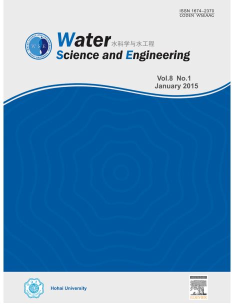
Accepted Manuscript



Impacts of prior parameter distributions on Bayesian evaluation of groundwater model complexity

Saeideh Samani, Ming Ye, Fan Zhang, Yong-zhen Pei, Guo-ping Tang, Ahmed Elshall, Asghar A. Moghaddam

PII: S1674-2370(18)30049-8
DOI: 10.1016/j.wse.2018.06.001
Reference: WSE 136

To appear in: *Water Science and Engineering*

Received Date: 14 September 2017

Accepted Date: 10 January 2018

Please cite this article as: Samani, S., Ye, M., Zhang, F., Pei, Y.-z., Tang, G.-p., Elshall, A., Moghaddam, A.A., Impacts of prior parameter distributions on Bayesian evaluation of groundwater model complexity, *Water Science and Engineering* (2018), doi: 10.1016/j.wse.2018.06.001.

Impacts of prior parameter distributions on Bayesian evaluation of groundwater model complexity

Saeideh Samani [a, b], Ming Ye [b, c, d, *], Fan Zhang [e], Yong-zhen Pei [c], Guo-ping Tang [f], Ahmed Elshall [b], Asghar A. Moghaddam [a]

[a] *Department of Earth Sciences, University of Tabriz, Tabriz 5166616471, Iran*
[b] *Department of Scientific Computing, Florida State University, Tallahassee FL 32306, USA*
[c] *School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin 300387, China*
[d] *Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee FL 32306, USA*
[e] *Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China*
[f] *Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge TN 37831, USA*
Received 14 September, 2017; accepted 10 January, 2018
Available online ***

Abstract

This study used the marginal likelihood and Bayesian posterior model probability for evaluation of model complexity in order to avoid using over-complex models for numerical simulations. It focused on investigation of the impacts of prior parameter distributions (involved in calculating the marginal likelihood) on the evaluation of model complexity. We argue that prior parameter distributions should define the parameter space in which numerical simulations are made. New perspectives on the prior parameter distribution and posterior model probability were demonstrated in an example of groundwater solute transport modeling with four models, each simulating four column experiments. The models had different levels of complexity in terms of their model structures and numbers of calibrated parameters. The posterior model probability was evaluated for four cases with different prior parameter distributions. While the distributions substantially impacted model ranking, the model ranking in each case was reasonable for the specific circumstances in which numerical simulations were made. For evaluation of model complexity, it is thus necessary to determine the parameter spaces for modeling, which can be done by conducting numerical simulation and using engineering judgment based on understanding of the system being studied.

Keywords: Marginal likelihood; Posterior model probability; Advection-dispersion equation; Mobile-immobile model; Groundwater model

1. Introduction

While modelers have tended to avoid using over-complex models in groundwater and other fields of numerical modeling, identifying the appropriate level of model complexity is always challenging, as it involves many factors, such as complexity of the system, current understanding of system behaviors, quality and quantity of system observations, objectives of modeling projects, cost of model development, and mathematical/statistical analysis needed for identification (Clement, 2011; Hunt and Zheng, 1999; Simmons and Hunt, 2012). Model complexity can be explored in various ways by investigating the relation between model inputs and outputs (Jakeman and Hornberger, 1993; Hill, 2006; Hunt et al., 2007; Gómez-Hernández, 2006), evaluating variability of model outputs (Young et al., 1996; Arkesteijn and Pande, 2013), and/or examining model predictive performance (Brooks and Tobias, 1996; Schoups et al., 2008; Kumar, 2011). A common practice in groundwater modeling is to use model selection criteria, e.g., the Akaike information criterion (*AIC*) and Bayesian information criterion (*BIC*), to select the best model from a set of models of different levels of complexity (Dai et al., 2012; Elshall and Tsai, 2014; Engelhardt et al., 2014; Massoudieh et al., 2013; Ye et al., 2008a, 2016); the selected model is considered to have the appropriate level of model complexity. All the model selection criteria favor models with high goodness-of-fit to data, and disfavor the models with unnecessary complexity by including penalty terms to create a balance between the model's data-fitting ability and its complexity. Many of the penalty terms are a function of the number of model parameters. However, evaluating model complexity by only examining the number of calibrated parameters may inadequately quantify model complexity, because model complexity may be affected by other factors, such as parameter regions and model structures.

This work was supported by the U.S. Department of Energy Early Career Research Program Award (Grant No. DE-Sc0008272) and U.S. National Science Foundation (Grant No. 1552329).
* Corresponding author.
Email address: mye@fsu.edu (Ming Ye).