

**Working Report: Do Not Cite or Quote**

**Efficient Sensitivity/Uncertainty Analysis Using the  
High Dimensional Model Representation (HDMR) Method:  
Application to the Princeton Groundwater Model**

by

Genyuan Li, Herschel Rabitz

Department of Chemistry, Princeton University  
Princeton, NJ 08544

Sookyun Wang, Peter Jaffe

Department of Civil and Environmental Engineering, Princeton University  
Princeton, NJ 08544

Sheng-Wei Wang, Panos G. Georgopoulos

Computational Chemodynamics Laboratory  
Environmental and Occupational Health Sciences Institute (EOHSI)  
UMDNJ – R.W. Johnson Medical School & Rutgers, The State University of NJ  
170 Frelinghuysen Road, Piscataway, NJ 08854

CERM/CRESP-CEEA Technical Report CCL-TR-01.01/Working Version

CERM, the Center for Exposure and Risk Modeling at EOHSI, is funded by USEPA NERL  
CRESP-CEEA, the Center of Expertise in Exposure Assessment  
of the Consortium for Risk Evaluation with Stakeholder Participation,  
is funded by USDOE through IRM, the Institute for Responsible Management

January 7, 2002

# Contents

<b>Table of Contents</b>	<b>i</b>
<b>List of Figures</b>	<b>ii</b>
<b>List of Tables</b>	<b>iii</b>
<b>1 INTRODUCTION</b>	<b>1</b>
<b>2 THE PRINCETON GROUNDWATER MODEL</b>	<b>3</b>
<b>3 THE HIGH DIMENSIONAL MODEL REPRESENTATION (HDMR) METHOD</b>	<b>6</b>
<b>4 UNCERTAINTY ANALYSIS RESULTS AND DISCUSSION</b>	<b>9</b>
4.1 Results . . . . .	9
4.2 Discussion . . . . .	10

## List of Figures

1	The average value, $\sigma$ and $\sqrt{\sum \sigma_i^2}$ of output 2 at different locations. . . . .	22
2	The $\sigma_i$ of output 2 for inputs 1 to 7. . . . .	23
3	The $\sigma_i$ of output 2 for inputs 8 to 13. . . . .	24
4	The average value, $\sigma$ and $\sqrt{\sum \sigma_i^2}$ of output 3 at different locations. . . . .	25
5	The $\sigma_i$ of output 3 for inputs 1 to 7. . . . .	26
6	The $\sigma_i$ of output 3 for inputs 8 to 13. . . . .	27
7	The average value, $\sigma$ and $\sqrt{\sum \sigma_i^2}$ of output 4 at different locations. . . . .	28
8	The $\sigma_i$ of output 4 for inputs 1 to 7. . . . .	29
9	The $\sigma_i$ of output 4 for inputs 8 to 13. . . . .	30

## List of Tables

1	Ranges of the chosen input variables . . . . .	11
2	List of the output variables . . . . .	12
3	Convergence of the mean value $f_0$ for output 1 . . . . .	13
4	Differences between $\sigma$ and $\sqrt{\sum \sigma_i^2}$ for output 1 . . . . .	14
5	Order of standard deviations $\sigma_i$ for output 1 related to different input variables and different data sizes . . . . .	15
6	Differences between $\sigma$ and $\sqrt{\sum \sigma_i^2}$ for output 2 at 2 meters . . . . .	16
7	Order of standard deviations $\sigma_i$ for output 2 at 2 meters related to different input variables and different data sizes . . . . .	17
8	Differences between $\sigma$ and $\sqrt{\sum \sigma_i^2}$ for output 3 at 2 meters . . . . .	18
9	Order of standard deviations $\sigma_i$ for output 3 at 2 meters related to different input variables and different data size . . . . .	19
10	Differences between $\sigma$ and $\sqrt{\sum \sigma_i^2}$ for output 4 at 2 meters . . . . .	20
11	Order of standard deviations $\sigma_i$ for output 4 related to different input variables and different data sizes . . . . .	21

# 1 INTRODUCTION

Uncertainty/sensitivity analysis of complex environmental models can identify the effects of model input variations and uncertainties on the model outputs, and distinguish important input parameters from unimportant ones. Therefore, the outcomes of such an analysis can provide guidance for specific research needs and additional laboratory or field studies to reduce model prediction uncertainties. Traditional methods for characterizing/propagating uncertainty through models, such as standard Monte Carlo methods [1], and efficient modifications of Monte Carlo, such as Latin Hypercube Sampling [2], require a large number of model runs, which, for comprehensive environmental models, is a task often not feasible even on state-of-the-art computers.

There are two major approaches in making uncertainty analysis more efficient (or, depending on the model, simply feasible):

- Reducing the number of required model runs by “optimizing” sampling of the inputs for which simulations are performed such as the Deterministic Equivalent Modeling Method (DEMM) [3] and the Stochastic Response Surface Method (SRSM) [4].<sup>1</sup>
- Replacing the original model with a “fast equivalent” one (also often called a “secondary” model), the simplest form being a classical response surface, so that the computational requirements are reduced.

For very complex and computationally demanding models, a combination of both approaches may actually be needed. A new method for systematic mathematical model reduction, the High Dimensional Model Representation (HDMR) method, can be useful tool in both frameworks mentioned above.

This report describes an application involving the HDMR method to perform efficient uncertainty analysis on the Princeton Groundwater Model [6]. The Princeton Groundwater Model, developed by Smith and Jaffe, is used here to simulate a bioremediation scheme of trace metals and radionuclides in groundwater system. The critical uncertain model parameters in this model involve 20 biogeochemical reaction rate constants. The objective of performing the global uncertainty analysis through the HDMR method is to quantify the contributions of model output (e.g., the uranium U(IV) concentration) variation from the “uncertain” variation of these 20 model input parameters. Thus, the model input parameters that most significantly influence the model output can be identified, and then future research needs can be focused on reducing the uncertainty in determining the most important model inputs. This report is organized as follows:

- Section 2 describes briefs of the Princeton Groundwater Model.

---

<sup>1</sup>These two approaches in general do not require any modifications of the original model code: they can treat it as a “black box”. However, when the code is available and can be used to deduce derivatives and often sensitivity-related information on model variables, this information can be used to significantly increase computational efficiency of the uncertainty analysis [5].

- Section 3 summarizes the principles of the High Dimensional Model Representation (HDMR) method.
- Section 4 presents the results of the global uncertainty analysis for the Princeton Groundwater Model.

## 2 THE PRINCETON GROUNDWATER MODEL

Mobility of trace metals and radionuclides in the groundwater can be controlled to a significant extent via in-situ manipulation of several biogeochemical reactions that directly or indirectly alter their solubility. Specified redox conditions, therefore, which affect their speciation, dissolution, and precipitation, need to be established in the subsurface. Predictive tools are needed to assess the distribution of microbially mediated redox reactions in groundwater systems in response to the availability and transport of different electron acceptors, and their utilization by different bacteria during the degradation of an organic substrate in order to evaluate trace metal/radionuclides bioremediation schemes as well as their fate and transport.

For this purpose, a time-dependent one-dimensional reaction transport model has been developed by Smith and Jaffe [6]. The fate and transport of many trace metals and radionuclides in porous media is closely linked to the biogeochemical reactions that occur as a result of organic carbon being degraded by different microorganisms using a series of terminal electron acceptors such as  $O_2$ ,  $NO_3^-$ ,  $Mn(IV)$ ,  $Fe(III)$  and  $SO_4^{2-}$ . Throughout the redoxcline that develops in such environments, trace metals can be mobilized/immobilized via processes such as reduction/oxidation, sorption/desorption, precipitation/dissolution, and/or the formation of complex ions. It is possible, therefore, to represent the fate and transport of trace metals when every related species in the environments of concern, such as organic substrates, terminal electron acceptors and corresponding reduced species as well as contaminant metals/radionuclides of interest are considered simultaneously. In this model, generic mass balance equations are employed to represent dissolved and solid species in subsurface environments:

- dissolved species:

$$\frac{\partial}{\partial t} \left[ \phi(1 + K_i^{eff}) \right] C_i^{aq} = D_{hi} \frac{\partial^2 \phi}{\partial x^2} C_i^{aq} - V \frac{\partial \phi}{\partial x} C_i^{aq} + \sum R_i, \quad (1)$$

- solid species:

$$\frac{\partial}{\partial t} \left[ (1 - \phi) C_i^s \right] = \sum R_i. \quad (2)$$

where

- $t$  and  $x$  are time and the distance parallel to groundwater flow from the origin,
- $\phi$  is the porosity and  $K_i^{eff}$  is the effective partition coefficient for the equilibrium adsorption of species  $i$ ,
- $C_i^{aq}$  and  $C_i^s$  are the concentrations of the dissolved and solid species  $i$ , respectively,
- $V$  is the velocity of groundwater flow and  $D_{hi}$  is the coefficient of hydrodynamic dispersion (mechanical dispersion + molecular diffusion) of species  $i$ , and

- $\sum R_i$  is the sum of consumption/production of species  $i$  by biotic/abiotic reactions.

Although the work presented here is generic and applies to a wide range of trace metals, metalloids, and radionuclides, we will focus the simulations specifically on uranium as a model contaminant.

Uranium is a radioactive chemical element that is an important source of nuclear power. Uranium wastes can be dissolved into groundwater, migrate along groundwater and contaminate drinking water. The dominant forms of uranium in oxic waters is U(VI) as in the oxide  $UO_3$  and the yellow uranyl ion  $UO_2^{2+}$  which is highly soluble and mobile, but under reducing conditions it is precipitated as uranite and coffinite in the form of U(IV) which is highly insoluble. It also exhibits a U(III) and a U(V) state, but the respective ions are unstable, easily deprotonated to U(IV) and U(VI). The mobilization/immobilization of uranium by oxidation/reduction between U(VI) and U(IV) can be caused by both biotic and abiotic reactions. Some  $Fe(III)^-$  and sulfate-reducing bacteria are reported to be capable of reducing and precipitating uranium under anaerobic conditions by using U(VI) as a terminal electron acceptor coupling with enzymatic-mediated oxidation of organic carbons. The energy yield from the reduction of U(VI) to U(IV) is -63.3 kcal/mol and to U(III) is -52.2 kcal/mol, which is twice than that of Fe(III) reduction. Uranium reduction should occur in the sequence of  $Mn(IV) > U(IV) > Fe(III)$ . U(VI) can be also reduced by the abiotic mechanism which is coupled to the oxidation of Fe(II) sorbed on iron corrosion products such as hematite. In addition to uranium reduction, oxidation of U(IV) to U(VI) in oxic condition was also considered. Although various researches have reported biotic/abiotic uranium reduction/oxidation, information on the rate of each reaction is very limited. Therefore, in the model formulations,

- lumped rate coefficients for uranium reduction and oxidation are employed,
- it is assumed that enzymatic uranium reduction occurs simultaneously with Fe(III)- and sulfate-reduction after Mn(IV) is used up by biotic/abiotic reactions, and
- the sums of consumption/production of U(VI) and U(IV) can be expressed by

$$\sum R_{U6} = k_{ox,U4}(1 - \phi)C_0C_{U4} - \chi_{Fe3}k_{red,U6}C_{U6}, \quad (3)$$

$$\sum R_{U4} = \chi_{Fe3}k_{red,U6}\frac{\phi}{1 - \phi}C_{U6} - k_{ox,U4}\phi C_0C_{U4}. \quad (4)$$

where

- $k_{ox,U4}$  and  $k_{red,U6}$  are the lumped rate coefficients for oxidation from U(IV) to U(VI) and reduction from U(VI) to U(IV), respectively,
- $C_0, C_{U6}$ , and  $C_{U4}$  are the dissolved concentrations of oxygen and U(VI) and the solid concentration of U(IV), and
- $\chi_{Fe3}$  is the indicator coefficient which is 1 when Fe(III) can be used as an electron acceptor and 0 when it is not utilized.

The dynamic model that is composed of series of mass balance equations for the species of interest is iteratively calculated by coupling with the chemical equilibrium model, MINTEQA2 [7].

### 3 THE HIGH DIMENSIONAL MODEL REPRESENTATION (HDMR) METHOD

A general set of quantitative model assessment and analysis tools, termed High Dimensional Model Representations (HDMR), have been introduced recently [8–12] for improving the efficiency of deducing high dimensional input-output system behavior. HDMR represents a model output  $g(\mathbf{x})$  as a finite hierarchical correlated function expansion in terms of the input variables  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ :

$$g(\mathbf{x}) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} f_{ijk}(x_i, x_j, x_k) + \dots + f_{12\dots n}(x_1, x_2, \dots, x_n), \quad (5)$$

where  $f_0$  is a constant representing the mean value of  $g(\mathbf{x})$  in the entire domain of  $\mathbf{x}$ , and  $f_i(x_i)$  gives the independent contribution to  $g(\mathbf{x})$  by the  $i$ th input variable  $x_i$ ,  $f_{ij}(x_i, x_j)$  gives the pair correlated contribution of the input variables  $x_i$  and  $x_j$ , and so on. The last term  $f_{12\dots n}(x_1, x_2, \dots, x_n)$  contains any residual  $n$ th order correlated contribution over all of the input variables. The component functions of the HDMR expansion are constructed optimally, and tailored to  $g(\mathbf{x})$  over the entire domain of  $\mathbf{x}$  such that the expansion converges rapidly (in many cases, to 2nd order correlated terms in Eq. (5), as various trial results indicate [11;12]), i.e., a satisfactory approximation of  $g(\mathbf{x})$  can be given by

$$g(\mathbf{x}) \approx f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j). \quad (6)$$

Various forms of HDMR have been considered with applications to a variety of scientific problems [9–12]. Here we focus on what has been referred to as RS-HDMR. Without loss of generality, every input variable  $x_i$  can have a range  $[0, 1]$  after some proper transformation. Let  $C^k$  represent the  $k$ -dimensional unit hypercube. The formulas of RS-HDMR for the component functions in Eq. (5) are as follows:

$$f_0 = \int_{C^n} g(\mathbf{x}) d\mathbf{x}, \quad (7)$$

$$f_i(x_i) = \int_{C^{n-1}} g(\mathbf{x}) d\mathbf{x}^i - f_0, \quad (i = 1, 2, \dots, n) \quad (8)$$

$$f_{ij}(x_i, x_j) = \int_{C^{n-2}} g(\mathbf{x}) d\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0, \quad (1 \leq i < j \leq n) \quad (9)$$

.....

where  $d\mathbf{x} = dx_1 dx_2 \dots dx_n$ , and  $d\mathbf{x}^i$ ,  $d\mathbf{x}^{ij}$  are just  $d\mathbf{x}$  without elements  $dx_i$  and  $dx_i dx_j$ , respectively.

The component functions of RS-HDMR are mutually orthogonal to permit the model output variance  $\sigma_g^2$  to be decomposed into its input variable statistical contributions due to

the independent action of the variables  $\sigma_i^2$ , the pair correlated action  $\sigma_{ij}^2$ , etc.

$$\begin{aligned}
 \sigma_g^2 &= \int_{C^n} [g(\mathbf{x}) - \bar{g}]^2 d\mathbf{x} = \int_{C^n} [g(\mathbf{x}) - f_0]^2 d\mathbf{x} \\
 &= \sum_{i=1}^n \int_0^1 f_i^2(x_i) dx_i + \sum_{1 \leq i < j \leq n} \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j + \dots \\
 &= \sum_{i=1}^n \sigma_i^2 + \sum_{1 \leq i < j \leq n} \sigma_{ij}^2 + \dots
 \end{aligned} \tag{10}$$

The determination of  $\sigma_i^2$ ,  $\sigma_{ij}^2$ , ... requires the evaluation of the integrals  $\int f_i^2(x_i) dx_i$ ,  $\int \int f_{ij}^2(x_i, x_j) dx_i dx_j$ , etc. Monte Carlo integration is used for these evaluations [13]. For instance,  $N$  sets of  $n$ -dimensional vector  $\{x_1^s, x_2^s, \dots, x_n^s\}$  ( $s = 1, 2, \dots, N$ ) are randomly generated to yield

$$f_0 = \int_{C^n} g(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{s=1}^N g(\mathbf{x}^s), \tag{11}$$

$$\sigma_g^2 = \int_{C^n} [g(\mathbf{x}) - f_0]^2 d\mathbf{x} \approx \frac{1}{N} \sum_{s=1}^N g^2(\mathbf{x}^s) - f_0^2. \tag{12}$$

When  $N \rightarrow \infty$ , the limit values of  $f_0$ ,  $\sigma_g^2$  can be obtained. Very often the integrals converge sufficiently fast for modest values of  $N$ .

The functions  $f_i(x_i)$ ,  $f_{ij}(x_i, x_j)$ , ... can be approximately represented as linear combinations of orthogonal polynomials  $\varphi_k(x_i)$ ,

$$g(\mathbf{x}) \approx f_0 + \sum_{i=1}^n \sum_{k=1}^r c_k^i \varphi_k(x_i) + \sum_{1 \leq i < j \leq n} \sum_{k=1}^r \sum_{l=1}^r c_{kl}^{ij} \varphi_k(x_i) \varphi_l(x_j) + \dots \tag{13}$$

where

$$\varphi_1(x) = a_1 x + a_2, \tag{14}$$

$$\varphi_2(x) = b_1 x^2 + b_2 x + b_3, \tag{15}$$

$$\varphi_3(x) = c_1 x^3 + c_2 x^2 + c_3 x + c_4, \tag{16}$$

.....

are orthogonal polynomials with zero mean and unit norm. Their coefficients are determined by an optimization procedure for different random sample sizes to assure their orthogonality. In most cases, acceptable accuracy may be achieved using  $\varphi_1(x)$ ,  $\varphi_2(x)$  and  $\varphi_3(x)$  (i.e.,  $r \leq 3$ ).

Since  $\varphi_k(x_i)$  are mutually orthogonal for all  $i$  and  $k$ , the coefficients  $c_k^i$  and  $c_{kl}^{ij}$  can be determined by Monte Carlo sampling as

$$c_k^i \approx \frac{1}{N} \sum_{s=1}^N g(\mathbf{x}^s) \varphi_k(x_i^s), \tag{17}$$

$$c_{kl}^{ij} \approx \frac{1}{N} \sum_{s=1}^N g(\mathbf{x}^s) \varphi_k(x_i^s) \varphi_l(x_j^s). \tag{18}$$

Then,  $\sigma_i^2$  and  $\sigma_{ij}^2$  are simply the sums of  $(c_k^i)^2$  for all  $k$ , and the sums of  $(c_{kl}^{ij})^2$  for all  $k$  and  $l$ , respectively.

$$\sigma_i^2 = \sum_{k=1}^r (c_k^i)^2, \quad (19)$$

$$\sigma_{ij}^2 = \sum_{k=1}^r \sum_{l=1}^r (c_{kl}^{ij})^2. \quad (20)$$

Notice that only one randomly sampled set of  $g(\mathbf{x}^s)$  is needed to determine  $f_0$ ,  $f_i(x_i)$ ,  $f_{ij}(x_i, x_j), \dots$  and consequently,  $\sigma_{\bar{g}}^2$ ,  $\sigma_i^2$ ,  $\sigma_{ij}^2$ , etc, and the assessments are valid for the whole domain of  $\mathbf{x}$ . This makes global uncertainty assessments of a model very efficient. Moreover, RS-HDMR provides not only quantitative assessments  $\sigma_{\bar{g}}^2$ ,  $\sigma_i^2$ ,  $\sigma_{ij}^2, \dots$ , but also the qualitative behaviors of the independent and collective actions  $f_i(x_i)$ ,  $f_{ij}(x_i, x_j), \dots$  of the input variables on the output. Hence, the uncertainty assessment given by RS-HDMR is valuable for attaining a physical understanding of the origins of output uncertainty as well as suggestions for modification of the model and additional laboratory or field studies to best improve the quality of the model.

## 4 UNCERTAINTY ANALYSIS RESULTS AND DISCUSSION

A series of simulations based on hypothetical bioremediation scenarios were performed to illustrate the effect of biostimulation on the fate and transport of uranium in the subsurface and to identify the effects of various biological and/or chemical variables in the model. 300 random samples were taken from a 20 dimensional unit hypercube, and then converted to 300 sets of model inputs consisting of 20 biogeochemical reaction rate constants based on their plausible ranges of variations, as given in Table 1. These 20 reaction rate constants include maximum rates of oxidation by 8 electron acceptors, and rate constants for 12 second order reactions in the model. 300 random samples for 4 different outputs given in Table 2 were collected and then analyzed to identify the key input parameters among the 20 input parameters.

### 4.1 Results

The standard deviation  $\sigma$  caused by all input variables is calculated using Eq. (12) for different sizes of random data sets (100, 200 and 300). The square root of the sum of the variances  $\sqrt{\sum \sigma_i^2}$  contributed by each input variable, which is obtained from the first order RS-HDMR component functions, is also calculated by using Eqs. (17) and (19). If the difference between these two numbers for a given output variable is small, then the cooperative influences of the input variables either do not exist or are very small.

#### Output 1

The results for output 1 are given in Table 3 to 5. The linear, quadratic and cubic orthogonal polynomials are used for the approximation of the 1st order RS-HDMR component functions. The results show that the sample with 300 points gives satisfactory convergence for both the mean value and the standard deviation. Since the difference between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  is very small, the cooperative effects of the inputs on this model output do not exist or are very small.

#### Output 2

For output 2, the uncertainty assessment was performed for different distances (0-30 meters). One example, at 2 meters, is given in Tables 6 and 7. Similarly, the linear, quadratic and cubic orthogonal polynomials are used for the approximation of the 1st order RS-HDMR component functions. There are notable differences between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$ . This might be caused by either the existence of the cooperative influence of the input variables or the small sample sizes. The  $f_0$ ,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  determined by the 300 points sample at different locations are given in Figure 1. The standard deviation  $\sigma_i$  at different locations for all 13 input variables are given in Figures 2 and 3.

#### Output 3

For output 3 the uncertainty assessment was performed for different distances (0-30 meters). One example, at 2 meters, is given in Tables 8 and 9. The linear, quadratic and cubic orthogonal polynomials are used for the approximation of the 1st order RS-HDMR component functions. The  $f_0$ ,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  determined by the 300 points sample at different locations are given in Figure 4. The standard deviation  $\sigma_i$  at different locations for all 13 input variables are given in Figures 5 and 6.

## Output 4

For output 4 the uncertainty assessment was performed for different distances (0-30 meters). One example, at 2 meters, is given in Tables 10 and 11. The linear, quadratic and cubic orthogonal polynomials are used for the approximation of the 1st order RS-HDMR component functions. For output 4, the  $f_0$ ,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  at different locations are given in Figure 7. The  $\sigma_i$  at different locations for all 13 input variables are given in Figures 8 and 9.

## 4.2 Discussion

The above results show that the rate coefficient of uranium reduction (input variable 2),  $k_{red,U6}$ , has the most significant effect on all four outputs. After that, nitrate-reduction, manganese-reduction and oxygen-reduction rate coefficients,  $\mu_{C,N}$ ,  $\mu_{C,M}$  and  $\mu_{C,O}$  have bigger effects on uranium outputs. Because of the sequential uptake of terminal electron acceptor in the order of thermodynamic energy yield, uranium reduction takes place after manganese-reduction and can be affected by proceeding reactions. Even though oxygen-reduction begins before other reduction reactions, it has an effect less than nitrate- and manganese-reduction because of the fast oxygen depletion right after its introduction.

*This work has demonstrated that a highly efficient and thorough sampling of the overall model input variable space. for global uncertainty assessments, can be achieved using the HDMR approach. The results show the significant potential of this approach in characterizing and reducing model prediction uncertainty.*

**Table 1:** Ranges of the chosen input variables

	Input variables	Description	Ranges ( $M^{-1}yr^{-1}$ )
1	koxU <sub>4</sub>	uranium oxidation rate constant	$3.0 \times 10^6 - 1.2 \times 10^7$
2	kredU <sub>6</sub>	uranium reduction rate constant	65–260
3	umAO	oxygen/ammonia oxidation rate constant	15–60
4	umMeO	oxygen/methane oxidation rate constant	15–60
5	umCO	oxygen/organic carbon oxidation rate constant	0.5–2
6	umCN	nitrate/organic carbon oxidation rate constant	0.02–0.08
7	umCM	Mn(IV)/organic carbon oxidation rate constant	0.005–0.02
8	umCF	Fe(III)/organic carbon oxidation rate constant	0.0025–0.01
9	umCS	sulfate/organic carbon oxidation rate constant	0.085–0.34
10	umCMe	methane/organic carbon oxidation rate constant	0.025–0.1
11	umNO <sub>3</sub> Fe	2nd-order redox reaction rate ( $NO_3^-/Fe^{2+}$ )	800–3200
12	umOFe <sub>2</sub>	2nd-order redox reaction rate ( $O_2(aq)/Fe^{2+}$ )	$1.05 \times 10^7 - 4.2 \times 10^7$
13	umOMn <sub>2</sub>	2nd-order redox reaction rate ( $O_2(aq)/Mn^{2+}$ )	2.3–9.2
14	umOS <sub>2</sub>	2nd-order redox reaction rate ( $O_2(aq)/HS^{-1}$ )	$1.0 \times 10^5 - 4.0 \times 10^5$
15	umSO <sub>4</sub> CH <sub>4</sub>	2nd-order redox reaction rate ( $SO_4^{-1}/CH_4$ )	5000–20000
16	umOadsMn <sub>2</sub>	2nd-order redox reaction rate ( $O_2(aq)/Mn^{2+}(ads)$ )	$1.05 \times 10^7 - 4.2 \times 10^7$
17	umOadsFe <sub>2</sub>	2nd-order redox reaction rate ( $O_2(aq)/Fe^{2+}(ads)$ )	$2.5 \times 10^6 - 1.0 \times 10^7$
18	umMn <sub>4</sub> S <sub>2</sub>	2nd-order redox reaction rate ( $MnO_2/HS^{-}$ )	$4.0 \times 10^5 - 1.6 \times 10^6$
19	umMn <sub>4</sub> Fe <sub>2</sub>	2nd-order redox reaction rate ( $MnO_2/Fe^{2+}$ )	5000–20000
20	umFe <sub>3</sub> S <sub>2</sub>	2nd-order redox reaction rate ( $FeOOH/HS^{-}$ )	500–2000

**Table 2:** List of the output variables

Order #	Output variable
1	Total precipitation of U(VI) over entire simulation domain
2	Total precipitation of U(VI) within distances $x = 0 - 30$ m
3	Accumulative flux of U(VI) passing through $x = 0 - 30$ m
4	Concentration of U(VI) at $x = 0 - 30$ m

**Table 3:** Convergence of the mean value  $f_0$  for output 1

Mean value	Data size
$0.138650 \cdot 10^{-3}$	100
$0.138164 \cdot 10^{-3}$	200
$0.137543 \cdot 10^{-3}$	300

**Table 4:** Differences between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  for output 1

$\sigma$	$\sqrt{\sum \sigma_i^2}$	Data size
$0.263442 \cdot 10^{-4}$	$0.261378 \cdot 10^{-4}$	100
$0.257671 \cdot 10^{-4}$	$0.247393 \cdot 10^{-4}$	200
$0.273609 \cdot 10^{-4}$	$0.249839 \cdot 10^{-4}$	300

**Table 5:** Order of standard deviations  $\sigma_i$  for output 1 related to different input variables and different data sizes

100		200		300	
Input	$\sigma_i$	Input	$\sigma_i$	Input	$\sigma_i$
2	$0.159995 \cdot 10^{-4}$	2	$0.153508 \cdot 10^{-4}$	2	$0.171973 \cdot 10^{-4}$
6	$0.127834 \cdot 10^{-4}$	6	$0.113977 \cdot 10^{-4}$	11	$0.111347 \cdot 10^{-4}$
11	$0.996586 \cdot 10^{-5}$	11	$0.109497 \cdot 10^{-4}$	6	$0.898102 \cdot 10^{-5}$
7	$0.655194 \cdot 10^{-5}$	5	$0.841448 \cdot 10^{-5}$	5	$0.895993 \cdot 10^{-5}$
12	$0.519160 \cdot 10^{-5}$	7	$0.392106 \cdot 10^{-5}$	13	$0.357343 \cdot 10^{-5}$
5	$0.504785 \cdot 10^{-5}$	1	$0.325716 \cdot 10^{-5}$	7	$0.344827 \cdot 10^{-5}$
1	$0.450400 \cdot 10^{-5}$	13	$0.294819 \cdot 10^{-5}$	12	$0.295503 \cdot 10^{-5}$
10	$0.386426 \cdot 10^{-5}$	12	$0.265483 \cdot 10^{-5}$	9	$0.170344 \cdot 10^{-5}$
9	$0.375428 \cdot 10^{-5}$	9	$0.265321 \cdot 10^{-5}$	10	$0.149323 \cdot 10^{-5}$
13	$0.351408 \cdot 10^{-5}$	10	$0.167229 \cdot 10^{-5}$	4	$0.147647 \cdot 10^{-5}$
4	$0.192977 \cdot 10^{-5}$	4	$0.151147 \cdot 10^{-5}$	3	$0.103868 \cdot 10^{-5}$
8	$0.154248 \cdot 10^{-5}$	3	$0.102960 \cdot 10^{-5}$	1	$0.992904 \cdot 10^{-6}$
3	$0.115748 \cdot 10^{-5}$	8	$0.936095 \cdot 10^{-6}$	8	$0.870240 \cdot 10^{-6}$

**Table 6:** Differences between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  for output 2 at 2 meters

$\sigma$	$\sqrt{\sum \sigma_i^2}$	Data size
$0.306936 \cdot 10^{-4}$	$0.224788 \cdot 10^{-4}$	100
$0.293465 \cdot 10^{-4}$	$0.194446 \cdot 10^{-4}$	200
$0.288980 \cdot 10^{-4}$	$0.162046 \cdot 10^{-4}$	300

\*The mean value  $f_0$  for the 300 points sample is  $0.363884 \cdot 10^{-4}$ .

**Table 7:** Order of standard deviations  $\sigma_i$  for output 2 at 2 meters related to different input variables and different data sizes

100		200		300	
Input	$\sigma_i$	Input	$\sigma_i$	Input	$\sigma_i$
2	$0.170609 \cdot 10^{-4}$	2	$0.174678 \cdot 10^{-4}$	2	$0.144571 \cdot 10^{-4}$
11	$0.760822 \cdot 10^{-5}$	11	$0.364207 \cdot 10^{-5}$	6	$0.356782 \cdot 10^{-5}$
9	$0.500592 \cdot 10^{-5}$	12	$0.304761 \cdot 10^{-5}$	7	$0.333250 \cdot 10^{-5}$
7	$0.474022 \cdot 10^{-5}$	5	$0.295092 \cdot 10^{-5}$	4	$0.280828 \cdot 10^{-5}$
6	$0.447015 \cdot 10^{-5}$	4	$0.291934 \cdot 10^{-5}$	13	$0.224196 \cdot 10^{-5}$
13	$0.434843 \cdot 10^{-5}$	9	$0.263464 \cdot 10^{-5}$	5	$0.216179 \cdot 10^{-5}$
8	$0.422719 \cdot 10^{-5}$	6	$0.253386 \cdot 10^{-5}$	11	$0.204149 \cdot 10^{-5}$
12	$0.420436 \cdot 10^{-5}$	3	$0.233818 \cdot 10^{-5}$	9	$0.149770 \cdot 10^{-5}$
4	$0.332182 \cdot 10^{-5}$	8	$0.195082 \cdot 10^{-5}$	10	$0.138828 \cdot 10^{-5}$
3	$0.260819 \cdot 10^{-5}$	7	$0.189888 \cdot 10^{-5}$	3	$0.120802 \cdot 10^{-5}$
10	$0.257262 \cdot 10^{-5}$	13	$0.183943 \cdot 10^{-5}$	12	$0.105776 \cdot 10^{-5}$
5	$0.232084 \cdot 10^{-5}$	10	$0.153254 \cdot 10^{-5}$	8	$0.102406 \cdot 10^{-5}$
1	$0.212835 \cdot 10^{-5}$	1	$0.110115 \cdot 10^{-5}$	1	$0.441398 \cdot 10^{-6}$

**Table 8:** Differences between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  for output 3 at 2 meters

$\sigma$	$\sqrt{\sum \sigma_i^2}$	Data size
$0.255172 \cdot 10^{-4}$	$0.236598 \cdot 10^{-4}$	100
$0.247299 \cdot 10^{-4}$	$0.221269 \cdot 10^{-4}$	200
$0.255194 \cdot 10^{-4}$	$0.229067 \cdot 10^{-4}$	300

\*The mean value for the 300 points sample is  $0.247640 \cdot 10^{-4}$ .

**Table 9:** Order of standard deviations  $\sigma_i$  for output 3 at 2 meters related to different input variables and different data size

100		200		300	
Input	$\sigma_i$	Input	$\sigma_i$	Input	$\sigma_i$
2	$0.188752 \cdot 10^{-4}$	2	$0.195176 \cdot 10^{-4}$	2	$0.203930 \cdot 10^{-4}$
5	$0.647632 \cdot 10^{-5}$	5	$0.688022 \cdot 10^{-5}$	5	$0.683757 \cdot 10^{-5}$
6	$0.556381 \cdot 10^{-5}$	1	$0.438401 \cdot 10^{-5}$	1	$0.500147 \cdot 10^{-5}$
1	$0.482618 \cdot 10^{-5}$	6	$0.391443 \cdot 10^{-5}$	6	$0.321818 \cdot 10^{-5}$
12	$0.464542 \cdot 10^{-5}$	4	$0.267481 \cdot 10^{-5}$	7	$0.239809 \cdot 10^{-5}$
13	$0.451914 \cdot 10^{-5}$	7	$0.248339 \cdot 10^{-5}$	11	$0.199478 \cdot 10^{-5}$
7	$0.451057 \cdot 10^{-5}$	13	$0.246902 \cdot 10^{-5}$	9	$0.175954 \cdot 10^{-5}$
10	$0.412297 \cdot 10^{-5}$	11	$0.160146 \cdot 10^{-5}$	4	$0.169604 \cdot 10^{-5}$
4	$0.336268 \cdot 10^{-5}$	3	$0.131516 \cdot 10^{-5}$	10	$0.154081 \cdot 10^{-5}$
9	$0.241529 \cdot 10^{-5}$	9	$0.108412 \cdot 10^{-5}$	12	$0.153906 \cdot 10^{-5}$
8	$0.216174 \cdot 10^{-5}$	12	$0.101798 \cdot 10^{-5}$	13	$0.149555 \cdot 10^{-5}$
11	$0.187221 \cdot 10^{-5}$	10	$0.859862 \cdot 10^{-6}$	8	$0.148123 \cdot 10^{-5}$
3	$0.162968 \cdot 10^{-5}$	8	$0.348386 \cdot 10^{-6}$	3	$0.135680 \cdot 10^{-5}$

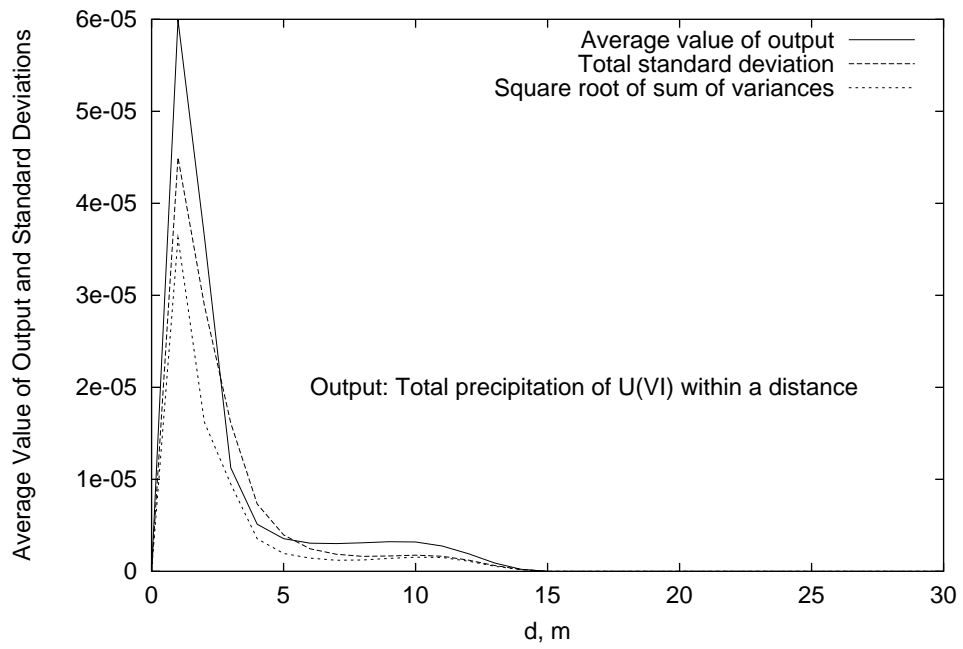
**Table 10:** Differences between  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  for output 4 at 2 meters

$\sigma$	$\sqrt{\sum \sigma_i^2}$	Data size
$0.279296 \cdot 10^{-7}$	$0.248749 \cdot 10^{-7}$	100
$0.277570 \cdot 10^{-7}$	$0.246952 \cdot 10^{-7}$	200
$0.289565 \cdot 10^{-7}$	$0.257628 \cdot 10^{-7}$	300

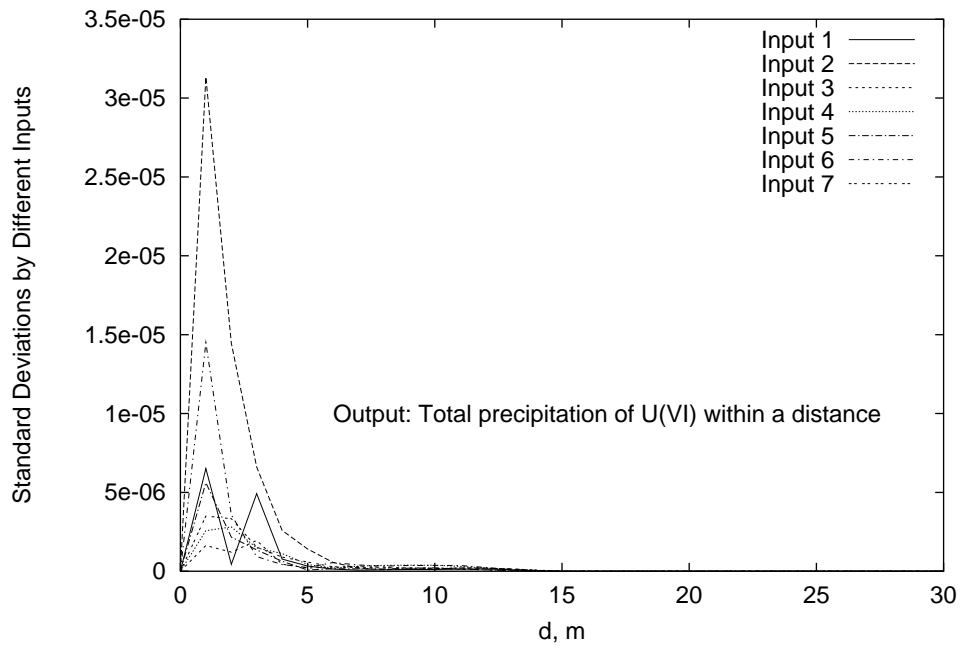
\*The mean value for the 300 points sample is  $0.163172 \cdot 10^{-7}$ .

**Table 11:** Order of standard deviations  $\sigma_i$  for output 4 related to different input variables and different data sizes

100		200		300	
Input	$\sigma_i$	Input	$\sigma_i$	Input	$\sigma_i$
2	$0.195915 \cdot 10^{-7}$	2	$0.219295 \cdot 10^{-7}$	2	$0.230027 \cdot 10^{-7}$
5	$0.671758 \cdot 10^{-8}$	5	$0.755725 \cdot 10^{-8}$	5	$0.789949 \cdot 10^{-8}$
1	$0.594292 \cdot 10^{-8}$	1	$0.517357 \cdot 10^{-8}$	1	$0.556782 \cdot 10^{-8}$
6	$0.535776 \cdot 10^{-8}$	6	$0.379047 \cdot 10^{-8}$	6	$0.290514 \cdot 10^{-8}$
10	$0.508676 \cdot 10^{-8}$	4	$0.259645 \cdot 10^{-8}$	10	$0.248300 \cdot 10^{-8}$
12	$0.503325 \cdot 10^{-8}$	13	$0.256021 \cdot 10^{-8}$	11	$0.242597 \cdot 10^{-8}$
13	$0.486064 \cdot 10^{-8}$	11	$0.228703 \cdot 10^{-8}$	3	$0.205531 \cdot 10^{-8}$
4	$0.393581 \cdot 10^{-8}$	9	$0.204112 \cdot 10^{-8}$	12	$0.203450 \cdot 10^{-8}$
7	$0.308192 \cdot 10^{-8}$	3	$0.164327 \cdot 10^{-8}$	9	$0.195489 \cdot 10^{-8}$
11	$0.278716 \cdot 10^{-8}$	10	$0.155516 \cdot 10^{-8}$	13	$0.193293 \cdot 10^{-8}$
9	$0.259582 \cdot 10^{-8}$	7	$0.127795 \cdot 10^{-8}$	4	$0.146985 \cdot 10^{-8}$
8	$0.259258 \cdot 10^{-8}$	12	$0.100461 \cdot 10^{-8}$	8	$0.136369 \cdot 10^{-8}$
3	$0.217571 \cdot 10^{-8}$	8	$0.498754 \cdot 10^{-9}$	7	$0.874389 \cdot 10^{-9}$



**Figure 1:** The average value,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  of output 2 at different locations.



**Figure 2:** The  $\sigma_i$  of output 2 for inputs 1 to 7.

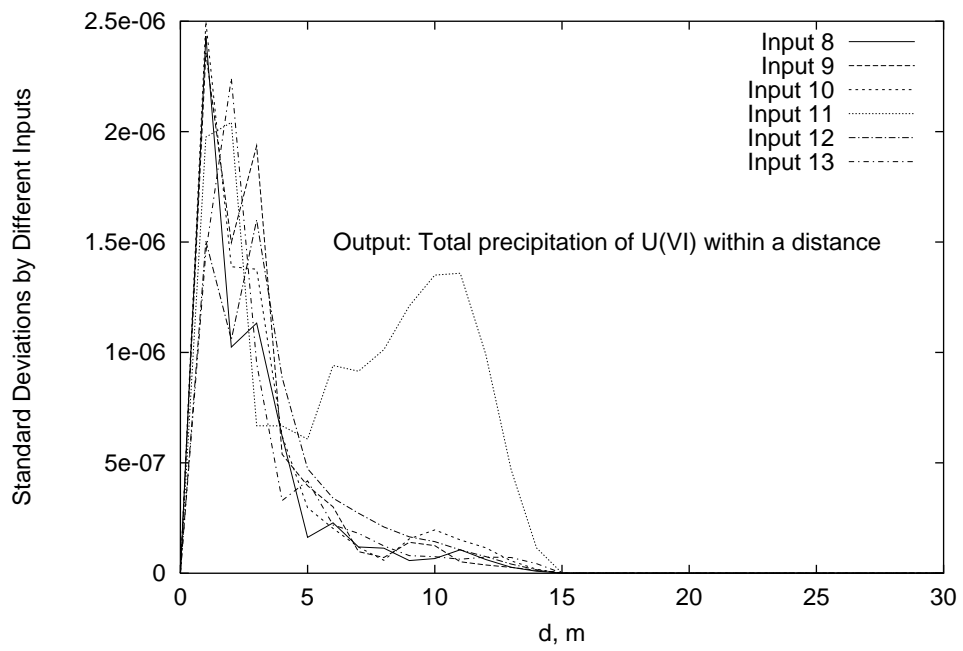


Figure 3: The  $\sigma_i$  of output 2 for inputs 8 to 13.

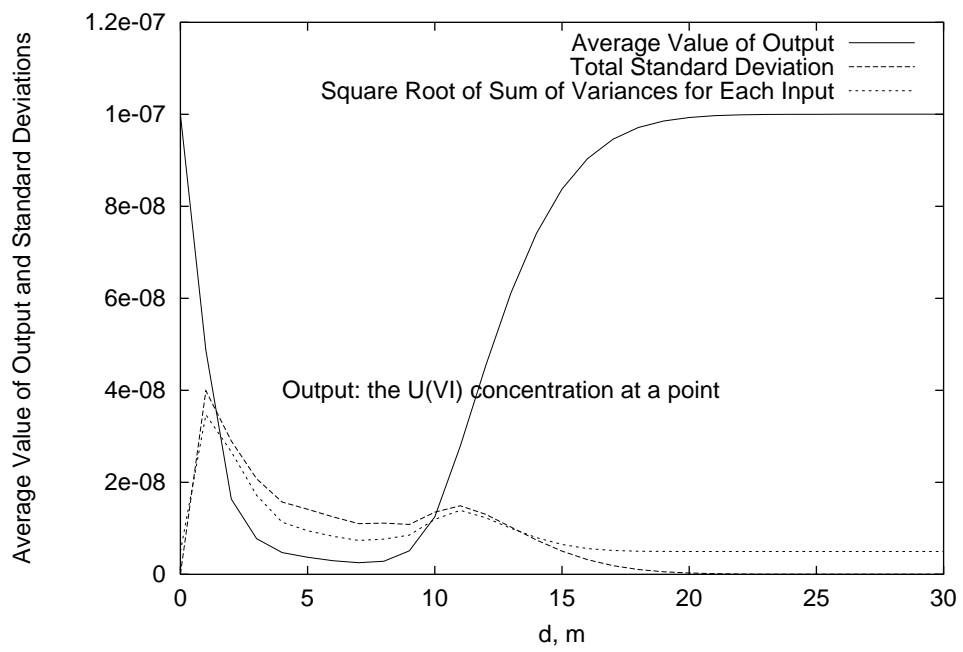


Figure 4: The average value,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  of output 3 at different locations.

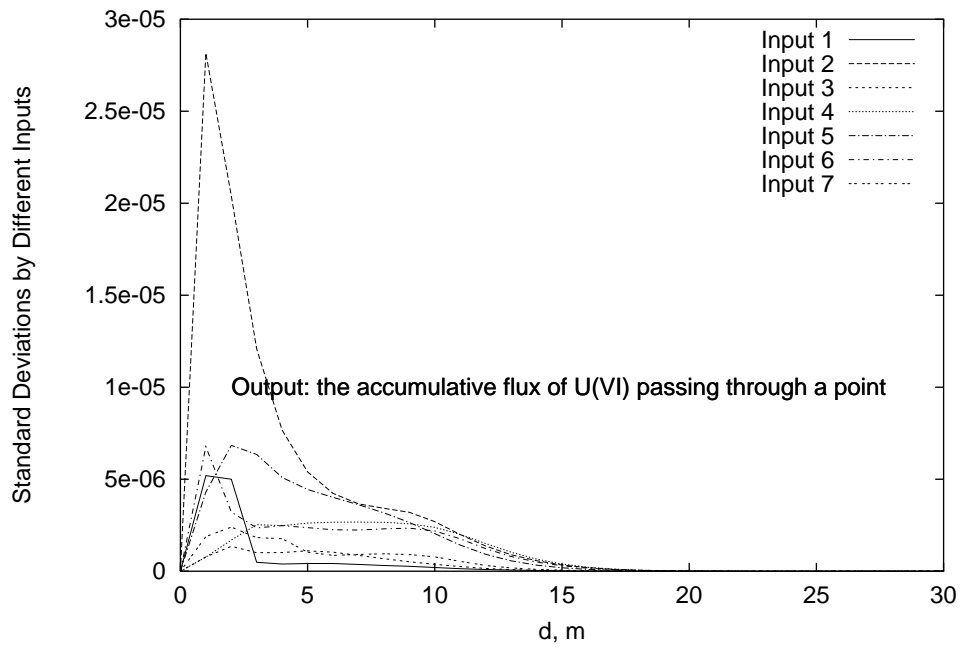


Figure 5: The  $\sigma_i$  of output 3 for inputs 1 to 7.

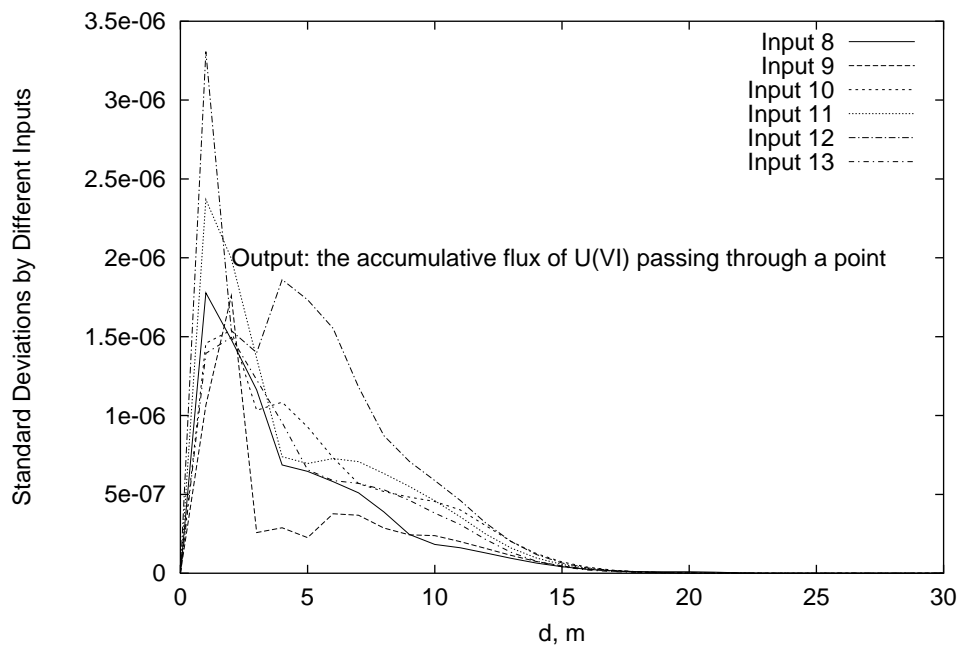


Figure 6: The  $\sigma_i$  of output 3 for inputs 8 to 13.

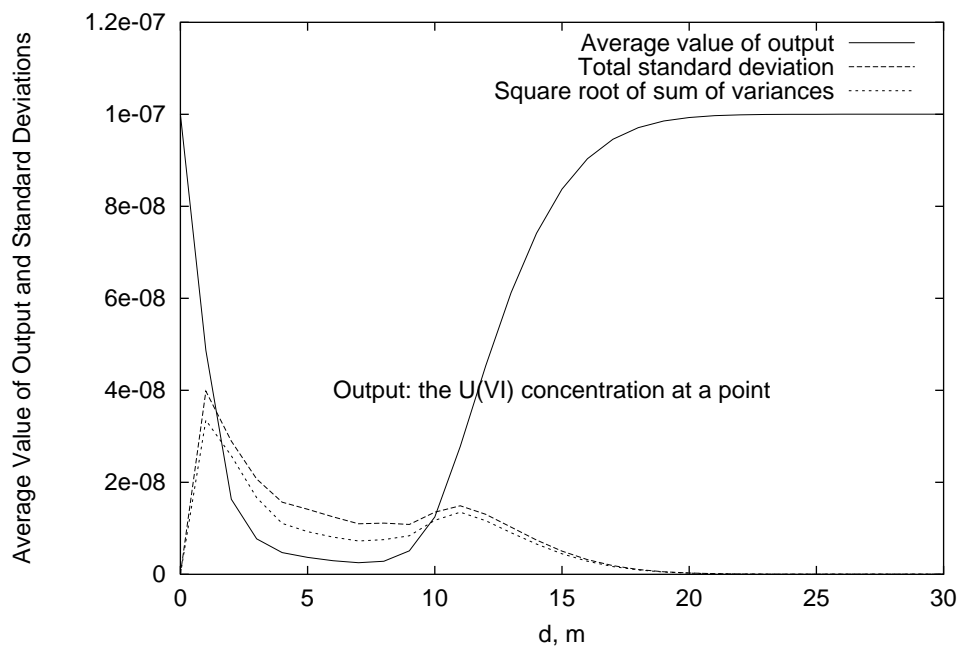


Figure 7: The average value,  $\sigma$  and  $\sqrt{\sum \sigma_i^2}$  of output 4 at different locations.

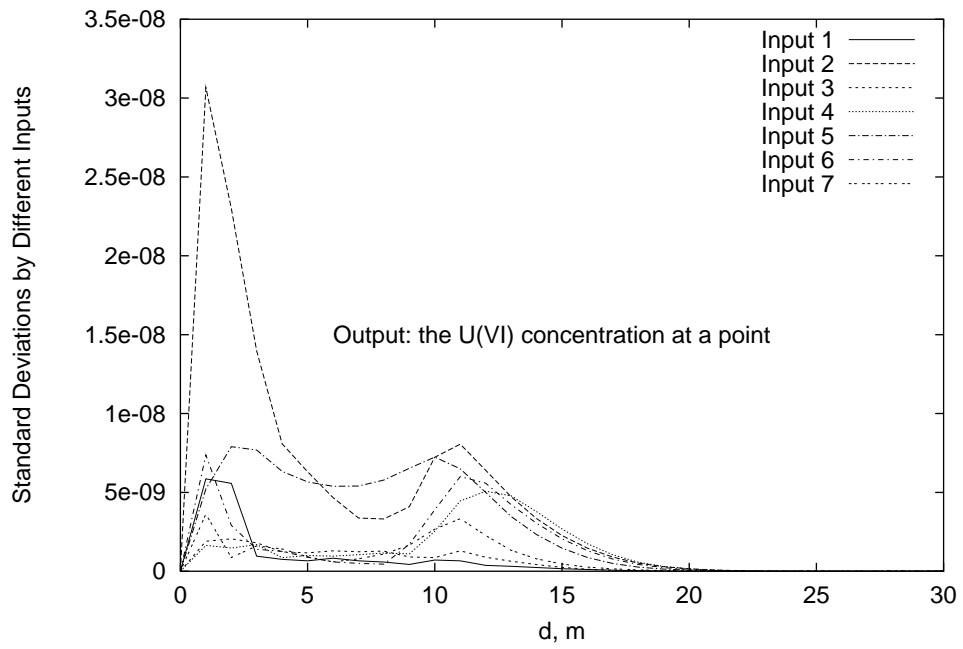
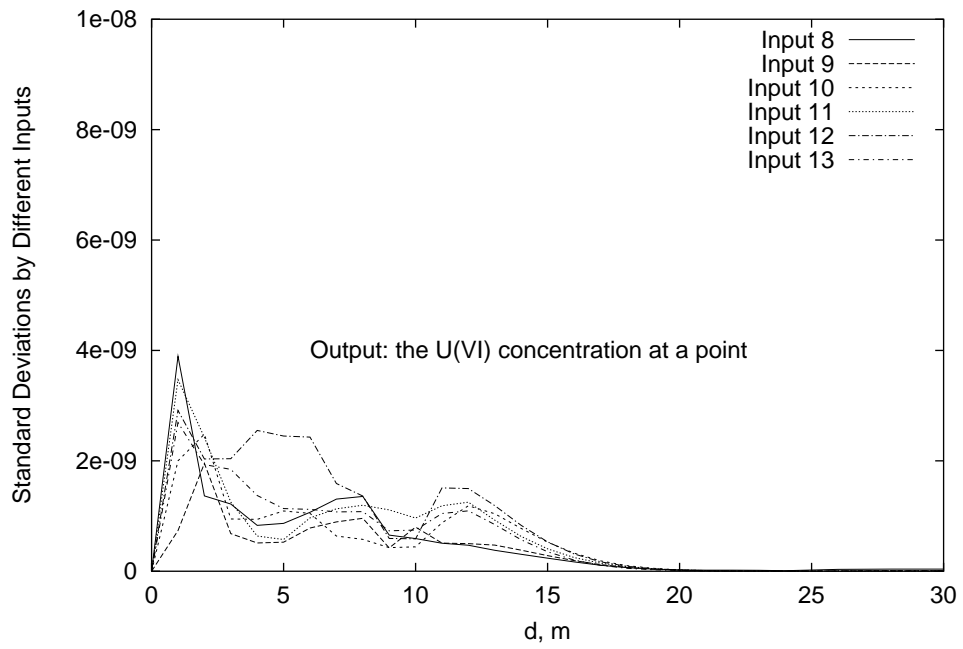


Figure 8: The  $\sigma_i$  of output 4 for inputs 1 to 7.



**Figure 9:** The  $\sigma_i$  of output 4 for inputs 8 to 13.

## REFERENCES

- [1] Doll, J.D., and Freeman D.L., Randomly exact methods, *Science*, **234**, 1356-1360(1986).
- [2] Iman, R.L., and Conover W.J., Small sample sensitivity analysis techniques for computer models, with an application to risk assessment, *Communications in Statistics, Part A. Theory and Methods*, **17**, 1749-1842(1980).
- [3] Tatang, M.A., Pan, W.W., Prinn, R.G., and McRae G.J., An efficient method for parametric uncertainty analysis of a numerical geophysical model, *J. Geophys. Res.-Atmospheres*, **102(D18)**, 21925-21932(1997).
- [4] Isukapalli, S.S., Roy, A., and Georgopoulos P.G., Stochastic Response Surface Methods (SRSMs) for uncertainty propagation: Application to environmental and biological systems, *Risk Analysis*, **18(3)**, 351-363(1998).
- [5] Isukapalli, S.S., Roy, A., and Georgopoulos P.G., Efficient Sensitivity/Uncertainty Analysis Using the Combined Stochastic Response Surface Method and Automated Differentiation: Application to environmental and biological systems, *Risk Analysis*, **20(5)**, 591-602(2000).
- [6] Smith, S.L., and Jaffe P.R., Modeling the Transport and Reaction of Trace Metals in Water-Saturated Soils and Sediments, *Water Resources Research*, **34(11)**, 3135-3147(1998).
- [7] National Technical Information Service (NTIS), MINTEQA2/PRODEFA2, A geochemical assessment model for environmental systems: Version 3.0 user's manual, EPA/600/3-91/021 (1991).
- [8] Rabitz, H., Alis, O. F., Shorter, J. and Shim K., Efficient Input-output Model Representations, *Computer Physics Communications*, **117**, 11-20(1999).
- [9] Alis, O. and Rabitz, H., General Foundations of High Dimensional Model Representations, *J. Math. Chem.*, **25**, 197-233(1999).
- [10] Shim, K., and Rabitz, H., Independent and Correlated Composition Behavior of Material Properties: Application to Energy Band Gaps for the  $Ga_{\alpha}In_{1-\alpha}P_{\beta}As_{1-\beta}$  and  $Ga_{\alpha}In_{1-\alpha}P_{\beta}Sb_{\gamma}As_{1-\beta-\gamma}$  Alloys, *Phys. Rev. B.*, **58**, 1940-1946(1998).
- [11] Shorter, J., Precila, C. Ip. and Rabitz, H., An Efficient Chemical Kinetics Solver using High Dimensional Model Representations, *J. Phys. Chem. A.*, **103(36)**, 7,192-7,198(1999).
- [12] Wang, S.W., Levy, II., H., Li, G. and Rabitz, H., Fully Equivalent Operational Models for Atmospheric Chemical Kinetics within Global Chemistry-transport Models, *J. Geophys. Res.*, **104**, D23, 30,417-30,426(1999).
- [13] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., Numerical Recipes in FORTRAN, Cambridge University Press, New York, 1992, p. 299-319.