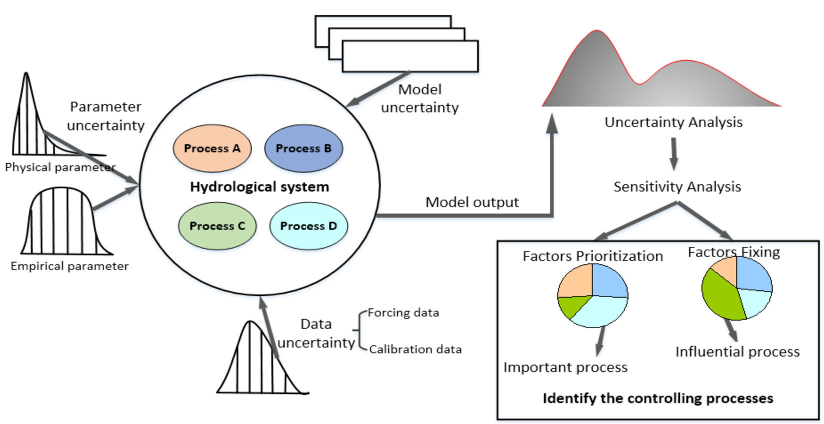
中国地质大学
CHINA UNIVERSITY OF GEOSCIENCES

博士学位论文答辩

地下水模型控制过程识别的多模型全局敏感性分析方法及应用

中国地质大学 China University Of Geosciences
环境学院 school of environmental studies



姓　　名：杨　静
专　　业：水利工程
指导教师：唐仲华　教授
合作导师：叶　明　教授

答辩日期：11/16/2021

Table of Contents 目录

Table of Contents 目录

1. Introduction

1.1 Research Background 研究背景

论文题目:

Two questions: 水文模型控制过程识别的多模型全局敏感性分析方法及应用

Q1. Why brother to identify the controlling processes?

为什么要识别控制过程?(什么是控制?)

Q2. Why multiple models should be considered?

为什么要考虑多模型? (什么是多模型?)

1. Introduction

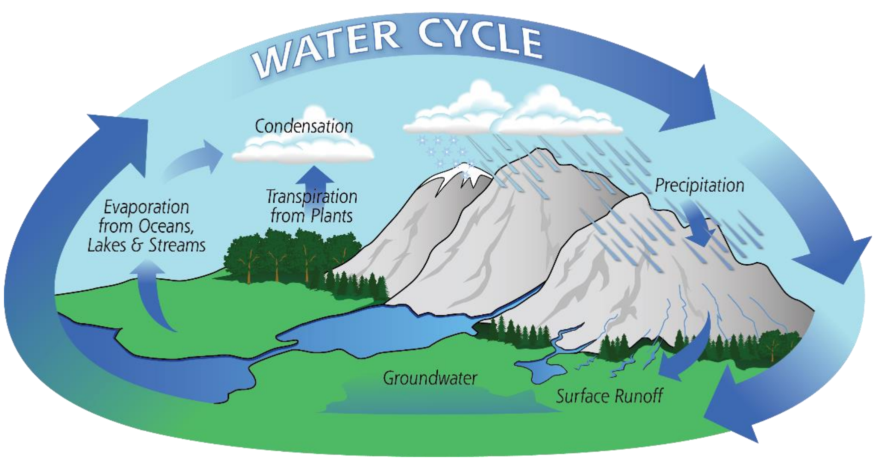
1.1 Research Background 研究背景 第一个问题

Q1. Why bother to identify the controlling processes? 控制过程?

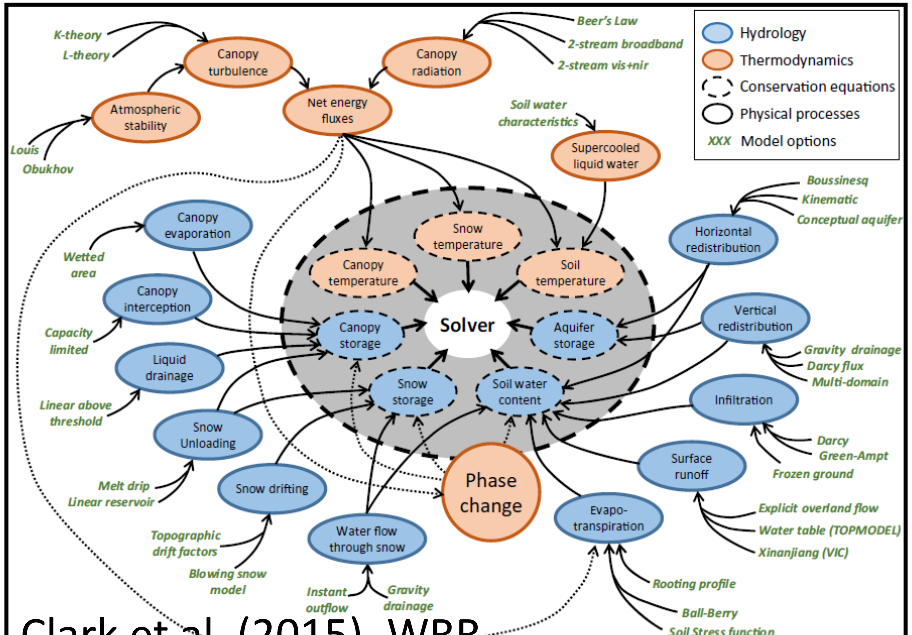
"…for many phenomena 80% of the output or consequences are produced by 20% of the input or causes."

80% 的影响来自于 20% 的投入.

—The Pareto Principle



https://gpm.nasa.gov/education/interactive/anime-webquest-water-cycle

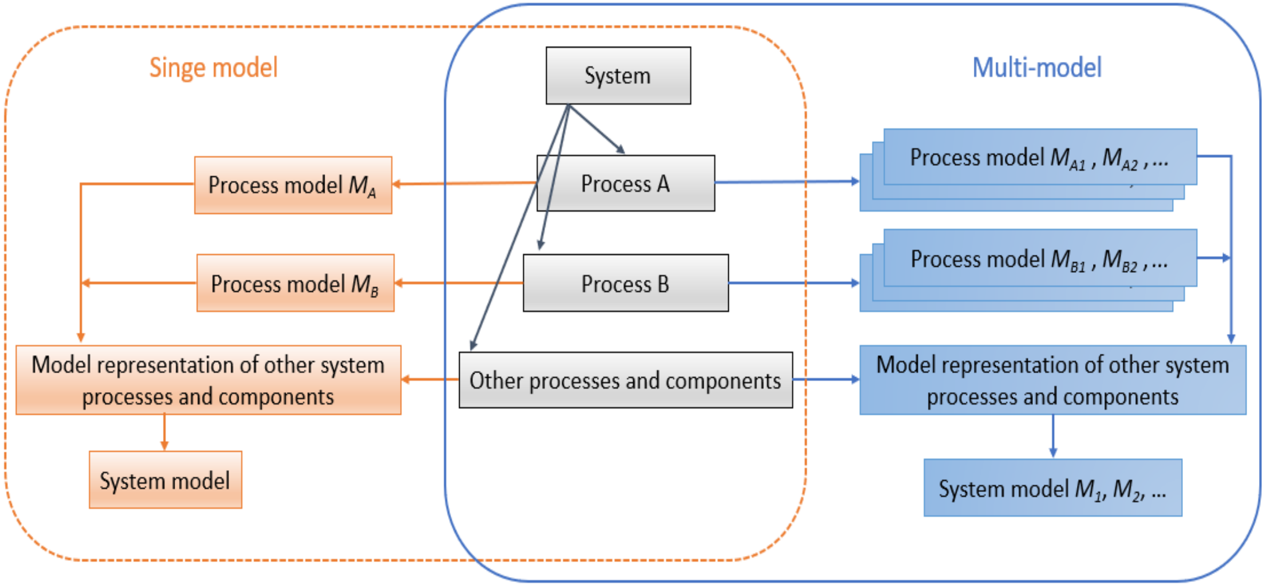


Clark et al. (2015), WRR

1. Introduction

1.1 Research Background 研究背景　　第二个问题

Q2. Why multiple models should be considered? 多模型?



系统与过程之间的关系和过程模型与系统模型之间的关系

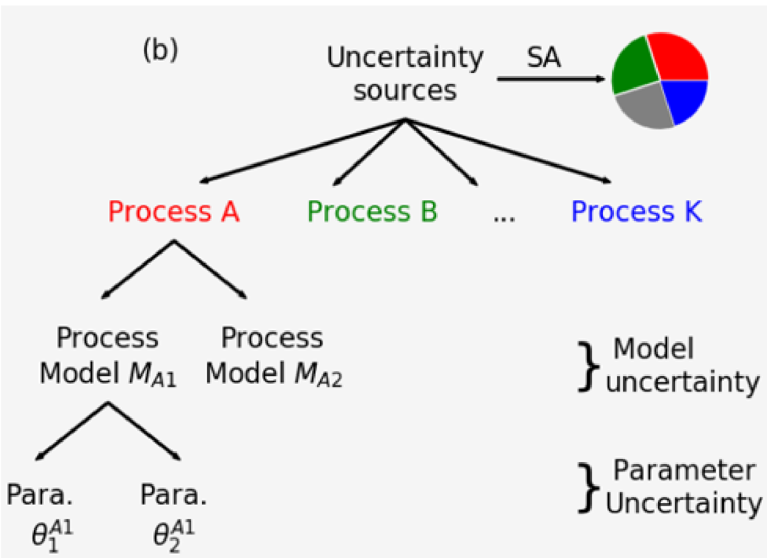
传统单模型：一个过程对应一个过程模型

多模型：一个过程对应多个过程模型

1. Introduction

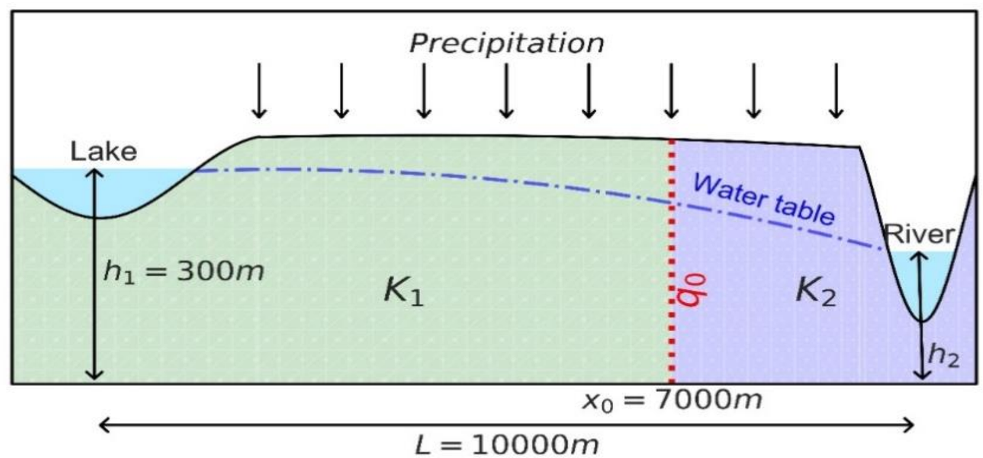
1.1 Research Background 研究背景 第二个问题

Q2. Why multiple models should be considered? 多模型?



qoi: x_0=7000 m处流量

降雨入渗（R_1, R_2）
渗透系数均值/非均质(G_1, G_2)
右侧河流水位（M_1, M_2）



1. Introduction

1.1 Research Background 研究背景

Q2. Why multiple models should be considered? 必要性2

$S_A - S_D$ denote fictitious values of "controlling index" for processes $A - D$ of a system

	S_R	S_G	S_M
Model $R_1G_1M_1$	50%	30%	20%
Model $R_2G_2M_2$	30%	20%	50%

假设已经找到一个控制指标, S, 可以刻画模型的控制程度.
由于多模型的存在，考虑这个系统可以由两个模型刻画。

Which one is really of controlling?

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

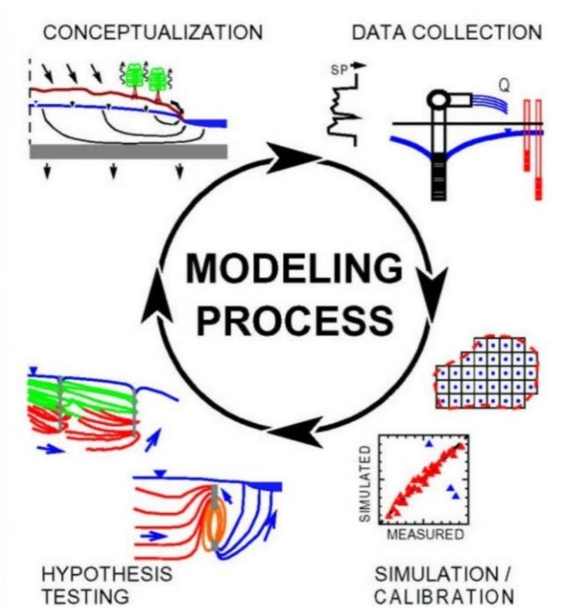
Q. Can we measure the degree of the controlling for a process?

给定一个过程，是否能测定它的控制程度?

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.1 Uncertainty in hydrology modeling 水文模型不确定性



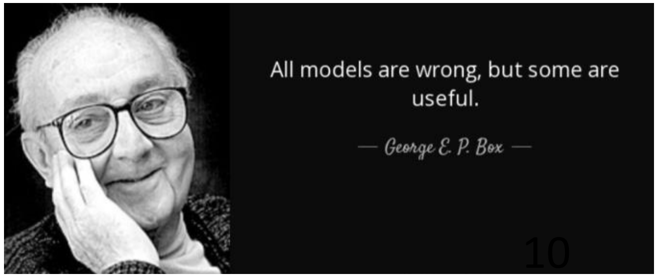
三大类不确定

1. Data 数据
2. Parameter 参数
3. Model 模型

http://inside.mines.edu/~epoeter/ **2006 Darcy Lecture**

"All models are approximations. Essentially,
All models are wrong, but some are useful".

—George E. P. Box



1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.1 Uncertainty in hydrology modeling

Uncertainty serves as the cornerstone that why and how the hydrology community needs to identify the controlling processes.模型不确定性是确定水文模型控制过程的基石.

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment

The relationship between Uncertainty Analysis (UA) and Sensitivity Analysis (SA)?不确定性分析和敏感性分析的关系与区别

Why sensitivity analysis?

Generally, **UA comes before SA and SA should precede UA**: uncertainty needs to be first estimated and before it can be apportioned.

UA -- “How uncertain is my model output?”. 模型输出有多不确定
SA -- “this factor alone is responsible for 80% of the uncertainty in the output”. 某个因子贡献了80%的不确定

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment

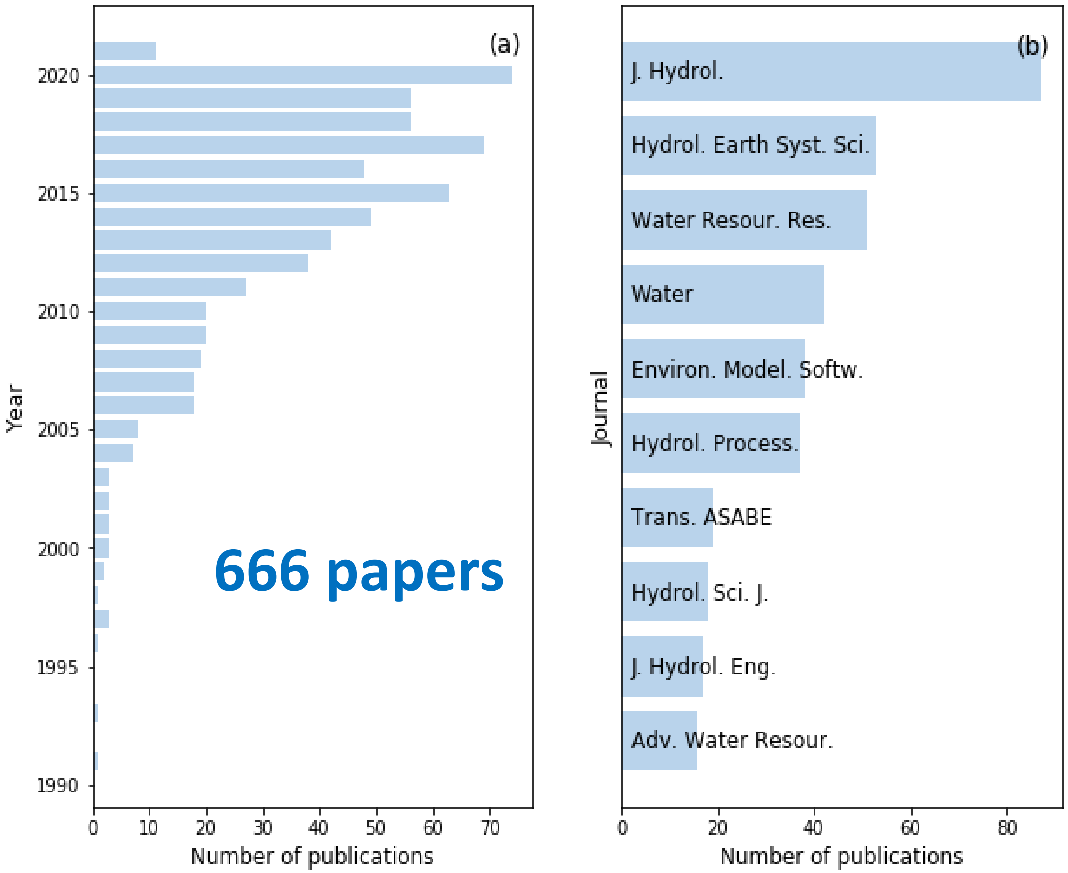


Figure 1-2 (a) Yearly publications of peer-reviewed journal articles (i.e., the publication type should be "J") on sensitivity analysis in the field of hydrologic modeling from the Web of Science Core Collection and (b) the top 10 publishing journals. The total number of records is 666 based on the search terms: "sensitivity analysis" AND ("hydrologic" OR "hydrological") AND ("model" OR "modeling") AND "uncertainty" on May 26, 2021.

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment

Classifications	LSA 局部敏感性分析	GSA全局敏感性分析
Descriptions	$\boxed{y = f(x), x \sim U[0,1]}$ Compute **local response** of model output	Evaluate the effect in the **entire ranges** of uncertain parameters
Characteristics	Easy of operation and interpret, relatively low computational cost	Estimating the effect of all the inputs or their combined effect on the variation of output based on numerous model runs
Typical methods	Finite difference method; First-order second moment (FOSM)	Regression method; Morris' method (Morris 1991);Sobol's method (Sobol' 2001); PAWN index; AMA indices; Delta moment-independent measure
Software	C++ based PSUADE software (Gan et al. 2014a); Sensitivity package for R; SAFE toolbox for MATLAB (Pianosi et al. 2015); SALib package for Python (Herman and Usher 2017).	

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment

An open question in SA is how the uncertainty of the model output can be apportioned. **怎么分配?**

Within the framework of GSA, Saltelli et al. (2004) argued that this must be linked to the nature of the question(s) the SA is intended to address. They introduced four "settings", i.e.,

(1) Factors Prioritization Setting 因子排序,
(2) Factors Fixing Setting 因子固定,
(3) Variance Cutting Setting, 方差阈值
(4) Factors Mapping Setting. 因子映射

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment 前人研究

(1). Factors Prioritization Setting 因子排序: The objective of this setting is to identify the most important factor 重要因子, which is defined as the one that, if determined would result in the greatest reduction in the uncertainty of the system model output. 一经确定，模型不确定性减少最多

(2). Factors Fixing Setting 因子固定: The objective of this setting, which could also be labelled "screening", is to identify the non-influential factor无影响因子 or the subset of input factors that we can fix at any given value within their range of uncertainty without significantly reducing the uncertainty in output. 可随意处置，对模型不确定性减少最少

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.2 Sensitivity Analysis for Uncertainty Apportionment 前人研究

A simple example:

$$y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_k). \quad f: \mathbb{R}^k \to \mathbb{R}$$

Factors Prioritization 因子排序

统计指标（方差 差分）衡量不确定性 ← $\zeta\left(P_Y, P_{Y|x_i^*}\right)$

找到这样的一个参数，当这个参数固定在参数真值的时候（假设我们知道），模型输出的不确定性减少最多 ----重要参数

Factors Fixing 因子固定

统计指标（方差 差分）衡量不确定性 ← $\zeta\left(P_Y, P_{Y|x_{\sim i}^*}\right)$

找到这样的一个参数，当其他参数固定在其真值，而这个参数在其范围任意变化时，模型不确定性基本不变 --- 无影响参数

where P_Y and $P_{Y|x\sim i*}$ are unconditional and conditional distributions (PDF or CDF) of the model output.

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.3 Sensitivity Analysis for Uncertainty Apportionment 本文研究

Factors Prioritization – Important processes 重要过程

We want to make a rational bet on what is the process that one should fix to achieve the greatest reduction in the uncertainty of the output. In line with this, controlling processes are those control reductions of predictive uncertainty. **找到一个过程，此过程一经确定，模型输出不确定减少最多**

Factors Fixing – Influential processes 影响过程

We try to screen the input processes by identifying process or sets of processes that non-influential. In this sense, controlling processes are those control how much uncertainty caused by themselves and their interactions with other processes. **找到一个过程，此过程无法确定，模型输出不确定最多**

1. General Background

1.2 Literature Review and Problems 研究现状与存在问题

1.2.3 Sensitivity Analysis for Uncertainty Apportionment

水文模型控制过程识别的多模型全局敏感性分析方法及应用

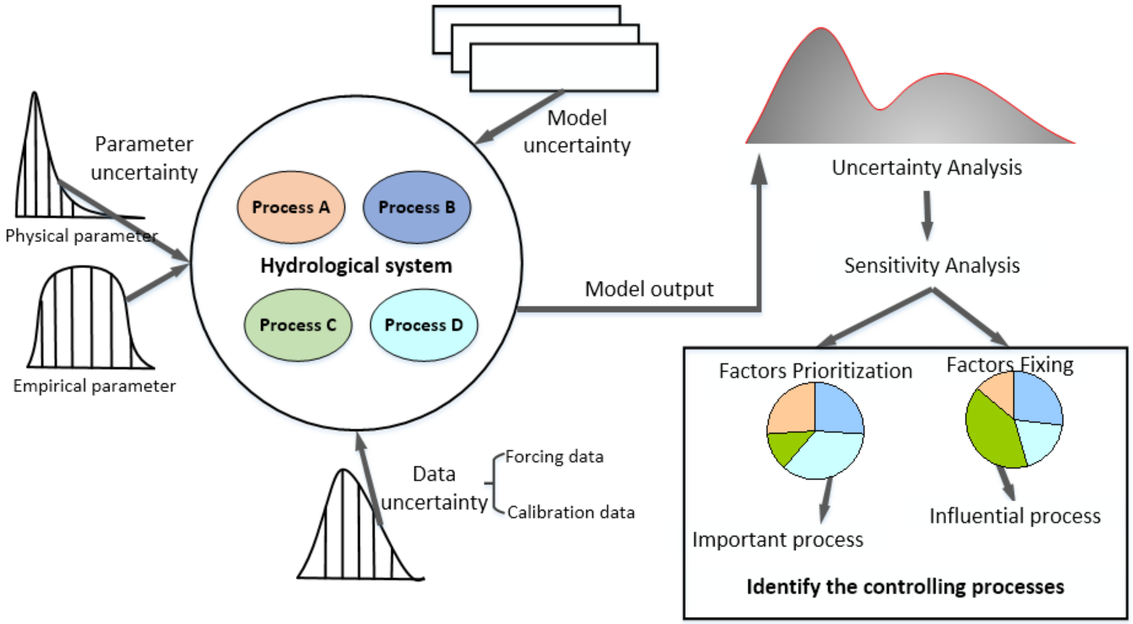


Figure 1-3. A big picture depicting the uncertainties in hydrologic modeling, uncertainty analysis, sensitivity analysis, and identifying the controlling process via sensitivity analysis.

1. Introduction

1.2 Literature Review and Problems 研究现状与存在问题

1.2.4 Main Problems 主要问题

- The existing methods ignore the inherent uncertainty in conceptualizing and modeling individual processes. 多忽略模型不确定性
- The existing methods are lack of a summary measure that explicitly quantifies the relative importance of individual processes. 缺乏同时考虑模型不确定性和参数不确定性方法
- The existing methods cannot be directly used for large-scale problems in practice, because the methods are computationally expensive due to their relying on Monte Carlo implementations. 计算量大，没有工具

1. Introduction

1.3 Research Objectives, Procedure, and Innovations

1.3.1 Research Objectives

The overarching scientific question to be answered in this project is as follows: 本文科学问题

If we are not certain about the choice of process models and model parameters, how can we correctly identify the controlling processes of a complex system?

如果水文模型过程模型和模型参数都不确定，
怎样正确识别水文模型的控制过程？

1. General Background

1.3 Research Objectives, Procedure, and Innovations

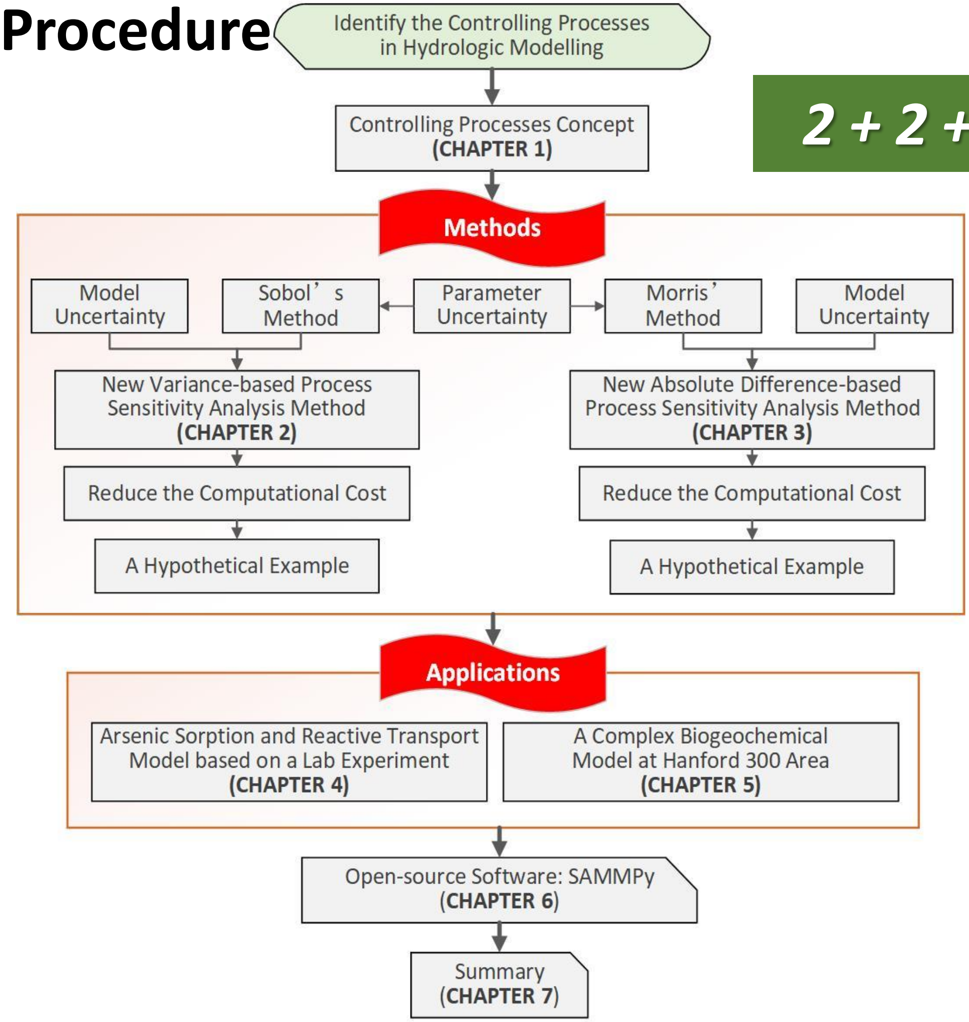
1.3.1 Research Objectives

- Introduce the concept of multiple working hypotheses and apply the new developed methods to real-word hydrologic models. 引入新概念，多模型
- Develop a set of new process sensitivity indices as summary measures to explicitly quantify relative importance and influence of individual processes. 提出新方法，控制过程
- Create new software or a package to put the ethos into action. 开发新软件，工具

1. Introduction

1.3 Research Objectives, Procedure, and Innovations

1.3.2 Research Procedure



2 + 2 + 1

Table of Contents 目录

2. New VBSA Method

2.1 Methodology

2.1.1 Traditional Sobol's Parameter Sensitivity Indices for a Single Model 前人研究：用方差衡量不确定性 --- 因子排序

Factors Prioritization

Sobol first-order sensitivity index一阶敏感性指数

$$S_i = E_{X_i}\left(\frac{V(Y) - V_{X_{\sim i}}(Y \mid X_i = x_i)}{V(Y)}\right)$$

重要参数
越大越重要

$$= \frac{V(Y) - E_{X_i}\left(V_{X_{\sim i}}[Y \mid X_i = x]\right)}{V(Y)} = \frac{V_{X_i}\left(E_{X_{\sim i}}[Y \mid X_i = x]\right)}{V(Y)}$$

因子排序
找到这样的一个参数，当这个参数固定在参数真值的时候（假设我们知道），模型输出的不确定性**减少最多** ----重要参数

单模型，仅考虑参数不确定

$$\zeta\left(P_Y, P_{Y|x_i^*}\right) \longrightarrow V(Y) - \boxed{V(Y \mid x_i = x_i^*)}$$

条件方差

2. New VBSA Method

2.1 Methodology

2.1.1 Traditional Sobol's Parameter Sensitivity Indices for a Single Model 前人研究：用方差衡量不确定性 --- 因子固定

有影响参数
越小影响程度就越小

Factors Prioritization

Sobol total-effect sensitivity index总效应敏感性指数

$$S_{Ti} = E_{X_{\sim i}}\left(\frac{V_{X_i}(Y\,|\,X_{\sim i})}{V(Y)}\right) = \frac{E_{X_{\sim i}}V_{X_i}(Y\,|\,X_{\sim i})}{V(Y)} = 1 - \frac{V_{X_{\sim i}}E_{X_i}(Y\,|\,X_{\sim i})}{V(Y)}$$

单模型，仅考虑参数不确定

因子固定

找到这样的一个参数，当其他参数(除了这个参数)固定在其真值，而这个参数在其范围任意变化时，模型**不确定性基本不变** --- 无影响参数

$$\zeta\left(P_Y, P_{Y|x_{\sim i}^*}\right) \longrightarrow V(Y\,|\,x_{\sim i} = x_{\sim i}^*)$$

2. New VBSA Method

2.1 Methodology

2.1.2 New Variance-based Process Sensitivity Indices for Multi-model

Factors Prioritization

因子排序：重要过程

First-order process sensitivity index by Dai et al. (2017, WRR) 一阶过程敏感性指数

$$PS_K = \frac{V_{\mathbf{M}_K}\left(E_{\mathbf{M}_{\sim K}}[\Delta \mid M_K]\right)}{V(\Delta)}$$

找到一个过程，此过程一经确定(过程模型固定在真值，过程模型涉及的参数固定在真值），模型输出不确定减少最多

Factors Fixing:

因子固定：影响过程

Total-effect process sensitivity index by Yang et al. (2021, WRR) 总效应过程敏感性指数

找到一个过程，此过程无法确定(过程模型无法确定，过程模型参数亦无法确定)，模型输出不确定最多

$$PS_{TK} = \frac{E_{\mathbf{M}_{\sim K}}\left(V_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]\right)}{V(\Delta)} = 1 - \frac{V_{\mathbf{M}_{\sim K}}\left(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]\right)}{V(\Delta)}$$

2. New VBSA Method

2.1 Methodology

2.1.2 New Variance-based Process Sensitivity Indices for Multi-model

Total-effect process sensitivity index

$$PS_{TK} = \frac{E_{\mathbf{M}_{\sim K}}(V_{\mathbf{M}_K}[\Delta \mid M_{\sim K}])}{V(\Delta)} = 1 - \frac{V_{\mathbf{M}_{\sim K}}(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}])}{V(\Delta)}$$

Expending it to include parameter uncertainty 推导扩展 $\boxed{V(\Delta) = E(\Delta^2) - E^2(\Delta)}$

$$V_{\mathbf{M}_{\sim K}}(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]) = E_{\mathbf{M}_{\sim K}}(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}])^2 - (E_{\mathbf{M}_{\sim K}} E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}])^2$$

$$\begin{aligned} V_{\mathbf{M}_{\sim K}}(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]) = {} & E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}}(E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K \mid M_K}[\Delta \mid M_{\sim K}, \theta_{\sim K}, \theta_K, M_K])^2 \\ & - (E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K \mid M_K}[\Delta \mid M_{\sim K}, \theta_{\sim K}, \theta_K, M_K])^2 \end{aligned}$$

2. New VBSA Method

2.2 Numerical Evaluation

Total-effect process sensitivity index 总效应过程敏感性指数

$$V_{\mathbf{M}_{\sim K}}(E_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]) = E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}} (E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid M_{\sim K}, \theta_{\sim K}, \theta_K, M_K])^2 - (E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}} E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid M_{\sim K}, \theta_{\sim K}, \theta_K, M_K])^2$$

- Expectation with model uncertainty – Model Averaging

$$E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}(\bullet) = \sum_{\mathbf{M}_{\sim K}} E_{\sim\boldsymbol{\theta}_K|\sim M_K}(\bullet) P(M_{\sim K})$$

模型求期望--模型平均法

- Expectation with parameter uncertainty – Monte Carlo

$$E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}} \quad E_{\boldsymbol{\theta}_K|M_{\sim K}}$$

参数求期望—蒙特卡洛方法

2. New VBSA Method

2.2 Numerical Evaluation

伪代码

Loop [1] over model combinations $M_{\sim K}$ for process ~K in $\boldsymbol{M}_{\sim K}$

Loop [2] over parameter realizations $\theta_{\sim K}$ of model $M_{\sim K}$ in $\boldsymbol{\theta}_{\sim K}$

Loop [3] over process models M_K for process K in $\boldsymbol{M}_K$

Loop [4] over parameter realizations θ_K of model M_K in $\boldsymbol{\theta}_K$

Compute $\Delta \mid M_{\sim K}, \theta_{\sim K}, M_K, \theta_K$

End Loop [4]

Compute $E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K]$

End Loop [3]

Compute $E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K]$ and

$(E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K])^2$ using model averaging

End Loop [2]

Compute $E_{\sim\boldsymbol{\theta}_K|\sim M_K} E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K]$ and

$E_{\sim\boldsymbol{\theta}_K|\sim M_K}(E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K])^2$

End Loop [1]

Compute $(E_{\sim\boldsymbol{M}_K} E_{\sim\boldsymbol{\theta}_K|\sim M_K} E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K])^2$ and

$E_{\sim\boldsymbol{M}_K} E_{\sim\boldsymbol{\theta}_K|\sim M_K}(E_{\boldsymbol{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \sim M_K, \sim \theta_K, \theta_K, M_K])^2$ using model averaging

To evaluate the total-effect process sensitivity index (as well as the first-order process sensitivity index proposed by Dai et al. (2017b)) for each process needs four loops, the corresponding total number of model executions is $(N_{\sim K} \times n_{\sim K}) \times (N_K \times n_K)$. The $n_{\sim K} \times n_K$ would be computationally demanding since it relies Monte Carlo simulations.

$O(n^2)$

n: Monte Carlo 采样次数

2. New VBSA Method

2.3 Efficient Design and Estimator to Reduce the Computational Cost

高效算法

$$\Delta = f(R, G, S)$$

$$R_1 = M_{R1}(\theta_{R1}) \qquad G_1 = M_{G1}(\theta_{G1}) \qquad S_1 = M_{S1}(\theta_{S1})$$

$$R_2 = M_{R2}(\theta_{R2}) \qquad G_2 = M_{G2}(\theta_{G2}^1, \theta_{G2}^2) \qquad S_2 = M_{S2}(\theta_{S2}^1, \theta_{S2}^2)$$

2. New VBSA Method

2.3 Efficient Design and Estimator to Reduce the Computational Cost

$$\mathbf{A}=\begin{bmatrix} \theta_{R1}^{(1)} & \theta_{G1}^{(1)} & \theta_{S2}^{1\,(1)} & \theta_{S2}^{2\,(1)} \\ \theta_{R1}^{(2)} & \theta_{G1}^{(2)} & \theta_{S2}^{1\,(2)} & \theta_{S2}^{2\,(2)} \\ \cdots & \cdots & \cdots & \cdots \\ \theta_{R1}^{(n)} & \theta_{G1}^{(n)} & \theta_{S2}^{1\,(n)} & \theta_{S2}^{2\,(n)} \end{bmatrix} \quad \mathbf{C}_1=\begin{bmatrix} \theta_{R1}^{\prime(1)} & \theta_{G1}^{(1)} & \theta_{S2}^{1\,(1)} & \theta_{S2}^{2\,(1)} \\ \theta_{R1}^{\prime(2)} & \theta_{G1}^{(2)} & \theta_{S2}^{1\,(2)} & \theta_{S2}^{2\,(2)} \\ \cdots & \cdots & \cdots & \cdots \\ \theta_{R1}^{\prime(n)} & \theta_{G1}^{(n)} & \theta_{S2}^{1\,(n)} & \theta_{S2}^{2\,(n)} \end{bmatrix}$$

$$\mathbf{B}=\begin{bmatrix} \theta_{R1}^{\prime(1)} & \theta_{G1}^{\prime(1)} & \theta_{S2}^{1\,\prime(1)} & \theta_{S2}^{2\,\prime(1)} \\ \theta_{R1}^{\prime(2)} & \theta_{G1}^{\prime(2)} & \theta_{S2}^{1\,\prime(2)} & \theta_{S2}^{2\,\prime(2)} \\ \cdots & \cdots & \cdots & \cdots \\ \theta_{R1}^{\prime(n)} & \theta_{G1}^{\prime(n)} & \theta_{S2}^{1\,\prime(n)} & \theta_{S2}^{2\,\prime(n)} \end{bmatrix}$$

$$\mathbf{C}_2=[\ldots]$$

$$\mathbf{C}_3=[\ldots]$$

特别设计采样矩阵

$O(n)$

2. New VBSA Method

2.3 Efficient Design and Estimator to Reduce the Computational Cost

$$S_i = \frac{V_{\boldsymbol{\theta}_i}(E_{\boldsymbol{\theta}_{\sim i}}[\Delta \mid \theta_i])}{V(\Delta)} = 1 - \frac{(1/2n)\sum_{j=1}^{n}\left(\boldsymbol{\Delta}_B^{(j)} - \boldsymbol{\Delta}_{C_i}^{(j)}\right)^2}{(1/n)\sum_{j=1}^{n}\left(\boldsymbol{\Delta}_A^{(j)}\right)^2 - \overline{\Delta}_A^{2}}$$

单模型一阶敏感性

$$S_{Ti} = \frac{E_{\boldsymbol{\theta}_{\sim i}}(V_{\boldsymbol{\theta}_i}[\Delta \mid \theta_{\sim i}])}{V(\Delta)} = \frac{(1/2n)\sum_{j=1}^{n}\left(\boldsymbol{\Delta}_A^{(j)} - \boldsymbol{\Delta}_{C_i}^{(j)}\right)^2}{(1/n)\sum_{j=1}^{n}\left(\boldsymbol{\Delta}_A^{(j)}\right)^2 - \overline{\Delta}_A^{2}}$$

单模型总效应敏感性

2. New VBSA Method

2.3 Efficient Design and Estimator to Reduce the Computational Cost

$$E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K|M_K}(E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}[\Delta|\Delta|\theta_K, M_K, \theta_{\sim K}, M_{\sim K}])^2$$ 多模型一阶敏感性

$$= \sum_{p=1}^{m_k}\sum_{q=1}^{m_{\sim k}}\frac{1}{n}\sum_{j=1}^{n}\left[\Delta_B^{(j)}|M_{K_p}, M_{\sim K_q}\right]\left[\Delta_{C_i}^{(j)}|M_{K_p}, M_{\sim K_q}\right]P\left(M_{\sim K_q}\right)^2 P\left(M_{K_p}\right)$$

$$+\sum_{p=1}^{m_k}\sum_{q=1}^{m_{\sim k}}\sum_{\substack{r=1\\ r\neq q}}^{m_{\sim k}}\frac{1}{n}\sum_{j=1}^{n}\left[\Delta_B^{(j)}|M_{K_p}, M_{\sim K_q}\right]\left[\Delta_{C_i}^{(j)}|M_{K_p}, M_{\sim K_r}\right]P\left(M_{\sim K_r}\right)P\left(M_{\sim K_q}\right)P\left(M_{K_p}\right)$$

$O(n)$

$$E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}(E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta|\theta_K, M_K, \theta_{\sim K}, M_{\sim K}])^2$$ 多模型总效应敏感性

$$= \sum_{q=1}^{m_{\sim K}}\sum_{p=1}^{m_K}\frac{1}{n}\sum_{j=1}^{n}\left[\Delta_A^{(j)}|M_{K_p}, M_{\sim K_q}\right]\left[\Delta_{C_i}^{(j)}|M_{K_p}, M_{\sim K_q}\right]P(M_{K_p})^2 P(M_{\sim K_q})$$

$$+\sum_{q=1}^{m_{\sim K}}\sum_{p=1}^{m_K}\sum_{\substack{r=1\\ r\neq p}}^{m_K}\frac{1}{n}\sum_{j=1}^{n}\left[\Delta_A^{(j)}|M_{K_p}, M_{\sim K_q}\right]\left[\Delta_{C_i}^{(j)}|M_{K_r}, M_{\sim K_q}\right]P(M_{K_r})P(M_{K_p})P(M_{\sim K_q})$$

2. New VBSA Method

2.4 A Hypothetical Example 河间地块模型

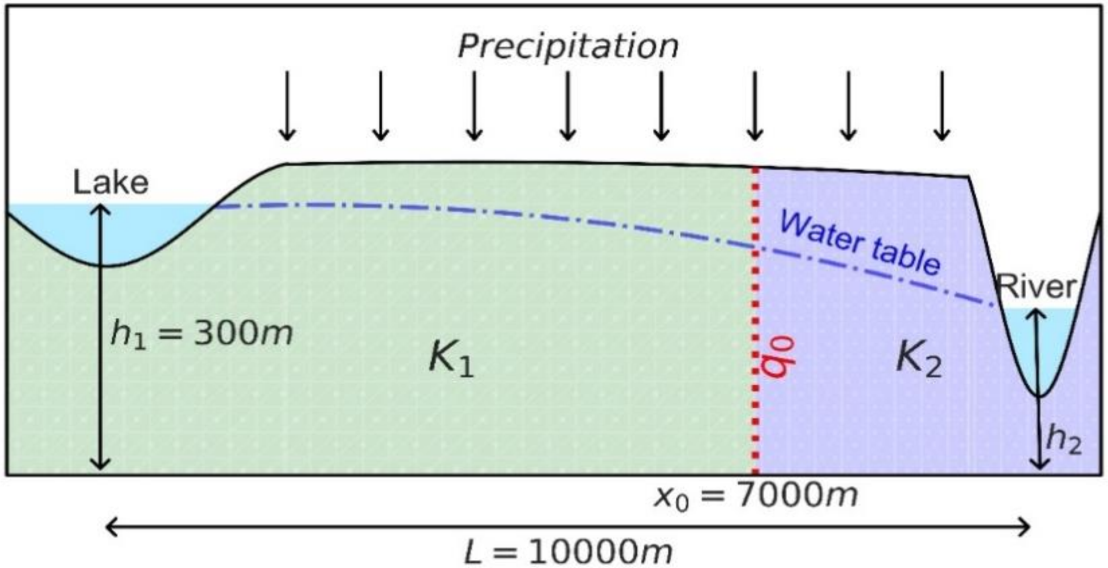


Figure 2-2 Sketch of the one-dimensional groundwater flow model. The aquifer is bounded with constant head boundary conditions. The left boundary is adjacent to a lake with a constant head of h1=300 m. The right boundary is adjacent to a river while no measuring data is available for the river stage. The divide of the two zones for the hydraulic conductivity at x0=7000 m is marked with dashed line and the groundwater discharge per unit width at this location (q_0) is our quantity of interest.

2. New VBSA Method

2.4 A Hypothetical Example

- Two recharge process models 降雨入深模型

$$R_1 : w = a(P-14)^{0.5} \times 0.0254/365$$
$$R_2 : w = b(P-15.7) \times 0.0254/365$$

线性/非线性入渗补给

where P = 60 inch/yr; a and b are scaling parameter, which follows $N(2.0, 0.4^2)$ and $U[0.2, 0.5]$, respectively.

- Two geology process models 地质模型

$$G_1 : K \text{ for any } x$$
$$G_2 : K = \begin{cases} K_1 & \text{for } x < 7000 \\ K_2 & \text{for } x \geq 7000 \end{cases}$$

均值/非均质渗透系数

where K follows $LN(2.9, 0.5^2)$, K_1 follows $LN[2.6, 0.3^2]$, and K_2 follows $LN[3.2, 0.3^2]$.

2. New VBSA Method

2.4 A Hypothetical Example

- Two snowmelt process models 融雪径流模型

$$h_2 = 0.3Q^{0.6} + 289$$ 河流水位-流量关系

where Q is the river discharge, which is estimated as

$$Q = C_{sn} \times M \times SVC \times A \times \frac{0.001}{86400}$$ 河道流量由上游融雪量控制

where C_{sn} = 0.8 is runoff coefficient, M = 0.7 mm/d is the snowmelt rate, SVC = 0.7 is the ratio of snow-covered area to the watershed area A = 2000 km^2.

不考虑/考虑太阳辐射

$$M_1 : M = f_1(T_a - T_m)$$ 度-日模型

$$M_2 : M = f_2(T_a - T_m) \boxed{+ rR_n}$$ 融入太阳辐射的度-日模型

where f_1 follows $LN(2.9, 0.5^2)$, f_2 follows $LN[2.6, 0.3^2]$, and r follows $LN[3.2, 0.3^2]$. Ta = 7°C is the average temperature, Tm = 0°C temperature threshold when snowmelt occurs, Rn = 80 W/m^{-2}.

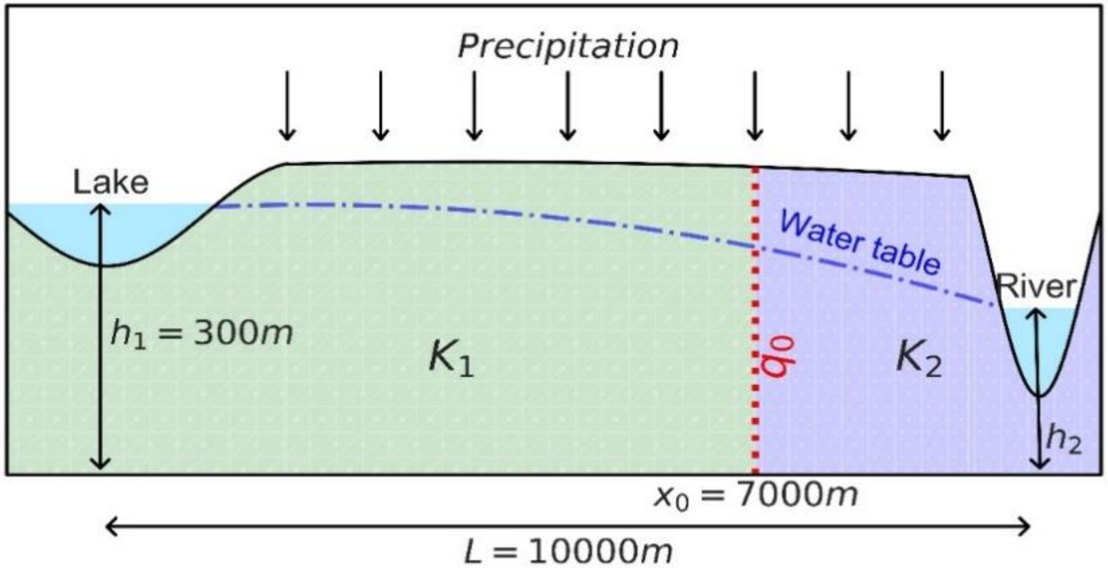
2. New VBSA Method

2.4 A Hypothetical Example

Analytical solutions for the groundwater discharge per unit width 解析解

$$\textbf{单宽流量}\quad q(x)=K_1\frac{h_1^{\,2}-h_2^{\,2}}{2\left(x_0-\lambda x_0+\lambda L\right)}-\frac{1}{2}w\frac{x_0^{\,2}-\lambda x_0^{\,2}+\lambda L^2}{x_0-\lambda x_0+\lambda L}+wx$$

where $\lambda = K_1/K_2$.



2. New VBSA Method

2.4 A Hypothetical Example

单模型下控制过程识别

Sensitivity Results – Eight individual system models

	Considering parametric uncertainty but not process model uncertainty														
Model	$R_1G_1M_1$				R1G1M2				R1G2M1				R1G2M2		
Parameter	a	K	f_1		a	K	f_2&r		a	K_1&K_2	f_1		a	K_1&K_2	f_2&r
S_i (%)	12	22	51		15	47	30		33	11	51		35	36	26
Rank	3	2	1		3	1	2		2	3	1		2	1	3
S_{Ti} (%)	12	37	66		15	56	38		34	16	55		36	39	28
Rank	3	2	1		3	1	2		2	3	1		2	1	3
Model	$R_2G_1M_1$				R2G1M2				R_2G2M_1				R2G2M2		
Parameter	b	K	f_1		b	K	f_1		b	K_1&K_2	f_1		b	K_1&K_2	f_2&r
S_i (%)	22	20	45		26	41	26		49	11	37		49	30	19
Rank	2	3	1		2	1	3		1	3	2		1	2	3
S_{Ti} (%)	22	33	59		26	48	33		50	14	40		50	32	20
Rank	3	2	1		3	1	2		1	3	2		1	2	3

2. New VBSA Method

2.4 A Hypothetical Example 多模型下控制过程识别

Sensitivity Results – Considering multiple models

	Considering parametric uncertainty and process model uncertainty		
Process	R	G	M
PS_K (%)	14.32	6.79	63.99
Rank	2	3	1
PS_{TK} (%)	14.67	21.68	78.54
Rank	3	2	1

The results were obtained by using a total number of 216,000,000 = 2 × 300 × 2 × 300 × 2 × 300 Monte Carlo simulation runs, corresponding to 2 recharge process models × 300 parameters × 2 geology process models × 300 parameters × 2 snowmelt process models × 300 parameters.

PS_K != PS_{TK} 跟Dai et al.（2017）最大的不同

The ranking changes!

2. New VBSA Method

2.4 A Hypothetical Example

Process Interaction

Second-order Third-order 过程间的相互作用

Decomposition: $PS_{TK} = PS_K + \sum_K \sum_V PS_{K,V} + \sum_K \sum_V \sum_W PS_{K,V,W} + \ldots$

Define: $$PS_{K,V} = \frac{V_{K,V}(E_{\sim(K,V)}[\Delta \mid K,V])}{V(\Delta)} - \frac{V_K(E_{\sim K}[\Delta \mid K])}{V(\Delta)} - \frac{V_V(E_{\sim V}[\Delta \mid V])}{V(\Delta)}$$

$$\begin{aligned}
&V_{K,V}(E_{\sim(K,V)}[\Delta \mid K,V]) \\
&= V_{\boldsymbol{M}_{K,V}}(E_{\boldsymbol{M}_{\sim(K,V)}}[\Delta \mid M_{K,V}]) \\
&= E_{\boldsymbol{M}_{K,V}} E_{\boldsymbol{\theta}_{K,V} \mid M_{K,V}} (E_{\boldsymbol{M}_{\sim(K,V)}} E_{\boldsymbol{\theta}_{\sim(K,V)} \mid M_{\sim(K,V)}} [\Delta \mid M_{K,V}, \theta_{K,V}, M_{\sim(K,V)}, \theta_{\sim(K,V)}])^2 \\
&\quad - (E_{\boldsymbol{M}_{K,V}} E_{\boldsymbol{\theta}_{K,V} \mid M_{K,V}} E_{\boldsymbol{M}_{\sim(K,V)}} E_{\boldsymbol{\theta}_{\sim(K,V)} \mid M_{\sim(K,V)}} [\Delta \mid M_{K,V}, \theta_{K,V}, M_{\sim(K,V)}, \theta_{\sim(K,V)}])^2
\end{aligned}$$

2. New VBSA Method

2.4 A Hypothetical Example

Weighted average? 多模型下控制程度（一阶和总效应）是否是单模型的平均?

- The variance left for total-effect process sensitivity index PS_{TK} under multiple models is not simply the weighted average.

$$
\begin{aligned}
& E_{\mathbf{M}_{\sim K}}\left(V_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]\right) \\
&= E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}\left(V_{\mathbf{M}_K}[\Delta \mid M_{\sim K}]\right) \\
&= \underbrace{\sum_{N_{\sim K}} \sum_{N_K} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}} V_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \theta_K, M_K, \theta_{\sim K}, M_{\sim K}] P(M_K) P(M_{\sim K})}_{\textit{within-model variance left}} \\
&\quad \boxed{+ \underbrace{\sum_{N_{\sim K}} \sum_{N_K} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}} \begin{pmatrix} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \theta_K, M_K, \theta_{\sim K}, M_{\sim K}] \\ -E_{\mathbf{M}_K} E_{\boldsymbol{\theta}_K|M_K}[\Delta \mid \theta_K, M_K, \theta_{\sim K}, M_{\sim K}] \end{pmatrix}^2 P(M_K) P(M_{\sim K})}_{\textit{between-model variance left}}}
\end{aligned}
$$

多模型下的控制指标并不是单模型下的简单平均！！！

2. New VBSA Method

2.4 A Hypothetical Example

Convergence and Computational Cost Analysis 高效算法对比

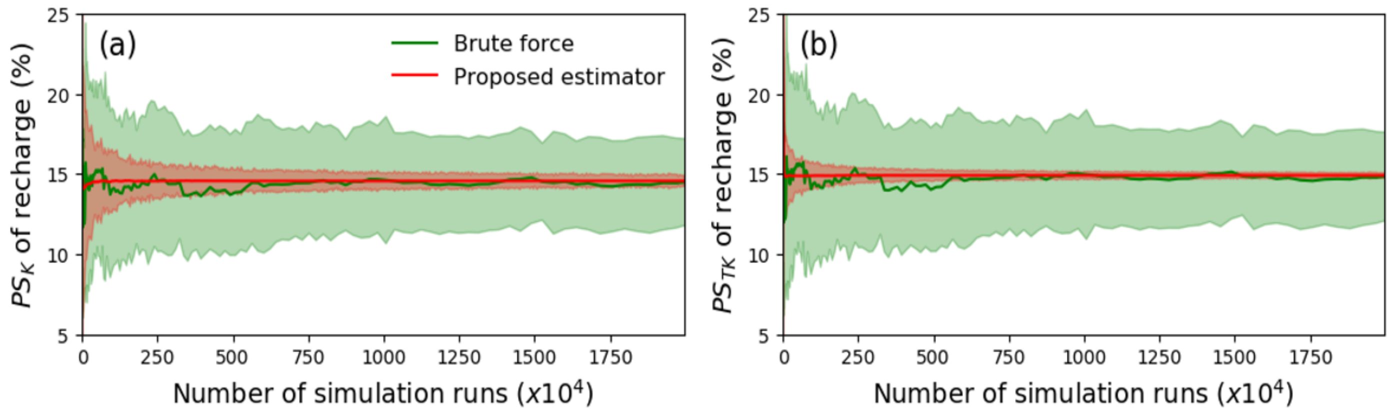


Figure 2-9. Convergence plots of (a) first-order process sensitivity index PS_K and (b) total-effect process sensitivity index PS_{TK} of recharge process in the groundwater flow modeling. The green and red areas denote the 95% confidence intervals estimated using bootstrap sampling for the brute force method and proposed efficient estimator, respectively.

Table of Contents 目录

3. New MMADS Method

3.1 Methodology

3.1.1 Traditional Morris Screening Method for a Single Model

Based on the Elementary Effect (EE) 基效应 （Morris, 1991）

$$EE_i = \frac{[y(x_1,\dots,x_{i-1},x_i+\delta,x_{i+1},\dots,x_k) - y(\mathbf{x})]}{\delta}$$

- Assume each parameter ranges in **interval [0,1]** and partitioned into **(p−1)** equally-sized intervals.

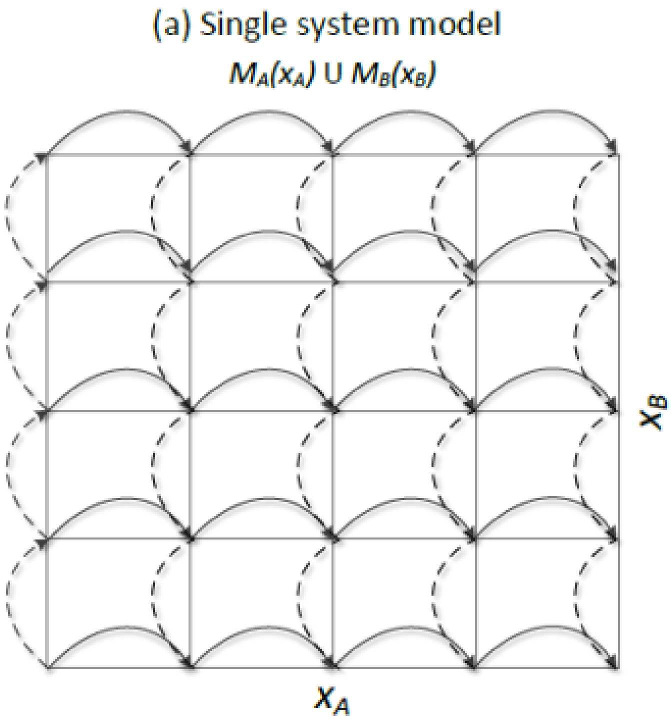
- There reference value of each parameter is selected randomly from the set [0, 1/(p−1), 2/(p−1),…, 1−Δ].

- The fixed increment Δ=p/2(p−1) is added to each parameter in order to compute the **EE.**

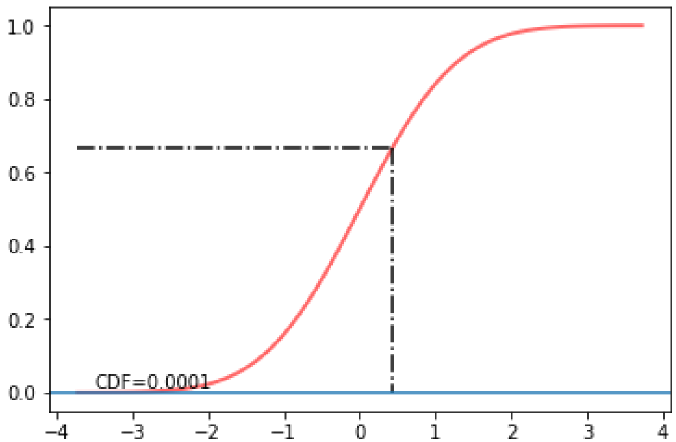
3. New MMADS Method

3.1 Methodology

3.1.1 Traditional Morris Screening Method for a Single Model



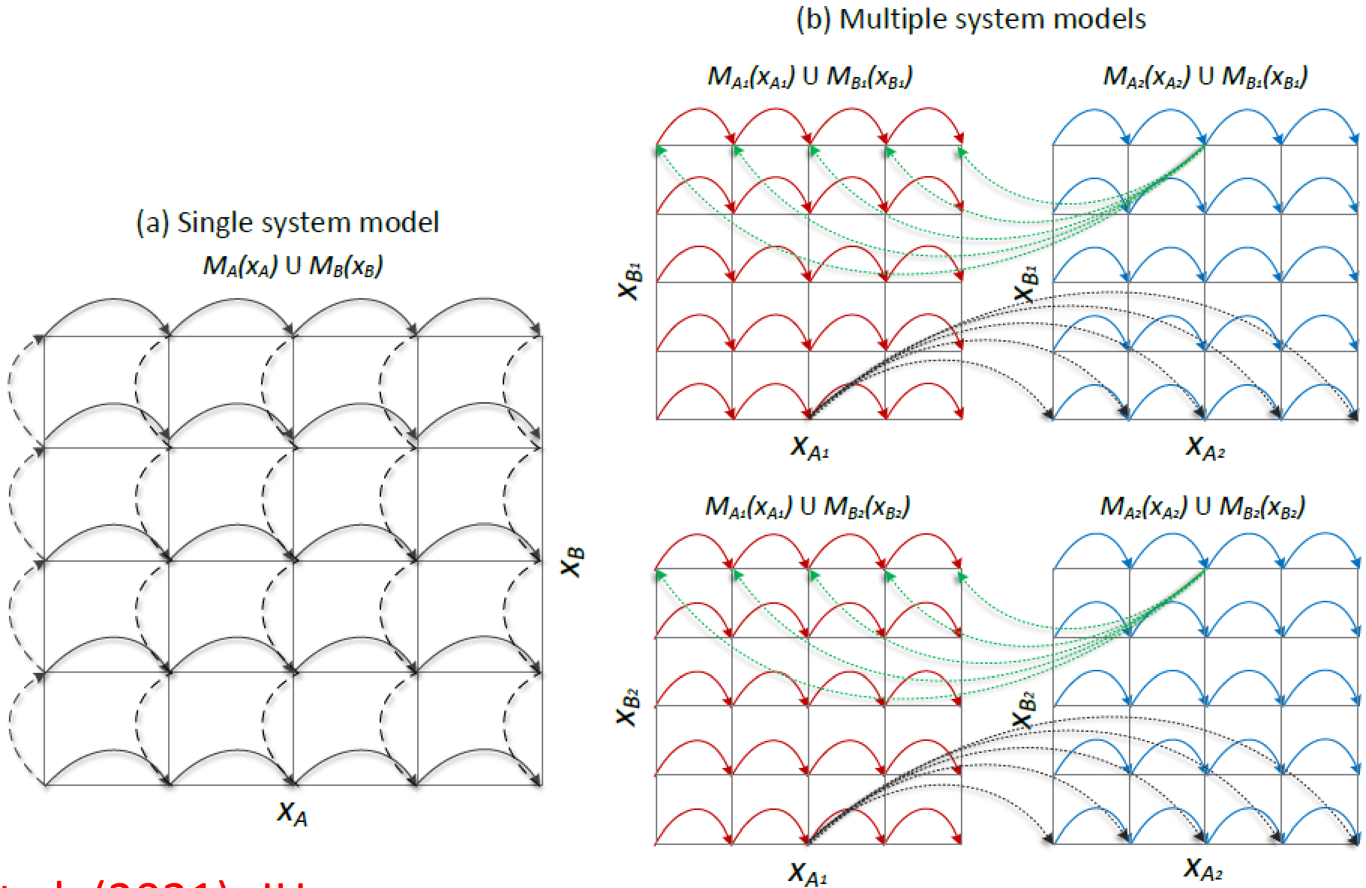
- Mean of EE_i
- Standard deviation of EE_i
- Mean of $|EE_i|$



3. New MMADS Method

3.1 Methodology

3.1.2 Basic Ideas of MMADS for Mluti-model



Yang et al. (2021), JH

3. New MMADS Method

3.1 Methodology

3.1.2 Basic Ideas of MMADS for Mutli-model

The basic ideas of MMADS are to evaluate the absolute difference, $d\Delta$, of the system model output, Δ, in the following two situations under process model uncertainty (assuming that parametric uncertainty does not exist at this stage):

(1) Transition between the two models of process *A*, conditioning on model M_{B1}, , of process *B*. There are four transitions between the two models of process *A* as follows: $T_{\mathbf{MA}} = \{M_{A1} \text{->} M_{A1}, M_{A1} \text{->} M_{A2}, M_{A2} \text{->} M_{A1}, M_{A2} \text{->} M_{A2}\}$ where $\mathbf{M}_A = \{M_{A1}, M_{A2}\}$. The difference, $d\Delta$, of model output for each transition is evaluated, and then averaged over all the transitions to yield $E_{T_{MA}}(d\Delta|M_{B1})$.

(2) Transition between the two models of process *A*, conditioning on model M_{B2}, , of process *B*. There are four transitions between the two models of process *A* as follows: $T_{\mathbf{MA}} = \{M_{A1} \text{->} M_{A1}, M_{A1} \text{->} M_{A2}, M_{A2} \text{->} M_{A1}, M_{A2} \text{->} M_{A2}\}$ where $\mathbf{M}_A = \{M_{A1}, M_{A2}\}$. Repeat the calculation above, we have $E_{T_{MA}}(d\Delta|M_{B2})$.

3. New MMADS Method

3.1 Methodology

3.1.2 Basic Ideas of MMADS for Mutli-model

(1) For $ET_{MA}(d\Delta|M_{B1})$ and $ET_{MA}(d\Delta|M_{B2})$, calculate the average to yield $E_{\mathbf{M}B}ET_{MA}(d\Delta|M_B) = E_A(d\Delta)$ for process A, which measures how the changes between and within the two models of process A affect model output.

Similarly, we can also have $E_{\mathbf{M}A}ET_{MB}(d\Delta|M_A) = E_B(d\Delta)$ for process B, which measures how the changes between and within the two models of process B affect model output.

3. New MMADS Method

3.1 Methodology

3.1.2 Basic Ideas of MMADS for Multi-model

差值均值 $$E_K(d\Delta) = E_{\mathbf{M}_{\sim K}} E_{T_{\mathbf{M}_K}}(d\Delta \mid M_{\sim K})$$

差值方差 $$V_K(d\Delta) = E_{\mathbf{M}_{\sim K}} V_{T_{\mathbf{M}_K}}(d\Delta \mid M_{\sim K}) + V_{\mathbf{M}_{\sim K}} E_{T_{\mathbf{M}_K}}(d\Delta \mid M_{\sim K})$$

Expanded to include parameter uncertainty

$$\begin{aligned} E_K(d\Delta) &= E_{\mathbf{M}_{\sim K}} E_{T_{\mathbf{M}_K}}(d\Delta \mid M_{\sim K}) \\ &= E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{M_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K}) \end{aligned}$$

$$\begin{aligned} V_K(d\Delta \mid K) &= E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{M_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K})^2 \\ &\quad - \left(E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{M_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K}) \right)^2 \end{aligned}$$

3. New MMADS Method

3.2 Numerical Evaluation

- Expectation with model uncertainty – Model Averaging

$$E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}(\bullet) = \sum E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}(\bullet) P(M_{\sim K})$$

$$E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K}|T_{M_K}}(\bullet) = \sum E_{T_{\boldsymbol{\theta}_K}|T_{M_K}}(\bullet) P(T_{M_K})$$

模型求期望--模型平均法

- Expectation with parameter uncertainty – Monte Carlo

$E_{\boldsymbol{\theta}_{\sim K}|M_{\sim K}}$ $E_{T_{\boldsymbol{\theta}_K}|T_{M_K}}$

参数求期望—蒙特卡洛方法

3. New MMADS Method

3.2 Numerical Evaluation

Use LHS method to generate the sampling matrix of all parameters

Loop [1] over model combinations $M_{\sim K}$ for process $\sim K$ in $\mathbf{M}_{\sim K}$

  Combine parameter values for parameters associate with $M_{\sim K}$ to yield $\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}$

  Loop [2] over parameter realizations $\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}$

    Loop [3] over process model M_K for process K in $\mathbf{M}_K$

      Combine parameter values for parameters associated with M_K to yield $\boldsymbol{\theta}_K \mid M_K$

      Loop [4] over parameter realizations $\boldsymbol{\theta}_K \mid M_K$

        Compute $\Delta \mid \theta_{\sim K}, M_{\sim K}, \theta_K, M_K$

      End loop [4]

    End loop [3]

    Loop [5] over process model transitions in $T_{\mathbf{M}_K}$

      Loop [6] over parameter realizations $T_{\boldsymbol{\theta}_K \mid T_{\mathbf{M}_K}}$ associated with $T_{\mathbf{M}_K}$

        Compute $d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}$ and $\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)^2$

      End Loop [6]

      Compute $E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)$ and $E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)^2$

    End Loop [5]

    Compute $E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)$ and

    $E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)^2$ using model averaging

  End Loop [2]

  Compute $E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}\left(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_{\sim K}}, T_{M_K}\right)$ and

  $E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K})^2$

End Loop [1]

Compute $E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K})$ and

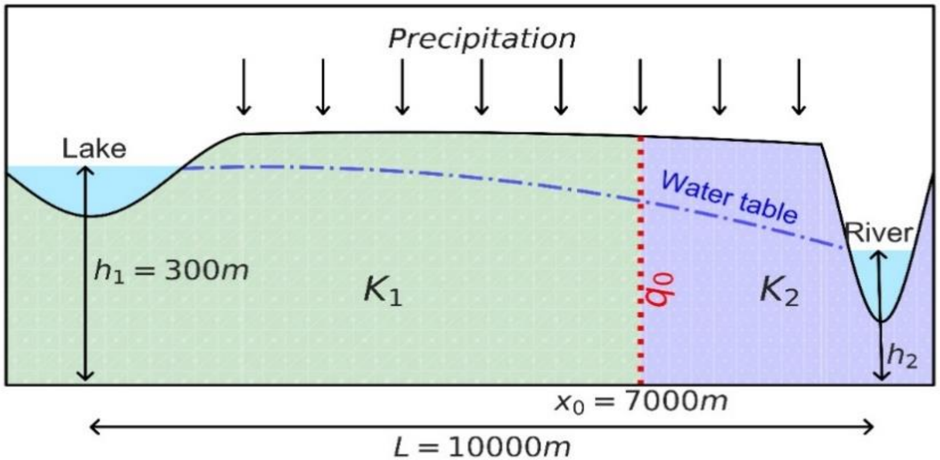
$E_{\mathbf{M}_{\sim K}} E_{\boldsymbol{\theta}_{\sim K} \mid M_{\sim K}} E_{T_{\mathbf{M}_K}} E_{T_{\boldsymbol{\theta}_K} \mid T_{\mathbf{M}_K}}(d\Delta \mid \theta_{\sim K}, M_{\sim K}, T_{\theta_K}, T_{M_K})^2$ using model averaging

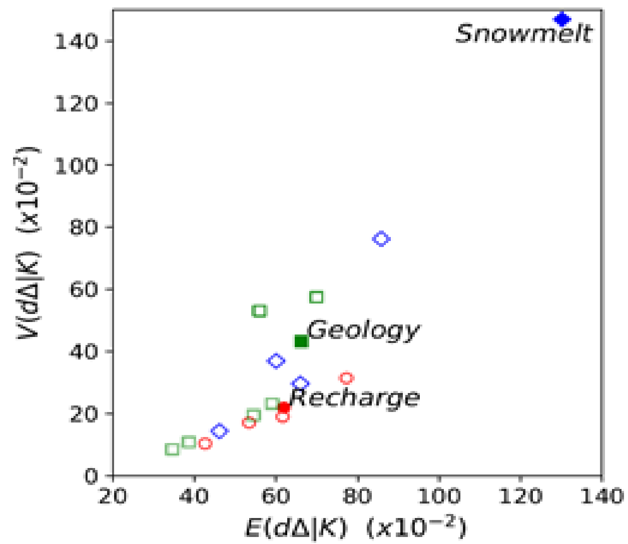
伪代码

Figure 3-2 Pseudo codes for evaluating the two sensitivity measures for process *K* using LHS. The number of model execution based on the pseudo code shown in Figure 3-2 is $(N_{M\sim K} \times n) \times (N_{MK} \times n)$, where the n^2 is caused by the nested loops [2] and [4], respectively.

3. New MMADS Method

3.4 A Hypothetical Example





Considering parametric uncertainty and process model uncertainty			
Process	R	G	M
PS_{TK}	14.67	21.68	78.54
Rank	3	2	1
Mean	61.9 60.8[a] 65.5[b]	66.2 65.8[a] 68.9[b]	130.1 125.8[a] 136.5[b]
Variance	22.0 21.1[a] 33.6[b]	43.3 37.9[a] 40.9[b]	146.9 129.7[a] 164.8[b]

3. New MMADS Method

3.4 A Hypothetical Example

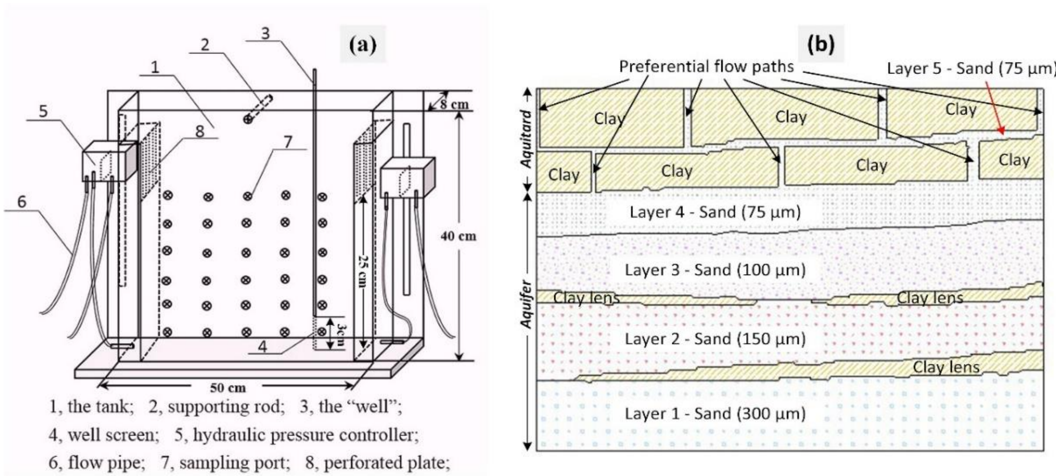
The MMADS results shown in Table 3-1 are obtained by using n = 5,000 random samples for each parameter. The total number of model executions is 200,000,000 = 2 × 2 × 5,000 × 2 × 5,000, which ensures that the Monte Carlo calculation of MMADS converges. The high computational cost is unnecessarily and can be substantially reduced. Table 3-1 shows that the MMADS results are still accurate enough when using n = 20, and the corresponding number of model executions is 3,200 = 2 × 2 × 20 × 2 × 20, which is computationally affordable to moderate models with the current computational resources.

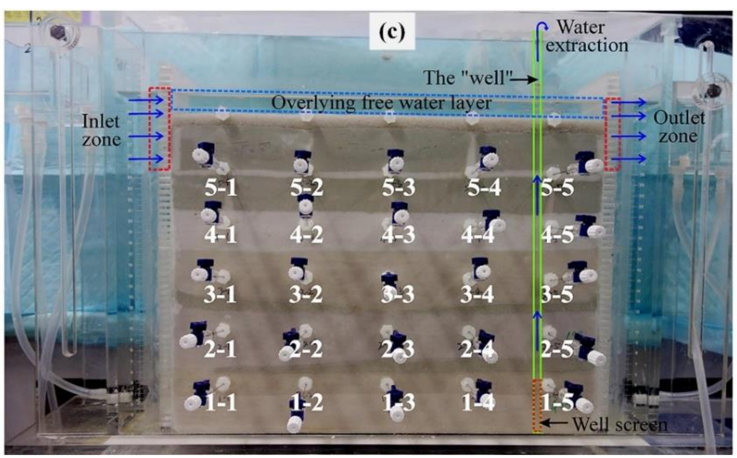
The number of model executions of the binning method is 800 = 2 × 2 × 2 × 100, substantially smaller than the number of 3,200 = 2 × 2 × 20 × 2 × 20 needed for the pseudo code shown in Figure 3-2.

Table of Contents 目录

4. Application to Arsenic Sorption and Reactive

4.1 Lab Experiment by Duan et al. (2020)





A synthetic heterogeneous aquifer was constructed in a tank and a flume experiment was performed by Duan et al. (2020) to investigate the effect of physico-chemical heterogeneity with preferential flow paths on As sorption and reactive transport under water extraction in a layered system.

Figure 4-1. Schematic diagram of the experimental setup: (a) experimental tank desgin; (b) schematic diagram of the layered system with complex preferential flow paths; (c) image of the packed flume showing the sampling ports with numeric labels (modified from Duan et al. (2020)).

4. Application to Arsenic Sorption and Reactive

4.2 Numerical Model Setup

A 2-D reactive transport model adopted from Duan et al. (2020) was built to simulate the As sorption and transport in the flume system. The numerical model was built using an open source, state-of-the-art massively **parallel reactive flow and transport model** for describing subsurface processes (Hammond et al., 2014).

40 cm × 8 cm × 30 cm

200 × 1 × 150 uniform grid cells

Constant hydraulic pressure head. Left head is 32.3 cm and the right head is 30.0 cm.

Simulation time 104 days: 0~7.75 day: no pumping; 7.75~99 day: pumping start; 99day ~ 104 day : end pumping

4. Application to Arsenic Sorption and Reactive

4.2 Numerical Model Setup

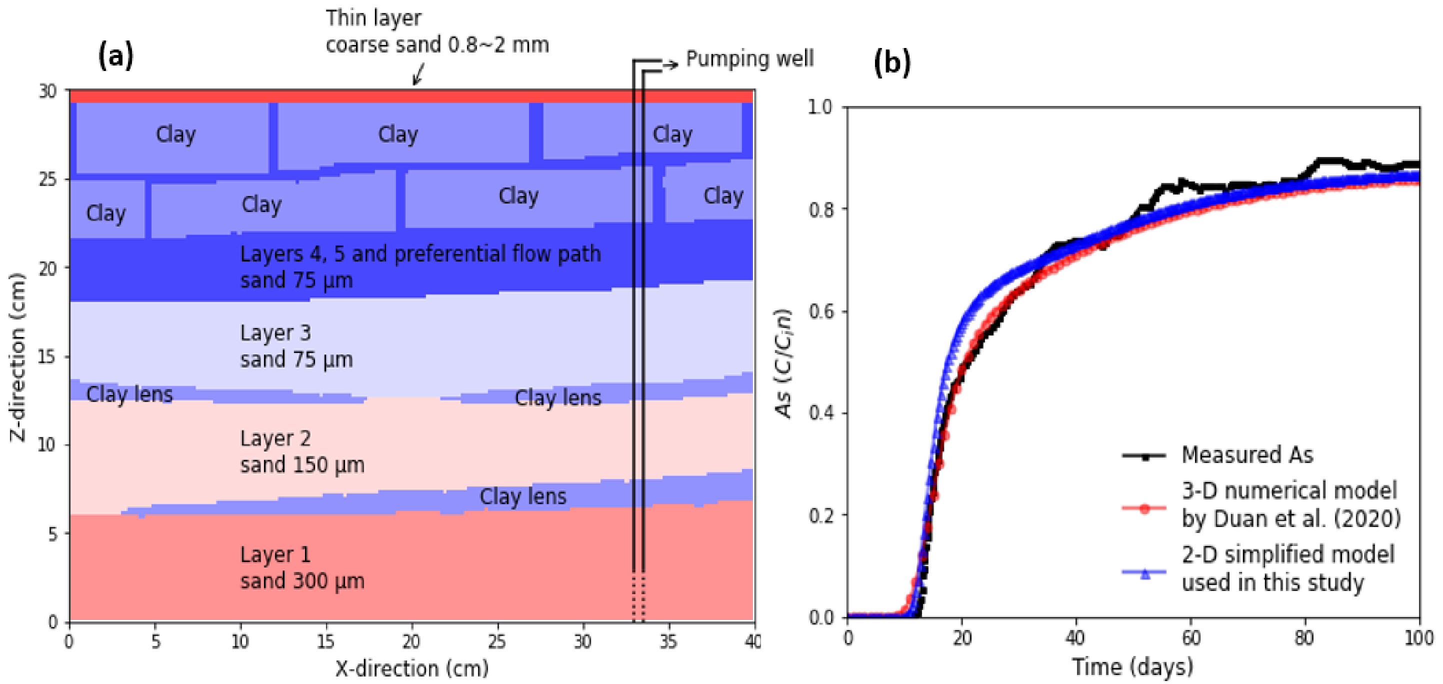
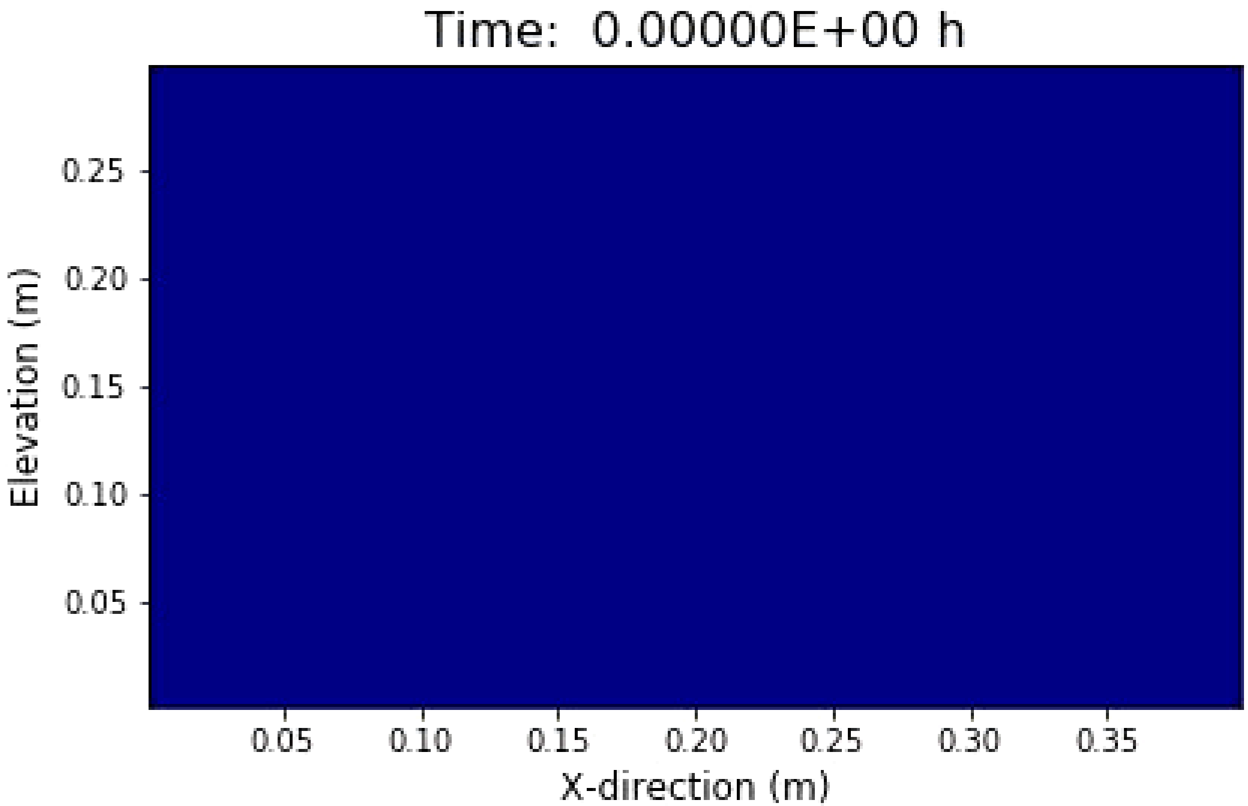


Figure 4-2 (a) Schematic diagram of the layered system with complex preferential flow paths and (b) time series of As concentration in the pumping well from the sampling measurements, the 3-D numerical model from Duan et al. (2020), and 2-D simplified model used in this study.

4. Application to Arsenic Sorption and Reactive

4.2 Numerical Model Setup

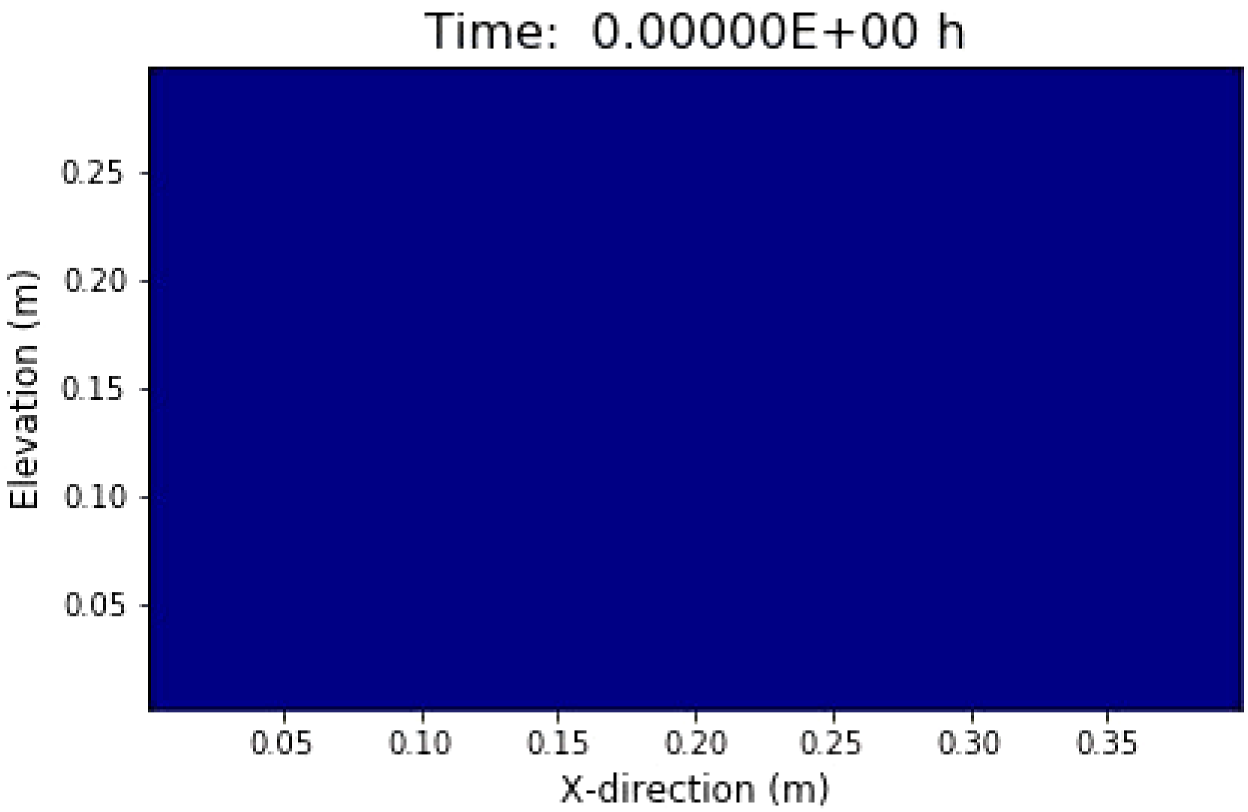
Total Tracer C/C_0 Pumping state at t=7.75d (186h)



4. Application to Arsenic Sorption and Reactive

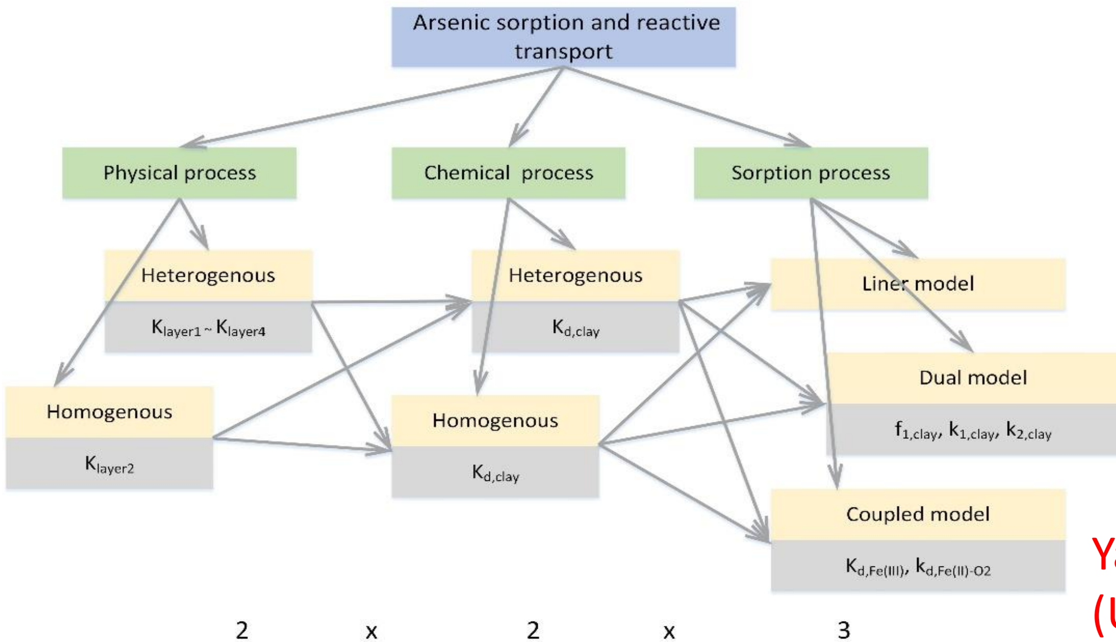
4.2 Numerical Model Setup

Total As C/C_0 Pumping state at t=7.75d (186h)



4. Application to Arsenic Sorption and Reactive

4.3 Modeling Uncertainty



Yang et al., 2021, WRR (Under preparation)

Figure 4-3 Overview of the 12 = 2 × 2 × 3 alternative models, corresponding to the heterogenous or homogenous permeability field (physical process), heterogenous or homogenous distribution coefficient (chemical process), and three sorption models (sorption process). The gray boxes show the uncertainty parameters associated with the corresponding process models. Note there is no parameter involved in the linear sorption model

4. Application to Arsenic Sorption and Reactive

4.3 Modeling Uncertainty

- Three sorption process models 三个吸附模型

Simple linear equilibrium model 线性平衡吸附模型 $q_{As} = K_d \times C_{As}$

Dual first-order kinetic sorption model 双室一级动力学模型

$$\frac{dq_{t,i}}{dt} = k_i(q_{e,i} - q_{t,i}) \qquad (i=1,2) \quad q_{e,i} = K_d \times C_{As} \times f_i \quad q_t = q_{t,1} + q_{t,2}$$

Linear equilibrium sorption model coupled with a kinetic model accounting for oxidation of Fe(II)-bearing clay minerals
线性平衡吸附模型耦合二价铁氧化吸附模型

$$\text{Fe(II)-clay+0.25O}_2 \rightarrow \text{Fe(III)-clay}$$

$$\mathrm{R}_{\mathrm{Fe(III)}} = k_{Fe(II)-O_2} c(Fe(II))c(DO)$$

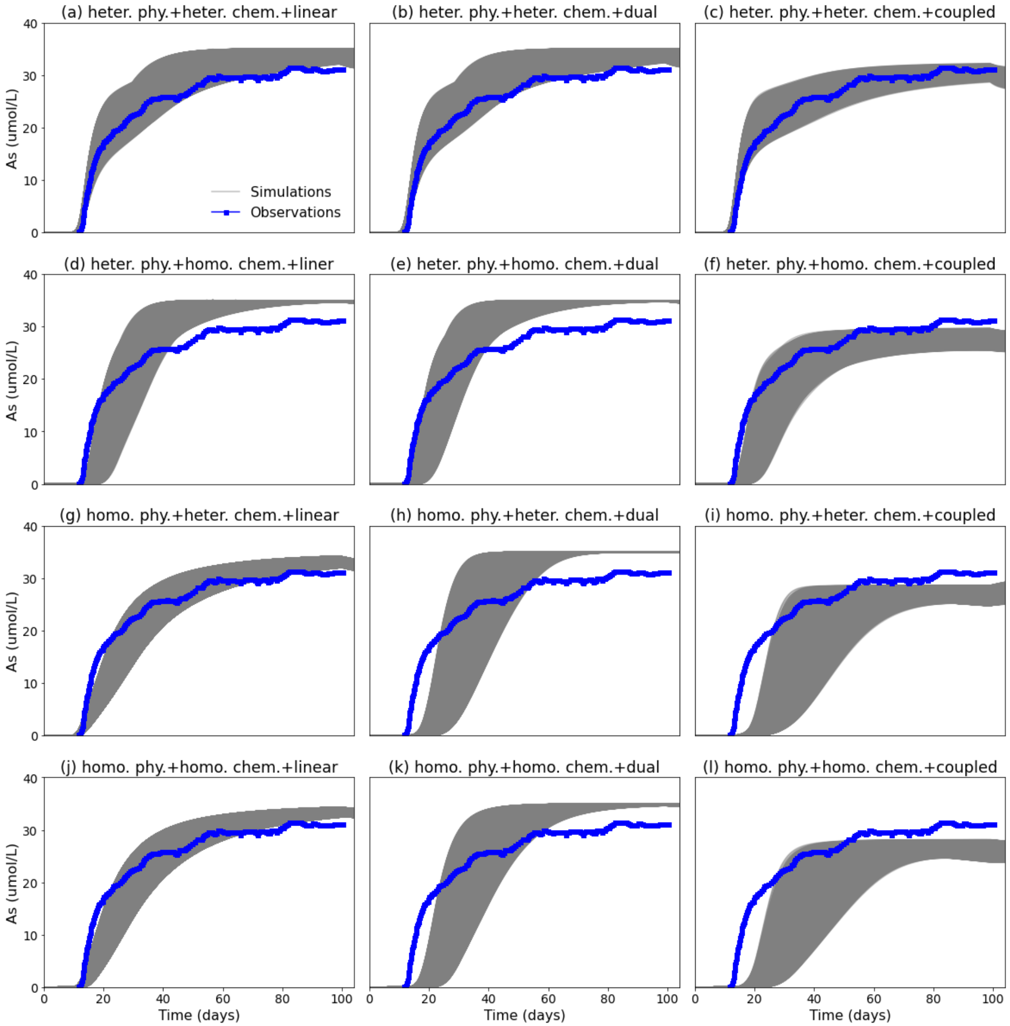
4. Application to Arsenic Sorption and Reactive

4.3 Modeling Uncertainty

Parameter	Unit	Nominal*	Description
K_{layer1}	m^2	1.86×10^{-11}	Permeability of layer 1
K_{layer2}	m^2	4.66×10^{-12}	Permeability of layer 2
K_{layer3}	m^2	2.07×10^{-12}	Permeability of layer 3
K_{layer4}	m^2	1.16×10^{-12}	Permeability of layer 4
$K_{d,clay}$	mL/g	5.65	Sorption distribution coefficient of clay
$f_{1,clay}$	-	0.89	Site fraction of clay at sorption site 1
$k_{1,clay}$	h^{-1}	118.07	First-order rate constant of clay at site 1
$k_{2,clay}$	h^{-2}	0.23	First-order rate constant of clay at site 2
$K_{d,Fe(III)}$	mL/g	3.01×10^{3}	Sorption distribution coefficient of Fe(III)
$K_{Fe(II)-O2}$	L/mol/s	0.5	Reaction rate constant

4. Application to Arsenic Sorption and Reactive

4.4 Simulation Results 基于方差的方法+高效算法

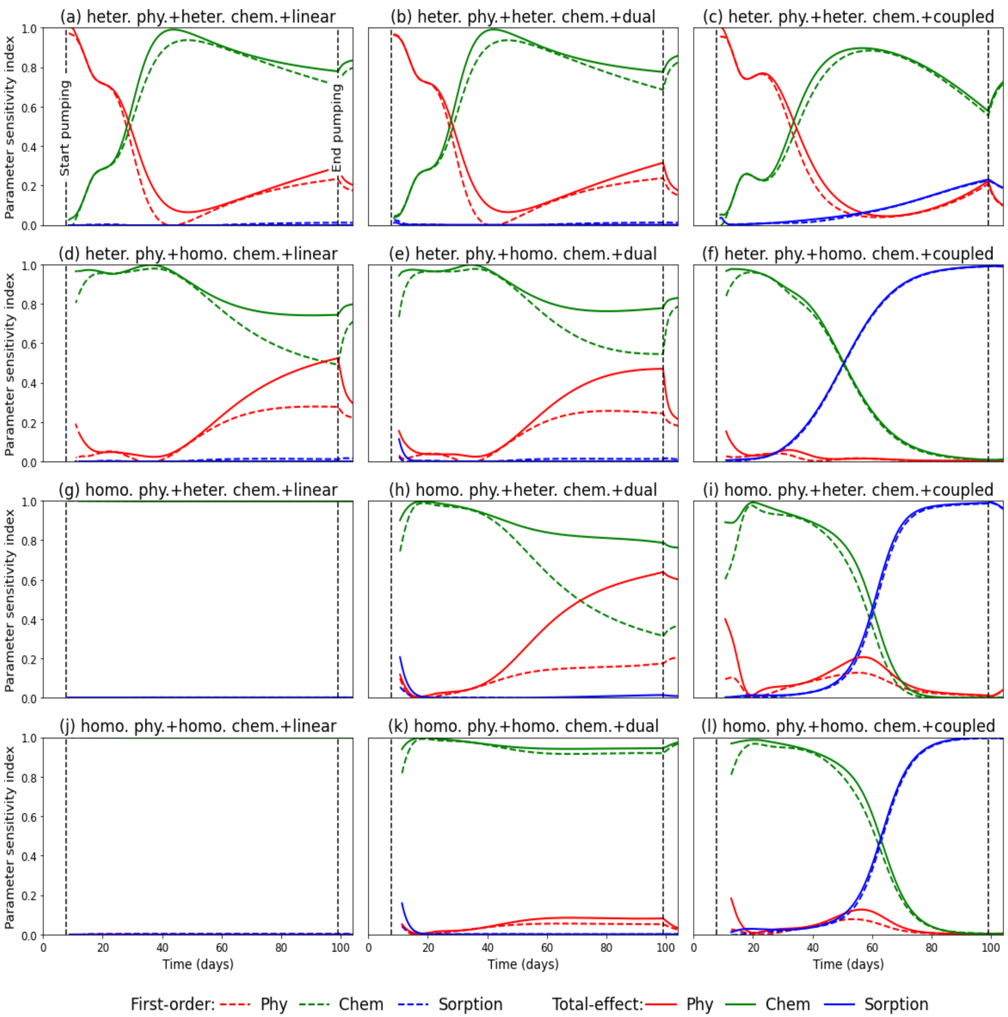


Using the efficient design proposed in Chapter 2, PFLOTRAN simulations were ran on a total of 2 × 2 × 3 × (3 + 2) × 3,600 = 216,000 model simulations considering all possible combinations of model configuration and input parameters. 总运行次数

The wall clock time for each simulation run was approximately 2 ~ 5 mins using 8 cores on the high-performance computer system at the Florida State University. 佛罗里达州立大学超算平台

4. Application to Arsenic Sorption and Reactive

4.4 Simulation Results



单模型下控制过程识别

(1) For individual system model , Importance/influence changes dramatically along with the simulation time.

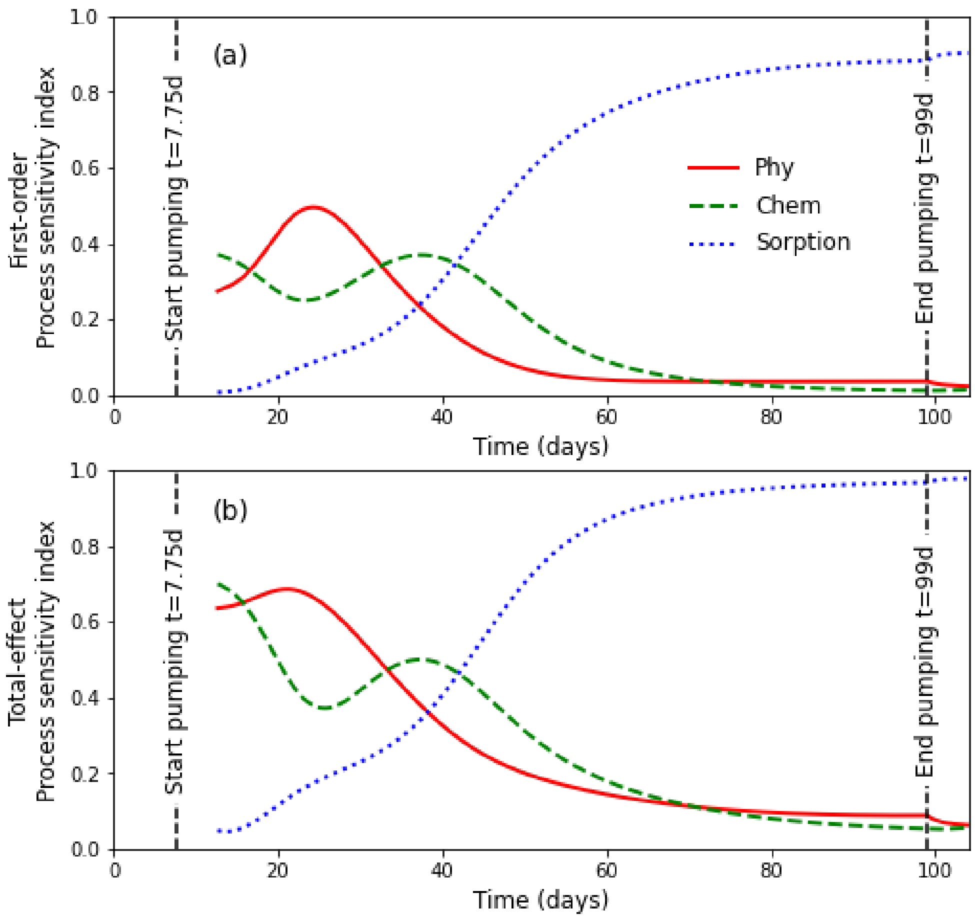
(2) Within individual system model, total-effect always higher than the first-order.

(3) Relative importance/influence of the three process changes dramatically among the twelve models.

(4) Physical process is only identified to be controlling at the beginning simulation period with heterogenous permeability filed and diffusion coefficient field.

4. Application to Arsenic Sorption and Reactive

4.4 Simulation Results



多模型下控制过程识别

(1) At very beginning of the simulation period, the chemical process is the most important process and the sorption process is not important at all.

(2) After the simulation time of t =16 d, the physical process becomes more important than the chemical process and becomes the most important process.

(3) The chemical process becomes the most important process once again after t = 32 d.

(4) The sorption process becomes the largest one among the three process after t = 41d.

Table of Contents 目录

5. Application to a Complex Biogeochemical Model

5.1 Site Description

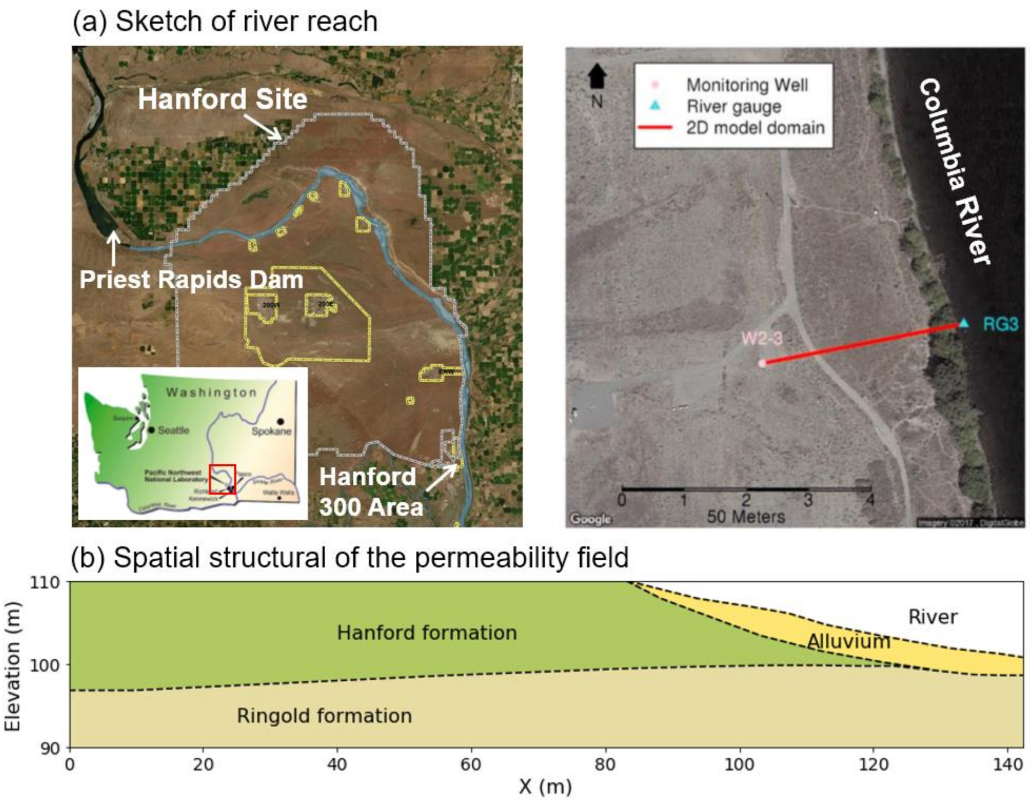
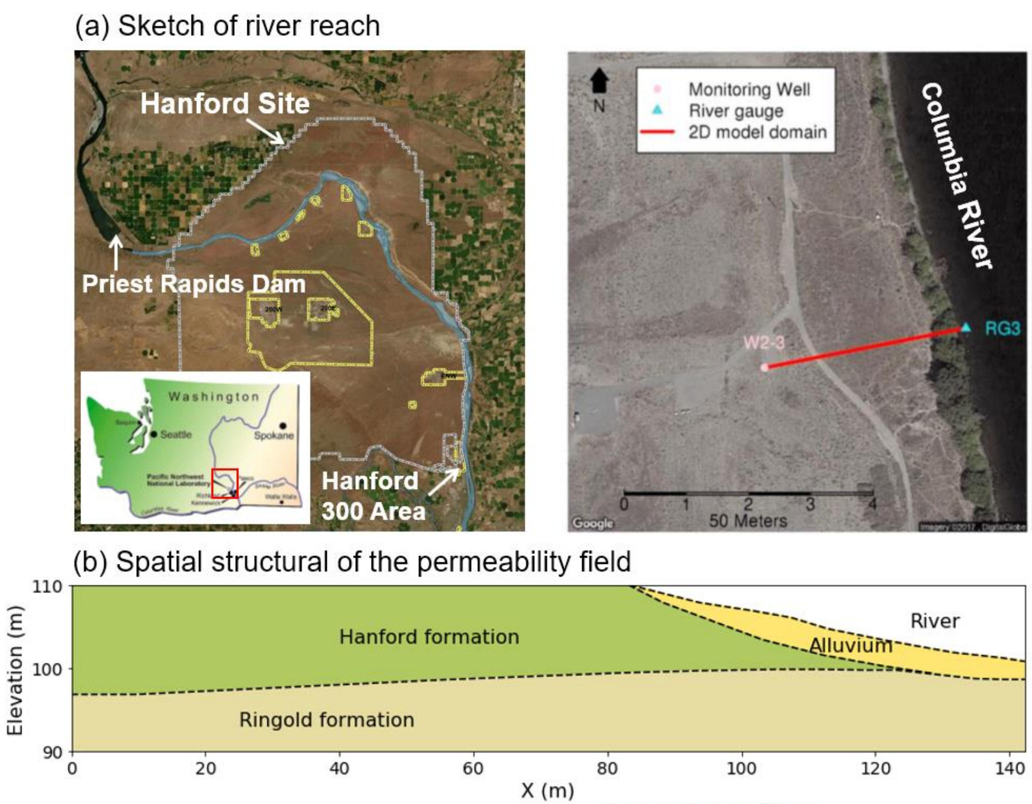


Figure 5-1. Location map and conceptual model. (a) Sketch of the river reach. The redline is the profile on the 2-D model between inland groundwater monitoring well W2-3 and the river gauge RG3 with continuous level and temperature data. (b) Geological structure of the Hanford formation, Ringold formation, and alluvium layer beneath the Columbia River.

The Columbia River is the second largest river in the contiguous United States in terms of total flow. The river stage at the study site is highly dynamic due to the upstream dam operations, fluctuating ~0.5 m daily and up to 2–3 m annually.

5. Application to a Complex Biogeochemical Model

5.1 Numerical Model Setup



A 2-D groundwater biogeochemical reactive transport model has been recently built in the 300 Area by Song et al. (2018).

143.2 m × 20 m

580 × 325 nonuniform grid cells

Lateral boundaries: a transient hydraulic pressure head boundary and a Dirichlet temperature boundary .

Top and bottom: no flow no heat

Simulation period: 2010-2015 with hourly step

Song et al., 2018 WRR


AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR022586

Key Points:
- High-frequency flow variations enhance hydrologic exchange and create long-term alterations to thermal regimes and biogeochemical reactions
- High-frequency flow variations have the largest impact on thermal

Drought Conditions Maximize the Impact of High-Frequency Flow Variations on Thermal Regimes and Biogeochemical Function in the Hyporheic Zone

Xuehang Song[1], Xingyuan Chen[1], James Stegen[1], Glenn Hammond[2], Hyun-Seob Song[1], Heng Dai[1], Emily Graham[1], and John M. Zachara[1]

[1]Pacific Northwest National Laboratory, Richland, WA, USA, [2]Applied Systems Analysis and Research, Sandia National Laboratories, Albuquerque, NM, USA

5. Application to a Complex Biogeochemical Model

5.1 Numerical Model Setup

Reaction Network

A two-step denitrification for the dissolved organic carbon (CH_2O) and oxidative respiration : 有机碳参与下的反硝化反应和有氧呼吸

$$CH_2O + 2NO_3^- \rightarrow 2NO_2^- + CO_2 + H_2O$$

$$CH_2O + 4/3NO_2^- + 4/3H^+ \rightarrow 2/3N_2 + CO_2 + 5/3H_2O$$

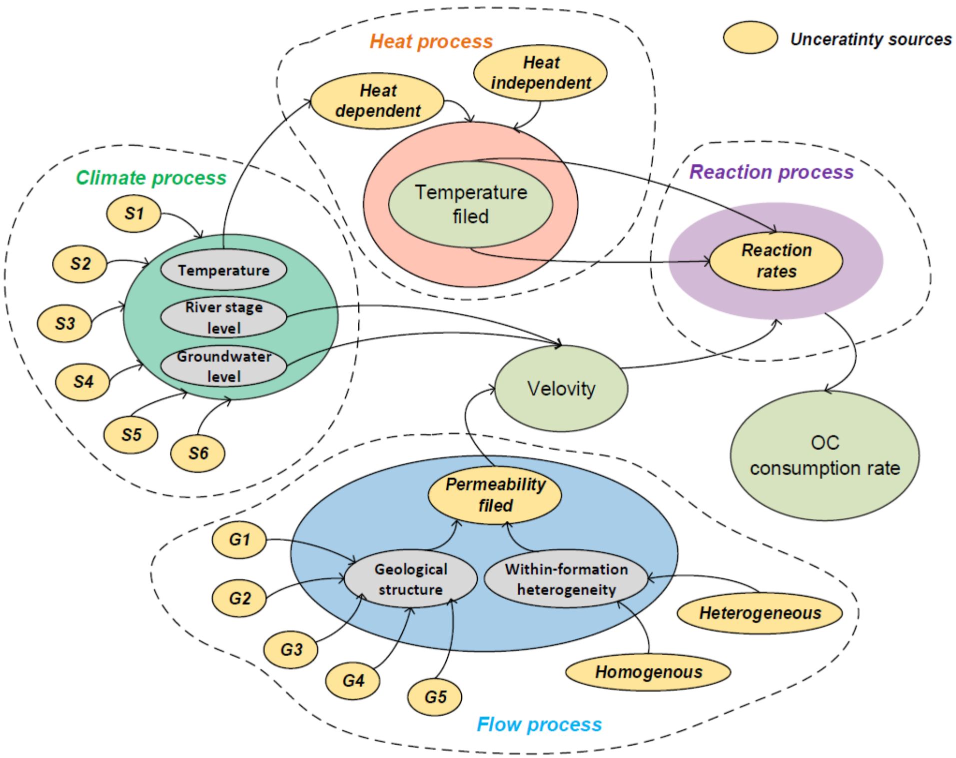
$$CH_2O + O_2 \rightarrow CO_2 + H_2O$$

Microbial biomass (BM; $C_5H_7O_2N$) synthesis was considered as

微生物生物量 $CH_2O + 1/5NH^+ \rightarrow 1/5C_5H_7O_2N + 5/3H_2O + 1/5H^+$

5. Application to a Complex Biogeochemical Model

5.2 Uncertainty Sources



Four processes

(1) Climate 气象场景: six climate scenarios.

(2) Flow**水流**: The permeability and the thickness.

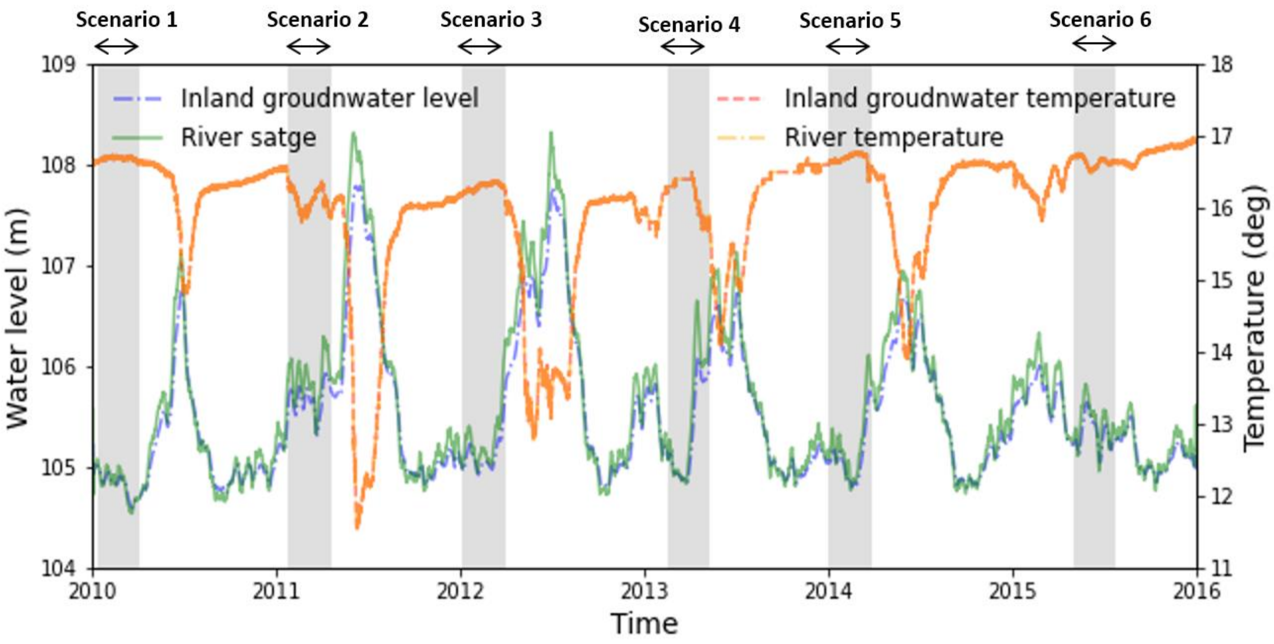
(3) Heat**热**: with and without heat.

(4) Reaction **反应**: reaction rates.

Yang et al., 2021, WRR
(Under preparation)

5. Application to a Complex Biogeochemical Model

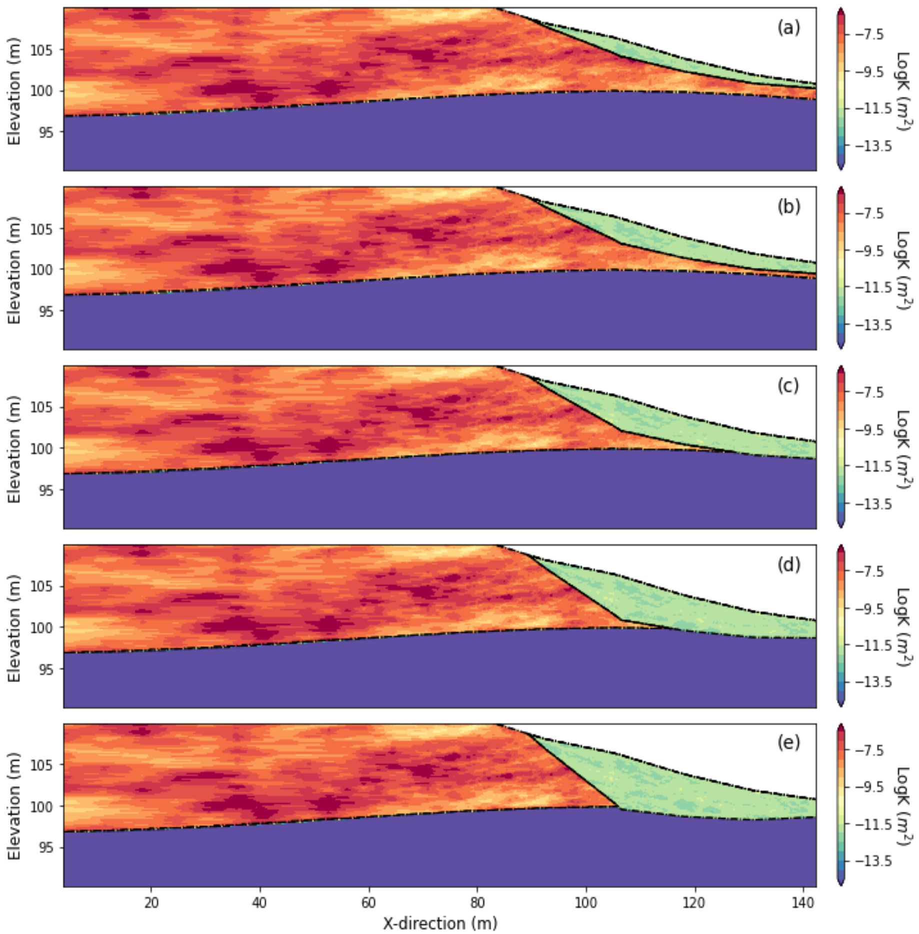
5.2 Uncertainty Sources



Climate: six climate scenarios.

5. Application to a Complex Biogeochemical Model

5.2 Uncertainty Sources



The five geological configurations shared the same geostatistical realizations of permeability formation.

5. Application to a Complex Biogeochemical Model

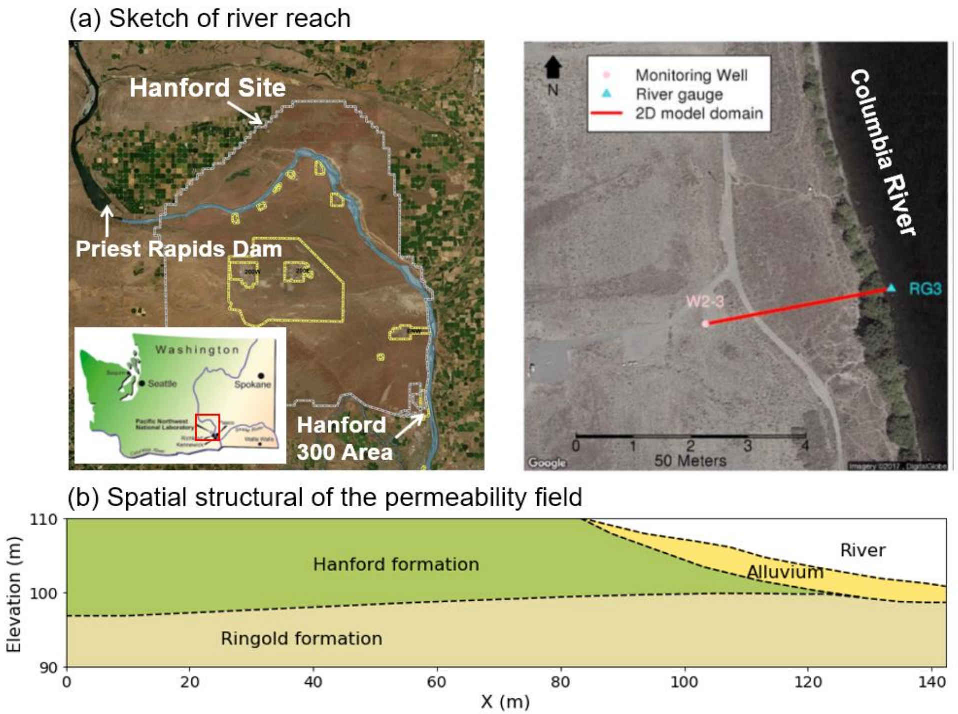
5.2 Simulation Results

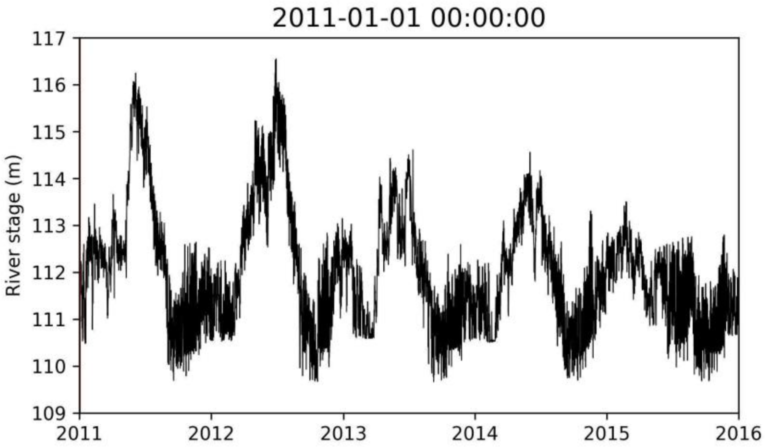
The total number of model simulations considering all possible combinations of model inputs and model configuration is 6 × 5 × 2 × 2 × 100 = 12,000, which represents six scenarios, five thicknesses of alluvium layer, heterogeneous/homogeneous formations, with/without heat transport process, and 100 reaction rates and permeability fields. 总运行次数12,000

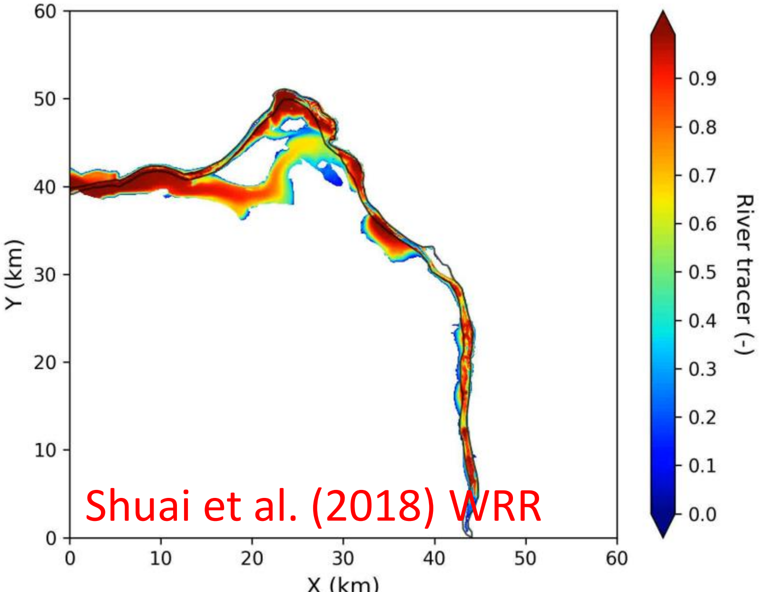
The carbon consumption rate is the most important indicator for the biogeochemical processes because it is the energy source for all biogeochemical reactions in the reaction network of the underlying system. The wall clock time for each realization was approximately 15 min using 128 cores on the Hopper supercomputer at the National Energy Research Scientific Computing Center (NERSC). 美国国家能源研究科学计算中心

5. Application to a Complex Biogeochemical Model

5.2 Simulation Results







Shuai et al. (2018) WRR

5. Application to a Complex Biogeochemical Model

5.2 Simulation Results

Total River Tracer C/C_0

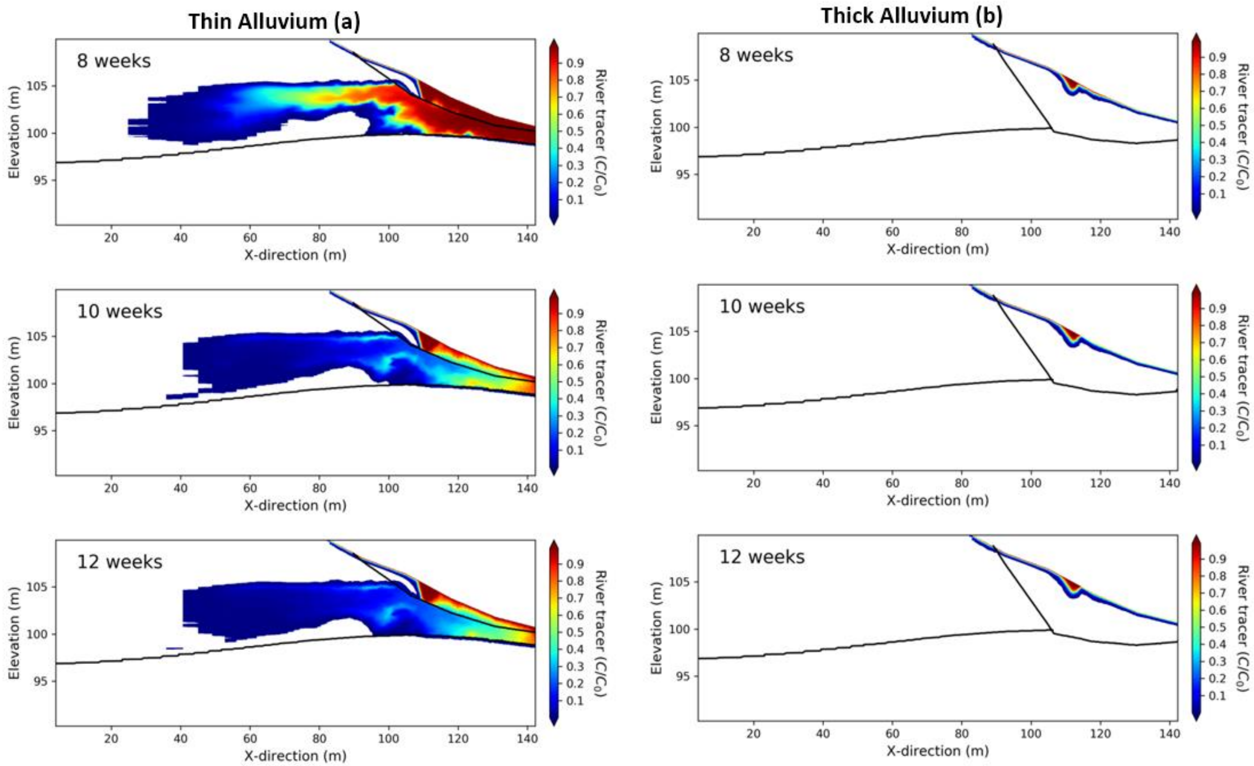


Figure 5-5 Snapshots of normalized river tracer concentration at the three different simulation times under scenario 1 using two distinct thicknesses of the alluvium layer. The homogeneous formation and heat transport process were used.

5. Application to a Complex Biogeochemical Model

5.2 Simulation Results

OC consumption rate

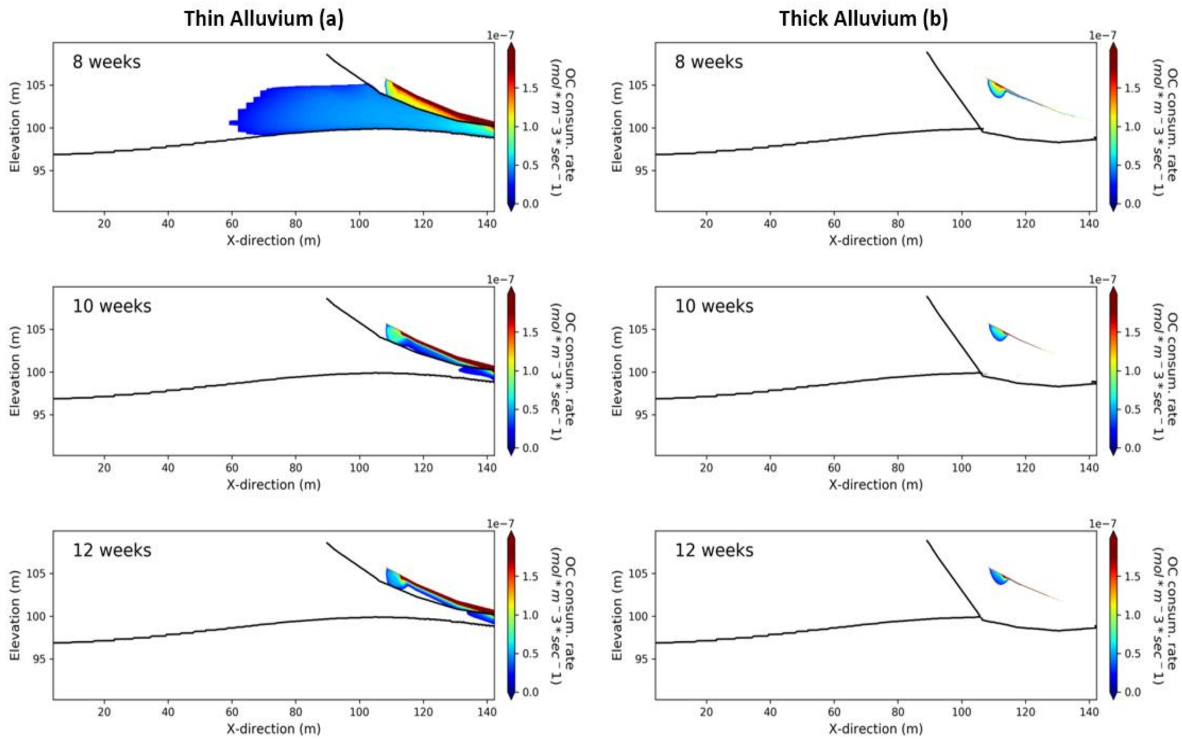


Figure 5-6 OC consumption rate at the three different simulation times under scenario 1 using two distinct thicknesses of the alluvium layer. The homogeneous formation and heat transport process were used.

5. Application to a Complex Biogeochemical Model

5.3 Sensitivity Results of VBSA

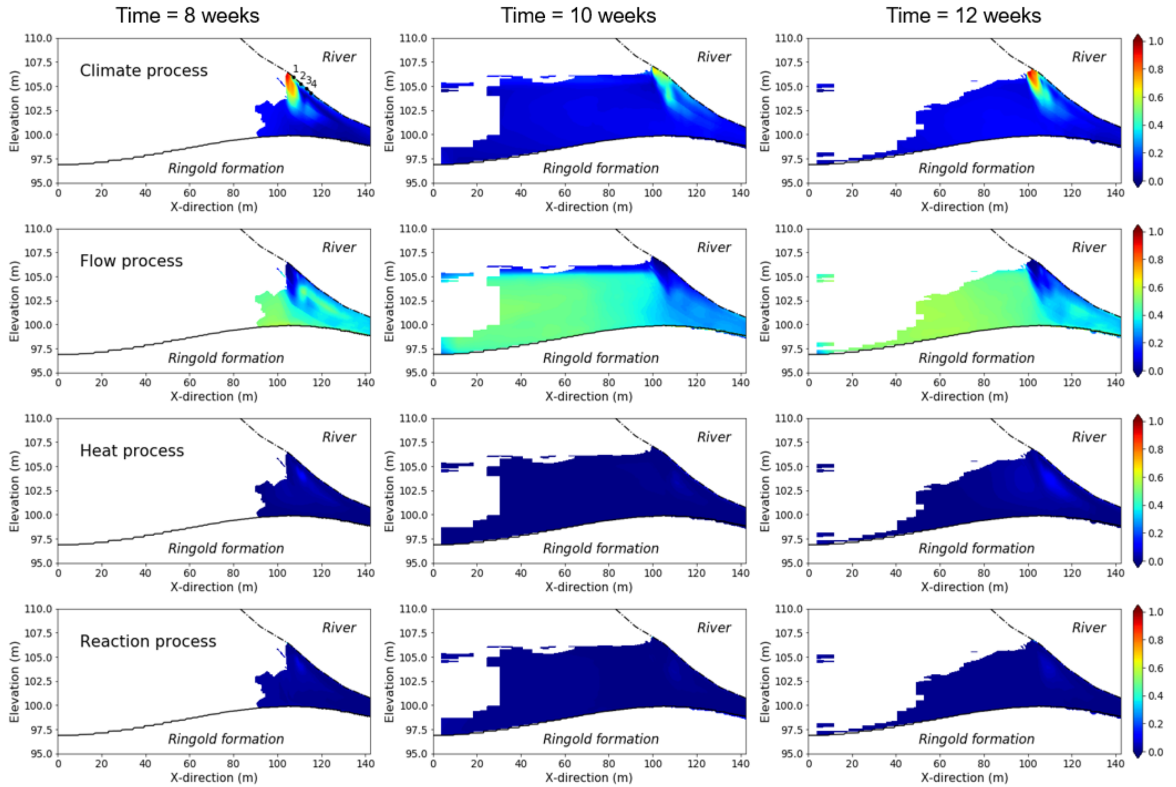


Figure 5-7 Spatial distribution of first-order process sensitivity analysis indices for climate, flow, heat, and reaction processes at the simulation times of 8, 10, and 12 weeks. The four rows and three columns are corresponding to the four processes and three simulation times, respectively.

5. Application to a Complex Biogeochemical Model

5.3 Sensitivity Results of VBSA

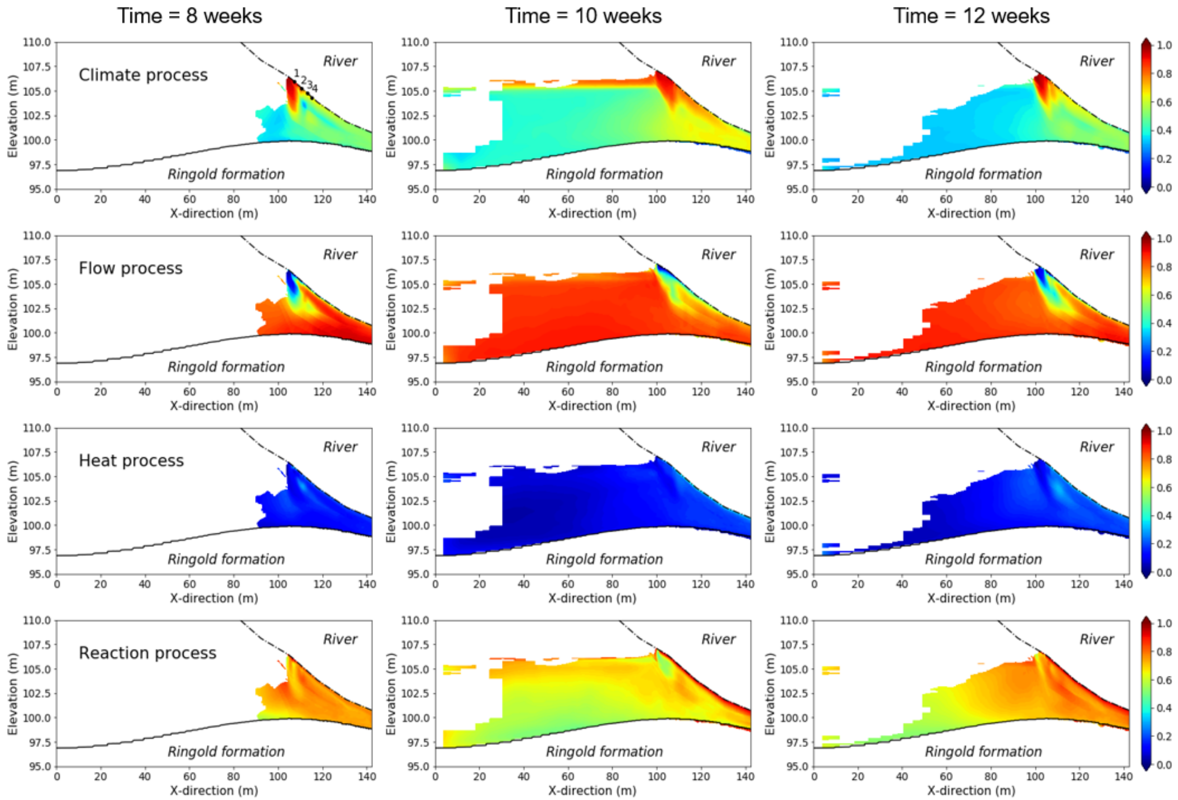


Figure 5-8 Spatial distribution of total-effect process sensitivity analysis indices for climate scenario, groundwater flow, heat transport, and reactive transport processes at the simulation times of 8, 10, and 12 weeks.

5. Application to a Complex Biogeochemical Model

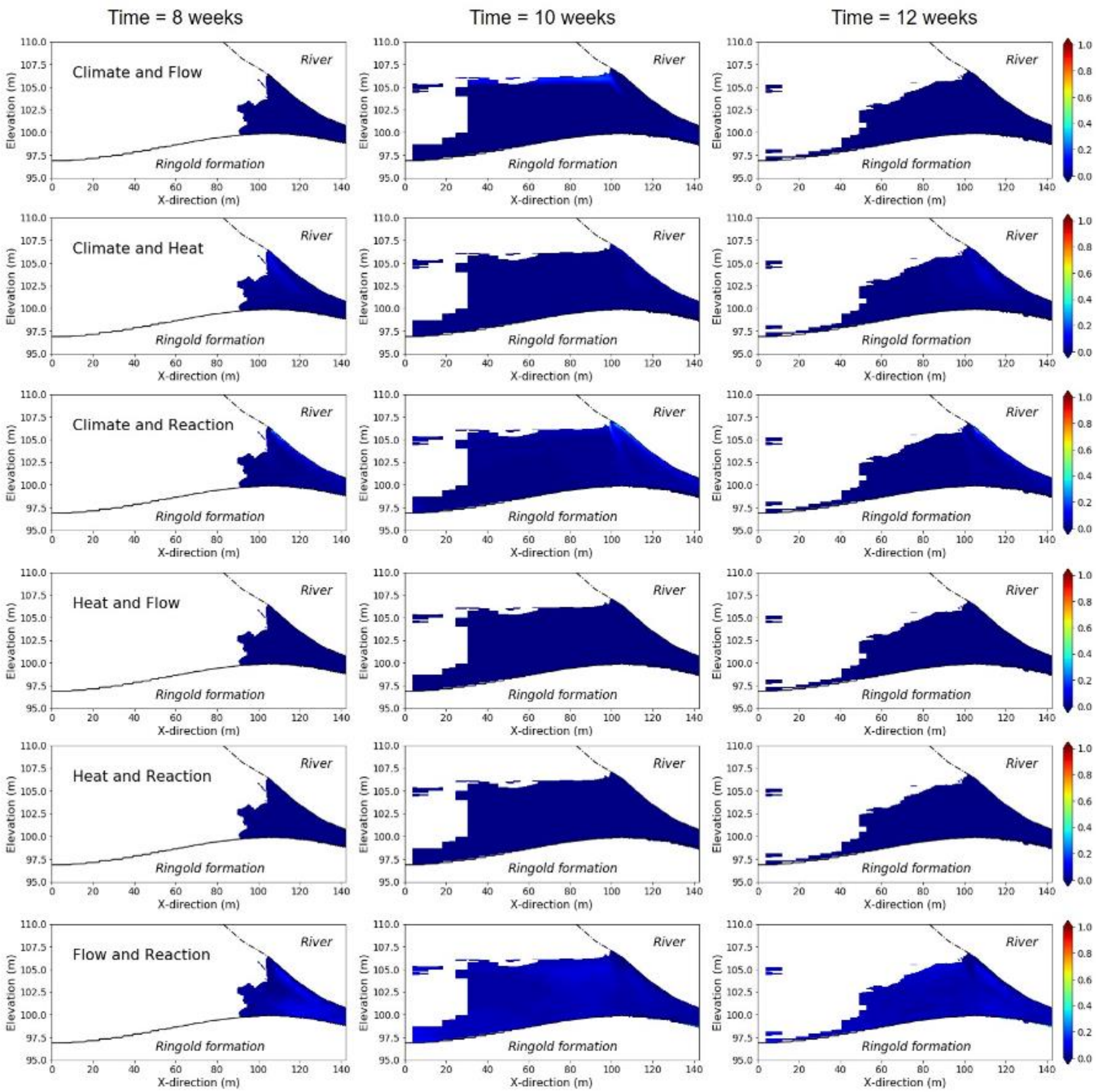


Figure 5-9 Spatial distribution of second-order process sensitivity analysis indices between climate scenario and groundwater flow, climate scenario and heat transport, climate scenario and reactive transport, heat transport and groundwater flow, and groundwater flow and reactive transport at the simulation times of 8, 10, and 12 weeks.

5. Application to a Complex Biogeochemical Model

5.3 Sensitivity Results of VBSA

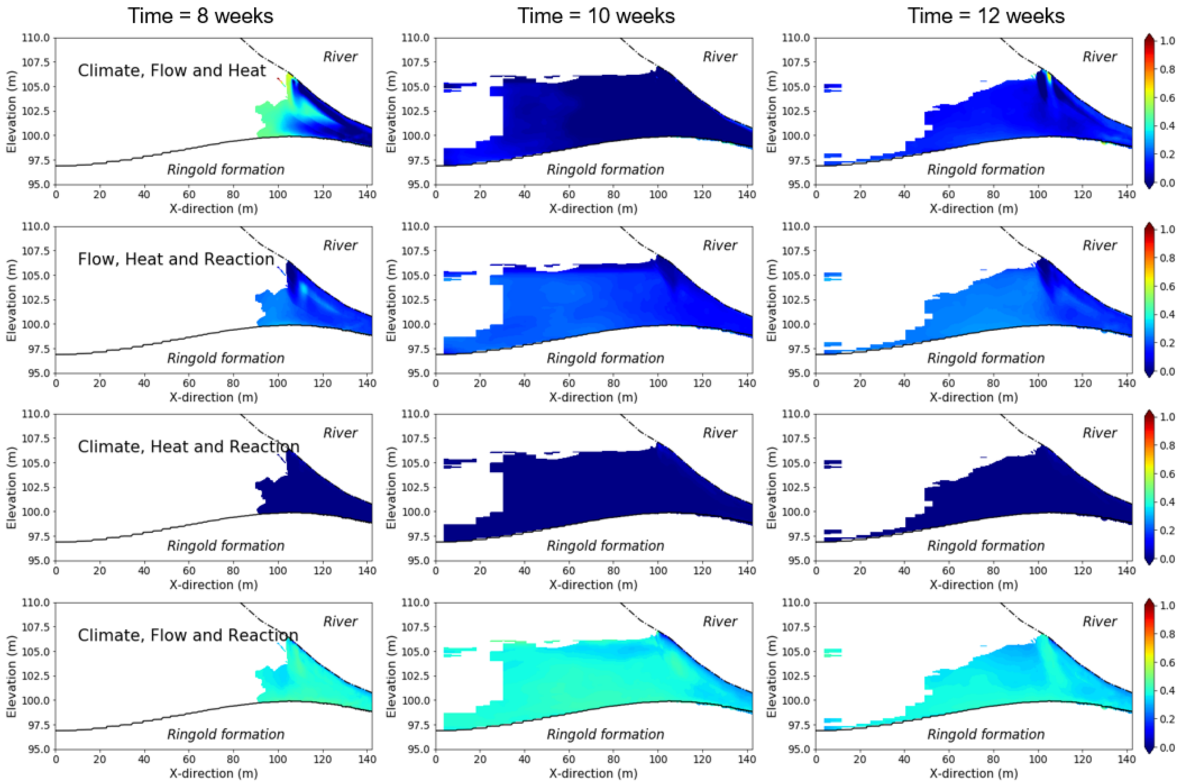


Figure 5-10 Spatial distribution of third-order process sensitivity analysis indices for climate scenario, groundwater flow, heat transport, and reactive transport processes at the simulation times of 8, 10, and 12 weeks.

5. Application to a Complex

5.3 Sensitivity Results of MMADS

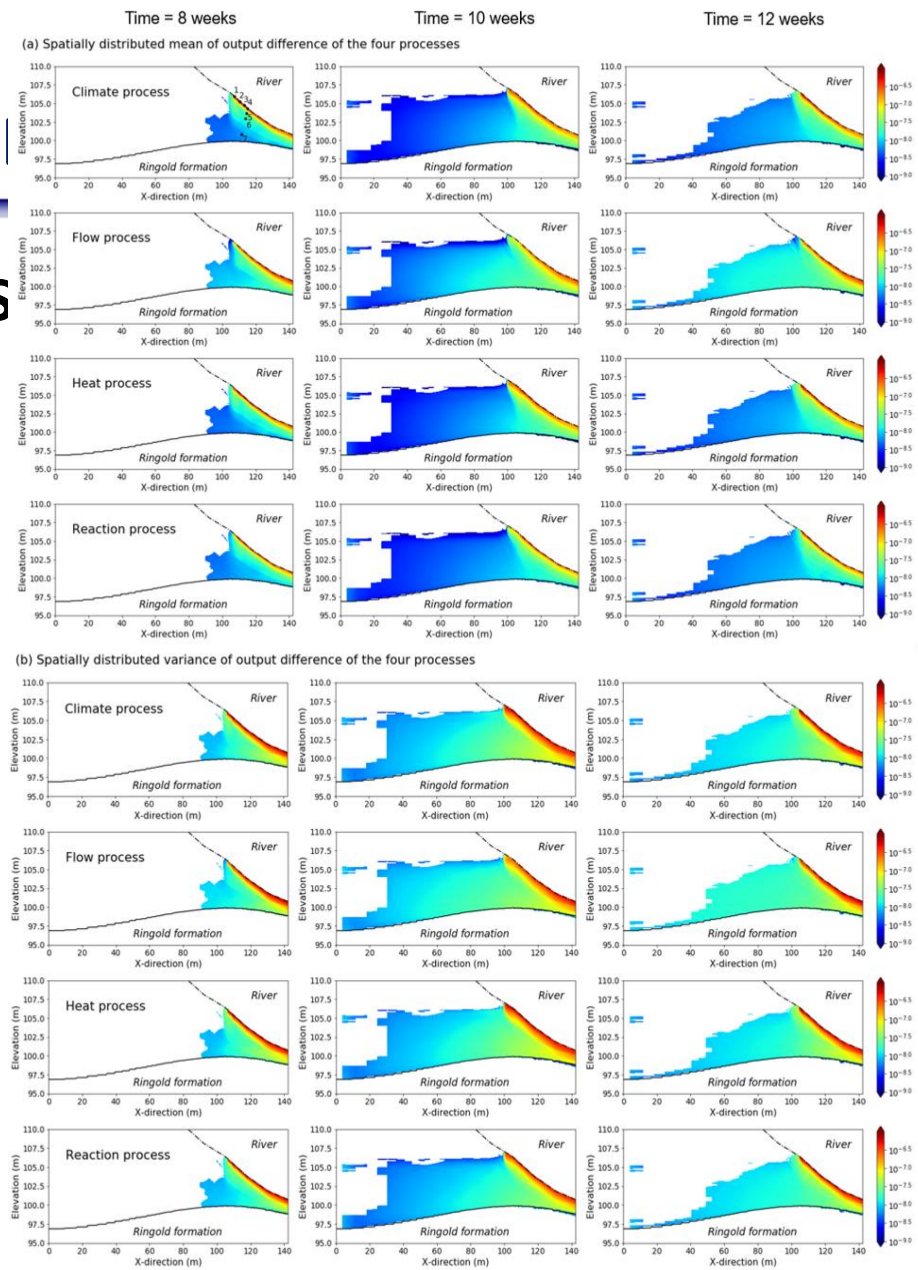


Figure 5-11 Spatial distribution of the two sensitivity analysis measures of (a) mean and (b) variance of the output difference of climate scenario (P_S), groundwater flow (P_F), heat transport (P_H), and reactive transport processes (P_R) to the OC consumption rate at simulation times of 8, 10, 12 weeks. Black dotted line denotes the river-sediment interface and dashed line denotes the upper bound of the Ringold formation.

5. Application to a Complex Biogeochemical Model

5.4 Comparison of the two methods

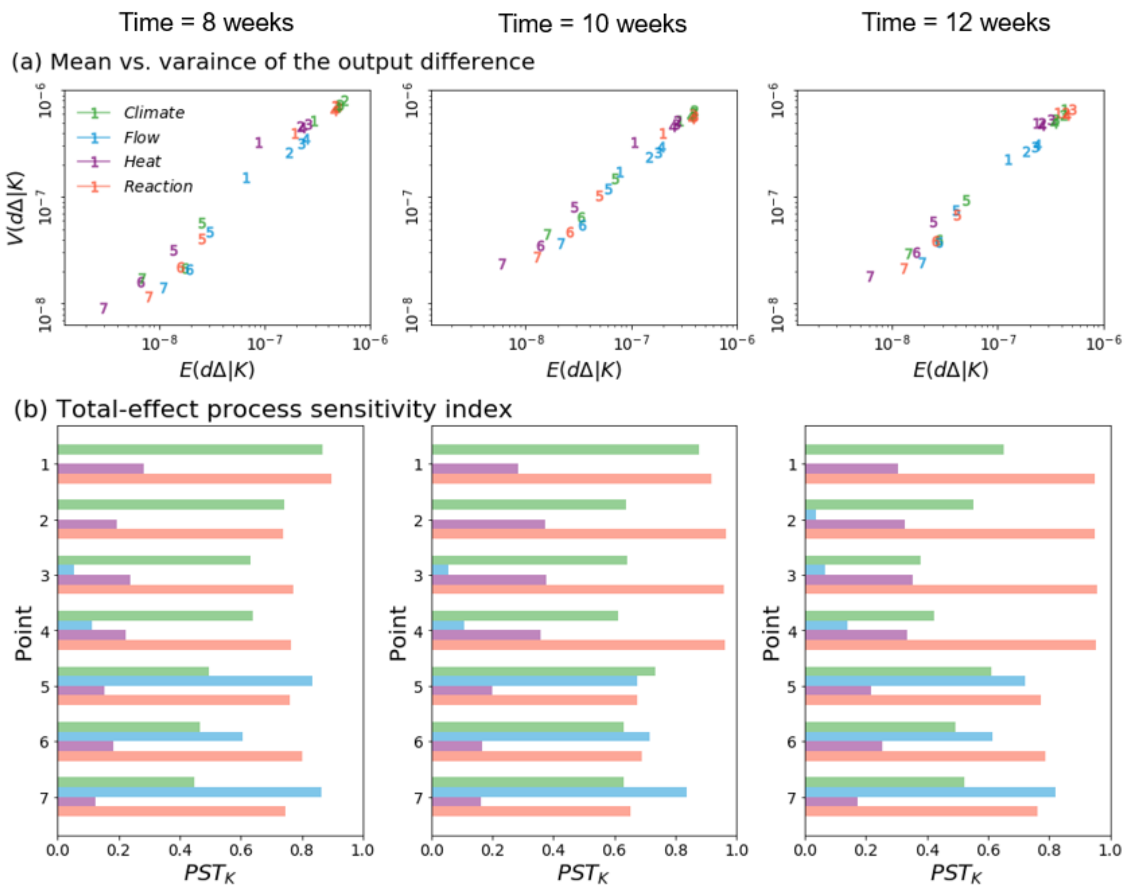


Figure 5-12 Comparison of the total-effect process sensitivity indices and the results of MMDS for the seven selected points

5. Application to a Complex Biogeochemical Model

5.4 Comparison of the two methods

Simulation time	8 weeks	10 weeks	12 weeks
Number of cells computed	37,664	55,157	52,008
	Percentage of highest PS_K (%)		
Climate process	11.24	14.83	11.85
Flow process	88.31	84.88	87.74
Heat process	0.00	0.00	0.15
Reaction Process	0.45	0.29	0.26
	Percentage of lowest PST_K (%)		
Climate process	0.18	0.11	0.20
Flow process	1.83	1.34	0.96
Heat process	97.99	98.55	98.84
Reaction Process	0.00	0.00	0.00
	Percentage of lowest $E(d\Delta\|K)$ and $V(d\Delta\|K)$ (%)		
Climate process	0.15	0.10	0.09
Flow process	0.59	0.82	0.75
Heat process	96.29	97.76	96.55
Reaction Process	0.00	0.00	0.00

Table of Contents 目录

6. SAMMPy: An Open-source Python Package

6.1 Structural of SAMMPy

- analyze
  - mmds.py
  - vbsa.py
- examples
  - workflow_1D_groundwater_flow.ipynb
  - workflow_Sobol_G_star.ipynb
- plotting
  - bar.py
  - dotty.py
  - hist.py
- util
  - results.py
- __init__.py
- LICENCE
- README.MD
- setup.py

Yang et al., 2021
(under preparation)

```
>>>import sammpy as sm

>>>model = sm.model()

>>>model.name = 'gwmodel'

>>>model.frames = {'names' : ['rechrg', 'geol', 'snomlt’],
                   'options': [['rechrg_lin', 'rechrg_power’],
                               ['geol_single', 'geol_double’],
                               ['snomlt_degree', 'snomlt_restrcd']],
                   'weights': [[0.5, 0.5],
                               [0.5, 0.5],
                               [0.5, 0.5]]}
```

6. SAMMPy: An Open-source Python Package

6.1 Structural of SAMMPy

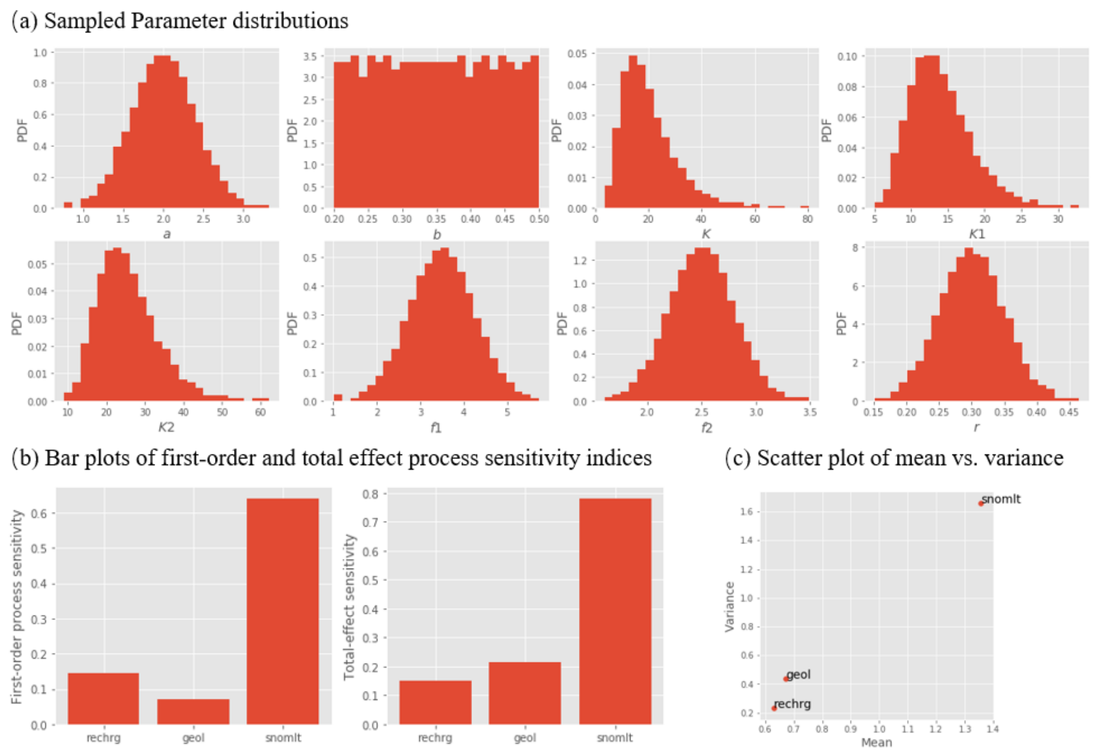


Figure 6-2 Examples of visualization tools implemented in the SAMMPy

6. SAMMPy: An Open-source Python Package

6.2 Software Available and Maintenance

Name of software: SAMMPy

Programming language: Python

Operating system: Windows, Linux and MacOS

Availability: https://github.com/jyangfsu/SAMMPy GitHub开源平台

Cost: Free for non-commercial academic research

Table of Contents 目录

7. Conclusions and Future Work

7.1 Thesis Summary

In this research, we developed two different methods, namely, the variance-based process sensitivity analysis method and the multi-model absolute difference-based sensitivity analysis method, to identify controlling processes in hydrologic modeling under process model and parameter uncertainty. We also presented a new Python package called SAMMPy: Sensitivity Analysis for Multiple Models, to evaluate the sensitivity indices of the two new methods.

The controlling processes are defined in the context of Factors Prioritization and Factors Fixing as discussed in Saltelli et al. (2004). For Factors Prioritization, the purpose is "make a rational bet on what is the factor that one should fix to achieve the greatest reduction in the uncertainty of the output". For Factors Fixing, the purpose is to "screen the input factors by identifying factors or sets of factors that are non-influential".

The developed methods were applied to two hydrologic models for identifying the controlling processes. The first one is a 2-D arsenic (As) sorption and reactive transport model based a laboratory experiment by Duan et al. (2020). The second one is 2-D complex biogeochemical model at Hanford's 300 Area focusing on spatio-temporal distribution of the organic carbon (OC) consumption rate in the aquifer.

7. Conclusions and Future Work

7.1 Thesis Summary

All the sensitivity analysis methods are implemented by developing python codes, and the codes are in a software called SAMMPy: a python package for process sensitivity analysis under multiple models.

7. Conclusions and Future Work

7.2 Thesis Contributions

The main contributions of this thesis are:

(1) The proposal of the two new multi-model global sensitivity analysis methods, which extends the traditional parameter sensitivity analysis methods for a single model and can be used to identify the controlling processes in hydrologic modeling with the presence of process model and parameter uncertainty. 两个方法

(2) The application of the two hydrologic models, which introduces the concept of multiple working hypotheses, evidences the need to include model uncertainty, and demonstrates how uncertainty estimates and sensitivity analysis can be combined to guide the identification of controlling processes, in hydrologic modeling to improve system robustness and reliability. 两个应用

(3) The development of new open-source Python package SAMMPy, which facilities the sensitivity analysis under multiple models. 一个软件

7. Conclusions and Future Work

7.3 Future Work

There are a number of possible extensions to the research presented in this thesis:

(1) The first is to incorporate observation data D to the current methods. In this thesis, the uncertainty of the parameters is characterized by using the prior distributions and the model weights of the plausible process models are set to be equally weighted. 整合观测数据

(2) A related issue is to link the data-worth analysis (or value-of-information) to the current methods. A major benefit of new data is its potential to help improve one's understanding of the system, in large part through a reduction in model predictive uncertainty and corresponding risk of failure (Neuman et al. 2012). 数据价值分析

(3) One last problem discussed here is to further break the computational barrier. 计算屏障

中國地質大學
CHINA UNIVERSITY OF GEOSCIENCES

博士学位论文答辩

谢谢！
恳请各位专家老师批评指正！

中國地質大學
China University Of Geosciences
环境学院
school of environmental studies