Parameter optimization and Bayesian inference in an estuarine eutrophication model of intermediate complexity

P. Ding
C. E. S. Hoyle
M. E. Borsuk

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference


This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Parameter optimization and Bayesian inference in an estuarine eutrophication model of intermediate complexity

P. Ding, C. E. S. Hoyle, M. E. Borsuk

Thayer School of Engineering, Dartmouth College
Hanover, NH 03755 USA (Peng.Ding@Dartmouth.edu)

Abstract: The Neuse River Estuary, North Carolina, has been experiencing severe consequences of eutrophication in recent years including excessive algal blooms, low levels of dissolved oxygen, declining shellfish populations, large fishkills, and outbreaks of toxic microorganisms. As in many other marine systems, nitrogen has been identified as the pollutant of concern in the estuary because it is believed to stimulate the excessive algal growth that is at the root of other ecological problems. A model incorporating the mechanisms of algal growth and nutrient consumption in the Neuse River Estuary is formulated mathematically and implemented in the computer software AQUASIM. Key model parameters of water quality interest are calibrated to observations of system variables by minimizing the sum of the squares of the weighted difference between actual measurements and simulated results. The calibrated model reproduces the observed seasonal patterns of key system variables and thus demonstrates a predictive capability that is of use to policy makers when they are making decisions for sustainable environmental management. As future work we will implement Bayesian parameter estimation, which would improve the robustness of decision support by accounting for parameter uncertainty using probability distributions. Eventually, our model will be linked with a Bayesian version of the SPARROW watershed model as a Bayesian Network to be used for developing an adaptive implementation modeling and monitoring strategy (AIMMS) for the Neuse River basin.

Keywords: eutrophication; process dynamics; mechanistic models; bayesian inference; predictive uncertainty; environmental management

1. INTRODUCTION

Calibrated mechanistic models, together with a characterization of uncertainty, are indispensable tools for understanding environmental system dynamics and for supporting environmental management decisions. In this paper, we will examine both optimization-based and Bayesian parameter estimation methods using a simulation model of eutrophication in the Neuse River Estuary (North Carolina, USA). The methodological framework is applicable to a wide variety of disciplines (e.g., hydrology, ecotoxicology and air pollution) [Arhonditsis, 2007].

The Neuse River Estuary (NRE, Figure 1) has been experiencing characteristic symptoms of eutrophication since the late 1970s. Investigations conducted by the Division of Water Quality (DWQ) of North Carolina indicated that algal blooms, stimulated by excessive nutrients, especially nitrogen, cause low dissolved oxygen levels contributing to the extensive fish kills. These fish kills, algal blooms, and correspondingly high levels of chlorophyll a prompted DWQ to place the NRE on the 303(d) list of impaired waters in 1994.
A Total Maximum Daily Load (TMDL) for nitrogen was then developed based on the Nutrient Sensitive Waters (NSW) Management Strategy and additional environmental modelling. This TMDL, calling for a 30% reduction in nitrogen loading, was approved by EPA in 2002, and rules to support the NSW Management Strategy were fully implemented by 2003.

However, at this point we have not yet observed any significant decrease in actual nutrient loading to the estuary, although nitrogen loads from point sources have been reduced by 65%. Thus, the goal of a 30% reduction in total nitrogen loading and the anticipated reduction of chlorophyll a standard violations have not yet been achieved. This phenomenon might be due to the accumulation and recycling of nutrients in riverine and estuarine sediments. Therefore, development of a more detailed model to better understand the complex nutrient dynamics in the NRE is necessary.

In this paper, we develop an NRE simulation model of intermediate complexity. On the one hand, our model is more complex than the zero-dimensional Bayesian probability network model (Neu-BERN) [Borsuk et al. 2003] since detailed process mechanism is incorporated. On the other hand, it is less complex than the two-dimensional Neuse Estuary Eutrophication Model (NEEM) (Bowen, 2003) in which daily values of response variables are predicted on both a longitudinal and vertical grid. The model we describe here includes two vertical sediment layers intended to capture the mechanisms of nutrient accumulation and release. In addition to providing model results calibrated by conventional optimization-based parameter estimation, we highlight the use of Bayes’ theorem to describe parameter uncertainty and update our knowledge as new data become available.

![Figure 1](image)

**Figure 1.** Neuse River and watershed (left), and estuary (right), showing the delineation into five sections for this analysis. Points indicate the location of water quality sampling stations. [Borsuk et al., 2004]

2. MODEL DESCRIPTION

The development of our model is based on the vertical 4-box lake model developed by Mieletitner and Reichert [2008]. The estuary model is essentially a combination of multiple lake models in a row. For our modeling purpose, the Neuse River Estuary is divided longitudinally into six sections (Figure 1). These divisions are based on the hydrological and water quality characteristics in different regions of the estuary as well as the available sampling points. The nutrient concentration decreases while the salinity concentration increases as we move downstream in the NRE. The middle four sections (Upper, Middle, Bend, Lower) are ones of interest since they are where most water quality violations take place. Each of the middle regions (Upper, Middle, Bend, Lower) are divided vertically into four “boxes” including two water layers and two sediment layers (Figure 2). The water body is divided into two parts, the epilimnion and hypolimnion, because the NRE has been observed to stratify into two layers between the frequent mixing events. For shallow water such as the NRE, sediments play a critical role in providing nutrients for algae because accumulated nutrients in sediments may support algal growth for a long period of time after external nutrient loading reductions. By including two sediment layers (Figure 2), the mechanisms of nutrient interactions are modelled more precisely, thus providing a better simulation and forecast in the long run.
The model also includes the River and Sound sections. However, the river is modeled as having just one water layer since the movement and mixing in the Neuse River does not allow it to stratify. The Sound does not have any sediment layers as we are not concerned with the sediment dynamics in the sound. The state variables we are tracking throughout the estuary and over time are listed in Table 1.

![Diagram of the model configuration and processes](image)

**Figure 2.** Schematic of four-box vertical model configuration and processes represented

**Table 1:** Description and units of state variables as well as the time interval their corresponding measurement.

<table>
<thead>
<tr>
<th>State Variable</th>
<th>Units</th>
<th>Description</th>
<th>Data Span</th>
<th>Approximate Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_ALG</td>
<td>gDM/m3</td>
<td>Algae (dry mass)</td>
<td>1994-2000*</td>
<td>Weekly</td>
</tr>
<tr>
<td>X_ZOO</td>
<td>gDM/m3</td>
<td>Zooplankton (dry mass)</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_S</td>
<td>gDM/m3</td>
<td>Degradeable particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_I</td>
<td>gDM/m3</td>
<td>Inert organic particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_inorganic</td>
<td>gDM/m3</td>
<td>Inorganic particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_HPO4</td>
<td>gP/m3</td>
<td>Phosphate-phosphorus - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_NH4</td>
<td>gN/m3</td>
<td>Ammonia-nitrogen - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_NO3</td>
<td>mgN/l</td>
<td>Nitrate-nitrogen - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_O2</td>
<td>mg/l</td>
<td>Oxygen - dissolved</td>
<td>1995-2000</td>
<td>Weekly</td>
</tr>
<tr>
<td>S_S</td>
<td>mg/m3</td>
<td>Salinity</td>
<td>1996-2000</td>
<td>Weekly</td>
</tr>
</tbody>
</table>

*We do not have direct measurement for algae. Instead, chlorophyll data are converted to algal concentration.

**We do not have direct measurement for zooplankton, degradable particulate and inert organic particulate. Instead, TSS measurement data is used as a aggregate measure.
The dynamics of algae (X_ALG) and zooplankton (X_ZOO) comprise the biological part of our model. The growth, death and respiration processes are modeled in both of the water compartments (Figure 2). However, unlike the lake model, we treat algae as one group, not four functional groups. This is reasonable for our model as we are interested in the effect nutrients have on the algal community rather than the different forms of algae. Phosphate, ammonium and nitrate are the most relevant nutrients and, together with dissolved oxygen and salinity, represent the dissolved state variables of our model. In the NRE we are particularly interested in modeling nitrogen, which is considered to be the limiting nutrient and is therefore subject to loading limitations. Biodegradable (X_S) and inert organic matter (X_I) summarizes organic particles resulting from death of algae and zooplankton and from zooplankton excretion as fecal pellets. The biological dynamics, chemical reactions and physical transportation are modeled within various compartments and by the links between the compartments (Figure 2). Table 2 summarizes these processes.

**Table 2:** Biological, chemical and physical processes simulated in NREM

<table>
<thead>
<tr>
<th>Biological dynamics</th>
<th>Chemical reactions</th>
<th>Physical processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Growth of Algae and Zooplankton</td>
<td>• Aerobic/anaerobic mineralization</td>
<td>• Exchange of oxygen between water and air</td>
</tr>
<tr>
<td>• Death of Algae and Zooplankton</td>
<td>• Nitrification</td>
<td>• Light absorption</td>
</tr>
<tr>
<td>• Respiration of Algae and Zooplankton</td>
<td></td>
<td>• Sedimentation and diffusion</td>
</tr>
</tbody>
</table>

Data sets representing measured values of the state variables are used as a basis for comparison to check how closely the model results reproduce measured results over the time period modeled. Measurements of chlorophyll a, phosphate, nitrate, ammonia and salinity are available in all compartments (River, Upper, Middle, Bend, Lower, Sound) in both water layers. Unfortunately, we do not have regularly measured data for the sediment layers, only snapshots from detailed investigations.

### 3. MODEL IMPLEMENTATION

Our model is implemented in AQUASIM [Reichert, 1998], a simulation and data analysis software for aquatic systems. Users are allowed to define state variables and processes within a configuration consisting of compartments and links of the available types (Figure 3). The partial differential equations describing the physical and biological dynamics of our estuary system are solved with the DASSL [Petzold, 1983] implementation of the backward-difference GEAR integration technique [Gear, 1971].

![Figure 3. The four functional units in AQUASIM and their relationships.](image-url)
4. PARAMETER CALIBRATION

4.1 Optimization-Based Parameter Estimation

To perform parameter calibration is essentially to solve a constrained optimization problem. The optimization objective function is to minimize the sum of the squares of the weighted difference between actual measurements and simulated results within the constraints of parameter ranges. Two numerical techniques implemented in AQUASIM are available for this parameter estimation task. The secant method [Ralston and Jennrich, 1978] converts our nonlinear least square fitting problem to a linear one by approximating the objective manifold with a secant plane through previous step objective function values. The downhill simplex method [Nelder and Mead, 1965] finds iteratively an improved searching direction in terms of decreasing the sum of squares of the weighted difference between actual measurements and simulated results using local information of function values only. Both secant and downhill simplex methods are derivative-free algorithms which are much less expensive computationally than derivative-based methods, such as gradient descent. The secant method is much more efficient in terms of convergence rate since it rapidly jumps to the position found by parabolic extrapolation, whereas the downhill simplex method slowly moves down the “gradient” of the objective function. Therefore, in a practical sense, parameter estimation usually starts with a secant algorithm. Also, we usually run multiple parameter estimation processes with different initial guesses of target parameters to check the convergence of the algorithm.

4.2 Bayesian Learning of Parameter Uncertainty

In water quality assessment and management, mechanistic simulation models are powerful in terms of understanding physical and biochemical processes, predicting aquatic ecosystem response to external nutrient loading changes and supporting the environmental policy making process [Reckhow and Chapra, 1999]. But these models are never “perfect” since models are always a simplified version of reality [Stow et al., 2003]. The imperfectness is described by the term uncertainty, usually characterized quantitatively by probability distributions. Parameter uncertainty is the type of uncertainty we are often interested in investigating with an established mechanistic simulation model. However, conventional optimization-based parameter calibration techniques fail to support this analysis in two aspects. One problem is that conventional parameter calibration providing a best fit of the model parameters to the dataset only computes a set of fixed parameter values. This procedure is formally referred as maximum likelihood estimation by frequentist statisticians. Equation (1) illustrates this mathematically, where \( \theta \) represents the data likelihood conditional on the parameter vector \( \theta \) of our model.

\[
\theta_{\text{optimal}} = \arg \max_{\theta} \left[ \text{likelihood}(\text{data} | \theta) \right]
\]  

The other problem is that conventional parameter calibration makes the model data-specific by fitting it to a given dataset at the moment. As new datasets become available, the model has to be recalibrated without a possibility of considering the previous parameter results. In other words, we do not update existing knowledge about model parameters, but rather we make our model “learn” from the very beginning whenever new information is available.

Fortunately, Bayes theorem provides us a means for addressing these two problems. Existing knowledge about parameters, or a prior distribution, is updated according to Bayes updating rule (equation 2) when new information becomes available, resulting in updated parameter knowledge, or a posterior distribution, of parameters.

\[
f_{\text{post}}(\theta | \text{data}) = \frac{f_{\text{like}}(\text{data} | \theta) f_{\text{pri}}(\theta)}{\int_{\theta} f_{\text{like}}(\text{data} | \theta) f_{\text{pri}}(\theta)}
\]
In this equation, \( f_{\text{post}}, f_{\text{like}}, f_{\text{pri}} \) denote the posterior distribution, model likelihood and prior distribution, respectively. To reduce the uncertainty about parameters, we can use the Bayesian updating rule iteratively, using the previous step’s posterior distribution as the prior distribution for the next step and obtaining a new posterior distribution with the new data provided. Computationally, the posterior parameter distributions are often approximated by their samples simulated based on Markov Chain Monte Carlo (MCMC) method. Bayesian inference changes our perspective of seeking a single “optimal” value for each model parameter, to finding a joint distribution of parameter sets. These probability distributions provide a straightforward way to quantify parameter uncertainty that can be easily used by decision makers/policy planners [Reckhow, 1994; Arhonditsis et al., 2006].

As future work, we will implement the Bayesian parameter estimation using UNCSIM [Reichert, 2004], a program package for statistical inference and uncertainty analysis. UNCSIM not only provides routines for frequentist (maximum likelihood) but also for Bayesian (Markov Chain Monte Carlo and importance sampling techniques) parameter estimation. Therefore, we can easily compare these two methods.

5. RESULTS

Optimization-based parameter estimation in AQUASIM requires two inputs: parameters of interest and measured data used as the fitting target. To reduce the computational time as well as to fit parameters more precisely, parameters of interest were divided into five groups and they are being calibrated sequentially based on relevant observed data (Table 3).

<table>
<thead>
<tr>
<th>Group Order</th>
<th>Parameter</th>
<th>Calibration Data</th>
<th>Additional Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dispersion constants</td>
<td>salinity and oxygen (epi* and hypo**)</td>
<td>none</td>
</tr>
<tr>
<td>2</td>
<td>Algal parameters (physical)</td>
<td>oxygen (epi), chlorophyll a (epi and hypo)</td>
<td>Fix phosphate, ammonia and nitrate concentrations at mean values</td>
</tr>
<tr>
<td>3</td>
<td>Zooplankton parameters</td>
<td>oxygen (epi), chlorophyll a (epi and hypo)</td>
<td>Fix phosphate, ammonia and nitrate concentrations at mean values</td>
</tr>
<tr>
<td>4</td>
<td>Sediment/water bacteria parameters</td>
<td>oxygen (epi and hypo), phosphate and nitrate (epi and hypo)</td>
<td>none</td>
</tr>
<tr>
<td>5</td>
<td>Algal parameters (nutrient)</td>
<td>all data</td>
<td>none</td>
</tr>
</tbody>
</table>

epi* represents epilimnion and hypo** represents hypolimnion

Dispersion parameters are important as they characterize the physical properties of the NRE. Thus, we calibrated these parameters according to salinity and oxygen measurements from both of the water layers as the first step. A comparison (Figure 4) of model results and measurements indicates that the simulations generally reproduce the seasonal and intermediate time patterns of salinity and oxygen levels and match the downstream trends. The simulated oxygen concentrations in the hypolimnion are currently too low relative to observed values, but we expect that to improve as we calibrate sediment and water bacterial parameters (Group 4). We are currently working on the remaining four groups of the parameter calibration.
Figure 4. Simulated vs. measured salinity (left) and oxygen (right) concentrations of the two water layers across the Upper, Middle, Bend and Lower sections with time range from 6/1/1996 to 12/27/2000. Concentration units of salinity: mg/m³. Concentration units of oxygen: mg/l.

Black lines: simulated model results in epilimnion, Yellow: measurements in epilimnion
Blue: simulated model results in hypolimnion, Magenta: measurements in hypolimnion

6. OUTLOOK

The next step will be to conduct Bayesian inference on the most significant parameters so that we have full parameter probability distributions for implementation in the Bayesian Network model. This will be linked with a Bayesian implementation of the SPARROW watershed nutrient delivery model [Smith et al., 1997] as part of a larger project. It is intended that these two linked sub models will facilitate a process of Bayesian adaptive learning and management on the basis of new evidence, including remotely sensed land use and water quality data (Figure 5).

ACKNOWLEDGEMENTS

This work was supported by a grant from the EPA Office of Research and Development’s Advanced Monitoring Initiative (AMI) Pilot Projects Focused on GEOSS (Global Earth Observation System of Systems) (PI: K.H. Reckhow, Duke University). The authors thank Martin Frey, Peter Reichert, Johanna Mieleitner, and Anne Dietzel of Eawag for model support and Ibrahim Alameddine of Duke University for assistance with data.
REFERENCES


