Interpretation of black box models with (or without) derivatives

ENBIS Webinar
2024 March 13

O. Roustant, INSA Toulouse

Synthesis of collaborative works from 2014 to now, with:

F. Barthe, F. Gamboa, A. Joulin (Inst. Math. Toulouse), B. Iooss (EDF),
N. Luethen, S. Marelli, B. Sudret (ETH Zürich)
J. Fruth, S. Kuhnt (Univ. Dortmund)

Part I

Interpretating black-box functions using Global Sensitivity Analysis (GSA)

Fundamental theorem: Sobol-Hoeffding decomposition

Framework. $X = (X_1, \ldots, X_d)$ is a vector of independent input variables with distribution $\mu_1 \otimes \cdots \otimes \mu_d$, and $f : \Delta \subseteq \mathbb{R}^d \to \mathbb{R}$ is such that $f(X) \in L^2(\mu)$.

Theorem [Hoeffding, 1948, Efron and Stein, 1981, Sobol, 1993]

There exists a unique expansion of *ff* of the form

$$f(X) = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{1 \leq i < j \leq d} f_{i,j}(X_i, X_j) + \cdots + f_{1,\ldots,d}(X_1, \ldots, X_d)$$

such that $E[f_I(X_I)|X_J] = 0$ for all $I \subseteq \{1, \ldots, d\}$ and all $J \subsetneq I$.

This *decomposition is orthogonal* $\to$ *variance decomposition*

Exploration tool: Sensitivity indices

Tool 1. Sobol indices

- Partial variances: $D_I = \mathbb{V}\text{ar}(f_I(X_I))$, and *Sobol indices* $S_I = D_I / D$

$$D = \sum_I D_I, \qquad 1 = \sum_I S_I$$

- $D_i^{\text{tot}} = \sum_{J \supseteq \{i\}} D_J$, $\quad S_i^{\text{tot}} = \frac{D_i^{\text{tot}}}{D}$ $\quad$ *Total index*
- $D_I^{\text{tot}} = \sum_{J \supseteq \{I\}} D_J$, $\quad S_I^{\text{tot}} = \frac{D_I^{\text{tot}}}{D}$ $\quad$ *Total interaction, superset importance*

Tool 2. Derivative Global Sensitivity Measure (DGSM)

$$\nu_i = \int \left(\frac{\partial f(x)}{\partial x_i} \right)^2 d\mu(x), \qquad \nu_I = \int \left(\frac{\partial^{|I|} f(x)}{\partial x_I} \right)^2 d\mu(x)$$

Usage for screening

Assume that:

- f is continuous on $\Delta = [0,1]^d$
- for all i, the support of μ_i contains $[0,1]$

- **Variable screening**
  If either $D_i^{tot} = 0$ or $\nu_i = 0$, then X_i is non influential

- **Interaction screening**
  If either $D_{i,j}^{tot} = 0$ or $\nu_{i,j} = 0$, then $(x_i, x_j) \mapsto f(x)$ is additive

  Total interactions can be visualized on the *FANOVA graph*, where the edge size is proportionnal to the index value [Fruth et al., 2014].

Illustration on a toy example

8D g-Sobol function, with uniform inputs on [0, 1]:

$$f(x) = \prod_{j=1}^{8} \frac{|4x_j - 2| + a_j}{1 + a_j}$$

with $a = c(0, 1, 4.5, 9, 99, 99, 99, 99)$.

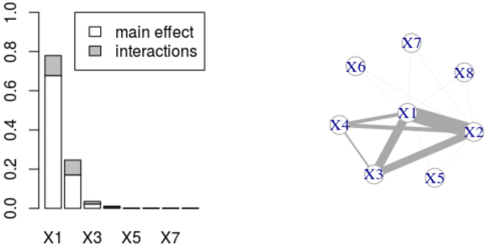


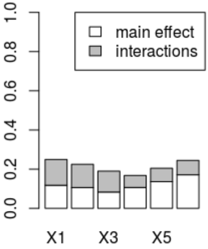
Figure: 1st order analysis (left) and 2nd order analysis (right) with 10^5 simulated data

Illustration on a toy example

A 6D block-additive function, with uniform inputs on $[-1, 1]$:

$$f(x) = \cos([1, x_1, x_2, x_3]^\top \beta) + \sin([1, x_4, x_5, x_6]^\top \gamma))$$

with $\beta = (-0.8, -1.1, 1.1, 1)^\top$ and $\gamma = (-0.5, 0.9, 1, -1.1)^\top$.



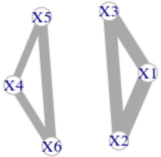


Figure: 1st order analysis (left) and 2nd order analysis (right) with 10^5 simulated data

Part II

Interpretating black-box functions using GSA and derivatives

Spectral expansions

Let $X_1, \ldots, X_d$ be independent input variables with probability distribution $\mu_1, \ldots, \mu_d$ and $f \in L^2(\mu)$ with $\mu = \mu_1 \otimes \cdots \otimes \mu_d$.

Polynomial chaos expansions consider the orthonormal basis of $L^2(\mu)$

$$e_{\underline{\ell}}(x) = \prod_{j=1}^{d} e_{j,\ell_j}(x_j), \quad \underline{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d$$

where $e_{j,0} = 1, e_{j,1}, e_{j,2}, \ldots$ is the basis of orthonormal polynomials in $L^2(\mu_j)$.

Once the basis decomposition obtained $f = \sum_{\underline{\ell}} c_{\underline{\ell}} e_{\underline{\ell}}$, with $c_{\underline{\ell}} = \langle f, e_{\underline{\ell}} \rangle$, *all variance-based sensitivity indices (Sobol indices) are available at once*, e.g.

$$D_1^{\text{tot}}(f) = \sum_{\ell_1 \geq 1,\, \ell_2, \ldots, \ell_d} c_{\underline{\ell}}^2$$

$\rightarrow$ *this nice property is not specific to polynomials, see e.g. [Antoniadis, 1984, Tissot, 2012]*

Poincaré chaos expansions

Now assume that $f \in H^1(\mu)$ i.e. f and $f' \in L^2(\mu)$.

A 1D Poincaré basis *diagonalizes the derivation operator, in the weak sense:*

$$\langle (e_{j,\ell_j})', g' \rangle = \lambda_{j,\ell_j} \langle e_{j,\ell_j}, g \rangle \qquad \forall g \in H^1(\mu_j)$$

With Poincaré chaos, all Sobol indices are available at once, *using derivatives*

$$D_1^{\text{tot}}(f) = \sum_{\ell_1 \geq 1, \ell_2, \dots, \ell_d} \langle f, e_{\underline{\ell}} \rangle^2 = \sum_{\ell_1 \geq 1, \ell_2, \dots, \ell_d} \frac{1}{(\lambda_{1,\ell_1})^2} \left\langle \frac{\partial f}{\partial x_1}, \frac{\partial e_{\underline{\ell}}}{\partial x_1} \right\rangle^2$$

$\rightarrow$ *Sobol indices can be estimated more accurately if f is smooth*.

Straightforward extension to *total interaction indices*

$$D_{1,2}^{\text{tot}}(f) = \sum_{\ell_1 \geq 1, \ell_2 \geq 1, \ell_3, \dots, \ell_d} \langle f, e_{\underline{\ell}} \rangle^2 = \sum_{\ell_1 \geq 1, \ell_2 \geq 1, \ell_3, \dots, \ell_d} \frac{1}{(\lambda_{1,\ell_1} \lambda_{2,\ell_2})^2} \left\langle \frac{\partial^2 f}{\partial x_1 \partial x_2}, \frac{\partial^2 e_{\underline{\ell}}}{\partial x_1 \partial x_2} \right\rangle^2$$

What does a Poincaré basis look like?

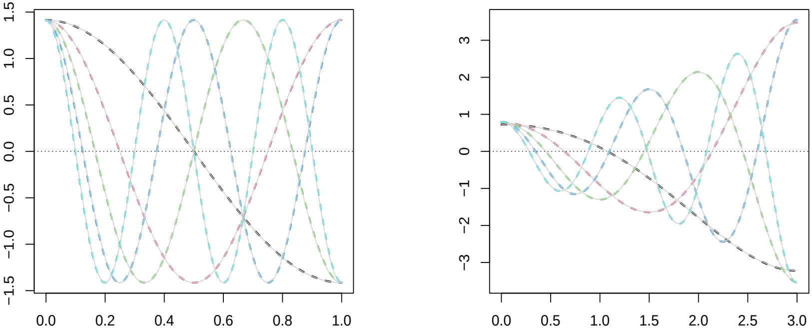


Figure: First Poincaré basis functions for $\mathcal{U}[0,1]$ and $\mathcal{E}(1)$ truncated on $[0,3]$. Solid line: analytic expression; Dotted curves: estimated by finite elements. Software used: `sensitivity` R package [Iooss et al., 2023].

Notice that the Poincaré basis function of 'degree' *n* has at most *n* zeros. Actually this property is stable by linear combination $\rightarrow$ *T-system*

Poincaré chaos expansion (PoinCE) vs polynomial chaos exp. (PCE)

On the flood model, using derivatives with PoinCE gives more accurate results especially for small indices, and outperforms PCE [Lüthen et al., 2023]

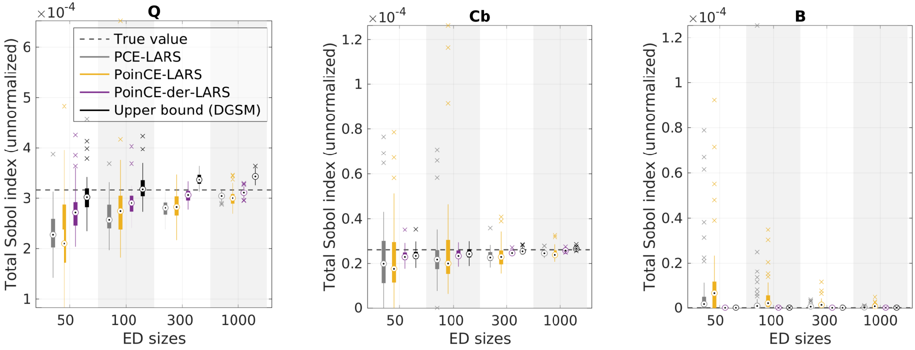


Figure: Estimates of unnormalized total Sobol' indices for the flood cost model, via sparse regression. Software used: UQLab [Marelli and Sudret, 2014].

Connections with Derivative-based Global Sensitivity Measures (DGSM)

For all f in $H^1(\mu)$, DGSM can be computed with the Poincaré basis coef. [Lüthen et al., 2023]:

$$\nu_1(f) := \int \left(\frac{\partial f}{\partial x_1}(x)\right)^2 d\mu(x) = \sum_{\ell_1 \geq 1, \ell_2, \dots, \ell_d} \lambda_{1,\ell_1} \langle f, e_{\underline{\ell}} \rangle^2$$

Furthermore, we have the optimal inequality [Lamboni et al., 2013]

$$D_1^{\text{tot}}(f) \leq C(\mu_1)\nu_1(f)$$

Extension to total interaction indices [Roustant et al., 2014]

$$D_{1,2}^{\text{tot}}(f) \leq C(\mu_1)C(\mu_2)\nu_{1,2}(f)$$

Low-cost screening of Sobol indices based on DGSM

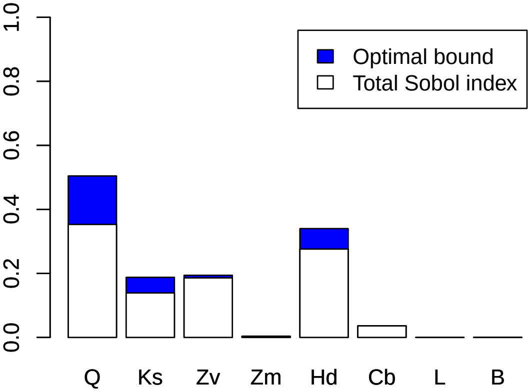


Figure: Total Sobol indices and their upper bounds $C(\mu_j)\nu_j/D$ for the flood model (subverse output), see [Roustant et al., 2017].

→ *Z_m, L, B (and maybe C_b) will be discarded, only knowing the upper bounds.*

Antoniadis, A. (1984).
Analysis of variance on function spaces.
Statistics: A Journal of Theoretical and Applied Statistics, 15(1):59–71.

Bakry, D., Gentil, I., and Ledoux, M. (2014).
Analysis and geometry of Markov diffusion operators, volume 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].
Springer, Cham.

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

Fruth, J., Roustant, O., and Kuhnt, S. (2014).
Total interaction index: A variance-based sensitivity index for second-order interaction screening.
Journal of Statistical Planning and Inference, 147:212 – 223.

Hoeffding, W. (1948).
A class of statistics with asymptotically normal distribution.
Ann. Math. Statist., 19(3):293–325.

Iooss, B., Veiga, S. D., Janon, A., and Pujol, G. (2023).
sensitivity: Global Sensitivity Analysis of Model Outputs.
R package version 1.28.1.

Lamboni, M., Iooss, B., Popelin, A.-L., and Gamboa, F. (2013).
Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests.
Mathematics and Computers in Simulation, 87:45–54.

Lüthen, N., Roustant, O., Gamboa, F., Iooss, B., Marelli, S., and Sudret, B. (2023).
Global sensitivity analysis using derivative-based sparse Poincaré chaos expansions.

International Journal for Uncertainty Quantification, 13(6).

Marelli, S. and Sudret, B. (2014).
UQLab: A framework for uncertainty quantification in Matlab.
In *Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom)*, pages 2554–2563.

Roustant, O., Barthe, F., and Iooss, B. (2017).
Poincaré inequalities on intervals – application to sensitivity analysis.
Electronic Journal of Statistics, 11(2):3081 – 3119.

Roustant, O., Fruth, J., Iooss, B., and Kuhnt, S. (2014).
Crossed-derivative based sensitivity measures for interaction screening.
Mathematics and Computers in Simulation, 105:105 – 118.

Sobol, I. (1993).
Sensitivity estimates for non linear mathematical models.
Mathematical Modelling and Computational Experiments, 1:407–414.

Tissot, J.-Y. (2012).
Sur la décomposition ANOVA et l'estimation des indices de Sobol'. Application à un modèle d'écosystème marin.
PhD thesis, Grenoble University.