Environmental Modelling & Software 101 (2018) 218–235

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A framework for uncertainty and risk analysis in Total Maximum Daily Load applications

Rene A. Camacho [a, *], James L. Martin [b], Tim Wool [c], Vijay P. Singh [d]

[a] Water Resources Engineer, Tetra Tech, Inc., 1899 Powers Ferry Road SE, Suite 400, Atlanta, GA 30339, United States
[b] Professor, Department of Civil and Environmental Engineering, Mississippi State University, MS, United States
[c] Environmental Scientist, US. Environmental Protection Agency, Atlanta, GA, United States
[d] Distinguished Professor, Regents Professor and Caroline & William N. Lehrer Distinguished Chair in Water Engineering, Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX, United States

ARTICLE INFO

Article history:
Received 11 July 2017
Received in revised form
2 November 2017
Accepted 12 December 2017

Keywords:
Total Maximum Daily Load
Margin of safety
Uncertainty analysis
Risk assessment
Bayesian analysis

ABSTRACT

In the United States, the computation of Total Maximum Daily Loads (TMDL) must include a Margin of Safety (MOS) to account for different sources of uncertainty. In practice however, TMDL studies rarely include an explicit uncertainty analysis and the estimation of the MOS is often subjective and even arbitrary. Such approaches are difficult to replicate and preclude the comparison of results between studies. To overcome these limitations, a Bayesian framework to compute TMDLs and MOSs including an explicit evaluation of uncertainty and risk is proposed in this investigation. The proposed framework uses the concept of Predictive Uncertainty to calculate a TMDL from an equation of allowable risk of non-compliance of a target water quality standard. The framework is illustrated in a synthetic example and in a real TMDL study for nutrients in Sawgrass Lake, Florida.

1. Introduction

Section 303(d) of the U.S. Clean Water Act identifies a Total Maximum Daily Load (TMDL) as the maximum pollutant load that a water body can assimilate without violating a specific water quality standard. A TMDL is computed as the sum of the allowable loads from point and non-point sources ($\sum WLA$ and $\sum LA$, respectively) plus a margin of safety (MOS) as follows (U.S. Environmental Protection Agency, 1999; Shirmohammadi et al., 2006):

$$TMDL = \sum WLA + \sum LA + MOS \qquad (1)$$

The MOS is a fraction of the TMDL which fundamentally accounts for the uncertainty in the modeling and calculation of the assimilative capacity of the water body. The main sources of this uncertainty are model structure uncertainty, input data uncertainty and model parameter uncertainty. The model structure uncertainty results from errors in model formulation and numerical solution of the equations describing a particular physical, biological or chemical process. Input data uncertainty results from errors in field and laboratory measurements used to force and calibrate the models. Finally, parameter uncertainty results from the use of inaccurate model parameter values. Given these multiple sources of uncertainty, the MOS represents a critical component of the TMDL. However, objective and standardized approaches for the computation of the MOS are limited.

Traditionally, the MOS has been either implicitly incorporated in a TMDL by using conservative assumptions for the estimation of the assimilative capacity of the water body or explicitly incorporated in the TMDL as an independent load allocation as in Eq. (1) (U.S. Environmental Protection Agency, 1991). The lack of an objective approach for the computation of the MOS has resulted, however, in a wide range of subjective and often arbitrary criteria for its computation which in most cases cannot be replicated nor used for comparative analyses between TMDL studies. In addition, the use of subjective approaches generally result in unclear relationships between the TMDL and the MOS and more importantly between the MOS and the water quality standards. The limitations of using subjective approaches for the computation of the MOS have been documented by several researchers including Dilks and Freedman (2004) in a review of 172 TMDLs performed in eight states, Langseth and Brown (2010) in a review of 50 TMDLs from New

* Corresponding author.
E-mail address: rene.camachorincon@tetratech.com (R.A. Camacho).

https://doi.org/10.1016/j.envsoft.2017.12.007
1364-8152/

England, and Crumpacker and Butkus (2009) in a review of 23 TMDLs from the states of Washington, Oregon, and California. Langseth and Brown (2010) point out that none of the TMDLs reviewed in their study explicitly consider the risk of violating the water quality standards as the basis to define the MOS.

To overcome the limitations of the subjective approaches the National Research Council recommends the use of objective uncertainty analyses as the basis for the MOS and TMDL calculation (NRC, 2001). This recommendation is also supported by several researchers and practitioners who also advocate the use of uncertainty analysis as a more transparent, reproducible and robust strategy to define the MOS and TMDL (Ames and Lall, 2008; Dilks and Freedman, 2004; Langseth and Brown, 2010; Liang et al., 2016; Reckhow, 2003; Shirmohammadi et al., 2006). Dilks and Freedman (2004) argued that an objective uncertainty-based-method to compute the MOS and TMDL should have four important attributes. First, the method should explicitly account for the impacts of uncertainty on the estimation of the MOS and TMDL. Second, the method should be reproducible. Third, the method should explicitly define the degree of protection expected from the TMDL as the probability that the water quality standard will be satisfied once the TMDL is implemented. And fourth, the method should identify data limitations and also implementation problems that could result from TMDLs computed under limited data availability or with the use of poor quality datasets. This latter aspect, however, is more related with policy making and requires stakeholder involvement during the definition of the TMDLs.

To incorporate explicit uncertainty analyses in the TMDL process, research has been conducted to compute the MOS and TMDL based on methods such as First Order Variance Analysis (Park and Roesner, 2012; Zhang and Yu, 2004), Point Estimation Methods (Franceschini and Tsai, 2008), Bayesian Networks (Alameddine et al., 2011; Ames and Lall, 2008; Patil and Deng, 2011) and Risk Assessments (Borsuk et al., 2002; Gronewold and Borsuk, 2009; Hantush and Chaudhary, 2014; Langseth and Brown, 2010). Methods based on risk assessments and Bayesian inference have been subject of increasing attention during the last decade because they can be used to explicitly calculate the probability of non-compliance or failure of the TMDL due to multiple sources of uncertainty. Borsuk et al. (2002) presented a probabilistic and Bayesian approach to calculate the risk of non compliance and MOS of TMDLs assuming the errors between the model predictions and observations are independent, normally distributed and unbiased. More recently Ames and Lall (2008) developed a Bayesian network to obtain uncertainty and risk estimates for TMDLs; Gronewold and Borsuk (2009) developed a software tool to estimate the probability of compliance of TMDLs from deterministic model results; and Langseth and Brown (2010) developed a strategy to compute the MOS using risk based concepts traditionally used in engineering design, although their strategy does not include an explicit method for the propagation of uncertainty to model predictions. Hantush and Chaudhary (2014) extended the method proposed by Borsuk et al. (2002) for more general cases where the errors between the model predictions and the observations are correlated and biased and also computed the MOS from an equation of risk of non-compliance.

Risk-based approaches apply the concept of performance failure to compute a TMDL. In engineering, a system experiences a performance failure when it is unable to perform as expected (Singh et al., 2007). In the TMDL context, the probability of failure of the TMDL after implementation is known as the risk of non-compliance. This probability can be computed using a mathematical model, if the target concentration (c^*) and also the allowed frequency of non-compliance (β) of a water quality standard are defined. Traditionally, the computation of a TMDL under a risk-based framework must satisfy:

$$P\left(Y > c^* \middle| \boldsymbol{\theta}, \mathbf{X}\right) \leq \beta \tag{2}$$

where $P(Y > c^* | \boldsymbol{\theta}, \mathbf{X})$ is the probability that a simulated water quality variable Y will exceed c^* given a vector of model parameters $\boldsymbol{\theta}$, and a matrix of input data $\mathbf{X}$ such as flows and contaminant loads from point and non-point sources (e.g. Borsuk et al., 2002; Hantush and Chaudhary, 2014). Risk based approaches based on Eq. (2) have, however, an important limitation. Eq. (2) assumes that the left side of the equation which represents the probability that the model predictions of a variable of interest (Y) will exceed the target concentration c^*, is equal to the probability that the actual concentrations (Z) will exceed the target concentration c^*, or $P(Z > c^*)$, which is the probability of interest for management purposes. The above is an inaccurate assumption because the model exceedance probability $P(Y > c^* | \boldsymbol{\theta}, \mathbf{X})$ and the actual exceedance probability $P(Z > c^*)$ can only be equal if the model is a perfect representation of the real world and is able to reproduce observed concentrations with a 100% accuracy i.e. an ideal case. In reality $P(Y > c^* | \boldsymbol{\theta}, \mathbf{X}) <> P(Z > c^*)$ and thus, a reformulation of Eq. (2) is necessary to have an accurate assessment of the risk of failure of the TMDL i.e. $P(Z > c^*)$ and also a more accurate basis for the MOS computation.

This investigation has three main objectives. The first objective is to reformulate Eq. (2) to obtain a more accurate assessment of the probability that the real world observations will exceed the target concentration $P(Z > c^*)$, i.e. the TMDL risk of failure. The second objective is to present a Bayesian strategy to solve the resulting equation for $P(Z > c^*)$; and the final objective is to propose a strategy to compute the MOS and TMDL that satisfy an allowable risk of non-compliance (β).

The proposed approach to calculate $P(Z > c^*)$ incorporates a Bayesian parameter inference strategy to explicitly account for the impacts of model and parametric uncertainty. The Bayesian parameter inference is based on the likelihood function recently proposed by Hantush and Chaudhary (2014) which is relatively general for most practical cases. The method is demonstrated in a theoretical biochemical oxygen demand TMDL using the estuarine Streeter-Phelps model, and in a real application of the Water Quality Analysis Simulation Program (WASP) (Ambrose et al., 1993) to determine a nutrient TMDL in Sawgrass Lake, Florida, USA. The paper is organized as follows. Section 2 formulates the Bayesian framework to compute $P(Z > c^*)$, MOS and TMDL. Section 3 and Section 4 present the case studies and results, and Section 5 presents the discussion and conclusions of the investigation.

2. Risk of non-compliance of a Total Maximum Daily Load (TMDL)

A critical piece of information for decision makers and stakeholders is the probability or risk of failure of the TMDL. This is the probability that the water quality of a receiving water body will exceed a particular standard following the TMDL implementation or $P(Z > c^*)$. The existing risk based approaches assume that this probability is equal to the probability that the model predictions Y will exceed the target standard c^* or $P(Y > c^* | \boldsymbol{\theta}, \mathbf{X})$. In practice, because of the existence of multiple sources of uncertainty, models are unable to reproduce observations with perfect accuracy and as a results $Y \neq Z$ and $P(Y > c^* | \boldsymbol{\theta}, \mathbf{X}) \neq P(Z > c^*)$. To formulate an alternative expression to compute $P(Z > c^*)$ it is necessary to bear in mind that decisions and inferences about Z (the future water quality concentrations under TMDL conditions) are 'conditional' on the information provided at present by the model predictions (Y), where $Y = g(\widehat{\boldsymbol{\theta}}, \mathbf{X})$, g represents a deterministic model and $\widehat{\boldsymbol{\theta}}$ is a vector of calibrated parameters. This conditionality can be explicitly taken into account to reformulate Eq. (2) as follows: