Sensitivity analysis. A coming of age

Stefano Tarantola, Andrea Saltelli, Paola Annoni

VARIANCE-BASED SENSIVITY ANALYSIS NEW STRTAGIES FOR COMPUTING TOTAL

Stefano Tarentola, Andrea Saltelli and Paola Annoni,
Joint Research Center
Institute for the Protection and Security of the Citizen
T.P. 361 Via E. Fermi, 2749 - 21027 Ispra (VA) – I

Stefano.tarantola@jrc.it andrea.saltelli@jrc.it paola.annoni@jrc.it

The aim of sensitivity analysis is to apportion the uncertainty of the output of a simulation model with respect to its inputs. In spite of the considerable developments which have taken place in this discipline, of the good practices that have emerged, and of existing guidelines for sensitivity analysis issued on both sides of the Atlantic, very primitive sensitivity analysis tools, based on "one-factor-at-a-time" (OAT) approaches, are still being used (Campbell et al., 2008; Hasmadi and Taylor, 2008; Stites et al., 2007; Murphy et al. 2004;).

In the context of model corroboration or falsification, this use of OAT methods is illicit and unjustified, unless the model under analysis is proved to be linear (Saltelli et al, 2008). In this talk we show that available good practices, such as variance based measures and screening methods, are able to overcome OAT shortcomings and are easy to implement. We illustrate new strategies for the efficient estimation of 'total order' sensitivity indices, which are effective variance-based tools to quantify at what extent an individual model input X_i can drive uncertainty in the model output Y by accounting for the presence of interactions between X_i and other model inputs.

A brief summary is given here to show the major features of variance-based methods. A Fourier implementation was developed in the seventies (Cukier et al. 1973), whereas antecedent works were made by Kolmogorov (see reviews Archer et al. 1997, Rabitz et al. 1999). However, the best systematization of the theory is due to Sobol' (Sobol 1990), while total sensitivity indices were introduced by Homma and Saltelli (1996). For reviews see also Saltelli et al. (2005), Helton et al. (2006) or Chapter 8 in Saltelli et al. (2000).

Prior to the introductino of total sensitivity indices, first and higher order effects are briefly discussed. In a variance based sensitivity framework the first order effect for a given model input X_i is written as

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y \mid X_i\right)\right) \tag{1}$$

The meaning of the inner expectation operator is that the mean of Y is taken over all possible values of non-X_i i.e. over all possible values of $\mathbf{X}_{\sim i}$, while keeping X_i fixed. The outer variance is taken over all possible values of X_i. The associated sensitivity measure (first order sensitivity coefficient) is written as:

$$S_i = \frac{V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y \mid X_i\right)\right)}{V(Y)} \tag{2}$$

Note that S_i is the expected fractional reduction of variance that would be achieved if X_i could be fixed. For instance, $S_i = 0.1$ implies that on average we can expect to have a 10% reduction in the variance $V(Y)$ of Y by fixing X_i. The terms 'expected' and 'on average' refer to the fact that we do not know where -- within its interval of definition - we should fix X_i. Alternatively it can be

said that the definition of S_i entails an averaging over the distribution of X_i in order to be independent from a particular point over this distribution.

Higher order sensitivity indices come from a functional decomposition of a function $Y = f(\mathbf{X})$ into main effects and interactions, $\mathbf{X}$ denoting the vector of model inputs X_i. When inputs are mutually independent, the functional decomposition can be translated into a variance decomposition which may be written in terms of first-order and higher order sensitivity indexes:

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \ldots + S_{123\ldots k} = 1 \tag{3}$$

where, for instance, the second-order interaction terms S_{ij} are defined as:

$$S_{ij} = \frac{V_{X_i X_j}(E_{\mathbf{X}_{\sim ij}}(Y \mid X_i, X_j) - V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y \mid X_i) - V_{X_j}(E_{\mathbf{X}_{\sim j}}(Y \mid X_j)}{V(Y)}$$

Another variance based measure is the total effect index (Homma and Saltelli 1996):

$$S_{T_i} = \frac{E_{\mathbf{X}_{\sim i}}\left(V_{X_i}\left(Y \mid \mathbf{X}_{\sim i}\right)\right)}{V(Y)} \equiv 1 - \frac{V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y \mid \mathbf{X}_{\sim i}\right)\right)}{V(Y)} \tag{4}$$

S_{T_i} measures the total effect (both first and higher orders - e.g. interactions - of X_i.

S_{T_i} is the expected fraction of variance that would be left if all inputs but X_i could be fixed. E.g. $S_{T_i} = 0.15$ would imply that fixing all inputs but X_i would still leave, on average, 15% of variance in Y. The average in this case is to be understood over all inputs but X_i. These techniques are available in the Simlab software, downloadable for free (Simlab v3. 2009).

For additive models the two measures S_i and S_{T_i} are equal to one another for each X_i. For an interacting input the difference $S_{T_i} - S_i$ is a measure of the strength of the interactions.

When the goal is to reduce model complexity the identification of non-influential model inputs comes into play. This setting is called *Factor Fixing* (*FF*) in Saltelli and Tarantola (2002). It can be seen that a necessary and sufficient condition for factor X_i to be totally non-influential is $E_{\mathbf{X}_{\sim i}}\left(V_{X_i}\left(Y \mid \mathbf{X}_{\sim i}\right)\right) = 0$, that is $S_{T_i} = 0$. The objective of the *FF* setting is to identify an input or a group of inputs which may be fixed to any given value within their range of uncertainty without significantly reducing the output variance. If this is possible, the remaining inputs will explain the most of the output variance. The implications on the process of simplification of complex models are evident. Besides, if there are prior beliefs about the importance of inputs, this setting may be used to endorse or disprove a given model representation.

Variance-based sensitivity analysis can be also used as 'diagnostic' tools to identify interactions among model inputs. In the presence of interacting inputs the first order sensitivity indices do not add up to one ($\sum_i S_i < 1$), and higher order terms have to be included to recover the total variance of Y, see expression (3). Specific interactions are measured by second and higher order terms S_{ij}, $S_{ijk}, \ldots$ while the total indices S_{T_i} only give the total effect for any given input.

Researchers in various application fields are more and more aware of the usefulness of total sensitivity indices (Chu et al., 2007; Helton et al., 2006) and are increasingly adopting them in testing their models.

References

Archer G, Saltelli A, Sobol' IM. 1997. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. Journal of Statistical Computation and Simulation 58: 99-120.

Campbell J.E. et al., 2008, Photosynthetic Control of Atmospheric Carbonyl Sulfide During the Growing Season, Science 322 1085-1088.

Chu PC, Ivanov LM, Margolina TM. 2007. On non-linear sensitivity of marine biological models to parameter variations. Ecological Modelling 206: 369-382.

Cukier RI, Fortuin CM, Schuler KE, Petschek AG, Schaibly JH. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. The Journal of Chemical Physics 59: 3873-3878.

Hasmadi M, Taylor JC. 2008. Sensitivity analysis of an optimal access road location in HillyForest area: a GIS approach. American Journal of Applied Sciences 5(12): 1686-1692.

Helton JC, Johnson JD, Salaberry CJ, Storlie CB. 2006. Survey of sampling based methods for uncertainty and sensitivity analysis, Reliability Engineering and System Safety 91: 1175-1209.

Homma T, Saltelli A. 1996. Importance measures in global sensitivity analysis of model output. Reliability Engineering and System Safety 52(1): 1-17.

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430: 768-772.

Rabitz H, Alis ÖF, Shorter J, Shim K. 1999. Efficient input-output model representations. Computer Physics Communication 117: 11-20.

Saltelli A, Chan K, Scott M. (Eds) 2000. Sensitivity Analysis. John Wiley and Sons publishers, Probability and Statistics series.

Saltelli A, Ratto M, Tarantola S, Campolongo F. 2005. Sensitivity Analysis for Chemical Models. Chemical Reviews 105(7): 2811-2828.

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. 2008. Global Sensitivity Analysis. The Primer, John Wiley and Sons.

Saltelli A, Tarantola S. 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. Journal of American Statistical Association 97: 702–709.

Simlab v3. 2009. Software for uncertainty and sensitivity analysis. Joint Research Centre of the European Commission (*http://simlab.jrc.ec.europa.eu/*)

Sobol' IM. 1990. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie 2: 112-118, in Russian, translated in English as: Sobol' IM. 1993. Sensitivity analysis for non-linear mathematical models. Mathematical Modelling and Computational Experiment 1: 407-414.

Stites E.C. et al., 2007. Network analysis of oncogenic Ras activation in cancer, Science 318, 463-467

Stefano Tarantola (short biography)

Stefano Tarantola is a scientific officer at the European Commission, Joint Research Centre with an MSc in Engineering and a PhD in Science and Technologies for Engineering at the Polytechnic of Milan.

He conducts methodological work in the field of statistical indicators and composite indicators for policy making at the European level. He has experience in systems analysis, modelling and in methods to perform robustness analysis of decision processes to policy assumptions. He develops methodologies for sensitivity analysis and combines these with participatory methods for the assessment of the robustness of composite indicators. He is the author of papers in the peer-reviewed literature, co-author of four books on uncertainty and sensitivity analysis and of a handbook on composite indicators development with the OECD.