Sensitivity analysis with stochastic models

B. Sanderse[a]

[a] *Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands*

Abstract

Abstract

Keywords: Stochastic model, sensitivity analysis

1. Introduction

Consider the following model (inspired by [1]):

$$Y(\boldsymbol{x}, \omega) = \sin(x_1) + 7\sin(x_2)^2 + \exp(Z_1 + (x_1/\pi) Z_2(\omega)), \tag{1}$$

where $\boldsymbol{x}$ denotes the input parameters $X_1, X_2 \sim \mathcal{U}(0, 2\pi)$, and Z denotes the latent variables that introduce 'internal' stochasticity into the model, $Z_1 \sim \mathcal{U}(0, 1)$, $Z_2 \sim \mathcal{N}(0, 1)$. The effect of Z on Y can be present in different ways, depending on how much is known about Z:

1. Treat Z as an additional parameter making the model deterministic:
$$Y(\hat{\boldsymbol{x}}) = \sin(x_1) + 7\sin(x_2)^2 + \exp(x_3 + (x_1/\pi) x_4), \tag{2}$$
with $X_3 \sim \mathcal{U}(0, 1)$ and $X_4 \sim \mathcal{N}(0, 1)$. This requires explicit knowledge of the internal stochasticity, i.e. the expression for Z. The input space is now four-dimensional. In this case one can use existing UQ techniques, such as polynomial chaos expansions.
2. Treat Z via the random seed. This is a slightly less intrusive approach only requiring access to the random seed. We take $X_3 \sim \mathcal{U}(2^{14}, 2^{16})$ (rounded to give integer values), which sets the value for the random seed: `seed`(X_3), and then $Z_1 \sim \mathcal{U}(0, 1)$, $Z_2 \sim \mathcal{N}(0, 1)$. In this case the input space is three-dimensional instead of four-dimensional.
3. Don't treat Z: in case no access is given to Z (a black-box code in which only X can be controlled), and the actual form (1) is not known, one can only sample a two-dimensional parameter space (X_1, X_2), and one fully depends on the stochasticity introduced by the black-box code.

Depending on the type of access to the latent variables Z, different approaches are possible to perform a global sensitivity analysis (Sobol' analysis):

- Classic analysis: extend $\boldsymbol{X}$ with the latent variables Z, turning the stochastic simulator in a deterministic one. This requires type 1 or type 2 access.

- QoI-based analysis: define a statistical quantity (called the quantity of interest in [1]) which eliminates the randomness due to Z. For example, take the mean $m(\boldsymbol{x})$ (expectation over Z), and then the sensitivity analysis is performed for the mean instead of for the output Y. In other words, one first integrates over Z to get a deterministic quantity, and then integrates over X for the Sobol' analysis.

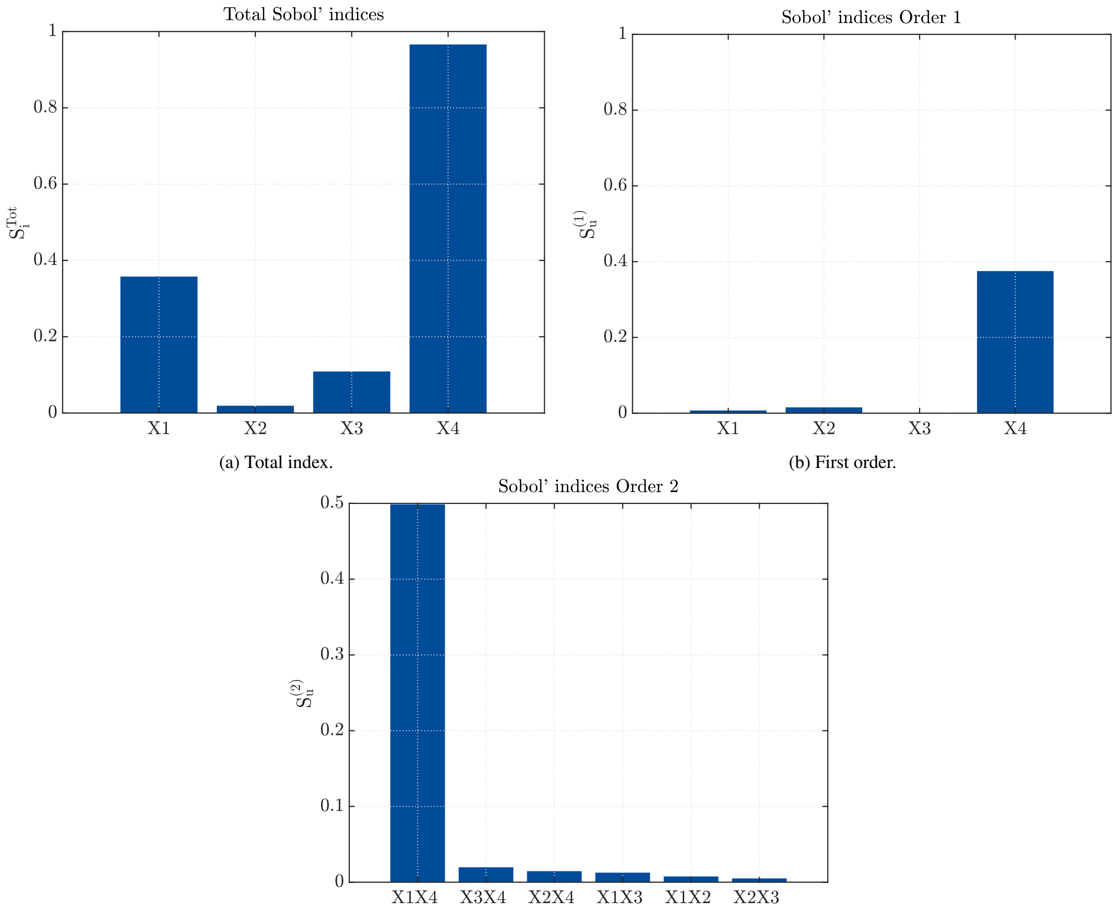
Email address: `B.Sanderse@cwi.nl, corresponding author` (B. Sanderse)

- Trajectory-based analysis. In this case the steps of the QoI-based analysis are reversed. In this case one first fixes a value for Z, does the Sobol' analysis with respect to X, and then repeats for different values of Z and takes the expectation.

2. Example results

2.1. Classical analysis

Classical analysis with type 1 access is shown in figure 1, based on Monte Carlo simulation. We set $M = 10^4$ samples, but in reality $N = 12M$ samples are needed to compute the first and second order indices. The total and first order indices require $M(p+2) = 6M$ ($p = 4$) evaluations (Saltelli method), and the second order indices require $M \cdot \binom{p}{2} = 6M$ evaluations.



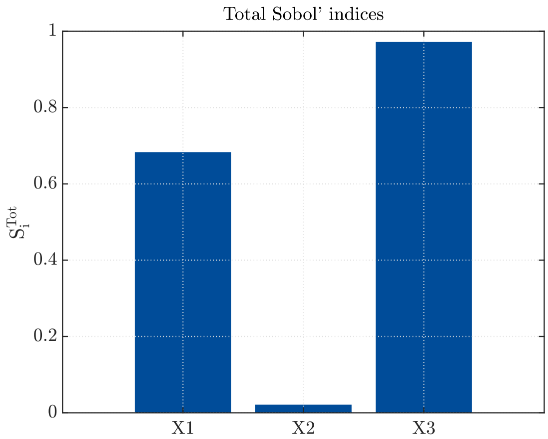
(a) Total index.

(b) First order.

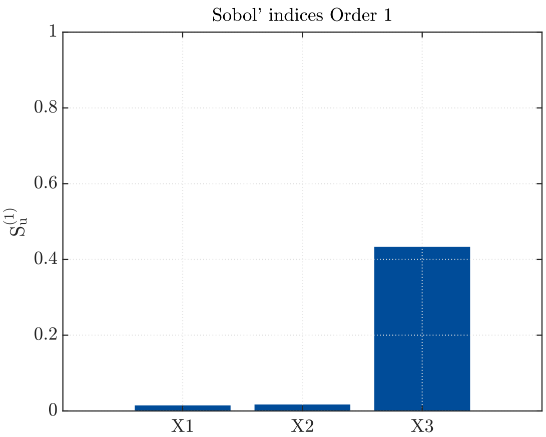
(c) Second order.

Figure 1: Sobol' indices for type 1.

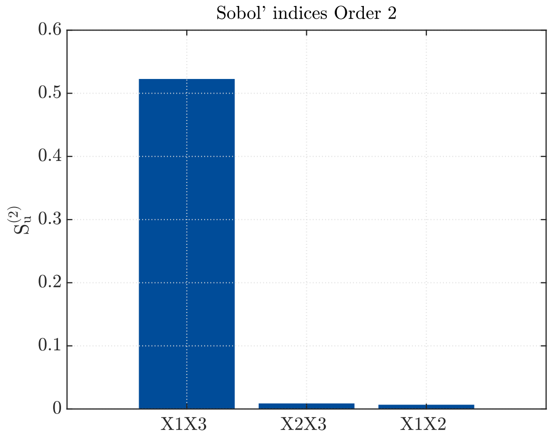
Classical analysis with type 2 access is shown in figure 2. In this case there are 3 input uncertainties, so $N = 8M$ samples are necessary for total, first, and second order indices. Although X_3 in this case is now responsible for both the effect of Z_1 and Z_2 as used in the previous (type 1) case, it is generally *not* true that the Sobol index for X_3 are the sum of those of X_3 and X_4 in the previous case.



(a) Total index.



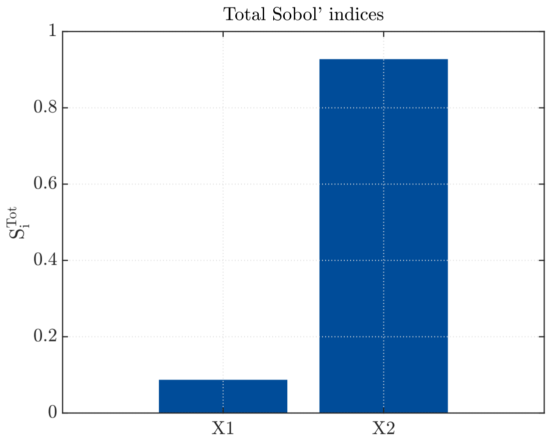
(b) First order.



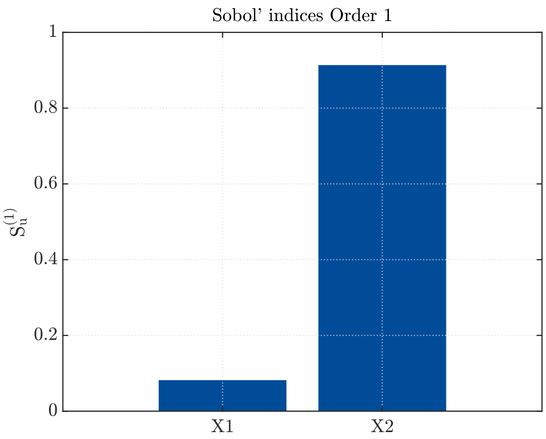
(c) Second order.

Figure 2: Sobol' indices for type 2.

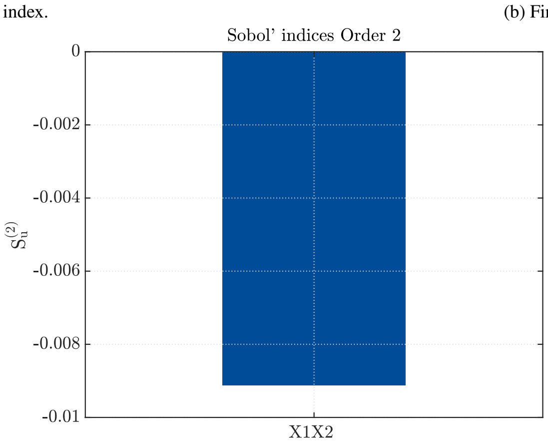
Classical analysis with type 3 access is shown in figure 2. In this case there are 2 input uncertainties, so $N = 5M$ samples are necessary for total, first, and second order indices. Combining classical analysis with type 3 access is questionable, as the effect of Z is completely ignored in computing the Sobol' indices.



(a) Total index.



(b) First order.

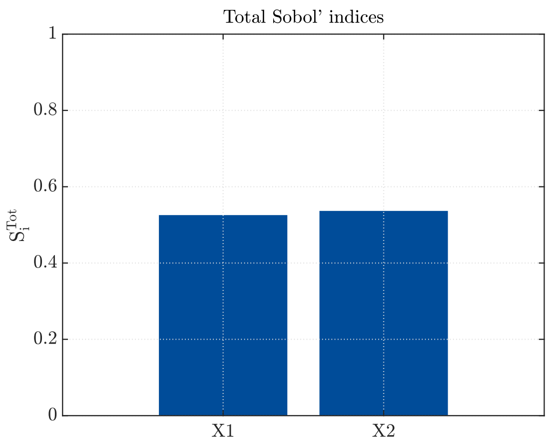


(c) Second order.

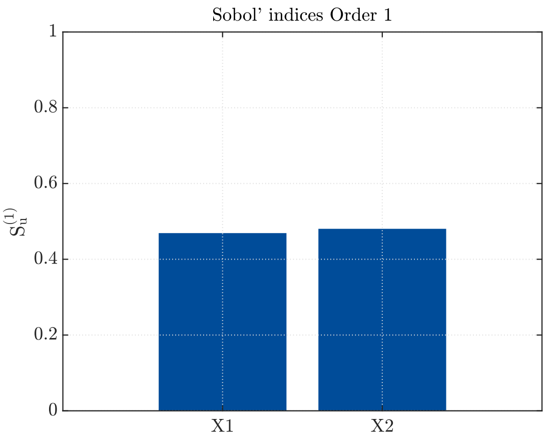
Figure 3: Sobol' indices for type 3.

2.2. *QoI-based analysis*

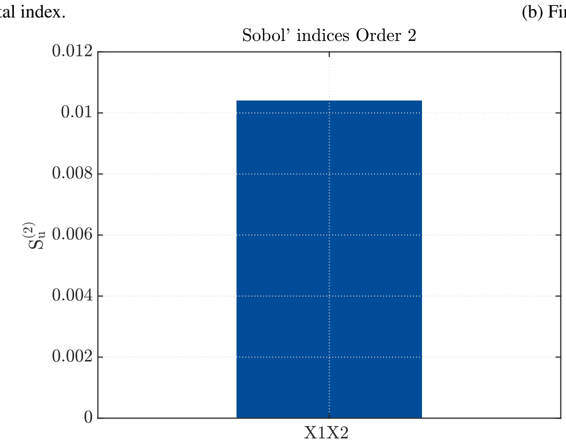
QoI-based analysis is essentially the same for type 1, 2 or 3, as the QoI-based analysis integrates over Z first, and this does not require knowledge about Z. In this case we take as QoI the mean of (1), where the mean is computed via a Monte Carlo sample of $\hat{M} = 10^3$ samples of Z. There are only two uncertainties remaining, namely X_1 and X_2. The Sobol' indices are then computed using $M = 10^4$ samples, requiring $5M$ samples. However, since for each model evaluation $\hat{M}$ samples are necessary, the total number of evaluations is $5M\hat{M} = 5 \cdot 10^7$. The resulting indices are shown in 4. Important to note that these figures can change drastically when a different QoI is taken (e.g. the standard deviation).



(a) Total index.



(b) First order.



(c) Second order.

Figure 4: Sobol' indices for QoI-based analysis, where the mean is taken as QoI.

3. References

[1] X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. (May):1–29, 2020.