A Bayesian Network Approach of Uncertainty Quantification for Groundwater Reactive Transport Modeling

Heng Dai[1] and Ming Ye[2]

[1]hd09@my.fsu.edu, [2]mye@fsu.edu, Department of Scientific Computing, Florida State University, Tallahassee, FL 32306

1. Introduction

➢ Bayesian networks (BNs), also known as belief networks, belong to the family of probabilistic graphical models (GMs). A Bayesian network consists of a graphical structure and a probabilistic description of the relationships among different variables of the system analyzed. The graphical structure explicitly represents cause and effect relationships that allow a complex causal chain linking actions to be factored into an articulated series of conditional relationships. Due to these characteristics, Bayesian networks can be particularly useful for uncertainty quantification of complex groundwater reactive transport models with multiple components related by different dependencies.

➢ A Bayesian network approach for quantifying the uncertainty of a groundwater reactive transport model is presented in this research. The uncertainty of future climate (i.e. precipitation variability and flooding) and different anthropogenic pressure such as land use are described as uncertain nodes in the Bayesian network. All the variables are characterized by multiple states, representing their uncertainty, in the form of continuous or discrete probability distributions that are propagated to the model endpoint which is the spatial distribution of certain chemical concentrations.

➢ After building the Bayesian network, uncertainty quantification is done through the probabilistic inferences which can obtain the posterior probability distributions over variables of interest. Most exact inference methods are not suitable for this study since the developed hybrid Bayesian network contains continuous nodes and their state values are described by partial differential equations. An approximate inference method: Monte Carlo (MC) is used to solve this problem through sampling. This Bayesian network can be widely applied into most contaminated sites with different nodes states and probability distributions. The uncertainty quantification results can be useful for the environmental managers and decision makers to formulate policies and strategies.

2. Bayesian Network Approach

2.1 Background

➢ A **directed graph** is defined as a pair (V, E), where V is a finite, nonempty set whose elements are called nodes (or vertices), and E is a set of ordered pairs of distinct elements of V. Elements of E are called edges (or arcs). If there is no path existed to start at some vertex v and follow a sequence of edges that eventually loops back to v itself, this directed graph is called a Directed Acyclic Graph (DAG).

➢ Considering n random variables $X_1,\ldots, X_n$, a directed acyclic graph with n numbered nodes, and suppose node j $(1 \le j \le n)$ of the graph is associated to the X_j variable. Then the graph is a **Bayesian network**, representing the variables $X_1,\ldots, X_n$, if:

$$P(X_1,\ldots,X_n)=\prod_{j=1}^{n} P(X_j|parents(X_j)) \qquad (1)$$

where $parents(X_j)$ denotes the set of all variables X_i, such that there is an arc from node i to node j in the graph.

2.2 Structure

In this research, a Bayesian network is constructed to quantify the uncertainties of a general groundwater reactive transport model. Figure 1 shows the Bayesian network structure.

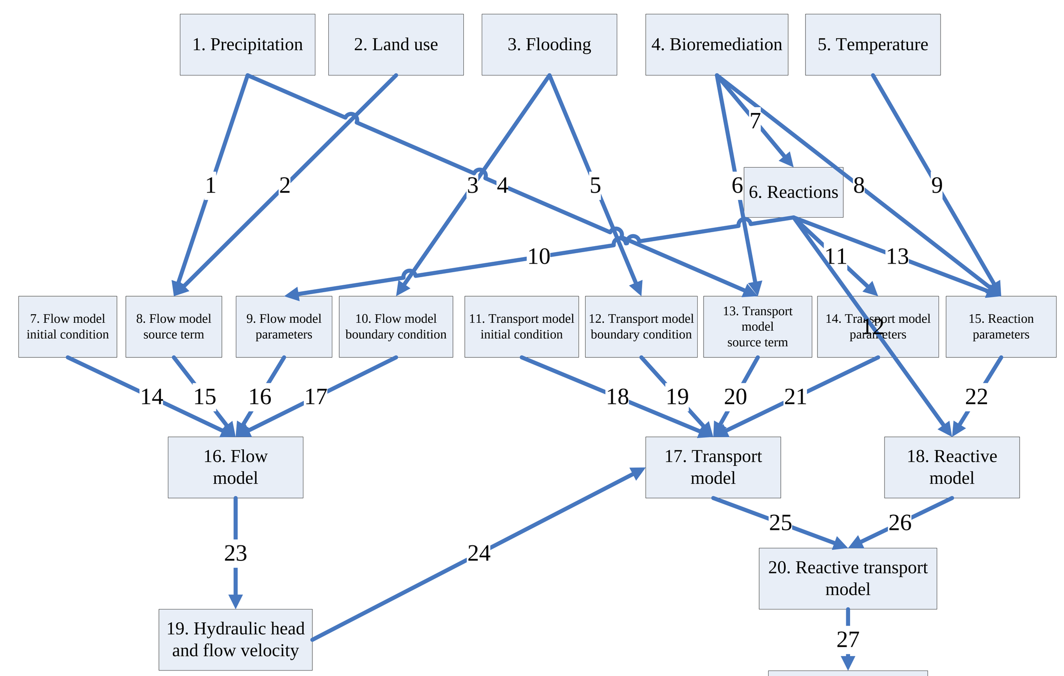

Figure 1. Bayesian network structure

As shown in Figure 1, each node in the graph represents a uncertain variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables.

2.3 Application

➢ For the purposes of testing and demonstration, the Bayesian network structure described above was simplified and applied into a test case. The land use, bioremediation and temperature nodes were deleted. Their related edges were also omitted.

➢ A simple synthetic one dimensional groundwater reactive transport case was built. The synthetic domain is illustrated in Figure 2. One set of single direction chemical reactions including five reactants were considered in this test case. The reactions are shown in Figure 3. Governing equations for this reactive transport system can be expressed as:

$$R_1\frac{\partial c_1}{\partial t}=D_x\frac{\partial^2 c_1}{\partial x^2}+D_y\frac{\partial^2 c_1}{\partial y^2}+D_z\frac{\partial^2 c_1}{\partial z^2}-v_s\frac{\partial c_1}{\partial x}-k_1c_1 \qquad (2)$$

$$R_i\frac{\partial c_i}{\partial t}=D_x\frac{\partial^2 c_i}{\partial x^2}+D_y\frac{\partial^2 c_i}{\partial y^2}+D_z\frac{\partial^2 c_i}{\partial z^2}-v_s\frac{\partial c_i}{\partial x}+y_{i-1}k_{i-1}c_{i-1}-k_ic_i,\ \text{for } i=2,3,\ldots,n \qquad (3)$$

where c_i is the concentration of certain reactant [mg/L]; D_x, D_y and D_z are the hydrodynamic dispersion coefficients [ft^2/yr]; v_s is the groundwater seepage velocity [ft/yr]; k_i is the first-order degradation coefficient [1/yr]; y_i is the yield coefficient; and R_i is the retardation factor; n is the reactant number which is five in this case.

➢ Although all the nodes in the network can have multi-states and corresponding probabilities. The test case only consider one unique flow and transport model. This means the nodes 16, 17, 18 and 20 only have one state.

➢ Uncertain nodes were described through both continuous and discrete probability distributions. For example, the precipitation values were assumed to follow normal distribution. Flooding events is assumed to happen with a 0.2 probability.

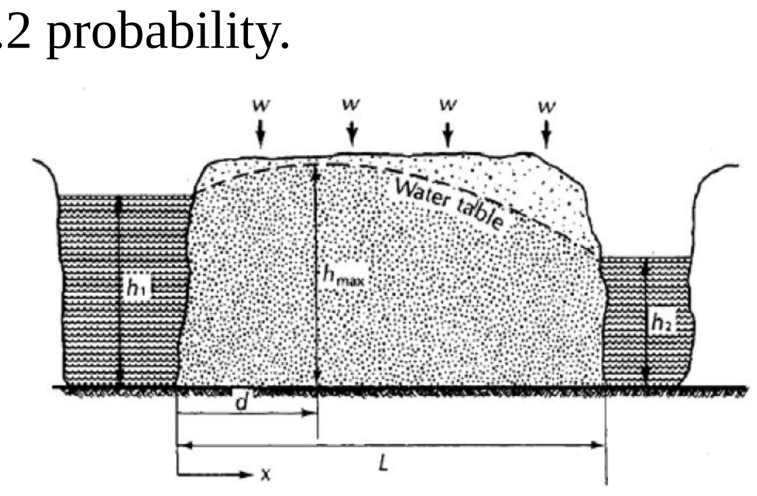

Figure 2. The synthetic domain illustration.

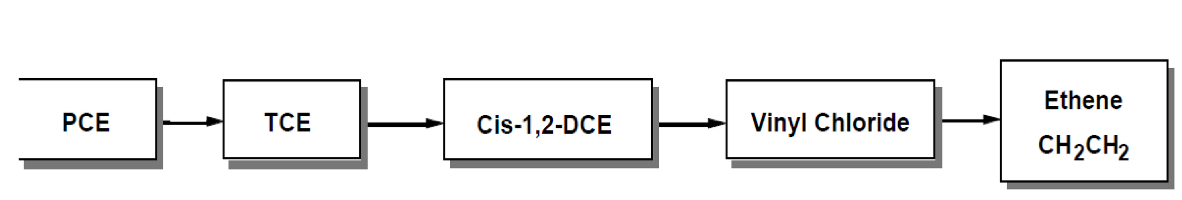

Figure 3. Chemical reactions involved in the test case.
PCE: Perchloroethylene, TCE: Trichloroethene, DCE: Dichloroethene.

3. Uncertainty Quantification

➢ The uncertainty quantification was implemented through the Bayesian network inference which represents the process of computing the posterior distribution of variables given evidence.

➢ Direct sampling Monte Carlo method was used in our Bayesian network inference. Ten thousands simulations were run and the histograms of hydraulic head and PCE concentration at location x = 1000 meters and 5000 meters are shown in Figure 4 and 5.

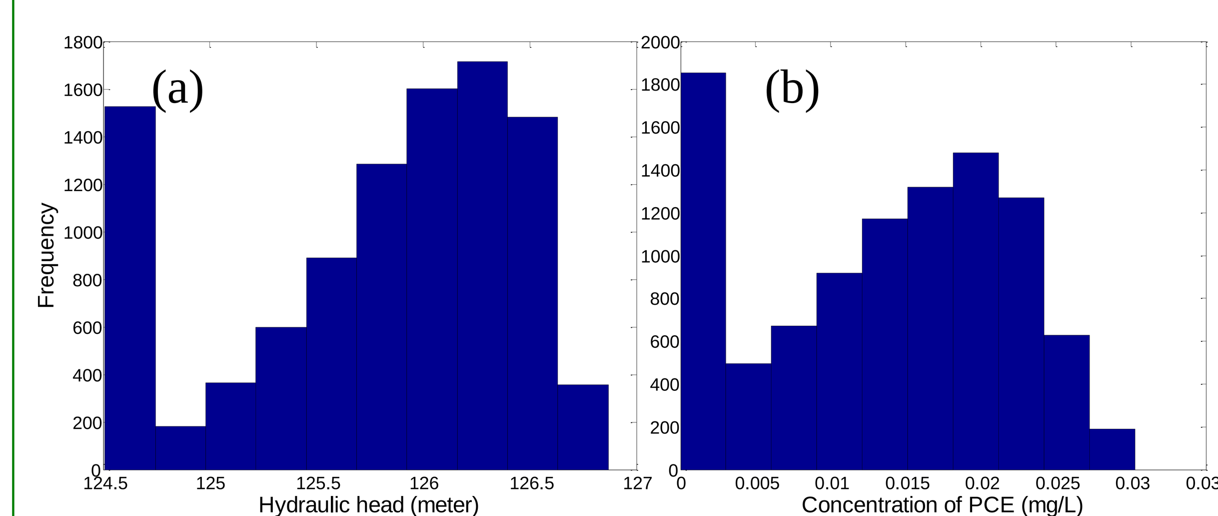

Figure 4. Hydraulic head (a) and concentration of PCE (b) distributions at location x = 1000 meters .

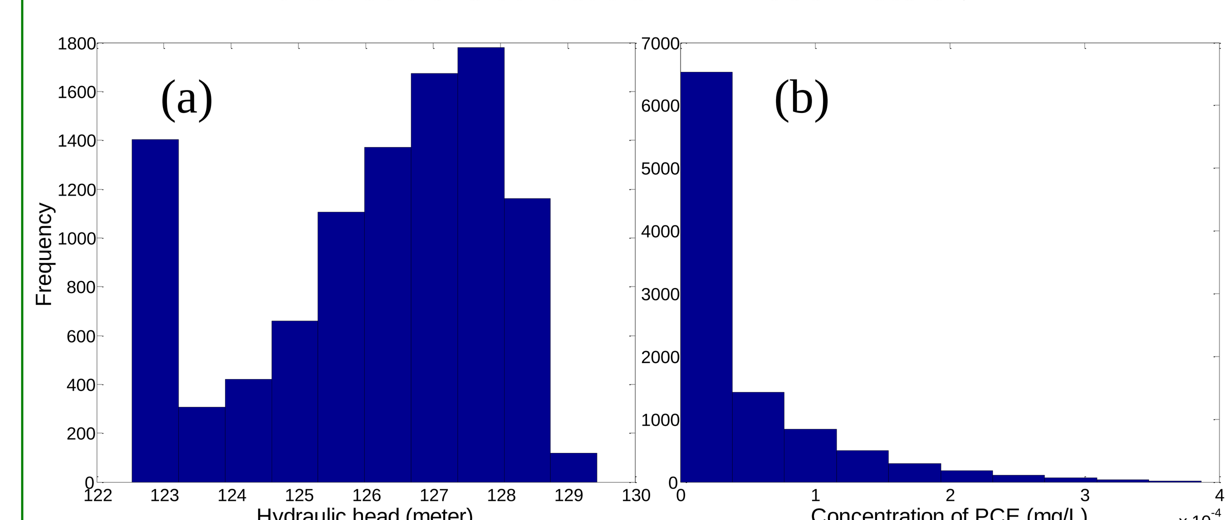

Figure 5. Hydraulic head (a) and concentration of PCE (b) distributions at location x = 5000 meters .

The histograms reveal that the distribution patterns are similar for hydraulic head but totally different for concentration of PCE at different locations.

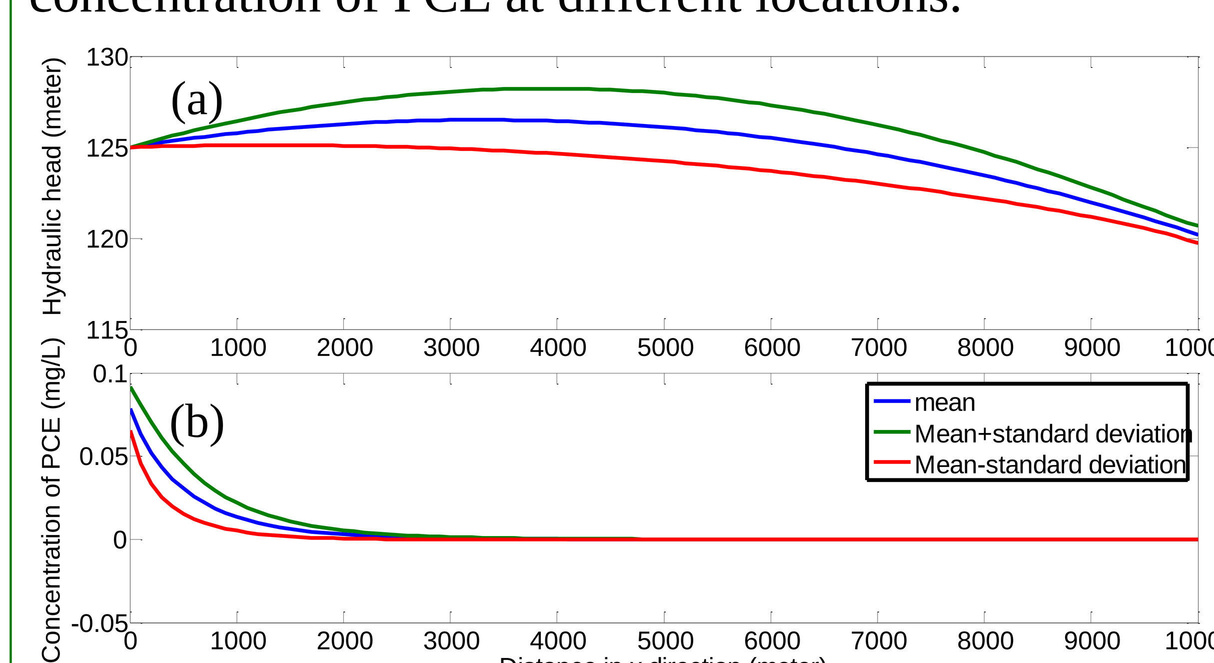

Figure 6. The mean and standard deviation of hydraulic head (a) and PCE concentration (b) plotted simultaneously for the whole domain.

The results show that the mean and standard deviation values of the two variables vary at the location. Larger uncertainties of the hydraulic head and the PCE concentration coincide with their larger mean values. This phenomenon fits with what Figure 5 shows.

4. Conclusions and Discussions

➢ The Bayesian network was built with considering uncertainties in the groundwater reactive transport model.

➢ The model results in form of probability distributions are useful to environmental protection and management.

➢ The Bayesian network has not been calibrated against existing data, i.e., evidence.

➢ Although the application described in this poster is based on a simple synthetic case, this Bayesian network can be applied into most general realistic cases.

Acknowledgement

The research support is provided by the DOE Early Career Award DE-SC0008272.