Research papers

A probabilistic appraisal of rainfall-runoff modeling approaches within SWAT in mixed land use watersheds

Ali Tasdighi⁎, Mazdak Arabi, Daren Harmel

⁎ Corresponding author.

E-mail addresses: ali.tasdighi@colostate.edu (A. Tasdighi), mazdak.arabi@colostate.edu (M. Arabi), daren.harmel@ars.usda.gov (D. Harmel).

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A probabilistic approach is presented to assess the performance validity of the empirical Curve Number (CN) and physically-based Green and Ampt (G&A) rainfall-runoff methods in the SWAT model. Specifically, the effects of modeling uncertainties on characterization of the hydrologic budgets and streamflow regimes at various spatial scales and upstream land use conditions are investigated. A Bayesian total uncertainty assessment framework, which explicitly accounts for uncertainties from model parameters, inputs, structure, and measurement data, was employed to explore uncertainties in streamflow simulation using SWAT with different rainfall-runoff methods in a mixed-land use watershed. While the models were trained for streamflow estimation only at the watershed outlet, the performances of the models were compared at different stream locations within the watershed. At the watershed outlet, the CN method had a slightly better, but not significant, performance in terms of streamflow error statistics. Similar results were obtained for the predominantly forested and agricultural tributaries. However, in tributaries with higher percentage of developed land, G&A outperformed the CN method in simulating streamflow based on various performance metrics. In general, the 95% prediction intervals from the models with G&A method covered a higher percentage of observed streamflow especially during the high flow events. However, they were approximately 20–45% wider than the corresponding 95% prediction intervals from the CN methods. Using 95% prediction interval for estimated flow duration curves, results indicated that the models with CN methods underestimated high flow events especially in tributaries with highly developed land use. However, the CN methods generated higher water yields to streams than the G&A method. The results of this study have important implications for the selection and application of appropriate rainfall-runoff methods within complex distributed hydrologic models particularly when simulating hydrologic responses in mixed-land use watersheds. In the present study, while CN and G&A methods in the SWAT model performed similarly at the outlet of a mixed-land use watershed, G&A captured the internal processes more realistically. The subsequent effects on the representation of internal hydrological processes and budgets are discussed.

1. Introduction

Watershed models are increasingly used to assess hydrologic responses of watersheds to changes in land use, climate variability and change, and other alterations of system characteristics. Faced with myriad hydrologic models, hydrologists ought to select an appropriate model and demonstrate its performance validity for the desired assessments. The performance of watershed models is often evaluated based on a deterministic approach, i.e. calibrate → validate → predict, where calibration is conducted to obtain a parameter set that provides the best fit between model responses and observations at the watershed outlet (Seibert, 1999; Santhi et al., 2008; Niraula et al., 2015). Such assessments can be inadequate, and in many cases misleading for selecting an appropriate distributed hydrologic model, particularly when the model is used to simulate interior hydrologic processes or to assess responses at various locations within the watershed (Beven, 2001; Ahmadi et al., 2014). Probabilistic approaches can address the equifinality and nonuniqueness issues in parameter estimation (Moradkhani et al., 2005), input uncertainty (Kavetski et al., 2003), and measurement errors (Harmel and Smith, 2007) when assessing competing model structures (Ajami et al., 2007).

Literature is replete with studies in which a model calibrated at the watershed outlet is used to predict streamflows at interior locations within the watershed under varying upstream land use conditions (Ahearn et al., 2005; Baker and Miller, 2013; Yan et al., 2013). Similarly, studies continue to use the deterministic approach to explore the
hydrologic effects of land use change (Zhou et al., 2013; Sunde et al., 2016; Zuo et al., 2016), climate change (Xu et al., 2013; Maeruo et al., 2017), or implementation of management practices (Taylor et al., 2016; Jang et al., 2017; Motallebi et al., 2017). Interestingly, calibrated model responses at the watershed outlet are also used to compare the performance of competing models (Seibert, 1999; Santhi et al., 2008; Niraula et al., 2015). The validity of these studies remains unclear, mainly because a model could produce responses at the watershed outlet that adequately match observations (i.e. right answer) while representing the internal processes and responses incorrectly (i.e. wrong reasons) (Beven, 2006).

Two common methods for rainfall-runoff modeling are the Curve Number method (CN; USDA-NRCS, 2004) and the Green and Ampt method (G&A; Green and Ampt, 1911). The CN is an empirical method that provides estimates of runoff under varying land use and soil types using total volume of rainfall. Conversely, the G&A is a physically-based method that uses rainfall intensity and duration along with soil physical characteristics such as hydraulic conductivity to simulate infiltration. A major limitation of the CN method is that it only uses the total volume of rainfall and does not account for rainfall intensity and duration. Hence, the applicability of the method is limited to simulations at daily to annual time steps, and cannot be extended to resolve processes at sub-daily time steps.

Such limitations can result in erroneous simulations of runoff processes in areas with inherently quick hydrologic responses to rainfall events, such as small catchments, areas with relatively low soil permeability, and developed areas with large impervious surfaces (Miller et al., 2014). Yet, many studies continue to use the CN method to quantify runoff in mixed-land use watersheds with considerable areas of developed land (Zhou et al., 2013; Yan et al., 2013). Similarly, several studies have used the CN method calibrated in a dominantly agricultural or forested watershed to predict changes in runoff or streamflow under projections of urban growth (Du et al., 2012; Zhou et al., 2013; Niraula et al., 2015; Wagner et al., 2016).

The Soil and Water Assessment Tool (SWAT; Arnold et al., 1998, 2012) is a semi-distributed watershed model, which includes both the CN and G&A methods to simulate rainfall-runoff processes. The large majority of studies conducted with SWAT have used the CN method for hydrologic and water quality assessments (Gassman et al., 2014; Bauwe et al., 2016). A few recent studies have compared the performance of the CN or G&A methods. The results from these studies are often inconsistent and in some cases contradictory. Some studies concluded better performance of the CN method (Wilcox et al., 1990; Kannan et al., 2007; Bauwe et al., 2016; Cheng et al., 2016), while other studies demonstrated better performance of the G&A (King et al., 1999; Ficklin and Zhang, 2013; Yang et al., 2016). However, the generalizability of the conclusions from these studies remains limited because they neglected three important considerations. First, comparison of models based on a single calibrated parameter set is subject to biases in model selection and modelers’ expertise. Second, a model calibrated for responses at the outlet of the watershed may not produce reliable simulations of interior locations with different geospatial characteristics such as land use. Third, modeling uncertainties must be incorporated in comparison and selection of models. Modeling uncertainty includes the uncertainty in model parameters, algorithms, inputs, and measurement data (Beven and Binley, 1992; Vrugt et al., 2003; Ajami et al., 2007; Harmel et al., 2014; Yen et al., 2014). Not accounting for any of these sources of uncertainty when comparing the performance of models can mask real differences in model performance.

The overall goal of this study is to probabilistically investigate the performance validity of the CN and G&A methods within SWAT for simulating hydrologic responses under varying land use conditions. The specific objectives are to: (i) evaluate the uncertainties from different sources (parameters, input data, and model structure) under the CN and G&A methods when simulating streamflow; (ii) compare the streamflow prediction uncertainty for SWAT with different rainfall-runoff methods; and (iii) quantify the total hydrologic regime and components of streamflow simulated using CN and G&A methods at locations with various dominant upstream land use. While a number of previous studies have compared the performance of CN and G&A methods, applying a total uncertainty estimation framework and accounting for upstream land use variations to assess the performance of the methods is novel. Considering that different climate inputs (daily vs. subdaily) are used for CN vs. G&A, the study reveals the importance of accounting for input data uncertainty when comparing the performance of the methods in simulating hydrologic responses. Also, using full flow statistics via predictive flow duration curve uncertainty for assessing the performance of the methods sheds light on the benefits and deficiencies of each method. This study provides useful insight into the benefits and limitations of different runoff simulation methods in the widely applied SWAT model.

2. Materials and methods

Three separate SWAT models were developed. The models were identical except for the rainfall-runoff mechanism and the precipitation time step accommodating the mechanism. The analysis period was 2002-2012 of which 2002–2008 was used for training (2000–2001 was used for model warmup), and 2009–2012 was used for testing the models. The uncertainty assessment framework developed was used with the likelihood function set to consider the errors only at the outlet of the watershed. At each model realization, the error statistics including likelihood, sum of squared errors (SSE), and Nash-Sutcliffe Coefficient of Efficiency (NS) as well as time series of simulated streamflow were stored for the outlet of the watershed and five other stream locations (USGS gauges) inside the watershed. The stream locations inside the watershed were selected such that they included a variety of sizes and dominant land use types (agriculture, developed, forest) of upstream subwatersheds. By comparing model performance at these locations inside the watershed, we assessed how models with CN and G&A mechanisms perform at subwatersheds with different dominant land use types while training the model at the watershed outlet.

2.1. Study watershed

The Haw watershed in central North Carolina within the Piedmont region (Fig. 1) drains 3280 km² (1270 miles²). The land use within the watershed is: 43% forest, 20% urban/suburban, and 27% agriculture, of which more than 90% is pasture (National Land Cover Database; NLCD2011; USGS-TNM, 2016). Average annual precipitation is 1060 mm (mm) with summer precipitation being the highest and autumn the driest. From 2002 to 2012, the driest year was 2002 with 780 mm precipitation, and 2003 had the highest precipitation (1815 mm) (National Oceanic and Atmospheric Administration; NOAA, 2016).

2.2. Hydrologic model

SWAT is a continuous-time, distributed-parameter, process-based watershed model, which has been used extensively for hydrologic and water quality assessments under varying climatic, land use, and management conditions in small watersheds to large river basins (Gassman et al., 2007; Arnold et al., 2012; CARD Staff, 2016). The model has the capability to run on daily or smaller time steps. In SWAT, the watershed is split into smaller subwatersheds, which are further discretized into Hydrologic Response Units (HRUs). HRUs are the smallest spatial units in SWAT, and are defined as areas within each subwatershed with unique combinations of land use, soil, and slope class. SWAT does not have a routing component between HRUs. The runoff from HRUs within each subwatershed is aggregated and then routed through the stream network.

Climate inputs drive hydrologic responses and provide moisture and
energy inputs in SWAT. Hydrologic processes simulated in the model include canopy storage, surface runoff, infiltration, evapotranspiration, lateral flow, tile drainage, redistribution of water within the soil profile, return flow, and recharge (Arnold et al., 2012). Surface runoff is simulated using either the modified G&A (Mein and Larson, 1973) with subdaily rainfall or CN method with daily rainfall. The rainfall-runoff methods examined in this study are infiltration-excess methods. Saturation-excess was not an important consideration in this study watershed. However, the probabilistic framework for assessment of rainfall-runoff methods developed and illustrated in this study could be extended to other versions of the SWAT model that do incorporate saturation-excess processes (e.g. Easton et al., 2008).

The Green and Ampt and CN methods are available runoff estimation mechanisms (i.e. model structures) in SWAT. In the CN method, surface runoff is estimated with daily rainfall depth ($R_{day}$) and retention parameter ($S$):

$$ Q_{surf} = \frac{(R_{day}-0.2S)^2}{R_{day} + 0.8S} $$

where $Q_{surf}$ is the depth of the surface runoff. All parameters are values for the day in millimeter (mm). The retention parameter for CN in SWAT is often estimated based on antecedent soil moisture:

$$ S = S_{max} \left( 1 - \frac{SW}{[SW + \exp(w_1 - w_2 + SW)]} \right) $$

where $SW$ is the soil water content of the entire profile excluding the amount of water (mm) held in profile at wilting point, $w_1$ and $w_2$ are shape coefficients (explained in SWAT documentation, Neitsch et al., 2011), and $S_{max}$ denotes the maximum value of retention parameter (mm) calculated as:

$$ S_{max} = 25.4 \left( \frac{100}{CN} - 10 \right) $$

where $CN$ is the curve number for soil moisture condition I (explained in SWAT documentation, Neitsch et al., 2011). This approach tends to overestimate the runoff in shallow soils (Kannan et al., 2007). Hence, another option can be used to compute the retention parameter at a given time step $t$ based on plant evapotranspiration (Neitsch et al., 2011):

$$ S_t = S_{t-1} + E_t \cdot \exp \left( -\frac{CNCOEFS_{t-1}}{S_{max}} \right) \cdot R_{day} + Q_{surf} $$

where $S_{t-1}$ is the retention parameter from the previous time step (i.e. day), $E_t$ is the potential evapotranspiration for the day, and $CNCOEF$ is the weighting coefficient determined by calibration. At the beginning of the simulation (first day), the retention is defined as $S = 0.9 \cdot S_{max}$. Calculating the CN based on plant evapotranspiration instead of soil moisture makes its value more dependent on climate instead of soil storage (Neitsch et al., 2011). The CN method based on soil moisture and evapotranspiration will be referred to as CN I and CN II, respectively, hereafter.

The modified G&A method is the alternative model structure in SWAT for runoff estimation that incorporates rainfall duration and intensity to compute the infiltration rate on subdaily time steps:

$$ f_{inf,t} = K_i \left( \frac{\Psi_{inf,t} \Delta \delta_i}{F_{inf,t}} \right) $$

where $f_{inf,t}$ is the infiltration rate at time $t$ (mm per hour), and $F_{inf,t}$ is the cumulative infiltration at time $t$ (mm), $K_i$ denotes the effective hydraulic conductivity (mm per hour), $\Psi_{inf,t}$ denotes the wetting front potential (mm), and $\Delta \delta_i$ represents the change in volumetric moisture content across the wetting front (mm/mm). Equations for $F_{inf,t}$, $K_i$, $\Psi_{inf,t}$, and $\Delta \delta_i$ are explained in SWAT documentation (Neitsch et al., 2011).

### 2.3. Model inputs: Terrain, soils, land use, climate, and hydrography

The elevation data for building the SWAT model was the 1/3 arc-second (~10m) resolution digital elevation model (DEM) of the Haw watershed obtained from United States Geological Survey The National Map (USGS-TNM, 2016). The Soil Survey Geographic (SSURGO) database from United States Department of Agriculture Natural Resources Conservation Services (USDA-NRCS, 2016) was used to represent soil characteristics and variability in the watershed. The SSURGO soil data were not available for a small area in the north-eastern part of the Haw watershed (~3% of total watershed area). The State Soil Geographic (STATSGO; USDA-NRCS, 2016) data were used to cover the area with missing SSURGO information. The components within each map unit were aggregated using a weighted average scheme in which the weights
were the percentage of each component within the map unit. The National Land Cover Database (NLCD) for year 2006 was obtained from the USGS TMN and used as land use for building the model. The resolution of the SSURGO, STATSGO, and NLCD data was 1 arc-second (∼30 m).

Stream flowlines, subwatershed boundaries, and other hydrography information were obtained from USGS National Hydrography Dataset (USGS-NHD, 2016). The stream flowlines were used for more accurate stream delineation in SWAT. By superimposing the NHD flowlines on the DEM in the process of watershed delineation, the hydrographic segmentation and subwatershed boundary delineation was improved especially in locations where the DEM does not provide enough accuracy (Winchell et al., 2007). The Haw Watershed contains three major lakes whose operations altered the natural streamflow regime to some extent. These lakes were included in the models, and their operations were modeled by including their average monthly outflows.

Observed climate for three meteorological stations (Fig. 1) were obtained from the National Climatic Data Center (NCDC) Quality Controlled Local Climatological Data (QCLCD; NOAA, 2016) database. Daily and hourly precipitation, minimum and maximum temperature were collected for the period of 2000 to 2012. Wind speed, solar radiation, and relative humidity were simulated by SWAT.

2.4. Measurements: Stream discharge

Daily stream discharge data from USGS National Water Information System (USGS-NWIS, 2016) were obtained for six stations with daily measurements for the analysis period (2002–2012) (Fig. 1). The locations of gauges (Table 1) were carefully selected to enable adequate characterization of the predictive skill of different model structures under varying land use conditions, particularly developed (i.e. urban) versus undeveloped areas.

2.5. The SWAT model setup

The DEM was used along with the NHD flowlines to delineate the watershed and streams. The watershed was divided into 23 subwatersheds. The HRU definition was done using the land use and soil data with 10% threshold for delineation. Since the topographic variability was small, a single class slope was assumed within each subwatershed. Using these settings, 122 HRUs were defined for the watershed.

In highly urbanized watersheds, facilities such as wastewater treatment plants can alter the natural streamflow regime. Hence, their effects must be incorporated into the model. In Haw Watershed, there were 10 large facilities with average daily effluent larger than 0.1 MGD discharging into the streams. These facilities were included in the models through adding point sources in SWAT at stream locations where they discharged their effluents. Their effluents were discharged into streams at related stream locations as average monthly flows.

Another factor that can alter the natural streamflow regime in urban areas is the stormwater discharge into streams through pipes or conduits. The urban stormwater collection systems and control measures were not included in the models as it was beyond the scope of this study. This study was concentrated on aggregate responses of the developed areas at the subwatershed level. In other words, the model performance was evaluated probabilistically for stream locations that incorporate the aggregate effects of urban areas, including point sources and urban stormwater runoff. Considering that, we tried to define the subwatersheds such that each urban area falls within a subwatershed so that addition of stormwater in a diffuse manner or through a conduit or pipe has a minimal effect on the results.

Two separate SWAT models were developed using ArcSWAT 2012 (USDA-ARS, 2014). The models were completely identical except for the runoff estimation method and precipitation time step. One of the models was developed with daily precipitation and CN method using the soil moisture (CN I) and plant evapotranspiration (CN II), and the other model was developed using the hourly precipitation and G&IA method for runoff simulation. Therefore, three model setups were prepared for the analyses.

2.6. Probabilistic model assessment framework

A probabilistic approach was used to assess the predictive skill and performance validity of competing model structures under varying land use conditions. The Bayesian-based approach explicitly accounts for uncertainties from model parameterization, climate input data (i.e. precipitation), model structure (CN I, CN II, or G&IA), and measurement data (i.e. daily streamflow). The framework was developed in MATLAB (The MathWorks, Inc.). A Markov Chain Monte Carlo (MCMC) sampling scheme, the DREAM method, was used to sample the parameter space and derive the posterior distributions. A statistically-correct likelihood function, which explicitly accounts for streamflow error heteroscedasticity and autocorrelation, was employed to ensure the reliability of the search algorithm. Input data uncertainty was incorporated by using precipitation multipliers drawn from a Gaussian distribution with an uncertain mean and standard deviation for each meteorological station. Inferences on posterior distributions of precipitation multipliers were obtained simultaneously along with SWAT model parameters. Finally, Bayesian Model Averaging (BMA) was used to evaluate model structural uncertainty and to assess the predictive skill of the competing model structures. Similar frameworks have been developed in other studies (e.g. Ajami et al., 2007; Kavetski et al., 2003) in which they used other MCMC algorithms such as Shuffled Complex Evolution Metropolis (SCEM). However, employing the DREAM algorithm for conducting the MCMC within this total uncertainty estimation framework is novel. More importantly, application of such framework for probabilistic appraisal of different rainfall-runoff methods within distributed hydrologic models has not been conducted to the best of our knowledge.

2.6.1. Model parameter uncertainty

In a hydrologic model (M), streamflow is simulated (\( \hat{Q} \)) as a function of climatic inputs (i.e. precipitation, \( R \)), and a vector of model parameters (\( \theta \)):

\[
\hat{Q} = M(R, \theta)
\]  

The simulated streamflow is subject to a variety of errors stemming

<table>
<thead>
<tr>
<th>USGS Gauge ID</th>
<th>Subwatershed Outlet #</th>
<th>Area (km²)</th>
<th>% Developed</th>
<th>% Forest</th>
<th>% Agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>2093800</td>
<td>7</td>
<td>53</td>
<td>16</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td>209553650</td>
<td>9</td>
<td>230</td>
<td>74</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>2095550</td>
<td>12</td>
<td>96</td>
<td>85</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>2096590</td>
<td>13</td>
<td>606</td>
<td>24</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>2096846</td>
<td>19</td>
<td>19</td>
<td>4</td>
<td>73</td>
<td>17</td>
</tr>
<tr>
<td>2096960</td>
<td>23</td>
<td>1275</td>
<td>18</td>
<td>45</td>
<td>29</td>
</tr>
</tbody>
</table>
from measured climate inputs, model parameters, and insufficiency of the model conceptualization (model structural error). The streamflow error residual then becomes:

\[
e(R, \hat{\beta}, M) = Q - M(R, \hat{\beta})
\]

where \( Q \) is the observed streamflow, and \( e(R, \hat{\beta}, M) \) denotes the streamflow error residuals due to errors from observed precipitation, model parameters, and model structure. Applying the Bayes theorem, the parameter set \( \hat{\beta} \) is assigned posterior probability distribution, \( p(\hat{\beta}|Q) \), which is proportional to the product of the parameter prior probability distribution, \( p(\hat{\beta}) \), and a likelihood function, \( L(Q|\hat{\beta}) \).

The likelihood function assuming normally and independently distributed model residuals \( e \) with mean zero and variable standard deviation at each observation time step \( \sigma_i \), can be expressed as (Vrugt, 2016):

\[
L(Q|\hat{\beta}, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i^2} \exp \left[ \frac{1}{2\sigma_i^2}(Q_i - \hat{Q}_i(\hat{\beta}))^2 \right]
\]

(8)

where \( n \) is the number of time steps. Using a variable standard deviation at each observation time step \( \sigma_i \) allows accounting for error heteroscedasticity which often exists in streamflow simulations. However, in most cases it is infeasible to determine \( \sigma_i \) at each time step due to lack of repeated streamflow measurements. Different approaches have been proposed to circumvent this issue. Some studies have used different transformations including natural log or Box-Cox to stabilize \( \sigma \) and reduce heteroscedasticity (Sorooshian and Dracup, 1980). Others have proposed alternative forms for the likelihood function (Schoups and Vrugt, 2010). In this study, we applied a natural log transformation on measured and observed streamflow to reduce error heteroscedasticity. Often it is easier to maximize the logarithm of the likelihood function due to numerical stability and algebraic simplicity (Ajami et al., 2007; Vrugt, 2016). Hence, the natural log of the likelihood function was adopted for optimization.

The streamflow error residuals are usually not independently distributed and in most cases temporal autocorrelation exists in the residuals specifically when errors are estimated at daily or smaller time steps. A common approach to reduce the autocorrelation in model residuals is applying an auto-regressive scheme on the error residuals (Sorooshian and Dracup, 1980):

\[
e_i = \rho e_{i-1} + \nu_i
\]

(9)

where \( \rho \) is the lag serial correlation coefficient for the error residuals and \( \nu_i \) is a vector of random components, \( \nu_i \in N(0, \sigma_i) \). In this study we used a first-order autoregressive transformation (AR-1). Applying the natural log and AR-1 (\( |\rho|<1 \)), the log-likelihood function becomes (Vrugt, 2016):

\[
L(Q, \sigma, \rho) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_i^2) - \frac{1}{2\sigma_i^2}(Q_i - \hat{Q}_i(\hat{\beta}))^2 - \frac{1}{2}(1-\rho^2)(\sigma_i^2) + \sum_{i=2}^{n}(e_i - \rho e_{i-1}) \]

(10)

Parameters \( \rho \) and \( \sigma \) are determined along with model parameters at each model realization during the MCMC sampling algorithm.

The parameters for the uncertainty analysis were selected based on experience and sensitivity analysis performed previously (Arabi et al., 2007; Arnold et al., 2012; Tasdighi et al., 2017). Table 2 lists the parameters selected for uncertainty analysis along with their ranges. In this study, uniform (noninformative) prior distributions were assumed for parameters within predefined ranges. The same assumption has been used in many other hydrological modeling studies since prior knowledge of the model parameters is often not available and is case-specific (Ajami et al., 2007). The ranges for parameters were selected based on the SWAT user manual and experience from previous study (Tasdighi et al., 2017; Arnold et al., 2012).

### 2.6.2. Model input uncertainty

While uncertainty from model parametrization has been examined in many hydrologic modeling studies, there are only a few cases where input uncertainty is explicitly accounted for (Kavetski et al., 2003; Ajami et al., 2007). The majority of these studies incorporate input uncertainty by applying latent variables, which are basically multipliers for precipitation events drawn randomly from a predefined distribution along with model parameters (Kavetski et al., 2003; Leta et al., 2015). This approach can lead to dimensionality problems as the number of precipitation events increase. In this study a method proposed by Ajami et al. (2007) is implemented to account for precipitation uncertainty. Using this method, instead of iterating on each single multiplier, the iteration is performed on the mean and standard deviation of a random Gaussian distribution from which the multipliers are randomly drawn at each time step:

\[
\bar{R}_i = \phi_1 R_i; \quad \phi_1 \sim N(\mu_{\phi_1}, \sigma_{\phi_1}^2)
\]

(11)

where \( \bar{R}_i \) and \( R_i \) are the corrected and observed precipitation depths respectively, \( \phi_1 \) is the random multiplier drawn from a normal

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Input file</th>
<th>Description</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURLAG</td>
<td>.bsn</td>
<td>Surface runoff lag coefficient</td>
<td>0.001</td>
<td>15</td>
</tr>
<tr>
<td>CNCOEF</td>
<td>.bsn</td>
<td>Plant ET curve number coefficient</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>ALPHA, BF</td>
<td>.gw</td>
<td>Base flow alpha factor for recession constant</td>
<td>0.001</td>
<td>1</td>
</tr>
<tr>
<td>GW, DELAY</td>
<td>.gw</td>
<td>Groundwater delay time</td>
<td>0.001</td>
<td>100</td>
</tr>
<tr>
<td>GW, REVAP</td>
<td>.gw</td>
<td>Groundwater revap coefficient</td>
<td>0.02</td>
<td>0.2</td>
</tr>
<tr>
<td>GWQMN</td>
<td>.gw</td>
<td>Threshold depth of water for return flow to occur</td>
<td>0.01</td>
<td>5000</td>
</tr>
<tr>
<td>RCHRG, DP</td>
<td>.gw</td>
<td>Deep aquifer percolation fraction</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CANMX</td>
<td>.hru</td>
<td>Maximum canopy storage factor</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>ESDO</td>
<td>.hru</td>
<td>Soil evaporation compensation factor</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>OV, N</td>
<td>.hru</td>
<td>Manning’s n value for overland flow</td>
<td>0.01</td>
<td>0.6</td>
</tr>
<tr>
<td>SLSUBRSN</td>
<td>.hru</td>
<td>Average slope length</td>
<td>10</td>
<td>150</td>
</tr>
<tr>
<td>CN, F</td>
<td>.mgf</td>
<td>Fraction change in SCS runoff curve number</td>
<td>−0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>CH, KII</td>
<td>.rtc</td>
<td>Effective hydraulic conductivity in the main channel</td>
<td>0.025</td>
<td>150</td>
</tr>
<tr>
<td>CH, NII</td>
<td>.rtc</td>
<td>Manning’s n value for the main channels</td>
<td>0.01</td>
<td>0.3</td>
</tr>
<tr>
<td>SOL, AWC</td>
<td>.sol</td>
<td>Fraction change in available soil water capacity</td>
<td>−0.5</td>
<td>1</td>
</tr>
<tr>
<td>SOL, K</td>
<td>.sol</td>
<td>Fraction change in saturated hydraulic conductivity</td>
<td>−0.5</td>
<td>2</td>
</tr>
<tr>
<td>SOL, Z</td>
<td>.sol</td>
<td>Fraction change in soil depth</td>
<td>−0.5</td>
<td>1</td>
</tr>
<tr>
<td>SOL, BD</td>
<td>.sol</td>
<td>Fraction change in soil moist bulk density</td>
<td>−0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>CLAY</td>
<td>.sol</td>
<td>Fraction change in % Clay</td>
<td>−0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>SAND</td>
<td>.sol</td>
<td>Fraction change in % Sand</td>
<td>−0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>CH, KI</td>
<td>.sub</td>
<td>Effective hydraulic conductivity in tributary channels</td>
<td>0.025</td>
<td>150</td>
</tr>
<tr>
<td>CH, NI</td>
<td>.sub</td>
<td>Manning’s n value for the tributary channels</td>
<td>0.01</td>
<td>0.3</td>
</tr>
</tbody>
</table>
where \( p(M|Q) \) is the posterior probability of the model \( M \). This term can be assumed as a probabilistic weight \( w_i \) for model \( M_i \) in the BMA prediction \( \hat{Q}_{\text{BMA}} \). The constraint for BMA weights is: \( \sum_{i=1}^{n} w_i = 1 \). Higher values of \( w_i \) can be interpreted as higher predictive skill for a given model structure.

The model weights can be determined using different optimization techniques. The expectation-maximization (EM) algorithm (Dempster et al., 1977) is one such technique to estimate model weights used in several studies (Yen et al., 2014; Ajami et al., 2007). In this study the EM method was used to determine the model weights.

The Brier scores were also employed to compare the performance of the three models while incorporating parameter and input data uncertainties. The Brier score (BS) is a measure of the accuracy of the prediction and has been frequently used in the probabilistic forecast analysis (Georgakakos et al., 2004). BS is defined as:

\[
BS = \frac{1}{n} \sum_{i=1}^{n} (f(t) - o(t))^2
\]

(13)

where \( n \) is the number of time steps in the record, \( f(t) \) is estimated by the fraction of model simulations larger than the predefined streamflow threshold, and \( o(t) \) is a binary value equal to 1 if the observation at time step \( t \) is larger than the predefined threshold and equal to zero in all other cases. In this form (Eq. (13)), the lower the value of the BS the better the prediction skill of the model.

2.6.4. The DREAM algorithm for MCMC analyses

Several Bayesian algorithms are available which have been widely used for uncertainty assessment in hydrologic modeling including the Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992), the Shuffle Complex Evolution Metropolis (SCEM-UA; Vrugt et al., 2003), and the DiRential Evolution Adaptive Metropolis (DREAM; Vrugt et al., 2009). DREAM is a multi-chain MCMC method that randomly samples the parameter space and automatically tunes the scale and orientation of the sampling distribution to move toward the target distribution by maximizing the value of the likelihood function. The method has been used extensively for parameter estimation of complex environmental models (Vrugt, 2016). The convergence of the algorithm can be monitored using the procedure proposed by Gelman and Rubin (1992). In this procedure, a scale reduction score \( R \) is monitored to check whether each parameter has reached a stationary distribution (Gelman and Rubin, 1992). The common convergence criterion of \( R \leq 1.2 \) was used in this study as well. DREAM is specifically beneficial in the optimization of complex high dimensional problems. In this study, DREAM was employed to sample the parameter space and derive the posterior distributions.

2.7. The strategy for assessment of model performances

Several criteria were used to assess the performance of the models including: error statistics (likelihood, SSE, and NS) determined based on simulated and observed hydrographs, width of the band of uncertainty (spread), inclusion rate (coverage) determined based on the streamflow observations and 95% confidence interval of simulation ensembles, flow duration curves along with bands of uncertainty, Brier scores, and the water budget in the watershed was analyzed using the results obtained from different models at various locations.
3. Results and discussion

3.1. Evaluation of model performances

The purpose of this step of analyses was to monitor the variation of streamflow error statistics at different stream locations in the watershed during the training of the model at the watershed outlet using the Bayesian total uncertainty analysis framework. Fig. 2 illustrates the variation of error statistics at different stream locations. It is important to note that the training of the models was performed only at the watershed outlet (outlet 23) using the values of likelihood (Eq. (10)) as the objective function.

The performances of the three models were identical in terms of the likelihood function at the watershed outlet (outlet 23). However, CN I and CN II models had a slightly (although statistically significant) better performance in terms of SSE and NS (t-test $p < 0.01$). The two-sample Kolmogorov-Smirnov test revealed that the distributions of error statistics (e.g. NS) under different models were different at 0.01 significance level at all locations. At outlets 13 and 19 with mainly agricultural and forested land use, CN models performed slightly better (t-test $p < 0.01$) than the G&A model in terms of SSE and NS. While CN I and CN II had very close performance at different locations, CN II had a relatively better performance (t-test $p < 0.01$) compared to CN I and G&A in the highly forested subwatershed (outlet 19) showing lower values of SSE (on average 8% and 11% lower than CNI and G&A respectively) and higher values of NS (on average 21% and 29% higher than CNI and G&A respectively). This was foreseeable as with CN II, the curve number is determined based on plant evapotranspiration instead of soil moisture. Similar results were reported for better performance of the CN method over the G&A in other agricultural watersheds (Kannan et al., 2007; Cheng et al, 2016). In contrast, comparing the performance of the CN and G&A methods in an intensive agricultural watershed, Ficklin and Zhang (2013) concluded that the G&A model is more likely to generate better daily simulations. It should be noted that these
Fig. 5. BMA weights for the 3 models at different locations (Solid horizontal lines on the boxes show the median; the boxes show the range of values between 25th and 75th percentile; the whiskers show the 0.5 and 99.5 percentile).

Fig. 6. Brier scores for the 3 models at different locations. Note that we used $BS' = 1 - BS$ for illustration of model performances in this figure. In this form, higher values of $BS'$ indicate better performance of the model.

Table 3
Coverage rates and spread at various outlets for the three models. The coverage rates and spread are calculated using the 95% confidence interval of simulated ensemble hydrographs and corresponding observations.

<table>
<thead>
<tr>
<th>Outlet#</th>
<th>Area (km²)</th>
<th>% Developed</th>
<th>Model</th>
<th>Training Coverage (%)</th>
<th>Spread (m³/sec)</th>
<th>Testing Coverage (%)</th>
<th>Spread (m³/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>230</td>
<td>74</td>
<td>CN I</td>
<td>45</td>
<td>1.8</td>
<td>46</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CN II</td>
<td>52</td>
<td>2.3</td>
<td>56</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G&amp;A</td>
<td>71</td>
<td>3.9</td>
<td>67</td>
<td>2.9</td>
</tr>
<tr>
<td>12</td>
<td>96</td>
<td>85</td>
<td>CN I</td>
<td>46</td>
<td>0.65</td>
<td>54</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CN II</td>
<td>53</td>
<td>0.87</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G&amp;A</td>
<td>67</td>
<td>1.4</td>
<td>73</td>
<td>1.1</td>
</tr>
<tr>
<td>13</td>
<td>606</td>
<td>24</td>
<td>CN I</td>
<td>51</td>
<td>16.5</td>
<td>50</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CN II</td>
<td>61</td>
<td>22.3</td>
<td>58</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G&amp;A</td>
<td>70</td>
<td>26.1</td>
<td>66</td>
<td>18.3</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>4</td>
<td>CN I</td>
<td>38</td>
<td>0.35</td>
<td>41</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CN II</td>
<td>55</td>
<td>0.44</td>
<td>62</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G&amp;A</td>
<td>45</td>
<td>0.61</td>
<td>42</td>
<td>0.41</td>
</tr>
<tr>
<td>23</td>
<td>1257</td>
<td>18</td>
<td>CN I</td>
<td>64</td>
<td>27.6</td>
<td>55</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CN II</td>
<td>70</td>
<td>31.1</td>
<td>56</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G&amp;A</td>
<td>76</td>
<td>36.5</td>
<td>54</td>
<td>19</td>
</tr>
</tbody>
</table>
studies used a deterministic approach, and none compared the performance of the CN and G&A with regard to upstream land use variations. In highly developed subwatersheds (outlets 9 and 12), G&A had a substantially and significantly ($t$-test $p < 0.01$) better performance than the CN methods with all error statistics (Fig. 2). This is an important result as it demonstrates that while trained at the outlet of the watershed, the G&A model had a much better performance in urbanized subwatersheds inside the watershed. The better performance of the G&A compared to CN methods in urban areas is investigated and discussed more in Section 3.5. Outlet 7 had an erratic behavior in terms of likelihood which can indicate fundamental deficiency of the models in simulating streamflow for that subwatershed.

3.2. Model parameter uncertainty

The posterior Cumulative Distribution Functions (CDF) of parameters for the three models are illustrated in Fig. 3. The posterior distributions were generated using 10,000 parameter sets sampled after the convergence of the MCMC algorithm. It can be observed that using different rainfall-runoff methods, different posterior distributions were inferred for parameters. In general, most posterior parameter distributions showed some level of skewness which indicates deficiency in identifiability (Ajami et al., 2007). While for most of the parameters the rainfall-runoff methods determined the degree of the skewness, for some parameters (CH-NII, GW-REVAP, GWQMN, and RCHRG-DP) the skewness changed from positive to negative using different methods. The CN I and CN II resulted in similar posterior distributions for most parameters. However, for CH-KI, SOL-AWC, and SOL-K different distributions were inferred which conforms to intuition as CN I uses the soil water content for determining the curve number in runoff estimation. Parameter distributions that show great deviation from normality indicate some sort of deficiency in the combination of model structure, input, and training data.

An important observation in Fig. 3 was the lower values of SURLAG parameter sampled for the G&A method. SURLAG controls the surface runoff storage through the fraction of the total available water that will be allowed to enter the reach. Smaller values of SURLAG result in higher storage of water and delay in release of water to the reach. Looking at the posterior distribution of this parameter, it can be observed that compared to CN methods, smaller SURLAG values have been sampled for the G&A method which can indicate simulation of a flashier hydrograph and higher peak flows by G&A. This was further investigated in Section 3.5.

These findings have important implications with regard to parameterization of SWAT when using different rainfall-runoff mechanisms. The majority of studies that have compared the performance of the rainfall-runoff mechanisms within SWAT have used a deterministic approach for setting the model parameters using the CN or G&A methods (e.g. King et al., 1999; Ficklin and Zhang, 2013). The results obtained here suggest that this approach may mask the differences between the methods and result in misleading inferences regarding the performance of the models under the CN or G&A methods.

3.3. Model input uncertainty

Precipitation multipliers were drawn from normal distributions
with random mean and standard deviation sampled during each model run within the Bayesian total uncertainty analysis framework. Fig. 4 shows the CDF of the values of the mean ($\mu$) sampled for input error models under different rainfall-runoff mechanisms for the three climate stations. Since the posterior distributions of standard deviation ($\sigma$) for input error models were close to uniform, they were excluded from the figure. In general, the mean of input error model under the G&A method showed a distribution closer to normal with mean around one. In contrast, the CN methods produced posterior distribution skewed toward the higher bound, especially at stations WBAN93783 and WBAN93785. Skewness toward the higher bound indicates the sampling algorithm’s attempt to increase the magnitude of the precipitation events by using multipliers larger than one. This can be explained by the structural deficiency of CN methods in simulating the peak streamflows, which caused higher values of precipitation multipliers to be drawn to augment the runoff volume to capture the high flow events. Under these circumstances, it can be hypothesized that the CN methods will lead to systematic overestimation of streamflow compared to G&A (assessed subsequently).

3.4. Model structural uncertainty

Bayesian model averaging was used at each model realization to combine the three rainfall-runoff models. BMA weights were determined using EM optimization method. Fig. 5 shows the boxplots of BMA weights at different subwatershed outlets. Mean of the weights and the weight for the optimized solution (max likelihood) are also marked on the boxplots. The two-sample Kolmogorov-Smirnov test revealed that the distribution of BMA weights for models were different at 0.01 significance level. Model performances in the agricultural and forested subwatersheds were relatively close in terms of BMA weights. It is evident that in the highly urbanized subwatersheds, G&A substantially and significantly (based on the t-test, $p < 0.01$) outperformed the CN methods. However, at the outlet of the watershed, the CN methods showed a slightly, but not significant, better performance. These results are in accordance with findings in Section 3.1 where G&A had a better performance in the subwatersheds with dominant urban land in terms of various error statistics.

In addition to BMA, Brier scores were also employed to assess the skill of the models in streamflow simulation. In original form (Eq. (13)), BS varies between 0 and 1; the lower the value of the BS the better the prediction skill of the model. For illustration, we used $BS = 1 - BS$ to compare the performance of the models in Fig. 6. In this figure, higher values of $BS$ indicate better performance of the model. At the outlet of the watershed, the CN models showed a slightly better performance compared to the G&A model specifically at low flow events. With the agricultural and forested watershed, the performances of the models were very close in terms of BS. However, at the highly urbanized subwatersheds, the G&A mechanism clearly outperformed the CN methods during low flow as well as high flow events resulting in higher values of BS. These findings are also congruent with results from the BMA and error statistics.

3.5. Streamflow prediction uncertainty

Hydrographs with bands of uncertainty have been used frequently
for visual assessment of streamflow prediction uncertainty (e.g. Harmel et al., 2010). While visual inspection of these graphs can provide useful information about the quality of simulation, in cases where a relatively long record of streamflow is to be inspected, this method can result in graphs that are difficult to read. Also, hydrographs may not be very effective when comparing the predictive performance of several models specifically when bands of uncertainty are involved. In this situation, one has to either use a smaller time period or select some specific events for illustration of the hydrograph or use other measures for visualizing streamflow. We, therefore, resorted to flow duration curves (FDC) to compare the performance of the models. Graphs of FDC along with bands of uncertainty provide an easily-readable informative measure which can be used to assess the quality of streamflow prediction (Vogel and Fennessey, 1994). More importantly, such graphs are much easier for comparing the predictive performance of the competing models. It should be noted that FDCs can also be misleading since the serial structure and autocorrelation of the sequence of the streamflow record is removed in them (Vogel and Fennessey, 1994). Bearing that in mind, while we used FDCs to compare the predictive performance of the models, the coverage (percent of observations lying inside the 95% confidence interval of simulation ensembles) and spread (average width of the 95% confidence interval uncertainty band) were determined based on hydrographs instead of FDCs. Since the performances of the models were similar in the smaller agricultural and forested subwatersheds (outlets 7, 13, and 19), they were excluded from further analysis. The G&A method resulted in wider bands of uncertainty (wider spread) and higher coverage rates compared to the CN methods (Table 2) during both training (2002–2008) and testing (2009–2012) periods except for outlet 23 during testing where CN II resulted in higher spread. At the outlet of the watershed (outlet 23), the difference between the methods is small in terms of coverage and spread. However, at stations 9 and 12 (urban-dominant subwatersheds), G&A generated higher coverage rates compared to the CN models.

Fig. 7 shows the flow duration curves for the observed streamflow and 95% confidence interval bounds for the simulated streamflows for the three models at the three stream locations during the training period. Comparing the performance of the models at the outlet of the watershed, the CN method produced narrower bands of uncertainty than the G&A method. However, the G&A model showed a slightly better performance in capturing the higher streamflow values. In highly developed subwatersheds 9 and 12, the CN methods were unable to capture the high flow events. Also for medium/low flow events, the observations lay on the upper bound of the uncertainty band. In contrast, for the G&A model, the high flow events were mostly captured and the observations were near the center of the uncertainty band.

The better performance of the G&A model in more developed areas was attributed to its better capacity in capturing the flashier behavior of the hydrographs due to quicker hydrologic responses in these areas. The G&A method was used with hourly precipitation data taking into account the effects of rainfall intensity. Since the hydrologic responses of urban areas (areas with higher impervious surfaces and lower permeability) tend to be flashier, this consideration may significantly improve the model performance. In particular, the G&A method outperformed the CN approaches in high-intensity rainfall events. This results in better performance of the G&A in capturing the peak flows compared to the CN method which often caused high errors for the CN method in more developed subwatersheds.

Similar results were obtained for the testing period (Fig. 8). Slightly narrower bands of uncertainty were determined during the testing period specifically at the watershed outlet (outlet 23). Table 4 presents the summary of different error statistics during training and testing periods. Minimum, median, and maximum values for various error statistics during the training and testing are provided in this table. The better performance of the G&A method compared to the CN in more developed subwatersheds can be observed in Table 4 as well.

The assumptions of the likelihood function used were assessed using different diagnostics (Fig. 9). Homoscedasticity, normality, and autocorrelation of streamflow residuals at watershed outlet (outlet 23) where model training is performed were assessed and illustrated in Fig. 9. The figure reveals that the AR-1 and log-transformation were
Fig. 9. Distribution of residuals (top row), partial autocorrelation coefficients of residuals with 95% confidence intervals (middle row), and residuals as a function of simulated streamflows (bottom row) at training station (outlet 23).

Fig. 10. Water budget for the watershed under different models: Panels (a)-(c) show the hydrologic budgets (ET: evapotranspiration; WY: total water yield to streams; R: deep groundwater recharge) for Outlets 9, 12, and 23, respectively; Panels (d)-(f) show surface runoff (SQ), lateral flow (LATQ), and groundwater (GWQ) contributions to water yield (WY).
successful in fulfilling the assumptions of the likelihood function in terms of normality, independence, and homoscedasticity of residuals. It can be observed that the G&A method resulted in narrower normal distribution around zero for residuals.

3.6. Assessment of hydrologic budget and streamflow components

The mean annual total water budget and components of streamflow for the three subwatersheds were quantified to gain insight into the differences among the hydrologic processes generating the outcomes. Fig. 10 shows the cumulative bar plots for the overall water budget (a-c) and more detailed components of streamflow (e-f). The G&A method generated lower water yield (total amount of water contributing to streamflow) and higher evapotranspiration (ET) compared to the CN at all locations. In all models, ET is the major component of water loss (between 60 and 75%) while about 10–15% and 20–25% of water turns into water yield for the G&A and CN methods, respectively.

The G&A method predicted lower surface runoff (SQ) and ground-water flow (GWQ) contributing to streamflow compared to the CN method at all locations. However, it simulated higher lateral flow (LATQ) contribution to streamflow. With CN I and CN II, about 35% of the water yield was derived from the lateral flow, while in the G&A method lateral flow contributed up to 75% of the streamflow. Other studies have shown higher baseflow contribution from the G&A method compared to the CN (Bauwe et al., 2016; Kannan et al., 2007). Considering all components of subsurface flow, the G&A model infiltrated more precipitation. More water in soil profile resulted in higher ET for the G&A model as well. Based on the results from multiple model performance criteria, the G&A performed better and hence was the more suitable method in the study watershed.

4. Conclusions

SWAT has been used extensively in the literature for hydrologic simulations. While the model has the capability to employ either CN or G&A method for runoff estimation, almost all studies have employed the CN method. This may be partly due to lack of a rigorous comparison study to justify merits of using one method over the other along with simplicity of the CN method. This study attempted to address this shortcoming. Regarding the extreme popularity of the SWAT model, the findings of this study can shed light on selecting the rainfall-runoff method within SWAT that can potentially lead to more realistic streamflow simulations in mixed-land use watersheds.

In this study, a Bayesian total uncertainty assessment framework was implemented to compare the performance of the three runoff generation mechanisms within SWAT under different upstream land use conditions. Using the uncertainty assessment framework at the watershed outlet, model performances was assessed at several stream locations inside the watershed. At the watershed outlet and subwatersheds with dominant agricultural or forest land use, the models with CN methods performed slightly better. However, at the two subwatersheds with highly developed land use, models with G&A method had a much better performance in simulating the streamflow. The better performance of the G&A method in more developed areas was attributed to its better capacity in capturing the flashy behavior of the hydrographs due to quicker hydrologic responses in these areas. The G&A method had a much better performance in simulating the peak flows in the more developed subwatersheds.

Overall, the streamflow prediction intervals from models with G&A method covered more observations. However, they were slightly wider indicating higher uncertainty for streamflow prediction. The CN models were unable to capture the high flow events specifically in developed subwatersheds. Posterior distribution of mean for Gaussian distributions from which precipitation multipliers were randomly drawn were closer to normal using the hourly precipitation data and G&A method while using daily precipitation with CN methods resulted in substantial negative skewness. The deficiency of models with CN methods in simulating the peak streamflows caused higher values of precipitation multipliers to be sampled to augment the runoff volume.

The CN method simulated higher water yield volumes specifically at the urban-dominated subwatersheds while G&A method simulated higher ET values. The higher water yield volumes predicted by CN method in the highly urbanized subwatersheds can be explained by CN attempt to simulate the high flow events during the model training which results in overall overestimation of water yield. The G&A model resulted in lower surface runoff at all locations compared to the CN models; however, it simulated higher infiltration and subsurface flows.

The results of this study have important implications for determining which rainfall-runoff method performs better in simulating the hydrologic regime. The evaluation is specifically relevant for applying a distributed hydrologic model such as SWAT in a mixed-land use watershed where model training will be performed only at the watershed outlet but the model is to be used for simulating hydrologic responses at different locations inside the watershed. In summary, the results suggest that in the Haw Watershed, while trained at watershed outlet, the SWAT model with G&A method can potentially perform better in areas inside the watershed with higher percentage of developed land. The SWAT models with CN methods proved to have similar or slightly better performance in areas with agriculture or forest dominant land use. However, care should be taken in applying such inferences as further studies in other watersheds with different physiographic characteristics are needed to generalize such findings. It should also be noted that in highly urbanized areas, the discharge of stormwater runoff through drainage pipes or conduits can disturb the natural streamflow regime. In such watersheds, including the discharge from conduits or pipes at the actual stream locations should improve the streamflow simulation.

Acknowledgements

This publication was made possible by USEPA grant RD835570. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. The authors would like to thank the Associate Editor and reviewers for their insightful comments.

USDA is an equal opportunity employer and provider.

References

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth