VETOR

IPSAL: Implementation of the module to generate the Sobol sequence and indices☆

IPSAL: Implementação do módulo para geração da sequência e índices de Sobol

Luiz Felipe Alves Borges[1], Fábio Freitas Ferreira[2,†], Fábio Gonçalves[2], Antônio Espósito Junior[2], Aline Fernanda da Silva Oliveira[1], Wagner Rambaldi Telles[3]

[1]*Fluminense Federal University, Production Engineering - Rio das Ostras, Brazil*
[2]*Fluminense Federal University, Department of Natural Sciences, Rio das Ostras, Brazil*
[3]*Federal Fluminense University, Department of Exact, Biological and Earth Sciences, Santo Antônio de Pádua, Brazil*
[†]**Corresponding author:** fabiofreitasferreira@id.uff.br

Abstract

Sensitivity and uncertainty analysis hold significant importance across a range of applications, spanning from industrial problems to climate change, financial risk assessment, as well as mathematical and computational models. These analyses involve identifying influential input parameters in models to comprehend their impact on the output. Sensitivity analysis can be performed locally, examining parameter effects at a fixed value, or globally, evaluating the model across a range of parameter values. The Sobol method stands as a robust approach for global sensitivity analysis, employing a Sobol sequence to create samples more uniformly within the input parameter space, thus enabling efficient exploration of model inputs. This paper aims to introduce a computational implementation in Scilab to generate the Sobol sequence for utilization in sensitivity analysis through the Sobol method. A test case was applied to generate Sobol sequences and discuss the obtained results.

Keywords

sensitivity analysis • Sobol method • unusual operation • parameter correlation

Resumo

A análise de sensibilidade e da incerteza tem uma importância significativa numa série de aplicações, abrangendo desde problemas industriais até alterações climáticas, avaliação de riscos financeiros, bem como modelos matemáticos e computacionais. Essas análises envolvem a identificação de parâmetros de entrada influentes nos modelos para compreender seu impacto na saída. A análise de sensibilidade pode ser realizada localmente, examinando os efeitos dos parâmetros em um valor fixo, ou globalmente, avaliando o modelo em uma faixa de valores de parâmetros. O método de Sobol apresenta uma abordagem robusta para análise de sensibilidade global, empregando uma sequência de Sobol para criar amostras de forma mais uniforme dentro do espaço de parâmetros de entrada, permitindo assim a exploração eficiente das entradas do modelo. Este artigo tem como objetivo apresentar uma implementação computacional no Scilab para gerar a sequência e os índices de Sobol para utilização em análise de sensibilidade através do método Sobol. Um caso de teste foi aplicado para gerar sequências de Sobol e discutir os resultados obtidos.

Palavras-chave

análise de sensibilidade • Método de Sobol • operação não usual • correlação de parâmetros

☆This article is an extended version of the work presented at the Joint XXVI ENMC National Meeting on Computational Modeling, XIV ECTM Meeting on Science and Technology of Materials held in Nova Friburgo – Brazil, from October 25th to 27th, 2023

1 Introduction

In various applied problems, such as industrial scenarios, climate change, financial risk analysis, physical mathematical models, and computational models, input data and parameters shaping the model are considered. To question the significance of these pieces of information and their effects in such applications, a direct analysis is conducted to facilitate fundamental decision-making across diverse contexts. In this context, the utilization of sensitivity and uncertainty analysis emerges to pinpoint the most influential inputs in a model, allowing for an in-depth understanding of their effects on the model's output [1].

Engaging in sensitivity analysis is essential to comprehend the behavior of the studied model concerning its input parameters. For inverse problems, for instance, a parameter lacking influence on the model's output should be excluded from the optimization process [2], as any value assigned to this parameter will yield the same output. The sensitivity study of the model concerning input parameters can be approached from two perspectives: local and global sensitivity analysis [3]. According to Ref. [4], one approach complements the other. The local approach assesses the model's sensitivity for a fixed parameter value across the entire model domain. The global approach evaluates sensitivity at a specific point in the model domain for various parameter values. There are specific methods to address both approaches. In this article, we delve into the Sobol method for global sensitivity analysis.

A central point for applying Sobol sensitivity analysis is the use of the Sobol sequence, a pseudo-random sampling approach introduced by Sobol [5]. Unlike random sampling sequences, such as those generated through the Monte Carlo method, which is used to approximate solutions for problems analytically challenging or impossible to solve, the Sobol sequence guarantees a more uniform coverage of the sample space, rendering it more efficient and precise [6]. Through the Sobol sequence, the capability to adequately explore the model input is enhanced, generating valuable insights for decision-making, including process optimization, identification of influential variables in models, thereby enabling process enhancement to boost efficiency and production quality, social sciences modeling, extensively applied across scientific domains as a potent tool for sensitivity analysis in mathematical models [7], physicochemical models [8], biological models [9], among others.

This article presents the implementation of a portion of the sensitivity analysis module using the Sobol method for the Inverse Problem and Sensitivity Analysis Library (IPSAL). It encompasses the development of the Sobol sequence and a test case.

2 Materials and Methods

In this section, an algorithm for generating the Sobol sequence is presented. The process involves converting decimal numbers into binary numbers, including the fractional parts, and performing the unconventional bitwise operation *bit by bit* [10], denoted as $\boldsymbol{u} \oplus \boldsymbol{v}$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors containing each digit from the binary decomposition of the decimal number. The unconventional operation is defined as $1 \oplus 1 = 0$, $1 \oplus 0 = 1$, $0 \oplus 1 = 1$, and $0 \oplus 0 = 0$. The Sobol method for sensitivity analysis involves generating a sample space composed of Sobol sequences to evaluate the model's sensitivity concerning its inputs. This sample space is derived from the unit hyper-edges of the hypercube [11]. In this work, model parameters are considered as inputs, and d represents the number of input parameters. d signifies the dimension of the hypercube, i.e., $[0,1)^d$, wherein the goal is to create a set of values, the Sobol sequence, contained within each of the hyper-edges.

2.1 Sobol Sequence

The Sobol sequence is denoted as S_n, where $n = 1, \dots, r$, r is the total number of sequences, indicating the sequence position, and each of its components as S_{n_j}, $j = 1, \dots, d$. The set $S_n = (S_{n_1}, \dots, S_{n_d})$. Each component of the sequence satisfies $0 \leq S_{n_j} < 1$, $j = 1, \dots, d$. To calculate S_{n_j}, the following equation is used:

$$S_{n_j} = b_{n_1} v_{j_1} \oplus b_{n_2} v_{j_2} \oplus b_{n_3} v_{j_3} \dots b_{n_k} v_{j_k} \tag{1}$$

where $j = 1, \dots, d$, and k, the last term in the unconventional summation in equation (1), depends on the binary decomposition of the integer n, i.e., $(n)_{10} = (b n_1 b_{n_2} \dots b_{n_k})_2$.

In equation (1), the term v_{j_i} is introduced, representing binary fractions with a denominator of 2^i, i.e.,

$$v_{j_i} = \frac{m_{j_i}}{2^i} \tag{2}$$

The terms v_{j_i}, $j = 1, \dots, d$, are referred to as directional numbers, with the variation in i depending on the total number of bits in the decomposition of n in base 2. The numbers m_{j_i}, $j = 1, \dots, d$, are defined by

$$m_{j_i} = 2a_{j_1} m_{j_{i-1}} \oplus 2^2 a_{j_2} m_{j_{i-2}} \oplus \cdots \oplus 2^{q-1} a_{j_{q-1}} m_{j_{i-(q-1)}} \oplus 2^q m_{j_{i-q}} \oplus m_{j_{i-q}} \tag{3}$$

Where $q = d + 1$ as presented in Eq. (3), referring to the degree of the polynomial given in Eq. (4). The coefficients a_j, $j = 1, \ldots, d$, are obtained from an irreducible polynomial in the field $\mathbb{Z}_2$ over the set $\{0, 1\}$ with addition modulo 2 and the standard multiplication, given by

$$p(x) = x^{d+1} + a_1 x^d + a_2 x^{d-1} + \cdots + a_d x + 1 \tag{4}$$

Thus, the coefficients a_j take on values of either 1 or 0. It is necessary to initialize the first three terms m_{j_1}, m_{j_2}, and m_{j_3} as odd integers smaller than 2^i.

In Figure 1, the workflow diagram of the Sobol sequence generation algorithm is presented. The unconventional bitwise operation is performed twice but under different circumstances. Initially, when calculating m_j in Eq. (3), the conversion between binary and integer is performed. Subsequently, when calculating S_{n_j} in Eq. (1), the operation is carried out between the decimal and binary parts.

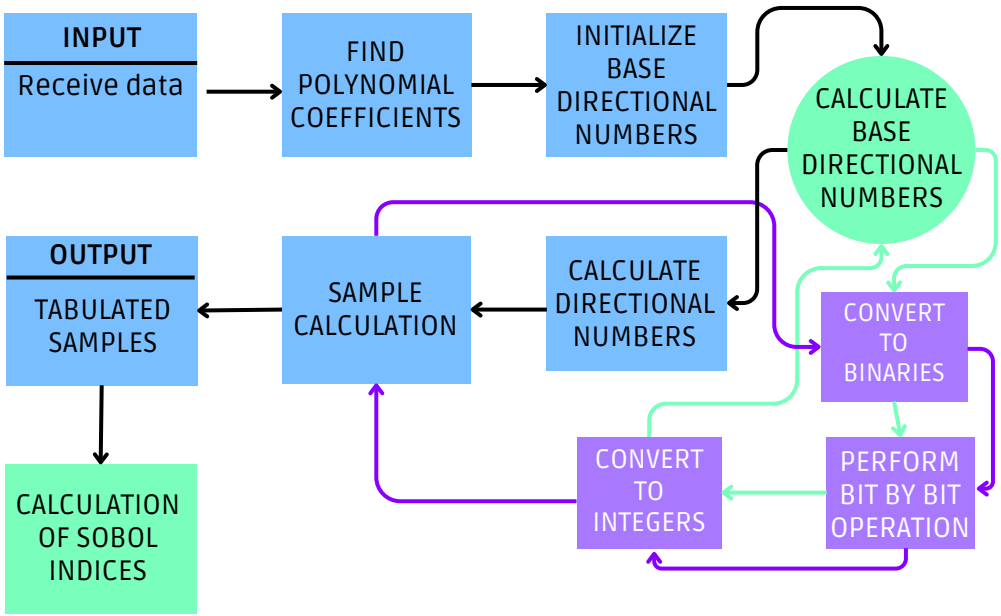


Figure 1: Flowchart for Sobol Sequence Generation

2.2 Sensitivity Analysis

Sobol sensitivity analysis is a technique used to evaluate the relative importance of input variables in relation to the output of a model. The order of Sobol sensitivity indices depends on the degree of interaction between the variables.

The most important factors for carrying out the sensitivity analysis are the table of values taken from the Sobol sequence, the expected value of the function and its average.

The Sobol sensitivity index is given by

$$S_i = \frac{V(f|x_i)}{V(f)} \tag{5}$$

where S_i is the sensitivity of the function with respect to the variable x_i, $V(f|x_i)$ is the variance conditioned on the variable x_i, and $V(f)$ variance of the function f. This coefficient is called the first-order coefficient, and tells us how sensitive the function is to each of the input variables.

2.3 Computer code

The IPSAL library should be downloaded and saved in the user's working folder. Afterward, the library functions for the Sobol method should be declared.

```
address  = get_absolute_file_path('sobol.sce')
exec(address+'decimal2binario.sci');
exec(address+'binario2decimal.sci');
exec(address+'construirM.sci');
exec(address+'polinomio.sci');
exec(address+'direcao.sci');
exec(address+'modelo.sci');
```

To generate the Sobol sequence, after calculating the directional numbers (direcao.sci), the following code is used. r input's with d parameters each will be generated. Each entry in this sequence will be used in the model to calculate the Sobol indices.

```
for i=1:r
    str = dec2bin(i)
    b   = zeros(1,length(str))
    for j=1:length(str)
        b(j)   = strtod(part(str,j:j))
    end
    for k=1:d
        w     = decimal2binario(0.0)
        for j=1:length(b)
            u = decimal2binario(b(j)*V(k,j))
            w = bitxor(u,w)
        end
        S(i,k) = binario2decimal(w)
    end
end
```

In Figure 2, the organizational structure of the code developed in Scilab is illustrated. The *results* folder will be responsible for storing all the outcomes obtained with the method.

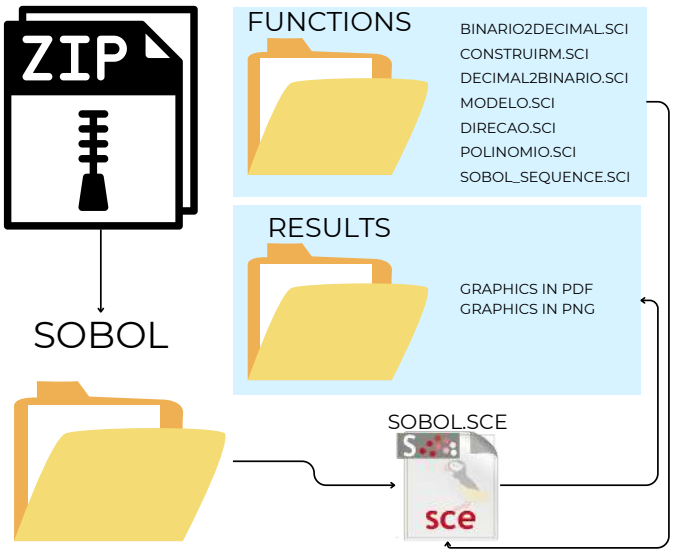


Figure 2: Organizational structure of the code.

The Sobol method module in Scilab can be accessed by the Computational Modeling, Inverse Problems and the Environment repository.

2.4 Test Case

The function given by

$$f(\mathbf{x}) = x_1 + x_2^2 + x_2 \cdot \sin(x_3) + x_4 \tag{6}$$

where $\mathbf{x} = (x_1, x_2, x_3, x_4)$ represents the input data, will be considered.

A Sobol sequence, S_n, $n = 1, \ldots, r$, will be constructed in the hypercube of dimension $d = 4$, $[0, 1)^4$, to evaluate the sensitivity of $f(\mathbf{x})$ (Eq. (6)) with respect to the inputs $\mathbf{x} = (x_1, x_2, x_3, x_4)$. To achieve this, four irreducible polynomials of degree 5 will be constructed within $\mathbb{Z}_2$. The polynomial is given by

$$p_j(\mathbf{x}) = x^5 + a_{j_1}x^4 + a_{j_2}x^3 + a_{j_3}x^2 + a_{j_4}x + c \tag{7}$$

where $j = 1, \ldots, 4$, $i = 1, \ldots, 3$, and each a_{j_i} is randomly obtained, taking on values of either 0 or 1, ensuring that at least one a_{j_i} is equal to 1.

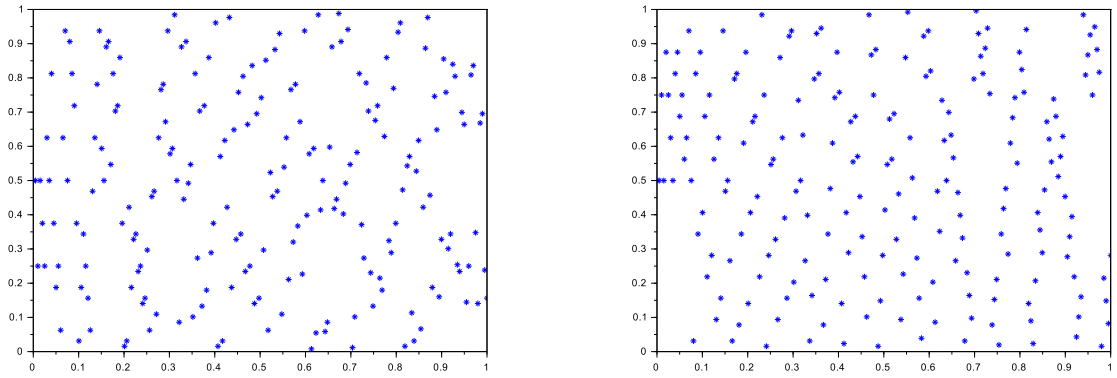
3 Results and Discussion

The test case described in Eq. (7) was used to generate the Sobol sequences. The construction will be to evaluate the four inputs, x_1, x_2, x_3 and x_4, with $d = 4$. After running the computational code elaborated in Scilab to generate 500 samples, in Table 1 we observe the first ten Sobol sequences obtained. Up to the sequence 4 there is a repetition of values [1].

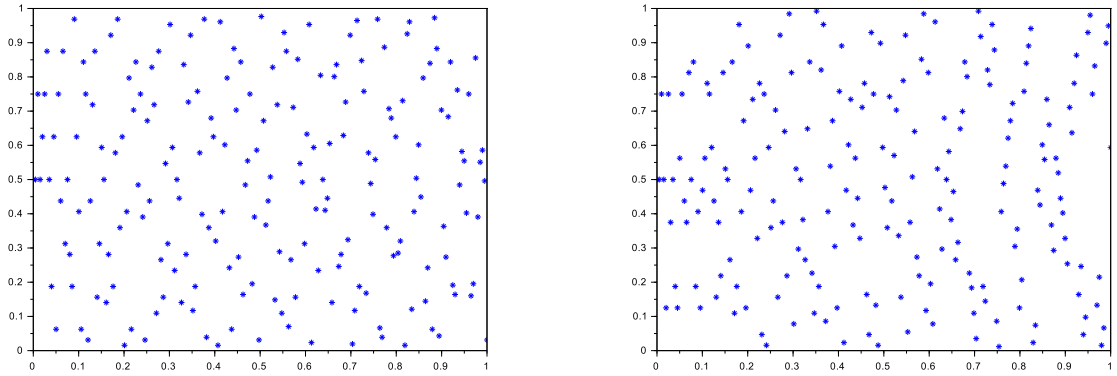
Table 1: Distribution of the first ten Sobol sequences S_n.

S_n	$n = 200$				$n = 500$				$n = 1000$			
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
S_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
S_2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
S_3	0.25	0.75	0.75	0.25	0.75	0.75	0.75	0.75	0.25	0.75	0.75	0.75
S_4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
S_5	0.375	0.875	0.625	0.625	0.625	0.125	0.875	0.875	0.875	0.875	0.625	0.375
S_6	0.25	0.75	0.75	0.25	0.75	0.75	0.75	0.75	0.25	0.75	0.75	0.75
S_7	0.625	0.625	0.875	0.375	0.875	0.375	0.625	0.625	0.125	0.625	0.875	0.125
S_8	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
S_9	0.8125	0.8125	0.1875	0.9375	0.4375	0.1875	0.9375	0.9375	0.4375	0.5625	0.6875	0.9375
S_{10}	0.375	0.875	0.625	0.625	0.625	0.125	0.875	0.875	0.875	0.875	0.625	0.375

To observe the distribution of Sobol sequence values for each input, the computational code was run for three different quantities. First, $n = 200$ was used, as can be seen in Fig. 3. For all inputs an almost uniform distribution was obtained.



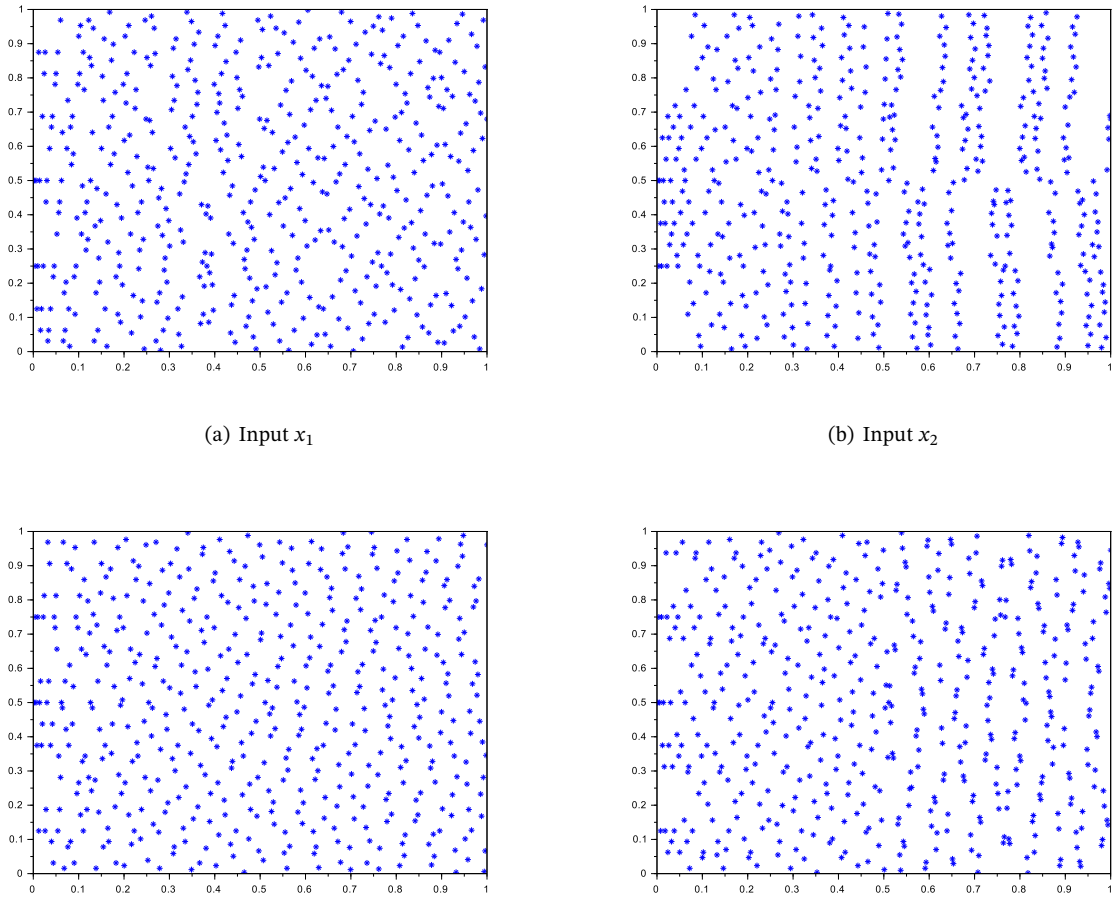
(a) Input x_1 (b) Input x_2



(c) Input x_3 (d) Input x_4

Figure 3: Distribution of Sobol sequences for each model input for $n = 200$.

Then, $n = 500$ was used, as presented in Fig. 4. The distribution of Sobol sequences remains almost uniform.

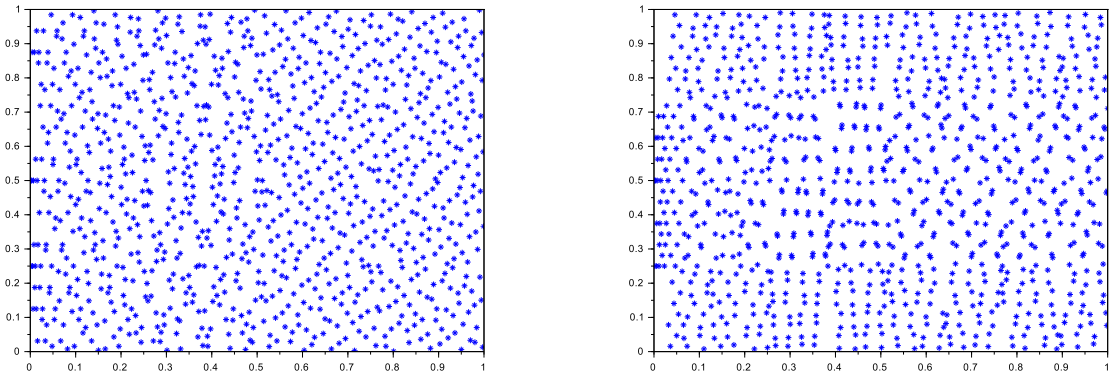


(a) Input x_1 (b) Input x_2

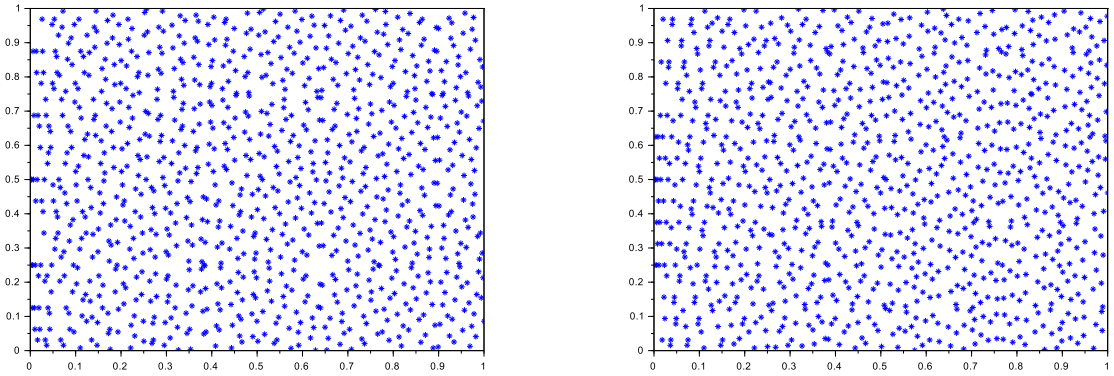
(c) Input x_3 (d) Input x_4

Figure 4: Distribution of Sobol sequences for each model input for $n = 500$.

Finally, $n = 1,000$ was used, according to Fig. 5. Again, the distribution of Sobol sequences remains almost uniform.



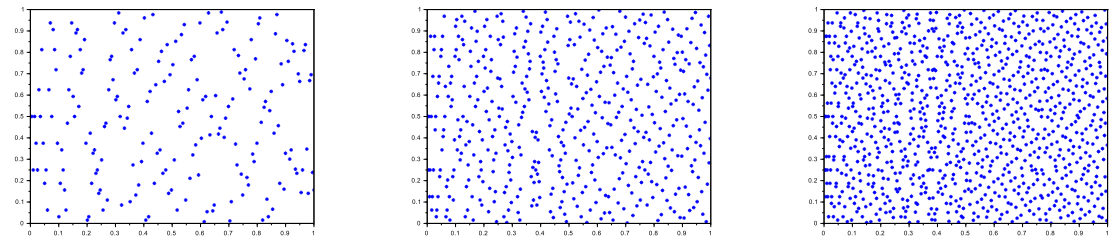
(a) Input x_1 (b) Input x_2



(c) Input x_3 (d) Input x_4

Figure 5: Distribution of Sobol sequences for each model input for $n = 1000$.

Figure 6 shows the distribution of the samples obtained in the generation of the Sobol sequence for the input x_1, for a total of 200, 500 and 1000 sequences. It is possible to observe the evolution of the distribution, which remains almost uniform.



(a) $n = 200$ (b) $n = 500$ (c) $n = 1.000$

Figure 6: Distribution of Sobol sequences for input x_1 adc $n = 200$, $n = 500$ e $n = 1.000$.

The first order Sobol indices present the sensitivity of the model with respect to each x_i, see Table 2. The function is sensitive to all inputs. The inputs x_1 and x_4 are less influential in the function due to their linear nature, as can be seen in Eq. (7).

Table 2: Results of the Sobol method with first order indices.

Sobol Index	Results			Evaluation
	$n = 200$	$n = 500$	$n = 1000$	
S_1	0.0115	0.0229	0.0018	Less influential
S_2	0.6294	0.2616	0.4177	Influential
S_3	0.3502	0.6878	0.5595	Influential
S_4	0.0089	0.0277	0.0209	Less influential

4 Conclusion

The IPSAL library allowed the application of sensitivity analysis techniques and inverse problems with just a few commands. It is currently being implemented in Scilab, but will also be translated into other programming languages. The implementation of Scilab currently allows sensitivity analysis using the Morris method and, as the study objective of this article, sensitivity analysis using the Sobol method. After generating the Sobol sequence, the first order indices are used to perform Sobol sensitivity analysis. The test case, Eq. (6), was used to observe the distribution of Sobol sequences for the model inputs, highlighting their almost uniform behavior. The Sobol indices show that the function is sensitive to all inputs, with inputs with a linear effect having lower values in the index. As future work, higher order indices will be implemented, which will allow analyzing the correlation between the function's inputs.

References

[1] T. Barbe, "A new algorithm for applications of Sobol sequences to the pricing of financial derivatives (in Portuguese)," Ph.D. dissertation, University of São Paulo, Faculty of Economics, Administration and Accounting, SP, 2005. Available at: https://doi.org/10.11606/T.12.2005.tde-14122022-095219

[2] F. F. Ferreira, A. J. da Silva Neto, G. B. Lyra, and S. P. de Queiroz, "van Genuchten sensitivity analysis by Morris method," in *Anais do XXIII Encontro Nacional de Modelagem Computacional*. Palmas, Brazil: ISSN: 2527-2357, 2020. Available at: http://enmc.ccam.uesc.br/

[3] F. Ferretti, A. Saltelli, and S. Tarantola, "Trends in sensitivity analysis practice in the last decade," *Science of The Total Environment*, vol. 568, pp. 666–670, 2016. Available at: https://doi.org/10.1016/j.scitotenv.2016.02.133

[4] D. M. Hamby, "A review of techniques for parameter sensitivity analysis of environmental models," *Environmental Monitoring and Assessment*, vol. 32, pp. 135–154, 1994. Available at: https://doi.org/10.1007/BF00547132

[5] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, no. 55, pp. 271–280, 2001. Available at: https://doi.org/10.1016/S0378-4754(00)00270-6

[6] D. P. Kroese and R. Y. Rubinsteino, "Monte Carlo methods," *Wires Computational Statistics*, vol. 4, no. 1, pp. 48–58, 2012. Available at: https://doi.org/10.1002/wics.194

[7] S. Joe and F. Y. Kuo, "Constructing Sobol sequences with better two-dimensional projections," *SIAM Journal on Scientific Computing*, vol. 30, no. 5, pp. 2635–2654, 2008. Available at: https://doi.org/10.1137/070709359

[8] D. Kumar, A. Singh, P. Kumar, R. K. Jha, S. K. Sahoo, and V. Jha, "Sobol sensitivity analysis for risk assessment of uranium in groundwater," *Environmental Geochemistry and Health*, vol. 6, no. 42, pp. 1789–1801, 2020. Available at: https://doi.org/10.1007/s10653-020-00522-5

[9] J. Nossent, P. Elsen, and W. Bauwens, "Sobol sensitivity analysis of a complex environmental model," *Environmental Modelling and Software*, vol. 26, no. 12, pp. 1515–1525, 2011. Available at: https://doi.org/10.1016/j.envsoft.2011.08.010

[10] K. F. Rocha, "Unusual numerical bases: A brief study (in portuguese)," Master's thesis, Federal University of Grande Dourados, Professional master's degree in mathematics, MS, Mato Grosso do Sul, Brazil, 2019. Available at: http://repositorio.ufgd.edu.br/jspui/handle/prefix/1375

[11] A. Navid, S. Khalilarya, and M. Abbasi, “Diesel engine optimization with multi-objective performance characteristics by non-evolutionary Nelder-Mead algorithm: Sobol sequence and Latin hypercube sampling methods comparison in DoE process,” *Fuel*, vol. 228, pp. 349–367, 2018. Available at: https://doi.org/10.1016/j.fuel.2018.04.142