Generalized Likelihood Uncertainty Estimation and Markov Chain Monte Carlo Simulation to Prioritize TMDL Pollutant Allocations

Anurag Mishra[1]; Ebrahim Ahmadisharaf, A.M.ASCE[2]; Brian L. Benham[3]; Mary Leigh Wolfe[4]; Scotland C. Leman[5]; Daniel L. Gallagher, M.ASCE[6]; Kenneth H. Reckhow[7]; and Eric P. Smith[8]

Abstract: This study presents a probabilistic framework that considers both the water quality improvement capability and reliability of alternative total maximum daily load (TMDL) pollutant allocations. Generalized likelihood uncertainty estimation and Markov chain Monte Carlo techniques were used to assess the relative uncertainty and reliability of two alternative TMDL pollutant allocations that were developed to address a fecal coliform (FC) bacteria impairment in a rural watershed in western Virginia. The allocation alternatives, developed using the Hydrological Simulation Program—FORTRAN, specified differing levels of FC bacteria reduction from different sources. While both allocations met the applicable water-quality criteria, the approved TMDL allocation called for less reduction in the FC source that produced the greatest uncertainty (cattle directly depositing feces in the stream), suggesting that it would be less reliable than the alternative, which called for a greater reduction from that same source. The approach presented in this paper illustrates a method to incorporate uncertainty assessment into TMDL development, thereby enabling stakeholders to engage in more informed decision making. **DOI: 10.1061/(ASCE)HE.1943-5584.0001720.**

Author keywords: Generalized likelihood uncertainty estimation; Markov chain Monte Carlo; Water quality modeling; Total maximum daily load (TMDL); Uncertainty analysis; Non-point-source pollution management.

Introduction

A total maximum daily load (TMDL) specifies the amount of a particular pollutant that a water body can receive and still meet water quality standards. Section 303(d) of the Clean Water Act (US Congress 1972) requires that TMDLs be developed for impaired water bodies. Development of a TMDL often includes the application of computational, watershed-scale water quality models. According to the TMDL Analysis and Modeling Task Committee (2017), commonly used models include the Soil and Water Assessment Tool (SWAT) (Neitsch et al. 2011) and the Hydrological Simulation Program—FORTRAN (HSPF) (Bicknell et al. 2005).

Uncertainty in water quality modeling may arise from several sources, including the spatiotemporal resolution of input and forcing data (e.g., meteorologic, hydrologic, geologic, geomorphologic, and pedologic), model structure and parameters, observed water quality and hydrologic data, as well as the choice of calibration and validation assessment and approach (Beck 1987; Harmel et al. 2006; Harmel and Smith 2007; Nguyen and Willems 2016; Shirmohammadi et al. 2006; van Straten 1998; Willems 2008). Without a rigorous assessment of uncertainty, analysts cannot reliably assess the probability of achieving, or the risk of exceeding, a given water quality criterion. A margin of safety (MOS) is included in TMDLs to account for uncertainty and to mitigate the risk of not achieving the applicable water quality criterion. The inclusion of a MOS is often done arbitrarily by making conservative assumptions or adding a fraction of the TMDL (TMDL Analysis and Modeling Task Committee 2017; Dilks and Freedman 2004; Reckhow 2003; Shirmohammadi et al. 2006) rather than explicitly estimating the uncertainty. The NRC (2001) encouraged the USEPA to establish a more rigorous, less subjective approach to assess TMDL development uncertainty (Ames and Lall 2008; Hantush and Chaudhary 2014). Failure to do so could result in wasting limited resources and not achieving water quality goals. Although researchers, such as Rode et al. (2010), have suggested that novel techniques are available to rigorously evaluate uncertainty in water-quality modeling efforts, the application of such techniques is rare.

The HSPF is frequently used to develop TMDLs (Singh and Woolhiser 2002). Although widely used, there have been limited documented attempts to assess the uncertainty associated with using a complex, computationally intensive, watershed-scale model like HSPF. Paul et al. (2004) and Wu et al. (2006) used first-order error analysis (FOEA) to analyze the uncertainty associated with a HSPF application that simulated in-stream water quality

[1]Senior Environmental Engineer, RESPEC Consulting and Services, 2672 Bayshore Pkwy., Suite 915, Mountain View, CA 94043.

[2]Postdoctoral Associate, Dept. of Biological Systems Engineering, Virginia Tech, 155 Ag Quad Ln., 400 Seitz Hall, Blacksburg, VA 24061 (corresponding author). ORCID: https://orcid.org/0000-0002-9452-7975. Email: eahmadis@vt.edu

[3]Extension Specialist and Professor, Dept. of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061.

[4]Professor, Dept. of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061.

[5]Associate Professor, Dept. of Statistics, Virginia Tech, Blacksburg, VA 24061.

[6]Professor, Dept. of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061.

[7]Emeritus Professor, Nicholas School of the Environment, Duke Univ., Durham, NC 27708.

[8]Professor, Dept. of Statistics, Virginia Tech, Blacksburg, VA 24061.

Note. This manuscript was submitted on August 30, 2017; approved on July 10, 2018; published online on October 13, 2018. Discussion period open until March 13, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

constituents. Wu et al. (2006) and Mitsova-Boneva and Wang (2007) applied Monte Carlo simulation, while Jia and Culver (2008) applied generalized likelihood uncertainty estimation (GLUE) (Beven and Binley 1992). Chin (2009) used a stochastic analysis of model residuals for uncertainty analysis of HSPF. Despite these limited efforts, application of Bayesian techniques such as GLUE and Markov chain Monte Carlo (MCMC) (Kass et al. 1998) for uncertainty analysis of watershed-scale water-quality modeling is rare. Bayesian inference, along with MCMC, provides a robust framework for quantifying the uncertainty within complex mathematical models parameterized with observations, either stochastically or deterministically. An important feature of Bayesian methods that makes them particularly advantageous in water-quality modeling applications is their independence from the degree of nonlinearity of a given model (Camacho et al. 2015). Examples of applications of Bayesian techniques to other water quality models (e.g., SWAT) are described by Shirmohammadi et al. (2006), Zheng and Keller (2007), Gardner et al. (2011), Abbaspour (2015), Zheng and Han (2016), and Camacho et al. (2018).

Uncertainty analysis provides stakeholders with additional information to compare the risk and reward of alternative TMDL pollutant allocations (Jia and Culver 2008). However, in practice, clearly communicating the risks to promote informed decision making can be challenging (Stow et al. 2007). Ocampo-Duque et al. (2013) and Xie and Huang (2014) pointed out the need for developing efficient methods to communicate the nature and implications of water-quality modeling uncertainty to decision makers and stakeholders.

The primary objective of this research is to present a probabilistic framework that considers both the water quality improvement capability and reliability of alternative TMDL pollutant allocations. A further objective is to demonstrate the application of two uncertainty analysis techniques, GLUE and MCMC, to a riverine fecal coliform (FC) TMDL developed previously (Benham et al. 2004). These two Bayesian techniques use observed data to refine input parameter prior to probability density functions (PDFs), thereby improving watershed model simulation capability. To illustrate the importance of incorporating uncertainty assessment into TMDL allocation analysis, the results of this research are compared with a TMDL developed using a deterministic analysis (Benham et al. 2004). The presented framework provides a well-informed routine to prioritize TMDL allocation.

Methodology

Generalized Likelihood Uncertainty Estimation

The basic premise of GLUE is that there is not a single optimal set of parameters for any given model (i.e., equifinality), so multiple sets of parameters can be used to satisfactorily represent a watershed response (Beven and Binley 1992). In GLUE, Monte Carlo simulation is used by generating multiple sets of model parameters from parameter-specific PDFs that are determined a priori. In the majority of previous GLUE applications, a priori or prior parameter-specific PDFs were uniform. When using GLUE, the model is run using a set of parameters or a parameter vector θ sampled from prior parameter-specific PDFs. A likelihood weight is calculated for each parameter vector. Likelihoods can be computed using any goodness-of-fit metric that compares observed data and model-simulated response variables (Stow et al. 2007). For this application, likelihoods were a function of the variance of the residuals [Eq. (1)]. Hydrologic calibration compared daily flow volume to assess model performance, while water quality calibration compared log-transformed simulated daily average and observed instantaneous FC concentrations. In-stream FC concentration can generally vary by orders of magnitude and, therefore, log transformation reduced the influence of extreme FC concentrations on estimated uncertainty

$$L_e = (\sigma_e^2)^{-N}$$
$$\sigma_e^2 = \frac{1}{n}\left(\sum_{i=1}^{n}[Y_i - Q_i]^2\right) \tag{1}$$

where L_e is the likelihood; σ_e^2 = variance of the residuals or mean square error; n = number of data points; Y_i = observed data; Q_i = simulated data; and N is the shaping parameter chosen by the user. Eq. (1) has been used frequently with other GLUE applications, such as Beven and Binley (1992). As the value of N increases, the magnitude of difference between the likelihood values of parameter sets with similar variance increases. Different values of N can lead to different confidence intervals (Ratto et al. 2001). This application used an N of 2 because the resulting model outputs bracket the observations (Mishra 2011).

All the parameter vectors that yield acceptable likelihood values, as determined subjectively by the analyst, are considered to be *behavioral* and are retained. Parameter vectors that produce unacceptable simulations are deemed *nonbehavioral* and are rejected. Previous applications of GLUE have reported a wide variety of parameter vector rejection criteria. For example, Beven and Binley (1992) considered parameter vectors with very low likelihood values to be nonbehavioral. In this study, the likelihood threshold used to differentiate behavioral and nonbehavioral parameter vectors was the inflection point on a cumulative distribution function (CDF) plot of likelihoods. The likelihood values of the retained parameter vectors are normalized to unity. The normalized likelihood values can be treated as a probabilistic weighting function for the simulated variables and can be used to assess the uncertainty associated with the simulations. The parameter-specific plots of the normalized behavioral likelihoods versus parameter values, also called dotty plots, defined the cumulative probability and posterior parameter distribution (Beven and Binley 1992). The normalized likelihoods for different parameter values are multiplied to the prior probability to calculate the posterior distribution of the model input parameters. As demonstrated in this application, the posterior distribution of parameters can be used to propagate uncertainty in the model outputs. The posterior PDFs provide a description of parameter uncertainty informed by the observed data, thereby leading to a better estimation of model output uncertainty.

Markov Chain Monte Carlo

MCMC is a broad class of algorithms which are used for sampling from complex distributions. To quantify the uncertainty in these parameter estimates, this study sampled from the Bayesian posterior $[\pi(\theta|Y)]$ distribution of the form

$$\pi(\theta|Y) = \frac{L(Y|\theta)\pi(\theta)}{\int L(Y|\theta)\pi(\theta)d\theta} \tag{2}$$

where $L(Y|\theta)$ is the likelihood function that expresses the dependence between observed data and unknown parameters θ; and $\pi(\theta)$ denotes the prior distribution. Similar to GLUE, the MCMC likelihood is based on the squared deviation between the observed data (Y_i) and the simulation model (Q_i). Because the unknown, estimable parameters (θ) control the physical specifications of the simulations, $Q_i = Q(Y_i|\theta)$ is used in Eq. (3) for clarity. Under this notation, likelihood can be expressed as

$$L_e(Y|\theta) = \frac{1}{(\sqrt{2\pi})^n \sigma_e^n} e^{\left(-(1/2\sigma_e^2)\sum_{i=1}^n (Y_i - Q(Y_i|\theta))^2\right)} \tag{3}$$

Eq. (3) assumes that the residuals between observed and simulated values are normally distributed. The resulting posterior distribution has this same form, up to proportionality, because $\pi(\theta) \propto 1$. Given the probability models sampling from the posterior requires an iterative algorithm, the Metropolis–Hastings algorithm (Hastings 1970), one of the most general MCMC variants, samples from $\pi(\theta|Y)$ by initializing $\theta^{(0)}$ and iterating over $(i = 1, \ldots, k)$:

1. Propose a new candidate θ: $\theta^{(*)} \sim g(\theta|\theta^{[i-1]}) \equiv \text{Normal}(\theta^{[i-1]}, s^2)$
2. Compute acceptance probability: $\alpha = \min\left(1, \frac{\pi[\theta^{(*)}|Y]}{g(\theta^{(*)}|\theta^{[i-1]})} \frac{g(\theta^{[i-1]}|\theta^{(*)})}{\pi[\theta^{[i-1]}|Y]}\right)$
3. Update sample: $\theta^{(i)} = \begin{cases} \theta^{(*)} & \text{with probability: } \alpha \\ \theta^{(i-1)} & \text{with probability: } 1-\alpha. \end{cases}$

The variable α is the acceptance probability for transitioning from our previous iteration to our proposed state and can take on any value between 0 and 1. Given that the proposal density function $g(\theta|\theta^{[i-1]})$ can be used to generate samples using $\pi(\theta|Y)$, this algorithm can be used to produce a sequence of random samples $(\theta^{[1]}, \theta^{[2]}, \ldots, \theta^{[k]})$. Eventually, after a sufficient number of iterations (denoted B for the burn-in period), these samples $(\theta^{[B+1]}, \theta^{[B+2]}, \ldots, \theta^{[k]})$ will be distributed according to the stationary target $\pi(\theta|Y)$. Note that in this application, θ is a correlated, multidimensional parameter vector. Due to the correlations of parameters within the vector θ, attempting to update every element of this vector for each iteration of the Metropolis–Hastings algorithm requires a large number of iterations for convergence. Therefore, in order to improve efficiency, this study applied the Metropolis within Gibbs (Gelfand and Smith 1990; Geman and Geman 1984) algorithm, which sequentially updates only single elements within θ at each iteration. Hence, for each scalar quantity, parameter update, the proposal density function $(g[\cdot \mid \cdot])$ is Gaussian, centered at the previous sample with the tunable scaling factor s, a Gaussian random walk. Choosing too large or too small of s will result in poor convergence. This study used the Gelman–Rubin statistic (Gelman et al. 2014) to tune and assess Markov chain convergence. The statistic is based on generating multiple Markov chains and calculating the mixture of chain variance, within chain variance and Gelman–Rubin statistic. The Gelman–Rubin statistic is the ratio of the mixture of chain variance and within chain variance multiplied by a correction factor. The Markov chain is considered to be converged when these variances stabilize and the Gelman–Rubin statistic approaches unity. The Gelman–Rubin statistic is useful in determining the convergence of a MCMC algorithm by monitoring the variances, both between and within m independent Markov chains. Letting B and W denote between and within chain variances for m independent Markov chains, each of length k, these variances are computed as

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\overline{\theta_j} - \bar{\theta})^2$$

$$W = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{i=1}^{k} (\theta_{ij} - \overline{\theta_j})^2 \tag{4}$$

where elements $\overline{\theta_j}$, $\bar{\theta}$, and θ_{ij} represent the mean of jth chain, the total mean, and the ith sample draw from chain j, respectively. The convergence is determined by analyzing the number of iterations until the following ratio, R, approaches unity

$$R = \frac{\frac{n-1}{n} W + \frac{1}{n} B}{W} \tag{5}$$

To implement the MCMC approach using the HSPF model, a Visual Basic software utility was developed to populate the HSPF, fixed-format, user control input (UCI) file with a new parameter vector for each simulation; the utility also analyzed the model output prior to the next iteration in the Markov chain. The utility also checked whether each new sampled parameter value was within the bounds of the prior PDF. If not, another parameter value was selected before going through the acceptance or rejection process. Each parameter vector and the simulation results were stored in a SQL Server database.

Probabilistic Comparison of Alternative TMDL Allocations

A probabilistic approach was used to compare a suite of feasible TMDL allocations. Selected uncertain watershed model parameters were characterized with PDFs (instead of deterministic values) and the uncertainty was propagated to the estimated model outputs and the TMDL using Monte Carlo simulation. An ensemble of simulated daily average in-stream FC concentration time series was used to determine quantiles to estimate the confidence intervals. A time series was generated for a lower and upper quantile to estimate the corresponding confidence interval. Two quantile sets were chosen (2.5%–97.5% and 10%–90%) to estimate the corresponding confidence interval (i.e., 95% and 80%). The water quality criterion, instantaneous violation rate (IVR) or exceedance rate for each quantile time series, was calculated by dividing the number of violations (days when the criterion was exceeded) by the number of days in the simulation period. The confidence interval (difference between the lower and upper quantiles) can also be considered to be a measure of reliability of or confidence in achieving the water quality criterion. In the proposed framework, an optimal TMDL allocation is one that (1) minimizes water quality criterion violations (lowest IVR); and (2) demonstrates the greatest reliability (narrowest confidence interval). The latter criterion means that the TMDL allocation performs better in a variety of plausible modeling scenarios (i.e., multiple parameter sets). Typically, in TMDL development, no formal calculation is performed to quantify uncertainty. Instead, a fraction of the total point and nonpoint sources is included as the MOS. In the uncertainty analysis framework presented here, selected parameters in HSPF are characterized with PDFs (instead of deterministic values) and the uncertainty associated with those parameters is propagated to model outputs (e.g., FC concentration) and the TMDL. The probabilistic analysis framework presented here avoids arbitrarily assigning a MOS, by providing a systematic approach to include uncertainty in TMDL development. Comparing alternative TMDL allocations probabilistically allows stakeholders to make more informed decisions between alternative TMDL allocations by clearly understanding the tradeoff between the specified pollutant load reductions, which directly influences implementation costs and the risk of violating the water quality criterion.

Case Study

The case study examined here used a bacteria TMDL developed for Mossy Creek (Benham et al. 2004), located in Rockingham and Augusta counties, in Virginia (Fig. 1). The watershed is 4,076 ha and is characterized as a rolling valley with the Blue Ridge

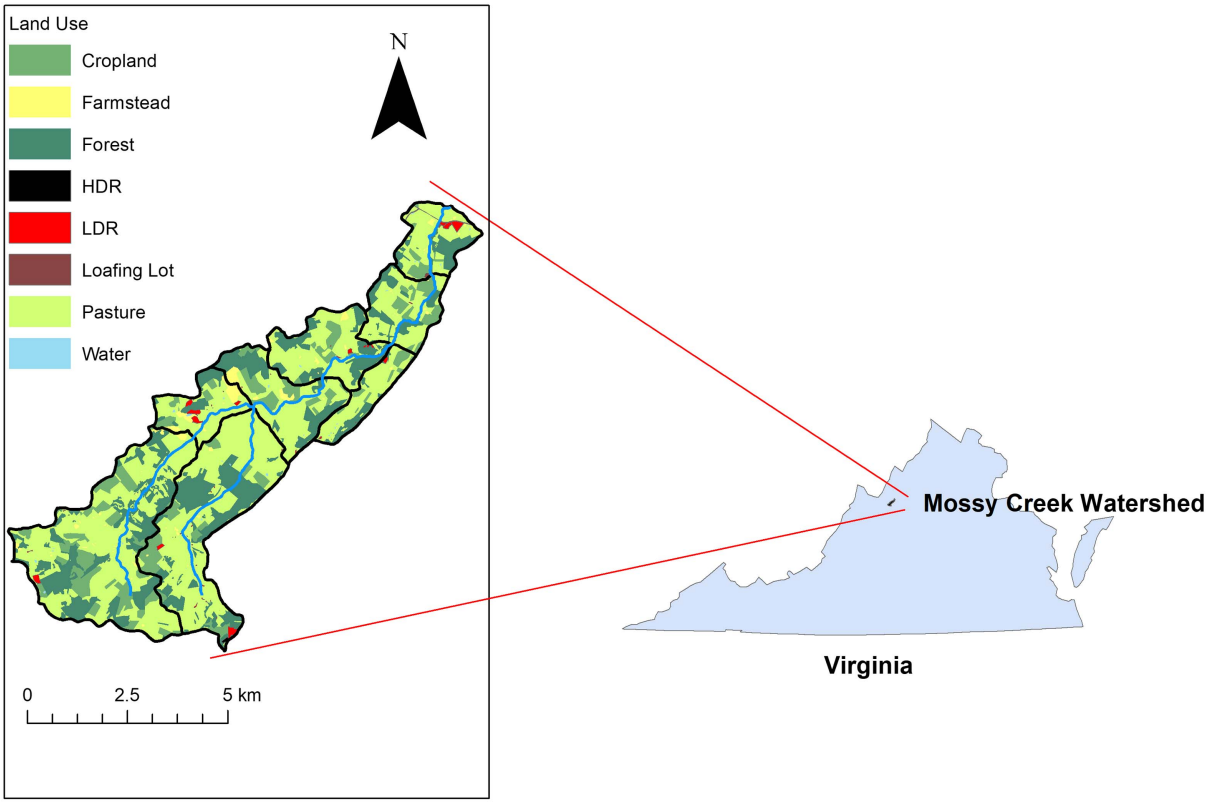

Fig. 1. Mossy Creek watershed. LDR = low-density residential; and HDR = high-density residential.

mountains to the east and the Appalachian mountains to the west. The predominant land uses are pasture (57.6%), forest (25.2%), and cropland (13.6%). The primary sources of FC are cattle lingering and defecating in the stream (direct deposit) and runoff from pastures where grazing animals defecate.

Watershed Model Implementation

Benham et al. (2004) developed the Mossy Creek bacterial TMDL using HSPF. The HSPF model is a comprehensive, lumped parameter model that continuously simulates the movement of water, sediment, and other water quality constituents on pervious and impervious surfaces, in soil profiles, and within streams and well-mixed reservoirs using three main modules PERLND, IMPLND, and RCHRES (Bicknell et al. 2005). The PERLND module represents pervious land, the IMPLND module represents impervious surfaces in which no infiltration occurs, and the RCHRES module represents the stream reaches and reservoirs in a watershed. When using HSPF to develop bacterial TMDLs, bacteria are typically simulated as a planktonic constituent. When modeling indicator bacterial with HSPF, the user typically specifies different build-up and wash-off relationships for the PERLND and IMPLND land uses. The daily FC loading rate to the land surface is specified by the user for each PERLND and IMPLND segment as a constant load (ACQOP) or a monthly variable load (MON-ACCUM). The concentration of FC in groundwater and interflow is specified for HSPF as a constant or monthly amount. Bacteria die-off on the land surface is represented indirectly by providing an asymptotic limit of bacteria build-up (SQOLIM-PERLND). The user specifies the asymptotic bacterial build-up limit for each PERLND and IMPLND land use. This limit is typically calculated using a first-order die-off rate. This asymptotic limit can vary dynamically by PERLND and IMPLND. Release of bacteria in overland runoff or FC wash-off potential is controlled by the WSQOP parameter that specifies the amount of runoff needed to wash off 90% of FC built up on the land surface. In-stream die-off is simulated using a temperature-dependent first-order relationship (Chick's law).

For this study, Mossy Creek was delineated into eight subwatersheds using the digital elevation model with spatial resolution of 30 m from the USGS (2002). The watershed was monitored monthly near the watershed outlet by the Virginia Department of Environmental Quality (VADEQ) between July 1992 and March 2003 for FC concentration and other water quality constituents. The Department of Biological Systems Engineering at Virginia Tech monitored Mossy Creek semi-monthly between February 1998 and December 2001 near the VADEQ station for selected water quality constituents, including in-stream FC concentration. Daily flow data were also collected from May 1998 to December 2002 at the same site. Hydrologic calibration and validation, and water quality calibration was conducted using data from both monitoring sites. The other data sets required to simulate in-stream FC using HSPF include rainfall, land use, FC loading rates by cattle and wildlife, inflows from springs, solar radiation, and temperature. Model and data development are described in Benham et al. (2004).

Uncertain Parameters

Using HSPF to simulate FC bacteria requires information about several hydrologic and water quality parameters. GLUE and MCMC procedures require the modeler to define Bayesian prior PDFs of these model parameters. Many previous studies, such as Beven and Freer (2001) and Makowski et al. (2002), assigned a uniform PDF to most input parameters to avoid bias. For this investigation, assigned model parameter prior PDFs were based on the literature (Al-Abed and Whiteley 2002; Lawson 2003; USEPA 2001), expert opinion, and GIS data available for the Mossy

Table 1. PDFs of the hydrologic parameters that were applied to all land uses

Parameter	Parameter description	Parameter PDF
LZSN (mm)	Lower zone nominal soil moisture storage	Uniform (76.2, 203.2)[a]
AGWRC[b]	Groundwater recession rate	Uniform (0.92, 0.99)[a]
DEEPFR	Fraction of infiltration into deeper aquifers	Uniform (0.0, 0.2)[a]
BASETP	Evapotranspiration by riparian vegetation as active groundwater enters streambed	Uniform (0.0, 0.05)[a]
AGWETP	Fraction of model segment that is subject to direct evaporation from groundwater storage	Uniform (0.0, 0.05)[a]
IRC	Interflow recession coefficient	Triangular (0.5, 0.7, 0.8)[c]
INTFW	Coefficient that determines the amount of water that enters the ground from surface detention and becomes interflow	Uniform (1.0, 3.0)[a]

[a]Numbers in parentheses show lower and upper limit of the uniform PDF, respectively. PDF limits based on values reported by USEPA (2000).
[b]Unless otherwise indicated parameters are dimensionless.
[c]Numbers in parentheses show lower limit, mode, and upper limit of the triangular PDF, respectively. PDF limits based on values reported by USEPA (2000).

Table 2. PDFs of the uncertain hydrologic parameters that vary according to the land use and time of year, for the month of January

Parameter	Land use	Parameter PDF
INFILT (mm h^{-1}) index to mean infiltration rate	Forest	Triangular (1.3, 2.5, 25.4)[a]
	Cropland	Triangular (0.8, 4.3, 6.1)[a]
	Pasture	Triangular (1.0, 2.3, 22.9)[a]
	LDR	Triangular (0.8, 4.3, 6.6)[a]
	HDR	Triangular (0.3, 0.3, 2.5)[a]
	Farmstead	Triangular (0.8, 3.8, 5.8)[a]
	Loafing lot[b]	Uniform (3.8, 5.8)[c]
UZSN (mm) nominal upper zone soil moisture storage	Forest	Uniform (5.1, 7.6)[c]
	Cropland	Uniform (1.5, 2.5)[c]
	Pasture	Uniform (1.5, 2.5)[c]
	Others[d]	Uniform (1.5, 2.5)[c]
CEPSC (mm) interception storage capacity	Forest	Uniform (1.3, 1.9)[c]
	Cropland	Uniform (1.3, 1.9)[c]
	Pasture	Uniform (1.3, 1.9)[c]
	Others[d]	Uniform (1.3, 1.9)[c]
LZETP[e] index to lower zone evapotranspiration	Forest	Uniform (0.1, 0.2)[c]
	Cropland	Uniform (0.1, 0.2)[c]
	Pasture	Uniform (0.1, 0.2)[c]
	Others[d]	Uniform (0.1, 0.2)[c]

Note: LDR = low-density residential; and HDR = high-density residential.
[a]Numbers in parentheses show lower limit, mode and upper limit of the triangular PDF, respectively. PDF limits and mode developed from histograms of gridded GIS soil and land use data.
[b]An area where livestock linger as part of the farming operation.
[c]Numbers in parentheses show lower and upper limit of the uniform PDF, respectively. PDF developed from histograms of gridded GIS soil and land use data.
[d]Other land uses include farmstead, LDR, HDR and loafing lot.
[e]Dimensionless.

Creek watershed. The uncertain hydrologic parameters were categorized into three groups on the basis of how each parameter potentially varied with land use and time. Table 1 lists the parameters that were not considered to be a function of land use or time. A uniform distribution was the most obvious choice for most parameters, because these are commonly used calibrated parameters, with no specific guidance, except the typical and possible limits in BASINS Technical Note 6 (USEPA 2000). The lower and upper limits of the uniform PDFs correspond to the typical minimum and maximum limits for these parameters from USEPA (2000). The interflow recession coefficient (IRC) was assigned a triangular PDF with lower limit, mode, and upper limit of 0.5, 0.7 (recommended value to start the model calibration process), and 0.8, respectively.

The second group was those parameters that were varied with land use and/or time (Table 2). The INFILT parameter is a function of land use only. An analysis of the histograms from gridded GIS soil and land use data suggested that a triangular PDF was appropriate for all the land uses except the loafing lot (an area where livestock linger as part of the farming operation) land use (Table 2). For those parameters that vary with time, i.e., including nominal upper zone soil moisture storage (UZSN), interception storage capacity (CEPSC), and index to lower zone evapotranspiration (LZETP), parameter PDF values (minimum and maximum) for the month of January were derived from USEPA (2000). The PDFs for the other months of the year were calculated by multiplying the January PDF by a monthly adjustment factor (Table S1) based on the TMDL (Benham et al. 2004).

Simulation of in-stream FC concentrations with HSPF requires the estimation of multiple water quality parameters including land-based constituent loading parameters (ACQOP and SQOLIM), the rate of surface runoff that results in a 90% wash-off in one hour (WSQOP), the first-order die-off rate for FC bacteria in the waterbody (FSTDEC), and the FSTDEC temperature correction coefficient (THFST). The FC loading rates depend on several factors, including species-specific feces production rates and fecal densities, die-off rates, animal density, and the fraction of time

Table 3. Prior PDFs of the water quality parameters used to simulate in-stream FC bacteria concentrations

Parameter	Land use	Parameter PDF
ACQOP-PERLND (cfu day^{-1}) Accumulation of FC on pervious land	Pasture	Log-triangular (1×10^9, 1×10^{10}, 1×10^{11})[a]
	Loafing lot[b]	Log-triangular (1.1×10^{11}, 1.1×10^{12}, 1.1×10^{13})[a]
	Cropland (for January)	Log-triangular (2×10^6, 2×10^7, 2×10^8)[a]
SQOLIM adjustment factor[c] multiplied by ACCUM values to obtain SQOLIM-PERLND	All	Uniform (2.5, 11.5)[d]
SQOLIM-PERLND maximum accumulation of FC on pervious land	All	ACQOP-PERLND (for each land use) multiplied by SQOLIM adjustment Factor
WSQOP-PERLND rate of surface runoff that will remove 90% of stored bacteria from pervious land surface	All	Uniform (0.5, 2.4)[d]
FSTDEC (day^{-1}) first order die-off rate	All	Triangular (0.12, 1.1, 2.52)[e]

[a]Numbers in parentheses show the logarithm of lower limit, mode and higher limit of the log-triangular PDF.
[b]An area where livestock linger as part of the farming operation.
[c]Unless otherwise indicated parameters are dimensionless.
[d]Numbers in parentheses show limits of the uniform PDF.
[e]Numbers in parentheses show lower limit, mode, and upper limit of the triangular PDF, respectively.

livestock are confined (Zeckoski et al. 2005). As cited in TMDL reports and in the literature (ASAE 2003; Geldreich 1978; Yagow 2001), the FC production rates, typically reported in colony forming units per day (cfu day^{-1}), for dairy cattle, beef cattle, and poultry can vary by several orders of magnitude. According to the Mossy Creek bacteria TMDL report, dairy cattle, beef cattle, and poultry are responsible for more than 94% of FC production in the watershed (Benham et al. 2004). Therefore, this study hypothesized that the uncertainty in production rates of dairy cattle, beef cattle, and poultry (present on the cropland and pasture) were likely to have the greatest impact on FC quantification uncertainty. To account for this uncertainty, the loading rates for pervious land areas were assigned a log-triangular PDF. The mode of the ACQOP PDF was determined from values generated using the Bacteria Source Load Calculator (BSLC), a software tool used in FC source characterization for TMDL development (Zeckoski et al. 2005), and the lower and upper limits of the ACQOP PDF were determined by multiplying the PDF mode by 0.1 and 10, respectively (Table 3).

Because the application of manure to cropland varies by month, the FC accumulation rate (MON-ACCUM) was varied monthly. To account for the temporal distribution of ACQOP, the FC cropland loading PDF for January was multiplied by an adjustment factor for each month (Table S1), with the factor developed using the trend of FC accumulation values generated by the BSLC for each month. Deterministic values for ACQOP calculated using the BSLC were used for the other land uses [forest; low-density residential (LDR); high-density residential (HDR); farmstead; and impervious areas].

The SQOLIM parameter is typically calculated by multiplying ACQOP by an adjustment factor of nine, based on the assumption that the die-off coefficient for FC bacteria on pervious land surface is 0.051 day^{-1} (Zeckoski et al. 2005). Crane and Moore (1986) reviewed several studies and reported a bacteria die-off rate ranging from 0.04 to 0.20 day^{-1}, which translates to the SQOLIM adjustment factor of 2.5–11.5. The SQOLIM adjustment factor was assigned a uniform PDF between 2.5 and 11.5 and was used to calculate the SQOLIM values for each land use (Table 3).

There is no guidance available on estimating WSQOP, FSTDEC, and THFST when simulating FC bacteria with HSPF. The values of these parameters are generally adapted from previous studies and further calibrated. A review of values used in previous FC TMDLs shows a range of 0.5–2.4 for WSQOP (Lawson 2003). Based on these reported values, a uniform PDF bounded by these two values was used for WSQOP for all land uses. In a review by Bowie et al. (1985), FSTDEC values ranging from 0.12 to 2.52 day^{-1} were reported for various streams. The average of the reported values was 1.1 day^{-1}. A similar value of FSTDEC has been used in TMDLs developed in Virginia, so a triangular PDF was assigned to FSTDEC with a mode of 1.1 and limits of 0.12 and 2.52 day^{-1}. In this study, it was assumed that any uncertainty in THFST would be masked by the uncertainty in FSTDEC. Hence, a deterministic value of 1.07 was used for THFST (Lawson 2003).

Fecal coliform bacteria directly deposited in a waterbody is input as an hourly time series to HSPF; this study used the BSLC-generated time series for the Mossy Creek TMDL. Because the FC discharge from the sole point source was negligible and the FC production by wildlife and straight pipes (unlawful sewers emptying directly into the stream) was estimated to be less than 1% and 2%, respectively, of the total FC direct deposit, the uncertainty in direct deposit FC is primarily due to cattle. To be consistent with the other PDFs of bacteria loads, the cattle direct deposit PDF was assigned a log-triangular distribution. To define this distribution, the cattle direct deposit time series was multiplied by a factor from a log-triangular PDF with a mode of 1 and limits of 0.1 and 10 (Table 3).

Watershed Model Calibration and Validation

The Mossy Creek model was calibrated and validated using the GLUE and MCMC techniques independently. For the GLUE application, this is a two-stage process in which the hydrologic calibration and validation is performed first, followed by water quality calibration and validation. Hydrologic calibration used the period from September 1998 to December 1999. The posterior parameter PDFs obtained using GLUE were used to validate the hydrologic model for the period from January 2000 to September 2002. Posterior PDFs of water quality parameters were estimated using GLUE for the period from October 1998 to December 2001. However, insufficient observed data prevented water quality validation (Benham et al. 2004).

For the MCMC application, each sampled parameter set is a function of the previous set, making separate hydrologic and water quality calibrations impractical. The Mossy Creek model for the MCMC application was calibrated using four years of concurrent hydrologic and water quality data (September 1998 to September 2002). In each iteration, two likelihood functions were calculated: one for hydrologic response based on the daily flow volume at the watershed outlet and one for water quality that compared observed instantaneous in-stream and simulated daily average FC concentrations. A log transformation was performed on the observed and

Table 4. TMDL pollutant allocations for FC bacteria

TMDL allocation	Required source-specific FC load reductions (%)							
	Single sample violation rate (%)	Cattle direct deposit	Cropland	Pasture	Loafing lot[a]	Wildlife direct deposit	Straight pipes[b]	All residential pervious land segments
Existing	48.0	0	0	0	0	0	0	0
S1	41.0	0	50	50	100	0	100	50
S2	0.1	94	95	97	100	0	100	95
S3	0.1	94	95	95	100	30	100	95
S4	0.1	99	95	95	100	99	100	95
S5	0	99	90	98	100	30	100	95
S6	0	94	95	98	100	0	100	95

[a]An area where livestock linger as part of the farming operation.
[b]Unlawful sewers emptying directly into the stream.

simulated flow volume and in-stream FC bacteria before deriving the likelihood function to normalize the scale of residuals.

TMDL Pollutant Allocations

The Mossy Creek TMDL (Benham et al. 2004) lists six pollutant load allocations, with only two meeting the FC water quality criterion of zero violations (S5 and S6, Table 4). These allocations were run on a 3.5-year period that includes a range of representative hydrological events in Mossy Creek. The S5 and S6 allocations both require 100% reduction in the FC loading from the loafing lot and straight pipes, 98% reduction from pasture, and 95% from residential pervious land segments. The differences between the two allocations are the reductions in cattle direct deposit (99% versus 94%), loading from cropland (90% versus 95%), and wildlife direct deposit (30% versus 0%).

Benham et al. (2004) recommended the S6 allocation because it calls for a lesser reduction in wildlife direct deposit than the S5 allocation. The TMDL allocations that focus on the reduction of sources from human activities are generally favored. No effort was made in the Mossy Creek TMDL study to consider uncertainty when prioritizing allocation alternatives. The comparison of the S5 and S6 allocations presented in Table 4 considered only selected sources of uncertainty. Uncertainty in wildlife direct deposit was not considered owing to the limited production of FC bacteria by wildlife (<1% of total FC load).

Results

Generalized Likelihood Uncertainty Estimation

Using GLUE, the posterior parameter-specific distributions were derived and were used to model Mossy Creek watershed. The prior and posterior CDFs of two hydrologic parameters, lower zone nominal soil moisture storage for pasture (LZSN-pasture) and fraction of infiltration into deep aquifers (DEEPFR) are presented in Fig. 2. Fig. 2(a) shows that the acceptable likelihoods for the LZSN-pasture occur at the lower end of the distribution, which results in the posterior distribution diverging from the prior. Fig. 2(b) illustrates that the acceptable likelihoods of DEEPFR occur for the entire range of the parameter, which results in a posterior distribution similar to the uniform prior distribution. The difference in prior and posterior distributions for the two parameters implies that the observed data provided greater information about LZSN-pasture than DEEPFR. The model output is more sensitive to LZSN-pasture than DEEPFR. A similar analysis was performed for all the uncertain hydrologic parameters in this study (Tables 1 and 2).

The posterior distributions of the hydrologic parameters were used to conduct Monte Carlo simulations for the validation period (January 2000 to September 2002). The daily flow volume resulting from each model parameterization was used to calculate the statistics typically used to assess HSPF calibration sufficiency (Lumb et al. 1994). The 2.5% and 97.5% quantiles for these measures were used to validate the posterior PDFs (Table 5). These dimensionless measures test the model performance in terms of mass balance, high- and low-flow distribution and seasonal variability. All the hydrologic calibration objectives, except 50% lowest

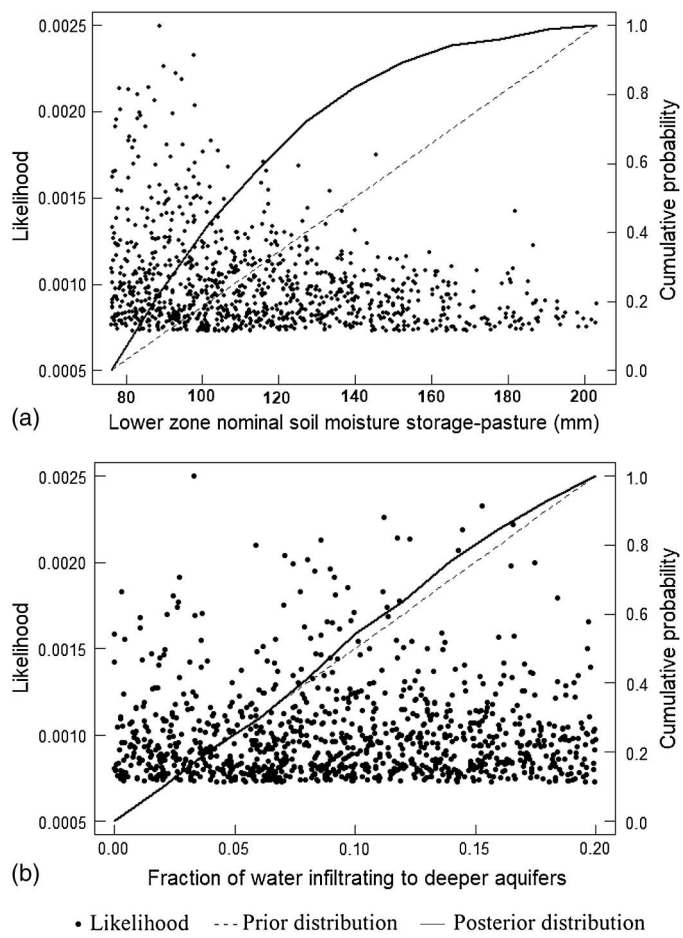

Fig. 2. Posterior cumulative distributions for two hydrologic parameters: (a) lower zone nominal soil moisture storage for pasture; and (b) fraction of infiltration into deep aquifers obtained using GLUE.

Table 5. Quantiles of the expert statistics for the validation period (January 2000 to September 2002) using posterior and prior PDFs

Expert statistic	Criterion (Lumb et al. 1994) (%)	Quantiles using prior PDFs 2.5%	Quantiles using prior PDFs 97.5%	Quantiles using posterior PDFs 2.5%	Quantiles using posterior PDFs 97.5%
Total volume error	±10	−13.1%	13.8%	−10.3%	8.2%
50% lowest flows error	±10	−8.8%	23.8%	−4.7%	16.8%
10% highest flows error	±15	−16.5%	19.0%	−16.9%	1.0%
Storm peaks error	±20	−15.4%	32.0%	−16.1%	1.7%
Seasonal volume error	±30	0.9%	16.3%	0.2%	11.1%
Summer storm volume error	±50	−19.1%	15.3%	−15.7%	7.0%

flows and 10% highest flows, were met 95% of the time when using the posterior distributions. A comparison of the expert statistics produced by prior and posterior PDFs indicate that including expert statistics in addition to the residuals between observed and simulated daily flow volume in the likelihood evaluation can improve model calibration. The posterior distributions obtained from GLUE were used for uncertainty analysis.

The model performance in terms of bacteria simulation was also evaluated using various goodness-of-fit measures, including absolute error of the geometric mean, average, and median of the simulated FC concentrations. Kim et al. (2007) suggested that the absolute error between observed and simulated values for these measures should be less than 100% for bacteria calibration. Here, the absolute errors were 3.1%, 5.9%, and 10.8%, respectively. The IVR (i.e., FC concentration >400 cfu/100 mL) of the observed data is 60.0%, while the simulations exhibit a rate of 50.6%. Since the difference is smaller than 10.0%, this criterion is also satisfactorily met (Kim et al. 2007). In addition, 91.1% of the observed data fall between the minimum and maximum simulated hourly values on the day the observation was made. The observed FC concentration ranged over many orders of magnitude (20–160, 000 cfu/100 mL), which could be due to various reasons such as the highly varying presence of livestock in the creek. The observed bacteria data are also subject to uncertainty (Hantush and Chaudhary 2014). Based on this assessment, we concluded that the model is sufficiently calibrated for bacteria modeling.

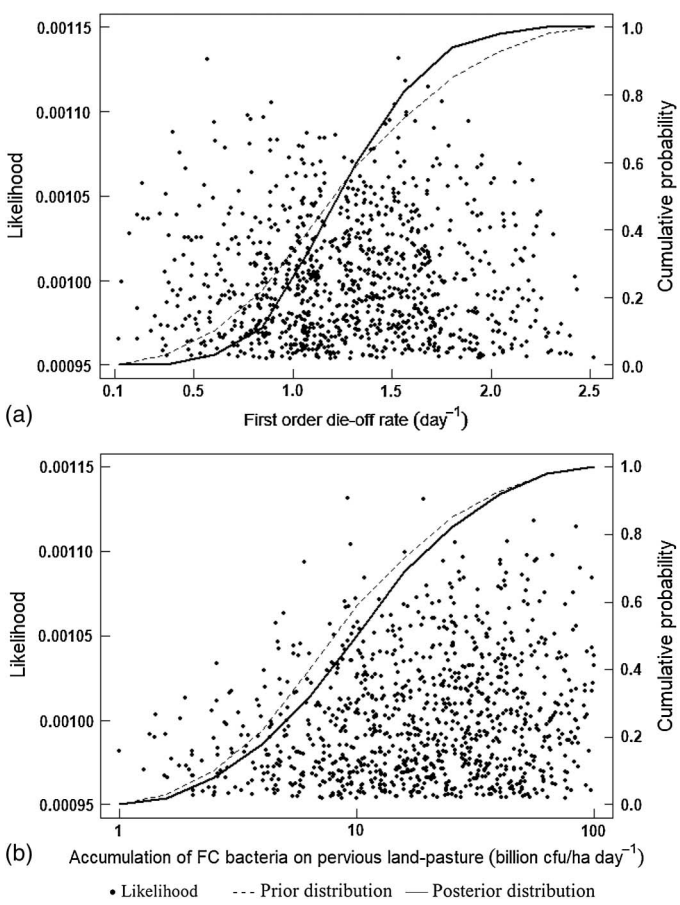

Fig. 3. Posterior cumulative distributions for two water quality parameters: (a) in-stream first-order die-off rate for FC bacteria; and (b) rate of FC accumulation on pasture obtained using GLUE.

Following the hydrologic calibration and validation as well as water quality calibration, GLUE was used to determine the posterior distributions of water quality parameters. The likelihood function for each water quality parameter was calculated using log-transformed observed FC bacteria in-stream concentrations and simulated daily average FC bacteria concentration for the days for which observed data were available.

Fig. 3 shows the prior CDF and the GLUE-generated posterior CDFs for two water quality parameters: FSTDEC and ACQOP-pasture. The posterior CDFs differ from their prior distributions, implying the influence of observed data on the prior distributions. However, the difference is not as large as that for the hydrologic parameters, which is likely because water quality data were not measured as frequently as streamflow data. The posterior CDF of all water quality parameters were developed similarly.

The posterior distributions obtained via GLUE were used to conduct Monte Carlo simulation for the two Mossy Creek TMDL allocations S5 and S6 (Table 4). The simulations were conducted for a period of approximately 3.5 years (1,218 days). The daily average simulated FC concentration time series from each Monte Carlo simulation was used to determine the 2.5%, 10%, 90%, and 97.5% quantiles (80% and 95% confidence intervals) for the two valid allocations S5 and S6 (Fig. 4). Scenario S6 exhibited greater uncertainty than S5 (i.e., larger difference between the lower and upper quantiles). The S6 scenario reduced cattle direct deposit less than the S5 scenario, but allowed more FC loading from cropland, illustrating that cattle direct deposit was a greater source of uncertainty than cropland FC loading in the Mossy Creek simulation, given that both sources had similar input uncertainty (section "Uncertain Parameters").

The ensemble average IVR (1.2% versus 1.3%) is comparable for the two TMDL allocations S5 and S6, suggesting that both allocations have a similar chance of improving water quality (Table 6). While the 80% confidence interval for S5 and S6 are similar, the 95% confidence interval for scenario S6 is wider than for S5. As a result, the probabilistic analysis based on the GLUE application suggests that S5 should be the preferred Mossy Creek TMDL allocation because S5 provides similar water quality improvement with greater reliability when compared with S6. This is not in agreement with the deterministic analysis by Benham et al. (2004). The inconsistency between deterministic and probabilistic

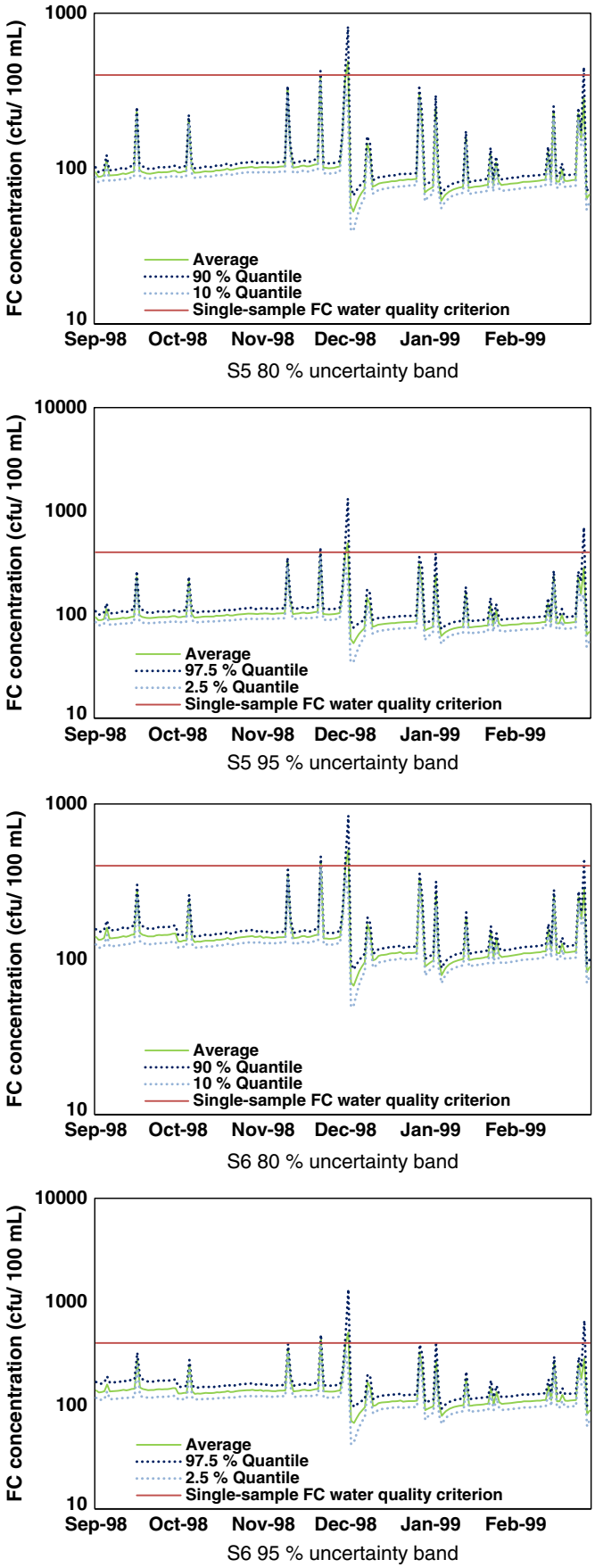

Fig. 4. Representative 80% and 95% confidence intervals for simulated in-stream FC for TMDL allocations S5 and S6 using GLUE.

Table 6. IVR of water quality criterion by the ensemble average time series and the confidence intervals for the two TMDL allocations using GLUE

TMDL allocation	Ensemble average IVR (%)	80% confidence interval of IVR (%)	95% confidence interval of IVR (%)
S5	1.2	(0.2, 1.6)[a]	(0.0, 2.5)[b]
S6	1.3	(0.4, 1.8)[a]	(0.1, 3.1)[b]

[a]10% and 90% quantile of the IVR over the simulation period.
[b]2.5% and 97.5% quantile of the IVR over the simulation period.

frameworks when prioritizing TMDL allocations has also been documented in previous research (Borsuk et al. 2002; Langseth and Brown 2011). However, the focus of decision making in the probabilistic application is on water quality improvement and reliability; technical feasibility and cost of implementation was not considered as a decision criterion. With respect to the two alternative allocation scenarios, S5 will likely be costlier than S6 and less feasible to implement. More expensive because S5 requires a greater reduction in cattle direct deposits (i.e., more streambank livestock exclusion fencing to achieve the required reduction), while S6 requires a larger cropland bacteria load reduction that can be achieved in a variety of less costly ways, and less feasible because S5 requires a greater reduction from cattle direct deposits than S6 in addition to reductions in wildlife bacteria loading. The implementation of either of these scenarios will be challenging given the large load reductions that are needed to achieve the applicable water quality criteria.

Markov Chain Monte Carlo

Markov chain Monte Carlo was used to generate posterior distributions of hydrologic and water quality parameters for the Mossy Creek TMDL model to estimate the uncertainty associated with in-stream FC concentrations. After burn-in (~70,000 iterations), 30,000 MCMC iterations were used to infer both hydrologic and water quality parameters. The distribution of hydrologic parameters exhibits less variation within the parameter space compared with the water quality parameters. The observed data had a greater effect on the hydrologic parameters LZSN-cropland and DEEPFR than the water quality parameters ACCUM-pasture and FSTDEC because the observed data for flow were available for the entire simulation period (1,218 days), whereas observed FC concentration data were available for only 90 days of the period.

When using the Gelman–Rubin statistic for assessing convergence, at least three independent Markov chains are required. To obtain these Markov chains, three independently initiated Markov chains were executed in parallel. Markov chain convergence is determined when the Gelman–Rubin statistic is approximately unity and estimates of the variance stabilize. This analysis was performed for all hydrologic and water quality parameters. All parameters converged around 70,000 iterations or earlier. Posterior distributions for the LZSN-cropland, DEEPFR, ACCUM-pasture, and FSTDEC parameters are shown in Fig. 5.

The posterior CDFs (solid lines in Fig. 5) differed from the prior CDFs (dashed lines in Fig. 5) showing the effect of observed data on the parameters. In general, the posterior CDFs of the hydrologic parameters differed more from the prior CDFs than did those of the water quality parameters. As expected, the GLUE analysis yielded similar results, but there are inconsistencies between the two techniques. The posterior CDFs for all hydrologic and water quality parameters were derived similarly. Using the posterior CDFs, Monte Carlo simulations were performed to yield average daily

J. Hydrol. Eng., 2018, 23(12): 05018025

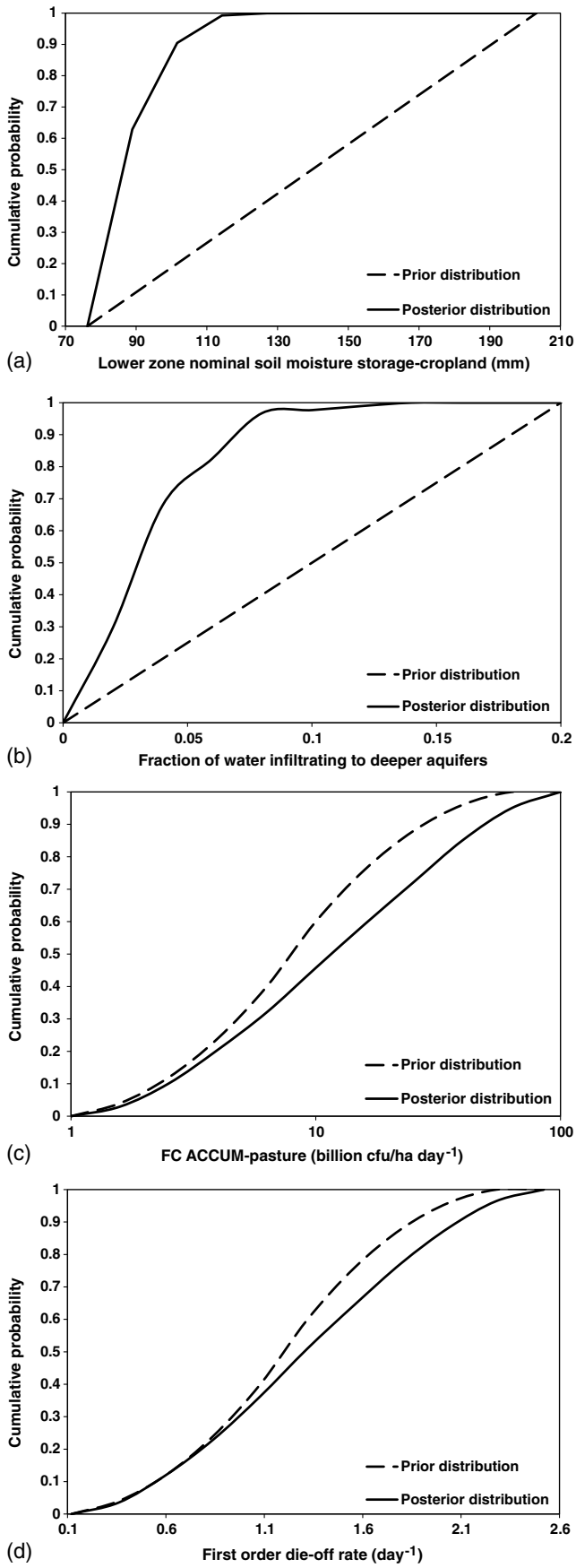

Fig. 5. Posterior cumulative distributions for (a) lower zone nominal storage for cropland; (b) fraction of infiltration into deep aquifers; (c) FC bacteria accumulation on pasture; and (d) in-stream first-order die-off rate for FC bacteria using Markov chain Monte Carlo analysis.

in-stream FC bacteria concentrations. As with the GLUE analysis, model output was used to generate 2.5%, 10%, 90%, and 97.5% quantiles (80% and 95% confidence intervals) for the Mossy Creek TMDL allocations S5 and S6 (Fig. 6).

The IVR for different quantiles for the S5 and S6 TMDL allocations (0.9% versus 1.1%) are similar suggesting that both allocations have a similar chance of improving water quality (Table 7). However, both the 80% and 95% confidence intervals for scenario S6 are wider than the respective bands for S5. As was the case with the GLUE analysis, this analysis suggests that S5 should be the preferred Mossy Creek TMDL allocation alternative since it provides similar water quality improvement with greater reliability when compared with S6. Similar to what was found in GLUE application, the optimal alternative is not in agreement with the deterministic analysis by Benham et al. (2004).

Discussion

In this study, the reliability (i.e., the risk of not meeting an applicable water quality criterion) of two TMDL pollutant allocation alternatives (S5 and S6) was assessed. Based on the analysis presented in this paper, the allocation recommended in the USEPA-approved TMDL (alternative S6) had a wider confidence interval than the allocation selected by the probabilistic framework (alternative S5) and thus, less reliability or higher risk. This suggests that while both allocations produced zero water quality criterion violations, the scenario that exerted more control (i.e., a greater reduction) on the source that contributed the most uncertainty (cattle direct deposition) is preferred. The cost of implementing the two allocation alternatives was not considered. The comparative valuation of reliability and cost should be considered by both water quality managers and analysts. A high-cost scenario with greater reliability might be preferable when an ecosystem is fragile (i.e., the risk of a water quality violation has potentially greater consequences). However, a less expensive, less reliable pollutant allocation scenario might be preferred in a situation when the ecosystem is more resilient or watershed management funding is limited. An explicit assessment of the uncertainty associated with watershed modeling and TMDL development enables stakeholders to engage in more informed decision making.

Both GLUE and MCMC allow uncertainty analyses of watershed-scale water quality simulations by estimating posterior distributions of model parameters using observed data and user-defined, parameter-specific prior distributions coupled with Monte Carlo simulation. The likelihood formulation is a key step in the application of GLUE that provides the analyst flexibility in selecting model responses that are important for a given application. Information used to develop a statistically rigorous likelihood function is often scarce in watershed water quality modeling. However, GLUE does not require a statistical likelihood function and is often criticized for that (Mantovan and Todini 2006; Stedinger et al. 2008). When using GLUE, the analyst also needs to define the model parameter acceptance or rejection criteria, which may affect the posterior parameter distribution. By contrast, MCMC facilitates statistical sampling even in the most complex of cases using a Bayesian posterior distribution (or in many cases, a likelihood function). Compared with GLUE, MCMC requires more computational resources. For the 4,076 ha Mossy Creek TMDL case study presented in this paper, the MCMC analysis (run with three chains to ensure model convergence) took about four times longer to run compared with GLUE (nearly a week versus 2 days). Implementation of GLUE is also simpler

J. Hydrol. Eng., 2018, 23(12): 05018025

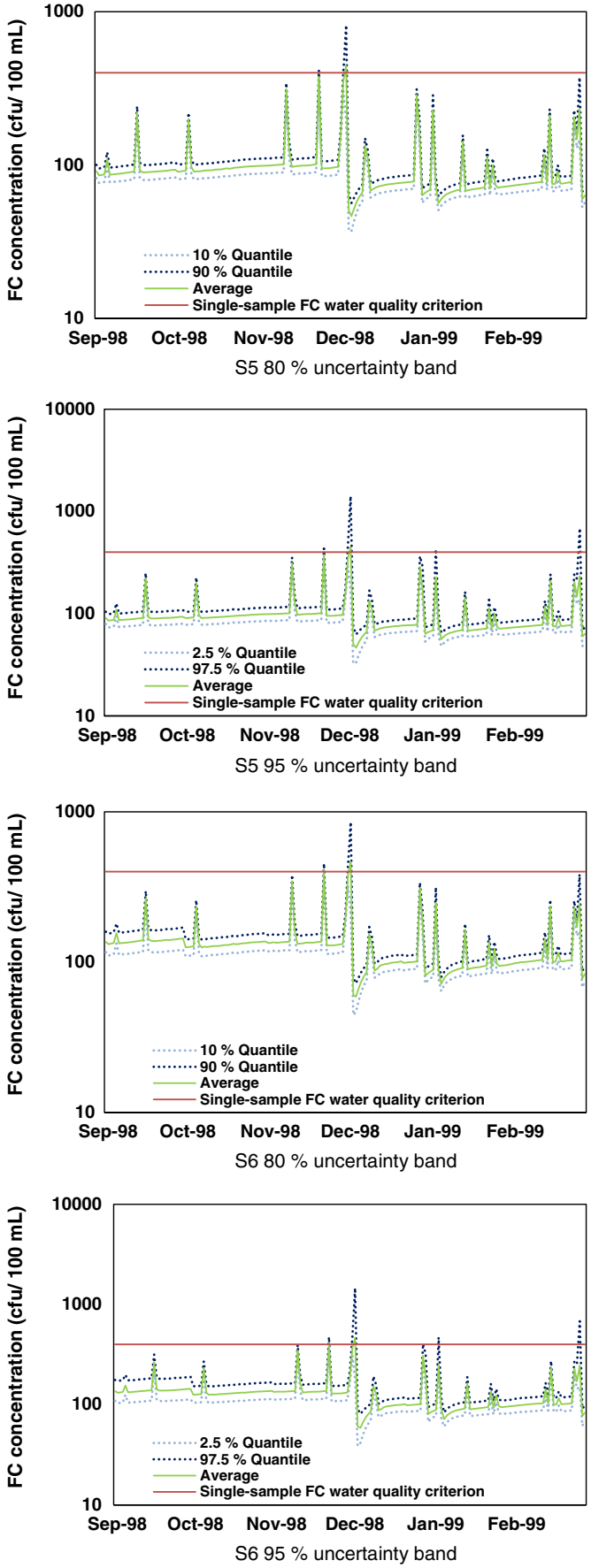

Fig. 6. Representative 80% and 95% concentration uncertainty bands for simulated in-stream FC bacteria for TMDL allocations S5 and S6 using Markov chain Monte Carlo analysis.

Table 7. IVR of water quality criterion by the ensemble average time series and the confidence intervals for two TMDL allocations using Markov chain Monte Carlo analysis

TMDL allocation	Ensemble average IVR (%)	80% confidence interval of IVR (%)	95% confidence interval of IVR (%)
S5	0.9	(0.0, 1.5)[a]	(0.0, 2.8)[b]
S6	1.1	(0.0, 1.6)[a]	(0.0, 3.5)[b]

[a]10% and 90% quantile of the IVR over the simulation period.
[b]2.5% and 97.5% quantile of the IVR over the simulation period.

and more straightforward than MCMC. Therefore, when considering computational requirements, GLUE might be viewed as a more practical technique for uncertainty estimation of HSPF-based bacteria modeling. However, with the increase in computational power and reduced run time for HSPF, these advantages of GLUE over MCMC may blur in the future. Further research on MCMC statistical sampling methods for HSPF applications are likely to decrease the computational cost. Additional research is also required to analyze the effects of using different likelihood functions and different parameter acceptance or rejection criteria when using GLUE for HSPF-based water quality applications (He et al. 2010). A potential method that was recommended by Hantush and Chaudhary (2014) for TMDL applications is Bayesian Monte Carlo Maximum Likelihood, which takes the advantages of both GLUE and MCMC. Despite its recent applications (Chaudhary and Hantush 2017), there has been no application to complex watershed-scale water quality models.

The performance measures and associated criteria reported by Lumb et al. (1994) were used as HSPF hydrologic calibration likelihood functions, primarily because these criteria are widely accepted and applied in HSPF applications. To derive parameter-specific posterior distributions of selected water quality parameters, published criteria for bacteria calibration (Kim et al. 2007) were used. Other model performance measures (e.g., Nash–Sutcliffe efficiency) and hypothesis testing that are sometimes used to assess model performance for hydrology and other water quality constituents could also be used, if criteria for those measures were established for bacteria simulation. Future applications applying uncertainty propagation techniques like GLUE and MCMC should focus on developing advanced likelihood functions for bacteria simulation. Water quality calibration procedures using Bayesian techniques (Camacho et al. 2018) show potential to improve water quality simulation model performance. However, application of this computationally intensive procedures in overly parameterized models such as HSPF will be both complicated and computationally expensive.

This study focused only on the uncertainty resulting from model parameters. In future research, other sources of uncertainty (e.g., model structure, the cost of implementing pollution control measures, and the willingness of stakeholders to implement nonregulated pollution control measures) should be incorporated into uncertainty analyses. Additionally, this study was conducted under the assumption of stationarity; the uncertainty triggered by nonstationarity, such as changes in climate and land use, that might have a major impact on water quality (Fonseca et al. 2015; Whitehead et al. 2009) was not considered. Continued efforts to investigate the impact of these nonstationarities are vital for water quality–based watershed management decisions and promotes adaptive management.

Summary and Conclusions

This study presented a probabilistic framework to consider both water quality improvement and reliability during the prioritization of alternative TMDL pollutant allocation scenarios. Two uncertainty analysis techniques, GLUE and MCMC, were applied to analyze the uncertainty of two alternative TMDL pollutant load allocations developed to address a bacteria impairment in a rural watershed in Virginia. Both GLUE and MCMC allow uncertainty analyses of watershed-scale water quality simulations by estimating posterior distributions of model parameters using observed data and user-defined, parameter-specific prior distributions coupled with Monte Carlo simulation. This case study demonstrated the applicability of both GLUE and MCMC to estimate the uncertainty in simulated in-stream FC concentrations and resulting water quality criterion violations, and that the resulting information can be used to analyze options that reduce bacteria sources to improve water quality. The application of the probabilistic framework led to selection of a different pollutant allocation than that selected through the deterministic analysis by Benham et al. (2004), suggesting that missing uncertainty can misrepresent the TMDL pollutant allocations. The advantages of probabilistic frameworks for TMDL development were demonstrated. Such frameworks should be applied to improve the current practice of TMDL. The presented probabilistic framework provides a systematic approach to include uncertainty in TMDL development as a more rigorous alternative to the current practice of arbitrarily assigning a MOS. A probabilistic framework allows stakeholders to make more informed decisions when prioritizing alternative TMDL pollutant allocations. Future research should explore developing tools that examine the tradeoffs between the estimated implementation costs and relative reliability (i.e., the risk of not meeting an applicable water quality criterion) of alternative pollutant allocation scenarios.

Supplemental Data

Table S1 is available online in the ASCE Library (www.ascelibrary.org).

References

Abbaspour, K. C. 2015. *SWAT-CUP: SWAT calibration and uncertainty programs: A user manual*. Dübendorf, Switzerland: Swiss Federal Institute of Aquatic Science and Technology.

Al-Abed, N. A., and H. R. Whiteley. 2002. "Calibration of the Hydrological Simulation Program Fortran (HSPF) model using automatic calibration and geographical information systems." *Hydrol. Processes* 16 (16): 3169–3188. https://doi.org/10.1002/hyp.1094.

Ames, D. P., and U. Lall. 2008. "Developing total maximum daily loads under uncertainty: Decision analysis and the margin of safety." *J. Contemp. Water Res. Educ.* 140 (1): 37–52. https://doi.org/10.1111/j.1936-704X.2008.00027.x.

ASAE (American Society of Agricultural Engineers). 2003. *Manure production and characteristics*. St. Joseph, MI: ASAE.

Beck, M. B. 1987. "Water quality modeling: A review of the analysis of uncertainty." *Water Resour. Res.* 23 (8): 1393–1442. https://doi.org/10.1029/WR023i008p01393.

Benham, B., K. Brannan, K. Christophel, T. Dillaha, L. Henry, S. Mostaghimi, R. Wagner, J. Wynn, G. Yagow, and R. Zeckoski. 2004. *Total maximum daily load development for Mossy Creek and Long Glade Run: Bacteria and general standard (benthic) impairments*. Richmond, VA: VADEQ.

Beven, K., and A. Binley. 1992. "The future of distributed models: Model calibration and uncertainty prediction." *Hydrol. Processes* 6 (3): 279–298. https://doi.org/10.1002/hyp.3360060305.

Beven, K., and J. Freer. 2001. "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using GLUE methodology." *J. Hydrol.* 249 (1–4): 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Bicknell, B., J. Imhoff, A. Kittle Jr., T. Jobes, and A. Donigian Jr. 2005. *Hydrological simulation program: Fortran (HSPF) user's manual for release 12.2*, 845. Athens, GA: USEPA.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow. 2002. "Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development." *Environ. Sci. Technol.* 36 (10): 2109–2115. https://doi.org/10.1021/es011246m.

Bowie, G. L., W. B. Mills, D. B. Porcella, C. L. Campbell, J. R. Pagenkopf, G. L. Rupp, K. M. Johnson, P. Chan, S. A. Gherini, and C. E. Chamberlin. 1985. *Rates, constants, and kinetics formulations in surface water quality modeling*. EPA/600/3-85/040. Athens, GA: USEPA.

Camacho, R. A., J. L. Martin, W. McAnally, J. Díaz-Ramirez, H. Rodriguez, P. Sucsy, and S. Zhang. 2015. "A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling." *J. Am. Water Resour. Assoc.* 51 (5): 1372–1393. https://doi.org/10.1111/1752-1688.12319.

Camacho, R. A., J. L. Martin, T. Wool, and V. P. Singh. 2018. "A framework for uncertainty and risk analysis in total maximum daily load applications." *Environ. Modell. Software* 101: 218–235. https://doi.org/10.1016/j.envsoft.2017.12.007.

Chaudhary, A., and M. M. Hantush. 2017. "Bayesian Monte Carlo and maximum likelihood approach for uncertainty estimation and risk management: Application to lake oxygen recovery model." *Water Res.* 108: 301–311. https://doi.org/10.1016/j.watres.2016.11.012.

Chin, D. A. 2009. "Predictive uncertainty in water-quality modeling." *J. Environ. Eng.* 135 (12): 1315–1325. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000101.

Crane, S. R., and J. A. Moore. 1986. "Modeling enteric bacterial die-off: A review." *Water Air Soil Pollut.* 27 (3): 411–439. https://doi.org/10.1007/BF00649422.

Dilks, D. W., and P. L. Freedman. 2004. "Improved consideration of the margin of safety in total maximum daily load development." *J. Environ. Eng.* 130 (6): 690–694. https://doi.org/10.1061/(ASCE)0733-9372(2004)130:6(690).

Fonseca, A., C. Botelho, R. Boaventura, and V. Vilar. 2015. "Global warming effects on faecal coliform bacterium watershed impairments in Portugal." *River Res. Appl.* 31 (10): 1344–1353. https://doi.org/10.1002/rra.2821.

Gardner, K. K., B. L. McGlynn, and L. A. Marshall. 2011. "Quantifying watershed sensitivity to spatially variable N loading and the relative importance of watershed N retention mechanisms." *Water Resour. Res.* 47 (8): W08524. https://doi.org/10.1029/2010WR009738.

Geldreich, E. E. 1978. "Bacterial populations and indicator concepts in feces, sewage, stormwater and solid wastes." In *Indicators of viruses in water and food*, edited by G. Berg, 51–96. Ann Arbor, MI: Ann Arbor Science.

Gelfand, A. E., and A. F. M. Smith. 1990. "Sampling based approaches to calculating marginal densities." *J. Am. Stat. Assoc.* 85 (410): 398–409. https://doi.org/10.1080/01621459.1990.10476213.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall.

Geman, S., and D. Geman. 1984. "Stochastic relaxation, Gibbs' distribution and Bayesian restoration of images." *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6 (6): 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Hantush, M. M., and A. Chaudhary. 2014. "Bayesian framework for water quality model uncertainty estimation and risk management." *J. Hydrol. Eng.* 19 (9): 04014015. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000900.

Harmel, R. D., R. Cooper, R. Slade, R. Haney, and J. Arnold. 2006. "Cumulative uncertainty in measured streamflow and water quality data for small watersheds." *Trans. ASABE* 49 (3): 689–701. https://doi.org/10.13031/2013.20488.

Harmel, R. D., and P. K. Smith. 2007. "Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water

quality modeling." *J. Hydrol.* 337 (3): 326–336. https://doi.org/10.1016/j.jhydrol.2007.01.043.

Hastings, W. 1970. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika* 57 (1): 97–109. https://doi.org/10.1093/biomet/57.1.97.

He, J., J. W. Jones, W. D. Graham, and M. D. Dukes. 2010. "Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method." *Agric. Syst.* 103 (5): 256–264. https://doi.org/10.1016/j.agsy.2010.01.006.

Jia, Y., and T. B. Culver. 2008. "Uncertainty analysis for watershed modeling using generalized likelihood uncertainty estimation with multiple calibration measures." *J. Water Resour. Plann. Manage.* 134 (2): 97–106. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(97).

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal. 1998. "Markov chain Monte Carlo in practice: A roundtable discussion." *Am. Stat.* 52 (2): 93–100. https://doi.org/10.1080/00031305.1998.10480547.

Kim, S. M., B. L. Benham, K. M. Brannan, R. W. Zeckoski, and G. Yagow. 2007. "Water quality calibration criteria for bacteria TMDL development." *Appl. Eng. Agric.* 23 (2): 171–176. https://doi.org/10.13031/2013.22610.

Langseth, D. E., and N. Brown. 2011. "Risk-based margins of safety for phosphorus TMDLs in lakes." *J. Water Resour. Plann. Manage.* 137 (3): 276–283. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000097.

Lawson, L. G. 2003. *HSPF model calibration and verification for bacteria TMDLs*. Richmond, VA: VADEQ.

Lumb, A. M., R. B. McCammon, and J. L. Kittle. 1994. *Users manual for an expert system (HSPEXP) for calibration of the hydrological simulation program: Fortran*. US Geological Survey, Water-Resources Investigations Rep. No. 94-4168. Reston, VA: US Geological Survey.

Makowski, D., D. Wallach, and M. Tremblay. 2002. "Using a Bayesian approach to parameter estimation: Comparison of GLUE and MCMC methods." *Agronomie* 22 (2): 191–203. https://doi.org/10.1051/agro:2002007.

Mantovan, P., and E. Todini. 2006. "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology." *J. Hydrol.* 330 (1–2): 368–381. https://doi.org/10.1016/j.jhydrol.2006.04.046.

Mishra, A. 2011. "Estimating uncertainty in HSPF based water quality model: Application of Monte-Carlo based techniques." Ph.D. thesis, Dept. of Biological Systems Engineering, Virginia Tech.

Mitsova-Boneva, D., and X. Wang. 2007. "Exploring the variability in suspended sediment yield using BASINS-HSPF and probabilistic modeling: Implications for land use planning." *J. Environ. Inf.* 9 (1): 29–40. https://doi.org/10.3808/jei.200700085.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams. 2011. *Soil and water assessment tool theoretical documentation version 2009*. Texas Water Resources Institute, TR-406. College Station, TX: Texas Water Resources Institute.

Nguyen, T. T., and P. Willems. 2016. "The influence of model structure uncertainty on water quality assessment." *Water Resour. Manage.* 30 (9): 3043–3061. https://doi.org/10.1007/s11269-016-1330-x.

NRC (National Research Council). 2001. *Assessing the TMDL approach to water quality management*. Washington, DC: National Academies Press.

Ocampo-Duque, W., C. Osorio, C. Piamba, M. Schuhmacher, and J. L. Domingo. 2013. "Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems: Application to the Cauca River, Colombia." *Environ. Int.* 52: 17–28. https://doi.org/10.1016/j.envint.2012.11.007.

Paul, S., P. Haan, M. Matlock, S. Mukhtar, and S. Pillai. 2004. "Analysis of the HSPF water quality parameter uncertainty in predicting peak in-stream fecal coliform concentrations." *Trans. ASAE* 47 (1): 69–78. https://doi.org/10.13031/2013.15872.

Ratto, M., S. Tarantola, and A. Saltelli. 2001. "Sensitivity analysis in model calibration: GSA-GLUE approach." *Comput. Phys. Commun.* 136 (3): 212–224. https://doi.org/10.1016/S0010-4655(01)00159-X.

Reckhow, K. 2003. "On the need for uncertainty assessment in TMDL modeling and implementation." *J. Water Resour. Plann. Manage.* 129 (4): 245–246. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:4(245).

Rode, M., G. Arhonditsis, D. Balin, T. Kebede, V. Krysanova, A. Van Griensven, and S. E. Van der Zee. 2010. "New challenges in integrated water quality modelling." *Hydrol. Processes* 24 (24): 3447–3461. https://doi.org/10.1002/hyp.7766.

Shirmohammadi, A., et al. 2006. "Uncertainty in TMDL models." *Trans. ASABE* 49 (4): 1033–1049. https://doi.org/10.13031/2013.21741.

Singh, V. P., and D. A. Woolhiser. 2002. "Mathematical modeling of watershed hydrology." *J. Hydrol. Eng.* 7 (4): 270–292. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270).

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder. 2008. "Appraisal of the generalized likelihood uncertainty estimation (GLUE) method." *Water Resour. Res.* 44 (12): W00B06. https://doi.org/10.1029/2008WR006822.

Stow, C. A., K. H. Reckhow, S. S. Qian, E. C. Lamon, G. B. Arhonditsis, M. E. Borsuk, and D. Seo. 2007. "Approaches to evaluate water quality model parameter uncertainty for adaptive TMDL implementation." *J. Am. Water Resour. Assoc.* 43 (6): 1499–1507. https://doi.org/10.1111/j.1752-1688.2007.00123.x.

TMDL Analysis and Modeling Task Committee. 2017. *Total maximum daily load analysis and modeling: Assessment of the practice*. Reston, VA: ASCE.

US Congress. 1972." The clean water act." Accessed November 27, 2002. https://www.epa.gov/sites/production/files/2017-08/documents/federal-water-pollution-control-act-508full.pdf.

USEPA (US Environmental Protection Agency). 2000. *Basins technical note 6: Estimating hydrology and hydraulic parameters for HSPF*. EPA-823-R00-012. Washington, DC: USEPA.

USEPA (US Environmental Protection Agency). 2001. *The national costs of the total maximum daily load program*. EPA 841-D-01-003. Washington, DC: USEPA.

USGS (US Geological Survey). 2002. "The national elevation dataset." Accessed June 15, 2002. https://nationalmap.gov/elevation.html.

van Straten, G. 1998. "Models for water quality management: The problem of structural change." *Water Sci. Technol.* 37 (3): 103–111. https://doi.org/10.2166/wst.1998.0185.

Whitehead, P., R. Wilby, R. Battarbee, M. Kernan, and A. J. Wade. 2009. "A review of the potential impacts of climate change on surface water quality." *Hydrol. Sci. J.* 54 (1): 101–123. https://doi.org/10.1623/hysj.54.1.101.

Willems, P. 2008. "Quantification and relative comparison of different types of uncertainties in sewer water quality modeling." *Water Res.* 42 (13): 3539–3551. https://doi.org/10.1016/j.watres.2008.05.006.

Wu, J., R. Zou, and S. L. Yu. 2006. "Uncertainty analysis for coupled watershed and water quality modeling systems." *J. Water Resour. Plann. Manage.* 132 (5): 351–361. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:5(351).

Xie, Y., and G. Huang. 2014. "Development of an inexact two-stage stochastic model with downside risk control for water quality management and decision analysis under uncertainty." *Stochastic Environ. Res. Risk Assess.* 28 (6): 1555–1575. https://doi.org/10.1007/s00477-013-0834-7.

Yagow, G. 2001. *Fecal coliform TMDL: Mountain run watershed, Culpeper County, Virginia*. Richmond, VA: VADEQ.

Zeckoski, R., B. Benham, S. Shah, M. Wolfe, K. Brannan, M. Al-Smadi, T. Dillaha, S. Mostaghimi, and C. Heatwole. 2005. "BSLC: A tool for bacteria source characterization for watershed management." *Appl. Eng. Agric.* 21 (5): 879–889. https://doi.org/10.13031/2013.19716.

Zheng, Y., and F. Han. 2016. "Markov Chain Monte Carlo (MCMC) uncertainty analysis for watershed water quality modeling and management." *Stochastic Environ. Res. Risk Assess.* 30 (1): 293–308. https://doi.org/10.1007/s00477-015-1091-8.

Zheng, Y., and A. A. Keller. 2007. "Uncertainty assessment in watershed-scale water quality modeling and management. Part I: Framework and application of generalized likelihood uncertainty estimation (GLUE) approach." *Water Resour. Res.* 43 (8): W08407. https://doi.org/10.1029/2006WR005345.