Science of the Total Environment 702 (2020) 135022

Contents lists available at ScienceDirect

Science of the Total Environment

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

Risk-based decision making to evaluate pollutant reduction scenarios

Ebrahim Ahmadisharaf [a,*], Brian L. Benham [b]

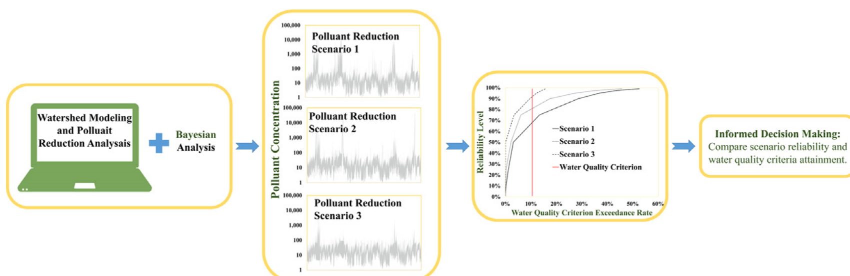
[a] DHI, Lakewood, CO 80228, USA

[b] Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

HIGHLIGHTS

- Bayesian framework for evaluating pollutant reduction scenarios.
- Framework allows decision makers to understand pollutant reduction risk/ reliability.
- Application for a bacteria TMDL developed with HSPF model.
- Aggressive load reduction is useful only if a medium-high reliability is desired.
- Achieving bacteria water quality goals with high reliability seems challenging.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 September 2019
Received in revised form 12 October 2019
Accepted 15 October 2019
Available online 3 November 2019

Editor: Jay Gan

Keywords:
Total maximum daily load (TMDL)
Risk-based decision making
Generalized Likelihood Uncertainty Estimation (GLUE)
Watershed modeling
Uncertainty analysis
Nonpoint source pollution

ABSTRACT

A total maximum daily load (TMDL) is required for water bodies in the U.S. that do not meet applicable water quality standards. Computational watershed models are often used to develop TMDL pollutant reduction scenarios. Uncertainty is inherent in the modeling process. An explicit uncertainty analysis would improve model performance and result in more robust decision making when comparing alternative pollutant reduction scenarios. This paper presents a risk-based framework for evaluating alternative pollutant allocation scenarios considering reliability in achieving water quality goals. We demonstrate a generic routine for the application of Generalized Likelihood Uncertainty Estimation (GLUE) to Hydrological Simulation Program-FORTRAN (HSPF) using existing softwares to evaluate two bacteria reduction scenarios from a recently developed TMDL that addressed a bacterial impairment in a mixed land use watershed in Virginia, U.S. Our probabilistic analysis showed that for reliability levels <25%, the recommended TMDL bacterial load reduction scenario had the same exceedance rate as the full reduction scenario (fully reducing all bacterial loads except wildlife), while for reliability levels between 25% and 50%, the exceedance rates for the two pollutant reduction scenarios were similar, with the TMDL recommended scenario violating the water quality criteria only slightly more often. The full reduction scenario performed better in higher reliability levels, although it could not meet the water quality criteria. Our results indicated that, in this case, achieving water quality goals with very high reliability was not possible, even with extreme levels of pollutant reduction. The risk-based framework presented here illustrates a method to propagate watershed model uncertainty and assess performance of alternative pollutant reduction scenarios using existing tools, thereby enabling decision makers to understand the reliability of a given scenario in achieving water quality goals.

* Corresponding author at: DHI, 141 Union Blvd., Lakewood, CO 80228, USA.
E-mail address: ebra@dhigroup.com (E. Ahmadisharaf).

https://doi.org/10.1016/j.scitotenv.2019.135022
0048-9697/

1. Introduction

Section 303(d) of the Clean Water Act (U.S. Congress, 1972) mandates that a total maximum daily load (TMDL) be developed for water bodies in the U.S. that do not meet applicable water quality standards. A TMDL specifies the amount of a particular pollutant that the water body can assimilate without violating applicable water quality standards. Computational models are widely used in TMDL development to simulate fate and transport of pollutants and to evaluate existing watershed conditions and proposed pollutant reduction scenarios (Borah et al., 2019). These models are subject to many sources of uncertainty, including the accuracy and spatiotemporal resolution of input datasets (e.g., climatic, hydrologic, geologic, geomorphologic and pedologic), the model structure and parameters, the model calibration and validation strategy (e.g., optimization method) and dataset (e.g., streamflow and water quality), the spatial resolution associated with the watershed delineation and the choice of fit measures (Ahmadisharaf et al., 2019; Shirmohammadi et al., 2006). A margin of safety (MOS) is mandatory in TMDLs to reflect these uncertainties. In practice, the MOS estimation is included by either making conservative modeling-related assumptions (implicit) or by explicitly adding a subjectively determined fraction of the estimated point and nonpoint source loads to the TMDL (ASCE-EWRI TMDL Analysis and Modeling Task Committee, 2017; Dilks and Freedman, 2004; Reckhow, 2003; Shirmohammadi et al., 2006).

While some studies have sought to apply uncertainty analysis techniques to watershed-scale water quality models used to simulate nutrient and sediment loads, other pollutants like bacterial pathogens have received limited attention. These include the application of: first-order error analysis by Paul et al. (2004); Monte Carlo simulation, Latin Hypercube Sampling (LHS) (McKay et al., 1979) and Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) by Shirmohammadi et al. (2006); Monte Carlo simulation by Wu et al. (2006) and Mitsova-Boneva and Wang (2007); GLUE by Jia and Culver (2008); stochastic analysis of model residuals by Chin (2009); Markov chain Monte Carlo (MCMC) (Kass et al., 1998) and GLUE by Mishra et al. (2018); and Monte Carlo simulation and two-phase Monte Carlo simulation by Mishra et al. (2019). The reader is referred to Ahmadisharaf et al. (2019) for a detailed review on uncertainty analysis in watershed modeling. Despite these limited efforts, uncertainty of watershed-scale bacteria modeling is not routinely performed to compare alternative pollutant reduction scenarios.

While uncertainty analysis has the potential to allow decision makers to make informed decisions about the relative risk/reliability of alternate pollutant reduction scenarios (Jia and Culver, 2008), in practice, clearly communicating the results of an uncertainty analysis to stakeholders is challenging (Stow et al., 2007). Ocampo-Duque et al. (2013) and Xie and Huang (2014) emphasized the need for more efficient methods to communicate the uncertainty associated with water quality modeling to decision makers. Simple transparent approaches, which could be effectively communicated to the public, need to be developed.

The primary objective of this research is to present a risk-based framework that provides the reliability of a given pollutant reduction scenario being able to achieve a given water quality criterion. We seek to answer the following questions: What is the risk that a pollutant allocation scenario achieves water quality goals? and How much will the risk of not achieving water quality goals be reduced with additional reductions in pollutant loads? A further objective is to present a generic framework for application of a Bayesian method (GLUE) for uncertainty analysis of a widely used watershed model, Hydrological Simulation Program-FORTRAN (HSPF) (Bicknell et al., 2005), when used for modeling *E. coli* (EC) fate and transport by leveraging the capabilities of an existing open-source software package, HSPF Enhanced Expert System (HSPEXP+) (Mishra et al., 2017). The risk-based framework presented here illustrates and approach to propagate watershed model uncertainty into the performance of pollutant reduction scenarios, thereby enabling decision makers to better understand the reliability level of a given scenario with respect to achieving a given water quality goal.

Our study has three major novel aspects compared to previous watershed-scale water quality modeling uncertainty analysis efforts. First, we use an objective likelihood function in Bayesian modeling, while past studies like Jia and Culver (2008) and Mishra et al. (2018) used subjective likelihood functions. As described by Camacho et al. (2018), water quality calibration using objective measures could improve the model performance. Second, we present a generic framework for application of GLUE with a watershed-scale water quality model using an existing open-source software package. The lack of accessible tools often hinders probabilistic watershed modeling and an explicit uncertainty analysis. Third, we present the performance of alternative pollutant load reduction scenarios as a quantitative transparent risk measure that enables modelers to communicate complex uncertainty analyses with stakeholders. This is relevant because lack of transparent measures often hinders watershed modelers as they try to effectively communicate model uncertainties to stakeholders.

2. Case study

The 20.1 km^2 Woods Creek watershed in Rockbridge County and the City of Lexington in west-central Virginia, U.S. (Fig. 1), was selected to demonstrate the risk-based decision making framework. The 9.7 km main stream of Woods Creek—from the headwaters to the confluence with Maury River—was declared impaired due to excessive EC loading. A bacteria TMDL was developed for Woods Creek and approved by the U.S. Environmental Protection Agency (USEPA) (Benham et al., 2018). This watershed is predominantly residential (48.1%), followed by pasture (28.1%) and forest (23.7%). Average annual precipitation is 1,029 mm and the average daily temperature ranges from −6°C to of 31 °C. Primary sources of EC are nonpoint sources, including runoff from pasture (60%) and residential (9%) as well as direct deposit by livestock (20%) and wildlife (10%) (Benham et al., 2018). Other EC sources are pets, sanitary sewer overflows, failing septic systems and a general sewage permit, infiltration and inflow (I&I) as well as exfiltration of the sewer pipes.

The Woods Creek bacteria TMDL (Benham et al., 2018) included two specific TMDL pollutant load reduction scenarios (Table 1). An implicit MOS was included by conservatively estimating the pollutant source loads. No effort was made to explicitly incorporate uncertainties when evaluating the recommended 'TMDL' pollutant reduction scenario. Benham et al. (2018) compared the recommended TMDL pollutant reduction scenario alongside existing conditions (no action scenario) and a hypothetical reduction scenario that required full reduction of all bacterial sources except wildlife ('full reduction' scenario; Table 1). The reason that wildlife load reductions were not considered in the full reduction scenario was that reducing this load is often not practical and the reduction of sources from human activities are generally favored. Both the TMDL and full reduction scenarios require that bacteria loading from the failing septic systems and sewer overflow be fully reduced, but the full reduction scenario requires additional reductions in other sources. The implementation of the proposed TMDL will be challenging, given the large load reductions, particularly from pasture and livestock direct deposit sources.

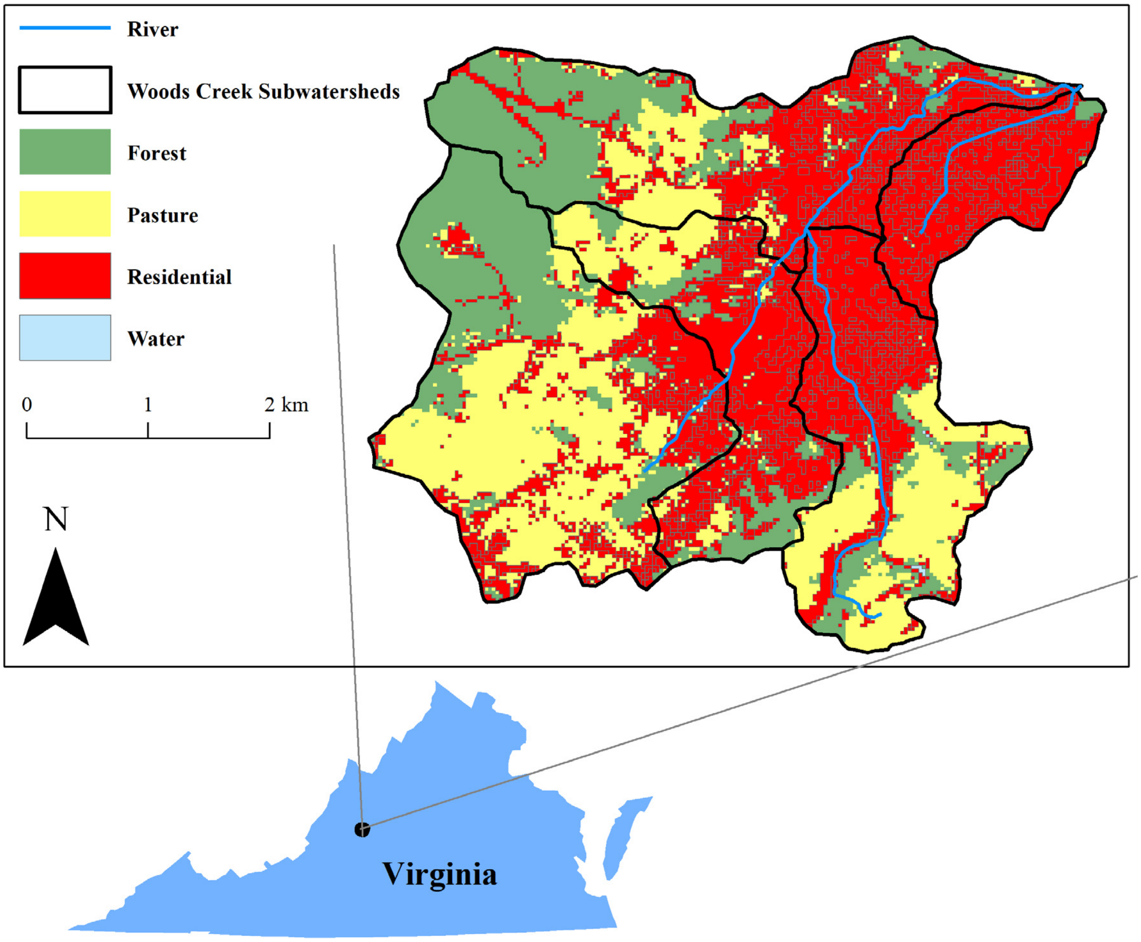


Fig. 1. Woods creek watershed.

Table 1
Source-specific *E. coli* (EC) load reductions (%) in pollutant allocation scenarios.

TMDL allocation	Source-specific EC load reductions (%)					
	Livestock direct deposit	Pasture	Failing septic systems	Sewer overflow	Residential	Wildlife direct deposit
Existing	0	0	0	0	0	0
TMDL	85	70	100	100	25	0
Full Reduction	100	100	100	100	100	0

TMDL: Total maximum daily load.

2.1. Watershed modeling in the study watershed

We used a semi-distributed watershed model, HSPF, to simulate the fate and transport of EC. HSPF simulates water quantity and quality processes on the land surface, in soil profiles and within stream reaches using three main modules: pervious land (PERLND), impervious land (IMPLND) and reach (RCHRES) (Bicknell et al., 2005). Bacteria is typically simulated as a planktonic constituent. Multiple parameters are used to simulate in-stream EC fate and transport. The EC loading rates depend on various factors, including species-specific feces production rates and fecal densities, animal density, die-off rates and the fraction of time livestock are confined (Zeckoski et al., 2005).

The deterministic HSPF model for the Woods Creek watershed (Benham et al., 2018) was used as the initial setup for this study. The watershed was delineated into five subwatersheds using a 10 m DEM from the U.S. Geological Survey's (USGS') National Elevation Dataset. Climate datasets, including rainfall, evaporation, percent sun, wind speed and solar radiation, were collected from two National Climate Date Center stations (COOP IDs 444876 and 445120). Dew point temperature, which was not reported in the two stations, was collected from NASA's (National Aeronautics and Space Administration's) Prediction of Worldwide Energy Resource project. Other datasets required to simulate in-stream EC using HSPF included soil, land use, EC daily loading rates by livestock, pets and wildlife, permitted point sources as well as inflows and withdrawals from springs, golf courses and country clubs. Animal population was initially taken from the U.S. Department of Agriculture census database and was later nominally adjusted during several meetings with local stakeholders. In-stream bacteria data were collected at watershed outlet by the Virginia Department of Environmental Quality (VADEQ) during 2007–2018 and 2015–2016 periods. Limited daily streamflow data were also collected in 2015–2016 near the outlet by USGS (gauge #

0202304110). All these observations were used to calibrate and validate the HSPF Woods Creek model. The reader is referred to Benham et al, (2018) for more information.

3. Methodology

3.1. Generalized Likelihood Uncertainty Estimation (GLUE)

Bayesian inference 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 non-linearity of a given model (Camacho et al., 2015). The basic premise of GLUE is equifinality; i.e., there is not a single optimal parameter set for a given model, rather, multiple parameter sets can satisfactorily represent a system (Beven and Binley, 1992). In GLUE, Monte Carlo simulation is often employed to generate numerous random sets of model parameters from prior distributions. The model is run using a set of parameters sampled from prior distributions. A likelihood weight is then calculated for each parameter set as a function of the variance of the residuals. In this case, the likelihood weight is a function of the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970):

$$L = (NSE)^{-N} \tag{1}$$

where, L is the likelihood value, NSE is Nash-Sutcliffe efficiency and N is the shaping parameter. As the value of N increases, the difference between the likelihood values of parameter sets with similar variance increases and simulations with a greater performance receive more accentuation (Freer et al., 1996). Different values of N can lead to different model output uncertainty (Ratto et al., 2001) and the greater the N, the narrower the uncertainty (Freer et al., 1996). Beven and Freer (2001) recommended against using very high values for N as these tend to underestimate the uncertainty due to the peakedness of likelihood surface. Our use of $N = 2$ was informed by previous GLUE applications with HSPF (Mishra et al., 2018).

Initial simulations from prior distributions are divided into behavioral and non-behavioral sets based on likelihood values. Behavioral simulations can be determined by using either a cutoff threshold (a.k.a. limits of acceptability approach) (Beven, 2006) or a percentage of the simulations with the 'best' likelihood values (i.e., greatest correlation or least error). The former approach was used in this study. Likelihood values of the behavioral parameter sets were normalized to unity. These 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. Parameter-specific plots of the normalized behavioral likelihoods versus parameter values (a.k.a. dotty plots) defined the cumulative probability and posterior parameter distribution (Beven and Binley, 1992). The normalized likelihoods for different parameter values were multiplied to the prior probability to derive the posterior distributions that provide a description of parameter uncertainty adjusted by the observations. Additional information about GLUE can be found in Beven and Freer (2001), Stedinger et al. (2008) and Beven and Binley (2014).

3.2. Probabilistic watershed modeling

3.2.1. Uncertain model parameters

HSPF uses several hydrologic and water quality parameters to simulate EC. Sensitive hydrologic and water quality parameters were initially determined from the literature (Ahmadisharaf et al., 2019; Chin et al., 2009; Im et al., 2007; Jia and Culver, 2008; Lawson, 2003; Mishra et al., 2019, 2018; Yagow, 2001). We then refined these parameters based on the deterministic HSPF modeling of the study watershed (Benham et al., 2018). The ultimate uncertain parameters were initially assigned uniform distributions that were adjusted via GLUE during model calibration. Plausible ranges of the selected uncertain parameters were derived based on the minimum and maximum values recommended in the literature (Mishra et al., 2019; USEPA, 2000). In HSPF, the EC accumulation rate on land is represented by two tables, ACCUM and SQOLIM, which vary monthly due to the application of animal manures to agricultural fields. The basic values of these two were derived from the Bacteria Source Load Calculator (BSLC), a tool for bacteria source characterization (Zeckoski et al., 2005). To

Table 2
Probability distribution of the uncertain parameters.

Parameter	Parameter description	Parameter distribution
AGWETP	Fraction of remaining evapotranspiration from active groundwater	Uniform (0.00, 0.20)[1]
AGWRC	Base groundwater recession	Uniform (0.85, 0.99)
BASETP	Fraction of remaining evapotranspiration from baseflow	Uniform (0.00, 0.20)
CEPSC (mm)	Interception storage capacity	Uniform (0.25, 10.16)
DEEPFR	Fraction of groundwater inflow to deep recharge	Uniform (0.00, 0.50)
INFEXP	Exponent in infiltration equation	Uniform (1.00, 3.00)
INFILD	Ratio of max/mean infiltration capacities	Uniform (1.00, 2.00)
INFILT	Index to infiltration capacity	Uniform (0.03, 12.70)
INTFW	Interflow inflow parameter	Uniform (1.0, 10.0)
IRC	Interflow recession coefficient	Uniform (0.30, 0.85)
KS	Weighting factor for hydraulic routing	Uniform (0.00, 0.99)
KVARY (mm^{-1})	Non-linear groundwater recession rate	Uniform (0.00, 0.20)
LZETP	Index to lower zone evapotranspiration	Uniform (0.10, 0.90)
LZSN (mm)	Lower zone nominal soil moisture storage	Uniform (50.80, 381.00)
UZSN (mm)	Upper zone nominal soil moisture storage	Uniform (1.27, 50.80)
ACCUM and SQOLIM adjustment factor	Adjustment factor for monthly ACCUM and SQOLIM tables (maximum bacteria accumulation on pervious land)	Uniform (0.10, 10.00)
FSTDEC (day^{-1})	First order die-off rate	Uniform (0.12, 2.52)
IOQC (cfu[2] m^{-3})	Bacteria concentration on the interflow outflow from pervious land surface	Uniform (1246, 124600)
WSQOP ($mm\ hr^{-1}$)	Rate of surface runoff that removes 90% of stored bacteria from pervious land surface	Uniform (0.25, 12.70)

[1] Numbers in parentheses show lower and upper limits of the uniform distribution.
[2] cfu: colony forming units.

perturb the uncertainty of ACCUM and SQOLIM, the basic values were multiplied by an adjustment factor, which was assumed to range from 0.1 and 10 (Jia and Culver, 2008; Mishra et al., 2018). Table 2 lists the prior distributions of the HSPF parameters.

3.2.2. Probabilistic watershed modeling

Ensemble parameter samples were generated by LHS with random point in strata. LHS was preferred to standard Monte Carlo simulation as past studies like Janssen (2013) have shown that it is more efficient in numerical convergence. We assumed no correlation between the model parameters because while some researchers like Stow et al. (2007) suggested that considering a correlation between the variables improves the sampling accuracy, others like Omlin et al. (2001) and Wu et al. (2006) argued that a correlation necessarily results in less sampling error.

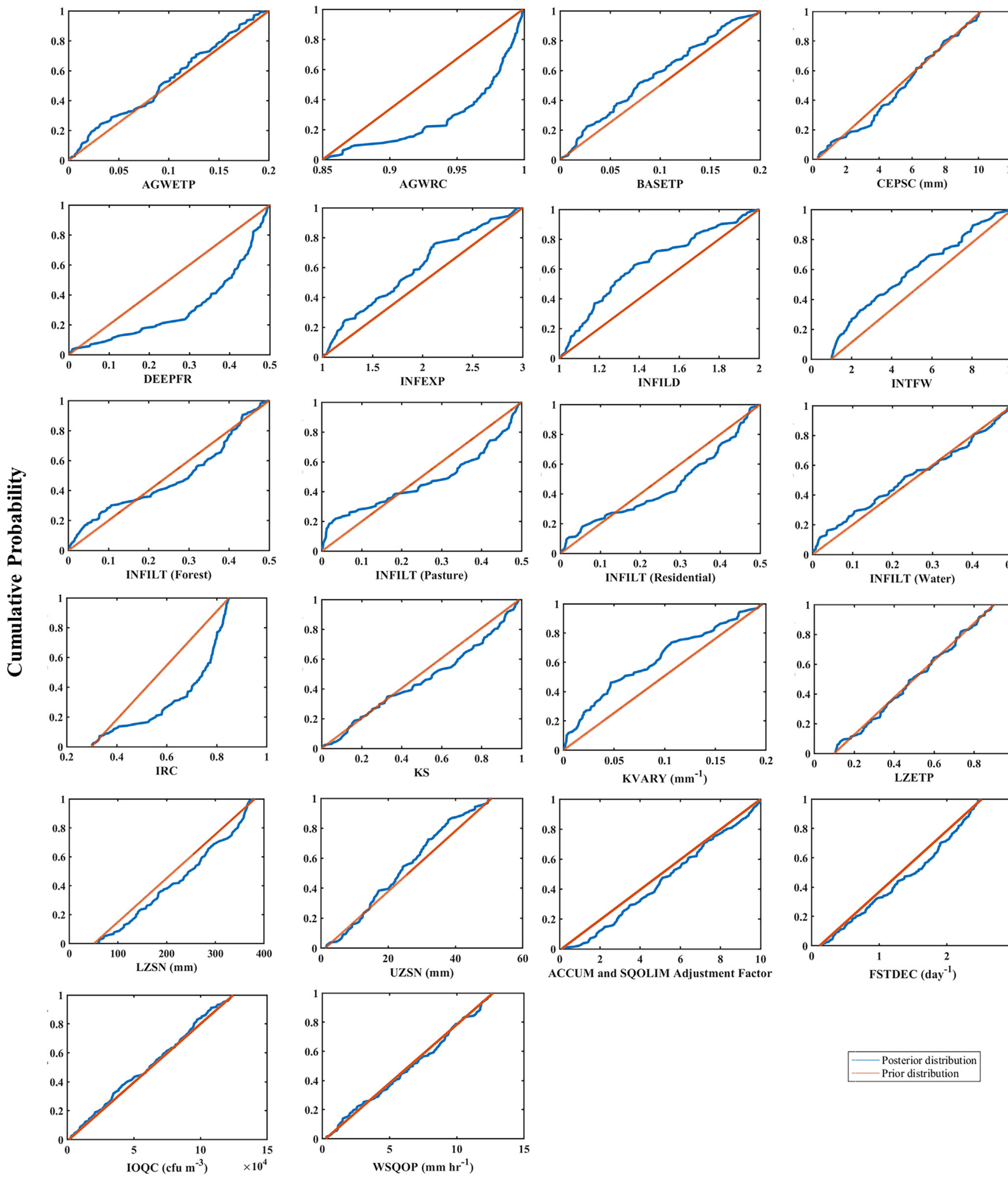


Fig. 2. Posterior cumulative distributions of the model parameters.

Generated random values from the prior parameter distributions were fed into a comma-separated values (CSV) file, the primary input for the HSPEXP+. We leveraged batch capabilities of HSPEXP+ to perform ensemble simulations. Numerous HSPF simulations were performed using the prior parameter distributions. Ensemble model output time series generated by HSPEXP+ were then post-processed to calculate NSE values and determine likelihoods.

A likelihood cutoff threshold was used to determine behavioral simulations. A NSE = 0.5 was used to determine behavioral hydrologic simulations (Ahmadisharaf et al., 2019; Moriasi et al., 2015). For water quality, the threshold was extracted from past studies where NSE was used to evaluate bacteria simulation model performance (Baffaut and Sadeghi, 2010; Cho et al., 2010; Coffey et al., 2013; Hernandez-Suarez et al., 2019; Niazi et al., 2015; Pandey et al., 2012; Parajuli et al., 2009). Because a wide range of NSE values has been applied in the past studies (from −6.0 to 0.8), we used a cutoff threshold of −0.4, which resulted in a reasonable number of behavioral simulations (about 10% of the total simulations). It is notable that while this threshold seems low for watershed-scale simulation of other constituents (e.g., streamflow, nutrients and sediment), it is reasonable for bacteria modeling. The reader is referred to (Ahmadisharaf et al., 2019) for detailed discussion of bacterial simulation calibration.

A generic code was developed to derive posterior distributions via GLUE. Implemented in MATLAB, the code can separate behavioral simulations using either a cutoff threshold or some percentage of the best simulations. The primary input is a spreadsheet that includes generated random parameters from the prior distributions, values of the likelihood function (NSE in this case), cutoff threshold for the likelihood function and the shaping parameter value. Taking this input file, the code produces the number of behavioral samples and parameter-specific posterior distributions. A number of random values can be then generated from the derived posterior distributions. The generated random numbers from posterior distributions were written to the CSV input file of the HSPEXP+ to perform ensemble watershed simulations. This process was repeated in model calibration-validation and simulation of pollutant allocation scenarios (Table 1).

We used a two-stage process to calibrate and validate the probabilistic watershed model via GLUE. A hydrologic calibration was first done to derive posterior distributions of the hydrologic parameters, which were used in subsequent water quality simulations. A water quality calibration was then performed to derive posterior distributions of the HSPF water quality parameters. These distributions were later validated against an independently observed bacteria dataset. Three metrics were used to evaluate model performance during the validation period: (a) the number of observations that fall inside the prediction intervals of average daily EC concentration (Mishra et al., 2018), (b) the number of observations that fall inside the prediction intervals of five-day average daily EC concentration (Kim et al., 2007) and (c) the number of median of ensemble simulated average daily EC that fall inside one-order of magnitude of the observations (Dorner et al., 2006). These three measures should be greater than 70% for satisfactory model performance (Dorner et al., 2006; Kim et al., 2007; Mishra et al., 2018; Pandey et al., 2012).

3.3. Risk-based decision making

The allocation scenarios were modeled using the probabilistic HSPF model. A six-year period (1999, 2001, 2003, 2009, 2015 and 2016) was used in allocation simulations. The simulation period time series included two average rainfall years, two dry years (below average rainfall), two wet years (above average rainfall) and a range of water quality conditions (low and high bacteria concentrations) (Benham et al., 2018). An ensemble of simulated daily average in-stream EC concentration time series was used to derive percentiles (e.g., 5th and 95th) and prediction intervals (e.g., 90%). These time series were used to estimate the level of reliability (as opposed to risk) of achieving two water quality criteria: (i) daily exceedance rate of the maximum water quality assessment (a.k.a. 'daily single-sample' criterion) calculated as the number of days with EC concentration >235 cfu/100 mL divided by the number of days in the simulation period; and (ii) a monthly geometric mean criterion calculated as the number of calendar months with geometric mean (of the average daily EC concentrations) > 126 cfu/100 mL. The former criterion must be less than 10.5% exceedance rate, while the latter criterion requires a zero exceedance rate. For each allocation scenario, an empirical cumulative distribution was constructed by computing different percentiles of the exceedance rates (a p^{th} percentile means p% reliability or [100-p]% risk in achieving the water quality criterion). This distribution visualizes reliability levels in achieving the desired water quality target. Using this distribution, decision makers can prioritize alternative pollutant reduction scenarios, selecting the scenario that meets the applicable water quality target at a desired level of reliability.

4. Results and discussion

4.1. Model calibration and validation

Hydrologic calibration was performed to derive posterior distribution of the 18 hydrologic parameters and four water quality parameters for the 2015–2016 period. Of the 5,000 LHS realizations that were settled for hydrologic calibration, only 110

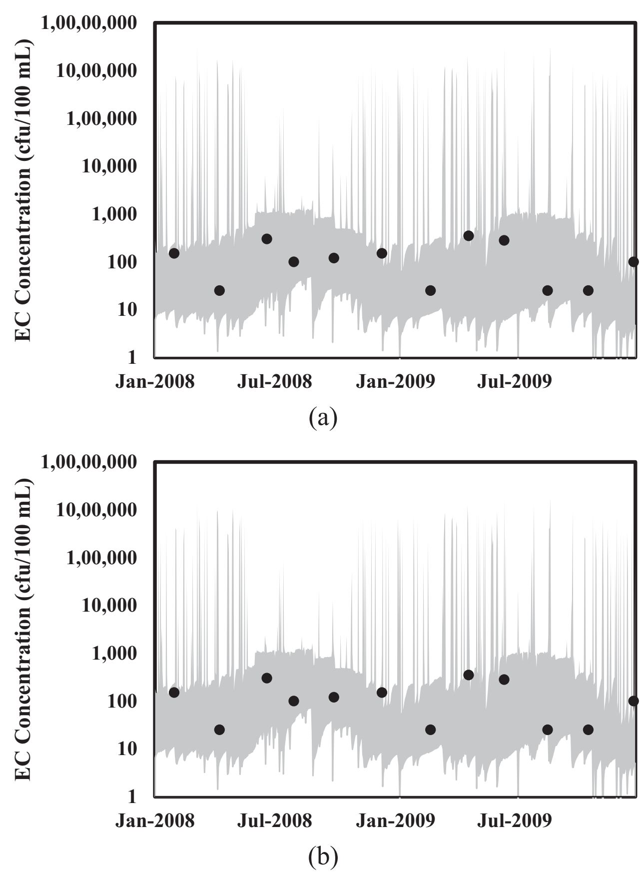


Fig. 3. (a) 95% and (b) 90% prediction intervals of the simulated in-stream *E. coli* (EC) bacteria alongside observations in the validation period.

simulations were behavioral (NSE > 0.5). Using posterior distribution of the hydrologic parameters, 1,000 LHS realizations were settled for water quality calibration. Only 91 of these simulations were behavioral (NSE > −0.4). The prior and posterior distributions of HSPF parameters, which were derived via GLUE, are presented in Fig. 2. The difference between prior and posterior distributions implied that the model was most sensitive to AGRWC, DEEPFR, INFILD, IRC and KVARY for the hydrologic simulation and ACCUM/SQOLIM adjustment factor and FSTDEC for the EC simulation. Similar to past probabilistic HSPF studies like Mishra et al. (2018), the difference was greater for the hydrologic parameters.

Next, we validated the posterior distributions of the water quality parameters against observed EC data in the 2007–2008 period. Using either 90% or 95% prediction interval, 91.7% of the observations fell inside the predicted daily EC (Fig. 3). This percentage reduced to 83.3% when 80% prediction interval was used. All the observations fell inside prediction intervals >80% of the five-day average daily EC concentration and 83.3% of the ensemble median of simulated average daily EC concentration were within an order of the magnitude of the observations. All the three metrics were therefore met and the validation was judged satisfactory.

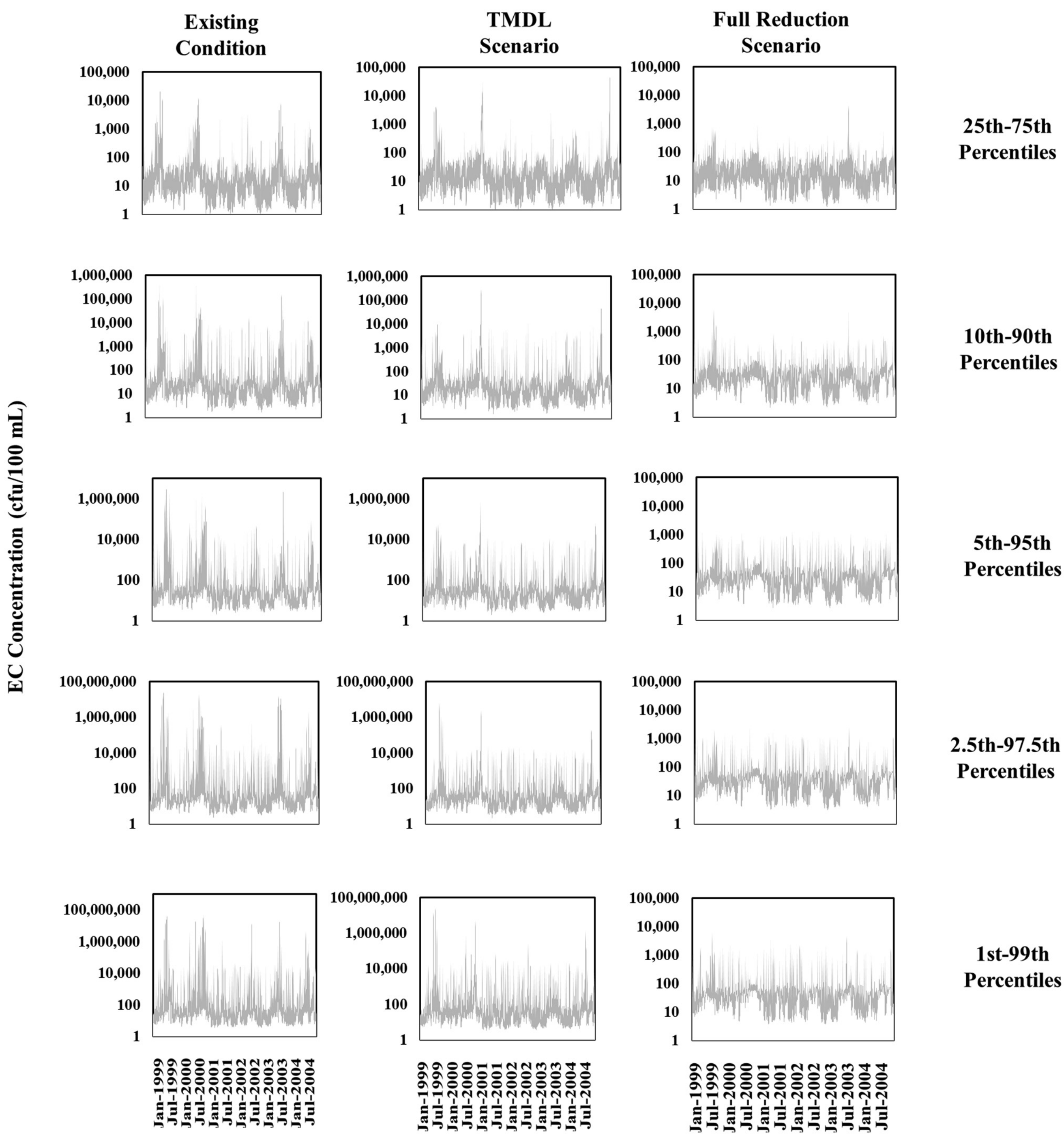


Fig. 4. Percentile time series of simulated in-stream *E. coli* (EC) bacteria for the existing, total maximum daily load (TMDL) and full reduction scenarios.

4.2. Risk-based decision making

Pollutant allocation scenarios (Table 1) were simulated using the probabilistic watershed model, which used posterior distributions. One-thousand LHS simulations were performed via HSPEXP + for the six-year simulation period. The daily average simulated EC concentration time series from each LHS simulation was used to determine the 1st, 2.5th, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 97.5th and 99th percentiles (50%, 80%, 90%, 95% and 98% prediction intervals) for the existing, TMDL and full reduction scenarios (Fig. 4).

We post-processed the model ensemble outputs to derive the exceedance rates of the daily single-sample and monthly geomean water quality criteria for the two pollutant reduction scenarios, the TMDL scenario and the full reduction scenario (Fig. 5). The reliability level at which the deterministically determined TMDL scenario achieved the daily single-sample and monthly geomean water quality criterion was 82.6% and 29.1%, respectively. For the full reduction scenario, the reliability of achieving the daily single-sample and monthly geomean water quality criterion was 91.7% and 53.2%, respectively. Achieving the applicable water quality criteria with a higher degree of reliability would require wildlife load reduction, which is generally not favored and might be impractical.

For reliability levels <25%, the TMDL scenario had the same exceedance rate as the full reduction scenario, while for reliability levels between 25% and 50%, the exceedance rates for the two pollutant reduction scenarios were similar, with the TMDL scenario violating both water quality criteria only slightly more. For reliability levels >50%, however, the TMDL scenario violated the water quality criteria more often. The exceedance rate for the daily single-sample criterion was as high as six times greater (80.6% vs. 12.5% at the 99th percentile) for the TMDL scenario when compared to the full reduction scenario, with a much greater difference for the monthly geomean criterion. These results suggest that higher levels of load reduction are more useful in situations where very high reliability (i.e., very low risk) is warranted.

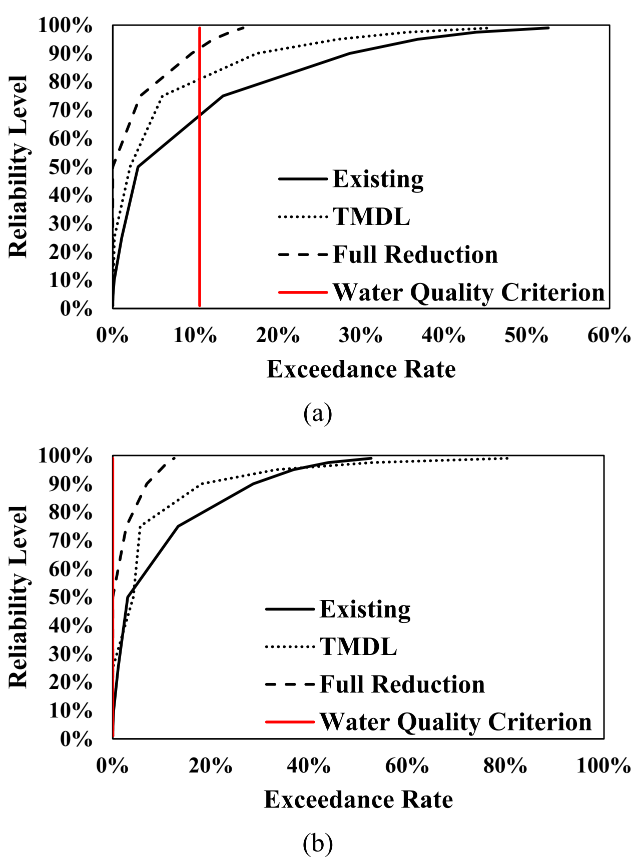


Fig. 5. (a) Daily single-sample and (b) monthly geomean exceedance rates for the existing conditions, total maximum daily load (TMDL) and full pollutant reduction scenarios.

We also compared our probabilistic analyses of the TMDL scenario from the deterministic analysis by Benham et al. (2018). For the monthly geomean water quality criterion, the deterministic model predicted an exceedance rate of 0.0%, while we predicted a range of 0% (2.5th percentile) to 36.1% (97.5th percentile), with a median of 11.1%. For the daily single-sample water quality criterion, the deterministic model predicted a violation rate of 0.1%, while the probabilistic model predicts a range of 0.0% (2.5th percentile) to 33.8% (97.5th percentile), with a median rate of 3.0%. This inconsistency between the reliability of pollutant reduction scenarios developed deterministically and probabilistically is consistent with the analysis of others (Borsuk et al., 2002; Langseth and Brown, 2011; Mishra et al., 2018). The risk-based framework presented here allows the decision maker to more effectively compare alternative pollutant reduction scenarios, and provides a systematic approach to include uncertainty in TMDL development using Bayesian inference, which has direct impacts on public and ecosystem health.

For nearly two decades, researchers have debated the practice of uncertainty analysis and MOS estimation in TMDLs. The National Research Council (2001) recommended the use of more rigorous, less subjective approaches in MOS estimation. More recently, ASCE-EWRI TMDL Analysis and Modeling Task Committee (2017) recommended the need for major improvements in MOS estimation using scientifically sound uncertainty analysis approaches. Despite these, the practice of uncertainty analysis in TMDLs has yet to advance substantially. The risk-based framework demonstrated in this study supports uncertainty analysis in TMDLs by using existing softwares and paves the way for the application of a scientifically defensible approach to MOS estimation. The simulation tools and implications for management are directly transferable to watershed managers. This is a major step forward in the application of Bayesian inference for TMDLs. Ensemble time series (Fig. 4) and ECDFs (Fig. 5) serve as examples to watershed modelers studying complex TMDLs in the U.S. and more broadly pollution mitigation worldwide. Decision makers could use this type of analysis and information to select a pollutant reduction scenario based on the level of risk and reliability associated with achieving specific water quality targets. The framework helps watershed managers answer the following question: At what risk, will you implement this pollutant reduction scenario?

Limitations exist in this study suggest the need for future research. The cost of implementation was not considered here. Rather, we focused on a method for analyzing and prioritizing alternative pollutant reduction scenarios during TMDL development. Cost is an important decision criterion that should be examined along with reliability. Prioritizing alternative pollutant load allocation scenarios by considering both reliability and implementation costs would aid in determining how a water quality target can be achieved with a limited budget at different reliability levels. An expensive scenario with high reliability level might be preferable in a more fragile ecosystem (i.e., water quality violation has large consequences), while a less expensive pollutant reduction scenario with lower reliability might be preferable where the ecosystem is more resilient or where funding is limited.

With the batch simulation capabilities of HSPEXP+, GLUE can be used in practice for uncertainty estimation of HSPF-based EC modeling. Additional research is also required to analyze the effects of using different likelihood functions and parameter acceptance criteria when applying GLUE for HSPF-based water quality applications (He et al., 2010). This work can be also replicated by using advanced Bayesian methods such as MCMC. However, implementing GLUE is simpler and more straightforward than complex Bayesian methods like MCMC and the results can be more effectively communicated to stakeholders.

We focused only on epistemic uncertainty resulting from model parameterization. Future research should focus on other epistemic uncertainties such as model inputs (e.g., climate datasets) and structure as well as observed data used for calibration and validation. The uncertainty triggered by a changing climate and land use change, which has been shown to have a major impact on predicted water quality (Fonseca et al., 2015; Whitehead et al., 2009), was also not considered in our study. Future efforts to investigate the impact of these nonstationary stressors are vital for proactive and adaptive watershed management.

5. Summary and conclusions

We presented a risk-based framework for evaluation of alternative pollutant reduction scenarios to meet the water quality criteria associated with a bacteria TMDL. A Bayesian method—GLUE—was applied for uncertainty analysis of watershed-scale HSPF model used to simulate instream bacteria concentrations. Our sole focus was on parametric uncertainty; other sources of uncertainties were not explored. A generic framework was presented for application of GLUE with HSPF-based bacteria modeling using a random generator software and HSPEXP+. Using this framework, we evaluated two alternative pollutant load reduction scenarios—TMDL and full reduction—that were included in a recently developed USEPA-approved bacteria TMDL (Benham et al., 2018) for Woods Creek in Virginia, U.S.

The reliability level at which the deterministically determined TMDL recommended scenario achieved the daily single-sample and monthly geomean water quality criterion was 82.6% and 29.1%, respectively. For the full reduction scenario, the reliability of achieving the daily single-sample and monthly geomean water quality criterion was 91.7% and 53.2%, respectively. Achieving the applicable water quality criteria with a higher degree of reliability would require wildlife load reduction, which is generally not favored and might be impractical. For reliability levels < 25%, the TMDL recommended scenario had the same exceedance rate as the full reduction scenario, while for reliability levels between 25% and 50%, the exceedance rates for the two pollutant reduction scenarios were similar, with the TMDL scenario violating the water quality criteria only slightly more often. For reliability levels >50%, the TMDL recommended scenario violated the water quality criteria more often. These results suggest that higher levels of load reduction are more useful in situations where very high reliability (i.e., very low risk) is warranted. Achieving water quality goals, particularly in this case, the geomean criterion, with very high reliability was not possible even with the full reduction scenario. Higher levels of reliability would require a wildlife source load reduction, which is not practical.

The risk-based framework presented here illustrates a method to propagate watershed model uncertainty into the performance of alternative pollutant reduction scenarios, thereby enabling decision makers to understand the reliability level of a given scenario in achieving water quality goals. To draw more generalized conclusions, future research should explore other sources of uncertainty such as input climate data and model structure. Examining the tradeoffs between the implementation costs and reliability of achieving water quality goals would also help in better understanding the role of uncertainty by translating it into monetary terms.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Anurag Mishra and Karen Kline for the help with HSPEXP+ and HSPF development in the Woods Creek watershed.

References

Ahmadisharaf, E., Camacho, R.A., Zhang, H.X., Hantush, M., Mohamoud, Y.M., 2019. Calibration and validation of watershed models and advances in uncertainty analysis in TMDL studies. J. Hydrol. Eng. 24, 03119001.
ASCE-EWRI TMDL Analysis and Modeling Task Committee, 2017. Total maximum daily load analysis and modeling: Assessment of the practice. Reston, VA: ASCE.
Baffaut, C., Sadeghi, A., 2010. Bacteria modeling with SWAT for assessment and remediation studies: A review. Trans. ASABE 53, 1585–1594.
Benham, B.L., Yagow, G., Tse, W.C.-K., Ahmadisharaf, E., Kline, K.S., 2018. Bacteria TMDL development and a proactive approach to address the benthic impairment for Woods Creek. Rockbridge County and City of Lexington, Virginia, Richmond, VA.
Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36.
Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.
Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897–5918.
Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using GLUE methodology. J. Hydrol. 249, 11–29.
Bicknell, B., Imhoff, J., Kittle, Jr., A., Jobes, T., Donigian, Jr., A., 2005. Hydrological Simulation Program - Fortran (HSPF) User's Manual for Release 12.2, Athens, GA, pp. 845.
Borah, D.K., Ahmadisharaf, E., Padmanabhan, G., Imen, S., Yusuf, M.M., 2019. Watershed models for development and implementation of total maximum daily loads. J. Hydrol. Eng. 24, 03118001.
Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. Environ. Sci. Technol. 36, 2109–2115.
Camacho, R.A., Martin, J.L., McAnally, W., Díaz-Ramirez, J., Rodriguez, H., Sucsy, P., et al., 2015. A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling. JAWRA J. Am. Water Resour. Assoc. 51, 1372–1393.
Camacho, R.A., Martin, J.L., Wool, T., Singh, V.P., 2018. A framework for uncertainty and risk analysis in total maximum daily load applications. Environ. Modell. Software 101, 218–235.
Chin, D.A., 2009. Predictive uncertainty in water-quality modeling. J. Environ. Eng. 135, 1315–1325.
Chin, D.A., Sakura-Lemessy, D., Bosch, D.D., Gay, P.A., 2009. Watershed-scale fate and transport of bacteria. Trans. ASABE 52, 145–154.
Cho, K.H., Pachepsky, Y., Kim, J.H., Guber, A., Shelton, D., Rowland, R., 2010. Release of Escherichia coli from the bottom sediment in a first-order creek: Experiment and reach-specific modeling. J. Hydrol. 391, 322–332.
Coffey, R., Dorai-Raj, S., O'Flaherty, V., Cormican, M., Cummins, E., 2013. Modeling of pathogen indicator organisms in a small-scale agricultural catchment using SWAT. Hum. Ecological Risk Assessment: Int. J. 19, 232–253.
Dilks, D.W., Freedman, P.L., 2004. Improved consideration of the margin of safety in total maximum daily load development. J. Environ. Eng. 130, 690–694.
Dorner, S.M., Anderson, W.B., Slawson, R.M., Kouwen, N., Huck, P.M., 2006. Hydrologic modeling of pathogen fate and transport. Environ. Sci. Technol. 40, 4746–4753.
Fonseca, A., Botelho, C., Boaventura, R., Vilar, V., 2015. Global warming effects on faecal coliform bacterium watershed impairments in Portugal. River Res. Appl. 31, 1344–1353.
Freer, J., Beven, K., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resour. Res. 32, 2161–2173.
He, J., Jones, J.W., Graham, W.D., Dukes, M.D., 2010. Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. Agric. Syst. 103, 256–264.
Hernandez-Suarez, J.S., Woznicki, S.A., Nejadhashemi, A.P., 2019. Multi-site watershed model calibration for evaluating best management practice effectiveness in reducing fecal pollution. Hum. Ecol. Risk Assess.: An Int. J. https://doi.org/10.1080/10807039.2019.1680526. In press.
Im, S., Brannan, K.M., Mostaghimi, S., Kim, S.M., 2007. Comparison of HSPF and SWAT models performance for runoff and sediment yield prediction. J. Environ. Sci. Health, Part A 42, 1561–1570.
Janssen, H., 2013. Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. Reliab. Eng. Syst. Saf. 109, 123–132.
Jia, Y., Culver, T.B., 2008. Uncertainty analysis for watershed modeling using generalized likelihood uncertainty estimation with multiple calibration measures. J. Water Resour. Plann. Manage. 134, 97–106.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M., 1998. Markov chain Monte Carlo in practice: a roundtable discussion. Am. Statistician 52, 93–100.
Kim, S.M., Benham, B.L., Brannan, K.M., Zeckoski, R.W., Yagow, G., 2007. Water quality calibration criteria for bacteria TMDL development. Appl. Eng. Agric. 23, 171–176.
Langseth, D.E., Brown, N., 2011. Risk-based margins of safety for phosphorus TMDLs in lakes. J. Water Resour. Plann. Manage. 137, 276–283.
Lawson, L.G., 2003. HSPF model calibration and verification for bacteria TMDLs. VADEQ, Richmond, VA, Guidance Memo No. 03-2012.
McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.
Mishra, A., Ahmadisharaf, E., Benham, B.L., Gallagher, D.L., Reckhow, K.H., Smith, E. P., 2019. Two-phase Monte Carlo simulation for partitioning the effects of epistemic and aleatory uncertainty in TMDL modeling. J. Hydrol. Eng. 24, 04018058.
Mishra, A., Ahmadisharaf, E., Benham, B.L., Wolfe, M.L., Leman, S.C., Gallagher, D.L., et al., 2018. Generalized likelihood uncertainty estimation and Markov chain Monte Carlo simulation to prioritize TMDL pollutant allocations. J. Hydrol. Eng. 23.
Mishra, A., Bicknell, B.R., Duda, P., Donigian, T., Gray, M.H., 2017. HSPEXP+: An enhanced expert system for hspf model calibration—a case study of the Snake River watershed in Minnesota. J. Water Manage. Modeling.
Mitsova-Boneva, D., Wang, X., 2007. Exploring the variability in suspended sediment yield using BASINS-HSPF and probabilistic modeling: Implications for land use planning. J. Environ. Informatics, 9.
Moriasi, D., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans. ASABE 58, 1763–1785.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10, 282–290.
National Research Council, 2001. Assessing the TMDL approach to water quality management. Washington, D.C: Water Science and Technology Board, Division of Earth and Life Studies.
Niazi, M., Obropta, C., Miskewitz, R., 2015. Pathogen transport and fate modeling in the Upper Salem River Watershed using SWAT model. J. Environ. Manage. 151, 167–177.
Ocampo-Duque, W., Osorio, C., Piamba, C., Schuhmacher, M., Domingo, J.L., 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.
Omlin, M., Brun, R., Reichert, P., 2001. Biogeochemical model of Lake Zürich: sensitivity, identifiability and uncertainty analysis. Ecol. Model. 141, 105–123.
Pandey, P.K., Soupir, M.L., Haddad, M., Rothwell, J.J., 2012. Assessing the impacts of watershed indexes and precipitation on spatial in-stream E. coli concentrations. Ecol. Ind. 23, 641–652.
Parajuli, P.B., Douglas-Mankin, K., Barnes, P.L., Rossi, C., 2009. Fecal bacteria source characterization and sensitivity analysis of SWAT 2005. Trans. ASABE 52, 1847–1858.
Paul, S., Haan, P., Matlock, M., Mukhtar, S., Pillai, S., 2004. Analysis of the HSPF water quality parameter uncertainty in predicting peak in-stream fecal coliform concentrations. Trans. ASAE 47, 69.
Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Comput. Phys. Commun. 136, 212–224.
Reckhow, K., 2003. On the need for uncertainty assessment in TMDL modeling and implementation. J. Water Resour. Plann. Manage. 129, 245–246.
Shirmohammadi, A., Chaubey, I., Harmel, R.D., Bosch, D.D., Munoz-Carpena, R., Dharmasri, C., et al., 2006. Uncertainty in TMDL models. Trans. ASABE 49, 1033–1049.
Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resour. Res. 44, W00B06.
Stow, C.A., Reckhow, K.H., Qian, S.S., Lamon, E.C., Arhonditsis, G.B., Borsuk, M.E., et al., 2007. Approaches to evaluate water quality model parameter uncertainty for adaptive TMDL implementation. JAWRA J. Am. Water Resour. Assoc. 43, 1499–1507.
U.S. Congress. The Clean Water Act, 1972
USEPA, 2000. BASINS technical note 6: Estimating hydrology and hydraulic parameters for HSPF. Office of Water U, Washington, DC.
Whitehead, P., Wilby, R., Battarbee, R., Kernan, M., Wade, A.J., 2009. A review of the potential impacts of climate change on surface water quality. Hydrol. Sci. J. 54, 101–123.
Wu, J., Zou, R., Yu, S.L., 2006. Uncertainty analysis for coupled watershed and water quality modeling systems. J. Water Resour. Plann. Manage. 132, 351–361.
Xie, Y., Huang, G., 2014. Development of an inexact two-stage stochastic model with downside risk control for water quality management and decision analysis under uncertainty. Stoch. Env. Res. Risk Assess. 28, 1555–1575.
Yagow, G., 2001. Fecal coliform TMDL: Mountain Run Watershed, Culpeper County, Virginia.
Zeckoski, R., Benham, B.L., Shah, S., Wolfe, M.L., Brannan, K., Al-Smadi, M., et al., 2005. BSLC: A tool for bacteria source characterization for watershed management. Appl. Eng. Agric. 21, 879–892.