

## Assessment of pathogen pollution in watersheds using object-oriented modeling and probabilistic analysis

Amin Elshorbagy, Ramesh S. V. Teegavarapu and Lindell Ormsbee

### ABSTRACT

A limited number of research trials have been reported in the past to model pathogenic organisms in streams and large water bodies at a watershed scale. In this paper, modeling of fecal coliform in streams is proposed from a management perspective at the watershed level. To model the fate and transport of fecal coliform in a watershed in Southeastern Kentucky, an object-oriented (OO) simulation model, based on the concepts of system dynamics (SD) approach, is proposed in this study. The approach combines both data-driven approaches and insights gained from a process-based approach. Different management scenarios, based on flow conditions and pollution sources, are generated and evaluated to validate the proposed approach. Deterministic and conceptually simple probabilistic analyses are carried out to understand several water quality management alternatives that aim to reduce pollutant loadings. Results point to the potential use of the proposed OO–SD framework in addressing environmental policy issues and also to the need for relying on probabilistic analysis to obtain more credible results and recommendations in data-poor conditions. The proposed approach helps direct limited funding and watershed management efforts to be focused on areas that have the greatest impact on the surface water quality conditions.

**Key words** | object-oriented modeling, pathogen impairment, probabilistic analysis, STELLA modeling environment, stream impairment index, system dynamics, water quality management

**Amin Elshorbagy** (corresponding author)  
Centre for Advanced Numerical  
Simulation (CANSIM),  
Department of Civil and Geological Engineering,  
University of Saskatchewan,  
Saskatoon, SK, Canada S7N 5A9,  
Tel: +1 306 966 5414,  
Fax: +1 306 966 5427,  
E-mail: [amin.elshorbagy@usask.ca](mailto:amin.elshorbagy@usask.ca)

**Ramesh S. V. Teegavarapu**  
**Lindell Ormsbee**  
Department of Civil Engineering,  
University of Kentucky,  
Lexington, KY 40506,  
USA

### INTRODUCTION

Continued pollution of water bodies and associated violation of the Clean Water Act (CWA) and the terms and conditions set by the National Pollution Discharge Elimination System (NPDES) are common problems in all states in the US. It is estimated that about 21 000 river segments, lakes and estuaries have been identified by states as being in violation of one or more water quality standards (NRC 2001). In the Commonwealth of Kentucky, especially the southeastern area, the region of interest in the current study, pathogenic contamination of river segments has been identified as a major problem (KWRRRI 2000). A significant number of streams in this region are impaired for primary and secondary

contact due to pathogenic contamination. The impairment is mainly caused by pipes discharging raw sewage directly into receiving streams (i.e. straight pipes), failing septic systems and effluent bypass from wastewater treatment plants.

Accurate assessment of spatial and temporal variation of pollutant loadings in streams within a watershed is essential from a water quality management perspective. Process-based models are often used to derive watershed management alternatives. However, under data-poor conditions, it is difficult to calibrate and validate process-based models. Conceptually simple simulation models that are not over-parameterized are, in many situations, the superior

method of modeling and understanding the fate and transportation of pollutants in streams (Jian & Yu, 1998). Also, in the absence of enough data, it would be difficult to characterize the pollutant loads and identify the effect of management practices in a watershed on such loads.

The extremely limited literature on modeling the fate and transport of pathogens in water bodies emphasizes the need for developing modeling-based tools for watershed management. Of those approaches that have been developed, many are fraught with significant difficulties. For example, Connolly *et al.* (1999) coupled a hydrodynamic model with a pathogen fate model to define how hydrodynamic conditions affect the transport of fecal contamination in a bay area. They have highlighted uncertainties due to inadequate understanding of the physical system. Other studies (e.g. Stow *et al.* 2000) point to difficulties associated with calibration of these models and estimation of parameters from limited data. In many situations, the parameters of these models cannot be uniquely obtained from the available field data and thus must be estimated from technical guidance documents (Bowie *et al.* 1985). In such situations, little confidence can be attached to the results of the models (NRC 2001) and therefore they may not be appropriate to evaluate future environmental management scenarios. In either approach, process-based or empirical, water quality data-poor conditions cast a major doubt on how representative the available data of the year-round and long-term conditions of the system are. The small window of available data makes it necessary to develop a methodology that can generate longer records that better represent the overall conditions of the system. That is achievable through the probabilistic analysis approach, which is used in this study.

In this paper, a simple object-oriented modeling approach to surface water quality is proposed to facilitate the following tasks: (i) combining empirical (data-driven) modeling techniques with easy-to-use process-based concepts of modeling transport and fate of pathogens through the watershed under data-scarce conditions, (ii) constructing an understandable model that can be easily executed, managed and modified as necessary and (iii) creating a simulation model that can easily handle and test future scenarios and policy analysis. Deterministic and probabilistic analyses to characterize pollutant loads and flow conditions are used to identify and quantify the relative

contribution of individual sub-basins to the total downstream pollutant load. Both types of analysis are used and compared to evaluate the outcomes of different possible future water quality management scenarios in the region, with the purpose of directing management efforts and limited resources towards more pressing issues.

---

## SURFACE WATER QUALITY MODELING AND MANAGEMENT APPROACHES

### Process-based modeling approach

Effective water quality management relies mainly on effective modeling that can predict the water body's health under different hydrologic and pollutant loading scenarios. Although mechanistic models, sometimes called process-based models, are usually data-intensive and frequently over-parameterized, they are used for a wide variety of applications in studying surface water quality. There is a large number of available mechanistic models. One of the frequently used models is the HSPF (Donigan *et al.* 1995), which covers wide areas of applications including point and nonpoint source pollution analyses. QUAL2E (Brown & Barnwell 1987) is also used for modeling the fate and transport of pollutants in streams. Both data-driven or empirical and mechanistic approaches have been used in the past for water quality modeling studies (Chapra 1994). In many situations when the availability of water quality data is limited, the empirical approach becomes essential to characterize the pollutant loadings. Many examples of the application of such models (e.g. Reckhow & Chapra 1983; Jian & Yu 1998) are available in the literature.

Although mechanistic modeling environments attempt to overcome the limitations associated with other modeling approaches, their data requirements can be overwhelming. A good modeling approach, which is recommended by Chapra (2003) is "an adaptive approach starting with simpler models at the initial phases and then progress to more complex frameworks as additional data are collected." The model selection criteria concerning cost, flexibility, adaptability and ease of understanding all tend to favor simple models and support research in the development of models that can be fully parameterized from the available data (NRC 2001).

### Data-driven or empirical approaches

Recently, researchers have advocated the development and use of more simple models. The most useful predictive models are often extremely simple (Hodges 1987). A similar conclusion has been made by Levin (1985), regarding ecological models: that overly detailed models are useless as predictive devices and techniques for aggregation and simplification are essential. However, data-driven approaches to modeling the transport and fate of pathogens in streams (Elshorbagy *et al.* 2005b) rely mainly on linking the pollutant load to a hydrologic parameter (e.g. stream-flow), which has its own shortcomings. Such approaches lack the link between the contamination level and different sources of pollution. Also, they do not take into account important factors such as the decay rate or the effect of sunlight and water salinity on the fate of the pollution. More importantly, from a management perspective, data-driven models are incapable of modeling and tracking the effects of human intervention and management practices in the watershed on the pollution level in streams.

Simonovic (1992) suggests that systems analysis has its own place in the field of water resources management and simulation as an essential tool for developing a quantitative basis for water management decisions. There is, however, a strong need to explore simulation tools that can represent the complex systems in a realistic way and where water resources managers and operators can be involved in model development to increase their confidence in the modeling process. In this paper, an object-oriented simulation environment, which adopts the system dynamics modeling approach (OO-SD), is employed. A case study is provided to assess the capabilities of the technique in surface water quality management using both deterministic and probabilistic analyses.

### OBJECT-ORIENTED MODELING BASED ON SYSTEM DYNAMICS APPROACH (OO-SD)

Object-oriented (OO) modeling is a way of thinking about problems using models organized around real-world concepts (Rumbaugh *et al.* 1991). It is a way to organize software as a collection of discrete objects that incorporates both data structure and system behavior (Simonovic *et al.*

1997). Data are organized into discrete, recognizable entities called objects. These objects could be concrete, such as a river reach, or conceptual, such as a policy decision. Numerous tools can be used for the implementation of an object-oriented modeling approach and the STELLA software (HPS Inc. 2001) is used for the work presented in this paper. STELLA employs the object-oriented simulation environment as an appropriate tool for the implementation of systems thinking (system dynamics – SD).

### Building blocks of OO-SD approach

*Stocks* and *flows* are the building blocks (objects) of an OO-SD model. A simple illustration of *stocks* and *flows* is the accumulation of interest in a bank account (Ford 1999) or inflow of water into a reservoir. Figure 1 shows a simple model to keep track of the inflow in a reservoir. The double line represents the flow of water from a source, represented by a cloud, into the reservoir (stock). The cloud can be viewed as a stock that is outside the system boundary. The single lines in Figure 1 connect the rainfall and runoff coefficients to the inflow. These are called *connectors*, which show the flow of information inside the model. *Converters* (such as rainfall and runoff coefficients) can represent any process as a function of time and they can also represent any input value or parameter.

The OO-SD simulation approach relies on understanding complex interrelationships existing between different elements within a system. This is achieved by developing a model that can simulate and quantify the behavior of the system. Simulation of the model over time is considered essential to understand the dynamics of the system. Some of the major steps carried out in the development of an SD model are developing an understanding of the system and its boundaries, identifying the key variables, establishing a

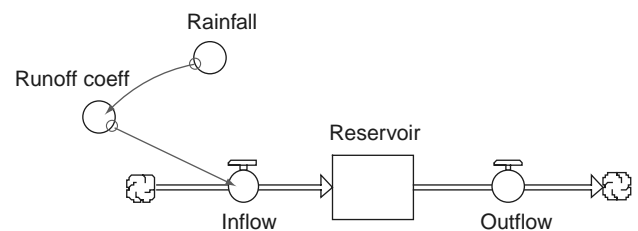


Figure 1 | Basic building blocks of the object-oriented simulation environment (STELLA).

representation of the physical processes or variables through mathematical relationships, mapping the structure of the model and simulating the model to understand its behavior. It is interesting to note that the central building blocks (objects) of the principles of the SD approach, referred to here as the OO–SD approach, are well suited for modeling any physical system (Ford 1999).

The governing equations in an OO–SD model are represented by finite difference expressions used for modeling different elements in a system and are solved using standard numerical schemes. For example, in the case of a *stock*, a continuity equation for mass balance is developed considering the inflows and the outflows, whereas a *converter* carries a functional relationship between different variables that can be represented in a mathematical or a graphical form.

## DEVELOPMENT OF PATHOGEN TRANSPORT MODEL

An area of seven watersheds is used as a case study region for this paper (Figure 2). All the watersheds lie within the North Fork of the Kentucky River basin which has an eight

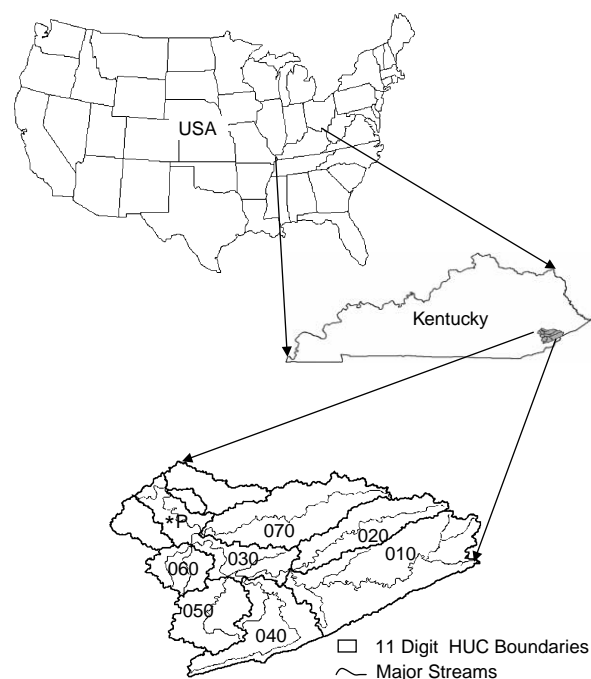


Figure 2 | Case study region in the North Fork river basin of Southeastern Kentucky.

digit identification number of 05100201. The identification number corresponds to a hydrologic unit code (HUC) designated by the US Geological Survey as part of their national watershed classification scheme. The water flows from six headwater sub-basins (HUC-11) 05100201-010, -020, -040, -050, -060 and -070 toward sub-basin -030. Continuous records of streamflow data are available only in sub-basins -010 and -030. Streamflows at the remaining sub-basins are estimated in proportion to their respective contributing drainage areas. Similarly, monthly samples are collected and analyzed for fecal coliform concentrations only in sub-basins -010 and -030. The historical data of fecal coliform concentrations (CFU/100 ml) are obtained from grab samples and are then used to estimate the instantaneous fecal loads (count/d) on the days when the measurements were conducted. Grab samples are collected mid-stream and at mid-depth of the stream channel and are delivered to the lab within six hours of collection. The standard analytical technique used to measure fecal coliform is method 9222 D of the *Standard Methods for Examination of Water and Wastewater* (Clesceri et al. 1989). The fecal coliform concentrations in this area were found to range from 10 to 78 000 CFU/100 ml.

The OO–SD modeling approach is used in this study to combine both process-based and data-driven techniques to handle the issue of pathogen transport and fate in the study area. No hydrodynamic conceptual model is used to route the streamflows; instead, the fecal loads at the HUC-11 sub-basin level are linked to the streamflows. A significant correlation found between the fecal coliform load and the streamflows in a recent study (Elshorbagy et al. 2005b) supports the approach that relates fecal coliform loads and streamflows (Equation (1)). The fecal contamination is attributed to straight pipes and failing septic systems (SF) in the area. A straight pipe is a pipe discharging raw sewage or human waste directly into receiving streams. Therefore, the regression equation linking flows to fecal load at HUC-11, -010, can be normalized using the total number of SF. This equation is found to be as follows:

$$L = 3060443.8 Q^{1.438} \quad (1)$$

where  $L$  is the fecal load (count/d) per one SF and  $Q$  is the streamflow ( $\text{ft}^3/\text{s}$ ). The regression coefficient of determi-

nation ( $R^2$ ) is found to be equal to 0.75. The land-use is predominantly deciduous forest type in all the sub-basins. The land-use percentages for deciduous forest type for sub-basins (HUC-11) 05100201-010, -020, -030, -040, -050, -060 and -070 are 98.08%, 97.76%, 96.15%, 99.29%, 98.79%, 99.21% and 96.21%, respectively. Available pathogen data collected through the PRIDE sampling effort suggests a similar form of load–flow relationship provided by Equation (1) in two watersheds. Also, the SF density (SF per area of the watershed) is approximately equal for all the watersheds. Considering these facts, the functional form of the flow–fecal coliform load relationship is assumed in this study to be similar in all the headwater HUC-11 sub-basins. There is no need for this assumption to be made in the model development, if adequate sampled data is available for developing individual load–flow relationships for each of the sub-basins.

The high number and the density of SF in the area may justify why the fate and transport of the pathogen loading behave in a way similar to a nonpoint source pollution. There is a possibility of having a different power (different from 1.438 in Equation (1)) in the regression equation for each 11-digit HUC. In this way, both the power of the regression and the number of SF can be treated as decision variables (i.e. changed for the purpose of generating future scenarios) in the proposed model. However, the power of the regression equation is kept unchanged in this study since Equation (1) is derived based on load per one SF and there is not sufficient evidence or data to support such a change of the power of the regression equation.

The exponential decay approach (Bowie *et al.* 1985; Mills *et al.* 1985; Thomann & Mueller 1987) is generally used to model fecal coliform in streams. Many water quality models relying on analytical formulations and computational methods are available in the literature for modeling the fate of fecal coliforms in streams. For example, QUAL2E allows direct input of fecal coliform bacteria concentration with temperature-dependent exponent decay. Other models that can be used for fecal modeling are CE-QUAL-RV1, CE-QUAL-W2, HSPF and WASP (USEPA 1997). However, fecal modeling is not directly possible using the WASP model. The exponential bio-degradation rate of fecal coliform can be specified in WASP with minor modifications. A comprehensive discussion on receiving models is

available in the report on tools for watershed assessment produced by USEPA (1997).

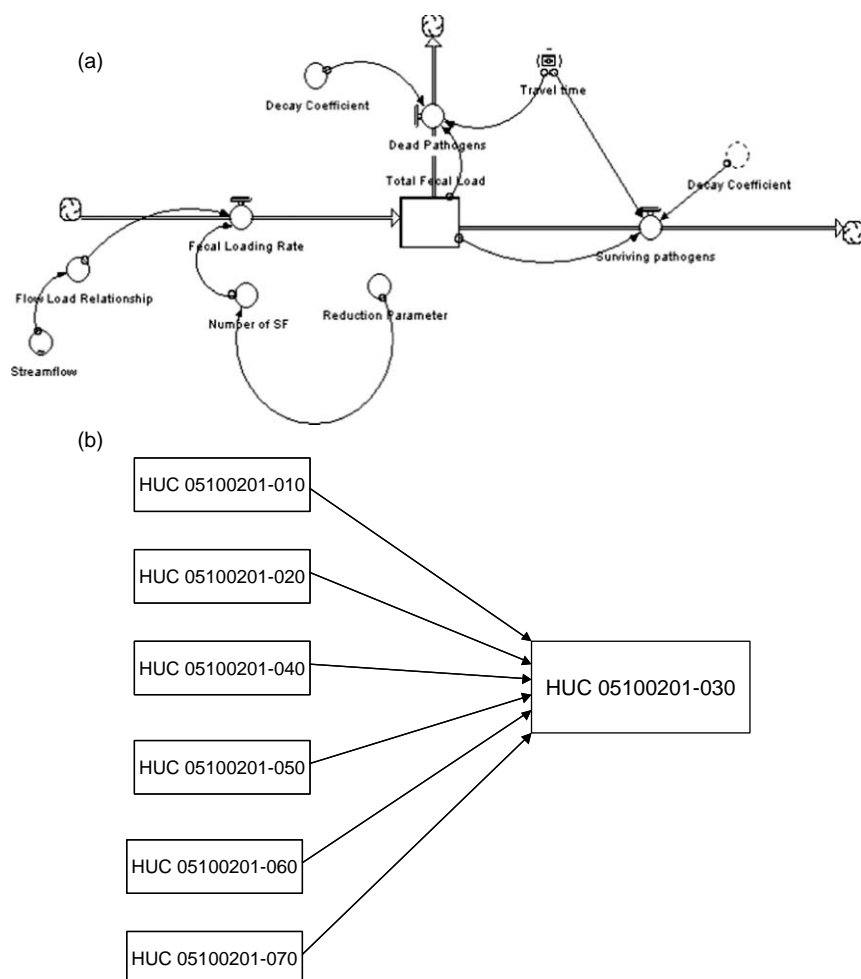
Fecal coliform load is modeled as a substance that declines (decays) as a first-order process (Thomann & Mueller 1987):

$$L_t = L_0 e^{-kt} \quad (2)$$

where  $L_t$  and  $L_0$  are the fecal load at downstream and upstream points (count/d) respectively,  $k$  is the decay rate ( $d^{-1}$ ) and  $t$  is the travel time (d) that is estimated from the length and velocity of flow in the stream. Factors such as sunlight, water temperature and salinity are not explicitly considered in the model. The effect of all these factors are implicitly considered and are lumped into one decay factor,  $k$ . However, inclusion of individual factors or relationships linking these factors to pathogen survival rate in the modeling approach is straightforward if data is available. The OO–SD model built for the study area is shown in Figure 3, where Flow blocks represent the fecal loading rate and Converters represent the flow–fecal load functional relationships. Stocks handle the pollutant load, where only surviving pathogens become outflow from the Stock block. Measured fecal concentrations at HUC-11-030 are used to validate the proposed model. Years 1999 and 2000 are used for model calibration and validation, respectively (Figure 4). The calibration process is carried out using mean squared error (MSE) as a criterion for model performance. The decay rate,  $k$ , of 0.5 ( $d^{-1}$ ) was obtained from the model calibration process.

## DETERMINISTIC ANALYSIS AND RESULTS

The developed OO–SD model is used in this study to test the impact of possible watershed future scenarios on stream pollution. In the case of deterministic analysis, pollutant load–streamflow relationships are developed using the regression relationships and observed values of streamflow are used in the model to develop the pollutant loads at the watershed scale. The developed model (Figure 3) is executed, simulating the daily fate and transport of the fecal coliform loadings through the different sub-basins from the headwaters until the downstream point (P) (see Figure 2). The object structure shown in Figure 3(a)



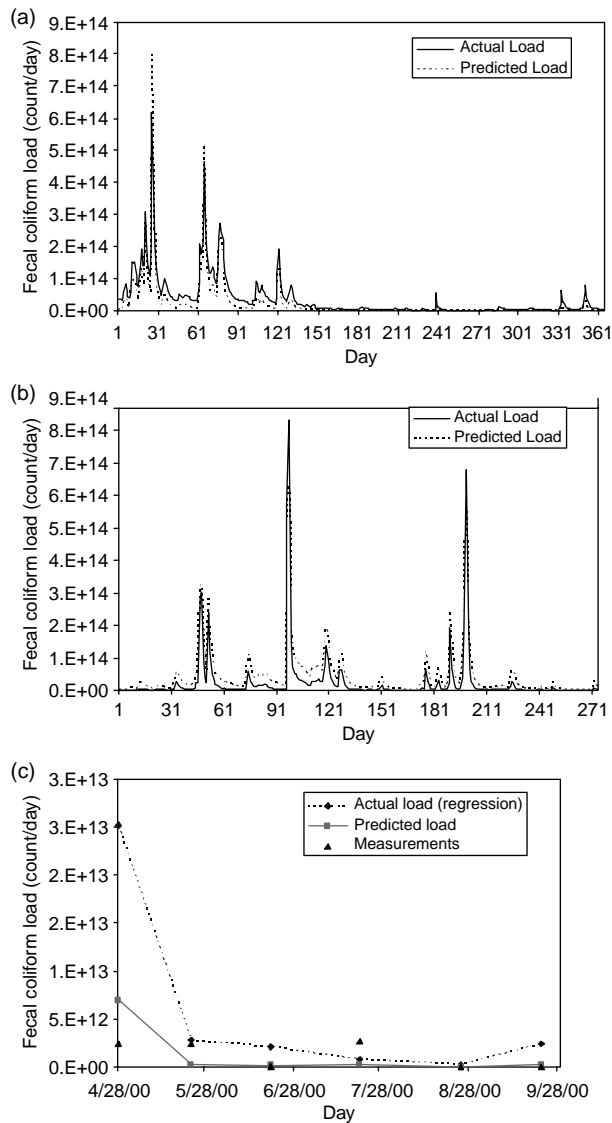
**Figure 3** | Structure of the object-oriented model. (a) Model of one sub-basin (e.g. 05100201-020). (b) Schematic of the entire watershed representing the study area.

represents only one sub-basin. The units 010, 020, 040, 050, 060 and 070 are parallel units that discharge to the downstream unit 030. After propagating the total daily pathogen loads (decay considered), along with the daily cumulative flows, the daily loads at point p and other outlets of the sub-basins are calculated by the model using the regression relationships. Total load at P is equivalent to the sum of the loads coming from all upstream sub-basins. Accordingly, daily fecal coliform concentrations are computed, at each sub-basin outlet as well as at P, by dividing the daily pathogen load by the streamflow at each outlet.

The number of days with concentrations higher than the pre-specified standards is counted in the model and marked as impaired days. The results are summarized in the first row of Table 1 as the baseline scenario B, which is based on

the data (flows and number of SF) from the year 1999. Finally, average annual loads at each point (outlet) are also computed based on the daily loads. Configuration of the model structure for deterministic analysis is provided in Figure 5.

Four scenarios are designed and investigated in order to help assess the effect of both streamflows and the number of straight pipes and failing septic systems (SF) on a downstream point under consideration (point P). Streamflow and number of SF are the variables, which are handled as potential candidates to shape future scenarios in the process of scenario generation. The scenarios are as follows: (1) a year of low flows with 30% decrease in the number of SF, (2) the same year of low flows used in scenario 1 but with 30% increase in the number of SF, (3) a year of high flows with 30%



**Figure 4** | Model performance in (a) calibration, (b) validation and (c) with respect to grab sampling.

decrease in the number of SF and (4) the same year of high flows used in scenario 3 with 30% increase in the number of SF. The baseline scenario is also analyzed and reported along with the ones indicated above. The low and high flow years selected are the driest and wettest years recorded in the last ten years. Scenarios 2 and 4 suggest an increase in the number of straight pipes and failing septic systems, which are a direct consequence of an increase in the population in the case study region. The percentages chosen for increase and decrease are arbitrary and are meant to illustrate the utility

of the simulation model and results. The simulation environment is flexible and would allow any percentage change in the contributing pollution sources.

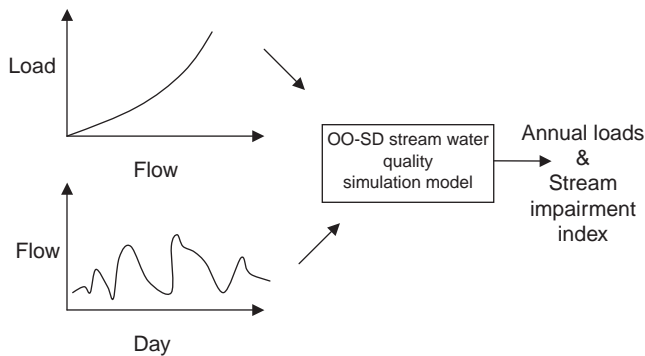
These scenarios are by no means exhaustive but they help address the effect of both natural conditions (wet and dry years) and human intervention (changing the number of SF in the basins) on the fecal coliform concentration in the streams. The scenarios are evaluated using two criteria: the average annual fecal load and the number of days during which the stream will be impaired. The model is set up to output those two criteria as the outputs of the run. Due to the large variation of fecal concentrations throughout the year, the average annual load of the pollutant might not be representative of the stream impairment. The average annual load could be high but concentrated in a lower number of days, which would suggest that the stream is not impaired most of the year, and vice versa. Therefore, assuming a water quality standard of 400 CFU/100 ml, the number of impaired days and the average annual load are estimated in each scenario. The results are summarized in [Table 1](#). A stream impairment index is defined in this context, which is referred to as vulnerability. Vulnerability is used to evaluate how vulnerable to pollution the stream is, i.e. how often the stream violates water quality standards (number of impaired days per year). Since concentrations are related, to a great extent, to flow values, vulnerability will be dependent on the flows.

The distribution of the fecal concentrations at point P is shown in [Figure 6](#). It should be noted that the two criteria used in this paper for scenario evaluation may lead to different conclusions. Although the average annual load in scenarios 3 and 4 is higher than that of scenarios 1 and 2, respectively, the number of impaired days is less. This is probably due to the fact that high flows cause more dilution and, therefore, lower concentrations.

The proposed modeling approach can help in the process of allocating funds to different sub-basins based on their contribution to the total pollution problem. The output of the model proposed in this study, based on the data of the validation year (i.e. 2000), is shown in [Figure 7](#). One can observe that the HUC-11-070 has the maximum pollution share and thus would be a candidate for the highest attention and allocation of resources, based on deterministic analysis. In addition to use in general analysis,

**Table 1** | Average annual fecal coliform load and number of impaired days (deterministic analysis)

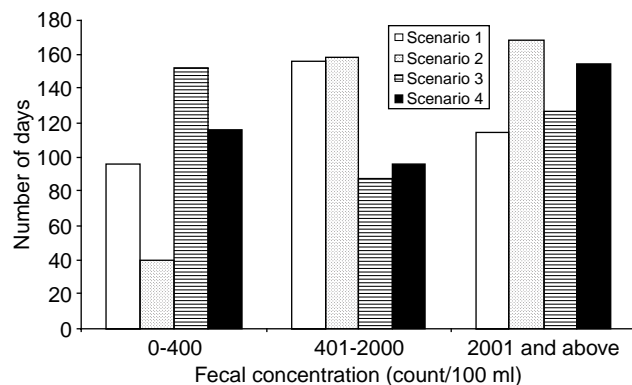
Scenario	Average annual flows (ft <sup>3</sup> /s)	Average annual fecal load (count/d)	Number of impaired days within a year
Baseline	503.4	$3.21 \times 10^{13}$	295
(1)	503.4	$2.25 \times 10^{13}$	269
(2)	503.4	$4.17 \times 10^{13}$	325
Baseline	974.7	$2.24 \times 10^{14}$	230
(3)	974.7	$1.57 \times 10^{14}$	214
(4)	974.7	$2.91 \times 10^{14}$	250

**Figure 5** | Model configuration for deterministic analysis.

the model can also be used to set or evaluate different management strategies.

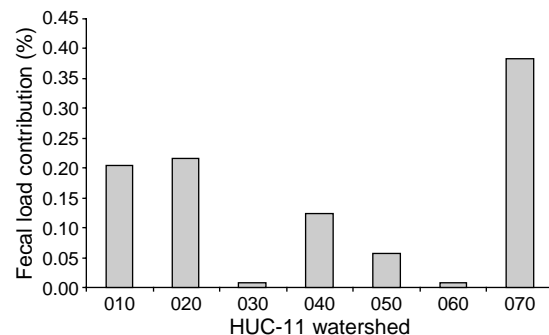
## PROBABILISTIC ANALYSIS AND RESULTS

The dilemma of water quality modeling is aggravated by uncertainties inherent in many steps throughout the

**Figure 6** | Number of impaired days under different management scenarios.

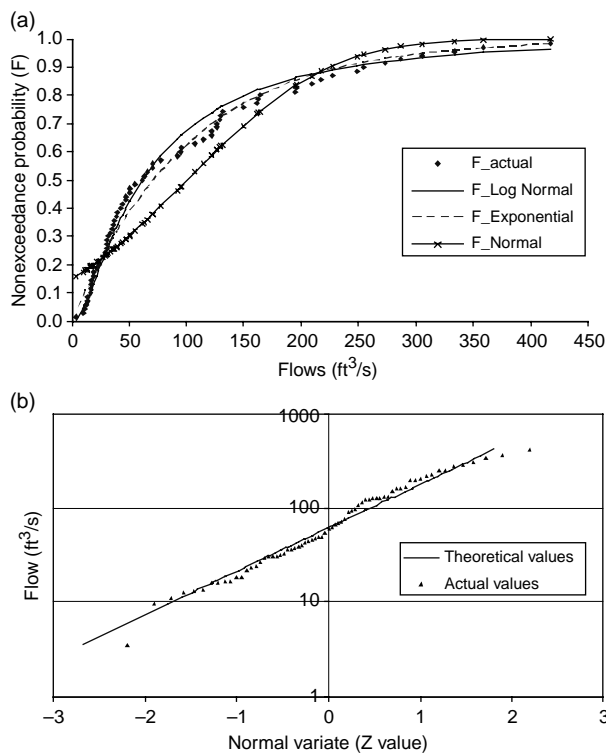
modeling exercise. First, the water quality measurements are usually insufficient for reliable calibration and validation of simulation models (Stow *et al.* 2000) and the case study under consideration is no exception. Second, the impairment, evaluated based on the number of days where concentrations exceed a certain threshold, is dependent on flows. Flows are random variables and those days where sampling occurred may not represent the hydrologic conditions over a long period. Third, the pathogen loads coming from straight pipes cannot be realistically considered constant every day, based only on the number of straight pipes. Therefore generating a large number of flow values as well as pollutant load values and passing them with random combinations into the simulation model can give a better representation of the situation at hand and a better characterization of the stream conditions.

An effective way of achieving this task is through identifying probabilistic distributions of both flows and

**Figure 7** | Pollutant loads of individual HUC-11 sub-basins.



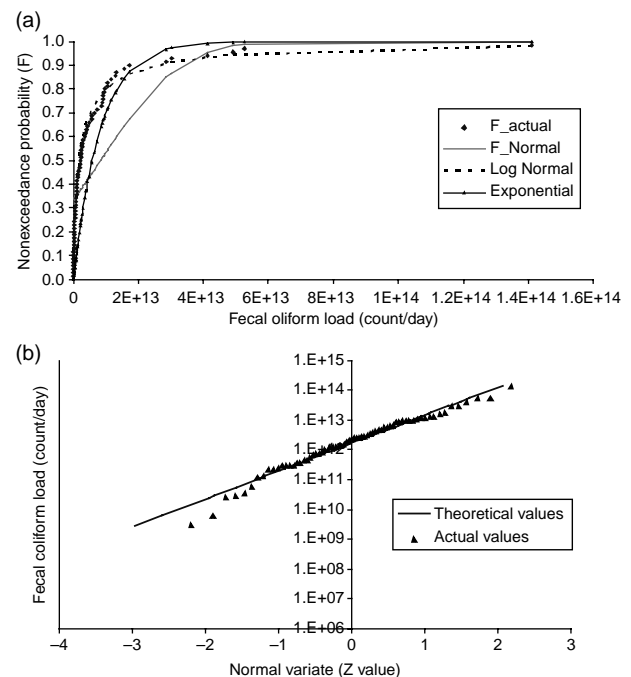
pollutant loads. The probabilistic distribution of measured flow values are compared to multiple theoretical distributions including normal, lognormal, logPearson III, gamma and exponential distributions. The probabilistic approach has been used to address issues of uncertainty with pollutant load allocation (TMDL) (Eheart & Ng 2004; Zhang & Yu 2004). Using commonly used visual inspection (Bedient & Huber 2002; McCuen 2003), lognormal distribution is selected to generate multiple realizations of flow values. Figure 8(a) shows the distribution of actual flows relative to three theoretical distributions. The lognormal Q-Q plot (Figure 8(b)) illustrates the fact that flow variables can be reasonably represented by a lognormal distribution. Similar procedures are followed regarding the distribution of the fecal coliform loads (count/d) coming from each SF. Lognormal distribution is also selected to generate multiple realizations of load values (Figure 9). Many studies have shown that, for many pollutants, wastewater loads follow the lognormal probability distribution



**Figure 8** | Statistical distribution of streamflows. (a) Actual vs. theoretical distributions. (b) Lognormal Q-Q plot of the streamflows.

(Novotny 2004). Probabilistic analysis uses the Monte Carlo method by developing several realizations of loads and flows using the provided statistical distributions and the statistical properties (mean and standard deviation) of the variables.

To characterize the stochastic (probabilistic) behavior of pollutant load over time, the expected values of loads and/or concentration values are evaluated. The configuration of the model structure used for stochastic analysis is given in Figure 10. The same decay rate ( $k$  value) obtained from the deterministic model is used in the probabilistic analysis. The model is executed to generate 5475 values (the equivalent of 15 years of daily values) of flows and loads using Monte Carlo simulation at each sub-basin, propagate flows and loads in the same way as in the deterministic model (the decay of pathogen loads is considered) and finally estimate total loads and flows at point P as well as the outlet of each sub-basin. Expected (mean) values of annual pathogen loads are estimated as the mean value of the 5475 load values generated by the model. The number of impaired days, based on concentration, is calculated and divided by 15 (15 years) to estimate the expected number of



**Figure 9** | Statistical distribution of fecal coliform loads. (a) Actual vs. theoretical distributions. (b) Lognormal Q-Q plot of the fecal coliform loads.

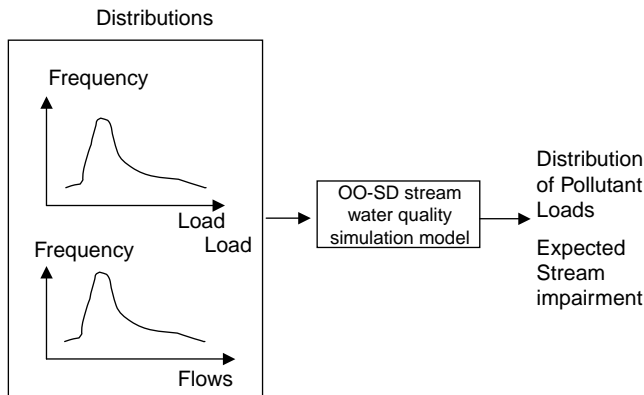


Figure 10 | Model configuration for probabilistic (stochastic) analysis.

impaired days per year. Figure 11 shows the relative load and flow contribution of each sub-basin at point P. It is evident that the expected value of the load contribution matches closely with the number of SF in the sub-basin. This provides more confidence in the probabilistic analysis.

Similar to the deterministic approach, two scenarios of 30% increase and decrease in the number of SF are investigated. It is important to note that assuming scenarios for different flow conditions is not necessary in probabilistic analysis since flow values are perturbed randomly over 5475 values. But the pathogen load perturbation is done with the average load per one SF, so changing the number of SF is expected to affect the expected annual value of pathogen loads as well as the number of impaired days and the distribution of pathogen concentration. The expected annual fecal coliform loads and number of impaired days per year at point P under the two scenarios are provided in Table 2. Comparing Tables 1 and 2 shows that expected

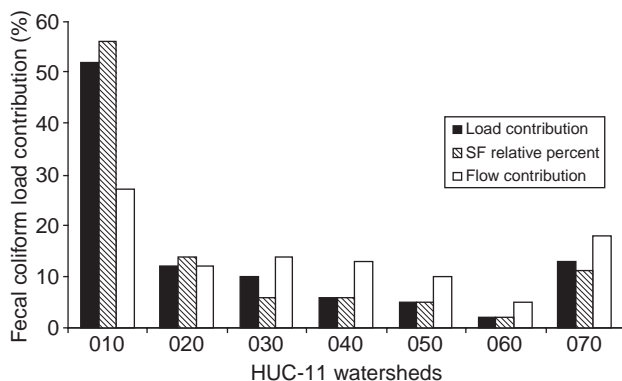


Figure 11 | Relative pollutant loads of individual HUC-11 sub-basins.

Table 2 | Average annual fecal coliform load and number of impaired days (probabilistic analysis)

Scenario	Average annual fecal load (count/d)	Number of impaired days per year
Baseline	$8.41 \times 10^{13}$	337
30% increase in SF	$1.21 \times 10^{14}$	348
30% decrease in SF	$6.62 \times 10^{13}$	315

annual loads (probabilistic analysis) under all scenarios are between the two extremes of dry years and wet years provided by the deterministic analysis. However, the number of impaired days, which is a decisive criterion for stream impairment, is higher in all cases based on probabilistic analysis. The fact that probabilistic analysis takes into account the inherent randomness of flow conditions and the uncertainties of the average pollutant loads provides more confidence in the probabilistic analysis. The distribution of fecal coliform concentrations at point P is provided in Figure 12. Decreasing the number of SF increase the frequency of low or acceptable concentrations as a tradeoff for decreasing the frequency of the very high concentrations ( $> 16\,000$  count/100 ml). However, no significant change in the frequency in the intermediate range within which everything is in violation of the standards is observed.

## DISCUSSION AND GENERAL REMARKS

In the proposed OO-SD modeling approach, flow-load relationships have been established using a regression

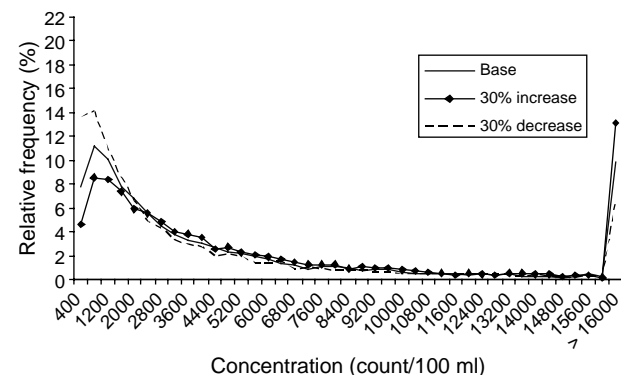
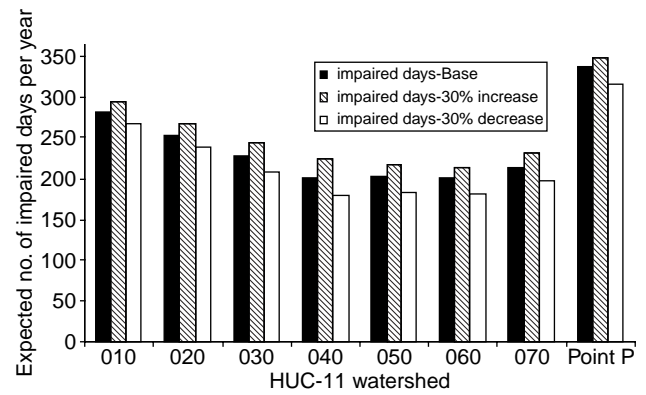


Figure 12 | The probability distribution of fecal coliform concentration at point P.

technique for the deterministic analysis. It should be noted that the adopted approach is flexible enough to allow for a more sophisticated rainfall–runoff hydrodynamic model to be included in place of the regression technique used in this study. The wide range of fecal concentrations observed throughout the year makes it extremely difficult for a single criterion (e.g. mean or median) to assess the conditions of the stream impairment throughout the year. The vulnerability criterion presented in this paper (number of impaired days in the year) is believed to address this issue by indicating how frequently the water quality in the stream is violated or is acceptable based on the definition of water quality standards. Such a criterion can also be useful in selecting the critical period or critical flow for different design problems, which helps, for instance, in developing total maximum daily loads (TMDLs) (Ormsbee *et al.* 2004).

The probabilistic analysis conducted in this study results in a conclusion, regarding the relative contribution of fecal coliform loads from upstream sub-basins, that is different from the conclusion achieved based on deterministic analysis. Probabilistic analysis reveals that the contribution of sub-basin -010 is significantly higher than those of the other sub-basins while the deterministic analysis referred to sub-basin -070 as the number one contributor to fecal coliform loads. Deterministic analysis generalizes and generates results based on a snapshot analysis using a small window of available data, but probabilistic analysis has the ability to provide a holistic picture of the possible outputs based on a possibly wide range of inputs. However, caution must be exercised before recommending that more resources be allocated to sub-basin -010, for two reasons. First, the stream impairment is dependent on the number of impaired days per year rather than the expected annual loads. The number of impaired days at the outlet of each sub-basin is provided in Figure 13. This figure provides a different perspective on the relative contributions of upstream sub-basins. It shows that almost all of the sub-basins play an important role in degrading the water quality at point P. This is due to the fact that the flow contribution, associated with the load, differs from one sub-basin to the other. Second, decreasing a certain number of SF distributed all over the watershed could be more effective than cleaning a certain sub-basin. For example, removing all SF from sub-basin -010 (1307 SF), which means 56% of the



**Figure 13** | Number of impaired days at different locations under different scenarios (probabilistic analysis).

total number of SF, has a different impact from removing the same number of SF distributed all over the watershed (seven sub-basins). The results of this comparison are provided in Table 3. Not only is there an improvement all over the watershed, with the exception of sub-basin -010, but also the number of impaired days at downstream point P is less when the cleaning effort is uniformly distributed.

This last conclusion, along with more comprehensive insights provided by the probabilistic analysis, highlights the probabilistic analysis approach as a viable tool to test different management alternatives. A final decision on a specific course of action may be made based on the expected number of impaired days per year at the point of concern. The effect of the course of action on the probability distribution of the pollutant concentration, such as Figure 12, may also be helpful in assessing the efficacy and risk of adopting a certain course of action.

One could argue against the concept of generating flows as a random variable since assessing the number of impaired days per year is related to daily flows, which are frequently autocorrelated. Flows and loads generated in this study are based on the statistical properties of the available record and samples, which were taken once a month for a few years. This eliminates, or at least reduces, the chances of autocorrelation in the flow record. However, in order to remove any doubt regarding this point, another run of the model has been executed with a record of the available natural daily flows, which are, as expected, autocorrelated. The expected annual fecal coliform load and number of impaired days per year are estimated and compared to the

**Table 3** | Number of impaired days as a result of localized and distributed cleaning efforts

Sub-basin	Base scenario		Clean sub-basin -010		56% decrease in SF (uniformly distributed)	
	No. of days	No. of days	% reduction of impaired days	No. of days	% reduction of impaired days	
010	281	0	100	245	13	
020	253	253	0	209	17	
030	228	228	0	183	20	
040	201	201	0	149	26	
050	202	202	0	154	24	
060	200	200	0	153	23	
070	215	215	0	168	28	
Point P	337	300	11	287	15	

results of the baseline scenario in which both flows and loads are treated as random variables. The results are as follows: expected annual fecal coliform load based on autocorrelated flow is  $8.6 \times 10^{13}$  count/d and the number of impaired days at point P is 333 d, while those values based on random flows are  $8.4 \times 10^{15}$  count/d and 337 d (Table 2), respectively. The small differences in these results marginalize the possible effects of the autocorrelation in the flow record on the model's outputs.

Finally, it is worth mentioning that the application presented in this paper is only an example of an approach that can be used to guide surface water quality management decisions. Visual inspection has been used in this study to select the probability distributions for both flow and pollutant load. A quantitative statistical approach could be adopted to improve the selection process. More data needs to be collected and subsequently more thorough calibration of the model should be conducted to increase the level of confidence in the generated outcomes and recommendations.

## CONCLUSIONS

The object-oriented simulation, based on the concepts of system dynamics (OO-SD) approach, is presented in this paper for simulating water quality management alternatives

at a watershed scale. The efficacy of the proposed approach in management aspects and its ability to handle and evaluate feasible potential scenarios are among its advantages. Conceptually simple approaches such as the OO-SD simulation provide good insights into the system and its behavior when the data relating to the system is scarce. Assessing the relative contribution of pollution from different sub-basins has been made possible using this approach. The use of the “number of impaired days in a year” criterion can help in better evaluation of management scenarios and can be more indicative of the water body conditions. The flexibility of the proposed approach allows for a more sophisticated model structure and also the use of probabilistic patterns of pollutant load. This is demonstrated by adopting a probabilistic framework to simulate the system. It is concluded that the probabilistic analysis provides a deeper insight and more credible predictions than the deterministic analysis.

## ACKNOWLEDGEMENTS

AE thanks the Natural Sciences and Engineering Research Council (NSERC) of Canada for its financial support through the Discovery Grants Program.

## REFERENCES

- Bedient, P. B. & Huber, W. C. 2002 *Hydrology and Floodplain Analysis*, Prentice Hall, Englewood Cliffs, NJ.
- Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., Johnson, K. M., Chan, P. W. H., Gherini, S. A. & Chamberlin, C. E. 1985 *Rates, Constants, and Kinetic Formulations in Surface Water Quality Modeling*. EPA/600/3-85/040, US Environmental Protection Agency, Washington, DC.
- Brown, L. C. & Barnwell, T. O. 1987 *The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User Manual*. EPA/600/3-87/007, Environmental Research Laboratory, EPA, Athens, GA.
- Chapra, S. C. 1994 *Surface Water Quality Modeling*. WCB/McGraw-Hill, New York.
- Chapra, S. C. 2003 Engineering water quality models and TMDLs. *J. Wat. Res. Plann. Mngmnt. ASCE*, **129**(4), 247–256.
- Clesceri, L. S., Greenberg, A. E. & Trussell, R. R. 1989 *Standard Methods for the Examination of Water and Wastewater*. APHA/AWWA/WPCF, Washington, DC.
- Connolly, J. P., Blumberg, A. F. & Quadri, J. D. 1999 Modeling fate of pathogenic organisms in coastal waters of Oahu, Hawaii. *J. Environ. Engng. ASCE*, **125**(5), 398–406.
- Donigian, A. S., Bicknell, B. R. & Imhoff, J. C. 1995 Hydrological Simulation Program – Fortran (HSPF). In *Computer Models of Watershed Hydrology*, (ed. V.P. Singh) Water Resources Publications, Littleton, CO.
- Eheart, J. W. & Ng, T. L. 2004 Role of effluent permit trading in total maximum daily load programs: overview and uncertainty and reliability implications. *J. Environ. Engng. ASCE*, **130**(6), 615–621.
- Elshorbagy, A., Teegavarapu, R. & Ormsbee, L. 2005a Total Maximum Daily Load (TMDL) approach to surface water quality management: concepts, issues and applications. *Can. J. Civil Engng.* **32**(2), 442–448.
- Elshorbagy, A., Teegavarapu, R. S. V. & Ormsbee, L. 2005b Framework for assessment of relative pollutant loads in streams with limited data. *Wat. Int.* **30**(4).
- Ford, A. 1999 *Modeling The Environment. An Introduction to System Dynamics Modeling of Environmental Systems*. Island Press, Washington, DC.
- HPS Inc. 2001 *Tutorial and Technical Documentation STELLA® II*, HPS Inc, Hanover, NH.
- Hodges, J. S. 1987 Uncertainty, policy analysis, and statistics (with discussion). *J. Am. Stat. Assoc.* **2**, 259–291.
- Jian, X. & Yu, Y. S. 1998 A comparative study of linear and nonlinear time series models for water quality. *J. Am. Wat. Res. Assoc.* **34**(3), 651–659.
- Kentucky Water Resources Research Institute (KWRI) 2000 *PRIDE Water Quality Assessment Report: I. Problems and Programs*. University of Kentucky, Lexington, KY.
- Levin, S. A. 1985 Scale and predictability in ecological modeling. In: T. L. Vincent, Y. Cohen, W. J. Grantham, G. P. Kirkwood and J. M. Skowronski (Eds.), *Modeling and Management of Resources Under Uncertainty. Lecture Notes in Biomathematics 72*, Springer-Verlag, Berlin, 1–8.
- McCuen, R. 2005 *Modeling Hydrologic Change. Statistical Methods*, Lewis Publishers, Boca Raton, FL.
- Mills, W. B., Borcella, B. B., Unga, M. J., Gherini, S. A., Summers, K. V., Iingsung, M., Rupp, G. L., Bowie, G. L. & Haith, D. A. 1985 *Water Quality Assessment: A Screening Procedure for Toxic and Conventional Pollutants in Surface and Ground Water*. EPA/600/6-85/002a-b. Environmental Research Laboratory, EPA, Athens, GA.
- Novotny, V. 2004 Simplified databased total maximum daily loads, or the world is log-normal. *J. Environ. Engng. ASCE*, **130**(6), 674–683.
- NRC 2001 *Assessing the TMDL Approach to Water Quality Management*. National Academy Press, Washington, DC.
- Ormsbee, L., Elshorbagy, A. & Zechman, E. 2004 A methodology for pH TMDLs: application to Beech Creek watershed. *J. Environ. Engng. ASCE*, **130**(2), 167–174.
- Reckhow, K. H. & Chapra, S.C. 1983 *Engineering Approaches for Lake Management, vol. Data Analysis and Empirical Modeling*, Butterworth, Woburn, MA.
- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. & Lorensen, W. 1991 *Object-oriented Modeling and Design*, Vol. 1. Prentice Hall, Englewood Cliffs, NJ.
- Simonovic, S. P. 1992 A knowledge-based advisory system for single multipurpose reservoir management. *Wat. Res. Develop.* **8**(3), 186–194.
- Simonovic, S. P., Fahmy, H. & Elshorbagy, A. 1997 The use of object-oriented modeling for water resources planning in Egypt. *Wat. Res. Mngmnt.* **11**, 243–261.
- Stow, A. C., Borsuk, Mark, E. & Reckhow K. H. 2000. *Nitrogen TMDL Development in the Neuse River Watershed: An Imperative for Adaptive Management*. Project report, Environmental Sciences and Policy Division, Nicholas School of the Environment and Earth Sciences, Duke University.
- Thomann, V. R. & Mueller, J. A. 1987 *Principles of Surface Water Quality Modeling and Control*. Harper and Row, New York.
- USEPA 1997 *Compendium of Tools for Watershed Assessment and TMDL Development*. EPA841-B-97-006. USEPA, Washington DC.
- Zhang, H. X. & Yu, S. L. 2004 Applying the first-order error analysis in determining the margin of safety for total maximum daily load computations. *J. Environ. Engng. ASCE*, **130**(6), 664–673.