New way of estimating Total Sensitivity Indices

Matieyendou Lamboni

Institute for Environment and Sustainability, JRC-Ispra, Itally

matieyendou.lamboni@jrc.ec.europa.eu / matieyendou.lamboni@gmail.com

Nice, July 2, 2013

Introduction

- Context : exploration of complex models
  - model often includes many input factors
  - only few inputs really act in the model [1]
- Global Sensitivity Analysis (GSA) [1,2] and Multivariate Sensitivity Analysis (MSA) [3]
  -apportion the variability of model output into inputs and their interactions:
  - objective way to evaluate the impact of inputs on the model output(s)
  - Total Sensitivty Indices (TSI) give the overall contribution of input factors
  - but: computation of TSI requires many model runs

Introduction

- **Objective:** investigate new methods of estimating TSI based on quadrature rules

  - the specific formulae for the commutation of the TSI

  - how to use these formulae for the commutation of TSI

- Plan

  - Recall the definitions of Total Sensitivity Indices

  - Give the main theorems for computing TSI

  - Show some results for counter and g-Sobol' functions and for Ishigami ' function

Methiods: Hoeffding decomposition

Let $Y = f(\mathbf{X})$ be a model output and $\mathbf{X} = (X_1, \ldots, X_d)$, d independent inputs (A1). If $\mathbb{E}\left(f^2(\mathbf{X})\right) < +\infty$ (A2), we have:

$$\begin{aligned} f(\mathbf{X}) &= f_0 + \sum_{j}^{d} f_j(X_j) + \sum_{i<j}^{d} f_{ij}(X_i, X_j) + \ldots + f_{1\ldots d}(X_1, \ldots, X_d) \\ &= \sum_{u \subset \{1,2,\ldots d\}} f_u(X_u), \end{aligned} \tag{1.1}$$

where $f_0 = \mathbb{E}\left[f(\mathbf{X})\right]$; $f_j(X_j) = \mathbb{E}\left[f(\mathbf{X})|X_j\right] - f_0$ and $f_u(X_u) = \mathbb{E}\left[f(\mathbf{X})|X_u\right] - \sum_{v \subset u} f_v(X_v)$ with $u \subset \{1, 2, \ldots, d\}$.

$$g(X_j, \mathbf{X}_{\sim j}) = \sum_{u \ni j} f_u(\mathbf{X}_u), \qquad \text{with } u \subseteq \{1, 2, \ldots d\}$$

Methods: Total Sensitivity Indices

Sobol and Kucherenko (2009) [4] and Lamboni $et\ al.$ (2013) [5]

$$\mathbf{f}(\mathbf{X}) = \mathbf{f_0} + \mathbf{g}(\mathbf{X_j}, \mathbf{X_{\sim j}}) + \mathbf{h}(\mathbf{X_{\sim j}}), \tag{1.2}$$

with $\mathbb{E}\left[g(X_j, \mathbf{X}_{\sim j}) | \mathbf{X}_{\sim j}\right] = 0$

Let $D = \int f^2(\mathbf{x}) d\mu(\mathbf{x}) - f_0^2$ and $D_u = \int f_u^2(\mathbf{x}_u) d\mu(\mathbf{x}_u)$

- Classical definition of TSI [1,2]

$$S_{T_j} = \frac{\sum_{u \supseteq j} D_u}{D} \ .$$

- Equivalent definition of TSI [5]

$$S_{T_j} = \frac{\int \left[g^2(X_j, \mathbf{X}_{\sim j} d\mu(\mathbf{x}))\right]}{D} \tag{1.3}$$

Methods: practical expression of $g(\cdot)$

Proposition

Let $Y = f(\mathbf{X})$ *be a model output. If A1 (independent input factors) and A2 (finite second moment) hold, the practical formulas to obtain* $g(X_j, \mathbf{X}_{\sim j})$ *is::*

$$g(X_j, \mathbf{X}_{\sim j}) = f(\mathbf{X}) - \mathbb{E}\left[f(\mathbf{X})|\mathbf{X}_{\sim j}\right] \quad (1.4)$$

Proof

Due to $f(\mathbf{X}) = f_0 + g(X_j, \mathbf{X}_{\sim j}) + h(\mathbf{X}_{\sim j})$ *and* $\mathbb{E}\left[g(X_j, \mathbf{X}_{\sim j})|\mathbf{X}_{\sim j}\right] = 0$, *we have* $\mathbb{E}\left[f(\mathbf{X})|\mathbf{X}_{\sim j}\right] = f_0 + h(\mathbf{X}_{\sim j})$

Methods: Numerical approximation of TSI

Theorem

Let $Y = f(\mathbf{X})$ *be a model output and* $\mathbf{x}_i,\ i = 1, 2 \ldots, N$ *be* N *sample points of* $\mathbf{X}$. *If* $f(\mathbf{X})$ *is a polynomial of order* $2p - 1$ *with respect to* X_j *(A3) and if A1, A2 hold we have:*

$(\mathbf{i})$ $g(x_j, \mathbf{x}_{\sim j})$ *is exactly obtained with* $p + 1$ *evaluations of* $f(\cdot)$*:*

$$g(x_j, \mathbf{x}_{\sim j}) = f(\mathbf{x}) - \sum_{i=1}^{p} c_i f(x^*_{i,j}, \mathbf{x}_{\sim j}), \tag{1.5}$$

$(\mathbf{ii})$ $\mathbf{N(dp + 1)}$ *is the total cost of model evaluations to estimate the* d *TSI indices*

Proof

$\mathbb{E}\left[f(\mathbf{X})|\mathbf{X}_{\sim j}\right] = \sum_{i=1}^{p} c_i f(x^*_{i,j}, \mathbf{x}_{\sim j})$, *for polynomials of order* $2p - 1$ *due to the quadrature rules (Abramowitz* $et\ al.$*, 1965).*

Methods: Numerical approximation of TSI

Theorem

Let $Y = f(\mathbf{X})$ *be a model output and* $\mathbf{x}_i,\ i = 1, 2 \ldots, N$ *be* N *sample points of* $\mathbf{X}$*. If* $f(\mathbf{X})$ *is* C^{2p} *with respect to* X_j *(A4) and if A1, A2 hold we have:*

$(\mathbf{i})$ $g(x_j, \mathbf{x}_{\sim j})$ *is approximated by:*

$$\widetilde{g(x_j, \mathbf{x}_{\sim j})} = f(\mathbf{x}) - \sum_{i=1}^{p} c_i f(x^*_{i,j}, \mathbf{x}_{\sim j}) \,, \tag{1.6}$$

$(\mathbf{ii})$ *and the error term* $Err(g)$ *is:*

$$Err(g) = \frac{f^{(2p)}(\epsilon_{x_j}, \mathbf{x}_{\sim j})}{(2p)!} P^2(x_j, , \mathbf{x}_{\sim j}) \tag{1.7}$$

Numerical Tests: functions

- Counter function ($d = 4$)

$$f(\mathbf{x}) = \sum_{j=1}^{d=4} (x_j - \frac{1}{2}) + 50(x_1 - \frac{1}{2})(x_2 - \frac{1}{2})^5$$

- g-Sobol ' function ($d = 10$) of type A, B, and C

$$f(\mathbf{x}) = \prod_{j=1}^{d=10} \frac{|4x_j - 2| + a_j}{1 + a_j},$$

Type A: $\mathbf{a} = (0, 0, 6.52, 6.52, 6.52, 6.52, 6.52, 6.52, 6.52, 6.52)$

Type B: $\mathbf{a} = (50, 50, 50, 50, 50, 50, 50, 50, 50, 50)$; $S_j = S_{T_j} = 0.1$

Type C: $\mathbf{a} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$, S$_j = 0.02$ and $S_{T_j} = 0.27$

- Ishigami' function ($d = 3$)

$$f(\mathbf{x}) = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4 \sin(x_1)$$

Numerical Tests: main choices

- Order of quadrature $p = 3$, Sobol' design with $N = 10, 20, \ldots$
- Gauss-Legendre quadrature rules for Counter and g-Sobol 'functions
- Gauss-Chebyshev (first kind) for Ishigami'function
- Effectiveness of methods: average of Mean Absolute Error

$$MAE = \frac{1}{R} \sum_{r=1}^{R=50} \sum_{j=1}^{d} |\hat{S_{T_j}} - S_{T_j}|$$

- Two ways of using evaluations of $g(\cdot)$

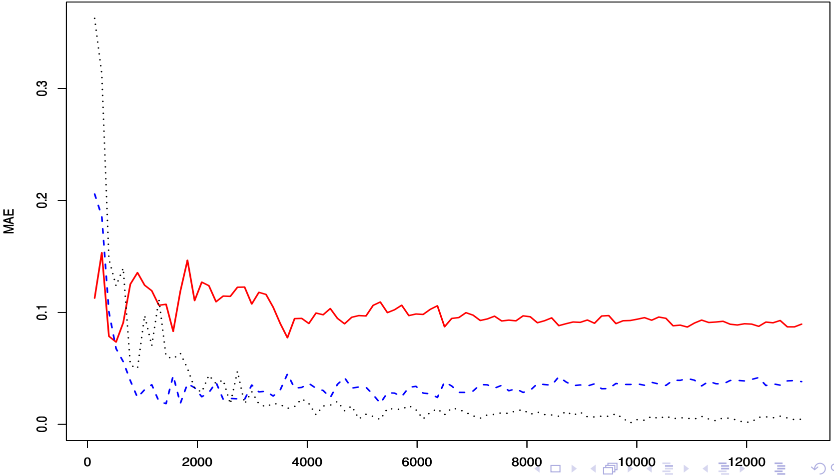
first way: use only N evaluations of

$g(x_j, \mathbf{x}_{\sim j}) = f(\mathbf{x}) - \sum_{i=1}^{p} c_i f(x_{i,j}^*, \mathbf{x}_{\sim j})$

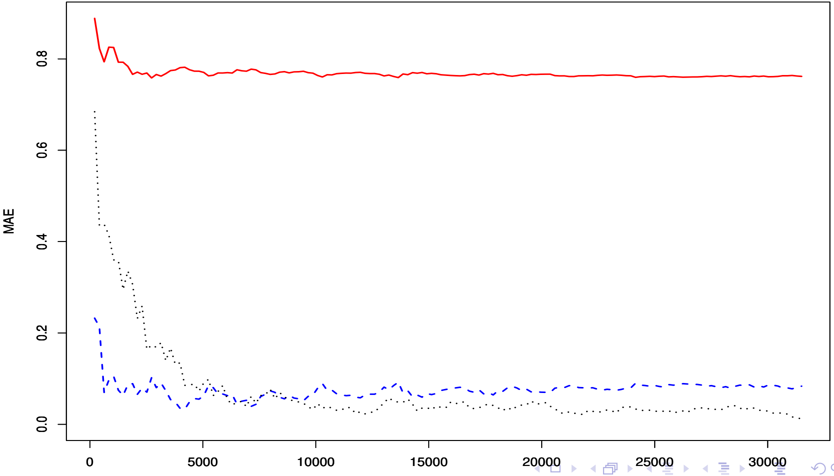
second way: first +

$g(x_{i,j}^*, \mathbf{x}_{\sim j}) = f(x_{i,j}^*, \mathbf{x}_{\sim j}) - \sum_{i=1}^{p} c_i f(x_{i,j}^*, \mathbf{x}_{\sim j})$

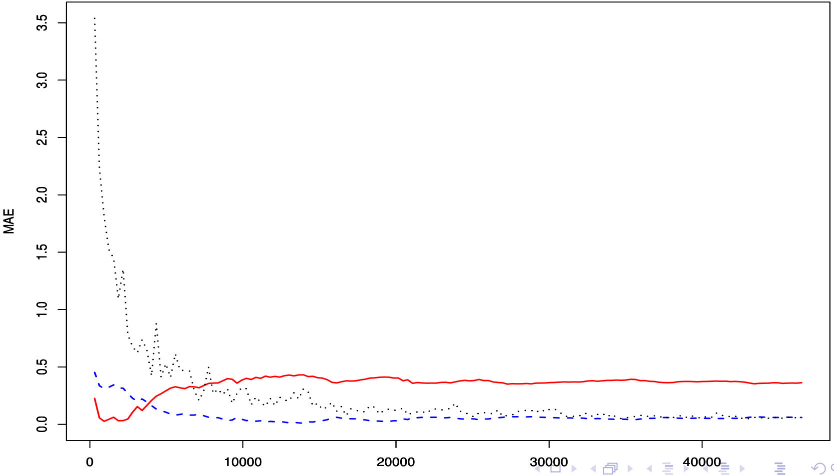
Numerical Tests: result for counter function



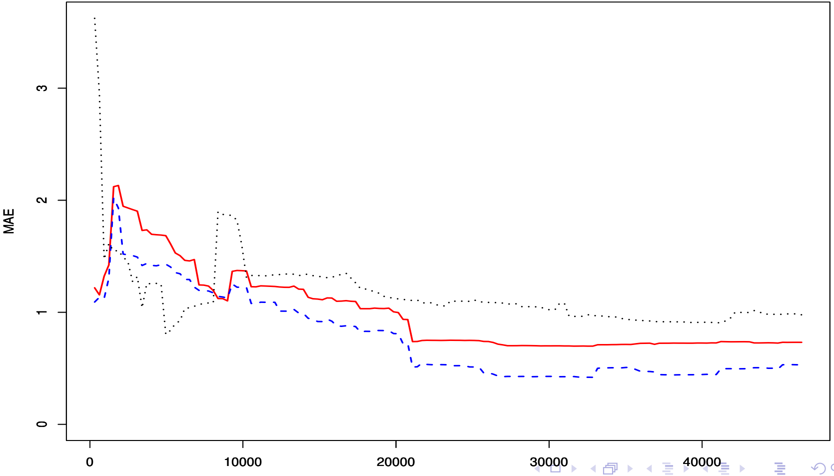
Numerical Tests: result for Sobol'function, type A



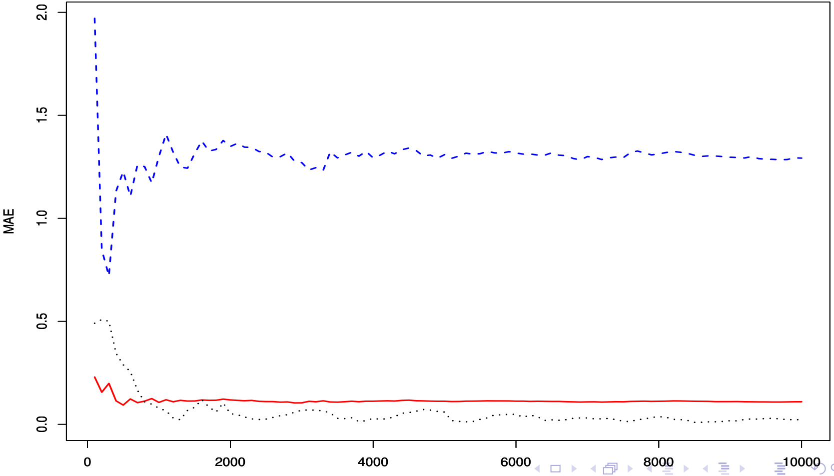
Numerical Tests: result for Sobol'function, type B



Numerical Tests: result for Sobol'function, type C



Numerical Tests: result for Ishigami'function



Conclusion and perspectives

- New methods of computing TSI based on quadrature rules
  - derive formulas for estimating TSI
  - derive the error of TSI based on quadrature rules
- Method, compared to Saltelli 2010 [6] + Sobol 2007 [7], is promising :
  - give reasonable TSI indices with few model runs
  - improves a bit TSI for functions with high effective dimension
  - method ' error is constant after some sample size
- Perspectives:
  - improve method' error: quadrature rules on subset
  - how to choose the order of quadrature rules (p)
  - other rules: Gauss is efficient for regular functions

Thanks for your attention!

Questions and propositions ?

Feel free to contact me!

References

-[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. Global Sensitivity Analysis: the Primer. Wiley, New York, 2008.
-[2] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. and Comp. in Simul 55, pp. 271280, 2001.
-[3] M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. RESS, 96, pp. 450-459, 2011.
-[4] I.M. Sobol, S. Kucherenko, Derivative based Global Sensitivity Measures and their link with global sensitivity indices, Mathematics and Computers in Simulation, V 79, Issue 10, pp. 3009-3017, June 2009.
-[5] M. Lamboni, B. Iooss, A-L. Popelin, and F. Gamboa. Derivative-based global sensitivity measures: general links with sobol' indices and numerical tests. Math. and Comp. in Simul. (2013)
-[6] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comp. Phy. Com. 181, pp. 259-270, 2010.
-[7] Sobol, I. Global sensitivity analysis indices for the investigation of nonlinear mathematical models. Matematicheskoe Modelirovanie 19 (2007), 23 - 24.