Stochastic And Deterministic Parameter Estimation Of Coupled Bacteria-Sediment Fate And Transport In Streams

Tri Le

Arash Massoudieh

Follow this and additional works at: http://academicworks.cuny.edu/cc_conf_hic

Part of the Water Resource Management Commons

Recommended Citation

This Presentation is brought to you for free and open access by CUNY Academic Works. It has been accepted for inclusion in International Conference on Hydroinformatics by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
STOCHASTIC AND DETERMINISTIC PARAMETER ESTIMATION OF COUPLED BACTERIA-SEDIMENT FATE AND TRANSPORT IN STREAMS

TRI LE (1), ARASH MASSOUDIEH (1)


In this presentation parameter estimation of a mechanistic model of bacteria-sediment interaction using a deterministic method through a hybrid genetic algorithm and also stochastically through Makov-Chain Monte Carlo (MCMC) approach will be presented. The physically-based model considers the advective-dispersive transport of sediments as well as both free-floating and sediment-associated bacteria in the water column and also the fate and transport of bacteria in the bed sediments. The bed sediments are treated as a distributed system which allows modeling the evolution of the vertical distribution of bacteria as a result of sedimentation and resuspension, diffusion and bioturbation in the sediments. The model is applied to sediment and E. coli concentration data collected during a high flow event in a small stream historically receiving agricultural runoff. The MCMC method is used to estimate the likeliest values as well as the joint probability density functions of model parameters including sediment deposition and erosion rates, critical shear stress for deposition and erosion, attachment and detachment rate constants of E. coli to/from sediments and also the effective diffusion coefficients of E. coli in the bed sediments. The uncertainties associated with the estimated parameters are quantified via the MCMC approach and the correlation between the posterior distribution of parameters have been used to assess the model adequacy and parsimony.

INTRODUCTION

E. Coli is widely used as an indicator organism to assess the risk of pathogenic bacteria in water bodies. Due to their strong association with suspended and bed sediments, the fate and transport of micro-organisms in water bodies is strongly controlled by sediment dynamics. It has been shown that bed sediments can contain orders of magnitude larger pathogen concentration than the water column and these sediment-associated bacteria can be released into the water column as a result of high flow velocities that causes sediment resuspension. Mechanistic (process-based) models can serve as predictive tools to perform risk assessment, and also can help us to test hypothesis about the role and importance of various processes in the fate and transport of bacteria in streams. In order to use models reliability it is required that the value of various model parameters controlling various processes governing the fate and
transport of bacteria to be know. In a heterogeneous stream system direct methods to measure these parameters are not always available due several factors including practical limitation or the fact that many of these parameters are in fact representative of model abstraction rather than measurable physical quantities or represent the interaction of various processes. Inverse modeling based on approaches like maximum likelihood or least squared error provides a way to determine estimate model parameters using observed data representing the output of the model. However, due to several sources of uncertainty including the uncertainty associated with measured data, model structural error and also the uncertainty associated with external forcing (e.g. boundary conditions, lateral inflows, etc.) there will be uncertainties associated with the parameters obtained through deterministic approaches. Bayesian inference provides powerful tool to infer the joint probability distribution of the parameters. Below development of the mechanistic model of bacterial transport in streams and the inverse modeling framework is described.

**METHODS**

*Mechanistic Model*

The model of bacterial transport in streams used in this study is consisted of three components including a 1-D hydrodynamic model, a 1-D sediment transport model and a coupled planktonic and sediment-associated bacterial fate and transport model. Saint-venant equations have been used to calculate the flow rate and velocity in the stream during the course of the simulation [1]. Sediment transport is modeled using a one-dimensional advection-dispersion-erosion-deposition:

\[
\frac{\partial C_s}{\partial t} + v \frac{\partial C_s}{\partial x} = \frac{1}{A} \frac{\partial}{\partial x} \left( D_s A \frac{\partial C_s}{\partial x} \right) + \frac{q_{in}}{A} (C_{s,in} - C_s) + \frac{P}{A} (E_r - w_p C_s)
\]  

(1)

Where \( C_s \) is the concentration of suspended particles (g m\(^{-3}\)), \( D_s \) is the dispersion coefficient for suspended particles (m\(^2\) s\(^{-1}\)), \( q_{in} \) is the amount of lateral water flux (m\(^2\) s\(^{-1}\)), \( P \) is the wetted perimeter of the stream (m), \( C_{s,in} \) is the concentration of suspended particles in the lateral inflow (g m\(^{-3}\)), \( E_r \) is the sediment entrainment rate (g m\(^{-2}\) s\(^{-1}\)), \( w_p \) is the deposition rate parameter (1/s). Chung et al. (2009) has calculated the rate of resuspension of sediment by the following equation:

\[
E_r = E \left( \frac{\tau_{u_b}(\tau) - \tau_{u_c}(\tau)}{\tau_{u_b}(\tau)} \right)
\]

(2)

Where \( \tau_{u_b}(\tau) \), the flow shear stress, is calculated as \( \rho C_f v^2 \) is the critical shear stress for erosion, and \( E \) is the bed erodibility coefficient. Mehta, et al., [2] and Partheniades [3] used the following relationship to indicate the deposition rate coefficient:

\[
w_p = \left( 1 - \frac{\tau_{u_b}(\tau)}{\tau_{u_c}(\tau)} \right) v_s
\]

(3)

Where \( \tau_{u_c}(\tau) \) is the critical shear stress for deposition and \( v_s \) is the settling velocity which can be calculated as:
Figure 1 illustrates the processes considered in pathogen transport in the stream. The pathogens can be considered to appear in two phases: planktonic phase and attached to particles (particle-bound) phase. The bacterial exchange between the two phases is represented by a first-order kinetic rate model where the rate can be considered as a function of the turbulent energy in the stream. Bacterial exchange with benthic sediments planktonic pathogens are assumed to be via transfer through the boundary layer represented as a linear kinetic exchange process whereas for particle-bound pathogens, bacterial mass exchange is considered to result from resuspension and deposition. The transport processes affecting both planktonic bacteria and particle-bound bacteria phases were simulated by a pair of coupled 1-D advection-dispersion equations [4]. assuming that the transport of pathogens in the overlying water is controlled by a 1-D advection-dispersion process, the fate and transport equation for planktonic bacteria in the water column can be written as:

\[
\frac{\partial A N_p}{\partial t} + v \frac{\partial A N_p}{\partial x} = \frac{\partial}{\partial x} \left( D A \frac{\partial N_p}{\partial x} \right) + P k_b [n_p(0) - N_p] - S(k_f N_p - k_f N_a) + P u_f \theta_0 n_p(0) + q_{in} N_{p, in} \tag{5}
\]

Similarly the population balance equation for the attached bacterial population can be written as:

\[
\frac{\partial A S N_a}{\partial t} + v \frac{\partial A S N_a}{\partial x} = \frac{\partial}{\partial x} \left( D_s A \frac{\partial N_a}{\partial x} \right) + P \cdot E \cdot R \cdot n_a(0) - P \cdot v \cdot S \cdot N_a + C_s (k_r N_p - k_f N_a) + \mu_a N_a + q_{in} S_{in} N_{in, a} \tag{6}
\]

where \( A \) is the cross-sectional area of the stream which can be a function of time and space (\( t \) and \( x \)), \( N_p \) is the number concentration (number per volume) of planktonic pathogens in the water columns, \( v \) is the mean flow velocity in the stream (m s\(^{-1}\)), \( D \) is the dispersion coefficient in the stream, \( P \) is the wetted perimeter of the stream which can be assumed to be equal to the width of the stream for shallow channels (m), \( k_b \) is the transfer rate of planktonic pathogens through the laminar boundary layer at the sediment water interface, \( n_p \) is the planktonic bacteria number concentration in the pore water in the benthic sediments and the argument (0) indicate the vertical location being on the sediment-water interface, \( S \) is the concentration of suspended solids, \( N_{a} \) is the attached concentration of pathogens expressed as the number of pathogens per mass of suspended sediments, \( u_f \) is the sum of hyporheic flow and upward flow.
velocity due to sediment consolidation in the benthic sediments, \( \theta_0 \) is the porosity of benthic sediments at the sediment-water interface, \( D_s \) is the dispersion coefficient for the sediments. In the benthic sediment, a similar first-order attachment-detachment governing equation was used to model the bacteria exchange between the pore water and the sediment, while a Fickian diffusive model was incorporated in order to simulate the mixing due to bioturbation [4]:

\[
\frac{\partial \theta n_p}{\partial t} + \frac{\partial (\theta u n_p)}{\partial z} = \frac{\partial}{\partial z} \left[ (D_m + D_B) \theta \frac{\partial n_p}{\partial z} \right] - B_d \left( k_r n_p - k_f n_a \right) \tag{7}
\]

\[
\frac{\partial B_d n_a}{\partial t} + \frac{\partial (u z B_d n_a)}{\partial z} = \frac{\partial}{\partial z} \left[ D_B \frac{\partial B_d n_a}{\partial z} \right] - B_d \left( k_r n_p - k_f n_a \right) \tag{8}
\]

where \( B_d \) and \( \theta \) are respectively the bulk density, and the porosity of the sediments that can be a function of time and depth, \( u \) is the pore water velocity as a result of hyporheic flow and sediment consolidation, \( D_m \) and \( D_B \) are respectively Brownian diffusion coefficient and diffusion coefficient indicating bioturbation, and \( u_s \) is the vertical velocity of sediments as a result of consolidation. It should be noted that the net growth rate coefficients \( u_p \) and \( u_a \) are ignored due to the above assumption.

**Parameter Estimation using Bayesian Inference and Markov Chain Monte Carlo (MCMC)**

Bayesian inverse modeling approach is based on the Bayes' theorem which able to uniformly treat the uncertainty at all levels of the modeling process as well as allow incorporation of prior knowledge and the seamless combination of such knowledge with observed data [5]. A program using the C++ language was developed to perform Bayesian inverse modeling to not only evaluating the credibility of estimated parameters but also calculating the appropriate level of complexity of the macro-scale transport model. Applying the Bayes' rules in sediment and E. coli transport modeling can be written as follows:

\[
P(\theta | \mathcal{C}) = \frac{P(\mathcal{C} | \theta)P(\theta)}{P(\mathcal{C})} \tag{9}
\]

where \( \theta \) is a set of model parameters, \( \mathcal{C} \) is the modeled BTC with parameter set \( \theta \), so \( O \) and \( \mathcal{C} \) are represented as matrices \( [c_{ij}] \) and \( [\mathcal{c}_{ij}] \) respectively, and these two matrices illustrate the measured and modeled concentrations of E. coli at sample \( i \) in column \( j \). \( P(\theta | \mathcal{C}) \) is the posterior distribution of model parameters \( \theta \) - given the observed BTCs \( \mathcal{C} \), \( P(\mathcal{C} | \theta) \) is the probability of seeing the observed BTCs - given model parameters \( \theta \), \( P(\theta) \) is the prior information about the model parameters, \( P(\mathcal{C}) \) is a normalizing factor. The likelihood function can be rewritten if the observation error structure is assumed to be additive and Gaussian:

\[
P(\mathcal{C} | \theta) = \frac{1}{(\sqrt{2\pi}\sigma^2)^{mn}} \exp \left[ - \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (c_{ij} - \mathcal{c}_{ij})^2}{2\sigma^2} \right] \tag{10}
\]

Where \( m \) is the number of observation, \( n \) is the number of samples, and \( \sigma \) is the standard deviation of observation error. The Metropolis-Hasting Markov Chain Monte-Carlo (MCMC) algorithm was used to draw a large number of samples from posterior distribution and the different statistics of the posterior distribution of the parameters was obtained from that.
RESULTS AND DISCUSSION

The modeling framework was applied to the data obtained during the experiments conducted on a stream at the Optimizing Production Inputs for Economic and watershed research site, USDA-BARC on the mid-Atlantic coastal plain of Maryland, USA by Cho et al. [Cho, et al. [6]]. Specifically, the watershed of the stream that was studied is 70 ha and contains 15% of deciduous forest and 75% of it is used for agricultural crop production. The width of creek is variable from 65 m to 100 m, while the entire length of it is approximately 1100 m. In the last 12 years, corn production has been operating in all four fields A, B, C and D surrounding the stream with the total area of 22.5 ha [6]. On August 12, 2008, 80 m$^3$ of city water was poured into the stream at station 1 to create the artificial flow at the rate of 60 L per second in four equal 20-m$^3$ allotments. The flow, turbidity and e-coli concentrations were sampled at three stations downstream of station 1 respectively referred to as stations 2, 3, and 4. These stations are respectively 140 m, 270 m and 630 m downstream of station number 1. In the other words, the creek was divided into 3 reaches: 140 m for reach 1, 130 m for reach 23 and 360 m for reach 34.

The Bayesian method was used to estimate sediment transport parameters including critical shear stress for resuspension ($\tau_{C_r}$) [N m$^{-2}$], critical shear stress for deposition ($\tau_{C_d}$) [N m$^{-2}$], the bed erodibility ($E_0$), deposition rate ($w_0$) [m s$^{-1}$], eta coefficient ($\eta$), $\gamma_0$ parameter as well as the parameter controlling E. coli fate and transport including sediment-water mass exchange coefficient for the dissolved bacteria ($k_b$), mass exchange coefficient between suspended particle and water ($k_s$), dispersion coefficient ($d_p$) [m$^2$s$^{-1}$] and the water-solid distribution coefficient ($k_f$). Figure 2 shows the posterior distribution of the parameters controlling sediment transport and the posterior distribution of the parameters governing the fate and transport of E. coli are depicted in Figure 3. As it can be noticed the range of the estimated parameters are large for some of the parameters which is mainly due to the fact that

![Figure 2](https://example.com/figure2.png)

*Figure 2. Histograms representing the posterior parameters obtained from BTCs of Sediment transport*
the model is over-parameterized relative to the amount of data and the possibility of lack of sensitivity of the model outputs to some of the parameters or colinearity between parameters.

The agreement between modeled and observed *E. coli* and sediment concentration in all three stations is represented in Fig. 4. In these figures the dashed line is the result of the deterministic (i.e. likeliest) value of the parameters while the grey lines represent 60 randomly sampled parameters from the posterior distribution of parameters. Although the modeled sediment concentration in both three stations approximately captured the trend of sediment dynamic during the high flow event, it should be pointed out that the calculated results were under-estimated compared to the observed data. On the other hand, the estimated values of *E. coli* concentration in station 4 efficiently simulated the tendency as well as magnitude of *E. coli* transport in downstream. However, the predicted data is overestimated in station 3 and underestimated in station 2.

Although the results of modeling flow rate, sediment concentration and *E. coli* concentration effectively captured the tendency and arrival time during or shortly after rainfall events, there are still several defects that lead to over or underestimation in term of the magnitude. However the level of such discrepancies seem acceptable considering the difficulties in accurately measuring the *E. coli* concentration and the heterogeneities present in the natural system that were not represented in the model. The fact that the sediment particle size is distributed and not uniform can affect the ability of the model to predict sediment transport.
Fig. 4. Modeled and Observed E. coli concentration (#/L) (left) and suspended sediment concentration (right) at station 2, 3 and 4 during artificial high-flow event.

REFERENCES


