Sensitivity Analysis of Model Output: numerical algorithms for evaluating Sobol' sensitivity indices

Rayna Georgieva

rayna@parallel.bas.bg

(in collaboration with Sofiya Ivanovska)

IPP - Bulgarian Academy of Sciences,
Sofia, Bulgaria

NATO Grant PDD(TC)-(ESP.EAP.CLG 982641)
"Monte Carlo Sensitivity Studies of Environmental Security",

The University of Reading, March 22, 2007

Outline

- Introduction
- Formulation of the problem
- Methods for evaluating global sensitivity indices
- Sobol' approach for evaluating global sensitivity indices
- Numerical algorithms for multidimensional integration
- Numerical experiments
- Conclusion

Motivation

"The real voyage of discovery consists not in seeking new landscapes but in having new eyes." (Marcel Proust)

Mathematical modelling is the use of mathematics to describe real-world phenomena:

- investigate important questions about the observed world
- explain real-world phenomena
- test ideas
- make predictions about the real world

Motivation

Jack Hardisty. ”Numerical experiments on the orthogonal profile with particular reference to changing environmental parameters.” *Landscape Sensitivity*, pp. 177–185, 1993.

”An appreciation of the stability, or otherwise, of an environmental system is achieved through the use of sensitivity analysis. This approach generally recognises that the mathematical representation of the system will respond to either changes in the input parameters (parameter sensitivity) or to changes in the representation of the internal transfer functions (process sensitivity).”

Motivation

Sensitivity analysis (SA) is used to determine:

- robustness, reliability, efficiency of a model

Applied fields:

- environmental modeling, financial applications, risk analysis, signal processing, neural networks

Air pollution modeling:

- initial conditions, boundary conditions, or chemical reaction rate constants

Introduction

SA includes:

- defining the model and its input parameters and output variable(s),
- assigning probability density functions to each input parameter,
- generating an input matrix through an appropriate random sampling method, evaluating the output,
- assessing the influences or relative importance of each input parameters on the output variable.

The mathematical model

- The mathematical model

$$y = f(\mathbf{x}), \quad \text{where} \quad \mathbf{x} = (x_1, x_2, \ldots, x_n) \in U^n \equiv [0,1]^n$$

is a vector of input parameters with joint p.d.f. $p(\mathbf{x}) = p(x_1, \ldots, x_n)$.

  - Local sensitivity: $\mathbf{x} = \mathbf{x}^* \in U^n, \quad y^* = f(\mathbf{x}^*)$.
  - Global sensitivity: $y = f(\mathbf{x}), \quad \mathbf{x} \in U^n$.

The mathematical model

- Total Sensitivity Index (TSI) of input parameter $x_i, \quad i \in \{1, \dots, n\}$:

$$TSI(x_i) = S_i + \sum_{l_1 \neq i} S_{il_1} + \sum_{l_1, l_2 \neq i, l_1 < l_2} S_{il_1 l_2} + \ldots + S_{il_1 \ldots l_{n-1}},$$

where

$$S_{il_1 \ldots l_{j-1}} - j^{\text{th}} \text{ order sensitivity index for parameter } x_i \ (1 \leq j \leq n),$$

$$j = 1: \ S_i - \text{"the main effect" of } x_i.$$

- Classification of input parameters:

$$\begin{array}{lcccll}
-\text{very important:} & 0.8 & < & TSI & & \\
-\text{important:} & 0.5 & < & TSI & < & 0.8 \\
-\text{unimportant:} & 0.3 & < & TSI & < & 0.5 \\
-\text{irrelevant:} & & & TSI & < & 0.3
\end{array}$$

An Illustration

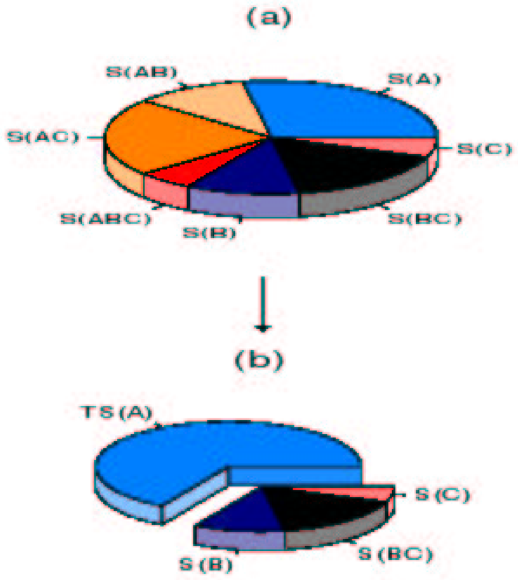


Figure 1: Graphical Representation of (a) Sensitivity Indices for the Three Parameters Case and (b) Total Sensitivity Indices of Parameter A

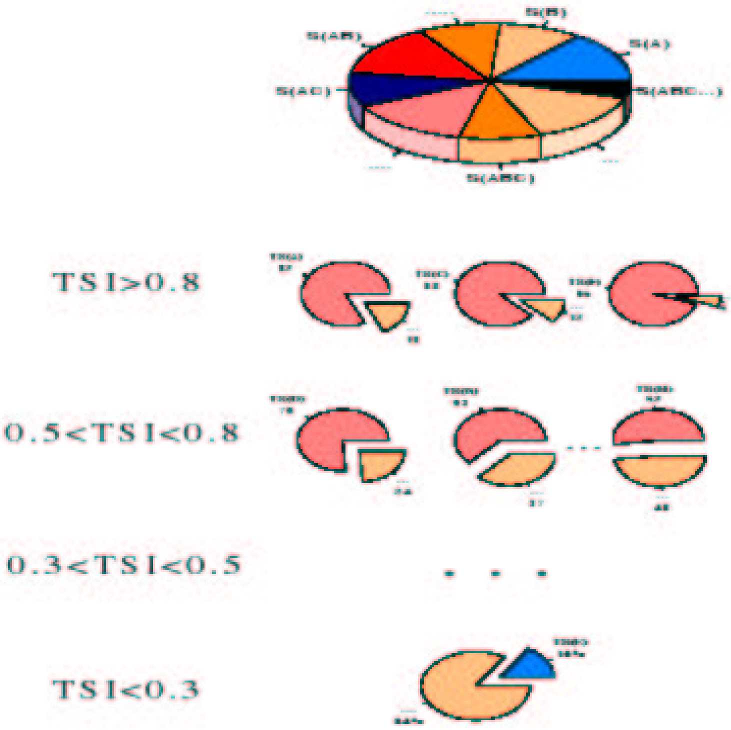


Figure 2: Graphical Representation of Grouping a Set of Input Parameters with the Total Sensitivity Indices

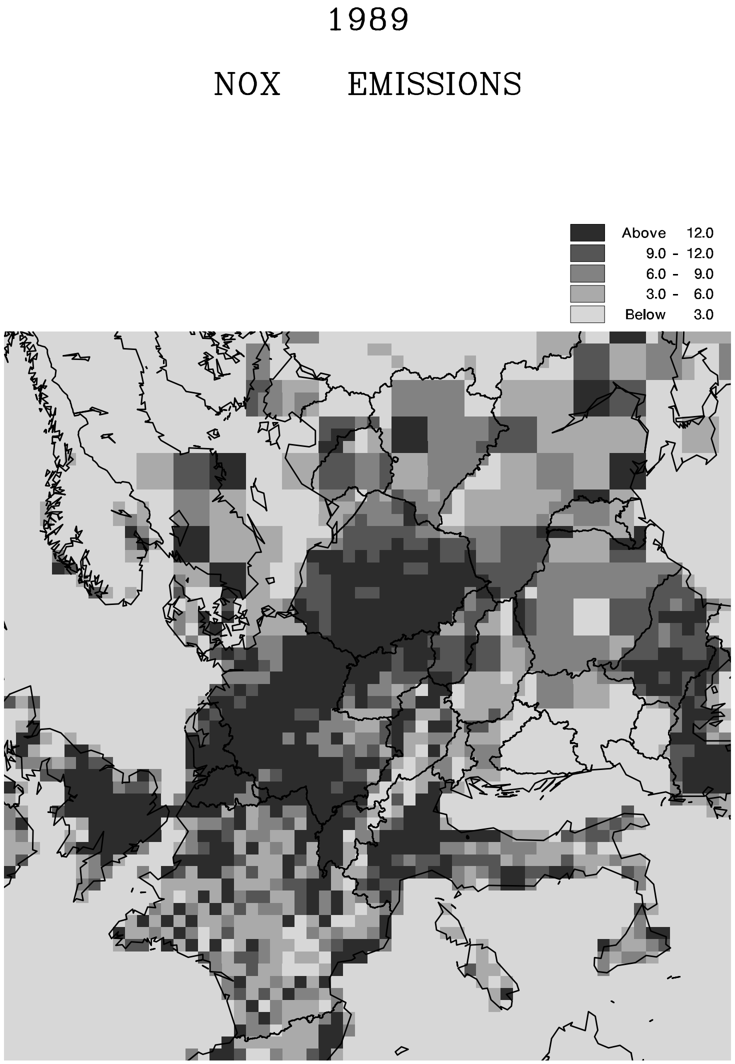
An Example in Air Pollution Transport

- Studing the sensitivity of the model output of the concentrations of some important pollutants
  (like NO_2 and O_3).

- The most important parameter: chemical rate constant of the reaction

$$O_3 + NO \Longrightarrow NO_2.$$

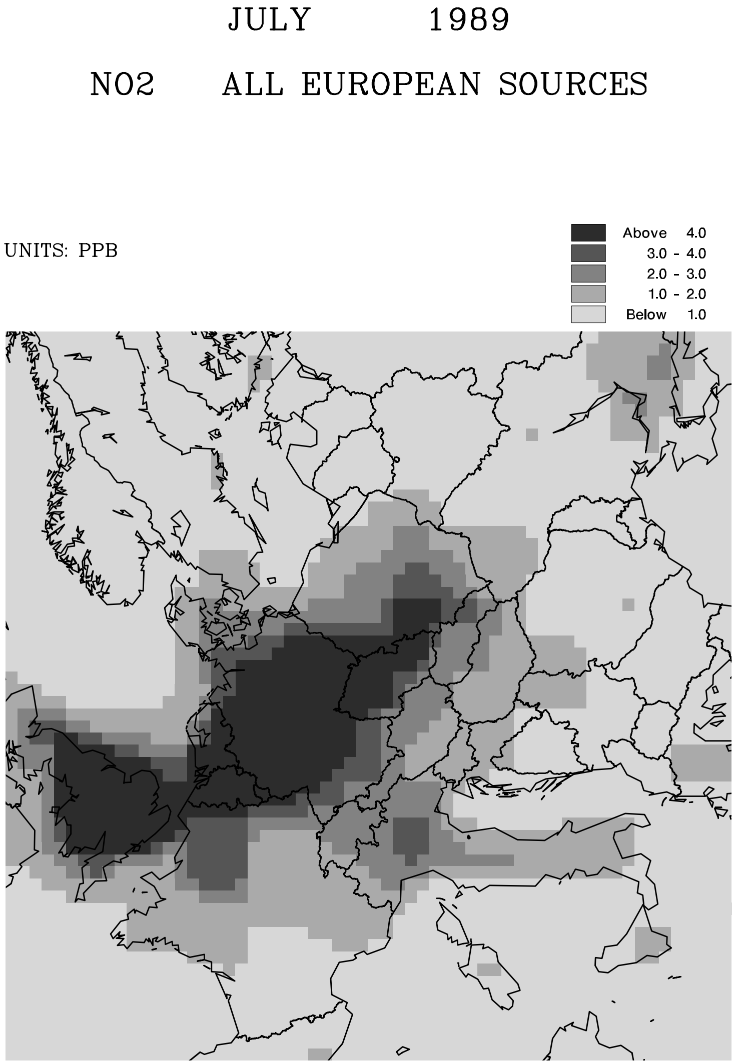
I. Dimov, Z. Zlatev, Testing the Sensitivity of Air Pollution Levels to Variations of Some Chemical Rate Constants, in: "Large-Scale Computations in Engineering and Environmental Sciences", Notes on Numerical Fluid Mechanics, Vol. 62, 1997, pp. 167–175.



1989

NOX EMISSIONS

Figure 1: European nitrogen oxides emissions



JULY 1989

NO2 ALL EUROPEAN SOURCES

UNITS: PPB

Figure 2: Nitrogen di-oxide concentrations in Europe

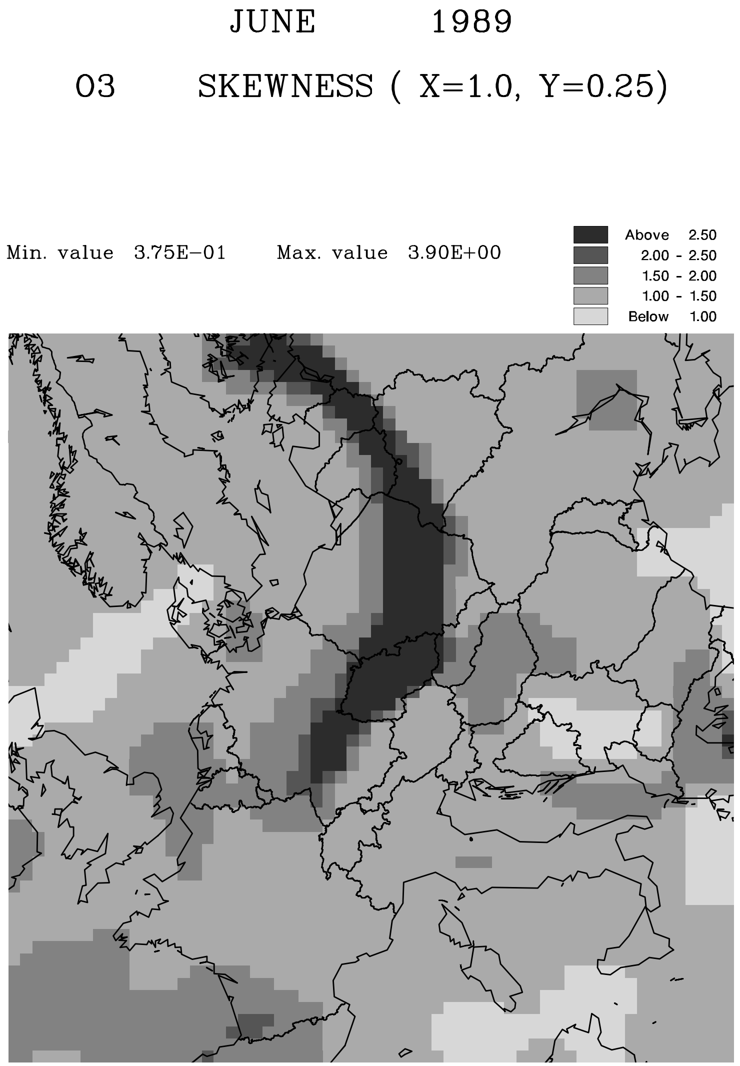


Figure 3: Skewness of the ozone concentrations (variance 0.50)

Methods for Evaluating Global Sensitivity Indices

Method	Cost (Model runs)	Sensitivities
FAST (1973)	$O(n^2)$	$S_i, \forall i$
Sobol (1993)	$N(2n+2)$	$S_i, TSI(x_i), \forall i$
EFAST (1999)	nN	$S_i, TSI(x_i), \forall i$
Saltelli (2002)	$N(n+2)$	$S_i, \forall i, S_{lj}, \forall l, j, l \neq i$

Numerical Example

- A function with separated variables:

$$g(\mathbf{x}) = \prod_{i=1}^{n} \varphi_i(x_i), \text{ where } \varphi_i(t) = \frac{(|4t-2|+a_i)}{(1+a_i)}.$$

- Let $n = 8,\ a_1 = a_2 = 0, a_3 = \ldots = a_8 = 3.$

$$S_1 = S_2 = 0.329, \quad S_3 = \ldots = S_8 = 0.021$$

$$S_{12} = 0.110, \quad S_{ij} = 0.007 \quad \text{if } i, j \in 1, 2, \quad S_{ij} = 0.0004, else$$

$$S_{12k} = 0.002, k \geq 3, \quad S_{ijk} \leq 0.00014, else.$$

A. Saltelli, I.M. Sobol , About the use of rank transformation in sensitivity analysis of model output, Reliab. Eng. Syst. Safety 50 (3) (1995) 225–239.

Sobol' Approach

ANalysis Of VAriances (ANOVA) HDMR of an integrable function $f(\mathbf{x})$:

$$f(\mathbf{x}) = f_0 + \sum_{s=1}^{n} \sum_{l_1<\ldots<l_s} f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s}),$$

where

- f_0 - constant,
- $\displaystyle\int_0^1 f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s})\mathrm{d}x_{l_k} = 0, \quad 1 \le k \le s.$

I.M. Sobol, Multidimensional Quadrature Formulas and Haar Functions, Nauka, Moscow, 1969.
I.M. Sobol, Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 1(4) (1993) 407–414.

Sobol' Approach

Therefore

- $\int_{U^n} f_{i_1,\dots,i_s} f_{j_1,\dots,j_l} \, \mathrm{d}\mathbf{x} = 0, \quad (i_1,\dots,i_s) \neq (j_1,\dots,j_l)$

and the functons in the right-hand side are defined in a unique way:

- $f_0 = \int_{U^n} f(\mathbf{x}) \mathrm{d}\mathbf{x}$
- $f_{l_1}(x_{l_1}) = \int_{U^n} f(\mathbf{x}) \prod_{k \neq l_1} \mathrm{d}\mathbf{x}_k - f_0, \; l_1 \in \{1, 2, \dots, n\}$
- $f_{l_1 l_2}(x_{l_1}, x_{l_2}) = \int_{U^n} f(\mathbf{x}) \prod_{k \neq l_1, l_2} \mathrm{d}\mathbf{x}_k - f_0 - f_{l_1}(x_{l_1}) - f_{l_2}(x_{l_2}),$
  $l_1, l_2 \in \{1, 2, \dots, n\}$

Global (Sobol') Sensitivity Indices

Definition (Sobol'):

$$S_{l_1 \ \ldots \ l_s} = \frac{D_{l_1 \ \ldots \ l_s}}{D}, \qquad s \in \{1, \ldots, n\},$$

where

- variances $D_{l_1 \ \ldots \ l_s} = \int f^2_{l_1 \ \ldots \ l_s} \mathrm{d}x_{l_1} \ldots \mathrm{d}x_{l_s}$,
- total variance $D = \int_{U^k} f^2(\mathbf{x}) \mathrm{d}\mathbf{x} - f_0^2$,

and the following properties hold:

- $S_{l_1 \ \ldots \ l_s} \geq 0, \qquad \sum_{s=1}^{k} \sum_{l_1 < \ldots < l_s}^{k} S_{l_1 \ \ldots \ l_s} = 1.$

Numerical Algorithms for Multidimensional Integration

- Numerical quadratures
- Stochastic methods - Monte Carlo methods, quasi-Monte Carlo methods
- Adaptive algorithms
- Software packages, tools - CUBPACK, CUBA, HIntLib, BASES/SPRING, Mathematica
- Sensitivity packages - SimLab

Numerical Experiments

Example1:

$$I_1 = \int_{U^4} \frac{4x_1x_3^2exp(2x_1x_3)}{(1+x_2+x_4)^2}\mathrm{d}\mathbf{x}.$$

Example2:

$$I_2 = \int_{U^5} \frac{4x_1x_3^2exp(2x_1x_3)exp(x_5)}{(1+x_2+x_4)^2}\mathrm{d}\mathbf{x}.$$

Example3:

$$I_3 = \int_{U^{10}} \frac{4x_1x_3^2exp(2x_1x_3)}{(1+x_2+x_4)^2}exp(x_5+\ldots+x_{10})\mathrm{d}\mathbf{x}.$$

Example4:

$$I_4 = \int_{U^{25}} \frac{4x_1x_3^2exp(2x_1x_3)}{(1+x_2+x_4)^2}exp(x_5+\ldots+x_{20})x_{21}\ldots x_{25}\mathrm{d}\mathbf{x}.$$

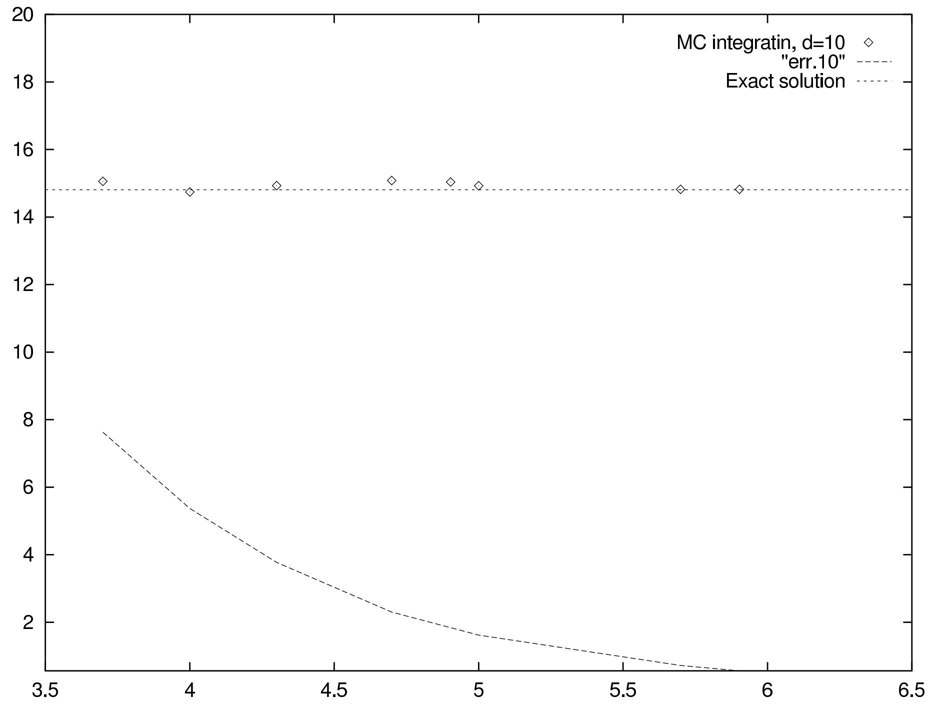


Figure 4: Superconvergent Adaptive MC integration for **Example 3.** and dependance of the error function on log n

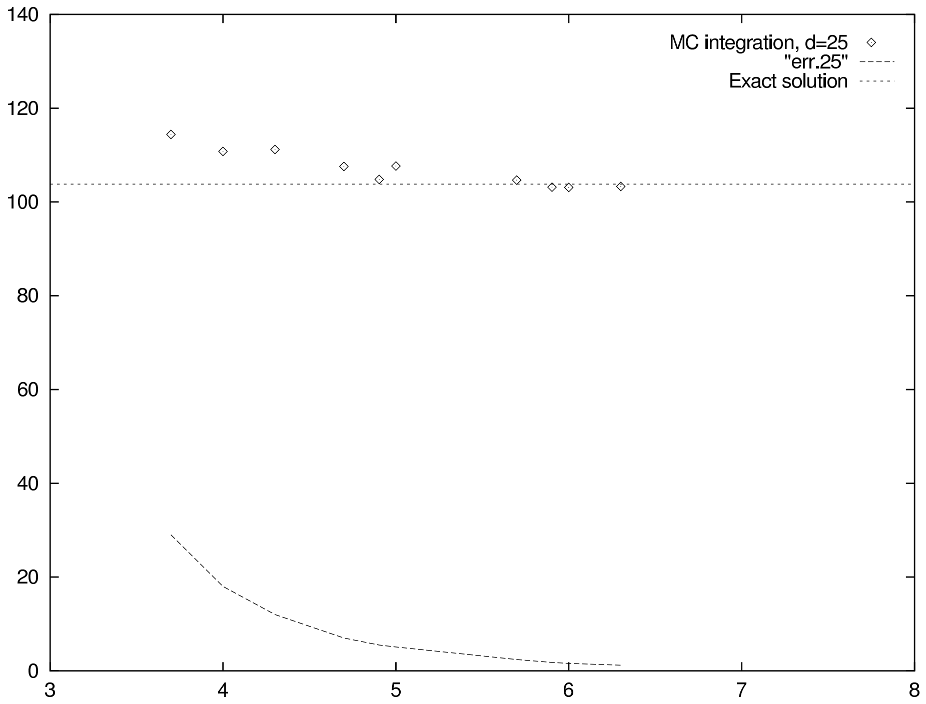


Figure 5: Adaptive MC integration for **Example 4.**

Numerical Experiments

dim	Exact value	Mathematica 5.1
4	0.57536	0.577898 (MC)
5	0.98864	0.9886378 (GC)
10	14.80844	14.9318 (MC)

N	Relative error	MC Solution
10^3	0.027	14.41426
10^4	0.0074	14.91811
10^5	0.0064	14.71296
10^6	0.00086	14.82073

Conclusion

- SA is a useful tool in the model building process
- Sobol' sensitivity indices measure "the main effect" of the input parameters as well as the interaction with other input parameters is taken into account