Sobol' indices and variance reduction diagram estimation from samples used for uncertainty propagation

Abed Benaïchouche (a.benaichouche@brgm.fr)
Jérémy Rohmer

BRGM, The French Geological Survey, France

1. Objective:

In this paper we present an efficient algorithm to estimate first order and second order Sobol' indices (SI) and its relationships with the inputs parameters variation range width. In the variance-based global sensitivity measure, SI are quantities defined by normalizing parts of variance in ANOVA decomposition (Sobol', 1993). They are estimated from the ratio between the variance of the conditional expectation of the output given the input and the unconditional variance of the output (eq .1). Many techniques have been proposed to estimate these indices. A recent review of these methods can be found in (Borgonovo & Plischke, 2015). Among others Monte-Carlo-based algorithm (Sobol' 1993, Saltelli et al., 1999, Jansen 1999, Monod et al., 2006, etc.) require a special random sampling scheme i.e. they cannot directly use samples used for the uncertainty propagation. Building on a similar idea than Plischkle (Plischke, 2010) we propose a methodology that allows the computation of SI form a set of given data. We propose to start from the local conditional variance to derive the global sensitivity indicator by estimating the variance of the conditional expectation (Eq. 2). The local information on sensitivity is summarized under the form of a Diagram of Expected Variance Reduction (DEVR), which relates the local reduction in uncertainty with the domain of variation of the considered input parameter with reduced width.

2. Methodology:

SI are given by eq. 1. The variance of the conditional expectation can be expressed as the expectation of conditional variance (eq. 2). In order to estimate the variance of the conditional expectation $V[E[(Y|x_i)]]$, we first estimate the expectation of the conditional variance $E[V[(Y|x_i)]]$ by following these steps: (1) Partition of the input parameters space into clusters (K-means algorithm with fixed number of samples for example), in which the variation of x_i is supposed weak (see fig. 1.A). (2) Computation of the local conditional variance $V[(Y|x_i)]$ (LCV) for each cluster (see fig. 1.B). (3) Estimation of the expectation of local conditional variance $E[V[(Y|x_i)]]$ (ELCV) with eq. 2 (fig. 1.C). (4) SI are obtained from the average of the expectation of local conditional variance. Higher order effects can be estimated following the same scheme.

$$S_i = \frac{V[E[(Y|x_i)]]}{V[Y]} \qquad \text{(eq. 1)}$$

$$V[E[(Y|x_i)]] = V[Y] - E[V[(Y|x_i)]] \qquad \text{(eq. 2)}$$

The described methodology allows also the computation of the relationship between SI and the inputs parameters variation range width (DEVR). In other words, this answers the question of what would be SI values if the initial variation range was reduced or true value of parameters was known: this can be achieved without new simulations run contrary to classical methods.

3. Results and Conclusions:

The proposed algorithm is used to estimate the first order and second order SI and the DEVR on the Ishigami function (eq. 3). This widely used test function exhibits strong nonlinearity and non-monotonicity (Sobol' & Levitan 1999) which make it challenging.

$$y = f(x_1, x_2, x_3) = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4 \sin(x_1), \text{ where } x_i \sim U(-\pi, \pi) \qquad \text{(eq. 3)}$$

We tested the developed algorithm termed ELVR against Jansen's algorithm (Jansen, 1999) and EASI (Plischke, 2010) for the first effects (5,000 model run) and against peek and freeze algorithm (Fruth & al 2014) and Monod's formulation (Monod et al. 2006) for the second order effects (10,000 model run). The results are reported in fig. 2 and fig. 3: these show a very good convergence for the first order SI index with less than 10,000 simulations as well as for the second-order indices especially for indices of high values.

From the DEVR (fig. 4) many information on local sensitivity can be extracted. Among others:

- In fig. 2, $S_1 = 0.31$, which means that if x_1 is fixed; a reduction of 31% on the total variance is expected. But in the DEVR (fig. 4) we see that depending on the chosen value the reduction can vary between 55% ($x_1^* = \{-\pi, 0, \pi\}$) and 0% ($x_1^* = \{-\pi/2, \pi/2\}$). But on average it's equal to 31%.

- The reduction of variation range of x_1 from $[-\pi, \pi]$ to $[-\pi/8, \pi/8]$ reduces the total variance of output by more than 50%. On the other hand, a reduction from $[-\pi, \pi]$ to $[3\pi/8, 5\pi/8]$ do not $V[Y]$.
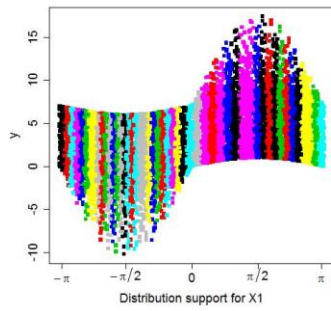
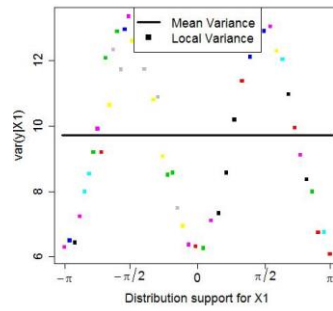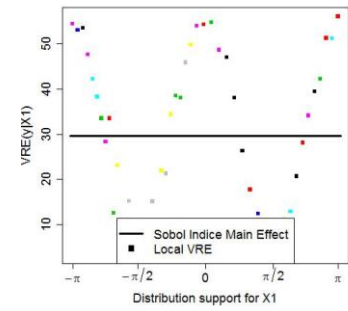**Fig.1.A.** Space partitioning for X1

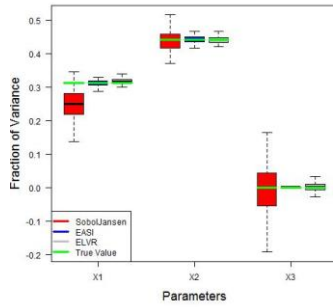**Fig.1.B.** LCV for X1

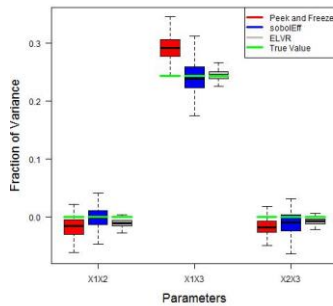**Fig.1.C.** LECV for X1

Fig.2. SI first order effects

Fig.3. SI second order effects
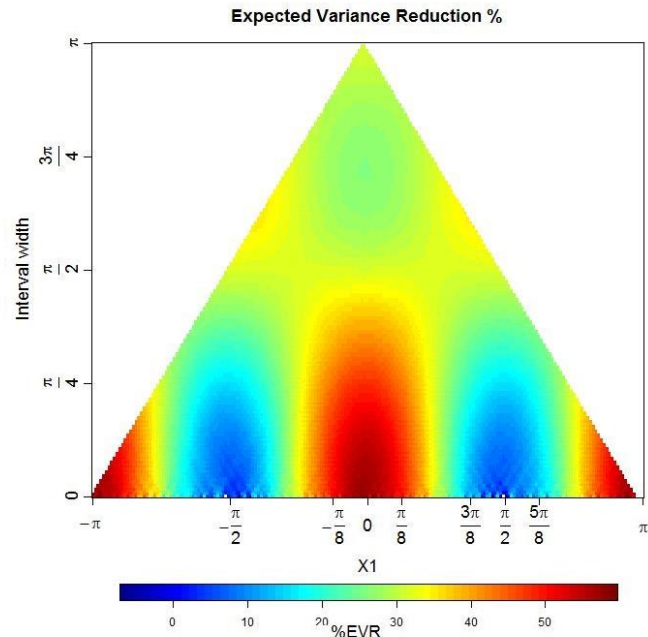
Fig.4. DEVR for X1

References:

Cukier, R. I., Levine, H. B., and Shuler, K. E. (1978). Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics, 26:1–42.

J. Fruth, O. Roustant, S. Kuhnt, 2014, Total interaction index: A variance-based sensitivity index for second-order interaction screening, J. Stat. Plan. Inference, 147, 212–223.

M.J.W. Jansen, 1999, Analysis of variance designs for model output, Computer Physics Communication, 117, 35–43.

Monod, H., Naud, C., Makowski, D. (2006), Uncertainty and sensitivity analysis for crop models in Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications, Elsevier.

E. Plischke, 2010, An effective algorithm for computing global sensitivity indices (EASI). Reliability Engineering & System Safety 95, 354-360.

Saltelli, A., Tarantola, S., and Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. Technometrics, 41:39–56.

Sobol', I. M. (1993). Sensitivity analysis for nonlinear mathematical models. Mathematical Modeling and Computational Experiment, 1:407–414.

Sobol', I. M., & Levitan, Y. L. (1999). On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index. Computer Physics Communications, 117(1), 52-61.

Tarantola, S., Gatelli, D., and Mara, T. A. (2006). Random balance designs for the estimation of first-order global sensitivity indices. Reliability Engineering and System Safety, 91:717–727.

E. Borgonovo, E. Plischke, Sensitivity analysis: A review of recent advances. European Journal of Operational Research. In press (2015).