arnold, 2008. Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods. *Hydrological Processes* 22(13):2114- 2121.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42: W03S04. DOI: 10.1029/ 2005WR004362.

Krause, P., D.P. Boyle, and F. Bäse, 2005. Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences* 5, 89-97, DOI: 10.5194/adgeo-5-89-2005.

Lam, Q.D., B. Schmalz, and N. Fohrer, 2010. Modeling point and diffuse source pollution of nitrate in a rural lowland catchment using the SWAT model. *Agricultural Water Management* 97(2):317-325.

Liu, Y. and H.V. Gupta, 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research* 43, W07401, DOI: 10.1029/2006WR005756.

Lyon, S.W., M.T. Walter, P. Gérard-Marchant, and T.S. Steenhuis, 2004. Using a topographic index to distribute variable source area runoff predicted with the SCS curve-number equation. *Hydrological Processes* 18: 2757-2771. DOI: 10.1002/hyp.1494.

Magnus, J.R. and H. Neudecker, 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. 2nd Ed., John Wiley & Sons, New York, ISBN-10: 047198633X.

Mantovan, P. and E. Todini, 2006. Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. *Journal of Hydrology* 330:368–381. DOI: 10.1016/j.jhydrol.2006.04.046

McMillan, H., D. Tetzlaff, M. Clark, and C. Soulsby, 2012. Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. *Water Resources Research* 48: W05501. DOI: 10.1029/2011WR011688.

Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resources Research* 41, 10 W08406, DOI: 10.1029/2004WR003826.

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Binger, R.D. Harmel, and T. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the American Society of Agricultural and Biological Engineers* 50(3):885-900. DOI: 10.13031/2013.23153

Moriasi, D.N., G.W. Gitau, N. Pai, and P. Daggupati, 2015. Hydrologic and water quality models: Performance measures and criteria. *Transactions of the ASABE* 58:1763–1785. DOI: 10.13031/trans.58.10715

Moussa, R., 2008. Effect of channel network topology, basin segmentation and rainfall spatial distribution on the GIUH transfer function. *Hydrological Processes* 22:395-419. DOI: 10.1002/hyp.6612.

Nash, J.E. and J.V. Sutcliffe, 1970. River flow forecasting through conceptual models part I -A discussion of principles. *Journal of Hydrology* 10(3):282-290. DOI: 10.1016/0022-1694(70)90255-6.

Neitsch, S.L., J.G. Arnold, J.R. Kiniry, J.R. Williams, and K.W. King, 2004. Soil and Water Assessment Tool (Version 2000)—theoretical documentation. GSWRL 02-01, BRC 02-05, TR-191, Texas Water Resources Institute, College Station, TX.

Nelder, J.A. and R.A. Mead, 1965. A simplex method for function minimization. *Computer Journal* 7:308-313.

Pourreza-Bilondi, M., S.Z. Samadi, A.M. Akhoond-Ali, and B. Ghahraman, 2017. Reliability of Semiarid Flash Flood Modeling Using Bayesian Framework. *Journal of Hydrologic Engineering* 22(4): DOI: 10.1061/(ASCE)HE.1943-5584.0001482.

Pourreza-Bilondi, M. and S.Z. Samadi, 2016. Quantifying the Uncertainty of Semiarid Runoff Extremes Using Generalized Likelihood Uncertainty Estimation. *Arabian Journal of Geosciences* 9:622. DOI: 10.1007/s12517-016-2650-0.

Priestly, C.H.B. and R.J. Taylor, 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* 100(2):81-92.

Riggs, S.R., V.D. Ames, D.R. Brant, and E.D. Sager, 2000. The Waccamaw Drainage System: Geology and Dynamics of a Coastal Wetland, Southeastern North Carolina, North Carolina Division of Water Resources Publication, Raleigh, NC.

Sadeghi-Tabas, S., A. Akbarpour, M. Pourreza-Bilondi, and S.Z. Samadi, 2016a. Toward Reliable Calibration of Aquifer Hydrodynamic Parameters: Characterizing and Optimization of Arid Groundwater System Using Swarm Intelligence Optimization Algorithm. *Arabian Journal of Geosciences* 9:719. DOI: 10.1007/s12517-016-2751-9.

Sadeghi-Tabas, S., S.Z. Samadi, A. Akbarpour, and M. Pourreza-Bilondi, 2016b. Sustainable Groundwater Modeling Using Single and Multi-Objective Optimization Algorithms. *Journal of Hydroinformatics* 19(1):97-114. DOI: 10.2166/hydro.2016.006.

Samadi, S.Z., 2016. Assessing the Sensitivity of SWAT Physical Parameters to Potential Evapotranspiration Estimation Methods over a Coastal Plain Watershed in the Southeast United States. *Hydrology Research* 48(2):395-415. DOI: 10.2166/nh.2016.034

Samadi, S.Z. and E.M. Meadows, 2017. The Transferability of Terrestrial Water Balance Components under Uncertainty and Non-stationarity: A Case Study of the Coastal Plain Watershed in the Southeastern United States. *River Research and Applications* 33(5):796-808. DOI: 10.1002/rra.3127.

Samadi, S., D. Tufford, and G.J. Carbone, 2014. Improving Hydrologic Predictions of Distributed Watershed Model via Uncertainty Quantification of Evapotranspiration Methods. South Carolina Water Resource Conference. 15-16 Oct. 2014, Columbia, SC, USA.

Schoups, G. and J.A. Vrugt, 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non‐Gaussian errors. *Water Resources Research* 46, W10531, DOI: 10.1029/2009WR008933.

Schneiderman, E.M., T.S. Steenhuis, D.J. Thongs, Z.M. Easton, M.S. Zion, A.L. Neal, G.F. Mendoza, and M.T. Walter, 2007. Incorporating variable source area hydrology into a curve-number-based watershed model. *Hydrological Processes* 21(25):3420-3430. DOI: 10.1002/hyp.6556.

Schroeter, H.O. and R.P. Epp, 1988. Muskingum-Cunge: A Practical Alternative to the HYMO VSC Method for Channel Routing**.** *Canadian Water Resources Journal* 13(4):68-79. DOI: 10.4296/cwrj1304068

Sellami, H., I. La Jeunesse, S. Benabdallah, and M. Vanclooster, 2013. Parameter and rating curve uncertainty propagation analysis of the SWAT model for two small Mediterranean watersheds. *Hydrological Sciences Journal* 58(8):1635−1657. DOI: 10.1080/02626667.2013.837222.

Setegn, S.G., R. Srinivasan, A.M. Melesse, and B. Dargahi, 2010. SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia. *Hydrological Processes* 24: 357–367. DOI: 10.1002/hyp.7457.

Shirmohammadi, A, J.M. Sheridan, and L.E. Asmussen, 1986. Hydrology of alluvial stream channels in southern Coastal Plain watersheds. *Transactions American Society of Agricultural Engineers* 29(1):135-142.

Sorooshian, S. and J.A. Dracup, 1980. Stochastic parameter estimation procedures for hydrologic rainfall-runoff models—correlated and heteroscedastic error cases. *Water Resources Research* 16(2): 430–442. DOI: 10.1029/ WR016i002p00430.

Tian, W., X. Li, X-S. Wang, and B.X. Hu, 2012. Coupling a groundwater model with a land surface model to improve water and energy cycle simulation. *Hydrology and Earth System Sciences* 16:4707–4723. DOI: 10.5194/hess-16-4707-2012.

Vrugt, J.A., H.V. Gupta, W. Bouten, and S. Sorooshian, 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39(8), 1201. DOI: 10.1029/2002WR001642

Vrugt, J.A. and Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of America* 104(3):708–711. DOI: 10.1073/pnas.0610471104.

Willmott, C.J., S.M. Robeson, K. Matsuura, and D.L. Ficklin, 2015. Assessment of three dimensionless measures of model performance. *Environmental Modelling and Software* 73:167–174. DOI: 10.1016/j.envsoft.2015.08.012

Winchell, M., R. Srinivasan, M. Di Luzio, and J. Arnold, 2007. ArcSwat interface for SWAT2005: User's guide. (USDA: Temple, Texas, U.S.A.).

Wu, K. and Y.J. Xu, 2006. Evaluation of the applicability of the SWAT model for Coastal watersheds in southeastern Louisiana. *Journal of the American Water Resources Association* 42(5):1247- 1260. DOI: 10.1111/j.1752-1688.2006.tb05298.x

Yang, J., K.C. Abbaspour, P. Reichert, and H. Yang, 2008. Comparing uncertainty analysis techniques for a SWAT application to Chaohe Basin in China. *Journal of Hydrology* 358, 1-23. DOI: 10.1016/j.jhydrol.2008.05.012.

Table 1. Distribution of land use defined with the SWAT model.

Land use type	Land use description	Area	
		ha	%
WATR	Open Water	4,890.5	1.6
URLD	Rural Residential	11,402.8	3.7
URMD	Medium Density Residential	2,902.8	0.9
URHD	High Density Residential	447.8	0.1
SWRN	South Western Range	421.4	0.1
FRSD	Forest and Woodland	248.8	0.1
FRSE	Evergreen Forest	84,253.9	27.0
FRST	Mixed Forest	4,362.3	1.4
RNGB	Range Shrubland	33,729.8	10.8
RNGE	Grasslands/Herbaceous	15,588.4	5.0
HAY	Pasture	249.6	0.1
AGRR	Agricultural Land-Row Crops	63,914.7	20.5
WETF	Forested Wetland	86,173.7	27.7
WETN	Non-forested Wetlands	2,990.9	1.0
UIDU	Industrial	107.4	<0.1
Watershed	Watershed area delineated by SWAT	311,684.9	100%

Table 2 Soil types in the Waccamaw River watershed.

Series Name	Hydrologic Group	Drainage Class	Area	
			ha	%
Meggett	D	poorly drained	55,985.4	18.0
Bohicket	D	very poorly	0.4	0.0
Kenansville	A	well drained	4,942.4	1.6
Leaf	D	poorly drained	14,495.1	4.7
Woodington	B/D	poorly drained	33,947.8	11.1
Croatan	D	very poorly	47,311.4	15.2
Rains	B/D	poorly drained	61,428.2	19.7
Norfolk	B	well drained	44,762.2	14.4
Autryville	A	well drained	8,903.4	2.9
Leon	B/D	poorly drained	4,246.0	1.5
Kureb	A	excessively well	1,561.4	0.5
Trebloc	D	poorly drained	16,343.5	5.2
Nansemond	A	moderately well	2,352.5	0.7
Montevallo	D	well drained	3,507.3	1.1
Osier	D	poorly drained	26.8	0.0
Chisolm	A	well drained	11,871.3	3.8

Table 3. Streamflow calibration and validation results for the Waccamaw watershed using SUFI-2, PSO, ParaSol and GLUE.

Objective Function		Freeland		Longs	
		Calibration	Validation	Calibration	Validation
NSE	SUFI-2	0.79	0.87	0.77	0.90
	PSO	0.79	0.85	0.76	0.91
	GLUE	0.66	0.87	0.81	0.87
	ParaSol	0.69	0.85	0.78	0.82
p-factor	SUFI-2	90%	61%	75%	52%
	PSO	97%	85%	92%	74%
	GLUE	100%	80%	97%	72%
	ParaSol	15%	18%	25%	21%
r-factor	SUFI-2	0.87	0.69	0.79	0.72
	PSO	1.47	1.39	1.23	1.40
	GLUE	1.85	1.20	1.53	1.27
	ParaSol	0.24	0.18	0.23	0.21
MSE	SUFI-2	57.89	94.74	325.70	140.02
	PSO	59.11	106.25	343.13	133.72

	GLUE	99.69	95.04	269.07	306.17
	ParaSol	85.59	102.75	309.76	261.32

Table 4. Selected parameters for uncertainty analysis and their prior distributions in SWATCUP and best parameter values estimated by SUFI-2 for the study watershed.

Aggregate parameter	Name of SWAT parameter	Sensitive Parameter Rank				SUFI-2 best value
		SUFI-2	GLUE	ParaSol	PSO	
v__CH_N2.rte	Manning's *n*-value for the main channel	1	1	1	1	0.1
r__CN2.mgt	SCS runoff curve number for moisture condition II	2	2	2	3	-0.11
r__SOL_K().sol	Saturated hydraulic conductivity (mm/hr)	3	15	15	8	-0.27
v__ESCO.hru	Soil evaporation compensation factor	4	16	3	17	0.88
r__SOL_AWC().sol	Available water capacity of the soil layer (mm H2O/mm soil)	5	13	17	11	0.15
r__CH_K2.rte	Effective hydraulic conductivity in tributary channel alluvium (mm/hr)	6	5	7	14	0.01
r__ALPHA_BF.gw	Base flow alpha factor (days)	7	8	9	5	0.36
r__GW_REVAP.gw	Groundwater "revap" coefficient	8	17	6	6	-0.1
v__OV_N.hru	Manning's *n*-value for overland flow	9	3	4	2	0.34
r__GWHT.gw	Initial groundwater height (m)	10	14	8	16	1.5
r__SOL_BD().sol	Moist bulk density (Mg/m3 or g/cm3)	11	7	13	12	0.13
v__GW_DELAY.gw	Groundwater delay time (days)	12	12	18	13	300.93
r__RCHRG_DP.gw	Deep aquifer percolation fraction	13	10	10	10	0.77
r__GW_SPYLD.gw	Specific yield of the shallow aquifer (m3/m3)	14	6	14	7	0.43
v__LAT_TTIME.hru	Lateral flow travel time (days)	15	9	12	18	156.6
r__SHALLST.gw	Initial depth of water in the	16	11	15	9	-0.18

	shallow aquifer (mm H2O)					
v__EPCO.hru	Plant uptake compensation factor	17	18	11	15	0.41
r__SLSUBBSN.hru	Average slope length (m)	18	4	5	4	0.54

Aggregate parameters are constructed based on (SWAT-CUP user manual, Abbaspour, 2013). ‘‘v__’’ and ‘‘r__’’ means a replacement, and a relative change to the initial parameter values, respectively.

Table 5. SUFI-2 calibrated curve numbers and soil hydraulic conductivity.

Land Cover (%) / Soil Series (%)	Runoff Curve Number				Saturated hydraulic conductivity			
	Forest Evergreen (27%)	Range/Shrub (11%)	Agri-row crop (21%)	Forested wetland (28%)	Soil layer			
					Layer 1	Layer 2	Layer 3	Layer 4
Meggett (18%)	71	72	82	77	130	7.3	3.8	63
Woodington (11%)	50	56	72	62	97	63	69	*
Croatan (15%)	77	80	89	83	250	110	20	450
Rains (20%)	51	56	72	61	110	18	15	14
Norfolk (14%)	50	55	72	60	450	13	6.2	500

*SWAT computed three layers only for the Woodington soil series.

LIST OF FIGURES

Figure 1 The 311,685 ha Waccamaw River Watershed used in SWAT. Model calibration, validation, and parameter sensitivity and uncertainty analyses were performed using data from USGS gaging stations at Freeland and Longs. Research data were retrieved on daily basis from the National Climatic Data Center (NCDC) and USGS portals on 06 and 13 September 2012, respectively. Selected climate stations (Loris, Whiteville 7 and Longwood) are distributed evenly through the watershed.

Figure 2. Rainfall-runoff processes in the SWAT model and its linkage to uncertainty optimization algorithm. Note that SWAT calculates baseflow by analyzing measured streamflow during periods of no recharge in the watershed (see Winchell et al., 2007). Baseflow varies as the exponential function of the number of low flow days in the SWAT model.

Figure 3. Daily calibration and validation streamflow in SUFI-2 at the Freeland station. The solid black line corresponds to the observed surface flow at the basin outlet, whereas the solid red line represents the best simulation obtained by SUFI-2. The grey and blue

shaded areas are the 95PPU of streamflow calibration and rainfall values – each large peak flow corresponds to high rainfall event (>50 mm) while small peak flows correspond to lower magnitudes rainfall events (< 20 mm).

Figure 4. Daily calibration and validation streamflow in SUFI-2 at the Longs station.

Figure 5. Daily calibration and validation streamflow in GLUE at the Freeland station.

Figure 6. Daily calibration streamflow in GLUE at Longs station.

Figure 7. Daily calibration and validation streamflow in the ParaSol model at Freeland station.

Figure 8. Daily calibration and validation streamflow in the ParaSol model at Longs station.

Figure 9. Daily calibration streamflow in the PSO model at Freeland station.

Figure 10. Daily calibration and validation streamflow in the PSO model at Longs station.

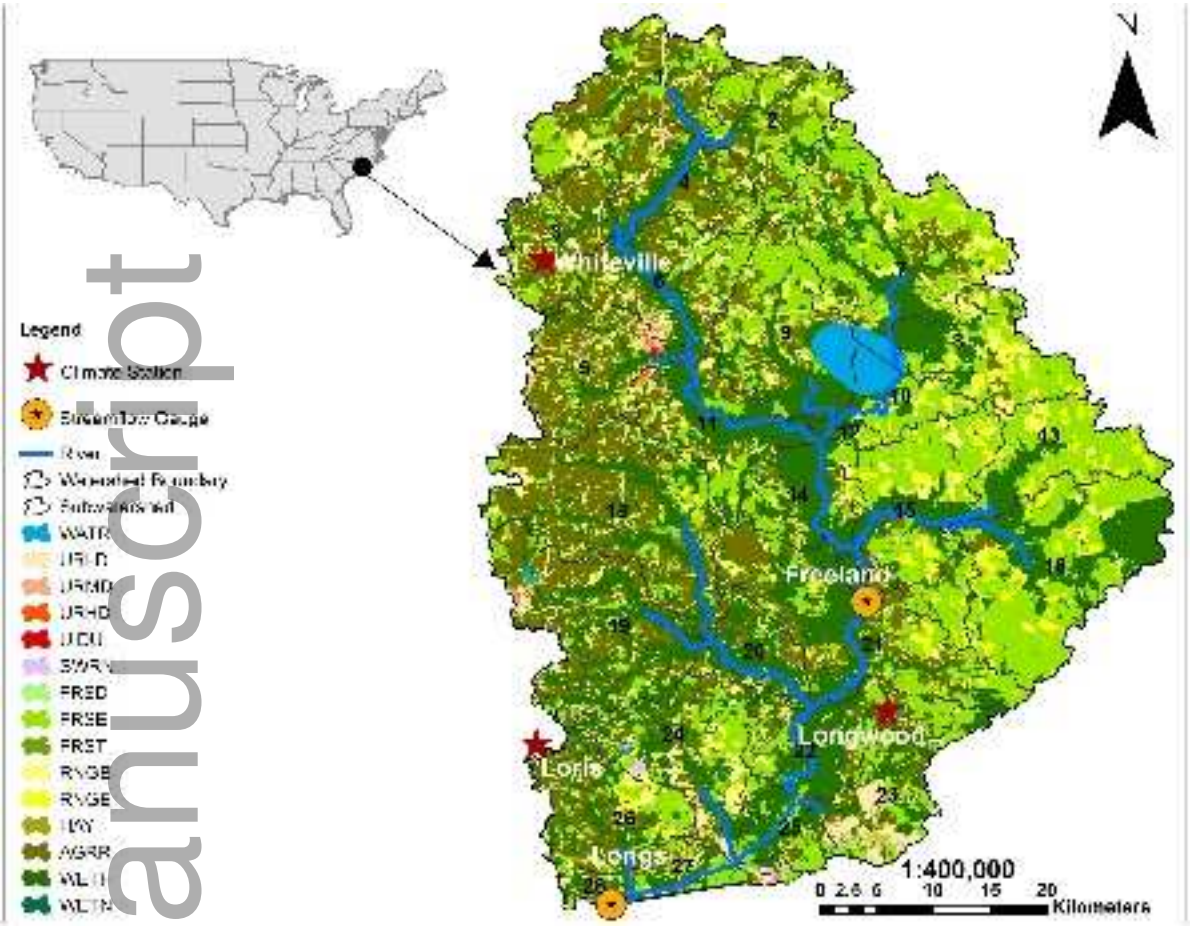
Figure 11. Marginal posterior probability distributions of sensitive parameters in the SWAT model inferred using PSO (the first row) and GLUE (the second row).

Figure 12. Marginal posterior probability distributions of sensitive parameters in the SWAT model inferred using SUFI-2 (the first row) and ParaSol (the second and the third rows).

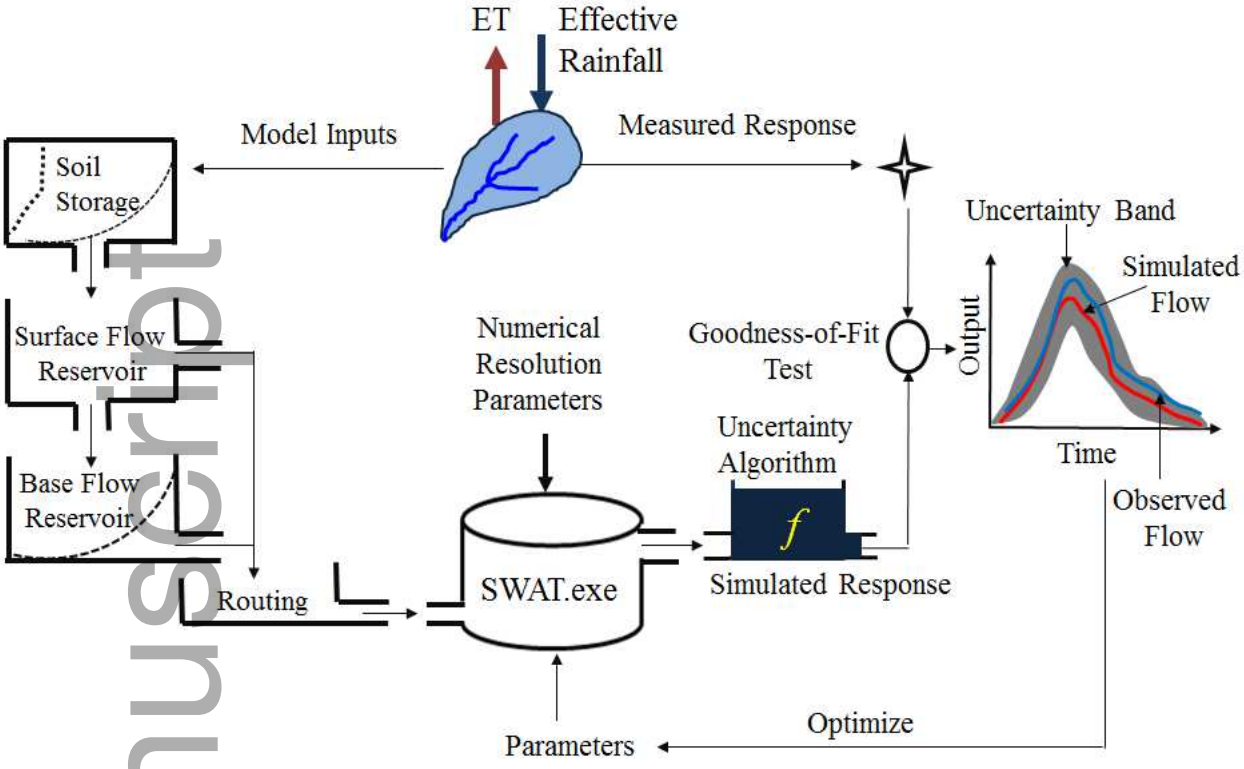
Figure 13. Violin plots of observed and simulated daily streamflow using all algorithms for Freeland outlet. Thick black line and white dot show the 25th and 75th percentile range and median, respectively, and thin black line shows the 5th and 95th percentile ranges of streamflow values.

Figure 14. Violin plots of observed and simulated daily streamflow using all algorithms for Longs outlet. Thick black line and white dot show the 25th and 75th percentile range and median, respectively, and thin black line shows the 5th and 95th percentile ranges of streamflow values.

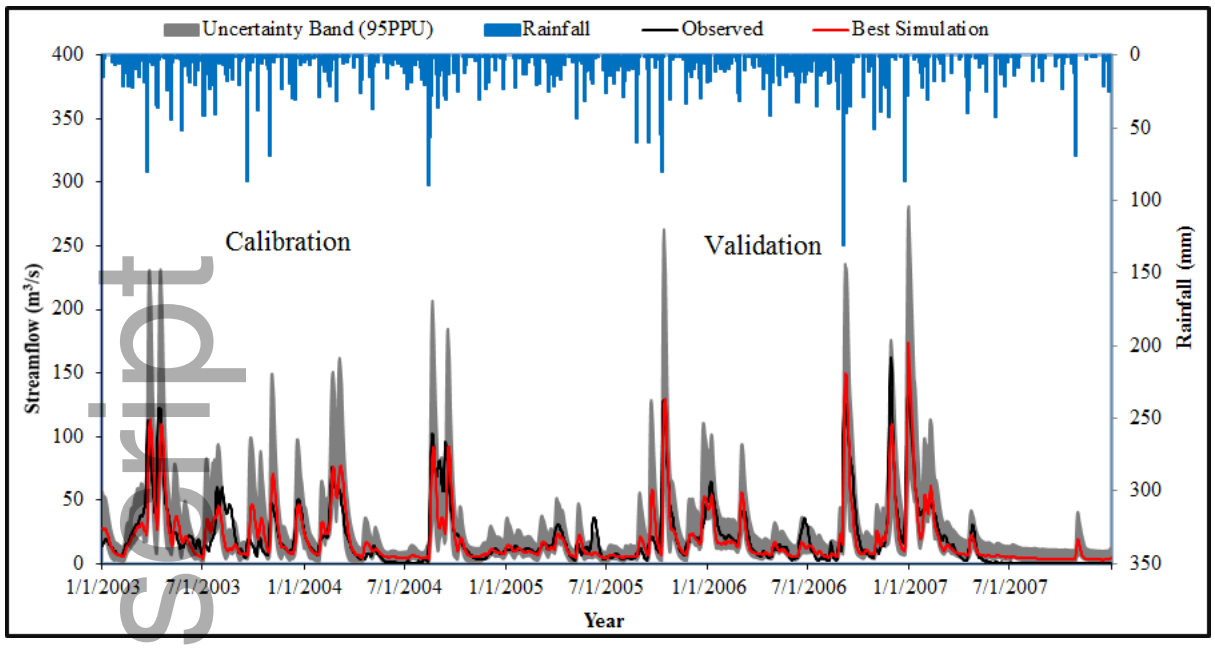
Figure15. Autocorrelation plots of residual errors using SUFI-2, PSO, GLUE and Parasol models for Freeland (left column) and Longs (right column)outlets.



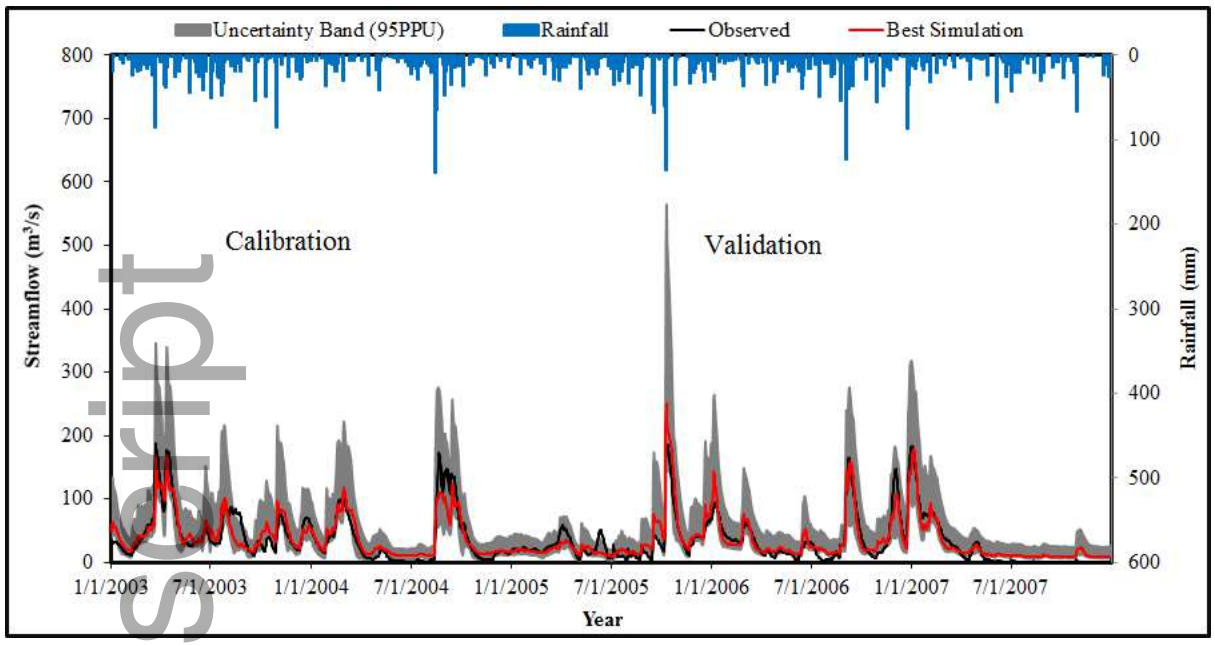
jawra_12596-16-0236_f1.tif



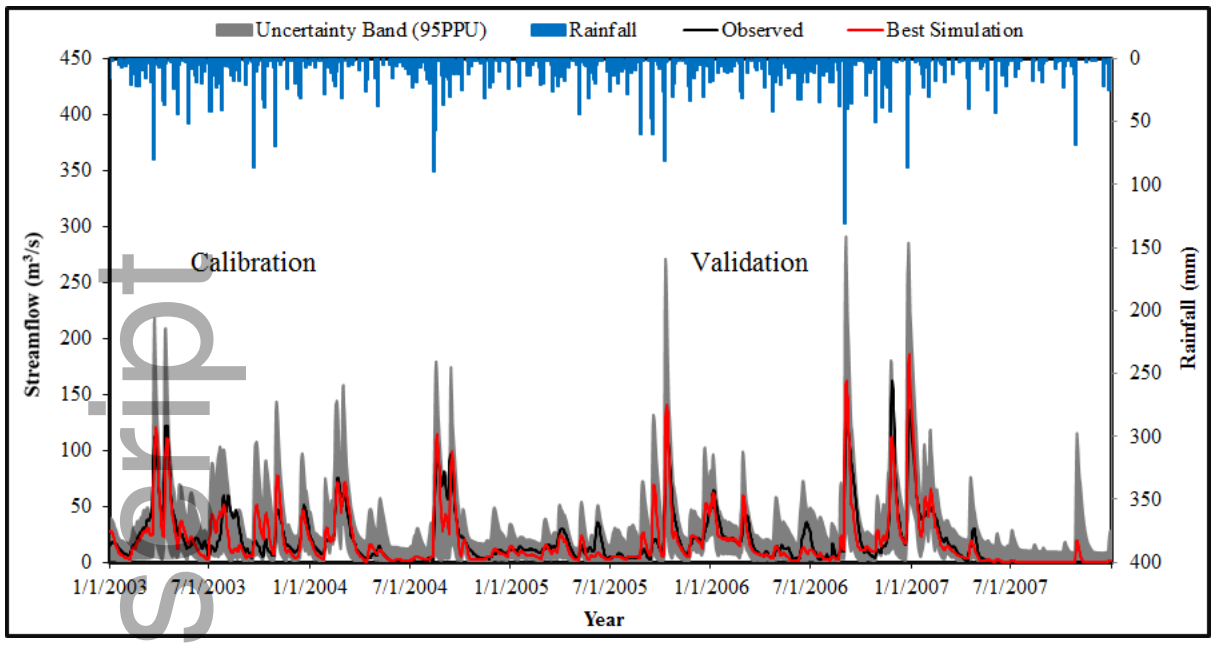
jawra_12596-16-0236_f2.tif



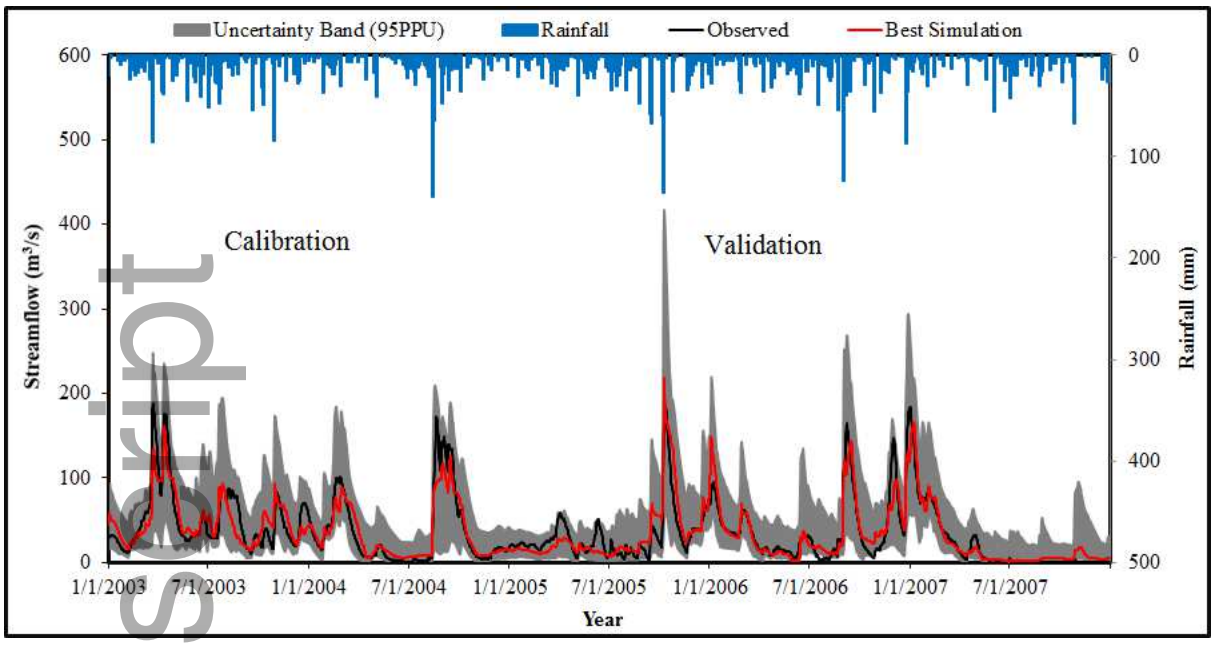
jawra_12596-16-0236_f3.tif



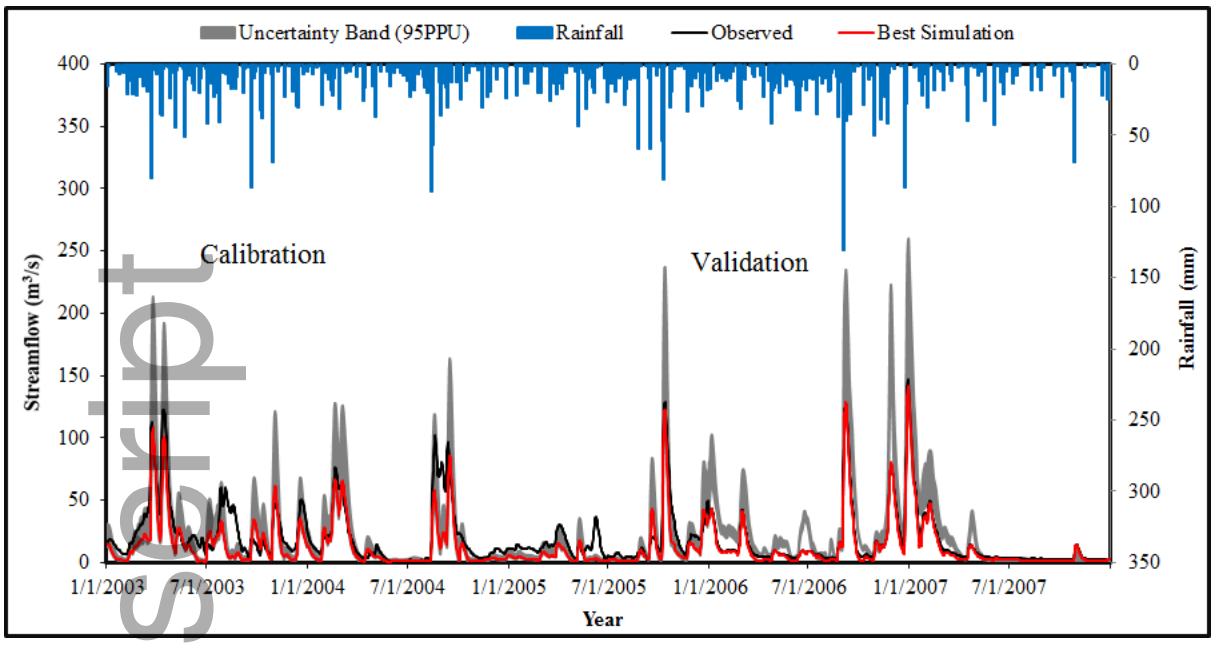
jawra_12596-16-0236_f4.tif



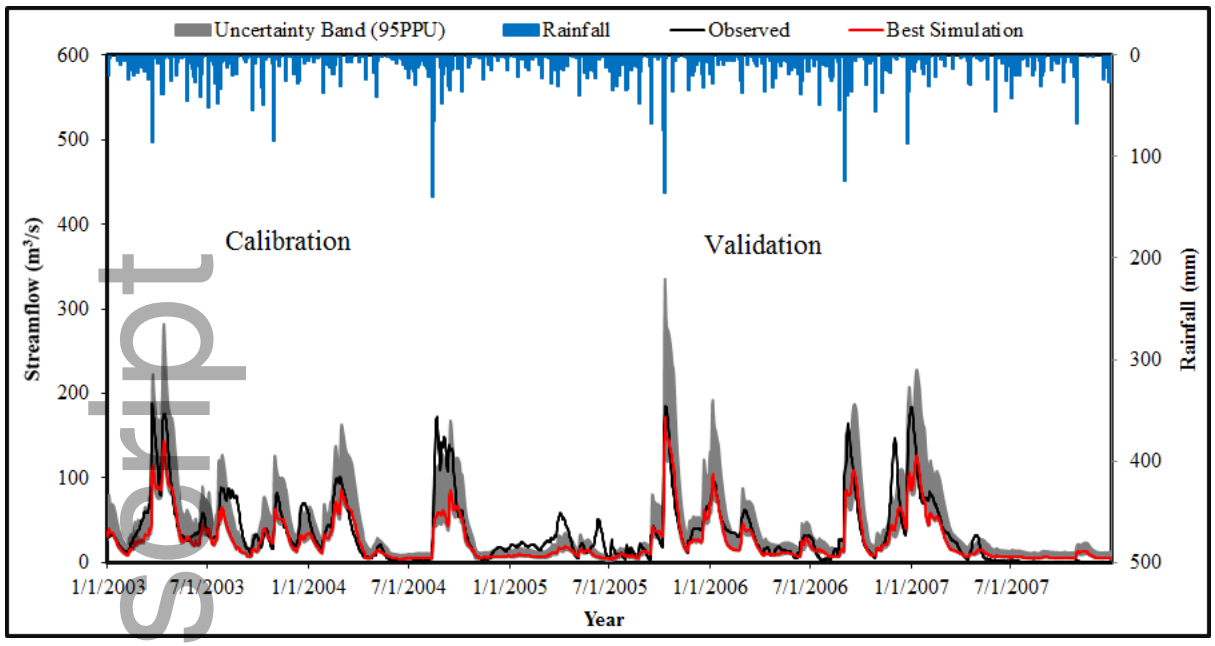
jawra_12596-16-0236_f5.tif



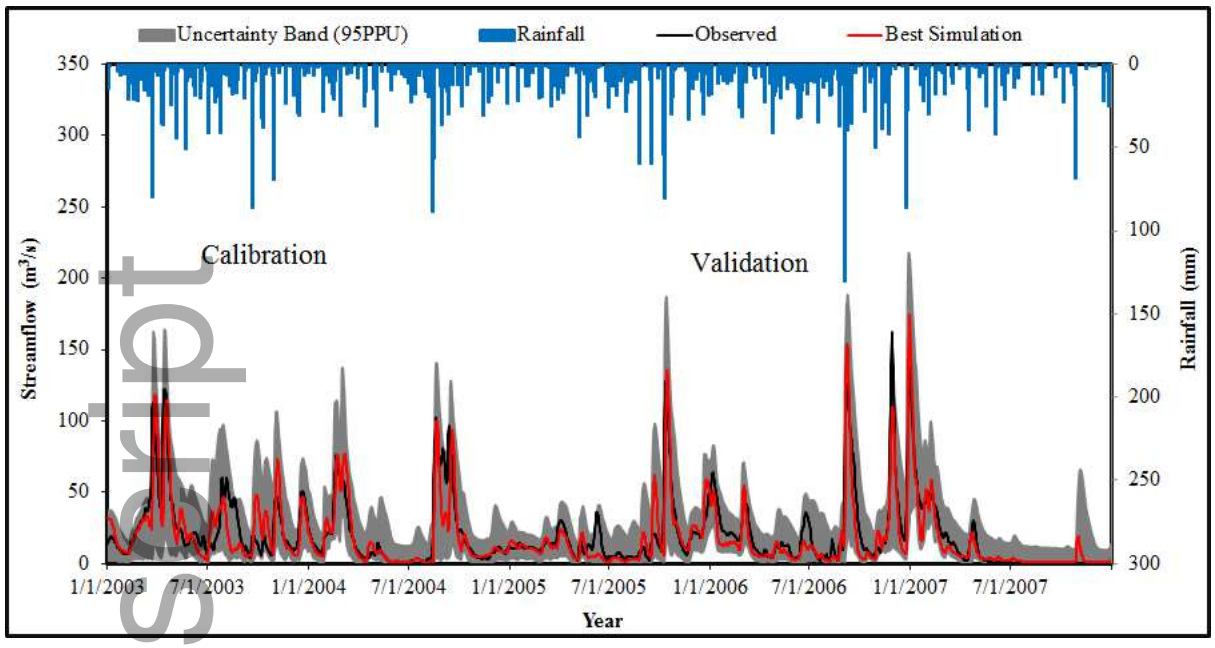
jawra_12596-16-0236_f6.tif



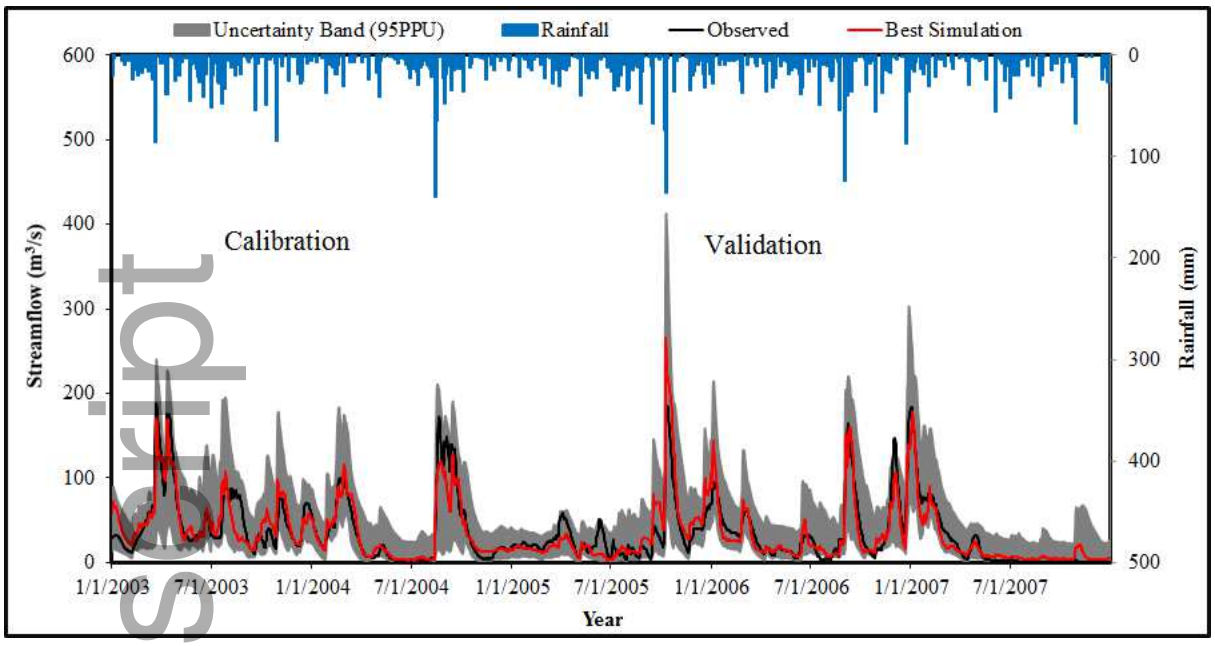
jawra_12596-16-0236_f7.tif



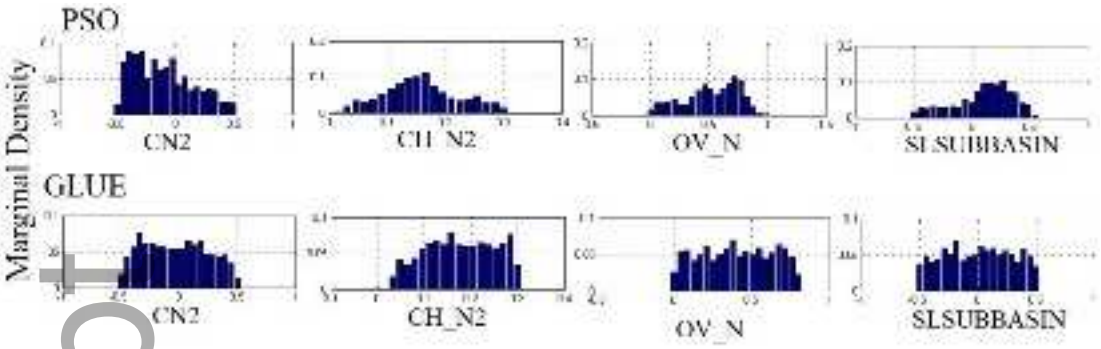
jawra_12596-16-0236_f8.tif



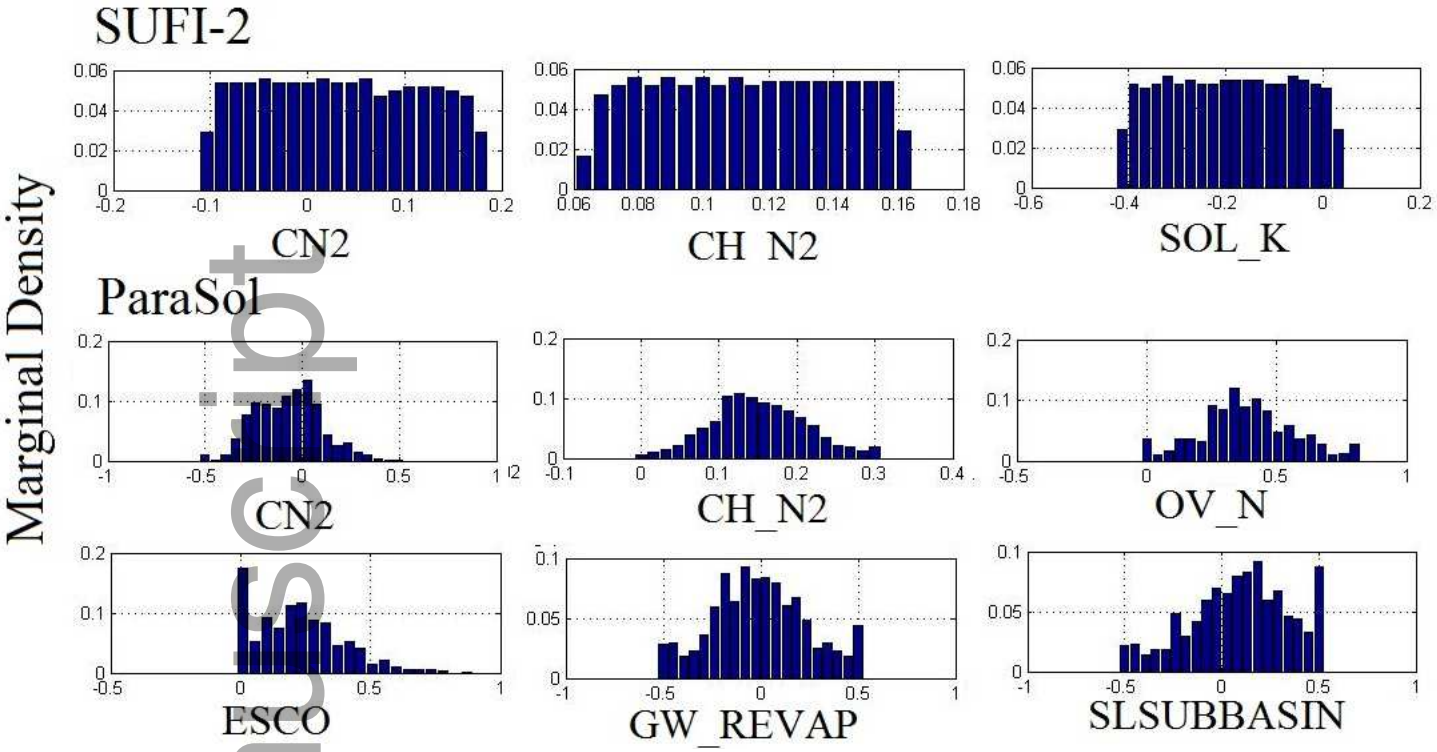
jawra_12596-16-0236_f9.tif



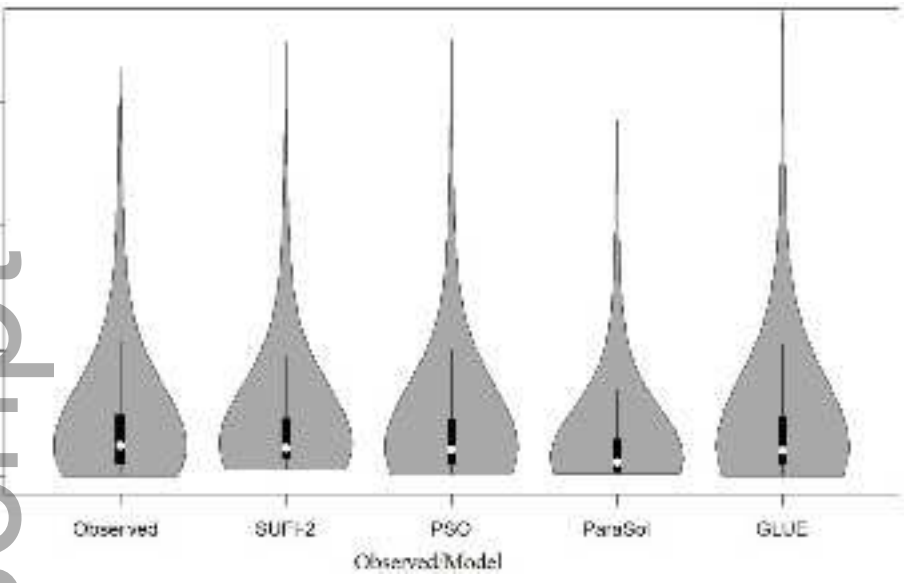
jawra_12596-16-0236_f10.tif



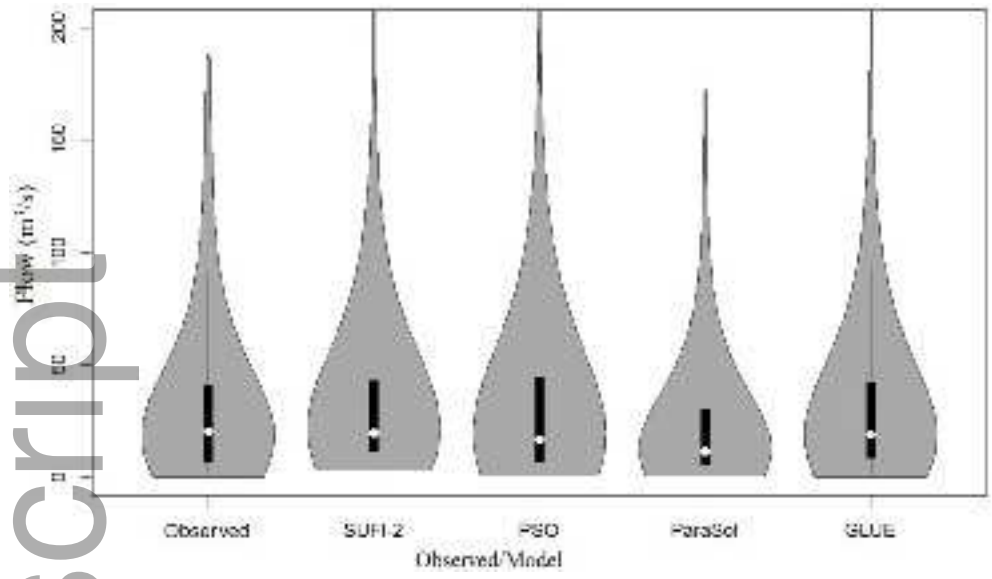
jawra_12596-16-0236_f11.tif



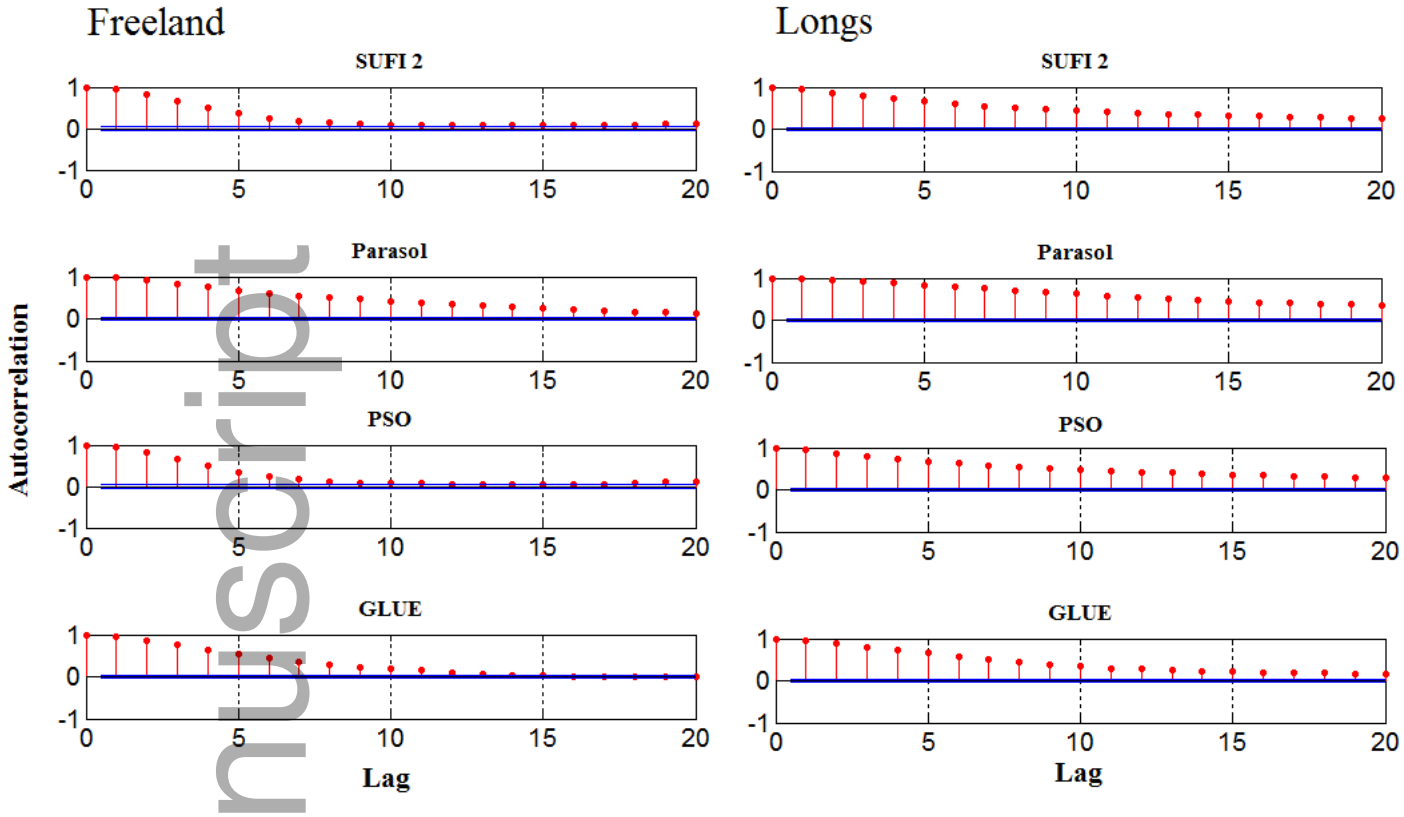
jawra_12596-16-0236_f12.tif



jawra_12596-16-0236_f13.tif



jawra_12596-16-0236_f14.tif



jawra_12596-16-0236_f15.tif