

Sixth International Conference on Sensitivity Analysis of Model Output

Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs

Herschel Rabitz*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

Abstract

A new unified global sensitivity analysis framework is introduced for systems whose input probability distribution can be independent and/or correlated. For correlated inputs, three sensitivity indices are defined to fully describe the total, structural (reflecting the system structure) and correlative (reflecting the correlated input probability distribution) contributions for an input or a subset of inputs. The magnitudes of all three indices need to be considered in order to quantitatively determine the relative importance of the inputs. For independent inputs, these indices reduce to a single index consistent with previous variance-based methods. This analysis is especially useful for the treatment of laboratory and field data.

Keywords: Sensitivity analysis; Variance analysis; Data correlation; Meta-modeling.

1. Main text

The objective of a global sensitivity analysis is to rank the importance of the system inputs (\mathbf{x}) considering their uncertainty and the influence they have upon the uncertainty of the system output y [1-3]. In global sensitivity analysis variance-based methods are commonly used [4-7] for quantifying the sensitivity of the output to the inputs in terms of a reduction in the variance of y :

$$S_i = V_i / V(y) = \text{Var}[E(y|x_i)] / \text{Var}(y) \quad (1)$$

$$S_{ij} = V_{ij} / V(y) = (\text{Var}[E(y|x_i, x_j)] - V_i - V_j) / \text{Var}(y) \quad (2)$$

...

where $E(\cdot)$ and $\text{Var}(\cdot)$ represent the expected value and variance; S_i and S_{ij} are referred to as the main and 1st-order interaction effects for x_i and x_i, x_j , respectively. When all of the inputs are independent, a simple decomposition of the output variance $V(y)$ is possible

$$V(y) = \sum_{i=1}^n V_i + \sum_{1 \leq i < j \leq n} V_{ij} + \dots + V_{12\dots n} = \sum_{j=1}^{2^n-1} V_{pj} \quad (3)$$

where V_{pj} are conditional variances of y for a fixed subset of inputs \mathbf{x}_{pj} . Then, the sensitivity index S_{pj} satisfies $0 \leq S_{pj} = V_{pj} / V(y) \leq 1$, and the importance rank of the inputs or subsets of inputs can be simply determined by comparing the magnitudes of the sensitivity indices.

* Corresponding author. Tel.: +1-609-258-3917; fax: +1-609-258-0967.
 E-mail address: hrabitz@princeton.edu.

When the inputs are correlated, some ambiguities arise in the definitions of sensitivity indices given by the variance-based method. The conditional variances will generally depend on the existence of correlations in the input variables. Adopting the same definition of sensitivity indices given by the variance-based methods for a given subset of inputs can lead to contributions from other correlated inputs contaminating the result [8]. This problem was also observed by Oakley and O'Hagan [1]. They demonstrated that $V(y)$ cannot be decomposed into a sum of squares as given in eq.(3), and the V_{pj} 's do not partition $V(y)$ for systems possessing a correlated input probability distribution. Therefore, the resultant relative importance of the inputs is questionable based on comparing the magnitudes of the sensitivity indices.

Here we introduce a new unified global sensitivity analysis framework for systems whose input probability distribution can be independent and/or correlated. In the ANOVA decomposition [5] an n -variate function can be represented as

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \dots + f_{12\dots n}(x_1, x_2, \dots, x_n) = f_0 + \sum_{j=1}^{2^n-1} f_{pj}(x_{pj}) \quad (4)$$

where $f_0 = E(y)$ and $E(f_{pj}(\mathbf{x}_{pj})) = 0$ for all the non-constant component functions above, the last term is determined by the difference of y and all other terms on the right, thus $f(\mathbf{x})$ is exactly equal to y . Suppose that y can be approximated by $n_p (< 2^n - 1)$ non-constant component functions in eq.(4)

$$y = f_0 + \sum_{j=1}^{n_p} f_{pj}(x_{pj}) + \varepsilon. \quad (5)$$

When all the f_{pj} 's are determined from a set of input-output data by an unbiased method (e.g., least-squares regression), the difference between y and its approximation $f_0 + \sum_{j=1}^{n_p} f_{pj}$ is orthogonal to the subspace spanned by all f_{pj} ($j=1, 2, \dots, n_p$) in the Hilbert space [9]. Using this condition, it can be readily proved that $V(y)$ can be decomposed as the sum of all the covariances, $\text{Cov}(f_{pj}, y)$, and the average square error ε^2

$$V(y) = \sum_{j=1}^{n_p} \text{Cov}(f_{pj}, y) + \overline{\varepsilon^2} = \sum_{j=1}^{n_p} \text{Cov}(f_{pj}, \sum_{k=1}^{n_p} f_{pk}) + \overline{\varepsilon^2} = \sum_{j=1}^{n_p} [\text{Var}(f_{pj}) + \text{Cov}(f_{pj}, \sum_{k=1, k \neq j}^{n_p} f_{pk})] + \overline{\varepsilon^2}. \quad (6)$$

If $\overline{\varepsilon^2}$ is sufficiently small compared to $V(y)$ (i.e., $f_0 + \sum_{j=1}^{n_p} f_{pj}(\mathbf{x}_{pj})$) is a good approximation for $f(\mathbf{x})$, the sum of the covariances forms a good decomposition of $V(y)$. When the f_{pj} 's are all of the component functions in eq.(4), $V(y)$ is exactly partitioned by all the $\text{Cov}(f_{pj}, y)$'s.

The covariance $\text{Cov}(f_{pj}, y)$ is the total contribution of f_{pj} composed of its structure piece (reflecting f_{pj} 's contribution in the system structure $y=f(\mathbf{x})$ and a correlation piece (reflecting the influence of the interaction between f_{pj} and the other f_{pk} 's through the correlated input probability distribution). Fixing some inputs may influence the distributions of the other inputs, and the total effect can decrease or increase the variance of the output, which yields a positive or negative $\text{Cov}(f_{pj}, y)$. For independent inputs all of the f_{pj} 's are mutually orthogonal, i.e., $\text{Cov}(f_{pj}, f_{pk}) = 0$, which gives $\text{Cov}(f_{pj}, y) = \text{Var}(f_{pj}) = V_{pj}$, and eq.(6) reduces to eq.(3). Therefore, eq.(3) is only a special case of eq.(6) for systems with independent inputs.

For systems with correlated inputs we define three sensitivity indices S_{pj} , S_{pj}^a , S_{pj}^b for f_{pj} as follows:

$$S_{pj} = \text{Cov}(f_{pj}, y) / V(y) \approx \sum_{s=1}^N f_{pj}(x_{pj}^{(s)}) (y^{(s)} - \bar{y}) / \sum_{s=1}^N (y^{(s)} - \bar{y})^2, \quad (7)$$

$$S_{pj}^a = \text{Var}(f_{pj}) / V(y) \approx \sum_{s=1}^N f_{pj}(x_{pj}^{(s)})^2 / \sum_{s=1}^N (y^{(s)} - \bar{y})^2, \quad (8)$$

$$S_{pj}^b = \text{Cov}(f_{pj}, \sum_{k=1, k \neq j}^{n_p} f_{pk}) / V(y) \approx S_{pj} - S_{pj}^a, \quad (9)$$

where s denotes the s th sample, and N is the total number of samples, \bar{y} is the average value of y . S_{pj}^a , which is always non-negative, reflects the contribution of \mathbf{x}_{pj} through f_{pj} in the system structure $f(\mathbf{x})$ (denoted as the *structural* contribution) and can be used to identify the important inputs in the system structure. S_{pj}^b , which can be positive or negative, reflects the contribution of \mathbf{x}_{pj} by the interaction of f_{pj} and other f_{pk} 's due to the correlated input probability distribution (denoted as the *correlative* contribution). S_{pj} is the total contribution of \mathbf{x}_{pj} through f_{pj} . When ε^2 is small, the sum of all S_{pj} 's will be close to unity. The magnitudes of all S_{pj} , S_{pj}^a , S_{pj}^b need to be considered in order to quantitatively determine the relative importance of the inputs acting either independently or collectively. When all

inputs are independent and the f_{pj} 's are mutually orthogonal, then $S_{pj}^b=0$, and $S_{pj}=S_{pj}^a$, which is the sensitivity index given by the variance-based methods.

The estimation of sensitivity indices is simple after the component functions f_{pj} 's in eq.(5) are determined by a suitable numerical method from a given set of data. The total sensitivity indices S_{Ti} , S_{Ti}^a and S_{Ti}^b also can be calculated by adding together all the sensitivity indices containing x_i . When $\sum_{j=1}^{np} S_{pj} \approx 1$, the resultant total sensitivity indices can be considered as reliable. This analysis technique can be applied to mathematical models, as well as measured data. The definition of sensitivity indices given by the variance-based methods for systems with independent inputs is a special case of the new unified treatment.

2. References

- [1] Oakley, J. E. and O'Hagan, A. 2004, *J R Statist Soc B* 66, Part 3, 751.
- [2] Saltelli, A.; Ratto, M.; Tarantola, S. and Campolongo, F. 2005, *Chem Rev* 105, 2811.
- [3] Ratto, M.; Pagano, A. and Young P. 2007, *Computer Phys Comm* 177, 863.
- [4] Cox, D. C. 1982, *IEEE Trans Reliab* 31, 265.
- [5] Sobol, I. M. 1993, *Mathematical Modeling and Computational Experiments* 1, 407.
- [6] Homma, T. and Saltelli, A. 1996, *Reliab Engng Syst Safety* 52, 1.
- [7] Saltelli, A.; Chan, K. and Scott, E. M.(eds) 2000, *Sensitivity Analysis*; John Wiley & Sons Ltd: New York.
- [8] Saltelli, A. and Tarantola, S. 2002, *J Amer Statis Assoc* 97, 702.
- [9] Deutsch, F. *Best Approximation in Inner Product Space*; 2001, Chapter 4, Springer-Verlag New York, Inc: New York.