UNIVERSITY OF CALIFORNIA SAN DIEGO

Efficient Global Sensitivity Analysis of Models with High-Dimensional Input

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Engineering Sciences (Aerospace Engineering)

by

Yikyung Yu

Committee in Charge:

Professor Daniel M. Tartakovsky, Chair
Professor Prabhakar R. Bandaru, Co-Chair
Professor Marcos Intaglietta
Professor Ratnesh Lal
Professor John S. McCartney

2025

The dissertation of Yikyung Yu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

DEDICATION

To my family and friends.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

2013	Bachelor of Science, University of California San Diego
2016	Master of Science, University of California San Diego
2016-2021	Teaching Assistant, University of California San Diego
2025	Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Engineering

Studies in Engineering Sciences (Aerospace Engineering)
Professor Daniel M. Tartakovsky

ABSTRACT OF THE DISSERTATION

Efficient Global Sensitivity Analysis of Models with High-Dimensional Input

by

Yikyung Yu

Doctor of Philosophy in Engineering Sciences (Aerospace Engineering)

University of California San Diego, 2025

Professor Daniel M. Tartakovsky, Chair
Professor Prabhakar R. Bandaru, Co-Chair

To make useful predictions about the behavior of a system, a model of the system is built. Complex models, such as environmental or electrochemical models, include numerous inputs that influence the model outputs. The true values of these inputs are often unknown and must be estimated using empirical data from experiments. The accuracy of these estimates improves as the uncertainty associated with the inputs decreases. While careful measurements can reduce input uncertainty, rigorous measurement of all inputs in complex models can be prohibitively expensive due to high experimental or computational costs,

making it essential to prioritize which inputs to measure precisely. Therefore, it is crucial to identify the inputs that have the greatest influence on the model outputs. Sensitivity analysis provides a framework for distinguishing important inputs by quantifying or qualifying the effects of inputs on model outputs. This allows researchers to focus resources on improving the accuracy of the most influential inputs, leading to more reliable model predictions.

This dissertation applied global sensitivity analysis (GSA) to two complex models: a groundwater flow model and a lithium-ion battery model. Two well-known GSA methods were employed: the Morris method and the Sobol' method. The Morris method assesses input influence by computing the sample mean and standard deviation of elementary effects of each input. The Sobol' method calculates first-order sensitivity indices and total sensitivity indices to quantify input importance.

For the groundwater flow model, the GSA results indicated that recharge flux was the primary driver of variations in the net flux of seawater intrusion. In the lithium-ion battery model, the thickness of the positive electrode had the greatest impact on battery life.

Chapter 1

Introduction

When observing phenomena, scientists strive to model them to predict their behavior. Models are constructed analytically, using systems of equations, or numerically, using computer code. It is challenging to create a perfect model for any phenomenon or system, as multiple models can often represent the same reality. Therefore, a model is built with simplifying assumptions and then calibrated with observed data to enhance its accuracy and usefulness. To improve model accuracy, numerous inputs that influence the model outputs were included. However, the true or exact values of these inputs are typically unknown, introducing uncertainty.

The accuracy of a model is significantly influenced by the degree of uncertainty associated with its inputs. To reduce input uncertainty, the values of the inputs must be measured or estimated as accurately as possible. Quantifying all inputs precisely can be challenging due to high experimental or computational costs, especially when dealing with models with numerous inputs. To address this challenge, sensitivity analysis (SA) has been developed by many mathematicians and scientists.

SA identifies the inputs that exert the most significant influence on the model's output. By applying SA to complex, high-dimensional models, the inputs to measure or estimate first to enhance model accuracy are prioritized. SA finds widespread application across diverse scientific disciplines.

1.1 Sensitivity Analysis

Uncertainty arises from imperfections and simplifications inherent in any model. A model can be a simple equation, like Newton's second law (F = ma), or a complex computer program. The model output is any value calculated from the model, such as force in F = ma, net flux or head in a groundwater model, or terminal voltage, current, and concentration in a battery model.

There are two main types of sensitivity analysis: local Sensitivity Analysis (LSA) and global sensitivity analysis (GSA). LSA focuses on a specific region near a point of interest within the input space. Since this method is well-suited for linear models, if a model is linear, LSA can be used instead of GSA. Common methods of LSA include calculating the partial derivative of the output (Y) with respect to the input $(X)(\delta_Y/\delta_X)$ and the standard deviation-normalized partial derivative $(\sigma_X\delta_Y/(\sigma_Y\delta_X))$.

However, applying LSA to a non-linear model will not yield reliable results. GSA considers the entire domain of the inputs. This method is more appropriate for non-linear models. Some common GSA methods include the Morris method and variance-based methods: Fourier amplitude sensitivity test (FAST), random balanced design based on FAST (RBD-FAST), and the Sobol method.

GSA helps choose and prioritize the most influential inputs for measurement. Since it can be applied to any model, GSA is a valuable tool in various fields for model improvement and calibration. GSA ranks the importance of inputs. Based on these rankings, which factors to measure or estimate first can be determined. For complex, high-dimensional models, GSA is typically employed.

1.2 Sampling for Global Sensitivity Analysis

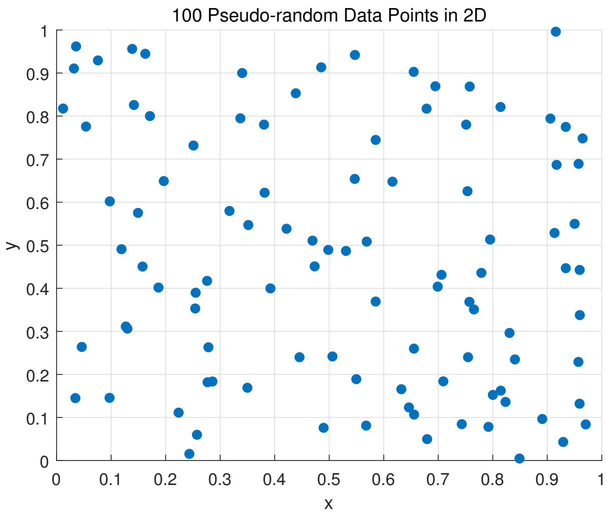
The convergence rate of Monte Carlo simulations with pseudo-random numbers is generally proportional to $1/\sqrt{N}$, where N is the sample size. To reduce the number of simulations required for a given level of accuracy, quasi-random sequences have been developed.

Quasi-random sequences, such as the Halton sequence and the Sobol' sequence, are generated systematically according to specific rules, unlike pseudo-random numbers, which are typically generated by algorithms that aim to mimic true randomness. Pseudo-random number generators often produce sequences that exhibit some degree of correlation or clustering. In contrast, quasi-random sequences exhibit low discrepancy, meaning they are more uniformly distributed across the sampling space, avoiding clusters and gaps.

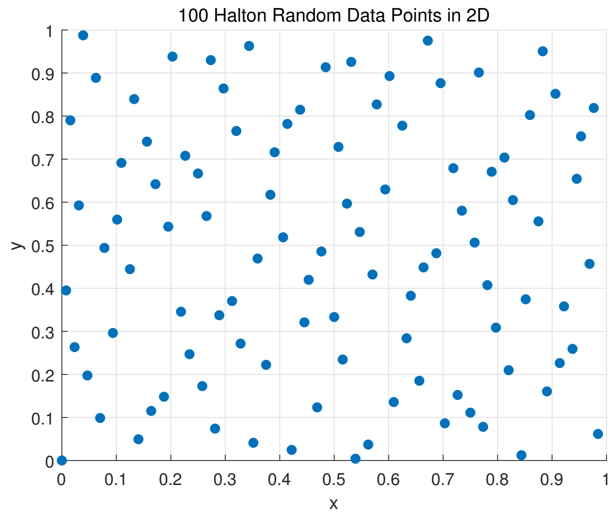
Figure 1.1 illustrates the distribution of points generated by pseudo-random, Halton, Sobol', and Latin Hypercube sequences in 2D and 3D space, demonstrating the improved uniformity of quasi-random sequences.

The convergence rate of quasi-random sequences can be significantly faster than that of pseudo-random sequences, approaching $1/N$ in optimal cases. The Sobol' sequence is widely used in many fields, including global sensitivity analysis, due to its efficient generation and good low-discrepancy properties.

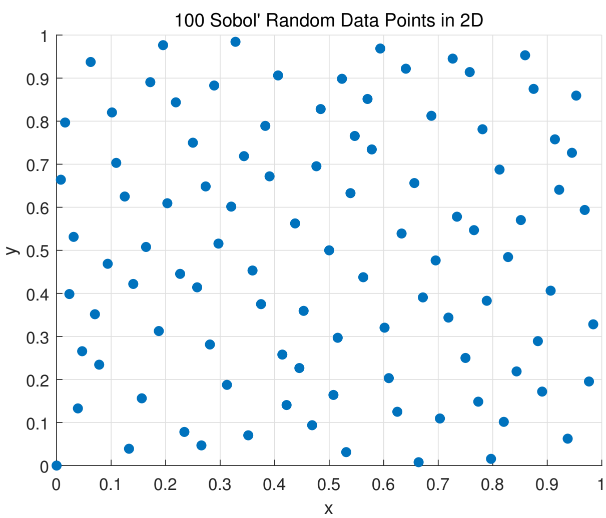
Due to the high computational cost associated with global sensitivity analysis, particularly when using methods like the Sobol' method, parallel computing techniques are recommended to accelerate the calculations.



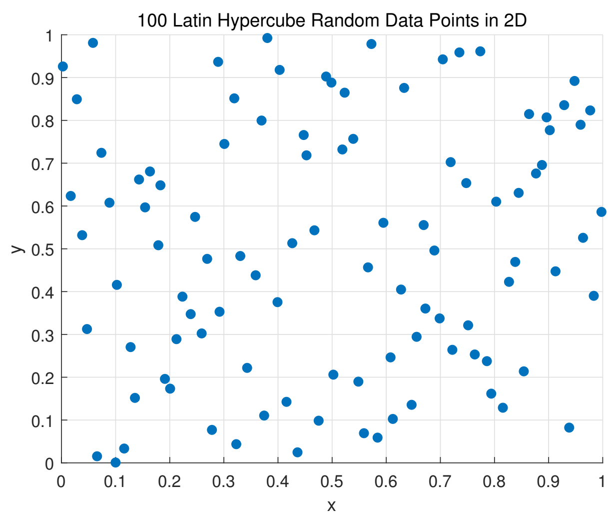
(a) 100 pseudo-random data points in 2D



(b) 100 Halton random data points in 2D

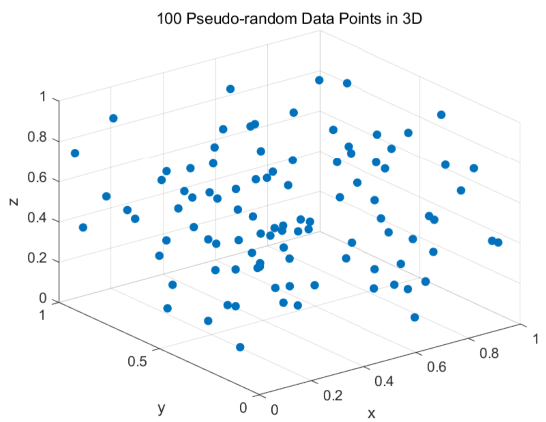


(c) 100 Sobol' random data points in 2D

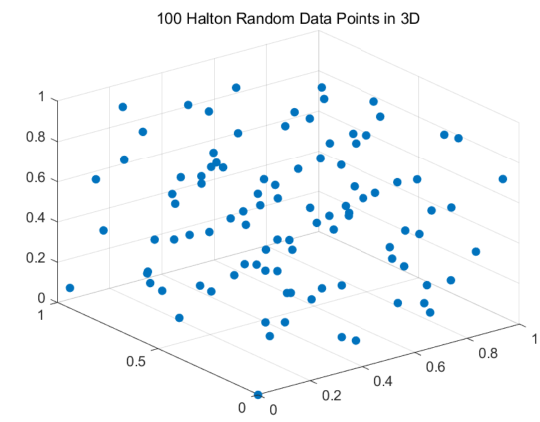


(d) 100 Latin hypercube random data points in 2D

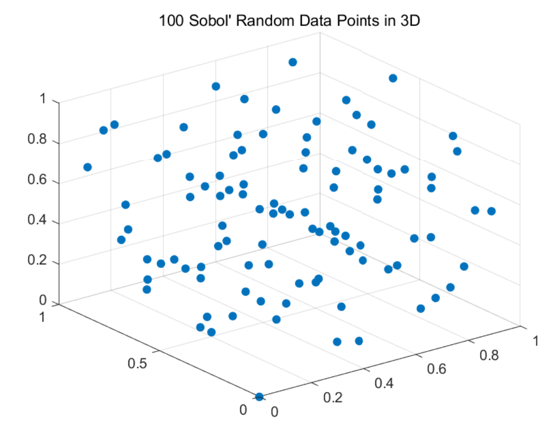
Figure 1.1: Scatter plots of different sets of random data points in 2D



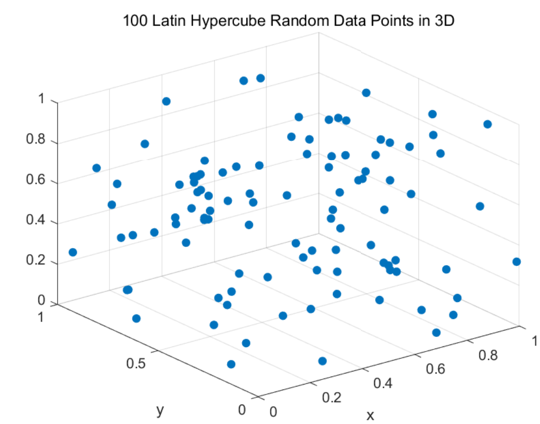
(a) 100 pseudo-random data points in 3D



(b) 100 Halton random data points in 3D



(c) 100 Sobol' random data points in 3D



(d) 100 Latin hypercube random data points in 3D

Figure 1.2: Scatter plots of different sets of random data points in 3D

Chapter 2

Two Methods for Global Sensitivity Analysis

Numerous methods for global sensitivity analysis have been developed. Prominent examples include the Morris method [9, 2] and variance-based methods such as the Sobol' method [19, 15, 16], the Fourier amplitude sensitivity test (FAST) [3, 1], and the random balanced designs based on FAST [23, 14, 24]. This chapter explores the details of the Morris method and the Sobol' method.

2.1 The Morris Method

The Morris method was proposed as an alternative to fractional factorial designs for analyzing deterministic computational models with a moderate-to-large number of inputs. It employs a series of randomized one-factor-at-a-time designs [9] and is also known as the elementary effects method because it is based on the analysis of so-called elementary effects. The sample mean and standard deviation of an input are used to measure the sensitivity of each input.

2.1.1 Elementary Effects

Suppose an output Y of a model is a deterministic scalar function of k independent random inputs of the model, denoted by the k-element vector $\mathbf{X} = (X_1, X_2, \ldots, X_i, \ldots, X_k)$

$$Y = f(\mathbf{X}).$$

The cumulative distribution function (CDF) values of the inputs $(X_{c1}, X_{c2}, \ldots, X_{ci}, \ldots, X_{ck})$ are sampled in a k-dimensional grid with p_i levels for the ith dimension, where X_{ci} takes on values from the set $\{0, 1/(p_i - 1), 2/(p_i - 1), \cdots, 1 - 1/(p_i - 1), 1\}$. For given values of $\mathbf{X}$, the elementary effect of the ith input X_i

$$D_i(\mathbf{X}) = \frac{f(X_1, X_2, \ldots, X_{i-1}, X_i + \Delta_i, X_{i+1}, \ldots, X_k) - f(\mathbf{X})}{\Delta_{ci}}, \tag{2.1}$$

where Δ_{ci} is a multiple of $1/(p_i - 1)$, $\Delta_{ci} \leq 1$, $X_i + \Delta_i$ is the inverse CDF value of $X_{ci} + \Delta_{ci}$, and $X_i + \Delta_i \leq sup(X_i)$ [9, 17].

2.1.2 Means and Standard Deviations

The effect of the ith input X_i on the output Y can be assessed by the population mean μ_i and standard deviation σ_i of the elementary effects of X_i.

- A large positive or negative μ_i and a small σ_i suggests X_i has a strong and consistent positive or negative effect on Y, respectively, compared to other inputs.

- A large absolute value of μ_i and a large σ_i indicates the effect of X_i on Y is significant and highly dependent on other inputs.

- A small $|\mu_i|$ and a small σ_i suggests X_i has no significant effect on Y.

- A small $|\mu_i|$ and a large σ_i suggests X_i has both positive and negative elementary effects which cancel each other out.

The influence of X_i on Y can also be assessed by the population mean $E(|D_i|)$ of the absolute values of the elementary effects of X_i.

- A large $E(|D_i|)$ suggests X_i has a significant influence on Y.

- A small $E(|D_i|)$ suggests X_i has no significant influence on Y.

If all p_i and Δ_i are equal to p and Δ, there are $p^{k-1}[p - \Delta(p-1)]$ elementary effects for each input [9]. For instance, when $k = 10, p = 4$, and $\Delta = p/[2(p-1)]$, there are 524,288 elementary effects for each input. Assuming it takes a minute to compute a single elementary, calculating all the elementary effects for all 10 inputs would take approximately 10 years. This computational cost is prohibitively high, rendering a full calculation of elementary effects for most real-world models impractical.

Therefore, μ_i, σ_i, and $E(|D_i|)$ are estimated by a sample mean $\overline{D_i}$, a standard deviation s_i, and a sample mean $\overline{|D_i|}$ of the absolute values of the elementary effects of X_i. These statistics serve as sensitivity measures within the Morris method [9, 2].

2.1.3 Trajectories

The Morris method employs an efficient sampling strategy. Instead of evaluating a model $2rk$ times, where r is the number of elementary effects sampled per input and k is the number of inputs, the concept of trajectories is introduced. A trajectory involves systematically varying one input at a time while holding others constant, generating k elementary effects with only $k + 1$ model evaluations. Then only $r(k + 1)$ evaluations are needed. This significantly reduces the computational burden.

Furthermore, efforts have been made to optimize trajectory selection. By strategically choosing a set of trajectories that effectively cover the input space from a larger set of random trajectories, the efficiency of the sampling process can be further improved. Maximizing the distance between trajectories has been explored to achieve this goal [2].

The distance between trajectories m and l

$$d_{m,l} = \begin{cases} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k} [X_i^m(z) - X_j^l(z)]^2} & \text{if } m \neq l, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

k is the number of inputs and $X_i^m(z)$ is the zth coordinate of the ith point of trajectory m [2].

The optimal set of r out of M random trajectories can be selected based on the root-mean-square (RMS) of $\binom{r}{2}$ distances for each of $\binom{M}{r}$ cases. For instance, if $r = 4$ trajectories are chosen from $M = 10$ random trajectories, the RMS of the $\binom{4}{2}$ distances

$$R_{i,j,k,l} = \sqrt{\frac{d_{i,j}^2 + d_{i,k}^2 + d_{i,l}^2 + d_{j,k}^2 + d_{j,l}^2 + d_{k,l}^2}{6}} \tag{2.3}$$

for each of the $\binom{10}{4}$ cases is calculated, where i, j, k, and l are distinct integers from 1 to M. Among the $\binom{10}{4}$ sets of 4 trajectories, the set with the highest $R_{i,j,k,l}$ would be considered optimal, as it indicates the greatest coverage of the input space.

Four random trajectories in a 2-dimensional 4-level grid with $\Delta = p/[2(p-1)]$ are shown in Figure 2.1. For example, trajectory t_4 starts from a random point in the grid $(X_1^{(4)}, X_2^{(4)}) = (0, 0)$. It then proceeds to $(X_1^{(4)}, X_2^{(4)}) = (2/3, 0)$, where $X_2^{(4)}$ remains unchanged while $X_1^{(4)}$ increases by $\Delta = 2/3$. The third point on this trajectory is $(X_1^{(4)}, X_2^{(4)}) = (2/3, 2/3)$.

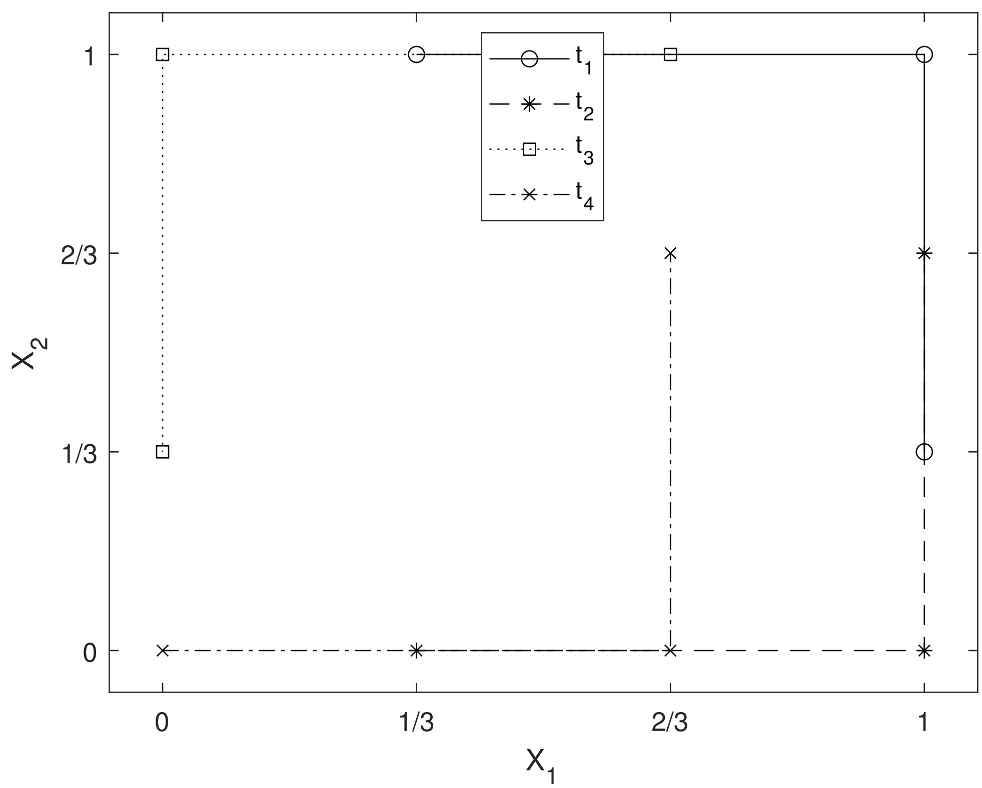


Figure 2.1: Four random trajectories in a 2-dimensional 4-level grid

One catch with the optimal trajectory selection strategy is the potential for selecting trajectories that are effectively the same but traverse the input space in the opposite direction. For example, two trajectories

- t_5: $(X_1^{(5)}, X_2^{(5)}) \rightarrow (X_1^{(5)} + \Delta, X_2^{(5)}) \rightarrow (X_1^{(5)} + \Delta, X_2^{(5)} + \Delta)$
- t_6: $(X_1^{(6)} = X_1^{(5)} + \Delta, X_2^{(6)} = X_2^{(5)} + \Delta) \rightarrow (X_1^{(6)}, X_2^{(6)} - \Delta) \rightarrow (X_1^{(6)} - \Delta, X_2^{(6)} - \Delta)$

explore the same sequence of input combinations but in the reverse order. Therefore, the strategy could limit the exploration of the input space.

Using these optimal trajectories, the sensitivity measures of the Morris method are calculated.

The sample mean of the elementary effects of the ith input X_i

$$\overline{D_i(\mathbf{X})} = \frac{1}{r}\sum_{j=1}^{r} D_i^j(\mathbf{X}), \tag{2.4}$$

where r is the number of trajectories and $D_i^j(\mathbf{X})$ is the elementary effect of X_i of the jth trajectory.

The sample standard deviation of the elementary effects of X_i

$$s_i(\mathbf{X}) = \sqrt{\frac{1}{r-1}\sum_{j=1}^{r}(D_i^j(\mathbf{X}) - \overline{D_i(\mathbf{X})})}. \tag{2.5}$$

The sample mean of the absolute values of the elementary effects of X_i

$$\overline{|D_i(\mathbf{X})|} = \frac{1}{r}\sum_{j=1}^{r} |D_i^j(\mathbf{X})|, \tag{2.6}$$

where $|D_i^j(\mathbf{X})|$ is the absolute value of the elementary effect of X_i of the jth trajectory.

2.2 The Sobol' Method

The Sobol' method is a variance-based method using variance decomposition. It calculates first-order and total sensitivity indices of model inputs to quantify the sensitivities of a model output to the inputs. The indices are estimated by the Monte Carlo integration using points in the Sobol' sequence.

2.2.1 Variance Decomposition

The variance of a square integrable scalar function of independent random variables having a uniform distribution on the interval [0, 1] can be decomposed into variances of functions in different dimensions [18].

Suppose that an output Y of a model is a square integrable scalar function of p independent random inputs uniformly distributed between 0 and 1, denoted by the p-element vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)$:

$$Y = f(\mathbf{X}), \quad f(\mathbf{X}) \in L_2, \quad \text{and} \quad X_i \sim U(0,1) \quad \text{for} \quad i = 1, \ldots, p.$$

Then the function $f(\mathbf{X})$ can be represented by the sum of a constant and $2^p - 1$ functions in different dimensions [4]:

$$\begin{aligned} f(\mathbf{X}) =& f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} f_{i,j}(X_i, X_j) + \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} \sum_{k=j+1}^{p} f_{i,j,k}(X_i, X_j, X_k) + \cdots \\ & + \sum_{i=1}^{p-n+1} \cdots \sum_{n=l+1}^{p} f_{i,\ldots,l,n}(X_i, \ldots, X_l, X_n) + \cdots + f_{1,2,\ldots,p}(X_1, X_2, \ldots, X_p), \end{aligned} \tag{2.7}$$

where the means of all the functions on the right side of Equation 2.7 are zero:

$$\int_0^1 \cdots \int_0^1 f_{i,\ldots,n}(x_i, \ldots, x_n) dx_i \cdots dx_n = 0. \tag{2.8}$$

The constant and functions can be calculated by integration. For example, integrating both sides of Equation 2.7 with respect to $\mathbf{x}$ gives

$$\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x} = f_0, \tag{2.9}$$

with respect to $\mathbf{x}$ except x_i gives

$$\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}_{\sim i} = f_0 + f_i(X_i), \tag{2.10}$$

and with respect to $\mathbf{x}$ except x_i and x_j gives

$$\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}_{\sim(i,j)} = f_0 + f_i(X_i) + f_j(X_j) + f_{i,j}(X_i, X_j). \tag{2.11}$$

Squaring and integrating both sides of Equation 2.7 with respect to $\mathbf{x}$ and taking f_0^2 on the right side to the left side gives

$$\begin{aligned}\int_0^1 \cdots \int_0^1 f^2(\mathbf{x})d\mathbf{x} - f_0^2 = &\sum_{i=1}^{p} \int_0^1 f_i^2(x_i)dx_i + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \int_0^1 \int_0^1 f_{i,j}^2(x_i, x_j)dx_i dx_j \\ &+ \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} \sum_{k=j+1}^{p} \int_0^1 \int_0^1 \int_0^1 f_{i,j,k}^2(x_i, x_j, x_k)dx_i dx_j dx_k + \cdots \\ &+ \sum_{i=1}^{p-n+1} \cdots \sum_{n=l+1}^{p} \int_0^1 \cdots \int_0^1 f_{i,\ldots,l,n}^2(x_i, \ldots, x_l, x_n)dx_i \ldots dx_l dx_n + \cdots \\ &+ \int_0^1 \cdots \int_0^1 f_{1,2,\ldots,p}^2(x_1, x_2, \ldots, x_p)d\mathbf{x}.\end{aligned} \tag{2.12}$$

By the definition of variance and Equations 2.8 and 2.9, the variance of Y

$$V(Y) = \int_0^1 \cdots \int_0^1 f^2(\mathbf{x})d\mathbf{x} - \left(\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}\right)^2 \tag{2.13}$$

$$= \int_0^1 \cdots \int_0^1 f^2(\mathbf{x})d\mathbf{x} - f_0^2, \tag{2.14}$$

the variance of $f_i(X_i)$

$$V(f_i(X_i)) = \int_0^1 f_i^2(x_i)dx_i - \left(\int_0^1 f_i(x_i)dx_i\right)^2$$
$$= \int_0^1 f_i^2(x_i)dx_i, \tag{2.15}$$

and the variance of $f_{i,\ldots,n}(X_i,\ldots,X_n)$

$$V(f_i(X_i)) = \int_0^1 \cdots \int_0^1 f_{i,\ldots,n}^2(x_i,\ldots,x_n)dx_i \ldots dx_n$$
$$- \left(\int_0^1 \cdots \int_0^1 f_{i,\ldots,n}(x_i,\ldots,x_n)dx_i \ldots dx_n\right)^2 \tag{2.16}$$
$$= \int_0^1 \cdots \int_0^1 f_{i,\ldots,n}^2(x_i,\ldots,x_n)dx_i \ldots dx_n. \tag{2.17}$$

Substituting Equations 2.14, 2.15, and 2.17 into Equation 2.12 gives

$$V(Y) = \sum_{i=1}^{p} V(f_i(X_i)) + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p} V(f_{i,j}(X_i,X_j)) + \sum_{i=1}^{p-2}\sum_{j=i+1}^{p-1}\sum_{k=j+1}^{p} V(f_{i,j,k}(X_i,X_j,X_k))$$
$$+ \cdots + \sum_{i=1}^{p-n+1} \cdots \sum_{n=l+1}^{p} V(f_{i,\ldots,l,n}(X_i,\ldots,X_l,X_n)) + \cdots + V(f_{1,2,\ldots,p}(\mathbf{X})). \tag{2.18}$$

$V(Y)$ is decomposed into the variances of $2^p - 1$ functions in different dimensions.

2.2.2 First-Order and Total Sensitivity Indices

The Sobol' method calculates first-order and total sensitivity indices, which quantify the influence of each input on the variance of an output.

The variances on the right side of Equation 2.18 can be expressed as variances of conditional expectations of Y:

$$
\begin{aligned}
V(f_i(X_i)) &= V\left(\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}_{\sim i} - f_0\right) \\
&= V\left(\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}_{\sim i}\right) && (2.19) \\
&= V\left(\int_0^1 \cdots \int_0^1 f(x_1, \ldots, X_i, \ldots, x_p)dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_p\right) \\
&= V(E(Y|X_i)), && (2.20) \\
V(f_{i,j}(X_i, X_j)) &= V\left(\int_0^1 \cdots \int_0^1 f(\mathbf{x})d\mathbf{x}_{\sim(i,j)} - f_i(X_i) - f_j(X_j) - f_0\right) && (2.21) \\
&= V(E(Y|X_i, X_j)) - V(E(Y|X_i)) - V(E(Y|X_j)), && (2.22) \\
V(f_{i,\ldots,n}(X_i, \ldots, X_n)) &= V(E(Y|X_i, \ldots, X_n)) - \cdots - V(E(Y|X_i)) - \cdots - V(E(Y|X_n)). && (2.23)
\end{aligned}
$$

$V(f_{i,\ldots,n}(X_i, \ldots, X_n))$ is called the nth-order effect of X_i through X_n on Y. For example, $V(f_i(X_i))$ is the first-order effect of X_i on Y and $V(f_{i,j}(X_i, X_j))$ is the second-order effect of X_i and X_j on Y. A higher-order effect does not include lower-order effects.

The sum of the effects including X_i is called the total effect of X_i on Y:

$$
V(f_i(X_i)) + \sum_{\substack{j=1 \\ j \neq i}}^{p} V(f_{i,j}(X_i, X_j)) + \sum_{\substack{j=1 \\ j \neq i}}^{p-1} \sum_{\substack{k>j \\ k \neq i}}^{p} V(f_{i,j,k}(X_i, X_j, X_k)) + \cdots + V(f_{i,j,\ldots,p}(X_i, X_j, \ldots, X_p)). \tag{2.24}
$$

It can also be expressed as the difference

$$
V(Y) - V(E(Y|\mathbf{X}_{\sim i})), \tag{2.25}
$$

where $\mathbf{X}_{\sim i}$ is all the inputs except X_i.

The first-order sensitivity index of X_i is defined as

$$S_i = \frac{V(E(Y|X_i))}{V(Y)}, \tag{2.26}$$

which is the ratio of the first-order effect of X_i on Y to $V(Y)$. The second-order and nth-order sensitivity indices are defined in similar ways with Equations 2.22 and 2.23.

The total sensitivity index of X_i is defined as

$$\begin{aligned} ST_i &= \frac{V(Y) - V(E(Y|\mathbf{X}_{\sim i}))}{V(Y)} && (2.27) \\ &= \frac{E(V(Y|\mathbf{X}_{\sim i}))}{V(Y)} \quad \text{(by the law of total variance)}, && (2.28) \end{aligned}$$

which is the ratio of the total effect of X_i on Y to $V(Y)$.

The sensitivity of Y to X_i is quantified by S_i and ST_i. X_i with a high S_i or ST_i greatly influences Y.

2.2.3 Sensitivity Index Estimation

First-order and total sensitivity indices are estimated by the Monte Carlo integration. Points in the Sobol' sequence are used for the estimation [19, 15] since the Sobol' sequence is a low-discrepancy quasi-random sequence [20]. The points are scrambled to avoid using the point at the origin [11].

The variances of Y, the first-order and total effects of X_i, and f_0 can be estimated by the Monte Carlo integration [18]:

$$
\begin{aligned}
V(Y) &= \int_0^1 \cdots \int_0^1 f^2(\mathbf{x}) d\mathbf{x} - f_0^2 \\
&\approx \frac{1}{N} \sum_{s=1}^{N} f^2(\mathbf{x}^{(s)}) - f_0^2 && (2.29) \\
V(E(Y|X_i)) &= \int_0^1 \left(\int_0^1 \cdots \int_0^1 f(\mathbf{x}) d\mathbf{x}_{\sim i} \right)^2 dx_i - f_0^2 && (2.30) \\
&\approx \frac{1}{N} \sum_{s=1}^{N} f(x_i^{(s)}, \mathbf{x}_{\sim i}^{(s)}) f(x_i^{(s)}, \mathbf{x}_{\sim i}^{\prime(s)}) - f_0^2 && (2.31) \\
V(Y) - V(E(Y|\mathbf{X}_{\sim i})) &= V(Y) - \left(\int_0^1 \cdots \int_0^1 \left(\int_0^1 f(\mathbf{x}) dx_i \right)^2 d\mathbf{x}_{\sim i} - f_0^2 \right) \\
&\approx V(Y) - \frac{1}{N} \sum_{s=1}^{N} f(x_i^{(s)}, \mathbf{x}_{\sim i}^{(s)}) f(x_i^{\prime(s)}, \mathbf{x}_{\sim i}^{(s)}) + f_0^2 && (2.32) \\
f_0 &= \int_0^1 \cdots \int_0^1 f(\mathbf{x}) d\mathbf{x} \\
&\approx \frac{1}{N} \sum_{s=1}^{N} f(\mathbf{x}^{(s)}) && (2.33)
\end{aligned}
$$

where $\mathbf{x}^{(s)}$ is the sth point of a set of random N points, $x_i^{(s)}$ is the ith element of the sth point, $\mathbf{x}_{\sim i}^{(s)}$ is $\mathbf{x}^{(s)}$ except $x_i^{(s)}$, and $\mathbf{x}_{\sim i}^{\prime(s)}$ is the sth point of another set of random N points except $x_i^{\prime(s)}$.

For a model with p inputs, N points in the $2p$-dimensional Sobol' sequence are sampled to form a $N \times 2p$ matrix:

$$
\mathbf{S} = \begin{bmatrix}
x_{1,1} & x_{1,2} & \cdots & x_{1,i} & \cdots & x_{1,p} & \cdots & x_{1,p+i} & \cdots & x_{1,2p} \\
x_{2,1} & x_{2,2} & \cdots & x_{2,i} & \cdots & x_{2,p} & \cdots & x_{2,p+i} & \cdots & x_{2,2p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
x_{N-1,1} & x_{N-1,2} & \cdots & x_{N-1,i} & \cdots & x_{N-1,p} & \cdots & x_{N-1,p+i} & \cdots & x_{N-1,2p} \\
x_{N,1} & x_{N,2} & \cdots & x_{N,i} & \cdots & x_{N,p} & \cdots & x_{N,p+i} & \cdots & x_{N,2p}.
\end{bmatrix} \qquad (2.34)
$$

$\mathbf{S}$ is then partitioned into two $N \times p$ matrices, $\mathbf{A}$ and $\mathbf{B}$, where

$$\mathbf{A} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,i} & \cdots & x_{2,p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1,1} & x_{N-1,2} & \cdots & x_{N-1,i} & \cdots & x_{N-1,p} \\ x_{N,1} & x_{N,2} & \cdots & x_{N,i} & \cdots & x_{N,p} \end{bmatrix} \quad (2.35)$$

contains the first p columns of $\mathbf{S}$ and

$$\mathbf{B} = \begin{bmatrix} x_{1,p+1} & x_{1,p+2} & \cdots & x_{1,p+i} & \cdots & x_{1,2p} \\ x_{2,p+1} & x_{2,p+2} & \cdots & x_{2,p+i} & \cdots & x_{2,2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1,p+1} & x_{N-1,p+2} & \cdots & x_{N-1,p+i} & \cdots & x_{N-1,2p} \\ x_{N,p+1} & x_{N,p+2} & \cdots & x_{N,p+i} & \cdots & x_{N,2p} \end{bmatrix} \quad (2.36)$$

contains the remaining p columns of $\mathbf{S}$.

A matrix

$$\mathbf{B}_{\mathbf{A}}^{(i)} = \begin{bmatrix} x_{1,p+1} & x_{1,p+2} & \cdots & x_{1,i} & \cdots & x_{1,2p} \\ x_{2,p+1} & x_{2,p+2} & \cdots & x_{2,i} & \cdots & x_{2,2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1,p+1} & x_{N-1,p+2} & \cdots & x_{N-1,i} & \cdots & x_{N-1,2p} \\ x_{N,p+1} & x_{N,p+2} & \cdots & x_{N,i} & \cdots & x_{N,2p} \end{bmatrix} \quad (2.37)$$

is obtained from $\mathbf{B}$ by replacing its ith column with the ith column of $\mathbf{A}$. Similarly, another matrix

$$\mathbf{A}_{\mathbf{B}}^{(i)} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p+i} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p+i} & \cdots & x_{2,p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1,1} & x_{N-1,2} & \cdots & x_{N-1,p+i} & \cdots & x_{N-1,p} \\ x_{N,1} & x_{N,2} & \cdots & x_{N,p+i} & \cdots & x_{N,p} \end{bmatrix} \tag{2.38}$$

is obtained from $\mathbf{A}$ by replacing its ith column with the ith column of $\mathbf{B}$.

The first-order sensitivity index of X_i can be estimated by

$$\begin{aligned} S_i &= \frac{V(E(Y|X_i))}{V(Y)} \\ &\approx \frac{\frac{1}{N}\sum_{s=1}^{N} f(\mathbf{B}(s,:))(f(\mathbf{A}_{\mathbf{B}}^{(i)}(s,:)) - f(\mathbf{A}(s,:)))}{\frac{1}{2N}\sum_{s=1}^{N}(f^2(\mathbf{A}(s,:)) + f^2(\mathbf{B}(s,:))) - \left(\frac{1}{2N}\sum_{s=1}^{N}(f(\mathbf{A}(s,:)) + f(\mathbf{B}(s,:)))\right)^2} \end{aligned} \tag{2.39}$$

and the total sensitivity index of X_i can be estimated by

$$\begin{aligned} ST_i &= \frac{E(V(Y|\mathbf{X}_{\sim i}))}{V(Y)} \\ &\approx \frac{\frac{1}{2N}\sum_{s=1}^{N}(f(\mathbf{A}(s,:)) - f(\mathbf{A}_{\mathbf{B}}^{(i)}(s,:))^2}{\frac{1}{2N}\sum_{s=1}^{N}(f^2(\mathbf{A}(s,:)) + f^2(\mathbf{B}(s,:))) - \left(\frac{1}{2N}\sum_{s=1}^{N}(f(\mathbf{A}(s,:)) + f(\mathbf{B}(s,:)))\right)^2}, \end{aligned} \tag{2.40}$$

where $\mathbf{A}(s,:)$ is the sth row of $\mathbf{A}$ [16, 6, 8]. $\mathbf{A}_{\mathbf{B}}$ is preferred over $\mathbf{B}_{\mathbf{A}}$ due to its better distribution of points [16].

2.2.4 A Simple Linear Model

For a simple linear model, first-order and total sensitivity indices can be calculated analytically and empirically.

Suppose Y is the sum of four independent random variables X_i:

$$Y = \sum_{i=1}^{4} X_i,$$

where $X_i \sim U(0,1)$.

Then the mean of X_i

$$\begin{aligned}
E(X_i) &= \int_a^b x_i f(x_i) dx_i \\
&= \frac{1}{b-a} \int_a^b x_i dx_i \\
&= \frac{b-a}{2} \\
&= \frac{1}{2} \quad \text{for} \quad i = 1, 2, 3, 4.
\end{aligned}$$

Therefore, the mean of Y

$$\begin{aligned}
E(Y) &= E(X_1 + X_2 + X_3 + X_4) \\
&= E(X_1) + E(X_2) + E(X_3) + E(X_4) \\
&= 2.
\end{aligned}$$

The variance of X_i

$$\begin{aligned}
V(X_i) &= E((X_i - E(X_i))^2) \\
&= E(X_i^2) - [E(X_i)]^2 \\
&= \int_a^b x_i^2 f(x_i) dx_i - \left(\frac{b-a}{2}\right)^2 \\
&= \frac{1}{b-a} \int_a^b x_i^2 dx_i - \left(\frac{b-a}{2}\right)^2 \\
&= \frac{(b-a)^2}{12} \\
&= \frac{1}{12} \quad \text{for} \quad i = 1, 2, 3, 4.
\end{aligned}$$

Therefore, the variance of Y

$$\begin{aligned}
V(Y) &= V(X_1 + X_2 + X_3 + X_4) \\
&= V(X_1) + V(X_2) + V(X_3) + V(X_4) \\
&= \frac{1}{3}.
\end{aligned}$$

The first-order sensitivity index of X_i

$$\begin{aligned}
S_i &= \frac{V(E(Y|X_i))}{V(Y)} \\
&= \frac{V(E(X_i + X_j + X_k + X_l))}{V(Y)} \\
&= \frac{V(X_i + 3/2)}{V(Y)} \\
&= \frac{V(X_i)}{V(Y)} \\
&= \frac{1}{4} \quad \text{for} \quad i = 1, 2, 3, 4
\end{aligned}$$

and the total effect sensitivity index of X_i

$$\begin{aligned}
ST_i &= \frac{E(V(Y|X_{\sim i}))}{V(Y)} \\
&= \frac{E(V(X_i + X_j + X_k + X_l))}{V(Y)} \\
&= \frac{E(1/3)}{V(Y)} \\
&= \frac{1}{4} \quad \text{for} \quad i = 1, 2, 3, 4.
\end{aligned}$$

The first-order sensitivity indices were estimated for four different sample sizes using points in the Sobol' sequence. As the sample size increases, estimated first-order sensitivity indices converge to the analytical indices as illustrated in Figure 2.2.

The total sensitivity indices were estimated for four different sample sizes using points in the Sobol' sequence. As the sample size increases, estimated total sensitivity indices

Table 2.1: Estimated first-order sensitivity indices of all four inputs ($\hat{S}_i$) for different sample sizes

	2^3	2^7	2^{10}	2^{13}
$\hat{S}_1$	0.0545	0.2501	0.2502	0.2500
$\hat{S}_2$	0.1938	0.2569	0.2502	0.2500
$\hat{S}_3$	0.1745	0.2510	0.2502	0.2500
$\hat{S}_4$	0.0335	0.2565	0.2503	0.2501

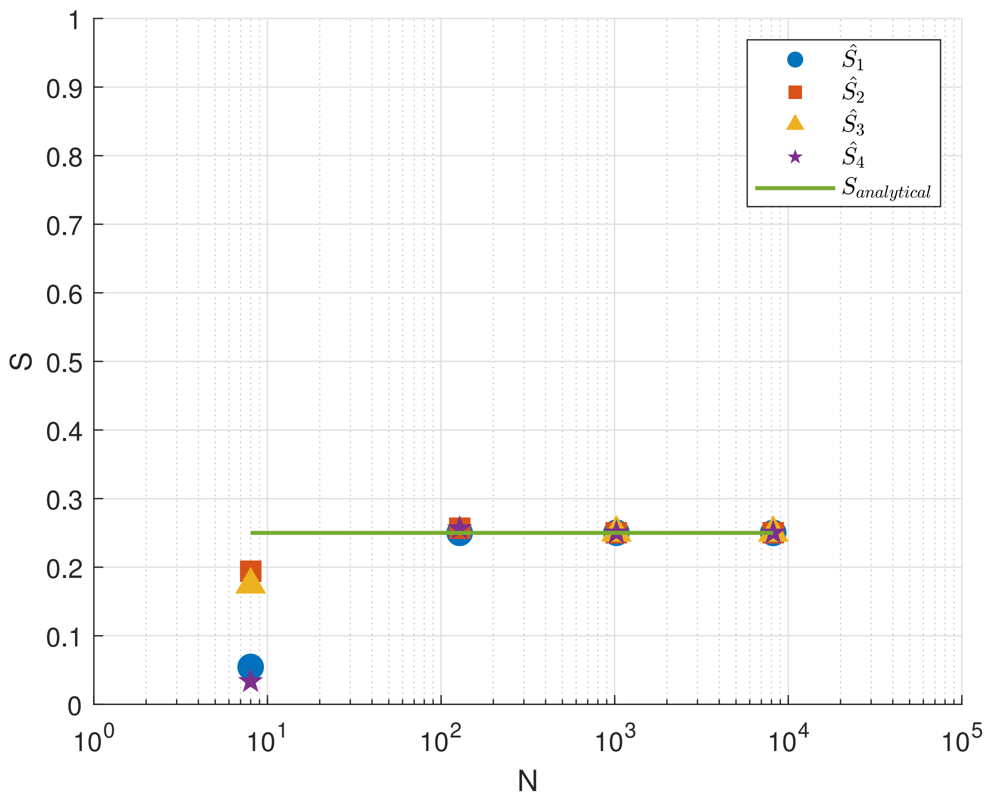


Figure 2.2: Convergence of estimated first-order sensitivity indices of all four inputs (S_i) ($S_{analytical}$ is 0.25.)

converge to the analytical indices as illustrated in Figure 2.3.

Table 2.2: Estimated total sensitivity indices of all four inputs ($\hat{ST_i}$) for different sample sizes

	2^3	2^7	2^{10}	2^{13}
$\hat{ST_1}$	0.4333	0.2497	0.2502	0.25
$\hat{ST_2}$	0.2566	0.2509	0.2502	0.25
$\hat{ST_3}$	0.2430	0.2512	0.2502	0.25
$\hat{ST_4}$	0.2606	0.2509	0.2501	0.25

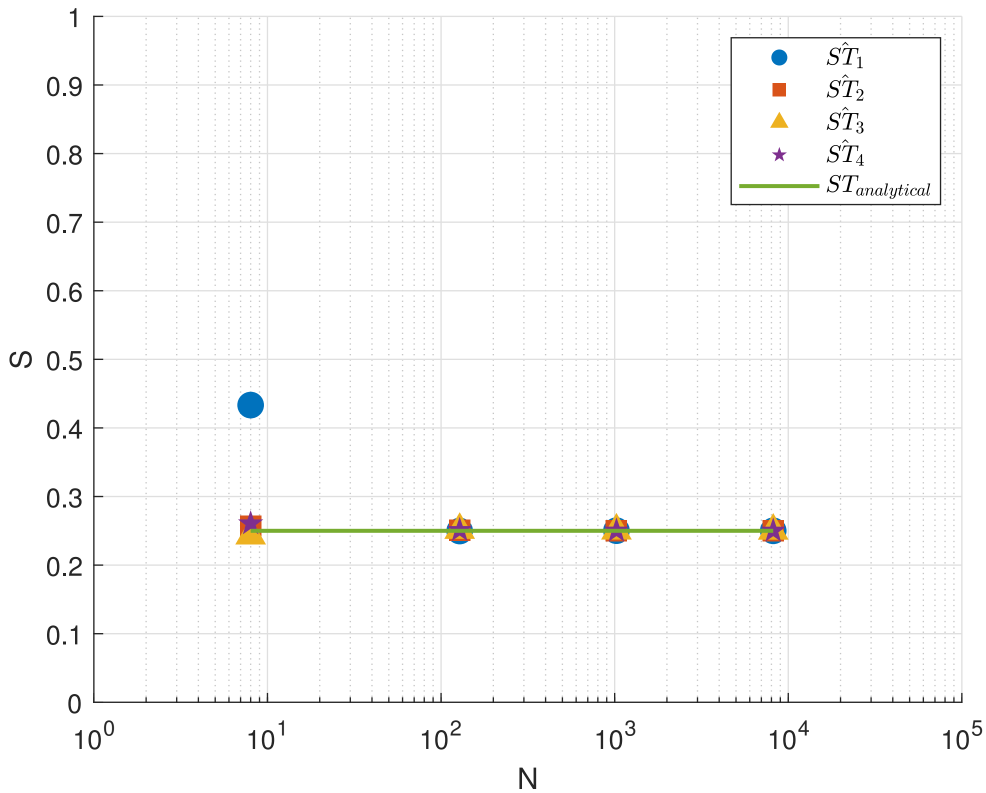


Figure 2.3: Convergence of estimated total sensitivity indices of all four inputs (ST_i) ($ST_{analytical}$ is 0.25.)

Chapter 3

Preventing Seawater Intrusion with Global Sensitivity Analysis

3.1 Introduction

Freshwater is essential for agriculture. However, the majority of Earth's freshwater is stored as seawater, which is unusable for irrigation due to its high salinity. Groundwater is the most readily accessible freshwater source in many regions, but its availability is limited.

Overpumping of groundwater can lead to land subsidence, including the formation of sinkholes. Near coastlines, freshwater depletion can induce seawater intrusion, contaminating freshwater aquifers.

Groundwater flow models are used to simulate and predict groundwater movement, incorporating factors such as geology, recharge rates, and pumping. While valuable, these models contain inherent uncertainties due to measurement errors and simplifications of real-world complexities. Data used to construct these models can have varying levels of accuracy, and models may not perfectly represent all real-world processes.

Global sensitivity analysis (GSA) can identify the most influential inputs in a model. Applying GSA to a groundwater flow model allows for the determination of the input parameters that most significantly affect the risk of seawater intrusion.

Identifying the most important input parameters in groundwater models enables the development of more targeted management strategies, informs decisions regarding sustainable groundwater use, and contributes to the protection of this vital resource for future generations.

3.2 Methods

The Morris method and the Sobol' method were employed to perform global sensitivity analysis of a groundwater flow model.

Sensitivity Analysis Library in Python (SALib) [6, 7] was used to sample points for both methods.

Modular Three-Dimensional Finite-Difference Ground-Water Flow Model with a Newton Formulation (MODFLOW-NWT) from USGS [10] was used to calculate the heads of the model after 372 monthly stress periods from January 1985 using the points from SALib as the input values.

Zonebudget (ZONBUD) [5] was used to calculate inflow and outflow across the boundary after the periods, December 2015 using the heads from MODFLOW-NWT. The sum of inflow and outflow is the net flux of seawater intruding into the aquifers.

MODFLOW-NWT and ZONBUD were run in MATLAB with Parallel Computing Toolbox to save time.

SALib was used to calculate the sensitivity measures of both methods.

3.2.1 The Groundwater Flow Model

Ventura Region Groundwater Flow Model (VRGWFM) [22], which is a model of aquifers in Ventura, California, was used in MODFLOW-NWT for the analysis.

MODFLOW-NWT uses the finite different method to calculate heads. The governing equation it solves is the groundwater flow equation.

The three-dimensional groundwater flow equation with anisotropic and heterogenous porous medium is derived by combining Darcy's law with the equation of continuity [25, 12]:

$$S_s \frac{\partial h}{\partial t} + \frac{\partial q}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial q}{\partial z} + W = 0, \tag{3.1}$$

where h is the hydraulic head $[L]$, $W = W(x, y, z, t)$ is the volumetric flux per unit volume, and $S_s = S_s(x, y, z)$ is the specific storage of the porous material $[L^{-1}]$.

From Darcy's law, the specific discharge or flux in each direction is

$$q_x = -K_x \frac{\partial h}{\partial x} \tag{3.2}$$

$$q_y = -K_y \frac{\partial h}{\partial y} \tag{3.3}$$

$$q_z = -K_z \frac{\partial h}{\partial z}, \tag{3.4}$$

where K_x, K_y, K_z are the hydraulic conductivity in the x, y, and z directions $[L/T]$.

Substituting Equations 3.2, 3.3, and 3.4 into Equation 3.1 and rearranging results in

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) - W = S_s \frac{\partial h}{\partial t}. \tag{3.5}$$

3.2.2 The Output and Inputs of the Model

The output of the model for global sensitivity analysis is the net flux of seawater intruding into the aquifers of the model across the general head boundary specified in the General-Head Boundary file for MODFLOW-NWT.

The inputs are 17 horizontal hydraulic conductivity in several layers and zones, 4 vertical hydraulic conductivity in different layers and zones, 4 storativity in different layers and zones, the recharge flux, the interface hydraulic conductance, and the streambed hydraulic conductance. All 28 inputs are listed in Table 3.1.

3.2.3 Settings for the Morris Method

All the inputs were assumed to be uniformly distributed within ±10% of their nominal values because the distributions of the inputs are unknown. 500 trajectories were randomly generated, and 10 optimal trajectories were selected from them. The number of grid levels were set to four. Since there are 28 inputs, $10 \times (28 + 1) = 290$ points were sampled.

The sample means of the absolute values of elementary effects of all 28 inputs and the sample standard deviations of the elementary effects of all 28 inputs were calculated.

Table 3.1: The 28 inputs of the groundwater flow model

Input name	Input description
X_1	Horizontal hydraulic conductivity in layer 1 zone 11 $[ft \cdot day^{-1}]$
X_2	Horizontal hydraulic conductivity in layer 3 zone 9 $[ft \cdot day^{-1}]$
X_3	Horizontal hydraulic conductivity in layer 3 zone 12 $[ft \cdot day^{-1}]$
X_4	Horizontal hydraulic conductivity in layer 5 zone 9 $[ft \cdot day^{-1}]$
X_5	Horizontal hydraulic conductivity in layer 5 zone 10 $[ft \cdot day^{-1}]$
X_6	Horizontal hydraulic conductivity in layer 5 zone 12 $[ft \cdot day^{-1}]$
X_7	Horizontal hydraulic conductivity in layer 6 zone 11 $[ft \cdot day^{-1}]$
X_8	Horizontal hydraulic conductivity in layer 6 zone 12 $[ft \cdot day^{-1}]$
X_9	Horizontal hydraulic conductivity in layer 6 zone 13 $[ft \cdot day^{-1}]$
X_{10}	Horizontal hydraulic conductivity in layer 6 zone 19 $[ft \cdot day^{-1}]$
X_{11}	Horizontal hydraulic conductivity in layer 7 zone 4 $[ft \cdot day^{-1}]$
X_{12}	Horizontal hydraulic conductivity in layer 7 zone 5 $[ft \cdot day^{-1}]$
X_{13}	Horizontal hydraulic conductivity in layer 7 zone 10 $[ft \cdot day^{-1}]$
X_{14}	Horizontal hydraulic conductivity in layer 7 zone 12 $[ft \cdot day^{-1}]$

3.2.4 Settings for the Sobol' Method

All the inputs were assumed to be uniformly distributed within ±10% of their nominal values because the distributions of the inputs are unknown. A base sample N was set to 1024 which is a power of 2. Since there are 28 inputs, $1024 \times (28 + 2) = 30720$ points in

Table 3.1: The 28 inputs of the groundwater flow model (Continued)

Input name	Input dsecripton
X_{15}	Horizontal hydraulic conductivity in layer 8 zone 12 $[ft \cdot day^{-1}]$
X_{16}	Horizontal hydraulic conductivity in layer 9 zone 4 $[ft \cdot day^{-1}]$
X_{17}	Horizontal hydraulic conductivity in layer 9 zone 5 $[ft \cdot day^{-1}]$
X_{18}	Vertical hydraulic conductivity in layer 6 zone 4 $[ft \cdot day^{-1}]$
X_{19}	Vertical hydraulic conductivity in layer 6 zone 11 $[ft \cdot day^{-1}]$
X_{20}	Vertical hydraulic conductivity in layer 6 zone 13 $[ft \cdot day^{-1}]$
X_{21}	Vertical hydraulic conductivity in layer 7 zone 12 $[ft \cdot day^{-1}]$
X_{22}	Storativity in layer 9 zone 12
X_{23}	Storativity in layer 10 zone 12
X_{24}	Storativity in layer 11 zone 12
X_{25}	Storativity in layer 13 zone 12
X_{26}	A multiplier for recharge flux
X_{27}	A multiplier for interface hydraulic conductance
X_{28}	A multiplier for streambed hydraulic conductance

the 56-dimensional Sobol' sequence were sampled.

The first-order and total sensitivity indices of all 28 inputs were estimated with 95% confidence intervals.

3.3 Results

Results of global sensitivity analysis, performed using the Morris method and the Sobol' method, are detailed in this section.

3.3.1 Results from the Morris Method

Figure 3.1 shows estimated means of the absolute values of elementary effects ($\overline{|D_i|}$) and estimated standard deviations of the elementary effects (s_i) for all 28 inputs ($i = 1, \ldots, 28$). Three inputs with estimated means greater than 1×10^7 are marked with their names (X_{26}, X_{27}, and X_{28}).

The estimated mean and standard deviation for X_{26}, $\overline{|D_{26}|}$ and s_{26}, are 1.2779×10^8 and 0.1239×10^8, respectively. $\overline{|D_{27}|}$ and s_{27} are 0.1749×10^8 and 0.0254×10^8. $\overline{|D_{28}|}$ and s_{28} are 0.2282×10^8 and 0.0434×10^8. The estimated means and standard deviations for the others are smaller than those for the three inputs. All estimated means and standard deviations are listed in Table A.1.

3.3.2 Results from the Sobol' Method

Of the 30720 evaluations, 168 fail to converge in at least two stress periods. However, since these failures represent only 0.5469% of the total (runs), their output values were included in the sensitivity index calculations.

Figures 3.2 and 3.3 show 95% confidence intervals for the first-order and total sensitivity indices, respectively, for all 28 inputs. Both figures look very similar to each other.

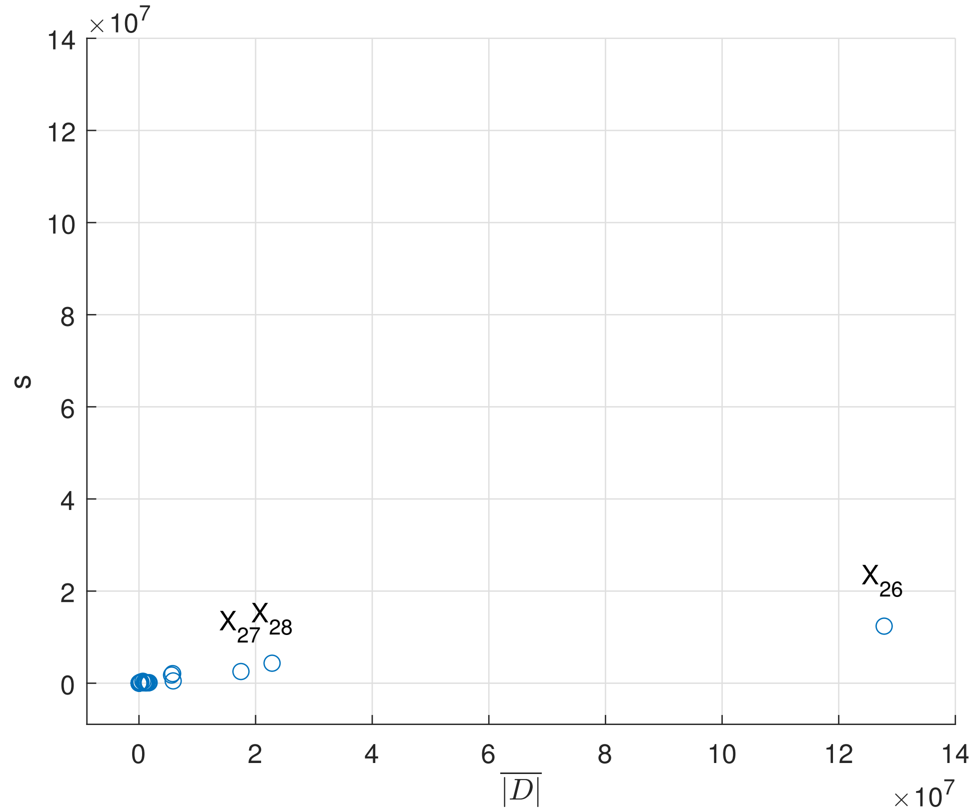


Figure 3.1: Estimated means ($\overline{|D_i|}$) and standard deviations (s_i) of the (absolute) values of elementary effects for all 28 inputs (Three inputs with estimated means greater than 1×10^7 are marked with their names.)

The 95% confidence interval for the first-order sensitivity index of X_{26}, S_{26}, is 0.9539 ± 0.0749. The 95% confidence intervals for S_{27} and S_{28} are 0.0182 ± 0.0126 and 0.0259 ± 0.0155, respectively. The 95% confidence intervals for the first-order sensitivity indices of the other inputs are around zero, including it. All confidence intervals for the first-order sensitivity indices are listed in Table A.2.

The 95% confidence interval for the total sensitivity index of X_{26}, ST_{26}, is 0.9443 ± 0.0635. The 95% confidence intervals for ST_{27} and ST_{28} are 0.0178 ± 0.0016 and 0.0250 ±

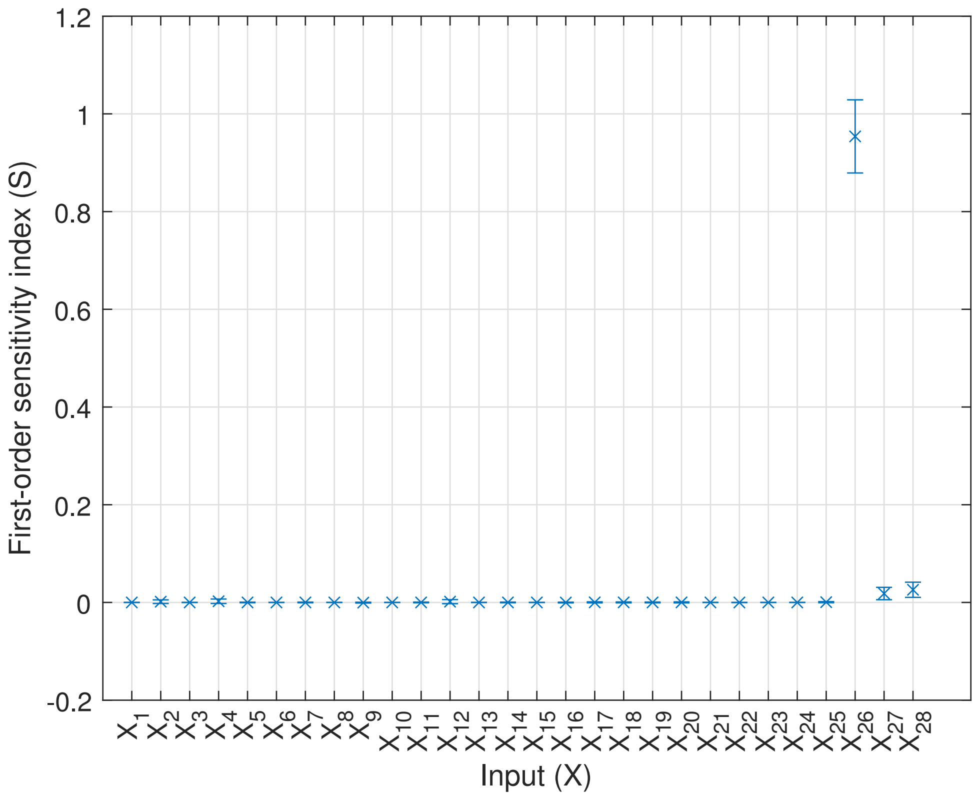


Figure 3.2: The 95% confidence intervals for the first-order sensitivity indices of all 28 inputs

0.0024, respectively. The 95% confidence intervals for the total sensitivity indices of the other inputs are around zero, including it. All confidence intervals for the total sensitivity indices are listed in Table A.2.

3.4 Discussion

Both the Morris method and the Sobol' method yield the same results. When the output of interest is the net flux of seawater intruding into the adjacent aquifers, X_{26}, the recharge flux, is the most influential input to the groundwater flow model. X_{27}, the

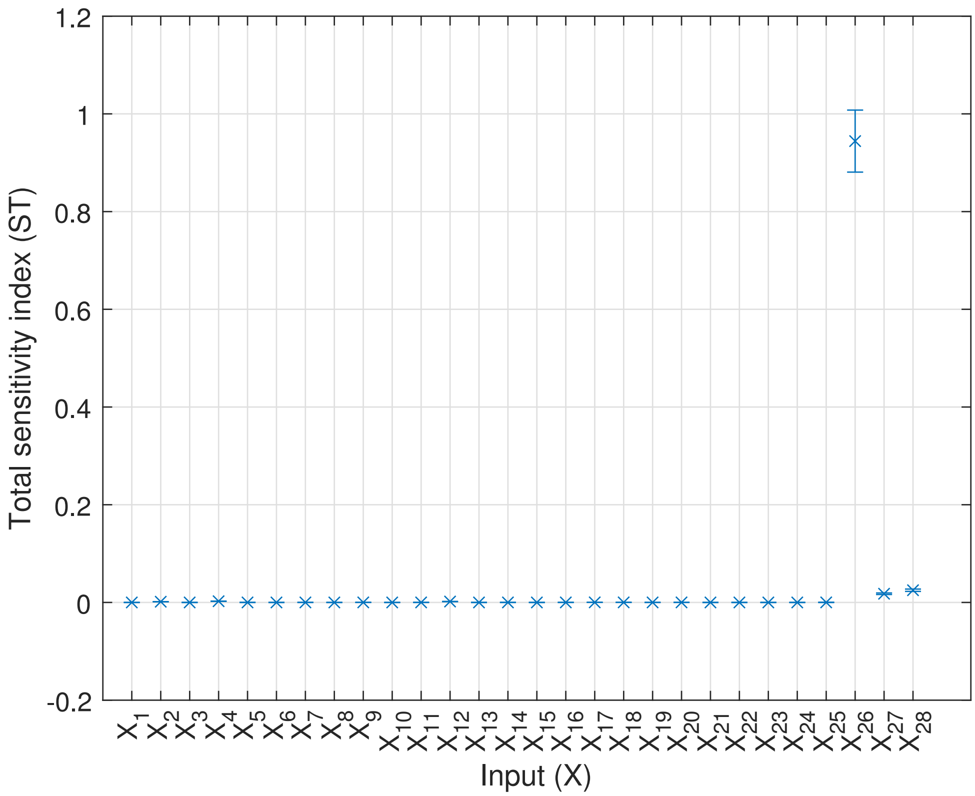
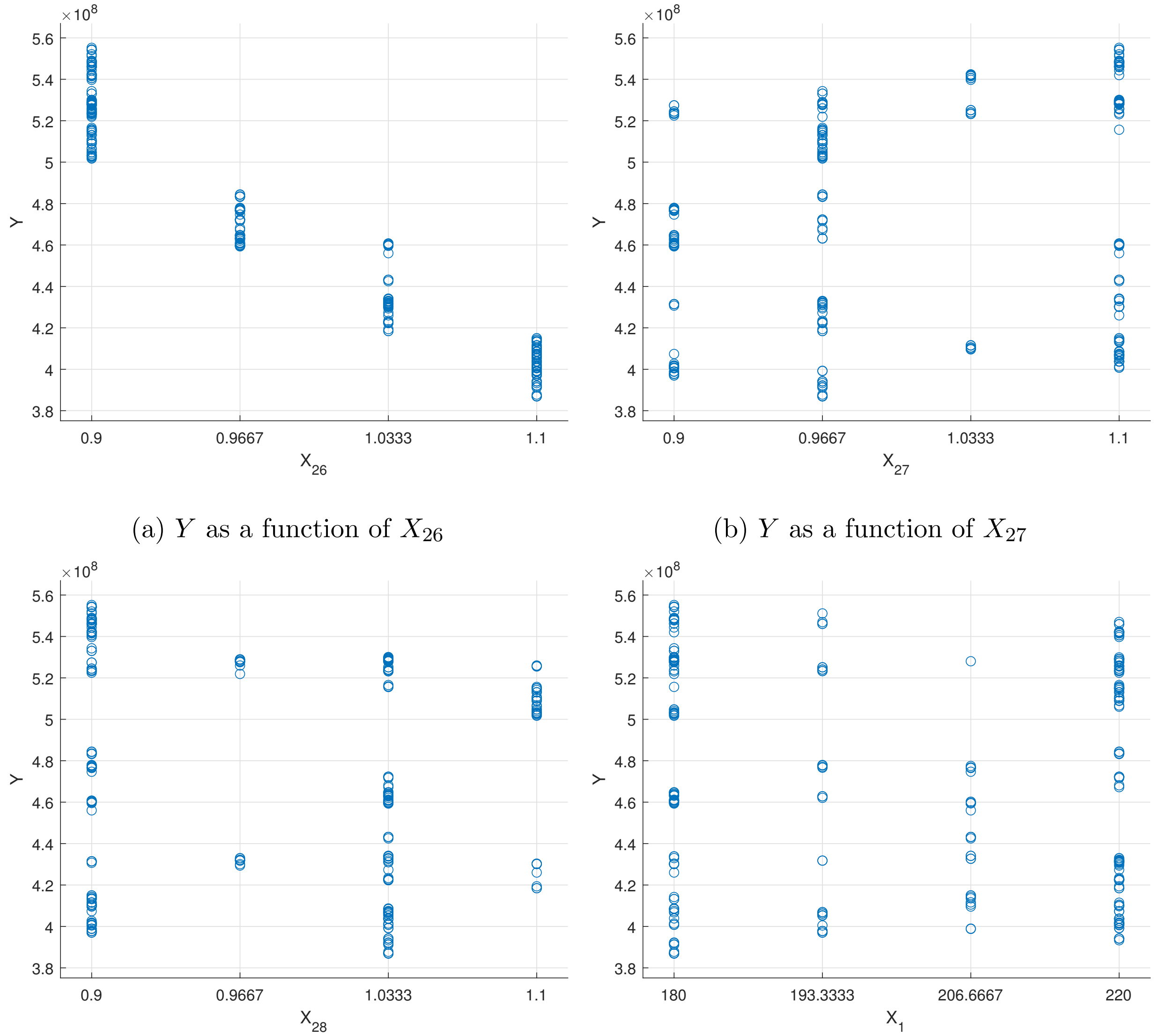


Figure 3.3: 95% confidence intervals for the total sensitivity indices of all 28 inputs

hydraulic conductance of the interface between the aquifer cells and the boundary, and X_{28}, the streambed hydraulic conductance, are relatively important inputs. The other 25 inputs have negligible influence on the output.

Four scatter plots were created to see the correlation between the output Y and each of four inputs: one important input (X_{26}), two relatively important inputs (X_{27} and X_{28}), and one unimportant input (X_1). Figure 3.4a indicates a negative linear correlation between Y and X_{26}. This suggests that increasing recharge flux reduces seawater intrusion into the aquifers.

The first-order and total sensitivity indices of the first 25 inputs are zero since the 95%



(a) Y as a function of X_{26}

(b) Y as a function of X_{27}

(c) Y as a function of X_{28}

(d) Y as a function of X_1

Figure 3.4: Scatter plots of output Y as a function of input X_i (i = 26, 27, 28, and 1)

confidence intervals for the first-order and total sensitivity indices of the first 25 inputs include zero, which means they have no effect on the output. Therefore, their values can be fixed at their nominal values to simplify the model.

3.5 Conclusion

If aquifers are located near the ocean, excessive groundwater pumping can lead to seawater intrusion. To prevent seawater from infiltrating the underground aquifers, the recharge flux must be carefully monitored and controlled. The flux can be monitored by measuring the hydraulic heads along the boundary and controlled by limiting groundwater extraction and implementing artificial recharge.

Additionally, more detailed information on the hydraulic conductance of the interface between the aquifer cells and the boundary and the streambed hydraulic conductance would improve predictions of the net seawater intrusion flux.

The global sensitivity analysis was conducted assuming independent inputs uniformly distributed. Further research is needed to investigate the impact of correlated inputs.

Chapter 4

Extending Battery Life with Global Sensitivity Analysis

4.1 Introduction

Batteries are ubiquitous in modern life, powering everything from smartphones and laptops to electric vehicles. Among various battery technologies, lithium-ion batteries have become dominant due to their high energy density. However, accurately predicting the performance and lifespan of lithium-ion batteries remains a significant challenge.

The increasing demand for electric vehicles and the growing emphasis on renewable energy sources necessitate advancements in battery modeling. Accurate battery models are crucial for several key areas: optimizing battery design and manufacturing; developing advanced battery management systems (BMS) that enable real-time monitoring and control of battery operation for safe and efficient use; and predicting battery degradation to accurately forecast battery performance over its lifetime, optimizing maintenance schedules and ensuring safe operation.

4.2 Methods

The Morris method and the Sobol' method were employed to perform global sensitivity analysis of a lithium-ion battery model.

Sensitivity Analysis Library in Python (SALib) [6, 7] was used to sample points for both methods.

The Doyle-Fuller-Newman (DFN) model in Python Battery Mathematical Modelling (PyBaMM) [21] was used to simulate 1C constant-current discharge using the points from SALib as the input values.

SALib was used to calculate the sensitivity measures of both methods.

4.2.1 The Doyle-Fuller-Newman Model

There are many lithium-ion battery models: single particle model, multiple particle model, the Doyle-Fuller-Newman (DFN) model, and so on. The DFN model was used for the analysis.

A set of governing equations for the DFN model is followed.

The volume-average approximation for conservation of charge in the solid phase of the porous electrode can be described as follows [13]

$$\nabla \cdot (\sigma_{eff} \nabla \bar{\phi}_s) = a_s F \bar{j}, \tag{4.1}$$

where σ_{eff} is the effective conductivity in the solid phase, $\bar{\phi}_s$ is the average of electric potential in the solid phase, a_s is the specific interfacial surface area, F is the Faraday's constant, and $\bar{j}$ is the average of the Butler-Volmer flux density.

The solid-phase mass conservation equation:

$$\frac{\partial c_s}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(D_s r^2 \frac{\partial c_s}{\partial r}\right), \tag{4.2}$$

where c_s is the lithium-ion concentration in the solid phase, r is the radius of spherical particles the solid phase, and D_s is the diffusion coefficient in the solid phase.

The volume-average approximation for charge conservation in the electrolyte phase of the porous electrode, which is

$$\nabla \cdot (\kappa_{eff}\nabla\bar{\phi}_e + \kappa_{D,eff}\nabla ln\bar{c}_e) + a_s F\bar{j} = 0, \tag{4.3}$$

where i_0 is the exchange current density The volume-average approximation for mass conservation in the electrolyte phase of the porous electrode:

$$\frac{\partial(\epsilon_e\bar{c}_e)}{\partial t} = \nabla \cdot (D_{e,eff}\nabla\bar{c}_e) + a_s(1 - t_+^0)\bar{j}. \tag{4.4}$$

The Bulter-Volmer equation couples these four equations and is expressed as

$$\bar{j} = \frac{\bar{i}_0}{F}\left[exp\left(\frac{(1-\alpha)F}{RT}\eta\right) - exp\left(-\frac{\alpha F}{RT}\eta\right)\right] \tag{4.5}$$

The boundary conditions for the conservation of charge in the solid phase of the porous electrode equation are

$$-\sigma_{eff}\frac{\partial\bar{\phi}_s}{\partial x}\Big|_{x=0} = \frac{i_{app}}{A_{surf}} \tag{4.6}$$

$$\sigma_{eff}\frac{\partial\bar{\phi}_s}{\partial x}\Big|_{x=\delta-} = 0 \quad \text{at the negative electrode} \tag{4.7}$$

$$\sigma_{eff}\frac{\partial\bar{\phi}_s}{\partial x}\Big|_{x=L-\delta+} = 0 \tag{4.8}$$

$$-\sigma_{eff}\frac{\partial\bar{\phi}_s}{\partial x}\Big|_{x=L} = \frac{i_{app}}{A_{surf}} \quad \text{at the positive electrode} \tag{4.9}$$

The boundary conditions for the solid phase mass conservation equation (concentration in the electrode active material) are

$$\frac{\partial c_s}{\partial r}\Big|_{r=0} = 0 \tag{4.10}$$

$$-D_s\frac{\partial c_s}{\partial r}\Big|_{r=R_s} = \bar{j} \tag{4.11}$$

The boundary conditions for the conservation of mass in the electrolyte phase of the porous electrode are

$$D_{e,eff}\frac{\partial c_e}{\partial x}\Big|_{x=0} = D_{e,eff}\frac{\partial c_e}{\partial x}\Big|_{x=L} = 0. \tag{4.12}$$

The boundary conditions for the conservation of charge in the electrolyte phase of the porous electrode are

$$\kappa_{eff}\frac{\partial\bar{\phi}_e}{\partial x}\Big|_{x=0} = \kappa_{eff}\frac{\partial\bar{\phi}_e}{\partial x}\Big|_{x=L} = 0. \tag{4.13}$$

4.2.2 The Output and Inputs of the Model

The output of the model for global sensitivity analysis is the time it takes until the terminal voltage hits the lower cut-off voltage which is set to 3.105V.

The inputs are the current collector thickness, the electrode thickness, the current collector conductivity, the current collector density, the current collector specific heat capacity, and the current collector thermal conductivity for both positive and negative electrodes, as well as the separator thickness. All 13 inputs are listed in Table 4.1.

Table 4.1: The 13 inputs of the lithium-ion battery model

Input name	Input description
X_1	Negative current collector thickness $[m]$
X_2	Negative electrode thickness $[m]$
X_3	Separator thickness $[m]$
X_4	Positive electrode thickness $[m]$
X_5	Positive current collector thickness [m]
X_6	Negative current collector conductivity $[S \cdot m^{-1}]$
X_7	Positive current collector conductivity $[S \cdot m^{-1}]$
X_8	Negative current collector density $[kg \cdot m^{-3}]$
X_9	Positive current collector density $[kg \cdot m^{-3}]$
X_{10}	Negative current collector specific heat capacity $[J \cdot kg^{-1} \cdot K^{-1}]$
X_{11}	Positive current collector specific heat capacity $[J \cdot kg^{-1} \cdot K^{-1}]$
X_{12}	Negative current collector thermal conductivity $[W \cdot m^{-1} \cdot K^{-1}]$
X_{13}	Positive current collector thermal conductivity $[W \cdot m^{-1} \cdot K^{-1}]$

4.2.3 Settings for the Morris Method

All the inputs were assumed to be uniformly distributed within $\pm 10\%$ of their nominal values because the distributions of the inputs are unknown. 500 trajectories were randomly generated, and 10 optimal trajectories were selected from them. The number of grid levels were set to four. Since there are 13 inputs, $10 \times (13 + 1) = 140$ points were sampled.

The sample means of the absolute values of elementary effects of all 13 inputs and the sample standard deviations of the elementary effects of all 13 inputs were calculated.

4.2.4 Settings for the Sobol' Method

All the inputs were assumed to be uniformly distributed within $\pm 10\%$ of their nominal values because the distributions of the inputs are unknown. A base sample N was set to 1024 which is a power of 2. Since there are 28 inputs, $1024 \times (13 + 2) = 30720$ points in the 56-dimensional Sobol' sequence were sampled.

The first-order and total sensitivity indices of all 13 inputs were estimated with 95% confidence intervals.

4.3 Results

Results of global sensitivity analysis, performed using the Morris method and the Sobol' method, are detailed in this section.

4.3.1 Results from the Morris Method

Figure 4.1 shows estimated means of the absolute values of elementary effects ($\overline{|D_i|}$) and estimated standard deviations of the elementary effects (s_i) for all 13 inputs ($i = 1, \ldots, 13$). One input with an estimated mean greater than 100 is marked with its name (X_4).

The estimated mean and standard deviation for X_4, $\overline{|D_4|}$ and s_4, are 724.6988 and 6.4513, respectively. The estimated means and standard deviations for the other inputs are zero or close to zero. All estimated means and standard deviations are listed in Table A.3

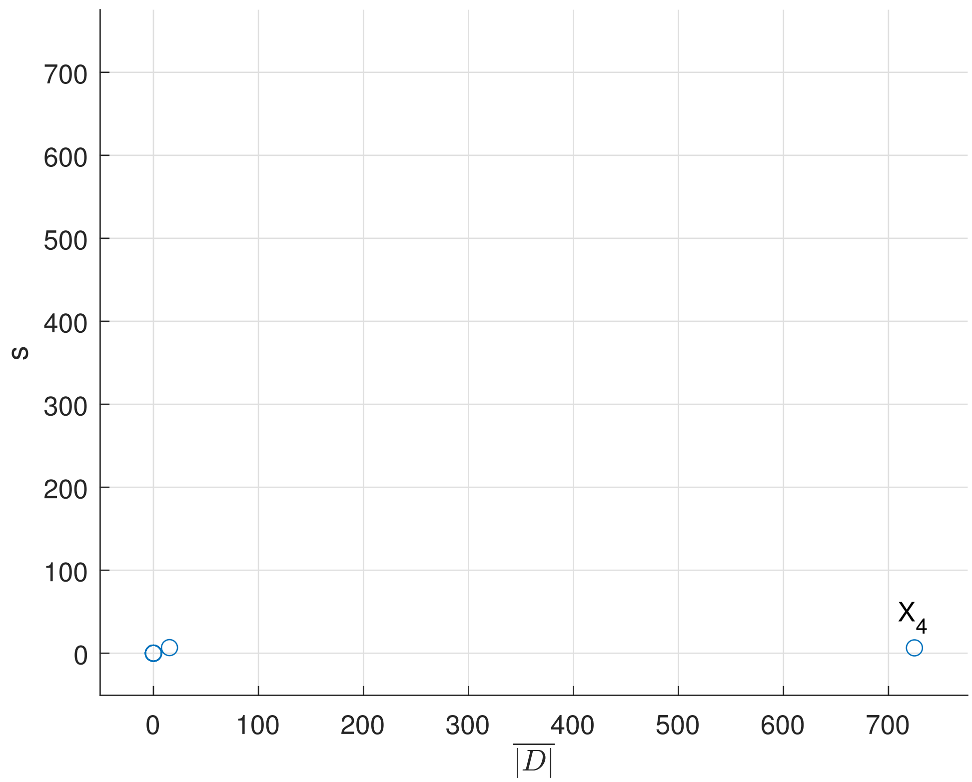


Figure 4.1: Estimated means ($\overline{|D_i|}$) and standard deviations (s_i) of the (absolute) values of elementary effects for all 13 inputs (One input with an estimated mean greater than 100 is marked with its name.)

4.3.2 Results from the Sobol' Method

Figures 4.2 and 4.3 show 95% confidence intervals for the first-order and total sensitivity indices of all 13 inputs, respectively. Both figures look very similar to each other.

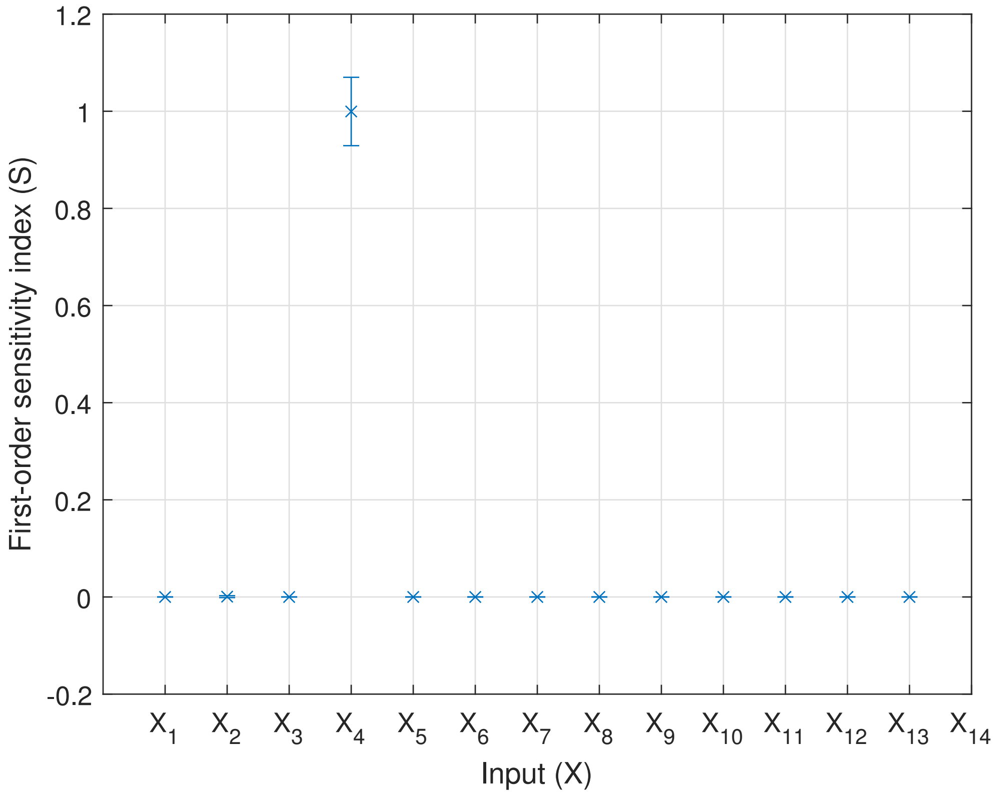


Figure 4.2: 95% confidence intervals for the first-order sensitivity indices of all 13 inputs

The 95% confidence interval for the first-order sensitivity index of X_4, S_4, is 0.9995 $\pm$ 0.0704. The 95% confidence intervals for the first-order sensitivity indices of the other inputs are around zero, including it. All confidence intervals for the first-order sensitivity indices are listed in Table A.4.

The 95% confidence interval for the total sensitivity index of X_4, ST_4, is 0.9995 $\pm$ 0.0599. The 95% confidence intervals for the total sensitivity indices of the other inputs

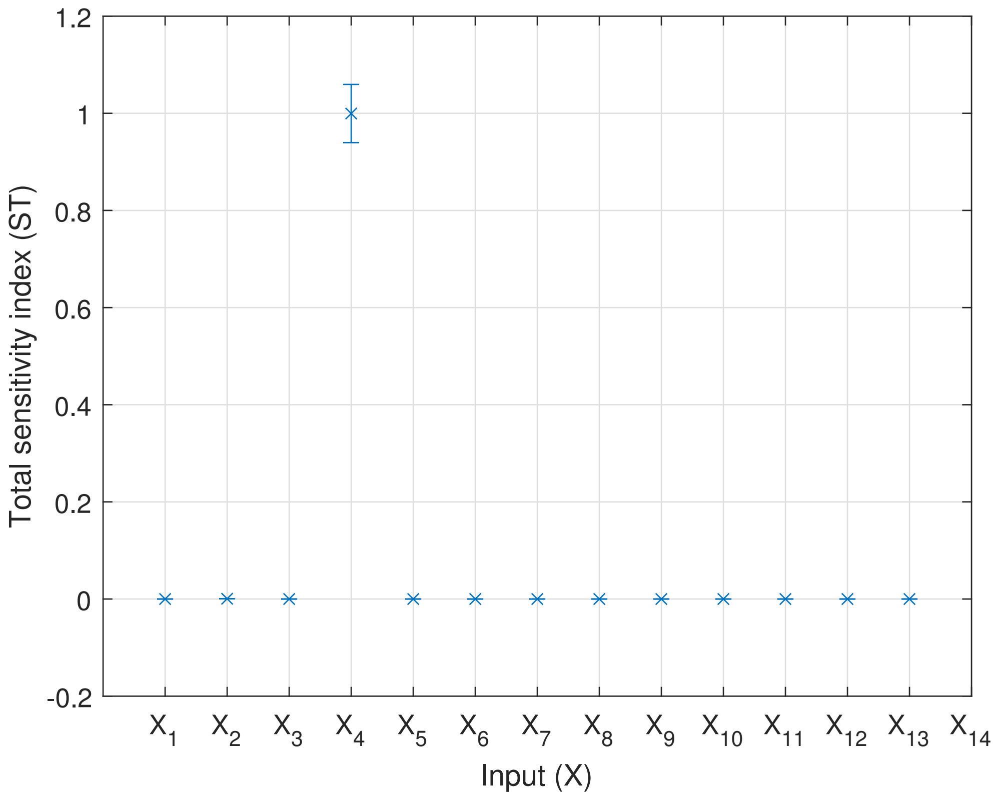


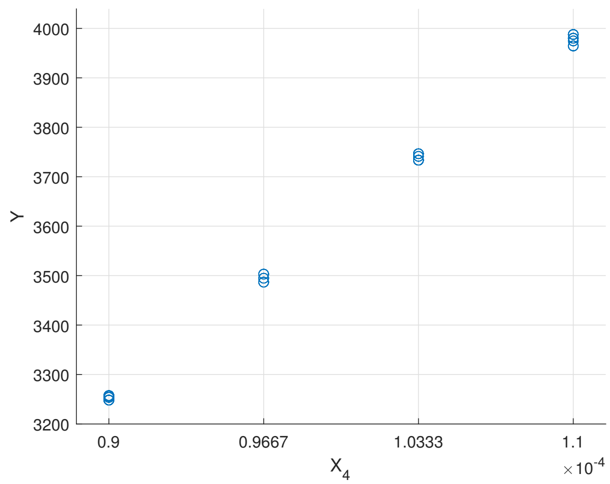
Figure 4.3: 95% confidence intervals for the total sensitivity indices of all 13 inputs

are around zero, including it. All confidence intervals for the total sensitivity indices are listed in Table A.4.

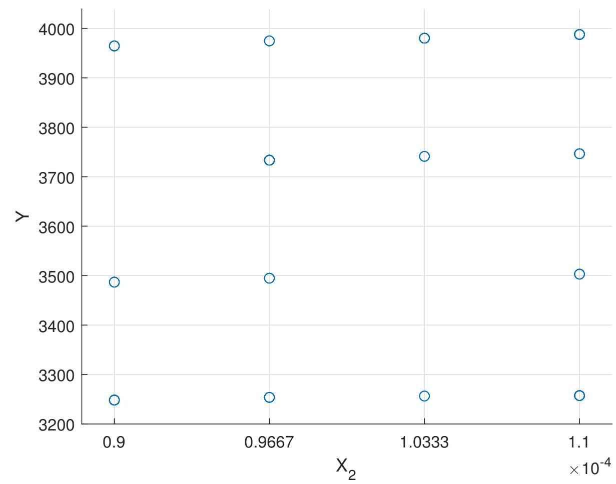
4.4 Discussion

Both the Morris method and the Sobol' method yield the same results. When the output of interest is the battery life, X_4, the positive electrode thickness, is the most influential input to the lithium-ion model. The other 12 inputs have negligible influence on the output.

Two scatter plots were created to see the correlation between the output Y and each of two inputs: one important input (X_4) and one unimportant input (X_2). Figure 4.4a indicates a positive linear correlation between Y and X_4. This suggests that increasing the thickness of the positive electrode increases the battery life.



(a) Y vs. X_4



(b) Y vs. X_2

Figure 4.4: Scatter plots of output Y as a function of input X_i ($i = 4$ and 2)

The first-order and total sensitivity indices of the 12 inputs are zero since the 95% confidence intervals for the first-order and total sensitivity indices of the 12 inputs include zero, which means they have no effect on the output. Therefore, their values can be fixed at their nominal values to simplify the model.

A comparison of the terminal voltage curves having the minimum and maximum life is shown in 4.5.

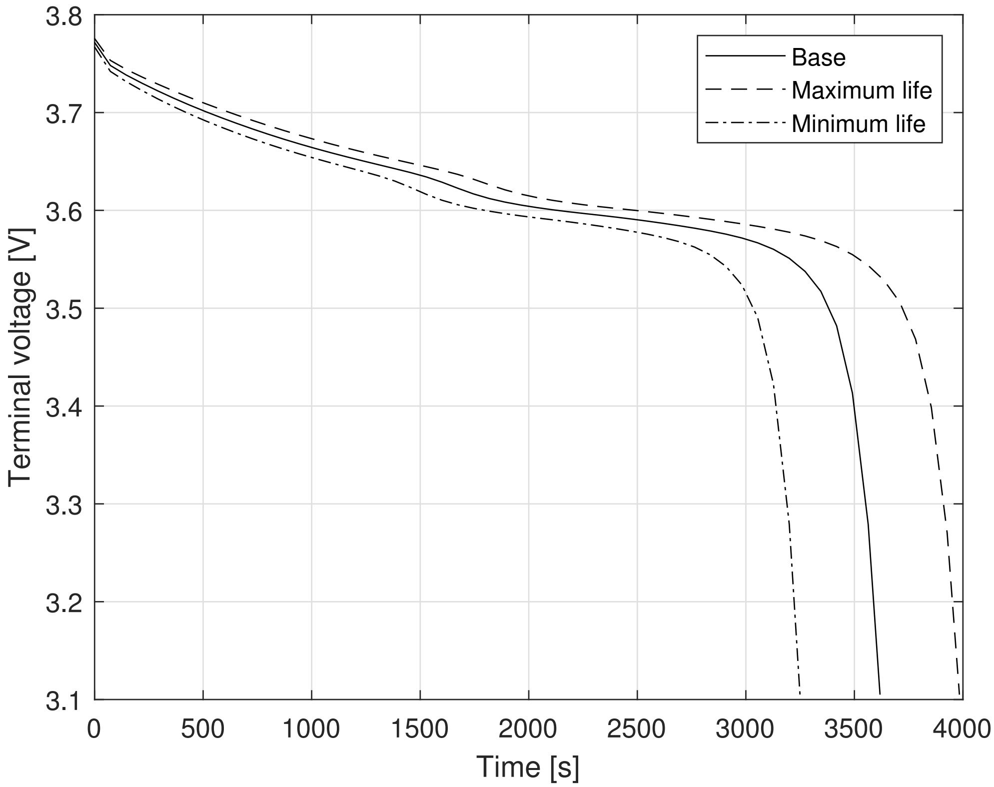


Figure 4.5: A comparison of the terminal voltage curves having the minimum and maximum life

4.5 Conclusion

For the lithium-ion battery model, the influence of 13 inputs on battery life during a 1C discharge rate was investigated. Specifically, the thickness of the positive electrode emerged as the most influential input. These results suggest that thickening the positive electrode can lead to increased battery life.

The global sensitivity analysis was conducted assuming independent inputs uniformly distributed. Further research is needed to investigate the impact of correlated inputs.

Chapter 5

Conclusion

Global sensitivity analysis (GSA) was performed on two complex models—a groundwater flow model and a lithium-ion battery model—to rank the importance of their input parameters. Two well-established GSA methods were employed: the Morris method and the Sobol' method.

For the groundwater flow model, both methods consistently identified recharge flux as the most influential input with respect to seawater intrusion.

In the lithium-ion battery model, the thickness of the positive electrode was identified as the most influential input for battery life.

A key assumption of the Sobol' method is the independence of input parameters. However, real-world input parameters may exhibit dependencies or correlations. Future research could explore advanced techniques for handling correlated inputs within the Sobol' framework.

Applying GSA to models across various disciplines enables the effective identification and prioritization of key inputs, leading to improved model reliability and predictive capabilities. GSA results can improve model structure optimization, reduce computational cost, and enhance model accuracy.

Appendix A

Detailed Research Data

The data used for the global sensitivity analysis are presented in this appendix. All values in the tables, except for zeros, are presented in fixed-decimal format with four decimal places, consistent with the MATLAB 'short' format.

In the SALib implementation of the Morris method, the random number generator seed was set to 1. This ensures reproducibility of the results, as setting the seed to 0 did not consistently generate the same sample across different runs.

sample(problem, 500, num_levels=4, optimal_trajectories=10, seed=1)

A.1 The Groundwater Flow Model

Table A.1 presents the sample means and standard deviations of the (absolute) values of the elementary effects. Table A.2 presents the 95% confidence intervals for the first-order and total sensitivity indices.

Table A.1: The sample means ($\overline{|D_i|}$) and standard deviations (s_i) of the (absolute) values of the elementary effects for all 28 inputs of the groundwater flow model

i	$\overline{\lvert D_i\rvert}$	s_i
1	0.0032×10^8	0.0027×10^8
2	0.0559×10^8	0.0177×10^8
3	0.0005×10^8	0.0003×10^8
4	0.0578×10^8	0.0208×10^8
5	0.0067×10^8	0.0042×10^8
6	0.0002×10^8	0.0002×10^8
7	0.0104×10^8	0.0008×10^8
8	0.0006×10^8	0.0005×10^8
9	0.0136×10^8	0.0007×10^8
10	0.0006×10^8	0.0006×10^8
11	0.0107×10^8	0.0007×10^8
12	0.0586×10^8	0.0050×10^8
13	0.0017×10^8	0.0016×10^8
14	0.0091×10^8	0.0022×10^8

A.2 The Lithium-Ion Battery Model

Table A.3. presents the sample means and standard deviations of the (absolute) values of elementary effects. Table A.4 presents the 95% confidence intervals for the first-order

Table A.1: The sample means ($\overline{|D_i|}$) and standard deviations (s_i) of the (absolute) values of the elementary effects for all 28 inputs of the groundwater flow model (Continued)

i	$\overline{\|D_i\|}$	s_i
15	0.0002×10^8	0.0000×10^8
16	0.0117×10^8	0.0007×10^8
17	0.0174×10^8	0.0013×10^8
18	0.0160×10^8	0.0008×10^8
19	0.0103×10^8	0.0008×10^8
20	0.0133×10^8	0.0009×10^8
21	0.0012×10^8	0.0007×10^8
22	0.0026×10^8	0.0006×10^8
23	0.0006×10^8	0.0001×10^8
24	0.0002×10^8	0.0002×10^8
25	0.0011×10^8	0.0005×10^8
26	1.2779×10^8	0.1239×10^8
27	0.1749×10^8	0.0254×10^8
28	0.2282×10^8	0.0434×10^8

and total sensitivity indices.

Table A.2: The 95% confidence intervals (CI) for the first-order (S_i) and total sensitivity indices (ST_i) for all 28 inputs of the groundwater flow model

i	CI for S_i	CI for ST_i
1	0.0000 ± 0.0002	0.0000 ± 0.0000
2	0.0016 ± 0.0037	0.0015 ± 0.0001
3	-0.0000 ± 0.0000	0.0000 ± 0.0000
4	0.0024 ± 0.0046	0.0025 ± 0.0002
5	0.0000 ± 0.0006	0.0000 ± 0.0000
6	-0.0000 ± 0.0000	0.0000 ± 0.0000
7	0.0001 ± 0.0007	0.0001 ± 0.0000
8	0.0000 ± 0.0000	0.0000 ± 0.0000
9	-0.0004 ± 0.0009	0.0001 ± 0.0000
10	0.0000 ± 0.0001	0.0000 ± 0.0000
11	0.0001 ± 0.0007	0.0001 ± 0.0000
12	0.0017 ± 0.0042	0.0019 ± 0.0002
13	0.0000 ± 0.0001	0.0000 ± 0.0000
14	0.0001 ± 0.0007	0.0001 ± 0.0000

Table A.2: The 95% confidence intervals (CI) for the first-order (S_i) and total sensitivity indices (ST_i) for all 28 inputs of the groundwater flow model (Continued)

i	CI for S_i	CI for ST_i
15	0.0000 ± 0.0000	0.0000 ± 0.0000
16	-0.0002 ± 0.0008	0.0001 ± 0.0000
17	0.0002 ± 0.0011	0.0002 ± 0.0000
18	0.0001 ± 0.0009	0.0001 ± 0.0000
19	0.0001 ± 0.0008	0.0001 ± 0.0000
20	0.0001 ± 0.0010	0.0001 ± 0.0000
21	0.0000 ± 0.0001	0.0000 ± 0.0000
22	-0.0001 ± 0.0001	0.0000 ± 0.0000
23	-0.0000 ± 0.0000	0.0000 ± 0.0000
24	0.0000 ± 0.0000	0.0000 ± 0.0000
25	0.0006 ± 0.0011	0.0001 ± 0.0003
26	0.9539 ± 0.0749	0.9443 ± 0.0635
27	0.0182 ± 0.0126	0.0178 ± 0.0016
28	0.0259 ± 0.0155	0.0250 ± 0.0024

Table A.3: The sample means ($\overline{|D_i|}$) and standard deviations (s_i) of the (absolute) values of the elementary effects for all 13 inputs of the lithium-ion battery model

i	$\overline{\|D_i\|}$	s_i
1	0	0
2	15.3023	6.7669
3	0.0563	0.0107
4	724.6988	6.4513
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0

Table A.4: The 95% confidence intervals (CI) for the first-order (S_i) and total sensitivity indices (ST_i) for all 13 inputs of the lithium-ion battery model

i	CI for S_i	CI for ST_i
1	0 ± 0	0 ± 0
2	0.0007 ± 0.0020	0.0006 ± 0.0000
3	-0.0000 ± 0.0000	0.0000 ± 0.0000
4	0.9995 ± 0.0704	0.9995 ± 0.0599
5	0 ± 0	0 ± 0
6	0 ± 0	0 ± 0
7	0 ± 0	0 ± 0
8	0 ± 0	0 ± 0
9	0 ± 0	0 ± 0
10	0 ± 0	0 ± 0
11	0 ± 0	0 ± 0
12	0 ± 0	0 ± 0
13	0 ± 0	0 ± 0

References

[1] S. Tarantola A. Saltelli and K. P.-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

[2] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518, 2007.

[3] Rober I. Cukier, H. B. Levine, and Kurt E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics*, 26(1):1–42, 1978.

[4] B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

[5] A.W. Harbaugh. *A computer program for calculating subregional water budgets using results from the U.S. Geological Survey Modular Three-Dimensional Finite- Difference Ground-Water Flow Model*. 1990.

[6] Jon Herman and Will Usher. SALib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9), January 2017.

[7] Takuya Iwanaga, William Usher, and Jonathan Herman. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling*, 4:18155, May 2022.

[8] Michiel J.W. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1):35–43, 1999.

[9] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[10] Richard G. Niswonger, Sorab Panday, and Motomu Ibaraki. *MODFLOW-NWT, a Newton formulation for MODFLOW-2005*. 2011.

[11] Art B. Owen. On dropping the first sobol' point, 2021.

[12] George F Pinder and Michael A Celia. *Subsurface hydrology.* John Wiley & Sons, 2006.

[13] Gregory L. Plett. *Battery management systems volume I : battery modeling.* Artech House power engineering series. Artech House, Boston, Mass, 2015.

[14] Elmar Plischke. An effective algorithm for computing global sensitivity indices (easi). *Reliability Engineering & System Safety*, 95(4):354–360, 2010.

[15] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, May 2002.

[16] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, February 2010.

[17] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer.* John Wiley & Sons, Ltd, 2008.

[18] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1(4):407–414, 1993.

[19] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1–3):271–280, February 2001.

[20] I.M. Sobol. On quasi-monte carlo integrations. *Mathematics and Computers in Simulation*, 47(2):103–112, 1998.

[21] Valentin Sulzer, Scott G. Marquis, Robert Timms, Martin Robinson, and S. Jon Chapman. Python Battery Mathematical Modelling (PyBaMM). *Journal of Open Research Software*, 9(1):14, 2021.

[22] Jason Sun, John Lindquist, Dan Detmer, and Timothy Moore. Ventura regional groundwater flow model and updated hydrogeologic conceptual model: Oxnard plain, oxnard forebay, pleasant valley, west las posas, and mound basins. *United Water Conservation District Open-File Report*, July 2018.

[23] Stefano Tarantola, Debora Gatelli, and Thierry Alex Mara. Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety*, 91(6):717–727, 2006.

[24] Jean-Yves Tissot and Clémentine Prieur. Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering & System Safety*, 107:205–213, 2012. SAMO 2010.

[25] D.K. Todd and L.W. Mays. *Groundwater Hydrology.* Wiley, 2004.