A Bayesian total uncertainty analysis framework for assessment of management practices using watershed models

Ali Tasdighi\textsuperscript{a,b,*}, Mazdak Arabi\textsuperscript{b}, Daren Harmel\textsuperscript{c}, Daniel Line\textsuperscript{d}

\textsuperscript{a} Department of Civil and Environmental Engineering, Colorado State University, Campus Delivery 1372, Fort Collins, CO, 80523-1372, USA
\textsuperscript{b} Department of Civil and Environmental Engineering, University of California Irvine, 44130 Engineering Gateway, Irvine, CA, 92697-2175, USA
\textsuperscript{c} USDA-ARS, Center for Agricultural Resources Research, 2150 Centre Ave, Fort Collins, CO, 80526-8119, USA
\textsuperscript{d} Department of Biological and Agricultural Engineering, North Carolina State University, Campus Box 7625, Raleigh, NC, 27695-7619, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Watershed management
BMP effectiveness
Water quality conservation
Watershed modeling
Nonpoint source pollution
Bayesian uncertainty analysis

\textbf{ABSTRACT}

A Bayesian total uncertainty analysis framework is presented to assess the model estimates of the effectiveness of watershed management practices in reducing nonpoint source (NPS) pollution. The framework entails a two-stage procedure. First, various sources of modeling uncertainties are characterized during the period before implementing Best Management Practices (BMPs). Second, the effectiveness of BMPs are probabilistically quantified during the post-BMP period. The framework was used to assess the uncertainties in effectiveness of two BMPs in reducing daily total nitrogen (TN) loads in a 54 ha agricultural watershed in North Carolina using the SWAT model. The results indicated that the modeling uncertainties in quantifying the effectiveness of selected BMPs were relatively large. Assessment of measured data uncertainty revealed that higher errors were observed in simulating TN loads during high flow events. The results of this study have important implications for decision-making under uncertainty when models are used for water quality simulation.

\section{Introduction}

Watershed management practices are regularly used for abatement of nonpoint source (NPS) pollution. Yet their effectiveness has been a subject of debate (Arabi et al., 2007; Park et al., 2011). Studies have quantified the effectiveness of NPS conservation practices using monitoring campaigns (Clausen et al., 1996; Bishop et al., 2005; Byers et al., 2005; Line et al., 2016). However, this approach can be costly and the results obtained are often site-specific and inconsistent between studies. Hence, models are increasingly used along with the monitoring data to better assess the effectiveness of Best Management Practices (BMPs; Santhi et al., 2003; Arabi et al., 2006; Lin et al., 2009; Park and Roesner, 2012; Taylor et al., 2016; Jang et al., 2017).

The majority of studies that have investigated the effectiveness of BMPs using models used a deterministic approach which often entails calibrating a watershed model by changing the parameters to obtain a good fit between model simulations and measured observations (Lin et al., 2009; Ulrich and Volk, 2009; Liu et al., 2013; Jingyuan et al., 2014; Motallebi et al., 2017). The calibrated model is then checked against a record of observations from a different time period to perform a so-called validation process. The model is then used to quantify the effectiveness of a specific BMP or a combination of BMPs by changing some parameters that reflect the BMP operation (Arabi et al., 2008; Liu et al., 2013; Jang et al., 2017). This approach can be inadequate, and in many cases misleading, due to lack of accounting for different sources of modeling uncertainties (Ajami et al., 2007; Arabi et al., 2007; Tasdighi, 2017). Propagation of uncertainties from different sources into model predictions during any modeling practice may result in biased and unrealistic decisions.

Models are mere representation of reality which makes them subject to uncertainty. The uncertainty in any modeling practice stems from different sources including: model parameters, input data (climate, land use, etc.), model structure (conceptualization), and measurement data used for training/testing the model (streamflow, nutrient concentrations or loads, etc.). There are many studies that investigate the effects of modeling uncertainties in simulating streamflow or various water quality constituents (Beven and Binley, 1992; Vrugt et al., 2003; Ajami et al., 2007; Rojas et al., 2008; Harmel et al., 2014; Yen et al., 2014). Ajami et al. (2007) developed a Bayesian framework for estimating various sources of uncertainty in simulating streamflow with...
hydrologic models. Yen et al. (2014) also developed a similar framework for propagation of uncertainty from various sources to simulated streamflow and nitrate. While these studies have investigated the effects of modeling uncertainties on streamflow and water quality simulations, the impacts of modeling uncertainty when quantifying the effectiveness of BMPs has not been addressed sufficiently (Arabi et al., 2007). Specifically, a framework that incorporates various sources of modeling uncertainty and determines uncertainty bounds around BMP effectiveness has not been developed to the best of our knowledge.

An important source of uncertainty when assessing the effectiveness of BMPs using a model is the measured data used for training and testing the model. In studies where only water quantity is explored, the uncertainty from streamflow measurements is often the focus while, the uncertainties from measured constituent concentration are also relevant in water quality studies (Harmel et al., 2009). Water quality constituent loads are often derived by multiplying streamflow volume and corresponding constituent concentration, thus uncertainties in streamflow and concentration measurement contribute to uncertainty in load estimates.

Assessing the effectiveness of NPS BMPs often entails a two-stage approach. First, a model is developed for quantifying the NPS pollution. Second, by changing selected model parameters to mimic the operation of BMPs, effectiveness of BMPs is quantified. This two-stage approach makes the uncertainty analysis of BMPs cumbersome and time consuming. Within an uncertainty assessment framework, each model parameter has a probability distribution derived during model training instead of one optimal set of parameters (deterministic approach), and some of these parameters and/or other additional parameters are changed later to reflect the operation of BMPs. Hence, adopting a probabilistic approach for changing these BMP parameters can quickly result in a large number of iterations. In this regard, a technique that can be used to efficiently conduct the BMP uncertainty analysis without compromising the statistical inferences gained during any step of the analysis is essential.

The overall goal of this study was to develop a probabilistic approach to assess the effectiveness of NPS conservation practices in reducing nutrient loads. The specific objectives were to: (i) assess the effects of uncertainty in measured training/testing data on simulating nutrient loads and how it can impact the inferences about effectiveness of BMPs and (ii) determine the total uncertainty bounds (95% confidence interval of model simulations) around the model estimates of BMPs’ effectiveness in reducing nutrient loads. While previous studies have determined the effectiveness of NPS BMPs through monitoring campaigns or deterministic modeling approaches, using a total uncertainty assessment framework (i.e. accounting for various sources of uncertainties including model parameters, input, structure, and observation data) to probabilistically assess the effectiveness of BMPs is novel. Specifically, incorporation of measurement data uncertainty which plays an important role in realistic simulation of NPS nutrient loads can enhance assessing the water quality benefits of BMPs using watershed models. Application of the framework developed in this study can produce useful information for decision-making in water quality improvement through application of NPS conservation practices.

2. Material and methods

Three separate SWAT models were developed. The models were identical except for the rainfall-runoff mechanism (accounting for different model structures) and the precipitation time step accommodating the mechanism. The analysis period was 2008–2015 of which 2008 to 2011 was the pre-BMP and 2012 to 2015 represented the post-BMP period. A 2yr warmup period was used when running the models. A two-stage Bayesian total uncertainty assessment framework was developed and applied to assess the effectiveness of the BMPs (nutrient management and cattle exclusion fencing; described further in 2.7) in reducing daily total nitrogen (TN) loads at the outlet of a 54 ha agricultural watershed in North Carolina.

In the first stage, modeling uncertainties were characterized during the pre-BMP period (2008–2011). A Markov Chain Monte Carlo (MCMC) sampling scheme along with a statistically correct likelihood function which accounts for various statistical conditions including normality, lack of autocorrelation, and heteroscedasticity of error residuals (Stedinger et al., 2008; Vrugt, 2016) set to consider the TN load errors at the outlet of the watershed was used to derive the posterior distribution of model parameters while incorporating various sources of modeling uncertainties.

In the second stage, the effectiveness of BMPs was assessed by comparing the simulated TN loads before and after implementing BMPs during the post-BMP period (2012–2015). The size of the posterior distribution of model parameters was reduced while maintaining their statistical characteristics using a uniform random sampling algorithm combined with a linear interpolation technique. The models were then run using each set of reduced model parameters to generate prediction intervals of TN loads before implementing BMPs. Then for each set of model parameters, BMP parameters were sampled using a uniform random sampling algorithm. The models were run using each combination of model and BMP parameters generating TN load after implementing BMPs. Using the post-BMP field observation data, the value of the likelihood function was also computed for each parameter set. The posterior distributions of BMP parameters were then derived and TN load prediction intervals after implementing BMPs were determined. Finally the TN load reduction prediction intervals were quantified to determine bands of uncertainty around effectiveness of BMPs.

2.1. Study watershed

The study watershed was a 54 ha pasture-dominated watershed located in central North Carolina (Fig. 1). The land use composition within the watershed is 78% pasture, 14% forest, 6% developed, and 2% cultivated crops (National Land Cover Database; NLCD 2011). The watershed was selected due to availability of comprehensive monitoring data (streamflow, nutrients, and sediments) from a published paired watershed study (Line et al., 2016). More importantly, the monitoring dataset provides a record of event-based measurements for 4yr before installing BMPs, and another 4yr after BMPs installation.

The main stream in the watershed known as the Mud Lick Creek flows much of the year except some periods during late summer and early fall where it dries up. In general, the streams in this region are known for low baseflow (Line et al., 2016).

2.2. Watershed 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.

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 Green and Ampt method (G&A; Green and Ampt, 1911) with subdaily rainfall or Curve Number method (CN; USDA-NRCS, 2004) with daily rainfall.
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 (~10 m) resolution digital elevation model (DEM) 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 National Agricultural Statistics Service (NASS) land use data for year 2011 were obtained from the USDA (USDA-CropScape, 2016). The resolution of the SSURGO and NASS data was 1 arc-second (~30 m).

Stream flowlines 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 and stream delineation, the hydrographic segmentation, subwatershed boundary, and stream delineation is improved especially in smaller scales such as the watershed in this study or locations where the DEM does not provide enough accuracy (Winchell et al., 2007).

Observed climate data for the closest meteorological station (Burlington Alamance Regional Airport, GHCND: USW00093783) were obtained from the National Climatic Data Center (NCDC) Quality Controlled Local Climatological Data (QCLCD) database (NOAA, 2016). The station was located about 20 km from the watershed. Daily and hourly precipitation, minimum and maximum temperature were collected for 2005 to 2015.

2.4. Measurements: stream discharge and nutrient concentrations

The monitoring station was located at the outlet of the watershed (Fig. 1). Streamflow and nutrient measurements were available from 2008 to 2015. This period includes pre-BMP (1 Jan. 2008 to 5 Oct. 2011) and post-BMP (6 Oct. 2011 to 31 Dec. 2015) periods. Stream discharge and nutrient concentrations were sampled during storm events. The discharge measurement and nutrient sampling procedures (collecting, storing, and analyzing) are explained in details in Line et al. (2016). Mean daily nutrient loads were computed by multiplying the mean daily storm discharge and corresponding nutrient or sediment concentration.

2.5. Cattle grazing and manure deposition

There were 75 beef cows grazing on various parts of the pasture within the watershed. Based on the observations and communication with land owners, Line et al. (2016) reported a density of 1.2 cows per ha which was used in this study for computing grazing and manure deposition rates. Manure deposition rate for beef cows was set at 29.5 kg/day/500-kg-animal (USDA-NRCS, 2016). Assuming an average weight of a beef cow equal to 600 kg, the average manure deposition rate was computed as 35.4 kg/ha/day. The cattle deposit their manure either on land or in streams. On average, cattle spend 7% of their time in the streams (Byers et al., 2005). Therefore it was assumed that 2.5 kg/ha/day of manure was directly deposited into streams while the remaining fraction (32.9 kg/ha/day) was deposited on the land. Each
beef cow consumes on average 45 kg/ha/day of grass (USDA-NRCS, 2016), which along with the 1.2 cows/ha density gives a grazing rate of about 54 kg/ha/day.

2.6. The SWAT model setup

ArcSWAT 2012 (USDA-ARS, 2014) was used to develop three SWAT models. The models were completely identical except for the runoff estimation method and precipitation time step. A model was developed with daily precipitation and CN method based on (i) soil moisture (CN I) and (ii) plant evapotranspiration (CN II), and the other model was developed using the hourly precipitation and G&A method for runoff simulation. Therefore, three model setups were prepared for the analyses corresponding to three different model structures.

The DEM was used along with the NHD flowlines to delineate the watershed and streams. The watershed was divided into 4 subwatersheds. The HRU definition was based on land use and soil data. Since the topographic variability was small, a single class slope was assumed within each subwatershed. Using these settings, 29 HRUs were defined for the watershed (Fig. 2).

2.7. Representing BMPs in the SWAT model

The BMPs implemented in the watershed were nutrient management and cattle exclusion fencing. While nutrient management can be readily modeled by changing the rate of fertilizer application (Arabi et al., 2008; Ahmadi et al., 2014), incorporating cattle exclusion fencing in the model is more complicated. Few studies have discussed methods for representing cattle exclusion fencing in SWAT. One such method is application of point sources to represent the presence of cattle in streams (Lin et al., 2009). This approach provides the capability to directly change the rate of nutrients or sediments introduced into the streams during the implementation of the exclusion fencing mimicking the operation of the BMP (limiting the access of the cattle to streams).

During the pre-BMP period, 336 kg/ha of 15-15-15 (N-P-K) fertilizer was applied uniformly to the pasture in subwatersheds 2, 3, and 4. Biosolids were applied to the pasture in subwatershed 1. Analysis of biosolids showed that 130 and 116 kg/ha of N and P were applied during each application. The 15-15-15 fertilizer is a standard fertilizer in SWAT databases and hence was readily incorporated in the model. However, for biosolids their elemental N and P content were used to incorporate them in the model. Nutrient management was implemented by replacing the biosolids and fertilizer with a granular N only fertilizer (21-0-0) at a rate of 70 kg/ha during the post-BMP period. This BMP was directly implemented in the SWAT model by replacing and changing the fertilizer application rate.

The cattle exclusion fencing was installed in October 2011. The fences were installed along approximately 520 m of the mainstream of Mud Lick Creek in the watershed. About half of the upper part of the mainstream was not fenced, so cattle were still able to have access to streams in the upper half part of the watershed. We modeled the cattle exclusion fencing by changing the rate of organic N introducing into the
streams via point sources in subbasins 3 and 4 where stream fencing was installed.

2.8. The BMP uncertainty assessment framework

A probabilistic framework was developed for assessing the effectiveness of NPS conservation practices. 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& A), and measured data for model training/testing (i.e. streamflow and nutrient loads). The framework was developed in MATLAB (The MathWorks, Inc.).

The framework was developed building on the work by Ajami et al. (2007) and Yen et al. (2014). Both of these studies have developed methods for estimating modeling uncertainties from various sources in simulating streamflow and/or water quality. However, neither addresses estimating the uncertainties in model predicted BMP effectiveness. In order to quantify the effectiveness of BMPs, the model should be run first before implementing the BMPs in the model. The BMPs are then represented in the model through manipulation of some parameters. The model is run again with the altered parameters to simulate the responses after implementing the BMPs. The difference between the simulated responses (e.g. streamflow, nutrient loads) before and after implementation of the BMPs quantifies their effectiveness.

The framework entails a two-stage procedure. First the modeling uncertainties in simulating nutrient loads are characterized during the pre-BMP period (2008–2011). During this stage, a Markov Chain Monte Carlo (MCMC) sampling scheme, the DREAM method (Vrugt et al., 2009), along with a statistically correct likelihood function (Sorooshian and Dracup, 1980; Stedinger et al., 2008; Vrugt, 2016) which accounts for various statistical conditions including normality, lack of autocorrelation, and heteroscedasticity of error residuals, is used to sample the parameter space and derive the posterior distributions. Input data uncertainty is incorporated by using precipitation multipliers drawn from a Gaussian distribution with an uncertain mean and standard deviation sampled along with model parameters during the MCMC procedure. The measured data uncertainty is also incorporated by applying correction factors on model residuals (residual = observed – simulated) computed based on a probable error range for each measurement (Harmel and Smith, 2007). Finally, Bayesian Model Averaging (BMA) is used to account for model structural uncertainty (Hoeting et al., 1999). Having characterized different sources of modeling uncertainties in the first stage, the model can be used for simulation of nutrient loads during the post-BMP period (2012–2015) without incorporating the BMPs in models. The outputs provide prediction intervals for nutrient loads during the post-BMP period assuming no BMPs were implemented.

The second stage entails iterating on BMP parameters for each model parameter set (from stage 1) to characterize the performance of the BMPs under various sources of modeling uncertainty during the post-BMP period (2012–2015). The first stage often results in too many parameter sets due to large number of iterations required for the MCMC algorithm to converge especially in highly parametrized complex models such as SWAT. A method was proposed to reduce the number of model parameter sets without compromising the statistical characteristics of the inferred posterior distributions. The method involves a uniform random sampling scheme along with a linear interpolation algorithm during which model parameter sets are drawn randomly from the posterior distributions while maintaining statistical characteristics of the distributions. The BMP parameters were then sampled for each set of the reduced model parameter sets using a uniform random sampling scheme. The model is then run using each new joint parameter set (model parameters and BMP parameters) and the posterior distribution of BMP parameters are derived with the new likelihood function during the post-BMP period. The outputs from this stage provide prediction intervals for simulated nutrient loads during the post-BMP period after BMPs are implemented which along with nutrient loads from stage 1 are used to compute prediction intervals for nutrient load reductions (i.e. BMPs effectiveness). Availability of monitoring data during the post-BMP conditions provided a unique opportunity to compute values of a new likelihood function in contrast to previous studies where the same likelihood function was used during both pre-BMP and post-BMP period due to lack of measurements during post-BMP conditions (Arabi et al., 2007).

2.8.1. Model and BMP parameters uncertainty

The generic equation for quantifying the effectiveness of a specific BMP using simulations from a watershed model is:

\[ \eta_{BMP} = \frac{Q_{pre-BMP} - Q_{post-BMP}}{Q_{pre-BMP}} \]  

where \( \eta_{BMP} \) denotes the efficiency of the BMP, and \( Q_{pre-BMP} \) and \( Q_{post-BMP} \) are the simulated response variables before and after application of the BMP respectively. The parameter uncertainty in quantifying the simulated response variables in Eq. (1) is due to (i) model parameters (\( \theta_M \)), and (ii) BMP parameters (\( \theta_{BMP} \)). It should be noted that the uncertainty from model parameters propagates to predictions during both pre-BMP and post-BMP periods. However, the BMP parameter uncertainty is only manifested in the post-BMP simulations. The simulated response variables are also subject to other errors stemming from measured climate inputs (\( R \)), insufficiency of the model conceptualization (\( M \); model structure), and measurement data used for training and testing the model (\( Q_{pre-BMP} \) and \( Q_{post-BMP} \)). The simulation error residuals (\( e \)) then take the form:

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

During the post-BMP conditions, the error residuals are a function of joint probability density of model and BMP parameters, \( e = f(\theta_M, \theta_{BMP}) \), with expected value of:

\[ E(e) = \int \int f(\theta_M, \theta_{BMP}) | f(\theta_{BMP}) | d\theta_M d\theta_{BMP} \]  

where \( E(e) \) denotes the expected value of the error residual as a function of both model and BMP parameters. Eq. (3) is conditioned on the independence of model and BMP parameters. Analytical solution to this equation is often infeasible in case of complex models. Hence, we resort to MCMC methods. Applying the Bayes theorem, the parameter set \( \hat{\theta} \) is assigned posterior probability distribution, \( p(\hat{\theta}|Q) \), which is proportional to the product of the parameter prior probability distribution, \( p(\hat{\theta}) \), and a likelihood function, \( L(\hat{\theta}|Q) \).

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

\[ L(\hat{\theta}|Q, \sigma) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_t^2} \exp\left(-\frac{1}{2\sigma_t^2}(Q_t - \hat{Q}_t(\hat{\theta}))^2\right) \]  

where \( n \) is the number of time steps. 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.

Streamflow and nutrient error residuals are often not independently distributed and in most cases temporal autocorrelation exists in the residuals (Evin et al., 2014). A first order auto-regressive scheme (AR-1) was employed to reduce autocorrelation in error residuals. Applying the natural log and AR-1 transformation, the log-likelihood function takes the form:

\[ l(\hat{\theta}|Q, \sigma, \rho) = -\frac{n}{2} \ln(2\pi\sigma) - \frac{1}{2} \ln \left( \frac{\sigma_t^2}{1 - \rho^2} \right) \]
\[ - \frac{1}{2\sigma_t^2} (1 - \rho^2)\sigma_t^2 + \sum_{i=1}^{n} (e_i - \rho e_{i-1})^2 \]

(5)
Table 1
Model parameters selected for uncertainty analysis in this study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Input file</th>
<th>Description</th>
<th>Unit</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA-BF</td>
<td>.gw</td>
<td>Base flow alpha factor for recession constant</td>
<td>day</td>
<td>0.001</td>
<td>1</td>
</tr>
<tr>
<td>GWHT</td>
<td>.gw</td>
<td>Initial groundwater height</td>
<td>m</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>CANMX</td>
<td>.hru</td>
<td>Maximum canopy storage</td>
<td>mm</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>ESCO</td>
<td>.hru</td>
<td>Soil evaporation compensation factor</td>
<td>–</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>QV-N</td>
<td>.hru</td>
<td>Manning's n value for overland flow</td>
<td>–</td>
<td>0.01</td>
<td>0.6</td>
</tr>
<tr>
<td>SLUBSBN</td>
<td>.hru</td>
<td>Average slope length</td>
<td>m</td>
<td>10</td>
<td>150</td>
</tr>
<tr>
<td>RSDIN</td>
<td>.hru</td>
<td>Initial residue cover</td>
<td>kg ha(^{-1})</td>
<td>0</td>
<td>10000</td>
</tr>
<tr>
<td>CH-KII</td>
<td>.re</td>
<td>Effective hydraulic conductivity in the main channel</td>
<td>mm hr(^{-1})</td>
<td>0.025</td>
<td>150</td>
</tr>
<tr>
<td>CH-NII</td>
<td>.re</td>
<td>Manning's n value for the main channel</td>
<td>–</td>
<td>0.01</td>
<td>0.3</td>
</tr>
<tr>
<td>EPPO</td>
<td>.bsn</td>
<td>Plant uptake compensation factor</td>
<td>–</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>SURLAG</td>
<td>.bsn</td>
<td>Surface runoff lag coefficient</td>
<td>day</td>
<td>0.001</td>
<td>15</td>
</tr>
<tr>
<td>CNCOEF</td>
<td>.bsn</td>
<td>Plant ET curve number coefficient</td>
<td>–</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>CDIN</td>
<td>.bsn</td>
<td>Exponential rate of denitrification</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CMN</td>
<td>.bsn</td>
<td>Rate factor for mineralization of organic nutrients</td>
<td>–</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>NPERCO</td>
<td>.bsn</td>
<td>Nitrogen percolation coefficient</td>
<td>–</td>
<td>0.01</td>
<td>1</td>
</tr>
<tr>
<td>RCN</td>
<td>.bsn</td>
<td>Concentration of nitrogen in rainfall</td>
<td>mg N l(^{-1})</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>RSDCO</td>
<td>.bsn</td>
<td>Residue decomposition coefficient</td>
<td>–</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>SINDCO</td>
<td>.bsn</td>
<td>Denitrification threshold water content</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SOL-AUB</td>
<td>.sol</td>
<td>Fraction change in moist soil albedo</td>
<td>–</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>SOL-AWC</td>
<td>.sol</td>
<td>Fraction change in available soil water capacity</td>
<td>mm mm(^{-1})</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>mm hr(^{-1})</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>mm</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>BIOMIN</td>
<td>.mgt</td>
<td>Minimum biomass for grazing</td>
<td>kg ha(^{-1})</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>BIOMIX</td>
<td>.mgt</td>
<td>Biological mixing efficiency</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CN-FF</td>
<td>.mgt</td>
<td>Fraction change in CVS runoff curve number</td>
<td>–</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>ORGN</td>
<td>.chm</td>
<td>Initial organic N in soils</td>
<td>kg N ha(^{-1})</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>SOLN</td>
<td>.chm</td>
<td>Initial NO(_3) in soils</td>
<td>mg N kg(^{-1})</td>
<td>0.1</td>
<td>5</td>
</tr>
</tbody>
</table>

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

The model 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 1 lists the model parameters selected for uncertainty analysis along with their ranges. The ranges for parameters were also selected based on the SWAT user manual and experience from previous study (Tasdighi et al., 2017; Arnold et al., 2012). 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 BMP parameters for uncertainty analysis were selected based on the SWAT capabilities in representing BMPs and studies where similar BMPs were implemented in SWAT (Arabi et al., 2008; Lin et al., 2009; Ahmadi et al., 2014). Table 2 lists the parameters selected for uncertainty analysis of BMPs. Similar to model parameters, uniform (noninformative) prior distributions were assumed for BMP parameters as well. The rate of fertilizer application was selected to represent nutrient management practices. The rate of organic N introducing into streams via point sources in subbasins 3 and 4 where cattle exclusion fencing was implemented were selected as parameters for simulating exclusion fencing in SWAT.

2.8.2. Model input uncertainty

A method proposed by Ajami et al. (2007) was implemented to account for precipitation uncertainty through application of multipliers on precipitation events. 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:

\[
\hat{\theta} = \hat{\theta} \cdot \hat{\phi}; \quad \hat{\phi} \sim N(\hat{\mu}_\phi, \sigma^2_\phi)
\]

where \(\hat{\theta}\) and \(\hat{\phi}\) are the corrected and observed precipitation depths respectively, \(\hat{\phi}\) is the random multiplier drawn from a normal distribution with a random mean \(\hat{\mu}_\phi\), \(\hat{\mu}_\phi \in [0.9, 1.1]\) and variance, \(\sigma^2_\phi\), \(\sigma^2_\phi \in [1e-5, 1e-3]\) (Ajami et al., 2007). Incorporating precipitation multipliers using this approach instead of iterating on each precipitation multiplier reduces the dimensionality issue and improves the identifiability.

2.8.3. Model structural uncertainty

Bayesian Model Averaging (BMA) was used to account for model structural uncertainty (Hoeting et al., 1999; Georgakakos et al., 2004). The BMA is a probabilistic algorithm for combining competing models based on their predictive skills (Ajami et al., 2007; Madadgar and Moradkhani, 2014). In this study, the three rainfall-runoff model structures in SWAT (M\(_1\), CN \(_I\), M\(_2\), CN \(_II\), M\(_3\); GkA) were used to explore the effects of model structural uncertainty. Using the BMA, the three model structures were combined using probabilistic weights to reduce the model structural uncertainty. The posterior distribution of the BMA prediction (\(\hat{\theta}_{BMA}\)) is:

\[
p(M|\hat{\theta}_{BMA}) = \sum_{i=1}^{3} [p(M|\theta_i) \times p(\hat{\theta}|M_i, Q)]
\]

where \(p(M|\theta_i)\) is the posterior probability of the model \(M_i\). 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: \( \sum \omega_i = 1 \). Higher values of \( \omega_i \) 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) was used in this study to estimate model weights.

2.8.4. Incorporating the uncertainty in measured data for model training/testing

The uncertainty inherent in measured data used for training and testing of models often stems from errors in monitoring design, instrumentation, data processing, storage, and operator (human) errors (Harmel et al., 2009). This type of uncertainty is rarely accounted for in evaluation of model performance (Harmel and Smith, 2007; Yen et al., 2014). The measured data uncertainty is often manifested in heteroscedasticity of error residuals (variable error variance). Some observations may be less reliable than others which results in their errors to have different variances. Different approaches have been proposed to circumvent this issue including application of different transformations such as natural log or Box-Cox to stabilize the error variances (Sorooshian and Dracup, 1980). Others have proposed alternative forms for the likelihood function (Schoups and Vrugt, 2010) which accounts for error heteroscedasticity by assuming a variable variance for each model residual. The main challenge in this approach is determination of the variance for each residual as it requires having repeats of each measurement which most often are not available. As a result, in such studies, variance was subject to inference along with model parameters at each model realization during the sampling procedure (Schoups and Vrugt, 2010; Vrugt, 2016). This approach can lead to dimensionality issues in case of highly parameterized models or when a relatively long record of measurements is used for model training.

In this study, we employed a method based on the study by Harmel and Smith (2007). In this method, each error residual is modified using a correction factor computed based on the properties of the probability distribution of each measured value. Previous experience and expert’s opinion can be used to make informed assumptions about probability distribution for each measured value in the record. Assuming a normal distribution for each measured value (\( Q_i \)), mean and median of the distribution is represented by \( Q \). The variance can then be computed based on a probable error range (PER) which can be assumed based on literature or professional judgment. PER can be constant or variable for all measurements depending on the experience and level of knowledge about the monitoring design. Once the PERs are determined, the variance \( (\sigma_i)^2 \) for each record of measurement can be computed as:

\[
(\sigma_i)^2 = \frac{\text{PER}_i \times Q_i}{3.9 \times 100}^2
\]

(8)

Eq. (8) is useful when the measured values are not transformed. Since we used a natural log transformation on streamflow and nutrient loads, the Taylor series expansion for the second moment of a function of random variable \( f(X) \) was used to estimate the variance \( \text{Var}(f(X)) \) as:

\[
\text{Var}(f[X]) = (f’(E[X]))^2 \text{Var}[X] = (f’(\mu_c))^2 \sigma^2
\]

(9)

Applying (9) on the natural log-transform function gives:

\[
\text{Var}[\ln(X)] \approx \left( \frac{(\sigma_X)}{x} \right)^2
\]

(10)

\[
\text{SD}[\ln(X)] \approx \frac{\sigma_X}{x} = \text{COV} = \text{PER}
\]

(11)

where SD stands for the standard deviation and COV denotes the coefficient of variation. Using this approach, \( \sigma \) for log-transformed observations, can be estimated with PER. It should be noted that this assumption holds for smaller PERs (<0.3) and higher approximation errors are introduced in higher values of PER. We used expert’s opinion to determine PERs for each streamflow or nutrient load measurement \( Q \). The correction factors were then computed as the area under the standard normal distribution:

\[
\begin{align*}
CF_i &= F(X|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{\mu_i - \frac{\sigma_i^2}{2}} e^{-\frac{x^2}{2\sigma_i^2}} \, dx - 0.5 \quad \text{if } \mu_i \leq X \\
CF_i &= F(X|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \int_{\mu_i + \frac{\sigma_i^2}{2}}^{\infty} e^{-\frac{x^2}{2\sigma_i^2}} \, dx + 0.5 \quad \text{if } \mu_i > X
\end{align*}
\]

(12a)

(12b)

where \( F(\cdot) \) denotes the normal cumulative distribution function, \( X \) denotes the simulated data (\( \hat{Q} \)) and \( \mu \) represents the observed data (\( Q \)) and \( \sigma \) is determined using Eq. (11). Using the correction factors, the residuals are adjusted based on the estimated measured data uncertainty. It should be noted that this method requires some level of knowledge about the characteristics of the measurement errors. In this case, this information was obtained using the description of measurement techniques in the study watershed by Line et al. (2016). These characteristics are often case-specific. Experience form previous studies can be used to enhance the assumptions and generate better estimates.

2.8.5. 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 Shuffled Complex Evolution Metropolis (SCEM-UA; Vrugt et al., 2003), and the DiffRental 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.

2.9. Evaluation of model performance

The performance of the models in simulating daily TN loads under the total uncertainty assessment framework was evaluated using different error statistics computed at each model realization. During the pre-BMP period (2008–2011), the uncertainty analysis was performed using the values of the likelihood function (Eq. (5)) as the objective function computed with daily TN loads at the monitoring station. The error statistics during the post-BMP (2012–2015) period were calculated using the joint distribution of reduced model parameters and BMP parameters.

3. Results and discussion

3.1. Evaluation of models during the pre and post-BMP periods

Table 3 summarizes the error statistics for models during the pre and post-BMP conditions. Compared to models with CN, the model with G&A method had a poor performance in terms of various error statistics during the pre-BMP period and it was excluded from further analysis for that reason. Similar results were reported on better performance of the CN method over the G&A in other agricultural watersheds (Kannan et al., 2007; Cheng et al., 2016). In contrast, Ficklin and Zhang (2013)
concluded that models with G&A are more likely to generate better daily simulations in agricultural watersheds. It should be noted that these studies have used a deterministic approach. Tasdighi et al. (2018) compared the performance of the CN and G&A methods based on upstream land use conditions using a probabilistic approach. They concluded that the G&A method had a better performance in highly developed subwatersheds while the CN method had a slightly better performance in agricultural watersheds.

Between models with CN I and CN II methods, CN II models had a slightly better performance during the pre-BMP period. During the post-BMP period the superiority of the CN II model was more accentuated generating better error statistics.

3.2. Characterizing the modeling uncertainties during the pre-BMP period (stage 1)

3.2.1. Model parameters uncertainty

The posterior Cumulative Distribution Functions (CDF) of parameters under the CN I and CN II model structures are illustrated in Fig. 3. The distributions are derived after the pre-BMP uncertainty analysis (2008–2011). Note that only the most sensitive parameters are included in this figure. As observed in Fig. 3, models with CN I and CN II resulted in different distributions for the same parameters. While the model structure determined the level of skewness for most parameters, for some parameters (ESCO, SOLZ, CDN) the skewness changed from positive to negative under different model structures. In general, CN I model showed higher sensitivity to parameters pertaining to soil characteristics (SOL-ALB, SOL-AWC, and SOL-Z) which conforms to intuition as the CN I method uses the soil water content for determining characteristics (SOL-ALB, SOL-AWC, and SOL-Z) which conforms to the CN I and CN II methods. The results showed that at the 0.05 significance level, BMA weights for CN II model were significantly higher than weights for CN I model. These results are congruent to the results in section 3.1 where CN II model showed superior performance in terms of various error statistics. One explanation for these results is that in CN II method, the curve number is determined for these results is that in CN II method, the curve number is determined based on plant evapotranspiration, and since the study watershed is a pasture-dominated agricultural watershed, the CN II has a better performance.

The distributions were derived using 20000 samples after the convergence of the DREAM algorithm. The uniform random sampling algorithm along with the linear interpolation technique discussed in section 2.8 was then used to reduce the number of model parameter sets while maintaining the statistical characteristics of the distributions to be used for post-BMP analysis. The algorithm was set to draw 1000 parameter sets from the posterior distributions. The reduced parameter sets act as priors for the post-BMP period. In other words, the prior distributions of model parameters for the post-BMP period are inferred from the posterior distribution of parameters from the pre-BMP period. For each of the new model parameter sets, the uniform random sampling technique discussed in section 2.8 was used to generate BMP parameter sets. Since only 3 parameters were used for BMPs, 30 random BMP parameter samples were assumed to adequately represent the BMP parameter space. The BMP parameter sets combined with the model parameter sets resulted in 30000 joint parameter sets for the second stage of the analysis (post-BMP).

3.2.2. Model input uncertainty

The posterior CDFs of mean and standard deviation for normal distributions from which precipitation multipliers were drawn, were almost similar and close to uniform. The uniformity of the distributions indicates that precipitation multipliers did not have a major impact on TN load simulations from each model. The similarity of the distributions on the other hand, could be predicted as the daily precipitation data for models with CN I and CN II were identical. Hence, any difference in distributions resulted from the model structural difference. Another explanation in this regard is the large number of sensitive model parameters which can diminish the impact of precipitation multipliers in the range assumed. Using a wider range for mean and standard deviation of the normal distributions from which precipitation multipliers were drawn could result in higher impact from multipliers and probably better performance of the input uncertainty estimation routine.

3.2.3. Model structural uncertainty

Bayesian model averaging was used at each model realization to probabilistically combine the models and reduce the model structural uncertainty. Fig. 4 shows the boxplots of BMA weights generated during the MCMC procedure. Based on BMA weights, CN II had a better performance in simulating daily TN loads. A two-sample t-test was performed to test the significance of the difference between the distribution of the BMA weights from the CN I and CN II models. The results showed that at the 0.05 significance level, BMA weights for CN II model were significantly higher than weights for CN I model. These results are congruent to the results in section 3.1 where CN II model showed superior performance in terms of various error statistics. One explanation for these results is that in CN II method, the curve number is determined based on plant evapotranspiration, and since the study watershed is a pasture-dominated agricultural watershed, the CN II has a better performance. Similar results were obtained in other studies (Tasdighi et al., 2018; Yen et al., 2014).

3.2.4. Measured training/testing data uncertainty

Measured data uncertainty was incorporated using correction factors. The correction factors were applied on the daily TN load residuals during the computation of the likelihood function at each model realization. Fig. 5 illustrates the probability distribution (PDF) of correction factors for each model during the pre-BMP period.

The correction factors were categorized based on flow conditions (low, medium, and high flows) before generating the PDFs to assess the effects of flow regime on measurement errors. The highest values of correction factors were observed during high flow events. Medium and low flow events resulted in relatively lower values for correction factors. This is an important finding as it indicates higher errors during high flow events compared to medium and low flows. The phenomenon of higher errors during high flow events is often described as the...
heteroscedasticity of error residuals which is deemed to be attenuated by applying correction factors. Other studies that have investigated the uncertainties in measurement data have reported similar behavior (Sorooshian and Dracup, 1980; Harmel and Smith, 2007). The results also conform to intuition as monitoring during high flow events often entails higher errors due to difficulties in measurements (Harmel et al., 2006).

3.3. Assessing the effectiveness of BMPs under various sources of modeling uncertainty (stage 2)

3.3.1. BMP parameters uncertainty

The posterior distribution of BMP parameters are illustrated in Fig. 6. Interestingly, the parameter pertaining to nutrient management BMP (FRT-KG) showed higher sensitivity when quantifying the TN loads. This is while the posterior distributions of parameters pertaining to the cattle exclusion fencing were close to uniform which indicates lower sensitivity. This outcome also indicates that the nutrient management had a higher impact on nutrient load reductions. A possible explanation for this can be the higher uncertainty in the nature of representing cattle exclusion fencing in the SWAT model. Several assumptions were used when representing the exclusion fencing such as changing the nutrient introducing into the streams via point sources and the rate adjustments for nutrients introducing into streams. A more rigorous approach for representing cattle exclusion fencing may enhance the performance of this BMP and result in more meaningful inferences. All BMP parameter posterior distributions showed high deviations from normality. Assuming wider prior distributions and larger number of iterations on BMP parameters may be effective in reducing these effects.

3.3.2. Estimating TN load prediction intervals before and after implementation of BMPs

The cumulative exceedance probability curves for daily TN loads were developed. These curves along with bands of uncertainty before and after implementing BMPs provide an easily-readable informative measure for assessing the effectiveness of BMPs in reducing TN loads under uncertainty. Fig. 7 illustrates the 95% confidence interval for cumulative exceedance probability curves for TN loads.

Comparing the prediction intervals before and after implementation
of BMPs, it is evident that the combination of BMPs (nutrient management and cattle exclusion fencing) was successful in reducing the TN loads in the watershed. Compared to CN I, the model with CN II had a better performance in capturing observed TN loads at all ranges especially the high and medium loads. However, they showed wider bands of uncertainty. The BMA had a close performance to the CN II model which was expected as the CN II method better simulated the TN loads resulting in higher BMA weights.

While the cumulative exceedance probability curves provide valuable insights into the statistical characteristic of BMPs’ effectiveness, they can be misleading too as the serial structure and autocorrelation of the sequence of the simulated and observed records are removed in them (Vogel and Fennessey, 1994). In this regard, the fraction of observations lying within the prediction intervals should be determined using the time series of simulations and observations. For this reason, 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) of simulations were determined based on the time series of simulations and observations of TN loads. Table 4 summarizes the coverage and spread for the models during the pre and post-BMP conditions.

### 3.3.3. Quantifying the effectiveness of BMPs in terms of TN load reductions

The effectiveness of BMPs in reducing the TN loads was computed by subtracting the TN loads from models before and after implementation of BMPs during the common post-BMP period (2012–2015). Fig. 8 depicts the cumulative exceedance probability curves for TN load reductions under different models. The highest reductions were observed for higher loads. In general, the results demonstrate high level of uncertainty in simulating the daily TN load reductions. For example for CN II model in Fig. 8, at exceedance probability of 25%, the TN load reduction from BMPs can be any number between 0.1 and 1 kg/ha. This outcome indicates the importance of accounting for various sources of uncertainty in modeling the performance of the BMPs as they directly affect the decision-making process. In general, modeling pollution loads from NPSs is subject to high levels of uncertainty. However, most often this uncertainty is ignored and models are used deterministically to compute pollution loads which can result in unrealistic and biased decisions.

The average annual TN load reduction from the ensemble of CN I, CN II models, and BMA were 1.5, 1.7, and 1.8 kg/ha respectively. In

![Fig. 4. BMA weights for the models (Solid horizontal lines in 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). 'Best Solution' corresponds to the BMA weight for the simulation that resulted in highest likelihood function.](image)

![Fig. 5. Histogram of TN load correction factors for CN I and CN II models under low, medium, and high flows conditions.](image)
terms of percentage, the average annual reductions were 58%, 65%, and 69% respectively. Fig. 9 shows the PDFs of average annual load reductions for each model. These values conform to the reductions determined by the paired watershed study on this watershed conducted by Line et al. (2016). They found statistically significant reductions in total Kjeldahl N (34%), and ammonia-N (54%) while changes in nitrate-N loads were not significant.

The TN load reductions were resulting from the performance of the nutrient management and cattle exclusion fencing combined. While it was not feasible to decompose the total TN load reductions between each BMP to determine the performance of each BMP, the nutrient management was determined to be more effective as the related BMP parameters showed higher sensitivities when quantifying TN loads.

### 4. Conclusions

A total uncertainty estimation framework was developed for assessing the water quality benefits of management practices using watershed models. The two-stage framework first characterizes various sources of modeling uncertainties during the pre-BMP period. The second stage of the framework uses the inferences on modeling uncertainties from the first stage and quantifies the effectiveness of BMPs using a probabilistic approach.

In general, the modeling uncertainties were large resulting in wide bands of uncertainty around both TN loads and load reductions. The framework however was successful in capturing the effects of different sources of modeling uncertainties on simulations and propagating them to the BMP effectiveness assessment stage. Between the three model structures (CN I, CN II, and G&A), CN II showed the best performance in terms of various performance measures including error statistics and BMA weights. This was attributed to the intensive agricultural land use in the watershed. The G&A method had an unsatisfactory performance in simulating the TN loads. Application of the BMA, slightly enhanced the quality of simulating TN load reductions under uncertainty as it resulted in higher coverage, 51% and 35% during the pre and post-BMP periods respectively. The distribution of the correction factors for measurement uncertainty indicated higher uncertainty for high flow events correctly capturing the heteroscedasticity of error residuals for daily TN load simulations.

Between the two BMPs, nutrient management had the highest impact on the TN load reductions. The parameters pertaining to cattle exclusion fencing did not show much sensitivity when quantifying daily

### Table 4

Coverage rates and spread for models. The coverage rates and spread are calculated using the 95% confidence interval of simulated TN loads and corresponding observations.

<table>
<thead>
<tr>
<th>Model</th>
<th>Pre-BMP</th>
<th>Post-BMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coverage (%)</td>
<td>Spread (kg ha(^{-1}))</td>
</tr>
<tr>
<td>CN I</td>
<td>47</td>
<td>1.05</td>
</tr>
<tr>
<td>CN II</td>
<td>41</td>
<td>1.30</td>
</tr>
<tr>
<td>BMA</td>
<td>51</td>
<td>1.45</td>
</tr>
</tbody>
</table>

Fig. 6. The posterior distribution of BMP parameters. ORGCNST (sub3) and ORGCNST (sub4) are the parameters pertaining to cattle exclusion fencing in sub-watersheds 3 and 4 respectively. FRT-KG is the BMP parameter for nutrient management.

Fig. 7. 95% confidence intervals for cumulative exceedance probability of TN loads before and after implementation of BMPs along with observed TN loads.
TN loads. A possible explanation for this is the higher uncertainty in the nature of representing cattle exclusion fencing in SWAT. Several assumptions were used when representing the exclusion fencing such as changing the nutrient introducing into the streams via point sources and the rate adjustments for nutrients introducing streams. It should be noted that the intent of this study was to develop a framework for probabilistic assessment of BMP effectiveness in reducing pollutants in streams and not to develop a method to represent specific BMPs in models.

The results of this study have important implications for decision-making when models are used for water quality simulation. While numerous uncertainty analysis frameworks have been developed to explore modeling uncertainties in quantification of streamflow and water quality components, the lack of pragmatic applications of such methods to tackle decision-making challenges is a major shortcoming. The framework presented in this study is deemed a pioneer attempt to fill this gap in the literature and add to the pragmatic aspects of the uncertainty analysis in hydrologic and water quality simulations.

**Software availability**

All the codes are developed within MATLAB and can be made available upon request from the first author.

**Acknowledgements**

This publication was made possible by U.S. EPA grant RD835570. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the U.S. EPA. Further, U.S. EPA does not endorse the purchase of any commercial products or services mentioned in the publication. USDA is an equal opportunity employer and provider.