

Estimating hydrologic model uncertainty in the presence of complex residual error structures

S. Samadi¹  · D. L. Tufford² · G. J. Carbone³

© Springer-Verlag GmbH Germany, part of Springer Nature 2017

Abstract Hydrologic models provide a comprehensive tool to estimate streamflow response to environmental variables. Yet, an incomplete understanding of physical processes and challenges associated with scaling processes to a river basin, introduces model uncertainty. Here, we apply generalized additive models of location, scale and shape (GAMLSS) to characterize this uncertainty in an Atlantic coastal plain watershed system. Specifically, we describe distributions of residual errors in a two-step procedure that includes model calibration of the soil and water assessment tool (SWAT) using a sequential Bayesian uncertainty algorithm, followed by time-series modeling of residual errors of simulated daily streamflow. SWAT identified dominant hydrological processes, performed best during moderately wet years, and exhibited less skill during times of extreme flow. Application of GAMLSS to model residuals efficiently produced a description of the error distribution parameters (mean, variance, skewness, and kurtosis), differentiating between upstream and downstream outlets of the watershed. Residual error distribution is better described by a non-parametric polynomial loess curve with a smooth transition from a Box–Cox t distribution upstream to a skew t type 3 distribution downstream. Overall, the fitted models show that low flow events more strongly influence the residual probability distribution, and

error variance increases with streamflow discharge, indicating correlation and heteroscedasticity of residual errors. These results provide useful insights into the complexity of error behavior and highlight the value of using GAMLSS models to conduct Bayesian inference in the context of a regression model with unknown skewness and/or kurtosis.

Keywords GAMLSS · SWAT · Bayesian model · Predictive uncertainty · Residual error distribution · Coastal plain watershed

1 Introduction

Epistemic uncertainty presents a great challenge to hydrology modeling. It arises from inadequate characterization of physical processes and their scaling to the watershed level (Vrugt et al. 2008; McMillan et al. 2012). Consequently, watershed modeling should include an uncertainty confidence range (e.g., Gupta et al. 1998; Beven and Freer 2001; Schoups and Vrugt 2010; Del Giudice et al. 2016; Samadi and Meadows 2017). If properly formulated, this range can help to evaluate different hypotheses about catchment functioning (Wagener and Gupta 2005; Wagener et al. 2007; Sivapalan 2009) and quantify inadequate characterization of system processes and scaling to the watershed level (e.g., Samadi et al. 2017).

In the past two decades, various uncertainty algorithms have been proposed in hydrology to enhance the simulation reliability and to compute hydrologic model errors. These approaches include multi-model averaging methods (Butts et al. 2004; Ajami et al. 2007; Sadegh and Vrugt 2013), state space filtering (Slater and Clark 2006), Bayesian approaches (Laloy and Vrugt 2012; Han et al. 2014),

✉ S. Samadi
samadi@cec.sc.edu

¹ Department of Civil and Environmental Engineering,
University of South Carolina, Columbia, SC, USA

² Department of Biological Sciences, University of South
Carolina, Columbia, SC, USA

³ Department of Geography, University of South Carolina,
Columbia, SC, USA

analysis of variance (e.g., Zhang and Yu 2004; Melching and Bauwens 2001), and stochastic modeling of residuals (Kim and Valdés 2005). These approaches propose different mathematical structures to treat uncertainty and to test different hypotheses and assumptions of residual errors.

Among these approaches, state space filtering, multi-modal averaging methods and Bayesian methods have been widely used in rainfall-runoff applications. Bayesian algorithms include Monte Carlo (MC; informal Bayesian model) and Markov chain Monte Carlo (MCMC; formal Bayesian approaches) samplers that specify the likelihood function in a parameter uncertainty scheme (Vrugt et al. 2008; Beven et al. 2008; Villarini et al. 2009; Zhenyao et al. 2013; Zheng and Han 2016). Model parameter inferences are then based on a likelihood function quantifying the probability that the observed data were generated by a particular parameter set (Box and Tiao, 1992; Schoups and Vrugt 2010; Moore et al. 2010; Wang et al. 2014). The mapping from parameter space to likelihood space results in the identification of a range of plausible parameter sets (Schoups and Vrugt 2010) that are then used to summarize parameter and predictive uncertainty.

In the formal Bayesian approach, one can assume a known statistical model (probability density function; PDF) for the residual errors during calibration. One of the possible assumptions of model errors is that they are independent and identically distributed according to a normal distribution with zero mean and a constant variance σ^2 that result in the standard least squares (SLS) approach for parameter estimation (Schoups and Vrugt 2010). Model errors are often modeled using an additive autoregressive error model (AR; e.g., Kuczera 1983; Bates and Campbell 2001), thus capturing correlation between subsequent model errors (e.g., Yang et al. 2007; Honti et al. 2013; Del Giudice et al. 2013; Sikorska et al. 2012). Statistical assumptions about the error characteristics of the residuals in formal Bayesian approaches are convenient in applying statistical theory, but are not often borne out in the actual calibration which may show changing bias, variance (heteroscedasticity), skewness, and correlation structures (e.g., Vrugt et al. 2008) under different hydrologic conditions. Alternatively, informal (MC-based algorithm) Bayesian models provide flexibility in specifying the form of the likelihood function, which makes this approach attractive in situations where traditional error assumptions are violated (Schoups and Vrugt 2010).

The formal Bayesian algorithm has been criticized for relying too strongly on residual error assumptions that do not hold in many applications in hydrology (Beven et al. 2008; Schoups and Vrugt 2010). In many cases residuals errors are correlated, nonstationary, and non-Gaussian

(e.g., Kuczera 1983; Schoups and Vrugt 2010). Non-stationarity is a common form of heteroscedasticity (Yang et al. 2007; Schoups and Vrugt 2010; Sikorska et al. 2012; Del Giudice et al. 2013), meaning that error variance changes with streamflow discharge (e.g., Sorooshian and Dracup 1980). Therefore, examination of the statistical properties of residual errors can provide insights about their nature and help to assess if the assumptions about the residual error distribution are fulfilled.

The purpose of this study was to compute the characteristics of residuals from a distributed hydrology model (i.e., soil and water assessment tool; SWAT, Arnold et al. 1993) that is coupled with an informal Bayesian algorithm (i.e., Sequential Uncertainty Fitting; SUFI2, Abbaspour et al. 2007). We used SWAT because of its prior successful use in the southeastern US (Amatya and Jha 2011; Zhang et al. 2009; Joseph and Guillaume 2013; Samadi 2016; Samadi and Meadows 2017), and because it is an open source model that can be coupled with various uncertainty algorithms.

Residual errors of SWAT modeling outputs were analyzed using the generalized additive models of location, scale and shape (GAMLSS; Rigby and Stasinopoulos 2005a, b) using the original “GAMLSS” R package (Stasinopoulos and Rigby 2007) and the add-on “gamlss.dist” and “gamlss.data” packages (Stasinopoulos and Rigby 2016). GAMLSS was fitted to highly skewed or kurtotic, continuous to discrete distributions with different additive terms such as smoothing and random-effect terms. The systematic part of the GAMLSS model is expanded to allow modelling not only the mean (or location), but also the scale (or variance), skewness, and kurtosis (or shape) of residual error distribution as linear parametric and/or additive non-parametric functions of an independent variable (i.e., observed streamflow) and/or random effects. Hence, GAMLSS is especially designed to model residual errors (i.e., response variables) that follow normal to highly skewed distributions, (e.g., leptokurtic or platykurtic and/or positive or negative skew response data, or over dispersed counts; Rigby and Stasinopoulos 2005a, b) or those that exhibit heterogeneity (changes in the scale or shape parameters with observed discharge data).

In this study, the model fitting of GAMLSS is achieved by two different algorithmic procedures. The RS algorithm of Rigby and Stasinopoulos (1996a, b) is used as the first algorithm for fitting the mean and dispersion additive models. The second algorithm (Cole and Green 1992; CG algorithm) was applied to model the first derivatives and the expected values of cross derivatives of the likelihood function with respect to distribution parameters. GAMLSS, as a higher order Markov chain model, allows combinations of various types of additive random-effects terms (e.g., cubic splines, multilevel hierarchical models and

crossed and spatial random effects) to be incorporated in the model using shrinking (smoothing) matrices within a backfitting algorithm (Rigby and Stasinopoulos 2005a, b). Using these algorithms, we followed Dunn and Smyth (1996) to account for the non-linearity in the location, the heteroscedasticity in the scale, the heterogeneity in skewness, and possible heterogeneity in the kurtosis parameters of residual probability distributions.

The main contribution of our work lies in the modeling of outliers and appropriate representation of total residual errors, i.e., a combination of measurement, model input, and model structural error, that accounts for modelling (positive or negative) skewness combined with (lepto or platy) kurtosis in error data, and ensures that the residuals probability distribution is accurately identified. Previous studies do not include in-depth analysis of the residual errors and the use of the GAM approaches in hydrologic prediction and modeling, nor do they explicitly address how these models can be coupled with Bayesian algorithmic approaches (see Beven et al. 2008; Vrugt et al. 2008; Schoups and Vrugt 2010; Laloy and Vrugt 2012). Our methodology does not make common assumptions of residual errors; rather it introduces a class of generalized additive models (GAM) models for residual error modeling of hydrologic simulation recently recommended by Serinaldi and Kilsby (2015).

The next section presents residual error estimation, describes the GAMLSS statistical approach, and introduces the SWAT and SUFI2 models. The methodology is then applied in Sect. 3 to identify sensitive parameters, to calibrate streamflow of a distributed hydrology model, and to analyze the residual errors using the GAMLSS model. Section 4 discusses our findings.

2 Methodology

2.1 Prediction error

Model error analysis was undertaken in this study to explore model weaknesses and to attribute differences in predictive ability to individual model components. Model error is actually the deviation of a simulated value from the true value, the analysis that focuses on the deviation of measurement from simulation (Hantush and Kalin 2008). This error is caused by random and/or systematic fluctuations in model dynamics, measurement errors, modeling assumptions, and representations of individual biophysical processes in the system (e.g., Clark et al. 2015).

Model error, ε_t is thus defined by Eq. 1.

$$\varepsilon_t = o_t - q_t, \quad t = 1, 2, \dots, T \quad (1)$$

where o_t , q_t , t and T denote for observed streamflow, model computed streamflow, time index (days) and the length of the time series record, respectively. Model error (ε_t) consists of a combination of input, model structural, output, parameter, and measurement errors.

2.2 Generalized additive models for location, scale, and shape

Different probability distributions have been used to account for skewness and deviation of data in hydrology (e.g., Mosaedi et al. 2015; Etemadi et al. 2015). Several models such as Markov chain models, linear parametric models (ARMA and their extensions; Hipel and McLeod 1994), generalized linear models (GLM; McCullagh and Nelder 1989), and generalized additive models (GAM; Hastie and Tibshirani 1990) are commonly used to fit various distributions to time series data and to estimate distribution parameters. Here, we test a higher order Markov chain model, GAMLSS, that deals with fat- to heavy-tailed datasets.

The GAMLSS package allows modeling of the location parameter (related to the mean/median), and scale and shape parameters (related to the dispersion, skewness, and kurtosis; Serinaldi and Cuomo 2011) of the distribution of Y as linear and/or non-linear, parametric and/or additive nonparametric functions of covariates (Sénégal et al. 2001) and/or random effects (e.g., Rigby and Stasinopoulos 2005a, b; Serinaldi 2011). The probability distribution of Y may be chosen from discrete to continuous and mixed distributions such as highly skewed and/or leptokurtic probability distributions. In this study, error standard deviation (ε_t from Eq. 1) is modeled as a function of observed streamflow; if daily streamflow has a large standard deviation, the magnitude of the random components will be large. Non-normality is then accounted for with the discrete to continuous and mixed probability distributions (overall, 80 probability distributions are available in GAMLSS).

GAMLSS assumes that independent observations (Y_i) for time series of $i = 1, \dots, n$, have $F_y(y_i; \theta^i)$ as a distribution function conditional on θ^i where $\theta^i = (\theta_1^i, \dots, \theta_p^i)$ is a vector of p parameters, each of which is related to an independent variable (e.g., observed streamflow; see more at <http://www.gamlss.org/>). Y is not restricted to follow a distribution from the exponential family such as a Gaussian distribution (Villarini et al. 2009; Serinaldi 2011). This allows different distribution functions, like highly skewed and/or kurtotic continuous and discrete distributions, apply to model the complex residuals common in hydrologic modeling. Usually p is less than or equal to four, because 1-, 2-, 3-, and 4-parameter families better describe the tail

behavior of distribution in most applications. The R implementation specifies these parameters as $(\mu_i, \sigma_i, v_i, \tau_i)$. μ_i and σ_i are known as location and scale parameters, while (v_i, τ_i) are characterized as shape parameters. The location parameter specifies the center of the distribution, the scale parameter identifies the size of deviations around the location parameter, and the shape parameter demonstrates the tail behavior of fitted probability distribution (e.g., Katz 2010).

Given an n as the length vector of the response variable, $y^T = (y_1, y_2, \dots, y_n)$ denotes the vector of the response variable observations. If $g_k(\cdot)$ for $k = 1, 2, 3, 4$, denotes as monotonic link functions relating the k th parameter to θ_k explanatory variables, the semi-parametric additive models is given by;

$$\begin{aligned} g_1(\mu) &= \eta_1 = X_1\beta_1 + \sum_{j=1}^{J_1} h_{j1}(x_{j1}) \\ g_2(\sigma) &= \eta_2 = X_2\beta_2 + \sum_{j=1}^{J_2} h_{j2}(x_{j2}) \\ g_3(v) &= \eta_3 = X_3\beta_3 + \sum_{j=1}^{J_3} h_{j3}(x_{j3}) \\ g_4(v) &= \eta_4 = X_4\beta_4 + \sum_{j=1}^{J_4} h_{j4}(x_{j4}) \end{aligned} \quad (2)$$

where $\mu, \sigma, v, \tau, \eta_k$ and x_{jk} , for $j = 1, 2, \dots, J_k$ and $k = 1, 2, 3, 4$, are vectors of length n . The function h_{jk} is a non-parametric additive function of the independent variable X_{jk} explained by x_{jk} . X_{jk} is a known design matrix (design matrix for the independent variable x (i.e., observed streamflow) and the response variable (i.e., error time series) of order $n \times J_k$) while β_k are the parameter vectors. X follows a distribution from different family (e.g., Gaussian distribution, skewed distribution, etc.) that allows for highly skewed and/or kurtotic continuous and discrete distributions. x_{jk} is assumed fixed and known. Equation 2 is called the semi-parametric GAMLSS model. By including random effect terms Eq. 2 can be extended as Eq. 3.

$$\begin{aligned} g_1(\mu) &= \eta_1 = X_1\beta_1 + \sum_{j=1}^{J_1} Z_{j1}\gamma_{j1} \\ g_2(\sigma) &= \eta_2 = X_2\beta_2 + \sum_{j=1}^{J_2} Z_{j2}\gamma_{j2} \\ g_3(v) &= \eta_3 = X_3\beta_3 + \sum_{j=1}^{J_3} Z_{j3}\gamma_{j3} \\ g_4(v) &= \eta_4 = X_4\beta_4 + \sum_{j=1}^{J_4} Z_{j4}\gamma_{j4} \end{aligned} \quad (3)$$

where γ_{jk} have independent (prior) normal distribution with $\gamma_{jk} \sim N_{q_{jk}}(0, G_{jk}^{-1})$ and G_{jk}^{-1} is the (generalized) inverse of a $q_{ik} \times q_{jk}$ symmetric matrix $G_{jk} = G_{jk}(\lambda_{jk})$, which may depend on a vector of hyper-parameters and/or non-linear parameters (λ_{jk}) . The parametric vectors β_k and the random effects parameters γ_{jk} , are estimated for $j = 1, 2, \dots, J_k$ and $k = 1, 2, 3, 4$ for fixed values of the smoothing hyper-parameters λ_{jk} by maximizing a penalized likelihood function ℓ_p (Eq. 4; Rigby and Stasinopoulos 2005a, b). Z_{jk} is a fixed known $n \times q_{jk}$ design matrix (n is number of time series data and q_{jk} is driven by the fitted probability distribution). Note that in the GAMLSS package different sub models were involved including: parametric, semi-parametric, non-parametric and random-effects terms. For instance, if $Z_{jk} = I_n$, where I_n is an $n \times n$ identity design matrix, and $\gamma_{jk} = h_{jk} = h_{jk}(X_{jk})$ is a specific combinations of j and k in the model, then the resulting model contains a semi-parametric additive formulation of GAMLSS as below:

$$\ell_p = \ell - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} G_{jk} \gamma_{jk} \quad (4)$$

where ℓ_p is the penalized log-likelihood. $\ell = \sum_{i=1}^n \log f(y_i/\theta^i)$ is the log likelihood function of the distribution parameters for $j = 1, 2, \dots, J_k$ and $k = 1, 2, 3, 4$. GAMLSS uses two main algorithms to fit different probability distributions. These are the CG and RS algorithms. The CG algorithm uses the first derivatives and the expected values of the second and cross derivatives of the likelihood function with respect to $\theta = (\mu, \sigma, v, \tau)$ (e.g., Rigby and Stasinopoulos 1996a, b). Although in many applications, in the population density functions $f(y/\theta)$ the parameter θ is orthogonal since the expected values of the cross derivatives of the likelihood function are zero (e.g., the location and scale parameters and dispersion family models) or approximately so (Rigby and Stasinopoulos 2005a, b). In such cases, a more suitable choice is the RS algorithm that used by Rigby and Stasinopoulos (1996a, b) to fit mean and dispersion additive models (MADAM).

The smoothing hyper-parameters λ_{jk} or the additive functions can be parametric and/or non-parametric smoothing terms. Example includes cubic splines proposed by Green and Silverman (1994), penalized splines developed by Eilers and Marx (1996), fractional polynomials of Royston and Altman (1994), or non-linear power polynomials. The power parameter can take any real values; for instance, the fractional polynomial generates the right design matrix algebra for fitting a power polynomial of the type $a + b_1 x_1^{p_1} + b_2 x_2^{p_2} + \dots + b_k x_k^{p_k}$. For given powers p_1, p_2, \dots, p_k , the argument "powers" in the GAMLSS code can be used to fit orthogonal or piecewise

polynomials. The function of power is used to fit the best fractional polynomials among a specific set of power values (Rigby and Stasinopoulos 1996a, b) and GAMLSS ultimately determines whether one, two, or three fractional polynomials should employ in the fitting (argument *npoly* in the GAMLSS package). GAMLSS employs different combinations of additive terms in the model using shrinking (smoothing) matrices within a backfitting algorithm (Rigby and Stasinopoulos 2005a, b). Table 1 shows a list of the different additive terms available in the GAMLSS models.

Model selection was based on the significance of the fitting improvement using Bayesian statistics including the Akaike information criterion (AIC; Akaike 1974), the Schwarz bayesian criterion (SBC; Schwarz 1978), or the generalized AIC (GAIC; Rigby and Stasinopoulos 2005a, b) and testing the significance of the differences between the deviance values of different models (e.g., Serinaldi and Kilsby 2015).

The kurtosis parameter (τ) determines the shape of probability distribution. Positive kurtosis indicates the possibility of a leptokurtic distribution (too tall) and a negative value determines the possibility of a platykurtic distribution (too flat). The skewness (ν) parameter, on the other hand, affects asymmetry ($\nu > 0$), as illustrated in Fig. 1. The distribution is symmetric if $\nu = 0$ (the mean is equal to the median) and positively (negatively) skewed if $\nu > 0$ ($\nu < 0$). Figure 1 is an illustration of the Box–Cox t (BCT) and the skew t type 3 distribution (i.e., “spliced-scale” distribution, Fernandez et al. 1995) parameters that showed the best fit for Freeland and Longs outlets. In this case, a Normal (Gaussian) distribution results when $\tau = 0$, a Chi square distribution when $\tau = 1$, and a skew t type 3 distribution when $\tau > 0$ (log *sigma*, log *nu*, and log *tau*).

Table 1 Additive terms implemented within the GAMLSS package in R (adapted from Stasinopoulos and Rigby 2007)

Additive terms	R function names
Cubic splines based	cs(), scs()
Decision trees	tr()
Fractional and power polynomials	fp(), pp()
Free knot smoothing (break points)	fk()
Loess	lo()
Neural networks	nn()
Non-linear fit	nl()
Penalized beta splines based	pb(), ps(), cy(), pvc()
Random effects	random()
Ridge regression	ri(), ridge()
Simon wood's	gam ga()

Hence, ν and τ lead us to relax the assumption of error distribution.

In particular, values for $\tau > 0$ result in taller densities with longer tails compared to a Gaussian/Normal PDF, which is useful for making parameter inference robust against outliers (e.g., Schoups and Vrugt 2010). The probability density function of a distribution with larger kurtosis (maximum value) has a high peak and fatter tails, compared with the probability density function of a distribution with smaller kurtosis.

2.3 Study area and data

The Waccamaw River watershed spans 311,685 ha on the coastal plain of North and South Carolina (Fig. 2). Annual average temperature (1946–2009) of this watershed system is 16.9 °C; annual precipitation averages a little more than 1300 mm. During the simulation period (2003–2007), precipitation ranged from an extremely wet year in 2003 (320 mm above average rainfall) to an extremely dry year in 2007 (around 330 mm below average). Soils are typically sandy loam and sandy clay loam—moderately drained in uplands and poorly drained in riparian zones and floodplains.

The watershed can be divided into two major landscape units with heterogeneous characteristics. The downstream part is characterized by low elevation (< 20 m) coastal region with Rains soil type (poorly drained) and forested and non-forested wetlands as the major land use categories (around 30% of the study area). The upstream part is somewhat higher elevation (up to 50 m) dominated by Meggett soil (poorly drained) and mixed and evergreen forest. The entire watershed has low slope gradient. According to SWAT soil outputs, 93.6% of the watershed has low storage soils (hydrologic groups B and D). Agriculture, rangeland, forested-rangeland, and forested wetland are the major land use categories.

There are no major dams or flow control structures in the watershed. There is a weir dam at the outlet to Lake Waccamaw; a shallow (mean depth 1.5 m) permanently flooded Carolina Bay in the headwaters of the watershed (Riggs et al. 2000). Though large for a Carolina Bay (3622 ha), its shallow depth and position in the headwaters suggest a minor influence on streamflow response at the calibration points.

Five years of approved daily streamflow data ranging from January 2000 to December 2007 were collected from the USGS website for the Freeland and Longs stations. Three years (2003–2005) were used to calibrate SWAT and the remaining 2 years (2006–2007) were used to validate the performance of the model in predicting daily streamflow. Meteorological data (daily precipitation, maximum and minimum air temperatures, wind speed, humidity, and

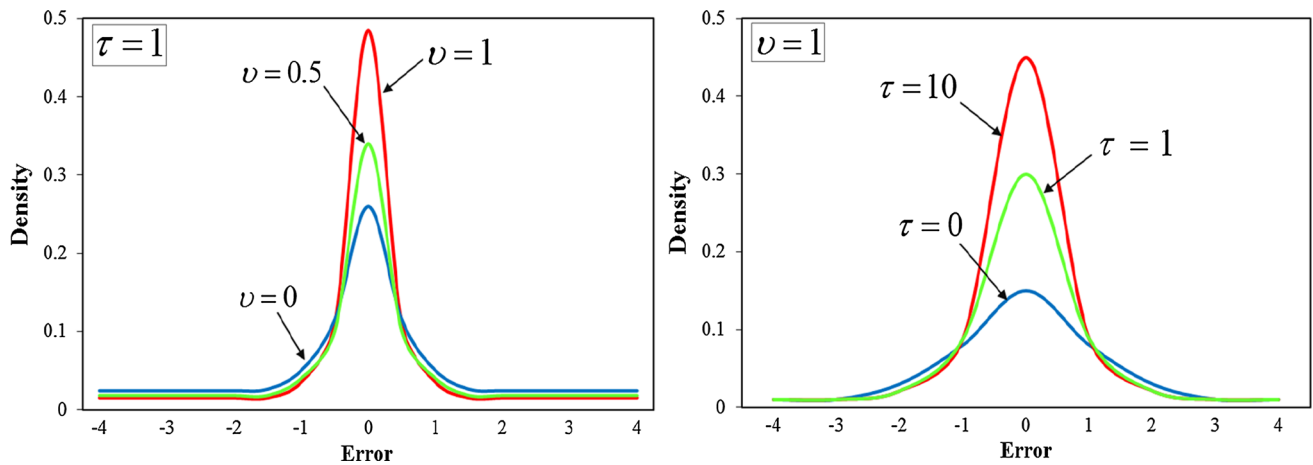


Fig. 1 Densities of the BCT (left, $\nu = 1$) and the skew t type 3 (right, $\tau = 10$) distributions with different values of τ and ν . Note that the location parameter in the BCT and the skew t distributions is equal to 5 and 0, respectively

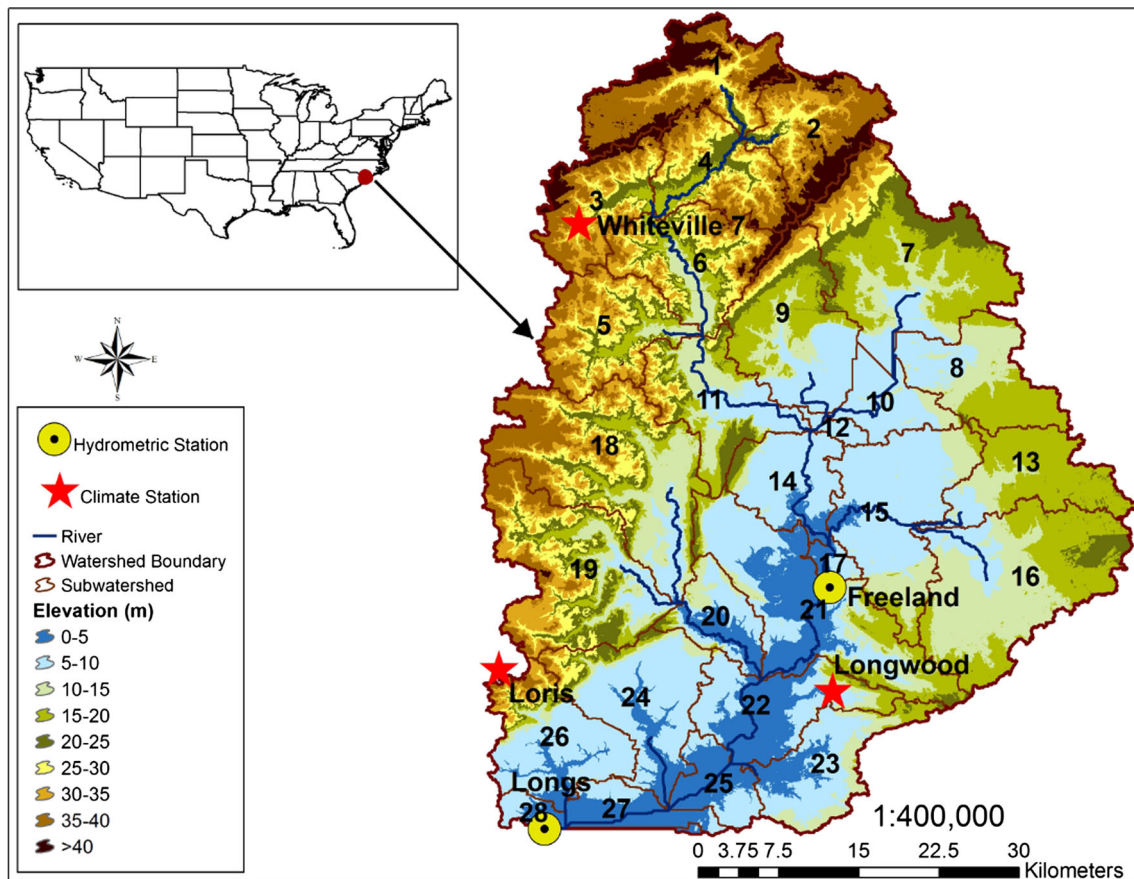


Fig. 2 The 311,685 ha Waccamaw River Watershed delineated by SWAT. Model calibration, validation, and sensitivity/uncertainty analysis were performed using data from US Geological Survey (USGS) gaging stations at Freeland and Longs

solar radiation), and spatial data inputs (digital elevation model (DEM), land use, and soil coverage) were acquired from the National Climatic Data Center (NCDC) and USGS portals on 06 and 13 September 2012, respectively. Whiteville7, Longwood, and Loris climatic data, located

inside the watershed boundary, were used to build SWAT climate forcing data (see Fig. 2). We interpolated data from these stations using Thiessen polygon and Inverse Distance Weighting (IDW) methods to capture the spatial continuity of rainfall fields.

2.4 SWAT model

SWAT is a semi-distributed watershed hydrologic and water quality model developed by Arnold et al. (1993). It estimates the surface runoff volume from each hydrologic response unit (HRU) using either the soil conservation service (SCS) curve number (CN) method (USDA-SCS 1972) or a newer CN method that includes a one-parameter soil moisture depletion curve to account for antecedent soil moisture conditions (CN ranges 25–98 in the SWAT model). SWAT uses Manning's equation to estimate flow rates and velocity in the stream. It routes water in the reach using the variable storage method (Williams 1969) or the Muskingum routing method, both variations of the kinematic wave model. SWAT has three options for estimating potential evapotranspiration (PET) using well-known empirical models: Priestley–Taylor (Priestley and Taylor 1972), Penman–Monteith (Monteith 1965), and Hargreaves (Hargreaves et al. 1985). Users also can incorporate observed evapotranspiration values in the model.

In this research, a setup of the Priestley–Taylor method for PET estimation (see Samadi 2016), Muskingum channel routing method, and the improved one-parameter depletion coefficient for adjusting the CN based on plant ET method provided the best results. HRU definitions were established at a 0% threshold for slope, soil, and land use layers to avoid missing small land use units. SWAT ultimately subdivided the Waccamaw watershed into 28 sub-watersheds (Fig. 2) and 2020 HRUs.

2.5 Uncertainty and sensitivity analyses of SUFI2

This research examined a Bayesian framework (i.e., SUFI2) as an uncertainty experiment linked with the SWAT model. SUFI2 uses a multiple regression method to compute parameter sensitivity. Sensitivity analysis of this algorithm defines which parameters have significant influences on the model simulations by presenting the smallest p -value (values close to zero signify a high level of significance) and the largest t-Stat (a measure of sensitivity, larger in absolute values are more sensitive), respectively. The t-Stat represents the precision of the regression coefficient. It is the coefficient of a parameter divided by its standard error. If a coefficient is large compared to its standard error, then it is probably different from 0 and the parameter is sensitive (e.g., Abbaspour 2015). SUFI2 then compares the t-Stat of a parameter with the values in the Student's t-distribution (defines the behavior of the mean of a sample) table to determine the p value. The p value for each term tests the null hypothesis that the coefficient is equal to zero (no effect, see Abbaspour 2015) a low p value (< 0.05) rejects the null hypothesis; thus, predictors (e.g., Manning's n -value) with low p values are likely to be

meaningful additions to the model. Conversely, predictors with large p values are more weakly associated with changes in the response variable (the parameter is not very sensitive).

SUFI2 initially assumes large uncertainty within physically meaningful ranges, then narrows the uncertainty band in sequential manner while monitoring the P -Factor and R -Factor (Eqs. 5, 6). The P -Factor is a measure of parameter uncertainty, specifically the percentage of data bracketed by 95% prediction uncertainty (95 PPU). The R -Factor is the ratio of the average thickness of 95 PPU bound to the standard deviation of the corresponding observed variable (see Abbaspour 2015). Using these two measures, a perfect simulation would have a P -Factor of 1 and an R -Factor of zero.

$$\overline{d_x} = \frac{1}{k} \sum_{i=1}^k (X_U - X_L) \quad (5)$$

$$R - factor = \frac{\overline{d_x}}{\sigma_x} \quad (6)$$

SUFI2 obtains the posterior parameters from prior assumptions to calculate the likelihood function of each streamflow event. In SUFI2, 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. SUFI2 first defines the goal function $g(\theta)$ with meaningful parameter ranges $[\theta_{absmin}, \theta_{absmax}]$. Then, a Latin hypercube sampling is carried out in the initial hypercube $[\theta_{absmin}, \theta_{absmax}]$ 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) are then calculated. If those two factors have satisfactory values, the parameter range travels from $[\theta_{absmin}, \theta_{absmax}]$ 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 (Yang et al. 2008). According to this procedure, SUFI2 was calibrated for 2000 runs in the calibration period and 500 runs in the validation period, respectively.

The R -Factor and P -Factor can be used for statistical analysis of modeling skills along with other functions such as coefficient of determination (R^2) and Nash–Sutcliffe

efficiency (*NSE*; Nash and Sutcliffe 1970). Other statistical indices are coefficient of determination (R^2) multiplied by the coefficient of the regression line (BR^2 , $0 \leq BR^2 \leq 1$, Eq. 7), and sum of the squares of the differences of the observed and simulated values after ranking (*SSQR*, Eq. 8). BR^2 ($0 \leq BR^2 \leq 1$, Equation 7) is R^2 multiplied by the coefficient of the regression line. This function accounts for discrepancies in the magnitude of two signals (depicted by b) as well as their dynamics (depicted by R^2 ; see Abbaspour 2015). The *SSQR* (Eq. 8) is highly influenced by the magnitudes of the differences between observed ($Q_{i,observed}$) and simulated ($Q_{i,simulated}$) values, because it assesses the relative magnitude of the bias among model. n is the number of pairs and i represents the rank.

$$\varphi = \begin{cases} |b|R^2 & \text{if } |b| \leq 1 \\ |b|^{-1}R^2 & \text{if } |b| > 1 \end{cases} \quad (7)$$

$$SSQR = \sum_{i=1,n} [Q_{i,observed} - Q_{i,simulated}]^2 \quad (8)$$

These indices were tested in SUFI2 using different cut-off thresholds. When we selected *NSE* as the objective function, SUFI2 provided better simulation in terms of low and peak flow fit. As stressed by Sevat and Dezetter (1991), ASCE (1993), Legates and McCabe (1999) and Shrestha et al. (2016), *NSE* is the best objective function for reflecting the overall fit of a hydrograph. It provides better statistics on simulated values, and, thus, is widely used as an objective measure in hydrologic simulations (Beven and Freer 2001; McCuen et al. 2006; Yang et al. 2008; Etemadi et al. 2014, 2015).

2.6 Setting the GAMLSS framework

We input 5 years of daily error time series data simulated by the SUFI2 algorithm into the GAMLSS model to fit an appropriate distribution function (DF) and to estimate four parameters in a fitted distribution family, i.e., μ , σ , ν and τ . When modelling the distribution of the residual errors, we accounted for: (1) the non-linearity in the median or location, (2) the heteroskedasticity in the scale, (3) the heterogeneity in skewness, and (4) possible heterogeneity in the kurtosis (Stasinopoulos and Rigby 2007). Once the effective DF, along with four distribution parameters, were defined for both outlets, Bayesian statistics were used to compare the fit of different additive terms or smoothing terms based on the global deviance (see Rigby and Stasinopoulos 2005a, b).

The main advantage of residual analysis in GAMLSS is that the performance of the selected distribution can be evaluated by independence of residuals on the response variable as well as the residuals' normality diagnostics. The normalized (randomized) quantile residuals are those that, whatever the distribution of the response variable their true

values $r_i, i = 1, 2, \dots, n$ always follow a Gaussian family distribution (Stasinopoulos and Rigby 2007). Examining the normalized (randomized) quantile residuals allowed us an easy way to evaluate the fitted model. In this study, we visually inspected the normalized quantile residuals shown in four residual plots: (1) observed streamflow versus residual error, (2) quantile residuals against the fitted values, (3) a density plot versus quantile residuals, and (4) sample quantiles versus theoretical quantiles. Note that residuals are randomized only for discrete response variables as advocated by Dunn and Smyth (1996). They depend solely on explanatory variables (i.e., observed streamflow), whereas the standard deviations of the residuals (σ) depend on both explanatory variables and error probability distribution (e.g., Westra et al. 2014).

We further examined the residuals error distribution during different ranges of simulation period by analyzing worm plots (argument *wp()*) in GAMLSS. A worm plot is a series of detrended quantile–quantile (Q–Q) plots, split by covariate levels to diagnose how well a statistical model fits the data, to identify regions (intervals) of an independent variable within which the model does not fit adequately (called “model violation”, see van Buuren and Fredriks 2001), and to compare the fit of different models. If all points (i.e., error values) fall in the “acceptance” region inside the two elliptic curves (the point-wise 95% confidence regions) and no specific shape is detected in the points, the overall model appears to fit well (e.g., Stasinopoulos and Rigby 2007). In this study, we performed worm plots for short periods by dividing the residual error data into different seasons (argument *n.inter* = 9 for Freeland and *n.inter* = 20 for Longs). We then analyzed different shapes for the worm plots (Table 2) and interpreted the fitting of residual errors in each season.

Figure 3 illustrates our method of linking SWAT with GAMLSS and a Bayesian framework to compute uncertainty, to improve a priori information, and to specify model violation explicitly. We simulated streamflow in hydrologic response units based on the land use, soil, and slope layers. The HRUs are used to assess the varying hydrologic conditions between sub-watersheds. SWAT calculated the initial soil water content using effective rainfall, evapotranspiration, surface runoff, the amount of water entering the shallow aquifer from the soil profile, the amount of groundwater, and the amount of return flow. After saturating the shallow aquifer, predicted runoff was routed to a deep aquifer as well as the channel to obtain the total streamflow (base flow and runoff). Next, SWAT output was linked with SUFI2 and examined sequentially (overall 2500 runs for calibration and validation periods) to obtain simulated streamflow and 95PPU. The best values of sensitive parameters were then used to optimize SWAT simulation and to obtain water balance estimates (not

Table 2 different shapes of the worm plot for fitted curve (first column) and the corresponding deficiency in the residuals (second column) and response variable distribution (third column; adapted from Stasinopoulos and Rigby 2007)

Shape of worm plot (or its fitted curve)	Residuals (i.e., observed streamflow minus simulated streamflow)	Response variable (i.e., error values)
Level: above the origin	Mean to high	Location parameter too low
Level: below the origin	Mean too low	Location parameter too high
Line: positive slope	Variance to high	Scale parameter too low
Line: negative slope	Variance to low	Scale parameter too high
U-shape	Positive skewness	Skewness too low
Inverted U-shape	Negative skewness	Skewness too high
S-shape with left bent down	Lepto-kurtosis	Kurtosis too low

In this study, these shapes were tested and fitted to both outlets datasets

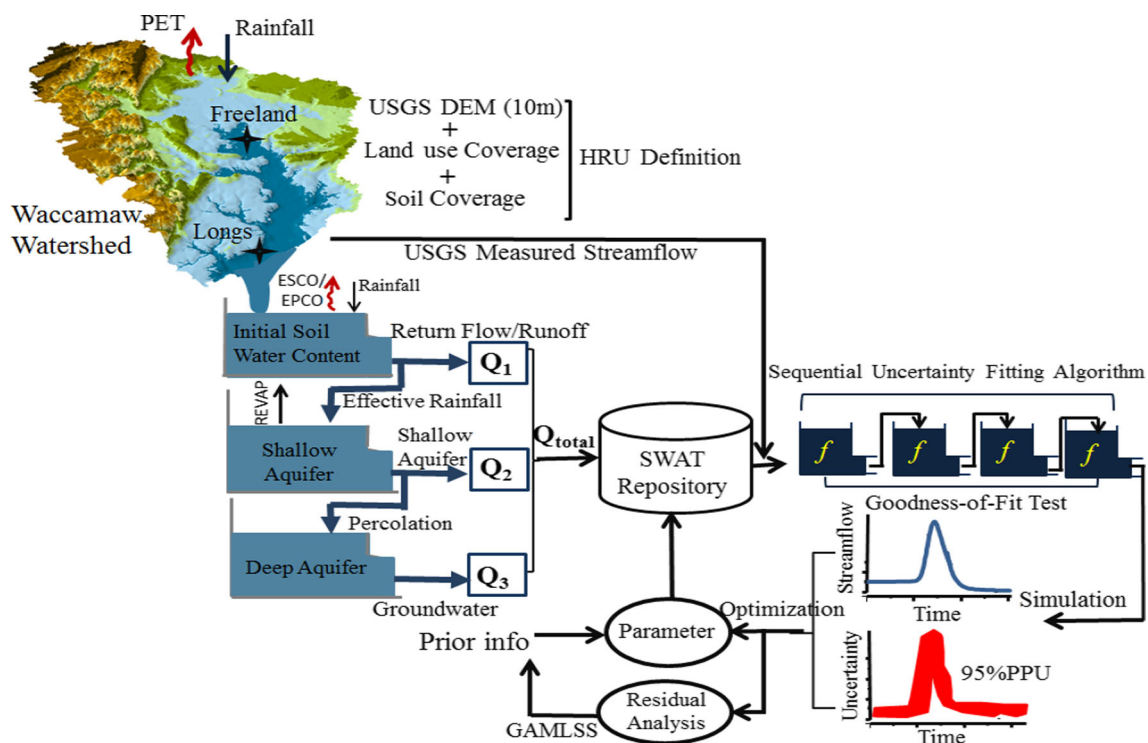


Fig. 3 A schematic framework of uncertainty and the residual error analysis for the Waccamaw watershed. SWAT simulates flow at each HRU using three reservoirs, i.e., initial soil water content (runoff), shallow aquifer, and deep aquifer (groundwater). SUFI2 was coupled

to SWAT using SWATCUP (SWAT calibration and uncertainty procedures; Abbaspour 2015) interface. See Table 3 for the definitions of ESCO, EPCO and REVAP

shown here). Various GAMLSS models were linked to the outputs of SUFI2 in order to compute the statistical distribution of residual errors, to describe model violation, and to assess at which locations the fit can be improved. These approaches can then improve our prior knowledge driving the Bayesian uncertainty algorithm.

3 Results

3.1 Sensitivity analysis

We tested the sensitivity of eighteen flow parameters; these were ranked qualitatively based on their sensitivity during a calibration period (Table 3). This study evaluated and compared parameter estimates and uncertainty using 3-year

Table 3 Candidate parameters for uncertainty analysis and their prior distribution in SUFI2

Aggregate parameter ^a	Name of SWAT parameter	Parameter's range		
		Sensitivity	Parameter Range	Optimal value
v__CH_N2.rte	Manning's <i>n</i> -value for the main channel	Most sensitive	0.06–0.16	0.11
r__SOL_K().sol	Saturated hydraulic conductivity (mm/hr)	Most sensitive	– 0.03 to –0.41	– 0.27
v__CH_K2.rte	Effective hydraulic conductivity in tributary channel alluvium (mm/hr)	Most sensitive	0.0–62.86	4.33
v__SLSUBBSN.hru ^b	Average slope length (m)	Moderate sensitive	110.37–150	132.84
v__ESCO.hru	Soil evaporation compensation factor	Moderate sensitive	0.63–1	0.80
v__OV_N.hru	Manning's <i>n</i> -value for overland flow	Moderate sensitive	0.0–0.28	0.17
r__SOL_BD().sol	Moist bulk density (Mg/m ³ or g/cm ³)	Moderate sensitive	– 0.11 to 0.19	0.035
r__GWHT.gw	Initial groundwater height (m)	Moderate sensitive	0.71–2.28	1.65
v__EPCO.hru	Plant uptake compensation factor	Moderate sensitive	0.55–0.97	0.91
r__SHALLST.gw	Initial depth of water in the shallow aquifer (mm H ₂ O)	Moderate sensitive	–0.72 to 0.14	– 0.57
v__ALPHA_BF.gw	Base flow alpha factor (days)	Less sensitive	0.16–0.49	0.30
r__CN2.mgt	SCS runoff curve number for moisture condition II	Less sensitive	– 0.10 to 0.17	– 0.002
v__RCHRG_DP.gw	Deep aquifer percolation fraction	Less sensitive	0.009–0.43	0.42
v__LAT_TTIME.hru	Lateral flow travel time (days)	Less sensitive	78.65–178.98	127.31
r__SOL_AWC().sol	Available water capacity of the soil layer (mm H ₂ O/mm) soil)	Less sensitive	– 0.06 to 0.36	0.15
v__GW_REVAP.gw	Groundwater “revap” coefficient	Less sensitive	0.07–0.16	0.1
r__GW_SPYLD.gw	Specific yield of the shallow aquifer (m ³ /m ³)	Less sensitive	– 0.65 to 0.35	– 0.3
v__GW_DELAY.gw	Groundwater delay time (days)	Less sensitive	321.03–500	412.48

^aAggregate parameters are constructed based on (SWACUP user manual, Abbaspour, 2015). “v__” and “r__” means a replacement, and a relative change to the initial parameter values, respectively

^bPrior distributions of aggregate parameters are based on Neitsch et al. (2001)

calibration periods with 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 simulated, and for each data set, parameter sensitivity was evaluated. Our analysis suggests that 3 years of daily streamflow data contains enough information about the parameters estimation of SWAT inputs for this watershed. Therefore, no significant variation in parameter estimates exists among different periods. Parameters with *p* value less than 0.05 are very sensitive while parameters with *p* values 0.05–0.075 are moderately sensitive. Parameters with *p* value 0.075–0.1 are less sensitive, and larger *p* value (> 0.1) indicates no sensitivity. In this study, various parameter combinations were tested during the calibration period and SUFI2 ultimately showed greatest sensitivity to soil and hydraulic properties.

3.2 Parameter calibration

3.2.1 Sensitive parameters

In this study, Manning's *n*-value for main tributary channels (CH_N), saturated hydraulic conductivity (SOL_K), and effective hydraulic conductivity in tributary channels (CH_K) were the most sensitive parameters during the calibration period. Sensitivity analysis further shows higher sensitivity of SWAT parameters during dry periods than wet periods, reflecting the challenges of simulating rainfall-runoff processes during dry period. We tested several simulation intervals (results not shown here) and found that SWAT is very sensitive to Manning's *n*-value during all calibration periods. Sensitivity analysis revealed that the uncertainty propagation due to hydraulic parameters is drastically increased with increasing floodplain width and

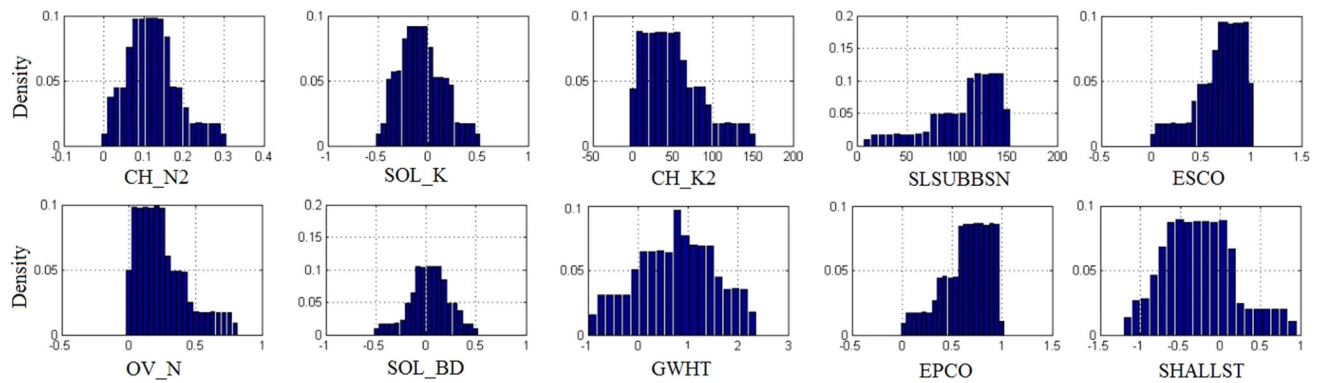


Fig. 4 Marginal posterior probability distributions of ten high to moderate sensitive parameters in the SWAT model inferred using SUFI2. SWAT was optimized using these ten sensitive parameters to update/optimize hydrologic variables at HRU level

showed more sensitivity during the summer season than during winter. We further examined the uncertainty associated with the Manning's n -value by calibrating the streamflow individually at each stream gauge. The results demonstrated that the downstream outlet (Long's) was more sensitive to changes in Manning's n -value than the upstream outlet (Freeland). Simulated flow at Freeland was relatively insensitive to the Manning's n -value as observed by a non-significant p value (0.34) and a lower t -State (0.37). Long's by contrast had a significant p value (0.000001) and a high t -State (8.48). The Manning's n -value included in SWAT is averaged across a small portion of the modeling domain; hence, model geometries are only an approximation to real system that increases the sensitivity of geophysical parameters. In this study, an average fitted value for Manning's n -value was 0.1 for the Waccamaw River watershed.

Saturated hydraulic conductivity was the second most sensitive parameter; it was assigned to be 110, 18, 15 and 14 mm/h for the first, the second, the third, and the fourth layers in Rains soil type (major soil type) and 130, 7.3, 3.8 and 6.3 for same layers in Meggett soil (second major soil). Parameter values ranged from 6.3 mm/h (Montevallo soil type) to 1050 mm/h (Leon soil type) in the first layer. We tested whether disregarding this parameter could influence the likelihood function during dry (summer) and wet (winter) seasons. The results indicated that this parameter alone could result in simulated streamflow errors of approximately 4% during wet seasons, and more than 17% during dry seasons. Our test further revealed that soil heterogeneity has a strong influence on low-flow estimates and almost negligible influence during high flow. Their influences might be stronger when there is less contribution from shallow aquifers, such as when distances from the water table to the land surface increase during summer.

The effective hydraulic conductivity in tributary channels was the third most sensitive parameter; it controls

surface runoff transmission losses to the main channel. This parameter is most effective in groundwater discharge/recharge areas. In the case of the Waccamaw River, this parameter is sensitive because of effluent channels in riparian wetland areas. The groundwater flow enters the river system through the side and bottom of the streambed and makes this parameter equal to zero in the perennial streams where groundwater continuously discharges to the river system. We tested this by delineating the Waccamaw watershed at the Freeland and Long's outlets individually, then conducting uncertainty analysis. The hydraulic conductivity parameter was greater than zero for the first order channel. In particular, the position of the water table to the ground surface and its depth significantly influences this value. This parameter, like other hydraulic properties, is an averaged value and its quantity is a key element in SWAT simulation. Optimized value of CH_K was equal to 4.33 for the Waccamaw River.

3.2.2 Posterior parameter histogram

Posterior density responses of the model parameters were visually investigated and the marginal posterior densities were constructed using ten parameters with high to moderate sensitivity (Fig. 4). The x-axis in each graph is fixed to the prior range of each individual parameter to facilitate comparison of sensitive parameters. In addition, most of high sensitive parameters showed more peaked densities in a narrower range of parameter values (see CH_N2 and SOL_K) while some moderately sensitive parameters have flat response surfaces (e.g., SLSUBBSN). The variations of the model parameter values significantly reflect the sensitivity of the associated parameters. For instance, Manning's n -value for the main channel showed a high variation. A high density of GWHT also represents slower groundwater flow recession and indicated that groundwater contribution to streamflow continued for a longer duration.

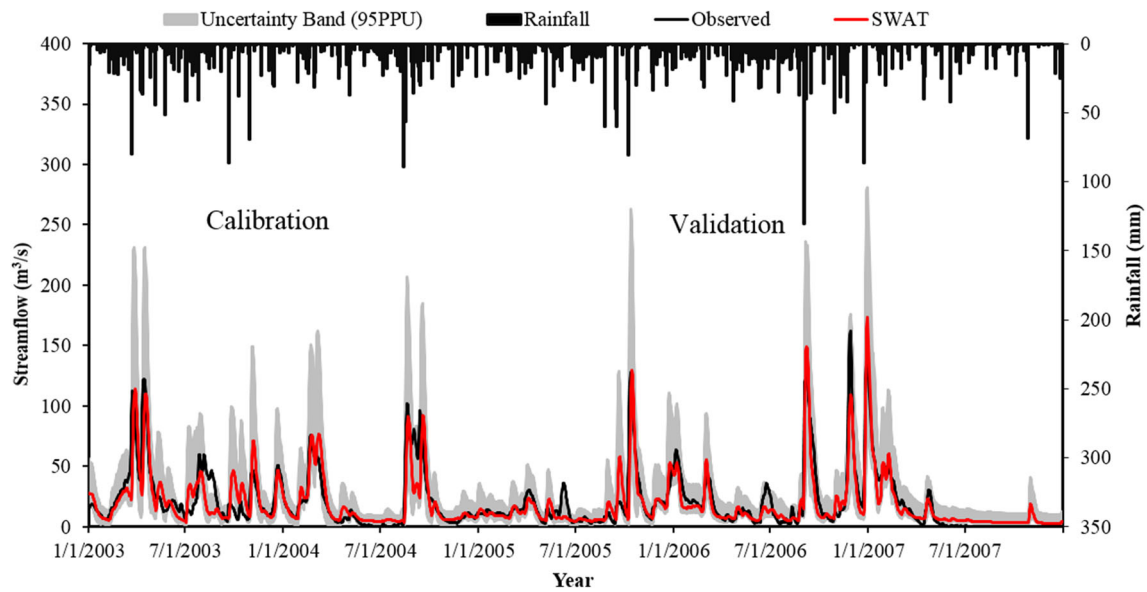


Fig. 5 Daily simulations during calibration and validation periods at the Freeland station. This outlet showed less sensitivity to river hydraulic and soil properties in the SWAT model

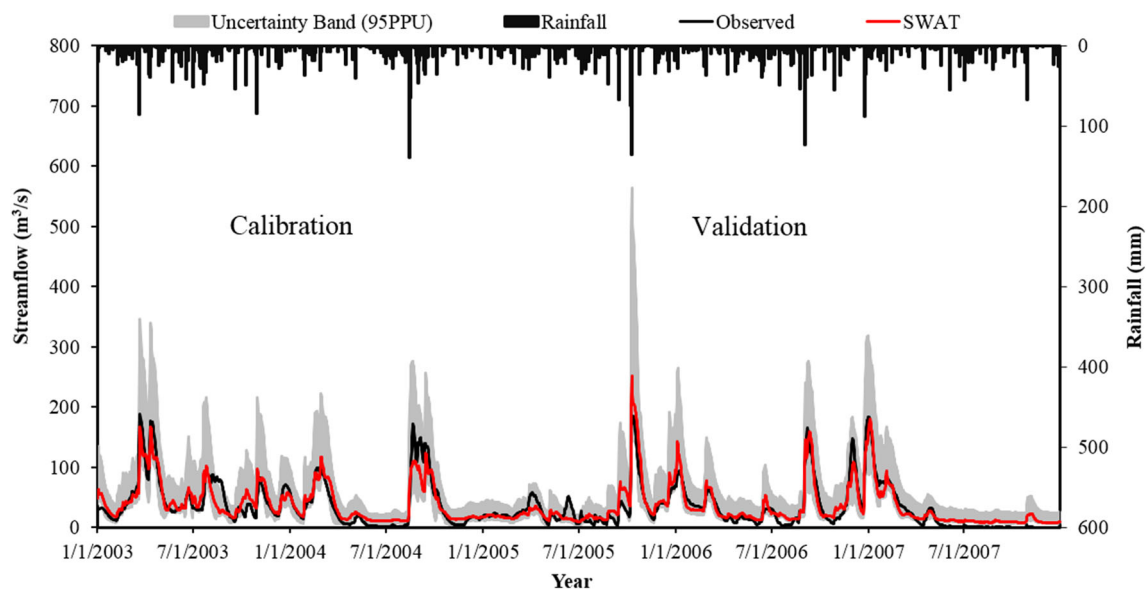


Fig. 6 Daily simulations during calibration and validation periods at the Longs station. At this location, SWAT almost overestimated low flow events while underestimated peak flows during moderate

hydrological conditions (i.e., 2005) and overestimated during very wet conditions (i.e., 2006)

In addition, SHALLST was not distributed uniformly across the calibration runs. This indicates that this parameter plays a significant role in describing subsurface flow contribution especially in dry periods. One possible explanation is that groundwater/shallow water contribution to streamflow during dry seasons is influentially high and lasts for a long time following intermittent rainfall periods. While SOL_BD illustrated higher densities in narrower

range of parameter values with a normal shape, but both ESCO and EPCO parameters further showed higher densities in a narrower range of parameter values with a skew to the right during simulation runs, reflecting the fact that the actual values of these two parameters are high and close to SWAT's maximum absolute range. This parameter density is highly correlated with groundwater and other soil properties.

Table 4 Daily streamflow calibration and validation performances

Objective function	Freeland		Longs	
	Calibration	Validation	Calibration	Validation
<i>NSE</i>	0.79	0.87	0.77	0.90
<i>R</i> ²	0.79	0.87	0.77	0.91
<i>P</i> -factor	90%	61%	75%	52%
<i>R</i> -factor	0.87	0.69	0.79	0.72
<i>Br</i> ²	0.68	0.72	0.59	0.77
<i>SSQR</i>	4.55	32.19	61.43	48.32

The SUFI2 analysis shows that uncertainty in model parameterization is significant. Further, the parameter posterior densities reflect that SUFI2 was well calibrated using high to moderate sensitive parameters with relatively tight uncertainty bounds. On the contrary, using a Bayesian approach resulted in parameter distributions that almost cover the entire prior defined hypercube of the individual parameters.

3.3 Calibration and validation results

Both calibration and validation modeling showed that SWAT simulated streamflow accurately at both outlets (Figs. 5, 6). While in some intervals, observed values are extrapolated out of the 95PPU bound, the 95PPU is quite suitable to bracket the observations in 2003, 2004, and 2006, while it is slightly overestimated in 2005 and 2007 due to heterogeneity of soil and hydraulic properties (see Bales and Pope 2001). Our analysis also demonstrated that SWAT more skillfully simulated streamflow during validation period compared to calibration (Table 4). One reason for the different results of calibration and validation simulations is the fact that SWAT underperforms in low flow conditions.

Specifically, the distribution of the uncertainty bound was wide during extreme events in both outlets, suggesting that parameter error produces high uncertainty and outliers during extreme events and that SWAT may have less skill in capturing high and low flow events. For example, daily streamflow of July 2003 in the Longs outlet was overestimated during a large storm event on July 18–20 (≈ 115 mm). By contrast, following repeated rainfall events on August 9–26, 2003, SWAT underestimated peak flow. In many cases, a large negative error (overestimation) is immediately followed by a large positive error (underestimation; see Hantush and Kalin 2008). This may result because SWAT has a simplified algorithm for disaggregating daily to hourly rainfall. For example, when a large rainfall event occurs at the end of a day, actual watershed response will occur a day later than simulated (Hantush and Kalin 2008; Samadi et al. 2017).

While the model reproduces most minor and major flow events, its reliability is diminished with very low and high flows. Over- and under-estimates of extreme flows may also result from the use of an average value for Manning's *n*-value, and sensitivity to soil properties. Despite these shortcomings, SWAT calibration results would be categorized as good to very good in Moriasi et al.'s. (2007) qualitative ranking scheme. Discussion of the results in the following sections will focus on fitting different probability distributions to modeling residuals and drawing conclusions about the best-fit model for error residuals at upstream and downstream outlets.

3.4 Evaluating error residuals

In this study, 21 probability distributions were fitted to residual error data (observed streamflow minus simulated streamflow) at both outlets and the residual error distribution parameters were computed. We fitted different distributions such as two-parametric distributions widely applied in streamflow time series analysis, i.e., Gumbel, Gamma, Lognormal, and Weibull (see El Adlouni et al. 2008), along with heavy-tailed (e.g., non-Gaussian) probability functions. We used both AIC and SBC Bayesian statistics to restrict the range of suitable μ and σ values of each probability distribution. We then selected a final value yielding a good visual fit to the streamflow time series with minimum effective DF. Both parametric and nonparametric smoothing terms were also examined using the additive smoothing functions in GAMLSS. Non-parametric smoothing additive terms, i.e., loess curve, showed the best fit to the residual error data at both streamflow gauges. In other words, the convergence criterion (tolerance level) for the backfitting algorithm at both Freeland and Longs error models occurred after only few hundreds runs when a polynomial loess curve was used as an appropriate smoothing term to the model.

At the Freeland outlet, the residual errors follow the four-parameter Box–Cox *t* (BCT) distribution (or a generalization of the Box–Cox Normal distribution; see Appendix A). Model error at this gauge had a mean of

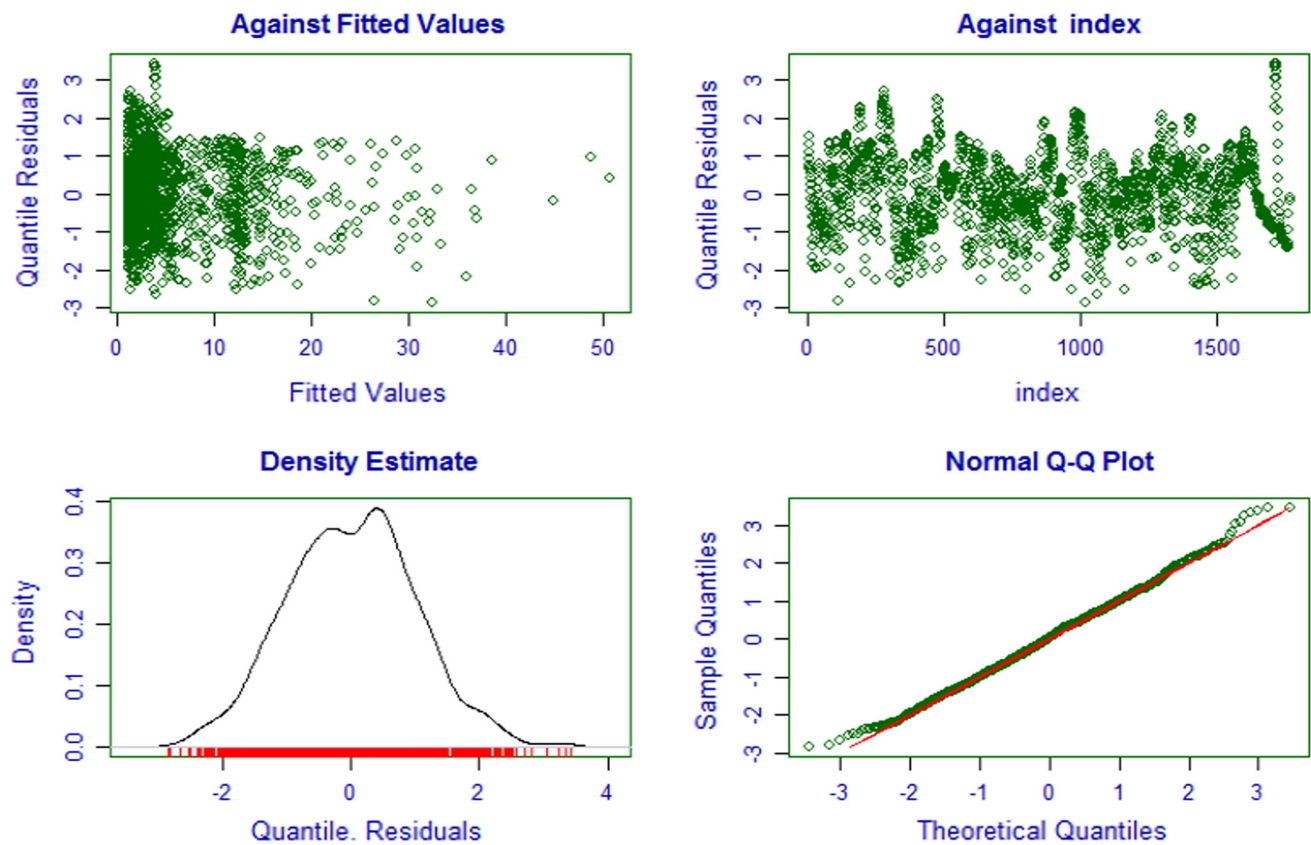


Fig. 7 Residual plot of Freeland outlet from BCT fitted model. The quantiles of the BCT distribution (theoretical) and error distributions agree well. The density plot exhibits multiple modes of varying error magnitudes

Table 5 Summary of the statistics of the distribution parameters for the residual errors in the Freeland and Longs outlets, respectively

Residuals	Freeland	Longs
Mean	0.0123	− 0.00091
Variance	1.001	1.002
Skewness coefficient	0.092	0.025
Kurtosis coefficient	2.95	2.95
Filliben correlation coefficient	0.99	0.99

0.012, a variance of 1.001, a coefficient of skewness of 0.092, a coefficient of kurtosis of 2.95, and a Filliben correlation coefficient of 0.99 (see Fig. 7; Table 5). The fitted model for μ at this outlet indicates that the variability of error mean decreases to near zero at about 10 cms then steadily increases at higher flow rates (see Fig. 9). Further, the fitted centile-based coefficient of variation of error data (σ) increases rapidly with increasing streamflow until about 15 cms, decreases slowly until about 30–40 cms, then increases afterward. The randomized quantile residuals of this outlet behaved well (Fig. 7), e.g., the median is nearly zero and the variance nearly one,

hence the residuals showed the best fit with the BCT distribution family. In addition, the corresponding kernel density plots in Fig. 7 adopted a non-parametric polynomial loess curve additive smoother, which centered most of error mass on zero with a significant correlation. At the Freeland outlet, the amount of probability mass is distributed on both tails of the PDF with a bell shaped multi-peak density plot. The shape parameters ν and τ are also bounded (almost fixed values, not shown here) in the BCT distribution, providing robustness to outlying error values. A summary of the fitted parameters is exhibited in Table 5.

A skew t type 3 distribution (see Appendix A) yielded improved estimates of parameters and significantly normalized the error time series at Longs gauge. A non-parametric analysis of the error probability density function indicates that a polynomial loess curve, as an appropriate additive term, showed the best fit to Longs residual time series data. In addition, a skew t type 3 distribution fitted the Longs error dataset well with a mean value near to zero (− 0.00091), a standard deviation or variance of 1.002, coefficients of skewness and kurtosis of 0.025 and 2.95, and Filliben correlation coefficient of 0.99 (see Table 5). Arguably, errors at this outlet have slightly heavier tails and were peaked compared to Freeland model (see Fig. 8).

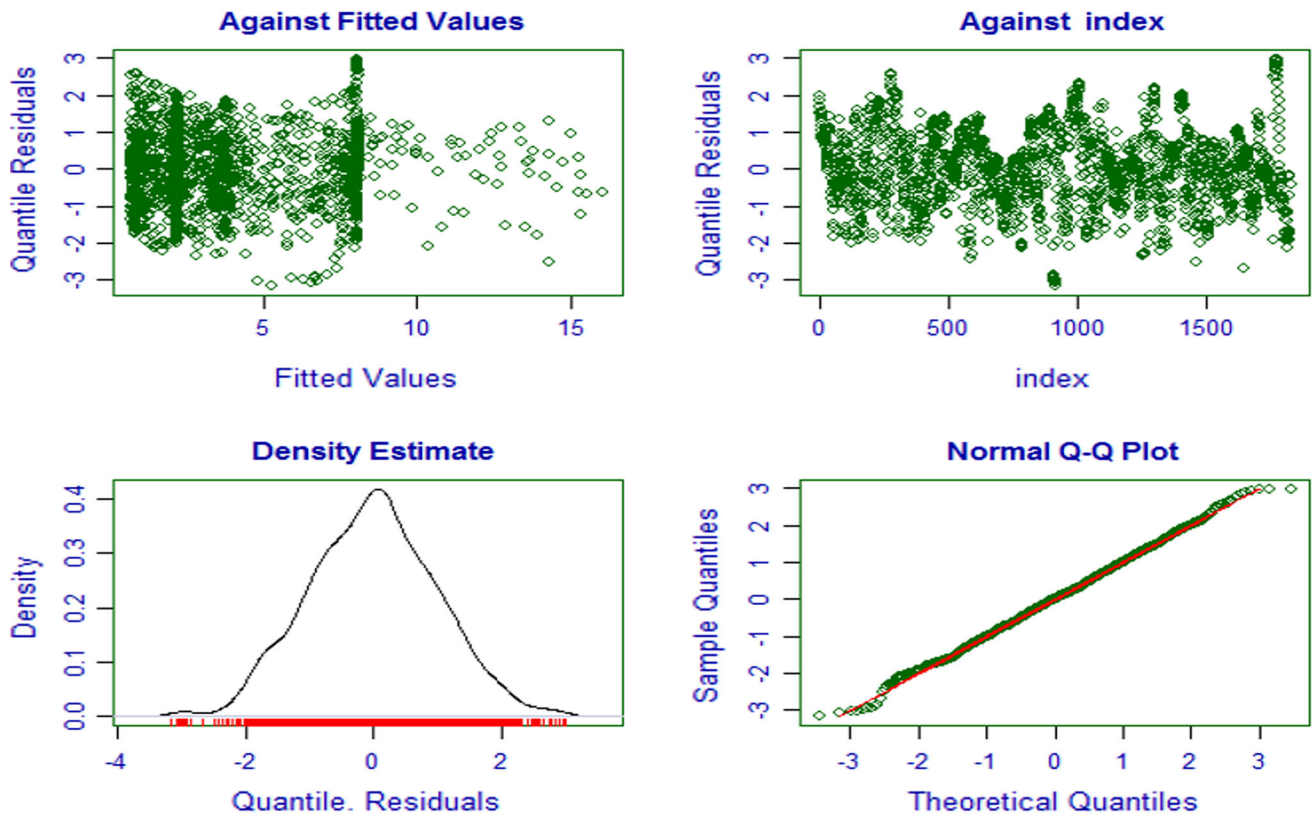


Fig. 8 Residual plot of Longs outlet from the skew t type 3 fitted models. The residuals show a symmetrical pattern and have a constant spread throughout the range. Error frequency is also in agreement with streamflow magnitudes (see quantile residual plot)

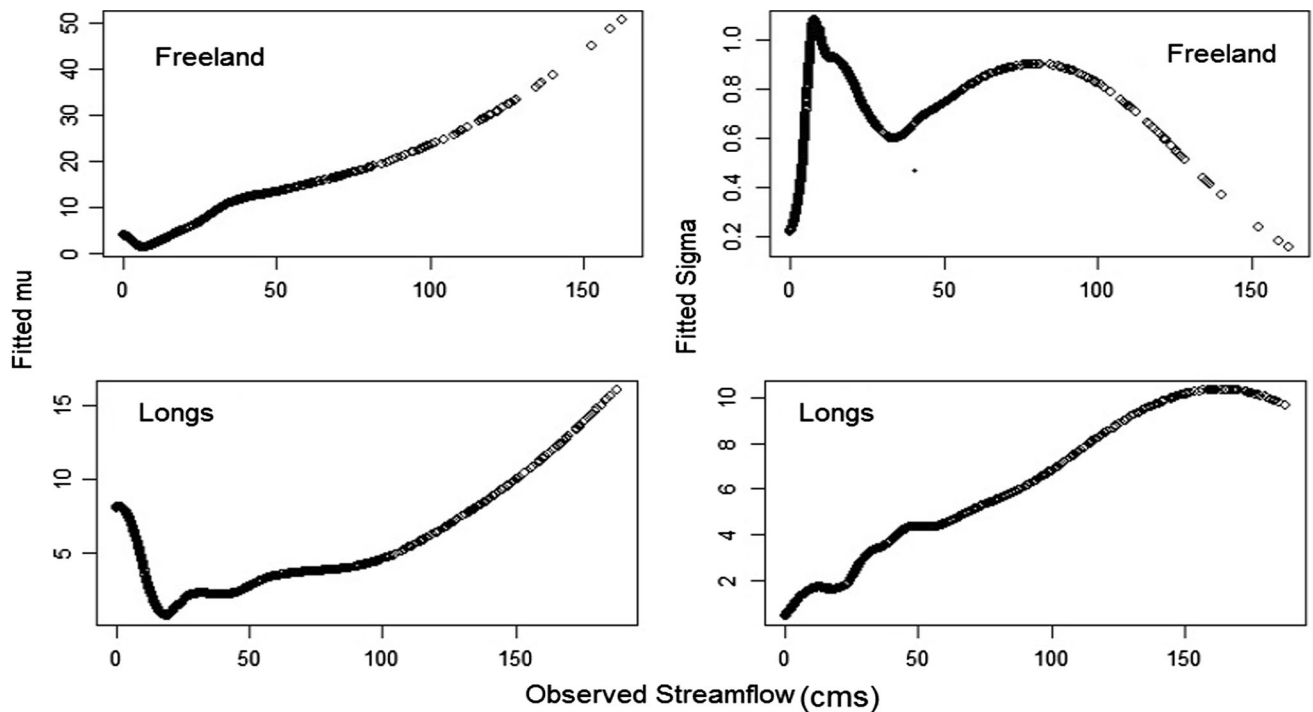


Fig. 9 The fitted error parameters (μ and σ) against observed streamflow for BCT (Freeland) and skew t type 3 (Longs) distributions. The location parameter (μ) and the scale parameter (σ)

are the same as the mean and the standard deviation of the fitted probability distribution of the random variate, respectively

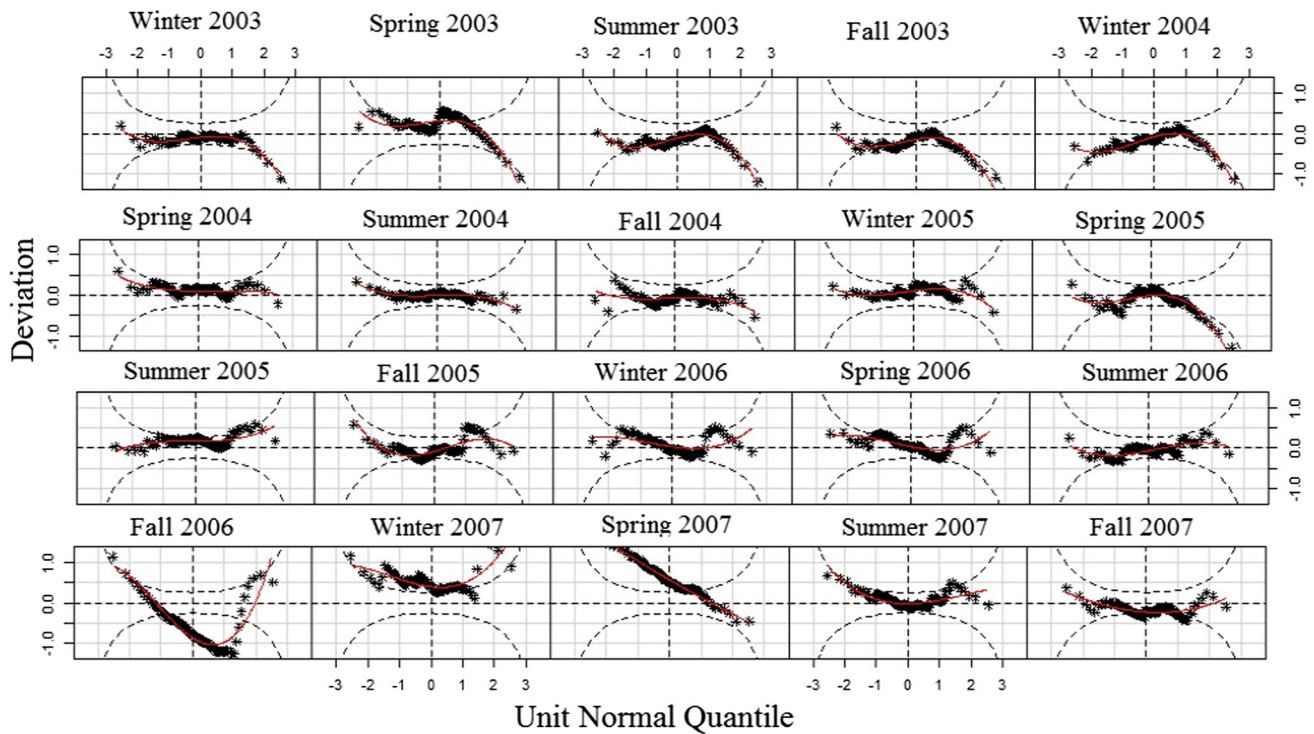


Fig. 10 Worm plots from the BCT distribution for the Freeland outlet. For a satisfactory fit, the data points should be within the two dash grey lines (95% confidence interval). The worm plots have a polynomial shape indicating that kurtosis has been accounted in fitting this model

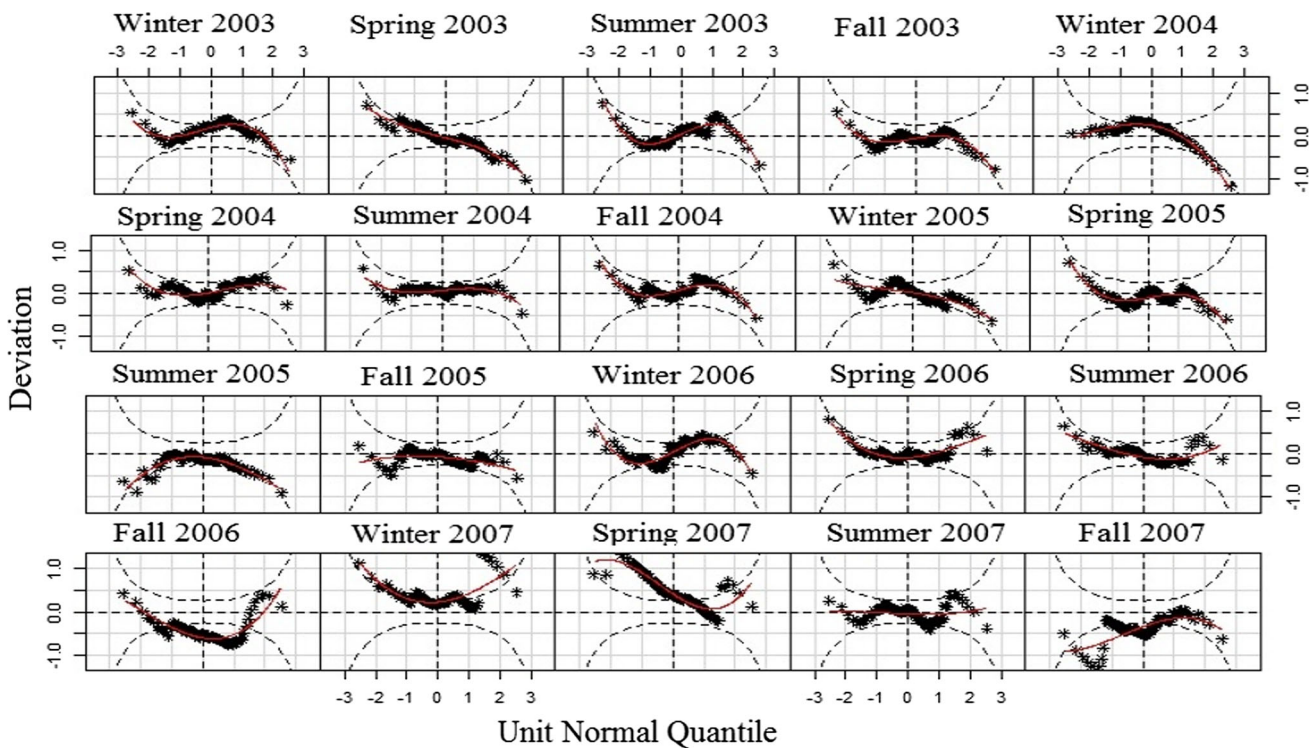


Fig. 11 Worm plots from the skew t type 3 distribution for the Longs outlet. For a satisfactory fit, the data points should be within the two dash grey lines (95% confidence interval)

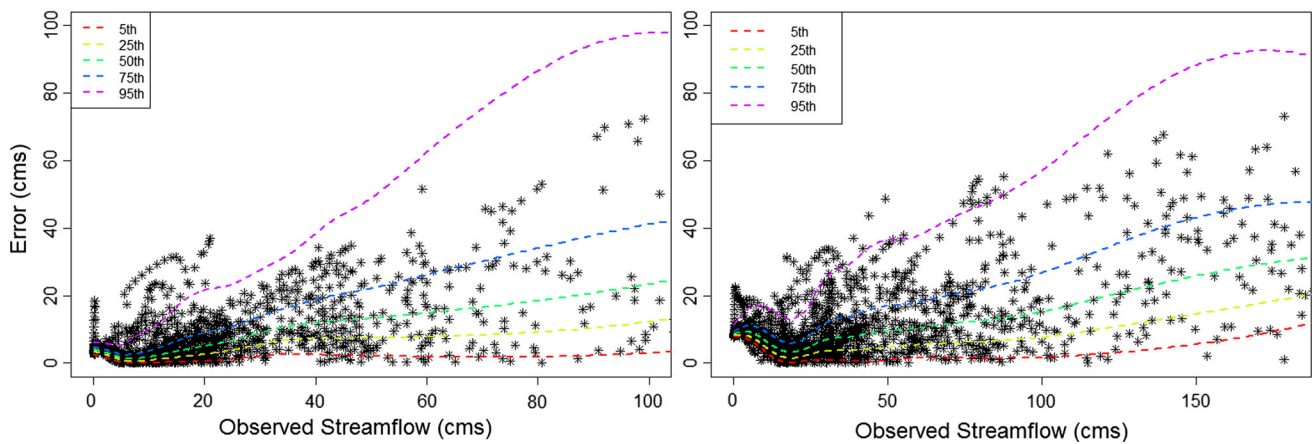


Fig. 12 Percentile curves using the Box–Cox t and the skew t type 3 distributions for the Freeland (left) and Longs (right) outlets. Both distributions showed best fits to non-parametric loess additive term smoothing function

The fitted model for μ also indicates a sharp decline of mean values during low flow events ($\sim <10$ cms) with significant variations around moderate flow (30–60 cms) and steadily increasing at high flow (see Fig. 9). The fitted σ value (error variance) steadily increases with streamflow discharge while greater deviation occurred during low to moderate flow events. Both median and variance of errors increase gradually with streamflow discharge, proposing heteroscedasticity and significant autocorrelation. The fitted model τ at this outlet also specifies the possibility of a leptokurtic distribution (too tall, $\tau \approx 3$). In addition, a high estimated kurtosis ($\tau = 2.95$) at both outlets reflects that values close to the mean/median are relatively more frequent and extreme values are also comparatively more persistent. The values in between are relatively less frequent. Therefore, the PDF plot is too tall in longs and too sharp in Freeland with long and fat tails.

To examine which range of the time series the model does not fit well, and to identify model violation in each season, we used an equal number of observations to produce worm plots of each fitted distribution. As is clear from Figs. 10 and 11, SWAT modeling showed a skillful forecast during moderately wet years while exhibited less skill during the 2007 dry period. For instance, most of the deviations of 2003, 2004 and 2005 are located in the 95% confidence interval inside the two elliptic curves, supporting the argument that SUFI2 skillfully predicted observed streamflow in these intervals. In addition, modeling results during summer 2004 showed the best fit at both outlets by bracketing most of deviations in the confidence interval. The error data showed more deviation during the drought period (low flow) from fall 2006 through spring 2007. The red curves in the plots are the fitted polynomial loess curves to the points (errors) at both streamflow gauges.

Computed area-average model bias (percentile) at the 5th, 25th, 50th, 75th, and 95th percentiles were driven according to the BCT and skew t type 3 distribution functions and fitted to the residual errors and observed time series (see Fig. 12). Two important findings can be deduced from these plots. First, some extreme outliers (0.1% of the data) in the upper and lower tails of the distributions are positioned outside the 95th percentile. This is most notable in the case of Longs, where simulation tends to expose larger bias. Although, the upper percentiles account for an extremely small portion of error (while having a large effect on daily streamflow simulation), their impact on model performances is significant. This error might relate to very low and/or very high streamflow events where SWAT less skillfully reproduces their dynamics (both magnitudes and amount).

The second finding in Fig. 12 is that the error model at the Freeland outlet tends to increase at the largest percentiles (95th). This is especially noticeable in error distribution between the 75th to 95th percentiles, where values are highly biased. Error values in this range might be related to extreme streamflow simulation (overestimation/underestimation of SWAT) and might be due to a combination of measurement, model input, and model structural errors (see Schoups and Vrugt 2010). For example, considering an average value for Manning's n -value or releasing water from the shallow aquifer to the reach based on the base flow alpha factor (ALPHA_BF) alone in the SWAT model, may cause overestimation/underestimation of streamflow and, consequently, heteroscedasticity and correlation in the residual errors. This result also suggests that SWAT has limited skill to compute the timing of daily high flow and has less skill with respect to the magnitude, especially during dry hydrological conditions. Together with posterior parameter histograms presented in Fig. 4, it

Table 6 Different distributions used for fitting error time series

Distribution	Freeland			Longs		
	GAIC	SBC	Rank	GAIC	SBC	Rank
Box–Cox t	9019.53	9133.74	1	10884.9	10995.8	8
Normal	10630.9	10735.7	2	10213.3	11905.9	19
Log normal	9257	9364.5	3	11096.3	11198.6	17
Gamma	9183.6	9290.9	4	10917.5	11019.8	12
Weibull	9282.3	9390.2	5	11106.1	11208.5	20
t family	9657.1	9765.7	6	11208.7	11314.2	6
Skew t type 3	9120.8	9233.576	7	10718.38	10828.8	1
Skew t type 1	9115.2	10844	8	10783.5	10844	9
Skew t type 2	9104.7	10831	9	10761.1	10891.5	4
Skew t type 4	9299.5	10973.9	10	10863.4	10973.9	2
Box–Cox Power Exp.	10570.72	10636.46	11	12330.92	12397.04	15
Logistic	11135.07	11189.85	12	12549.1	12604.2	11
Gumbel	11905	12010.9	13	12823.6	12924.21	18
Box–Cox cole and green	10692.82	10753.07	14	12450.19	12510.8	13
Exponential Gaussian	11905.05	12078.22	15	11284.6	12434.73	3
Logistic	10106.9	11595.4	16	11495.26	11595.45	14
Exponential	9685.9	9739	17	11934.9	11986.1	7
Beta	9682.3	9888.38	18	11864.7	11993.2	21
Inverse Gaussian	10228.2	10335.87	19	11796.7	11898.7	16
Reverse Gumbel	9544	9647.7	20	10975.9	11076	10
Power Exp.	10904.65	10964.91	21	12518.39	12578.99	5

The distributions with the lowest GAIC and SBC values are the best fit (bold values)

appears that uncertainty model puts less emphasis on fitting extreme flows.

Finally, this study investigated whether skewness and kurtosis of error data were due to the effect of extreme outliers. To explore this approach, the most extreme outliers (0.1% of the data) in the upper (95th and 99th) and lower (5th) tails of the distribution were removed from the error time series (193 values from Freeland and 207 values from Longs error datasets) and the models for each of 21 probability distributions in Table 6 were refitted to error datasets at both outlets. The BCT and the skew t type 3 distributions still provided the best fits according to Bayesian statistic; skewness and kurtosis were reduced, indicating that most errors were generated by extreme events. Their fit to the data was substantially improved, leading to improved centile estimates (results not shown here).

4 Discussion and conclusions

This study presents daily streamflow simulation and analysis of a coastal plain watershed in the southeastern United States. The modeling strategy relies on a two-step procedure—uncertainty assessment and residual error analysis.

Sensitivity and uncertainty analysis were performed using SUFI2 that computed parameter uncertainty associated with streamflow prediction of the SWAT model. Sensitivity analysis revealed that Manning's roughness coefficient, transmission loss, and the resistance of the soil matrix to water flow significantly increase parameter uncertainty associated with streamflow calibration. Sensitivity associated with these parameters induces relatively large uncertainty and errors during simulations of both high and low flow events in the SWAT model. This finding, made possible through analysis of the residual errors with the GAMLSS approach, indicated that error variance increases gradually during extreme events. Bayes law suffers from interaction between individual error sources (Vrugt et al. 2008), which leads to structured, non-stationary residuals (Beven 2008; Vrugt et al. 2009a; Pourreza-Bilondi et al. 2016a, b); therefore, results should be carefully interpreted.

In addition, the fitted probability distribution with non-parametric additive terms in the GAMLSS model helped to distinguish differences in error distribution parameters between upstream and downstream gauges. The fitted model for the location parameter (μ) indicates that the mean/median of the error model increases rapidly with increasing streamflow at an upstream outlet. At a downstream outlet, it decreases sharply during low flow events

and increases steadily thereafter, indicating that low flow events more strongly influence residual error probability distribution at the downstream outlet. In addition, the fitted centile-based coefficient of variation (*sigma*) showed a sharp increase at the upstream gauge whereas persistently skewed and more peaked at the downstream gauge.

Overall, our analysis reveals that residual errors in the watershed are complex and their distributions vary within upstream and downstream gauges. These errors may be caused by a combination of measurement errors, rainfall amount, and SWAT structural errors. For example, it could be hypothesized that runoff generation in this watershed is dominated by the saturation-excess mechanism. This would suggest that the runoff coefficient is largely independent of rainfall intensity. Rather, total rainfall and landscape characteristics such as local topography, upland-watershed area, and soil properties (i.e., available water storage capacity) are the important factors determining whether a particular area in this watershed will generate runoff. As rainfall continues, the saturated area grows in extent and increases the area generating runoff. Thus, the position and the extent of runoff generation areas may vary with each rainfall amount.

Some common locations where saturation occurs in this coastal plain watershed might be the areas with shallow soil and very gentle and flat topographic slope at the downstream gauge (Long's outlet). At these locations, a reduction in hydraulic transmissivity or hydraulic gradient restricts interflow capacity and saturates the soil (this might be the reason for sensitivity of hydraulic and soil properties). Hence, a switch to separate simulation of dry and wet periods, and/or individual simulation of each outlet, may be required to adequately predict streamflow behavior and dynamics in this complex landscape using SWAT. The use of higher vertical resolution topographic data, such as digital elevation models derived from Lidar (Light Detection and Ranging) data, also may improve results. Our initial attempts have been only moderately successful. This is because not only do wet and dry period models introduce different parameter sensitivities, but also because the interflow (Darcy flow) capacity cannot be adequately formulated by SWAT and similar models.

Generally, in the coastal plain watershed with persistent shallow aquifer contribution to streamflow, application of the appropriate Darcian method is more complex. With knowledge of saturated hydraulic conductivity and the gradient of matric pressure, Darcy's law straightforwardly calculates the flux density at a given time and position, and its downward component is interpretable as recharge (see Nimmo et al. 2005 for more details). Although, some SWAT groundwater parameters such as SHALLST represent temporal recharge averages, if SWAT calculates the long-term average recharge accurately, a detailed picture of

the behavior of groundwater can be computed in the unsaturated zone (or shallow aquifer system). Here the SUFI2-GAMLSS approach reveals some of these deficiencies, suggesting different error PDFs for upstream and downstream outlets, and greater errors during the period of shallow aquifer contribution (low flow) and flood events. This is one aspect of the broader issue of properly dealing with interactions between groundwater and surface water in locations where the fluxes between them are a significant part of the water budget accounting and streamflow dynamics. Various investigators have addressed this by coupling SWAT with groundwater models (e.g., Guzman et al. 2015). That effort was beyond the scope of this research.

In this study, error modelling reveals a smooth transition from a BCT distribution upstream to a skew *t* type 3 and too tall distribution downstream. The latter distribution also has heavier tails, accounting for the probability of larger errors. This induces robustness against outliers and random variations in low and high flow events as seen elsewhere (e.g., Schoups and Vrugt 2010; Tian et al. 2014; Tongal and Booij 2017). Overall, the fitted models show that extreme flow events more strongly influence residual error probability distribution and error variance increases with streamflow discharge, indicating correlation and heteroscedasticity of residual errors.

The modeling strategy proposed here differs from previous methods that typically have relied on a single known PDF model assumption for the entire watershed system (see Vrugt et al. 2009b; Schoups and Vrugt 2010). Although, use of a skewed distribution yielded good estimates of Bayesian statistics at the downstream location, we stress that the driven probability distribution can change case by case depending on how well the Bayesian algorithm calibrates different aspects of hydrograph (rising limb, peak flow, falling limb, time to peak, recession time, etc.). Error analysis helps us to identify specific reasons for model weaknesses and enables establishing alternative hypotheses of watershed behavior over time. We stress that the Generalized Additive Models described here led to a very efficient description of residual error behavior and that the SUFI2-GAMLSS approach showed flexibility, interpretability of parameters, and robustness in describing outliers. It is suggested that GAMLSS models enter into the toolbox of hydrologists in the future in order to improve theoretical and applied aspects of hydrological simulation.

Finally, the methodology presented here provides increased flexibility to describe residual errors behavior in rainfall-runoff applications using a Bayesian uncertainty approach. This flexibility can translate into improved estimates of parameter and total predictive uncertainty by coupling the generalized additive models with any Bayesian algorithm, introducing more reliable assumption of

residual error distribution. While the fitted distribution parameters (μ , σ , ν , and τ) have very clear influences on the error model, dependence between these parameters is equally important for Bayesian model prior assumptions. Moreover, if error standard deviation increases as both linear and nonlinear functions of streamflow, this information can be used when specifying a prior PDF. Since kurtosis describes the probability of tails and outlier, setting the Bayesian algorithm based on reaching the kurtosis parameter to convergence criteria is another approach, which can be used for future algorithmic development. These assumptions may simplify the process of choosing an appropriate prior error distribution in a Bayesian framework. Although our approach focused on streamflow error analysis, the methodology is general enough to examine the complexities of residual errors in other hydrology applications such as rainfall modeling.

Acknowledgements This research was partially supported by the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office (Grant # NA11OAR4310148) to the Carolinas Integrated Sciences and Assessments. The data and related code are available upon a request to the first author. The analyses were performed in R (R Development Core Team, 2013) by using the contributed package GAMLSS and other add-on packages. The authors and maintainers of this software are gratefully acknowledged.

Appendix A

Box-Cox t distribution

Let Y be a positive random variable (here observed streamflow time series) having a BCT distribution, denoted by BCT (μ, σ, ν, τ), defined through the transformed random variable Z (Eq. A1), which is given by (Rigby and Stasinopoulos 2005a, b)

$$Z = \begin{cases} \frac{1}{\sigma\nu} \left[\left(\frac{Y}{\mu} \right)^\nu - 1 \right], & \text{if } \nu \neq 0 \\ \frac{1}{\sigma} \log \left(\frac{Y}{\mu} \right), & \text{if } \nu = 0 \end{cases} \quad (\text{A1})$$

If $Y > 0$, where $\mu > 0$ and $\sigma > 0$, the random variable Z is then assumed to follow a t distribution with degrees of freedom, $\tau > 0$, treated as a continuous parameter. From the probability density function of Y , a BCT (μ, σ, ν, τ) random variable, is given by

$$f_{Y(y)} = f_{Z(z)} \left| \frac{dz}{dy} \right| = \frac{y^{\nu-1}}{\mu^\nu \sigma} f_{Z(z)} \quad (\text{A2})$$

where $f_{Z(z)}$ is the exact (truncated t) probability density function of Z .

Kurtosis parameter τ takes on values between -3 and $+3$ and determines the peakedness of the PDF, while

skewness (ν) parameters affects asymmetry ($\nu > 0$; Schoups and Vrugt 2010), as illustrated in Fig. 1. The density is symmetric if $\nu = 0$ and positively (negatively) skewed if $\nu > 1$ ($\nu < 1$).

Skew t type 3 distribution (ST3)

This is a “spliced-scale” distribution with PDF (see Fernandez et al. 1995; Rigby and Stasinopoulos 2005a, b), denoted by ST 3(μ, σ, ν, τ), defined by

$$f_{Y(y|\mu,\sigma,\nu,\tau)} = \frac{c}{\sigma} \left\{ 1 + \frac{z^2}{\tau} \left[v^2 I(y < \mu) + \frac{1}{v^2} I(y \geq \mu) \right] \right\} \quad (\text{A3})$$

For $-\infty < y < \infty$, where $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$, and $\tau > 0$, and where $z = (y - \mu)/\sigma$ and $c = 2\nu \left[\sigma(1 + \nu^2) B\left(\frac{1}{2}, \frac{\tau}{2}\right) \tau^{\frac{\tau}{2}} \right]$, Fernandez and Steel (1998).

Note that μ is the mode of Y . The mean and variance of Y are given by $E(Y) = \mu + \sigma E(Z)$ and $Var(Y) = \sigma^2 V(Z)$, where $E(z) = 2\tau^{\frac{1}{2}}(\nu^2 - 1)/[(\tau - 1)B(\frac{1}{2}, \frac{\tau}{2})\nu]$ and $E(z^2) = \tau(\nu^3 + \frac{1}{\nu^3})/[(\tau - 2)(\nu + \frac{1}{\nu})]$. See Fernandez and Steel (1998) for further information.

References

- Abbaspour KC (2015) User manual for SWAT-CUP, SWAT calibration and uncertainty analysis programs. Swiss Federal Institute of Aquatic Science and Technology, Eawag, Duebendorf, p 93
- Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Zobrist J, Srinivasan R (2007) Modelling hydrology and water quality in the pre-Alpine/Alpine Thur watershed using SWAT. J Hydrol 333:413–430
- Ajami NK, Duan Q, Sorooshian S (2007) An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour Res 43:W01403. <https://doi.org/10.1029/2005WR004745>
- Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Control 19(6):716–723
- Amatya KM, Jha MK (2011) Evaluating the SWAT model for a low-gradient forested watershed in Coastal South Carolina. Trans ASABE 54(6):2151–2163
- Arnold JG, Allen PM, Bernhardt G (1993) A comprehensive surface-groundwater flow model. J Hydrol 142:47–69
- ASCE (1993) Criteria for evaluation of watershed models. J Irrig Drain Eng 119(3):429–442
- Bales JD, Pope BF (2001) Identification of changes in streamflow characteristics. J Am Water Resour Assoc 37(1):91–104
- Bates BV, Campbell AEP (2001) Markov Chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resour Res 37(4):937–947
- Beven KJ (2008) On doing better hydrological science. Hydrol Process 22:3549–3553. <https://doi.org/10.1002/hyp.7108>
- Beven KJ, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems. J Hydrol 249:11–29

- Beven K, Smith PJ, Freer JE (2008) So just why would a modeler choose to be incoherent. *J Hydrol* 354:15–32
- Box GEP, Tiao GC (1992) Bayesian inference in statistical analysis. Wiley, New York, p 588
- Butts MB, Payne JT, Kristensen M, Madsen H (2004) An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J Hydrol* 298:242–266. <https://doi.org/10.1016/j.jhydrol.2004.03.042>
- Clark MP et al (2015) A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour Res* 51:2498–2514. <https://doi.org/10.1002/2015WR017198>
- Cole TJ, Green PJ (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 11:1305–1319
- Del Giudice D, Honti M, Scheidegger A, Albert C, Reichert P, Rieckermann J (2013) Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrol Earth Syst Sci* 17(2013):4209–4225
- Del Giudice D, Albert C, Rieckermann J, Reichert P (2016) Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation. *Water Resour Res* 52:3162–3186. <https://doi.org/10.1002/2015WR017871>
- Dunn PK, Smyth GK (1996) Randomised quantile residuals. *J Comput Gr Stat* 5:236–244
- Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Stat Sci* 11:89–121
- El Adlouni S, Bobée B, Ouarda TBMJ (2008) On the tails of extreme event distributions in hydrology. *J Hydrol* 355:16–33
- Etemadi H, Samadi S, Sharifikia M (2014) Uncertainty analysis of statistical downscaling techniques in an Arid region. *Clim Dyn* 42:2899–2920
- Etemadi H, Samadi S, Sharifikia M, Smoak JM (2015) Assessment of climate change downscaling and non-stationarity on the spatial pattern of a mangrove ecosystem in an arid coastal region of southern Iran. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-015-1552-5>
- Fernandez C, Steel MFJ (1998) On bayesian modelling of fat tails and skewness. *J Am Stat Assoc* 93:359–371
- Fernandez C, Osiewalski J, Steel MFJ (1995) Modeling and inference with v-spherical distributions. *J Am Stat Assoc* 90(432): 1331–1340
- Green PJ, Silverman BW (1994) Nonparametric regression and generalized linear models. Chapman and Hall, London
- Gupta HV, Sorooshian S, Yapo PO (1998) Toward improved calibration of hydrologic models: multiple and noncommensurate measures of information. *Water Resour Res* 34(4):751–763
- Guzman JA, Moriasi DN, Gowda PH, Steiner JL, Starks PJ, Arnold JG, Srinivasan R (2015) A model integration framework for linking SWAT and MODFLOW. *Environ Model Softw* 73:103–116
- Han JC, Huang GH, Zhang H et al (2014) Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of SLURP model applied to the Xiangxi River watershed, China. *Stoch Environ Res Risk Assess* 28:973. <https://doi.org/10.1007/s00477-013-0792-0>
- Hantush M, Kalin L (2008) Stochastic residual-error analysis for estimating hydrologic model predictive uncertainty. *J Hydrol Eng*. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:7\(585\)585-596](https://doi.org/10.1061/(ASCE)1084-0699(2008)13:7(585)585-596)
- Hargreaves GL, Hargreaves GH, Riley JP (1985) Agricultural benefits for Senegal River Basin. *J Irrig Drain E ASCE* 111:113–124
- Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall, London
- Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems. Elsevier, Amsterdam. <http://www.stats.uwo.ca/faculty/aim/1994Book/>
- Honti M, Stamm C, Reichert P (2013) Integrated uncertainty assessment of discharge predictions with a statistical error model. *Water Resour Res* 49(2013):4866–4884
- Joseph JF, Guillaume JHA (2013) Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environ Model Softw* 46:292–298. <https://doi.org/10.1016/j.envsoft.2013.03.012>
- Katz RW (2010) Statistics of extremes in climate change. *Clim Change* 100:71–76
- Kim T-W, Valdés JB (2005) Synthetic generation of hydrologic time series based on nonparametric random generation. *J Hydrol Eng* 105:395–404
- Kuczera G (1983) Improved parameter inference in catchment models, 1. Evaluating parameter uncertainty. *Water Resour Res* 19(5):1151–1162. <https://doi.org/10.1029/WR019i005p01151>
- Laloy E, Vrugt JA (2012) High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resour Res* 48:W01526. <https://doi.org/10.1029/2011WR010608>
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241
- McCuen R, Knight Z, Cutter A (2006) Evaluation of the Nash-Sutcliffe Efficiency Index. *J Hydrol Eng*. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:6\(597\)597-602](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597)597-602)
- McCullagh P, Nelder JA (1989) Generalized linear models, volume 37 of monographs on statistics and applied probability, 2nd edn. Chapman and Hall, London
- McKay MD, Beckman RJ, Conover WJ (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- McMillan H, Krueger T, Freer J (2012) Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol Process* 26:4078–4111
- Melching CS, Bauwens W (2001) Uncertainty in coupled nonpoint source and stream water-quality models. *J Water Resour Plann Manag* 127:403–413
- Monteith JL (1965) Evaporation and environment. In: *Proceedings of the 19th symposium of the society for experimental biology*. Cambridge University Press, New York, pp 205–233
- Moore C, Wöhling T, Doherty J (2010) Efficient regularization and uncertainty analysis using a global optimization methodology. *Water Resour Res* 46:W08527. <https://doi.org/10.1029/2009WR008627>
- Moriasi DN, Arnold JG, Van Liew MW, Binger RL, Harmel RD, Veith T (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50(3):885–900
- Mosaedi A, Zare Abyane H, Ghabaei Sough M, Zahra Samadi S (2015) Long-lead drought forecasting using equiprobability transformation function for reconnaissance drought index. *Water Resour Manag* 29:2451–2469
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models: part 1. A discussion of principles. *J Hydrol* 10(3):282–290
- Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2001) Soil and water assessment tool user’s manual, version 2000. Grassland, Soil and Water Research Laboratory, Agricultural Research Service, Blackland Research Center, Texas Agricultural Experiment Station
- Nimmo JR, Healy RW, Stonestrom DA (2005) Aquifer recharge. In: Anderson MG, Bear J (eds) *Encyclopedia of hydrological science: part 13. Groundwater*, vol 4. Wiley, Chichester, pp 2229–2246. <https://doi.org/10.1002/0470848944.hsa161a>

- Pourreza-Bilondi M, Samadi S (2016) Quantifying the uncertainty of semiarid runoff extremes using generalized likelihood uncertainty estimation. Special issues on water resources in arid areas. Arab J Geosci. <https://doi.org/10.1007/s12517-016-2650-0>
- Pourreza-Bilondi M, Samadi SZ, Akhoond-Ali AM, Ghahraman B (2016) On the assessment of reliability in semiarid convective flood modeling using bayesian framework. ASCE Hydrol Eng. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001482](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001482)
- Priestley CHB, Taylor RJ (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. Mon Weather Rev 100(2):81–92
- Rigby RA, Stasinopoulos DM (1996a) A semi-parametric additive model for variance heterogeneity. Statist Comput 6:57–65
- Rigby RA, Stasinopoulos DM (1996b) Mean and dispersion additive models. In: Hardle W, Schimek MG (eds) Statistical theory and computational aspects of smoothing. Physica, Heidelberg, pp 215–230
- Rigby RA, Stasinopoulos DM (2005a) Generalized additive models for location, scale and shape (with discussion). Appl Stat 54:507–554
- Rigby RA, Stasinopoulos DM (2005b) Generalized additive models for location, scale and shape. J R Stat Soc Ser C (Appl Stat) 54:507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Riggs SR, Ames DV, Brant DR, Sager ED (2000) The Waccamaw drainage system: geology and dynamics of a coastal wetland, Southeastern North Carolina. East Carolina University, Greenville, p 165
- Royston P, Altman DG (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). Appl Stat 43:429–467
- R Development Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. <http://www.R-project.org/>
- Sadegh M, Vrugt JA (2013) Approximate Bayesian computation in hydrologic modeling: equifinality of formal and informal approaches. Hydrol Earth Syst Sci Dis 10(4):p4739
- Samadi S (2016) Assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the Southeast United States. Hydrol Res. <https://doi.org/10.2166/nh.2016.034>
- Samadi S, Meadows EM (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 Res Appl. <https://doi.org/10.1002/rra.3127>
- Samadi S, Tufford DL, Carbone GJ (2017) Assessing parameter uncertainty of a semi-distributed hydrology model for a shallow aquifer dominated environmental system. J Am Water Resour Assoc (JAWRA) 1–22. <https://doi.org/10.1111/1752-1688.12596>
- Schoups G, Vrugt JA (2010) A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resour Res 46:W10531. <https://doi.org/10.1029/2009WR008933>
- Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
- Sénégas J, Wackernagel H, Rosenthal W et al (2001) Error covariance modeling in sequential data assimilation. Stoch Env Res Risk Assess 15:65. <https://doi.org/10.1007/PL00009788>
- Serinaldi F (2011) Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. Energy Econ 33(6):1216–1226
- Serinaldi F, Cuomo G (2011) Characterizing impulsive wave-in-deck loads on coastal bridges by probabilistic models of impact maxima and rise times. Coast Eng 58(9):908–926
- Serinaldi F, Kilsby CG (2015) Stationarity is undead: uncertainty dominates the distribution of extremes. Adv Wat Resour 77:17–36
- Sevat E, Dezetter A (1991) Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area. Hydrol Sci J 36(4):307–330
- Shrestha B, Cochran TA, Caruso BS, Arias ME, Piman T (2016) Uncertainty in flow and sediment projections due to future climate scenarios for the 3S Rivers in the Mekong Basin. J Hydrol 540:1088–1104. <https://doi.org/10.1016/j.jhydrol.2016.07.019>
- Sikorska AE, Scheidegger A, Banasik K, Rieckermann J (2012) Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. Hydrol Earth Syst Sci 16:1221–1236. <https://doi.org/10.5194/hess-16-1221-2012>
- Sivapalan M (2009) The secret to ‘doing better hydrological science’: change the question! Hydrol Process 23:1391–1396. <https://doi.org/10.1002/hyp.7242>
- Slater AG, Clark MP (2006) Snow data assimilation via an ensemble Kalman filter. J Hydrometeorol 7(3):478–493
- Sorooshian S, Dracup JA (1980) Stochastic parameter estimation procedures for hydrologic rainfall-runoff models—correlated and heteroscedastic error cases. Water Resour Res 16(2):430–442. <https://doi.org/10.1029/WR016i002p00430>
- Stasinopoulos DM, Rigby RA (2007) Generalized additive models for location scale and shape (GAMLSS) in R. J Stat Softw 23:1–46
- Stasinopoulos DM, Rigby RA (2016) Package ‘gamlss.dist’. <https://cran.r-project.org/web/packages/gamlss.dist/index.html>
- Tian Y, Booi MJ, Xu YP (2014) Uncertainty in high and low flows due to model structure and parameter errors. Stoch Environ Res Risk Assess 28:319. <https://doi.org/10.1007/s00477-013-0751-9>
- Tongal H, Booi MJ (2017) Quantification of parametric uncertainty of ANN models with GLUE method for different streamflow dynamics. Stoch Environ Res Risk Assess 31:993. <https://doi.org/10.1007/s00477-017-1408-x>
- USDA-SCS (United States Department of Agriculture–Soil Conservation Service) (1972) National engineering handbook, Section 4 Hydrology, Chapter 4–10, USDA-SCS, Washington
- Van Buuren S, Fredriks M (2001) Worm plot: a simple diagnostic device for modelling growth reference curves. Stat Med 20:1259–1277
- Villarini G, Smith JA, Serinaldi F, Bales J, Bates PD, Krajewski WF (2009) Flood frequency analysis for nonstationary annual peak records in an urban drainage area. Adv Water Resour 32:1255–1266
- Vrugt JA, ter Braak CJF, Clark MP, Hyman JM, Robinson BA (2008) Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour Res 44:W00B09. <https://doi.org/10.1029/2007WR006720>
- Vrugt JA, ter Braak CJF, Gupta HV, Robinson BA (2009a) Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch Environ Res Risk Assess 23:1011. <https://doi.org/10.1007/s00477-008-0274-y>
- Vrugt JA, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D (2009b) Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 10(3):271e288
- Wagener T, Gupta HV (2005) Model identification for hydrological forecasting under uncertainty. Stoch Environ Res Risk Assess 19:378. <https://doi.org/10.1007/s00477-005-0006-5>
- Wagener T, Sivapalan M, Troch P, Woods R (2007) Catchment classification and hydrologic similarity. Geogr Compass 1:901–931

- Wang G, Barber ME, Chen S et al (2014) SWAT modeling with uncertainty and cluster analyses of tillage impacts on hydrological processes. *Stoch Environ Res Risk Assess* 28:225. <https://doi.org/10.1007/s00477-013-0743-9>
- Westra S, Thyer M, Leonard M, Kavetski D, Lambert M (2014) A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resour Res* 50:5090–5113. <https://doi.org/10.1002/2013WR014719>
- Williams JR (1969) Flood routing with variable travel time or variable storage coefficients. *Trans ASAE* 12:100–103
- Yang J, Reichert P, Abbaspour KC, Yang H (2007) Hydrological modelling of the Chaohe Basin in China: statistical model formulation and Bayesian inference. *J Hydrol* 340(2007):167–182
- Yang J, Abbaspour KC, Reichert P, Yang H (2008) Comparing uncertainty analysis techniques for a SWAT application to Chaohe Basin in China. *J Hydrol* 358:1–23
- Zhang HX, Yu SL (2004) Applying the first-order error analysis in determining the margin of safety for total maximum daily load computations. *J Environ Eng* 1306:664–673
- Zhang X, Srinivasan R, Bosch D (2009) Calibration and uncertainty analysis of the SWAT model using genetic algorithms and Bayesian model averaging. *J Hydrol* 374:307–317. <https://doi.org/10.1016/j.jhydrol.2009.06.023>
- Zheng Y, Han F (2016) Markov Chain Monte Carlo (MCMC) uncertainty analysis for watershed water quality modeling and management. *Stoch Environ Res Risk Assess* 30:293. <https://doi.org/10.1007/s00477-015-1091-8>
- Zhenyao S, Lei C, Tao C (2013) The influence of parameter distribution uncertainty on hydrological and sediment modeling: a case study of SWAT model applied to the Daning watershed of the Three Gorges Reservoir Region, China. *Stoch Environ Res Risk Assess* 27:235. <https://doi.org/10.1007/s00477-012-0579-8>