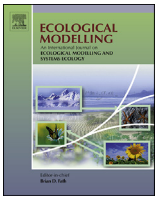
Ecological Modelling 351 (2017) 129–139

Contents lists available at ScienceDirect

Ecological Modelling

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



A sensitivity and uncertainty analysis of a continental-scale water quality model of pathogen pollution in African rivers

Klara Reder*, Joseph Alcamo, Martina Flörke

Center for Environmental Systems Research (CESR), University of Kassel, Wilhelmshöher Allee 47, 34171 Kassel, Germany

A R T I C L E   I N F O

Article history:
Received 18 July 2016
Received in revised form 10 February 2017
Accepted 13 February 2017
Available online 11 March 2017

Keywords:
Pathogen water pollution
River water quality
Africa
Fecal coliform bacteria
Water quality modeling
Sensitivity analysis
Uncertainty analysis
Latin Hypercube Sampling

A B S T R A C T

Continental-scale water quality modeling is a new scientific approach concerned with computing the level of water pollution for several river basins at once. Uncertainties in these models, and in models of smaller scale, arise especially from the specification of model parameters. To identify and analyze these uncertainties we perform a global sensitivity and uncertainty analysis using Latin Hypercube Sampling on the WorldQual water quality model. The focus of the analysis is the river pathogen model of WorldQual as applied to rivers in Africa. This is the first uncertainty and sensitivity analysis performed on a continental-scale pathogen river pollution model. The median output uncertainty of the model (coefficient of variation, based on log-transformed data), assuming plausible estimates of 42 parameter uncertainties, was 10.7%; 90% of grid cells had output uncertainties below 23%. The parameters making the largest contribution to this uncertainty (in order of importance) are the pathogen waste loading per capita, the in-stream settling velocity of pathogens, the percentage of population in a river basin connected to a sewer system, and the raw effluent concentration from the manufacturing sector. Over the continental study area, model output uncertainty and the most sensitive parameters were found to have a highly irregular spatial pattern. This finding suggests that model performance is a strong function of local and regional conditions and that reducing the uncertainty of a single parameter may not lead to large improvements in model performance over the entire continent. A more efficient approach would be to improve model performance region-by-region and improve the estimation of specific parameters known to have a large influence on model uncertainty in those regions. The analysis showed that only four parameters dominate output uncertainty over 93% of the study area, implying that model performance can be substantially improved by reducing the uncertainty of a small number of parameters.

1. Introduction

An urgent problem in developing countries is the increasing intensity of pathogen river pollution (UNEP, 2016) stemming from increasing population and economic activity, coupled with a lack of wastewater treatment facilities. Since surface waters are still used by a large fraction of the population in developing countries for bathing, washing clothes, irrigation and other purposes, it is also likely that large numbers of people are coming into physical contact with pathogen-contaminated water. Pathogen river pollution is a particularly important issue in Africa where hundreds of millions of people are estimated to be under health risk because of their contact with pathogen-contaminated surface waters (UNEP, 2016).

Water quality models can make an important contribution to dealing with the challenge of pathogen river pollution: They help establish the cause-effect relationships between pollution loads and the level of pathogen river pollution; they can be used to investigate future trends in this pollution, and they can be used to evaluate management actions (including the reduction in pollution loadings) that would reduce pathogen pollution to acceptable levels (Oliver et al., 2009; Voß et al., 2012). In data-poor areas water quality models can also fill in the spatial and temporal gaps in field measurements.

The existing set of models that simulate pathogen river pollution include only two continental-scale models, capable of analyzing and comparing several river basins at the same time (Hofstra et al., 2013; Reder et al., 2015). As with models of other scales, it is important to investigate the performance of these models in order to understand the operational limits of these tools, and how best to

---

* Corresponding author at: Center for Environmental Systems Research, Wilhelmshöher Allee 47, 34171 Kassel, Germany.
E-mail address: reder@usf.uni-kassel.de (K. Reder).

http://dx.doi.org/10.1016/j.ecolmodel.2017.02.008
0304-3800/

reduce their uncertainties. The customary way to assess performance of these and other environmental models is to validate their calculations against observations. While validation is a necessary minimum requirement for testing model performance, sensitivity and uncertainty analysis are also often used to extract additional information about model performance useful for further developing models. For example, as shown in this paper, sensitivity and uncertainty analysis provide quantitative estimates of model output uncertainty, spatial patterns of this uncertainty, and information about the parameters making the largest contribution to this uncertainty. This information is very useful for establishing the operating bounds of the model and for making changes to improve its performance.

The objective of this paper is to assess the performance of a continental-scale water quality model by conducting a detailed global sensitivity and uncertainty analysis of the model. This is the first analysis of its kind for a continental-scale pathogen water quality model. We aim to investigate the size of model output uncertainty as it is affected by parameter uncertainty, as well as how uncertainty is distributed spatially, and the relative contribution of particular parameters to output uncertainty. Although many different factors can have an impact on model uncertainty, in this paper we focus on the impact of parameter uncertainty because of its important role in environmental model uncertainties (Manache and Melching, 2004).

The object of our study is the river pathogen model of WorldQual as applied to Africa. The WorldQual model follows conventional practice and uses the concentration of fecal coliform bacteria as the indicator of pathogen pollution (DWA, 1996; European Commission, 1975; MEPPRC, 2002; USEPA, 1986; USEPA, 2012; WHO, 2000). River concentrations of fecal coliform bacteria are computed as a function of fecal coliform loadings, in-stream bacterial decay, and various river characteristics. Model equations, model inputs, and model assumptions and their application to European rivers are described in detail in Reder et al. (2015), Voß et al. (2012) and Williams et al. (2012). In this paper we apply the model to African rivers, which required a modification of some model inputs as described in Supplementary material A. The validation of the model against observations is also described in Supplementary material A.

2. Methods of uncertainty analysis and sensitivity analysis

There are many approaches to sensitivity and uncertainty analysis, among which are: screening methods (Morris, 1991), response surface methodologies (Box and Wilson, 1951; Deman et al., 2014), Monte Carlo methods (Freda et al., 1981; Saltelli et al., 2000) and variance-based methods (Cukier et al., 1978; Saltelli et al., 1999). Each approach has its advantages and drawbacks. Screening methods are conceptually simple, easy to implement and have low computational costs, but produce qualitative rather than quantitative results e.g. in form of parameter rankings. In addition, many screening methods do not account for interactions between parameters (Saltelli et al., 2000). Response surface methods can factor in strong interdependencies and non-linearities (Deman et al., 2014), but cannot effectively visualize interrelationships between numerous variables. Monte Carlo methods are powerful, robust and flexible as the whole range of input parameters is examined (Manache and Melching, 2008) but often have large computational requirements. Variance-based methods can be applied to any model including non-linear or non-monotonic models, but the computational cost is very high (Marino et al., 2008; Saltelli et al., 2000). Thus, all of these sensitivity approaches are equally valid but may work better in different problem-settings.

The two main types of sensitivity analysis are “local” and “global”. Local sensitivity analysis identifies the changes in model output resulting from a small perturbation of one of the model parameters (while holding the other parameters constant) (Oakley and O’Hagen, 2004; Saltelli et al., 2000). The advantages of this approach are its simplicity and generally lower computational costs, while its disadvantage is that it does not represent the full impact of the uncertainty of a parameter on model output (Oakley and O’Hagen, 2004). Global sensitivity analysis gives a better estimate of uncertainty (Saltelli et al., 2006) by using probability density functions to express the uncertainty of model parameters, rather than by perturbing these parameters. In global sensitivity analysis all parameters are varied at the same time within the range of their probability density functions (Saltelli et al., 2000), often taking into account their co-variance (Ogejo et al., 2010).

Local sensitivity analysis has been used in many studies to assess pathogen and other water quality models, including Cho et al. (2012), Coffey et al. (2010), Ferguson et al. (2007), Niazi et al. (2015) and Tian et al. (2002). Although global sensitivity analyses are less common (e.g. Cea et al., 2011) it is used here because of its above-mentioned advantages.

In this paper we perform a global sensitivity analysis on the river pathogen model of the WorldQual model using Latin Hypercube Sampling (LHS) (McKay and Beckman, 1979). Latin Hypercube Sampling is a type of stratified Monte Carlo Sampling (Iman and Helton, 1985) in which several random input sets are selected and for each input set individual model simulations are performed (Saltelli et al., 2000). The LHS was selected because (i) it is conceptually simple and easy to implement, (ii) its computational costs are low due to its dense stratification sampling strategy, and, (iii) uncertainty can be directly estimated from the simulation results (Helton and Davis, 2002).

We use global sensitivity analysis for two main purposes. Firstly, for “sensitivity analysis”, i.e. to investigate the influence of individual model parameters on the uncertainty of model output. Secondly, for “uncertainty analysis”, i.e. to estimate the total uncertainty of model output resulting from the uncertainties of all important model parameters.

2.1. Sampling approach

Latin Hypercube Sampling is an “importance” sampling approach meaning that the full sample space is divided into non-overlapping sub-regions from which random samples are taken (Helton and Davies, 2000; Saltelli et al., 2000).

To compute the sample LHS, the following steps are followed:

The input is a vector *x*

$$\mathrm{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_{\mathrm{nx}}].nx = \mathrm{numberofinputparameters} \tag{1}$$

The output is vector *y*

$$\mathrm{y} = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_{\mathrm{nx}}] \tag{2}$$

The cumulative probability distribution function is partitioned for each parameter in *I* intervals with equal probability. From each interval a value is selected randomly. To generate a vector from these samples one value from each sample set is selected randomly without replacement and paired up. Following this procedure vectors in form of an *nx*-tuple are generated:

$$\mathrm{x}_{\mathrm{k}} = [\mathrm{x}_{\mathrm{k}1}, \mathrm{x}_{\mathrm{k}2}, \ldots, \mathrm{x}_{\mathrm{knx}}]k = 1, \ldots, I \tag{3}$$

The result of this procedure is a matrix of possible variations and combinations of parameters. These parameters serve as input for model runs of the size *N*. In every realization of the Latin Hypercube Sampling each parameter is varied within the corresponding distribution function. Each sample *N* identifies the variation in the output