Science of the Total Environment 590–591 (2017) 287–296

Contents lists available at ScienceDirect

Science of the Total Environment

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

Bayesian inference for kinetic models of biotransformation using a generalized rate equation

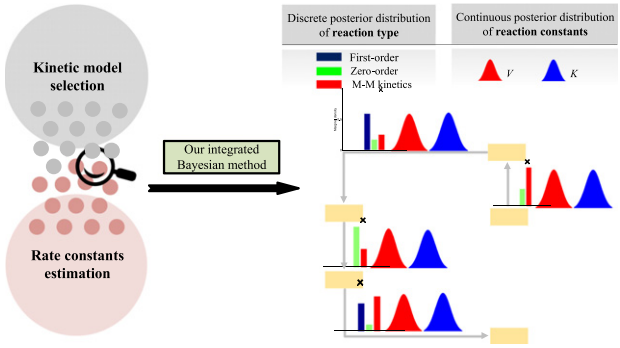
Shanshan Ying [a], Jiangjiang Zhang [a], Lingzao Zeng [a,*], Jiachun Shi [a], Laosheng Wu [a,b]

[a] College of Environmental and Natural Resource Sciences, Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, Zhejiang University, 310058 Hangzhou, China
[b] Department of Environmental Science, University of California, Riverside, CA 92521, United States

HIGHLIGHTS

- An integrated Bayesian method was developed based on a generalized rate equation.
- The method copes with model selection and parameter estimation simultaneously.
- Model hypotheses were generated by discrete parameters in rate equation.
- The method fully searches candidate models but avoids high computation cost.
- The method was validated through a numerical case and a complex experimental study.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 16 December 2016
Received in revised form 28 February 2017
Accepted 1 March 2017
Available online 6 March 2017

Editor: Jay Gan

Keywords:
Bayesian model selection
Parameter estimation
Rate equation
N transformation

ABSTRACT

Selecting proper rate equations for the kinetic models is essential to quantify biotransformation processes in the environment. Bayesian model selection method can be used to evaluate the candidate models. However, comparisons of all plausible models can result in high computational cost, while limiting the number of candidate models may lead to biased results. In this work, we developed an integrated Bayesian method to simultaneously perform model selection and parameter estimation by using a generalized rate equation. In the approach, the model hypotheses were represented by discrete parameters and the rate constants were represented by continuous parameters. Then Bayesian inference of the kinetic models was solved by implementing Markov Chain Monte Carlo simulation for parameter estimation with the mixed (i.e., discrete and continuous) priors. The validity of this approach was illustrated through a synthetic case and a nitrogen transformation experimental study. It showed that our method can successfully identify the plausible models and parameters, as well as uncertainties therein. Thus this method can provide a powerful tool to reveal more insightful information for the complex biotransformation processes.

1. Introduction

A variety of rate equations have been proposed to describe biotransformation processes in the environment (Aharoni et al., 1991). First-order, zero-order and Michaelis-Menten (M-M) rate equations were frequently applied in a wide range of reactions,

* Corresponding author.
E-mail address: lingzao@zju.edu.cn (L. Zeng).

http://dx.doi.org/10.1016/j.scitotenv.2017.03.003
0048-9697/

including the decay of POPs (Casey and Simunek, 2001; Weetman et al., 2015), carbon and N cycle (German et al., 2012; Müller et al., 2014), release and sorption of compounds like metal ions and pharmaceuticals (Fan et al., 2016; Inyang et al., 2016; Martínez-Hernández et al., 2016) and photocatalytic activities of nanoparticles (Xu et al., 2002; Murdoch et al., 2011). A first-order rate equation assumes that the maximum concentration of the substrate is much less than the half-saturation constant (Bekins et al., 1998). If the substrate concentration is high, a zero-order rate equation works better, while the M-M rate equation should be chosen under conditions between first-order and zero-order kinetics.

Traditionally, the form of a rate equation in a kinetic model is often chosen based on literature reports, and the corresponding reaction rates are then obtained by fitting the experimental data to the chosen rate equation. For example, micropollutant biotransformation in activated sludge is always modeled as pseudo first-order process (Kern et al., 2010; Jasper et al., 2014). Nevertheless, the consistency of the relationship between the rate equation and biotransformation properties has seldom been discussed in previous literatures. Thus, it is questionable to directly employ the rate equations from previous studies.

Recently, the process-based models have often been combined with inverse modeling methods to provide more insightful information for the complex biotransformation processes (Boisson et al., 2013; Molina-Herrera et al., 2016). While some studies attempted to select the rate equation from several candidates to obtain the best kinetic model (Müller et al., 2004; Wu et al., 2015), most other studies still chose their rate equations based on the previous literature, which may lead to the misuse of kinetic models (Scow et al., 1986; Sparks, 2013).

Bayesian inference can provide a rigorous framework for model selection and parameter estimation (Box and Tiao, 2011). However, when the number of candidate reactions is large, the computational cost in evaluating the alternative kinetic models is practically unaffordable. Ye et al. (2004) developed a maximum-likelihood-estimate (MLE) approximation for Bayesian model averaging in unsaturated fractured tuff. Toni et al. (2009) developed an Approximated Bayesian computation method for parameter estimation and model selection for dynamical processes. A Bayesian model selection approach was employed to develop a data-informed model to discuss the relationship between N_2O emissions and WFPS and soil temperature (Huang et al., 2013). Galagali and Marzouk (2015) imposed the point-mass mixture priors, which included a mass at zero mixed with a continuous density over rate constants.

To improve the full search of candidate models and simplify the process of model selection, we used a generalized rate equation in the kinetic models and for the rate constants estimation of biotransformation processes in this paper. The selection among reaction equations can be achieved by simply setting the indexing parameters as prescribed discrete values, and the indexing parameters can be simultaneously estimated with respective reaction rate parameters by Markov Chain Monte Carlo (MCMC), a sampling method to draw posterior realizations. Another related sampling approach, i.e., reversible-jump MCMC (Green, 1995), has been applied to infer kinetic models (Oates et al., 2012). In practice, however, it is difficult to design appropriate proposal distributions for between-model moves in a reversible-jump MCMC algorithm (Green and Hastie, 2009).

Comparing with other model selection techniques, the proposed approach generates the model hypotheses systematically and reduces the model selection to parameter estimation with discrete priors. Combined with the state-of-art adaptive MCMC, this approach provides a more direct and efficient tool for kinetic model selection and parameter estimation. In this work, the validity of this proposed method was illustrated by a synthetic numerical case and a complex nitrogen (N) transformation experimental study.

2. Methods

2.1. Generalized rate equation

The dynamics of a contaminant in the biotransformation processes can be defined as follows:

$$\frac{dc}{dt} = \sum_{i=1}^{m} v_i \tag{1}$$

where c is the concentration of the contaminant, t is time, v is reaction rate, i is the index of reaction pathway, m is the total number of reactions for the contaminant in the biotransformation processes. v_i denotes the rate of i-th reaction, which is defined by the following generalized rate equation:

$$v = \frac{Vc}{aK + bc} \tag{2}$$

where a and b are control parameters equal to either zero (0) or one (1). Three combinations of a and b are considered to realize the switch of rate equations among first-order ($a = 1$ and $b = 0$), zero-order ($a = 0$ and $b = 1$) and M-M ($a = 1$ and $b = 1$) rate equations.

When the rate equation uses the form of M-M kinetics, K is the substrate concentration at the half-maximal velocity and V is the maximal velocity. Here K, V, a and b are all unknown kinetic parameters to be estimated. By searching the optimal combinations of a and b for each reaction pathway, the optimal kinetic models of biotransformation processes can be identified, which is realized by parameter estimation with discrete priors.

For complex biotransformation processes with multiple pools, a set of ordinary differential equations (ODEs) are developed according to Eq. (1). These models were programmed with Matlab® in this work. The ODE solver (ode15s) was employed due to its capability in dealing with stiff problems (Ebert et al., 2012).

2.2. Bayesian inference

Bayesian model selection can be used to compare candidate models $M_1, \ldots, M_N$ by computing the posterior probability for each model (Wasserman, 2000). The posterior probability for model M_n is given by

$$p(M_n|\mathbf{Y}) = \frac{p(\mathbf{Y}|M_n)p(M_n)}{\sum_{l=1}^{N} p(\mathbf{Y}|M_l)p(M_l)} \tag{3}$$

where

$$p(\mathbf{Y}|M_n) = \int p(\mathbf{Y}|\boldsymbol{\theta}_n, M_n)p(\boldsymbol{\theta}_n|M_n)d\boldsymbol{\theta}_n \tag{4}$$

is the integrated likelihood of model M_n, $\boldsymbol{\theta}_n$ is the vector of parameters of model M_n, $p(\boldsymbol{\theta}_n|M_n)$ is the prior probability of $\boldsymbol{\theta}_n$ under model M_n, $p(\mathbf{Y}|\boldsymbol{\theta}_n, M_n)$ is the likelihood, and $p(M_n)$ is the prior probability that model M_n is the true model. The model that maximizes $P(M_n|\mathbf{Y})$ is considered as the optimal model. Therefore, the model selection relies on the computation of posterior probability $P(M_n|\mathbf{Y})$ for each model, which can result in high computational cost if the number of candidate models is large. On the other hand, limiting the number of candidate models would lead to potential bias in model selection.

To solve this problem, we converted the model selection into parameter estimation of a and b in Eq. (2). To simplify the implementation, we introduced an indexing parameter S to represent the three combinations of a and b (i.e., if $S = 1$, $a = 1$ and $b = 0$ for first-order; if $S = 2$, $a = 0$ and $b = 1$ for zero-order; if $S = 3$, $a = 1$ and $b = 1$ for M-M kinetics). Let $\boldsymbol{\theta}'$ denote the vector of all the parameters and assume that the kinetic parameters for each reaction are independent. Then,