

# A New Uncertainty Importance Measure

E. Borgonovo<sup>1</sup>

Institute for Quantitative Methods - Bocconi University

20135 Milano

Viale Isonzo 25, Italy

April 25, 2012

<sup>1</sup>[emanuele.borgonovo@unibocconi.it](mailto:emanuele.borgonovo@unibocconi.it)

## **Abstract**

Uncertainty in parameters is present in many risk assessment problems and leads to uncertainty in model predictions. In this work, we introduce a global sensitivity indicator which looks at the influence of input uncertainty on the entire output distribution without reference to a specific moment of the output (moment independence) and which can be defined also in the presence of correlations among the parameters. We discuss its mathematical properties and highlight the differences between the present indicator, variance-based uncertainty importance measures and a moment independent sensitivity indicator previously introduced in the literature. Numerical results are discussed with application to the probabilistic risk assessment model on which Iman (1987) first introduced uncertainty importance measures.

*Keywords: Importance Measures, Uncertainty Analysis, Global Sensitivity Analysis, Uncertainty Importance Measures, Probabilistic Risk Assessment.*

*Acknowledgments:* the author would like to thank Guido Consonni for very useful comments and suggestions during the development of this work. He would also like to thank the anonymous referees for the perceptive comments and the useful insights on the earlier version of the work.

## 1 Introduction

Dealing with uncertainty is one of the challenges of many quantitative risk assessment problems (Apostolakis (2005) [3]). As Hammit and Shiyakhter (1999) [12] underline, it is often “*the lack or sparsity of data*” which prevents the analyst/decision-maker from assigning a certain value to the parameters. Uncertainty in the inputs is reflected in uncertainty in model results and predictions (Apostolakis (1995) [2].)

Saltelli (2002) [42] defines sensitivity analysis (SA) as the study of how “*uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input (Saltelli (2002) [42]).*” With this respect, Saltelli (2002) [42] remarks that SA techniques to be utilized in the context of uncertainty analysis should satisfy the following three requirements: “*global, quantitative and model free.*” By global one means that the technique allows to take into consideration the entire input distribution. By model independent one means that no assumptions on the model functional relationship to its inputs is necessary in order for the SA method to produce accurate results.

Saltelli (2002) [42] shows that variance-based methods provide a set of tools that share the three above mentioned requirements. The works of Saltelli *et al* (2000) [40], Sobol’ (1993), (2001) and (2003) ([46], [47], [48]), Rabitz *et al* (1998) [34], Rabitz and Alis (1999) [35] and Alis and Rabitz (2001) [1] have established the theoretical and numerical background for the utilization of variance-based techniques. The demonstrated merits of variance-based global SA methods are the consideration of the entire range of variation of the inputs and the ability to identify individual parameter contributions and parameter interactions in a model independent fashion. Since Sobol’ decomposition method rests on the assumption of independent inputs, a limitation of a technical nature appears when one performs variance

decomposition in the presence of input correlations. More precisely, Bedford (1998) [4] proves that “*the values taken on by the indices depend on the ordering of the variables.*” The problem has next been tackled by Tarantola (2000) [49], Ratto *et al* (2001) [36] and Saltelli and Tarantola (2002) [41], who thoroughly discuss the use of variance based uncertainty importance measures in the presence of correlations among the parameters.

However, it has been recognized that the interpretation of global sensitivity analysis based on the sole variance as a way of indicating how “*the total uncertainty in model prediction is apportioned to uncertainty in the model input parameters (Homma and Saltelli (1995) [17])*” or “*the expected percentage reduction in the uncertainty ... that is attributable to each of the input variables (Iman *et al* (2005) [25])*” is not entirely satisfactory. In fact, Saltelli (2002) [42] underlines that variance-based methods “*implicitly assume that this moment (variance) is sufficient to describe output variability.*” Indeed, a decision-maker/analyst state of knowledge on a parameter or on a model output is represented by the entire uncertainty distribution (Savage (1972) [43]). With this respect, Helton and Davis (2003) ([16] Section 2, page 25) underline that any moment of a random variable “*provides a summary*” of its distribution with the inevitable “*loss of resolution*” that occurs when the information contained in the distribution is mapped into a single number. Thus, if an analyst aims at assessing which parameter influences the decision-maker state of knowledge the most, a sensitivity indicator should refer to the entire output distribution and not to one of its moments. With this respect, inspection of the whole decision-maker uncertainty requires to add a fourth feature to Saltelli’s three requirements, namely, moment independence.

In this work, we analyze how these issues can be addressed. To do so, we propose a global SA indicator (called  $\delta$ ) that considers the entire distribution both of the input and of the output (global) in a moment independent fashion (Figure 1). We define  $\delta$  so that its computation is well posed in the presence of correlations among the parameters. We derive analytically the main mathematical properties of  $\delta$ , showing that the importance of a parameter equals zero when the output is independent of the parameter and that  $\delta$  equals

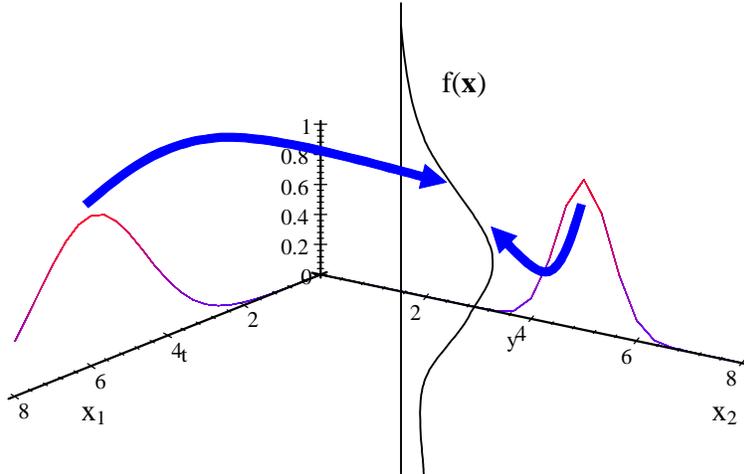


Figure 1:  $\delta$  aims at assessing the influence of the entire input distribution on the entire output distribution without reference to a particular moment of the output.

unity when the set of all inputs is considered. We propose an algorithm for its computation and analyze the numerical procedure.

We then compare  $\delta$  to both variance-based indicators and to the moment independent sensitivity indicator introduced by Chun *et al* (2000) [10]. With this respect, we show that the main difference between  $\delta$  and the Chun-Han-Tak (*CHT*) importance measure is that *CHT* requires to hypothesize a “*sensitivity case* (Chun *et al* (2000) [10], p. 314),” while  $\delta$  does not. Thus, *CHT* is investigating which of the parameters influence output uncertainty the most given the hypothesized change, while  $\delta$  does not require to presuppose a sensitivity case.

The application to the Ishigami test function (Nakashima and Iamato (1982) [30], Homma and Saltelli (1996) [17]) details the comparison of the ranking obtained with  $\delta$  to the ranking obtained with the importance indicators of Iman-Hora (1990) [24], the global sensitivity indices (Homma and Saltelli (1996) [18]) and *CHT* (Chun *et al* (2000) [10]).

We then discuss the application of  $\delta$  to the uncertainty and global SA of a probabilistic safety assessment model first introduced in Iman (1987) [22] and next utilized as a test case

in several works (Chun *et al* (2000) [10].) In this analysis, we also focus on the importance of parameter groups and interactions, which shall enable us to further highlight the difference between  $\delta$  and variance-based approaches.

Results of both applications show that variance-based global SA indicators and  $\delta$  agree in identifying the less relevant parameters with respect to (w.r.t.) the output uncertainty. However, differences in the ranking of the most relevant parameters emerge, due to the different scope of the indicators.

In Section 2, we present an overview of global SA as used in Risk Analysis, starting with variance based techniques and ending with a moment independent SA indicator introduced in Chun *et al* (2000). Section 3 proposes a new moment independent importance measure and discusses its mathematical properties. In Section 4, the application to the Ishigami test function is discussed with the purpose of illustrating the properties of the new indicator and comparing results with those of the other uncertainty importance measures presented in Section 2. In Section 5 the application to the uncertainty analysis of the probabilistic risk assessment model on which uncertainty importance measures were first introduced by Iman (1987) [22] is detailed. Section 6 deals with computational aspects and presents perspectives of future research. Section 7 offers conclusions.

## 2 Global Sensitivity Analysis

Global sensitivity analysis (SA) is the term utilized to denote the set of techniques aimed at determining which of the input parameters influence output the most when uncertainty in the parameters is propagated through the model (Iman (1987) [22], Iman and Conover (1987) [23], Iman and Hora (1990) [24], Helton (1993) [14], McKay (1996) [27], Saltelli (1997) [38], Helton and Davis (2003) [16].) In the family of global SA indicators one can include non-parametric techniques (Saltelli and Marivoet (1990) [37], Helton and Davis (2000) [15], Helton and Davis (2003) [16],) variance based techniques (Hora and Iman (1986)[19], Iman (1987) [22], Sobol' (1993) [46], (2001) [47] and (2003) [48], Rabitz *et al* (1998) [34], Rabitz and Alis (1999) [35] and Alis and Rabitz (2001) [1], Saltelli *et al* (1999) [39], Saltelli *et al*

(2000) [40],) and moment independent techniques (Park and Ahn (1994) [33], Chun et al (2000)[10].) Indicators created for global SA purposes are called global importance measures (Homma and Saltelli (1996) [17]) or uncertainty importance measures (Bier (1993) [5], Iman and Hora (1990) [24], Homma and Saltelli (1996) [17], Chun *et al* (2000) [10]) to differentiate them from local importance indicators (Borgonovo and Apostolakis (2001) [6], Cheok *et al* (1998) [9], Borgonovo *et al* (2003) [7],) and screening methods (Morris (1991) [29], Kelijnen (2005) [21].)

With reference to Saltelli's requirements, several authors have underlined that non-parametric methods often lack model independence. For example, regression based methods are appropriate when a linear input-output relationship exists (Frey and Patil (2002) [11] discuss limitations). Saltelli and Marivoet (1990) [37] and Hora and Helton (2003) [20] underline the fact that ranking provided for by the Spearman rank correlation coefficient would be significant if a monotone input/output relationship were to hold. To overcome this limitation, a test for non-monotone relationship is introduced in Hora and Helton (2003) [20].

As an alternative way of measuring uncertainty importance, after the early works of Nakashima and Yamato (1982) [30] and Bier (1983) [5], particular attention is deserved by the Iman and Hora indicator (Hora and Iman (1986) [19], Iman (1987) [22], Iman and Hora (1990) [24],) defined as follows:

$$IH_i = V[Y] - E\{V[Y|X_i]\} = V\{E[Y|X_i]\} \quad (1)$$

where  $V[Y]$  is the variance of the model output  $Y$ , and  $E\{V[Y|X_i]\}$  is the conditional expected value of  $V[Y]$  given  $X_i$  and the expectation is taken over the possible values of  $X_i$ , weighted by the appropriate density. It can be proven that the Iman-Hora uncertainty importance measure ( $IH_i$  from now on) is the expected reduction in output variance that can be achieved if uncertainty in  $X_i$  is eliminated (Saltelli (2002) [42], and Saltelli *et al* (2000) [40]).

As Saltelli *et al* (2000) [40] underline, robustness problems connected with  $IH_i$  have been solved through the global importance measures introduced in the works of Sobol' (1993) [46]

and further developed by Homma and Saltelli (1996) [17], Sobol' (2001) [47] and (2003), Rabitz and Alis (1999) [35] and Alis and Rabitz (2001) [1]. In these works it is shown that, letting  $\underline{X} \in [0, 1]^n$  be a set of random independent variables uniformly distributed in the unitary hypercube, and

$$Y = g(\underline{X}) \quad (2)$$

a square-integrable function, then  $g(\underline{X})$  can be uniquely decomposed as follows

$$g(\underline{X}) - g_0 = \sum_{i=1}^n g_i(X_i) + \sum_{i<j} g_{i,j}(X_i, X_j) + \dots + g_{1,2,\dots,n}(X_1, X_2, \dots, X_n) \quad (3)$$

and the variance of  $Y$  can be decomposed in the following sum

$$V[Y] = \sum_{i=1}^n V_i + \sum_{i<j} V_{i,j} + \sum_{i<j<m} V_{i,j,m,\dots} + V_{1,2,\dots,n} \quad (4)$$

where

$$\left\{ \begin{array}{l} V_i = \int \dots \int [g_i(\underline{X})]^2 dX_i \\ V_{i,j} = \int \dots \int [g_{i,j}(X_i, X_j)]^2 dX_i dX_j \\ \dots \\ V_{i,j,\dots,m} = \int \dots \int [g_{i,j,\dots,m}(X_i, X_j, \dots, X_m)]^2 \prod_{k=i,j,\dots,m} dX_k \end{array} \right. \quad (5)$$

Each of the integrands  $g_{i,j,\dots,m}(X_i, X_j, \dots, X_m)$  in eq. (5) is found by iterative expectations on  $Y$  (Sobol' (1993) [46], Bedford (1998) [4]). Eqs. (3) and (4) imply that, in the absence of input correlations, variance decomposition directly mirrors the model structure, evidencing the presence of interactions and the contribution to the model output due to individual and parameter groups.

It turns out that the first order terms,  $V_i$ , are the “*expected amount of variance reduction that would be achieved for  $Y$ , if we were able to specify  $X_i$  exactly*” (Bedford (1998) [4], Saltelli *et al* (2000) [40])” and, therefore, coincide with the  $IH_i$  indicator [eq. (1)].

Sobol' (1993) [46] introduced the “*sensitivity estimates*” of order  $r$ :

$$S_{i_1, i_2, \dots, i_r} \equiv \frac{V_{i_1, i_2, \dots, i_r}}{V[Y]} \quad (6)$$

$S_{i_1, i_2, \dots, i_r}$  are the ratios of the interaction terms of order  $r$ ,  $V_{i_1, i_2, \dots, i_r}$  in eq. (4), and  $V[Y]$ .

Homma and Saltelli (1996) [18] introduced the concept of global sensitivity indices. Particular

interest is deserved by the interpretation of the first order ( $S1_i$ ) and total order sensitivity indices ( $ST_i$ ) — or “total effects” in Saltelli (2002) [42]. — The first order indices:

$$S1_i \equiv \frac{V_i}{V[Y]} \quad (7)$$

represent the expected percentage reduction in  $V[Y]$  which is obtained when uncertainty in  $X_i$  is eliminated (Saltelli (2002) [42]). Note that if one selects  $S1_i$  as uncertainty importance measure of  $X_i$ , one would obtain the same ranking as with  $IH_i$ .

The total effects

$$ST_i \equiv \sum_{r=1}^n \sum_{l_1, \dots, l_r} S_{l_1, \dots, l_r} (l_1 = i) \quad (8)$$

represent the expected percentage of variance that remains if all parameters were known but  $X_i$  (Saltelli (2002) [42]). Selecting  $ST_i$  as uncertainty importance measure one would be measuring the importance of a parameter as the percentage of the output variance associated with the parameter (Homma and Saltelli (1995) [17] and (1996) [18], Saltelli et al (1999) [39], Saltelli *et al* (2000) [40]).

Several studies have been performed on the computation of the global sensitivity indices: estimation procedures are the Extended FAST (Saltelli et al. (1999) [39],) the method of Sobol’ (Sobol’ (2001) [47],) and others (see Homma and Saltelli (1995) [17] and (1996) [18], and Alis and Rabitz (2001) [1].)

We note that Sobol’ theorem holds under the assumption that inputs are independent. Oakley and O’Hagan (2004) [31] evidence that in the case of uncorrelated inputs “*the representation* (i.e. Sobol’ decomposition) *reflects the structure of the model itself*,” while it does not reflect the model structure when correlations emerge. In the case of dependent inputs, Bedford (1998) [4] shows that the function decomposition is no more unique, and “*the values taken on by the indices depend on the ordering of the variables*.” This problem has then been addressed by Saltelli and Tarantola (2002), which have identified two lottery settings for sensitivity analysis in the presence and absence of correlations. The first setting consists in identifying the factor that, if determined, would lead to the greatest reduction

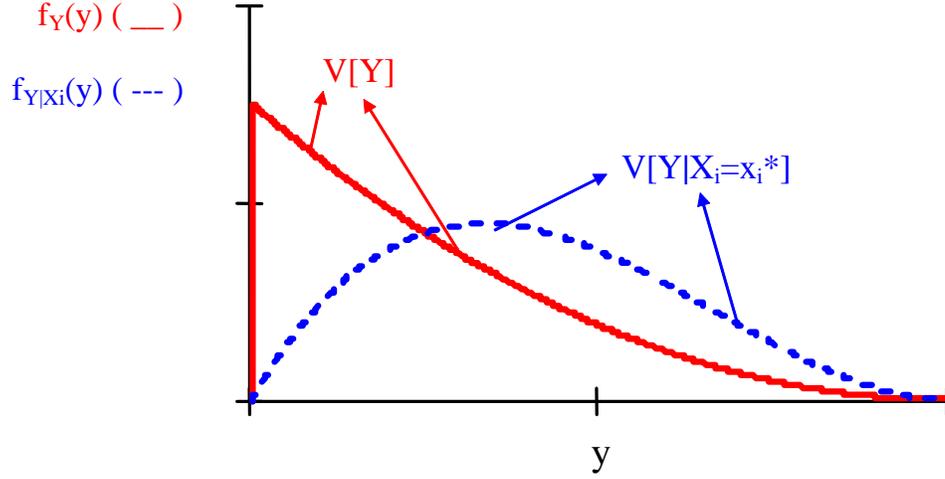


Figure 2: Unconditional ( $V[Y]$ ) and Conditional variance ( $V[Y|X_i = x_i^*]$ ) of  $Y$  when  $X_i$  is fixed at  $x_i^*$ .

in the variance of  $Y$ . The idea is that, by fixing  $X_i = x_i^*$ , one would obtain a new output distribution, namely  $f_{Y|X_i}(y)$ , that would produce a new output variance (Figure 2):

$$V[Y|X_i = x_i^*] \tag{9}$$

However, (see Saltelli and Tarantola (2002)), since  $X_i$  is a random variable,  $V[Y|X_i = x_i^*]$  is in its turn a random variable. Then

$$E_{X_i} \{V[Y|X_i]\} \tag{10}$$

is the expected remaining variance if one came to know  $X_i$  exactly. Utilizing eq. (1), one notes that the factor associated with the lowest  $E_{X_i} \{V[Y|X_i]\}$  are the more effective in reducing output variance (Saltelli et al (2000) [40].)

The second lottery setting of Saltelli and Tarantola (2002) [41] parallels the first, and consists in betting on the sets of factors that lead to

$$V[Y] < V_{tar} \tag{11}$$

where  $V_{tar}$  is a target variance. In this case, the terms  $V[Y | X_i, X_j, \dots, X_m]$  matter, and by extension of eq. (1), the sensitivity measures become:

$$V_{i,j,\dots,m} = V[E(Y | X_i, X_j, \dots, X_m)] \quad (12)$$

We note that, based on classical utility theory, variance is not sufficient to the determination of the decision maker state of knowledge in general. According to the theory, variance is sufficient to describe uncertainty in the following two cases: a) the decision maker possesses a quadratic utility function; b) the random variable is normally distributed (Huang and Litzenberger (1988) [13], Ch.3, pp. 61-62.) Hence, identifying which of the parameters reduces variance the most is not equivalent to identify which parameters influence the decision maker state of knowledge of the output the most, since  $V[Y]$  is just one of the moments of the output distribution.

Following this line of thought, Chun *et al* (2000) [10] introduced a global sensitivity indicator which is moment independent and looks at the entire distribution of the model output. The definition of the *CHT* indicator is as follows:

$$CHT_i = \left[ \int (y_\alpha^i - y_\alpha^0)^2 d\alpha \right]^{\frac{1}{2}} / E[Y^0] \quad (13)$$

where  $y_\alpha^i$  is the  $\alpha^{th}$  quantile of  $Y$  for the sensitivity case, and  $y_\alpha^0$  is the  $\alpha^{th}$  quantile of  $Y$  for the base case.  $CHT_i$  is expressed in terms of the cumulative distribution function of  $Y$  ( $F_Y$ ), and, intuitively speaking, represents the (square of the) area related to a shift in  $F_Y$  from the base case to the sensitivity case. By sensitivity case it is meant a recomputation of the model when: “1) the uncertainty in a parameter is completely eliminated; 2) the uncertainty range is changed; and 3) the type of distribution is changed” (Chun *et al* (2000) [10]). All three cases reflect a change in the state of knowledge of the analyst regarding the input parameters.

It is useful to remark two main differences in the definitions of  $CHT_i$  on the one side and  $ST_i/IH_i$  on the other side:

◁  $CHT_i$  requires the performance of a sensitivity case, while  $ST_i$  and  $IH_i$  do not;

$\triangleleft$   $ST_i$  and  $IH_i$  refer to a particular moment of the distribution of  $Y$ , namely,  $V[Y]$ , while  $CHT_i$  does not.

In other words, the question answered by  $CHT_i$  concerns the parameter that provokes the greatest change in the distribution of  $Y$  when, for example, the uncertainty in all the parameters is reduced by, say, a factor of 10 (Chun *et al* (2000) [10]).  $ST_i$  and  $IH_i$  measure the relevance of the parameter contribution to  $V[Y]$  given the current state of knowledge, *i.e.* without requiring to specify a sensitivity case that reflects an hypothetical decision-maker state of knowledge change.

In the next Section, we introduce a global sensitivity indicator, which is independent of the moments of the model output and independent of the sensitivity case.

### 3 A Moment Independent Importance Measure

In this Section, we present an uncertainty importance measure with the following characteristics: it does not refer to a particular moment of  $Y$  — and with this respect is similar to  $CHT_i$  but different from  $ST_i$  — and does not require a “sensitivity case” — with this respect it is similar to  $ST_i$  but different from  $CHT_i$ . — We also try and define the new indicator in such a way that its definition is properly posed in the presence of correlations among the parameters.

We start with the relevant notation. Let:

1.

$$\underline{X} = (X_1, X_2, \dots, X_n) \in R^n \tag{14}$$

the set of uncertain input parameters;

2.

$$Y = g(\underline{X}), g(\underline{X}) : E \subseteq \mathbb{R}^n \longrightarrow \mathbb{R} \tag{15}$$

be the functional relationship between output  $Y$  and input  $\underline{X}$ ;

3.  $\underline{x} = (x_1, x_2, \dots, x_n)$  a realization of  $\underline{X}$ ;

4.  $F_{\underline{X}}(\underline{x})$  the (subjective) cumulative distribution of  $\underline{X}$ , *i.e.* the joint cumulative distribution of the  $X_i$ .  $F_{\underline{X}}(\underline{x})$  expresses the analyst state of knowledge on  $\underline{X}$ ;
5.  $f_{\underline{X}}(\underline{x})$  the corresponding joint density of  $\underline{X}$ ;
6.  $f_{X_i}(x_i)$  the marginal density of  $x_i$ . As it is well known it is related to the joint density by:

$$f_{X_i}(x_i) = \int \dots \int f_{\underline{X}}(\underline{x}) \prod_{s \neq i} dx_s \quad (16)$$

7.  $F_Y(y)$  the cumulative distribution function of the model output  $Y$ ;
8.  $f_Y(y)$  the corresponding density;
9.  $f_{Y|X_i}(y)$  the conditional density of  $Y$  given that one of the parameters,  $X_i$ , assumes a fixed value.

The rationale behind the definition of the following moment independent importance indicator is as follows. The unconditional density/cumulative distribution of  $Y$  obtained with all parameters free to vary in their uncertainty range are  $f_Y(y)/F_Y(y)$  (the continuous line in Figure 3 shows an example of density  $f_Y(y)$ ). Suppose now that we are able to fix one of the inputs at, say,  $x_i^*$ . We would obtain the conditional density/distribution of  $Y$  given that  $X_i$  is fixed at  $x_i^*$ , namely  $f_{Y|X_i}(y)/F_{Y|X_i}(y)$  (dashed line in Figure 3).

The shift between  $f_Y(y)$  and  $f_{Y|X_i}(y)$  can be measured by the total area evidenced in Figure 3. Such an area is given by:

$$s(X_i) = \int |f_Y(y) - f_{Y|X_i}(y)| dy \quad (17)$$

Eq. (17) shows that  $s(X_i)$  is dependent on  $X_i$ , and as such, it is a function of random variable. The expected shift is given by:

$$E_{X_i}[s(X_i)] = \int f_{X_i}(x_i) \left[ \int |f_Y(y) - f_{Y|X_i}(y)| dy \right] dx_i \quad (18)$$

We then propose the following definition.

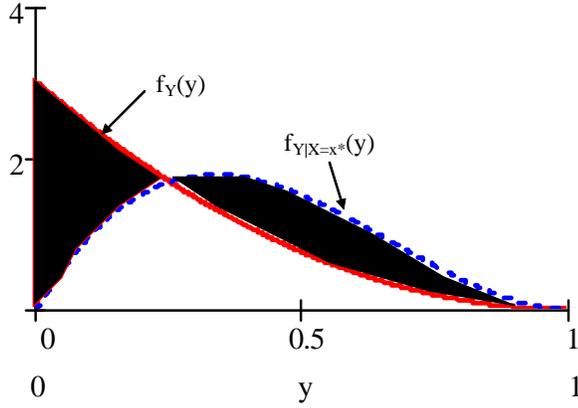


Figure 3:  $f_Y(y)$  (continuous) and  $f_{Y|X_j=x_j^*}(y)$  (dashed). The shift between the two densities is measured by the shaded regions.

**Definition 1** We name the quantity

$$\delta_i = \frac{1}{2} E_{X_i} [s(X_i)] \quad (19)$$

*moment independent sensitivity indicator of parameter  $X_i$  w.r.t. output  $Y$ .*<sup>1</sup>

$\delta_i$  represents the normalized expected shift in the distribution of  $Y$  provoked by  $X_i$ .

We now discuss some of the properties of  $\delta_i$  (Table 1.)

Property nr. 1 in Table 1 bounds the possible values that the  $\delta_i$  of an individual parameter can assume: it can be proven that  $\delta_i$  lies between 0 and 1 (the proof is in Section 8.1 in the Appendix). In particular, one finds that  $\delta_i$  is zero when  $Y$  is independent of  $X_i$  (Property 2 in Table 1.) In fact, if  $Y$  is independent of  $X_i$ , one would not get any change in  $f_Y(y)$  for any value  $x_i$  assumed by  $X_i$ . Thus,  $f_{Y|X_i}(y) = f_Y(y)$  and the integrand in eq. (18) is null for all  $x_i$ .

Property nr. 3 suggests that the  $\delta$  of all parameters equals unity. To prove the property, however, we need to extend the definition of  $\delta$  from an individual parameter to a group of

<sup>1</sup>A more technical definition of  $\delta_i$  is as follows:

$$\delta_i = \frac{1}{2} E_{X_i} \left[ \int |d\mu_Y - d\mu_{Y|X_i}| \right] \quad (20)$$

where  $\mu_Y$  and  $\mu_{Y|X_i}$  are, respectively, the unconditional and conditional measures of  $Y$ .

Nr.	Property	Meaning	Proof
1	$0 \leq \delta_i \leq 1$	Bounds the possible values $\delta_i$ can assume	§ 8.1
2	$\delta_i = 0$	If $Y$ is independent of $X_i$ then $\delta = 0$	Main body
3	$\delta_{1,2,\dots,n} = 1$	The importance of all parameters equals unity	§ 8.2
4	$\delta_{ij} = \delta_i$	If $Y$ is dependent on $X_i$ but independent of $X_j$ then $\delta_{ij} = \delta_i$	Footnote
5	$\delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j i}$	Bounds the possible values $\delta_{ij}$ can assume	§ 8.3

Table 1: Properties of the Uncertainty Importance Measure introduced in this work.

parameters. This is done as follows.

**Definition 2** Let  $\underline{R} = (X_{i_1}, X_{i_2}, \dots, X_{i_r})$  be any group of parameters. Then:

$$\begin{aligned} \delta_{i_1, i_2, \dots, i_r} &= \frac{1}{2} E_{\underline{R}}[s(\underline{R})] = \\ &= \int f_{X_{i_1}, X_{i_2}, \dots, X_{i_r}}(x_{i_1}, x_{i_2}, \dots, x_{i_r}) \left[ \int \left| f_Y(y) - f_{Y|X_{i_1}, X_{i_2}, \dots, X_{i_r}}(y) \right| dy \right] dx_{i_1} dx_{i_2} \dots dx_{i_r} \end{aligned} \quad (21)$$

where

$$f_{X_{i_1}, X_{i_2}, \dots, X_{i_r}}(x_{i_1}, x_{i_2}, \dots, x_{i_r}) = \int \dots \int f_{\underline{X}}(\underline{x}) \prod_{k \neq i_1, i_2, \dots, i_r} dx_k \quad (22)$$

The above definition then enables to prove Property 3 in Table 1 which states that

$$\delta_{1,2,\dots,n} = 1 \quad (23)$$

*i.e.* the joint importance of all parameters equals unity (for the proof, see Section 8.2 in the Appendix.)

One can summarize these three properties as follows. The  $\delta$  of an individual parameter or of a group can assume values between 0 and 1. It will equal 0 when  $Y$  is independent of the parameter or group of parameters at hand. It will equal 1 when the group including all inputs is considered.

A couple of remarks. As far as correlations are concerned, we observe that Definitions 1 and 2 hold independently of whether the parameters are correlated. In fact, eqs. (19) and

(21) require the specification of the joint density,  $f_{\underline{X}}(\underline{x})$ , without reference to the eventual independence of the parameters.

Let us now study the interpretation of Definitions 1 and 2 in terms of sensitivity analysis settings. One can see that Definition 1 resembles Setting 1 of Saltelli and Tarantola (2002) with a main similarity and a main difference. The similarity is that both settings involve conditioning with respect to  $X_i = x_i^*$ . The difference lies in the fact that Setting 1 of Saltelli and Tarantola (2002) [41] looks for the parameters that achieve the greatest reduction in the variance of  $Y$ , while the setting implied by Definition 1 is the identification of the parameters that influence the entire distribution the most. Similarly, Definition 2 parallels Setting 2 of Saltelli and Tarantola (2002) [41] insofar groups are concerned. However, we note that Setting 2 of Saltelli and Tarantola (2002) [41] again refers to variance reduction, while Definition 2 concerns influence with respect to the entire distribution.

We now discuss the computation of  $\delta$  for a simple example, with the purpose of illustrating its definition.

**Example 1** *Suppose that the unconditional density of  $Y = g(\underline{X})$  is:*

$$f_Y(y) = \text{Beta}(y; 1, 3) \quad (24)$$

*Suppose further that one of the parameters,  $X_i$  is a discrete random variable than can assume four values, namely,  $x_i^1, x_i^2, x_i^3, x_i^4$  with  $P(X_i = x_i^m) = 1/4$  ( $m = 1..4$ ). Suppose that either analytically or numerically, You are able to obtain the 4 conditional distributions of  $Y$  given that  $X_i = x_i^m$  and that they are as follows:*

$$f_{Y|X_i=x_i^m}(y) = \begin{cases} f_{Y|X_i=x_i^1}(y) = \text{Beta}(y; 2, 3) \\ f_{Y|X_i=x_i^2}(y) = \text{Beta}(y; 5, 3) \\ f_{Y|X_i=x_i^3}(y) = \text{Beta}(y; 7, 3) \\ f_{Y|X_i=x_i^4}(y) = \text{Beta}(y; 9, 3) \end{cases} \quad (25)$$

*Figure 4 shows the unconditional and conditional distributions named above. Let us compute*

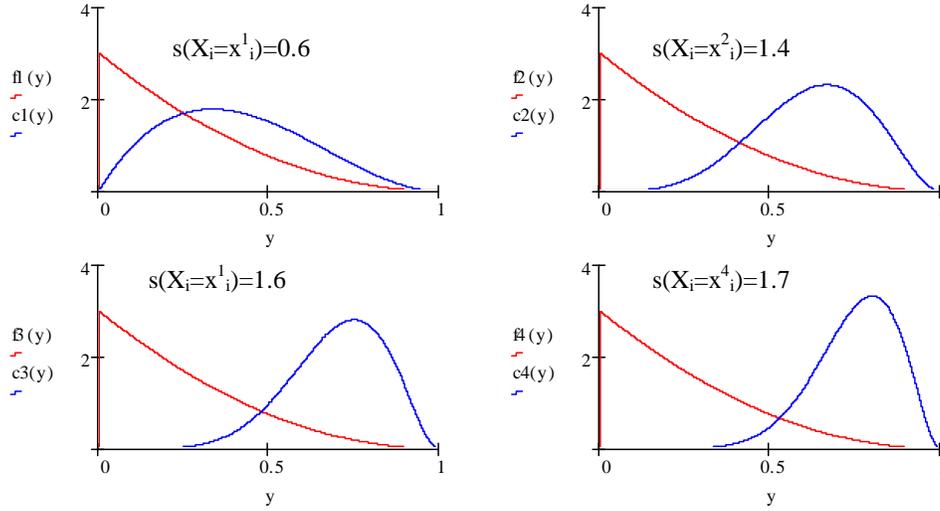


Figure 4:  $f_Y(y)$ ,  $f_{Y|X_i=x_i^*}(y)$  and  $s(x_i)$  for the example.

$\delta_i$ . For each of the conditional distributions we have [eq. (17)]:

$$s(x_i^m) = \int |f_Y(y) - f_{Y|X_i=x_i^m}(y)| dy, \quad m = 1 \dots 4 \quad (26)$$

For  $X_i = x_i^1$ , we have:

$$s(x_i^1) = \int_0^1 |\beta(y, 1, 3) - \beta(y, 2, 3)| dy = 0.6 \quad (27)$$

$s(x_i^2)$ ,  $s(x_i^3)$ ,  $s(x_i^4)$  are computed in a similar fashion (Figure 4.) The resulting value of  $\delta_i$  is found as  $\delta_i = \frac{1}{2} \left[ \frac{1}{4}s(x_i^1) + \frac{1}{4}s(x_i^2) + \frac{1}{4}s(x_i^3) + \frac{1}{4}s(x_i^4) \right] = 0.75$ .

Definitions 1, 2 and the example can be utilized to indicate a possible algorithm for the numerical computation of  $\delta_i$ . Preliminary step is an uncertainty propagation leading to the determination of the unconditional density of  $Y$ ; second step is the sampling of a value of  $x_i$  from  $f_{X_i}(x_i)$ ; third step is the sampling of the conditional distribution of  $Y$  given  $X_i$ , i.e.  $f_{Y|X_i}(y)$ ; fourth step is the computation of  $s(X_i)$ ; fifth step the estimation of  $\delta_i$  from the computed  $s(x_i)$ 's.

In the remainder of this Section, we detail some observation on properties of the importance of parameter groups ( $\delta_R$ , Definition 2). To do so, we begin with groups of two

parameters,  $\underline{R} = (X_i, X_j)$ .

According to Definition 2, the delta of  $X_i$  and  $X_j$  is given by:

$$\delta_{ij} = \frac{1}{2} E_{X_i X_j} [s(X_i, X_j)] \quad (28)$$

where

$$s(X_i, X_j) = \int \left| f_Y(y) - f_{Y|X_i X_j}(y) \right| dy \quad (29)$$

is the shift obtained fixing  $X_i$  at  $x_i$  and  $X_j$  at  $x_j$ . Based on the above discussion, it is immediate to observe that, if  $Y$  is independent of  $X_j$ , then (Property 4, Table 1):

$$\delta_{ij} = \delta_i \quad (30)$$

Eq. (30) simply re-states the fact that no contribution to model uncertainty comes from  $X_j$  if  $Y$  does not depend upon it<sup>2</sup>.

However, if there is a contribution to uncertainty coming from  $X_j$ , one would expect  $\delta_{ij}$  to increase. Indeed, let us think of  $s(X_i, X_j)$  as obtained in two steps. The first step is given by fixing  $X_i$  at  $x_i$  and the second step is obtained by then fixing of  $X_j$  at  $x_j$  (Figure 5). We limit ourselves to an intuitive explanation; a formal treatment is offered in Section 8.3. Note that from a mathematical viewpoint  $\delta$  shares the properties of a distance (see Malliavin (1995) [26] for definition of distance). Hence,  $\delta_i$  represents the expected distance between the density of  $Y$  and the conditional density of  $Y$  given  $X_i$ . Similarly,  $\delta_{ij}$  is the distance between the density of  $Y$  and the conditional density of  $Y$  given  $X_i$  and  $X_j$ . Now, geometrically, going from  $f_Y(y)$  to  $f_{Y|X_i X_j}(y)$  through  $f_{Y|X_i}(y)$  is the same as moving from point A to point C but first going through point B. The length of path AB-BC is greater than the length of AC, unless the three points lie on the same line.

With this in mind, one can write:

$$f_Y(y) - f_{Y|X_i X_j}(y) = [f_Y(y) - f_{Y|X_i}(y)] + [f_{Y|X_i}(y) - f_{Y|X_i X_j}(y)] \quad (31)$$

---

<sup>2</sup>In fact, if  $Y$  is independent of  $X_j$ , then  $\left| f_Y(y) - f_{Y|X_i X_j}(y) \right| = \left| f_Y(y) - f_{Y|X_i}(y) \right|$  that leads to eq. (30) by definition of  $\delta$ .

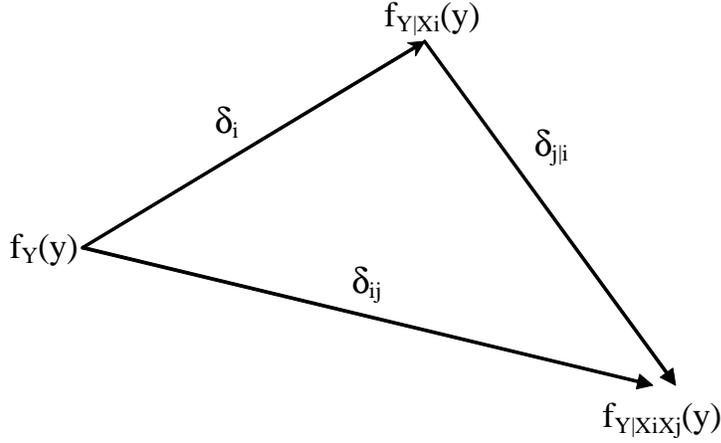


Figure 5: Geometric Interpretation of Property nr. 5.

and interpret  $[f_Y(y) - f_{Y|X_i}(y)]$  (first difference in the right hand side of eq. (31)) as the difference between the unconditional density of  $Y$  and the conditional density found fixing  $X_i$  and  $f_{Y|X_i}(y) - f_{Y|X_i X_j}(y)$  as the residual difference between  $f_{Y|X_i}(y)$  and  $f_{Y|X_i X_j}(y)$  obtained fixing  $X_j$  after  $X_i$  has been fixed. Now, defining the conditional  $\delta$  for the second step as

$$\delta_{j|i} = \frac{1}{2} E_{X_i X_j} \left[ \int |f_{Y|X_i}(y) - f_{Y|X_i X_j}(y)| dy \right] \quad (32)$$

Noting  $\delta_{j|i} \geq 0$  and that  $\delta_{j|i} = 0$ , if  $Y$  is independent of  $X_j$ , it is possible to see that (Appendix, Section 8.3):

$$\delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j|i} \quad (33)$$

which states that the joint importance of  $X_i$  and  $X_j$  is greater than the individual importance of  $X_i$ , but limited by the importance of the residual term  $\delta_{j|i}$ .

Suppose now that the observed shift in uncertainty due to  $X_j$  is always independent of the value assumed by  $X_i$ . In that case, one would expect  $\delta_{j|i} = \delta_j$ . If it happens that

$$\delta_{ij} = \delta_j + \delta_i \quad (34)$$

*i.e.* the three points lie on the same line, we say that the effects of the uncertainty in  $X_i$  and

$X_j$  on  $f_Y(y)$  are separable.

The next Section describes the application of  $\delta$  to the global sensitivity analysis of the Ishigami test function, highlights the computational aspects in greater detail and illustrates a first comparison of  $\delta$  with the previously introduced importance measures listed in Section 2.

#### 4 A Test Function Analysis

This Section describes the numerical computation of  $\delta$  and the comparison with variance-based techniques and the *CHT* indicator by studying the application of  $\delta$  to the Ishigami test function (Chun *et al* (2000) [10]). The mathematical expression of the function is:

$$Y = g(\underline{X}) = \sin X_1 + a \sin^2 X_2 + bX_3^4 \sin X_1 \quad (35)$$

and the  $X_i$  are assumed independent and uniformly distributed between  $-\pi$  and  $\pi$ . The input distributions, the sample size ( $N = 1000$ ) and the values of the constants  $a$  and  $b$  (5 and 0.1 respectively) are the same as in Chun *et al* (2000) [10] to allow for a direct comparison.

Let us first discuss the uncertainty analysis of  $Y$ . Uncertainty propagation (Figure 6) produces  $f_Y(y)$  best fitted by a logistic density, with a Kolmogorov-Smirnov statistics equal to 0.02.

We now describe the computation of  $\delta(X_1)$ . One first generates a value for  $X_1$ , namely  $x_1^1$  sampling from  $f_{X_1}(x_1)$ . In our first generation, we get  $x_1^1 = 1.029$ . Given this value, the conditional density of  $Y$  is obtained by propagating uncertainty in the model keeping  $X_1 = x_1^1$ . The resulting density,  $f_{Y|X_1=x_1^1}(y)$ , is shown in Figure 7.

$f_{Y|X_1=1.029}(y)$  is now fitted by a beta distribution (the parameters are illustrated in Figure 7) with a Kolmogorov-Smirnov statistics of 0.07.  $s(x_1^1)$  is, then, computed from a simple numerical integration of the absolute value of the difference between the unconditional (Figure 6) and the conditional density (Figure 7). In this case, it turns out that  $s(x_1^1) = 0.638$ . The next step is to repeat the procedure to produce  $x_1^2$  (equal to 2.84 in our second run), determine the new conditional density of  $Y$  and compute  $s(x_1^2)$ , which, in this case, turns

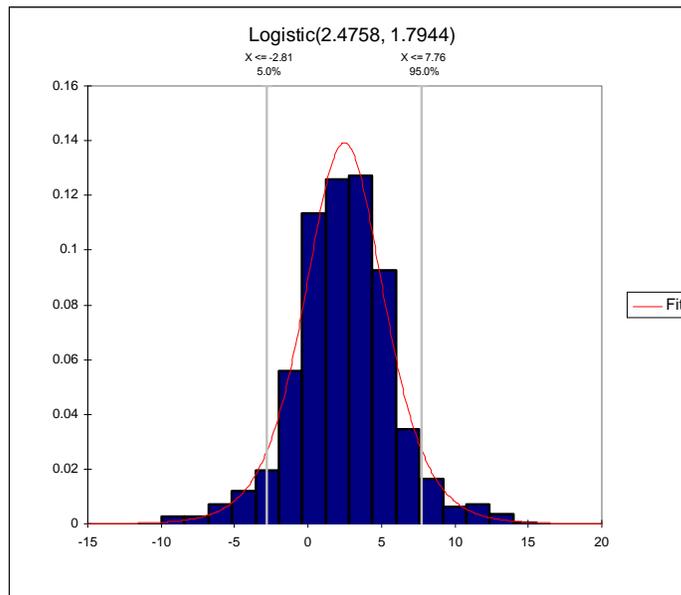


Figure 6: The uncertainty propagation in the example results in a logistic distribution for  $f_Y(y)$ .

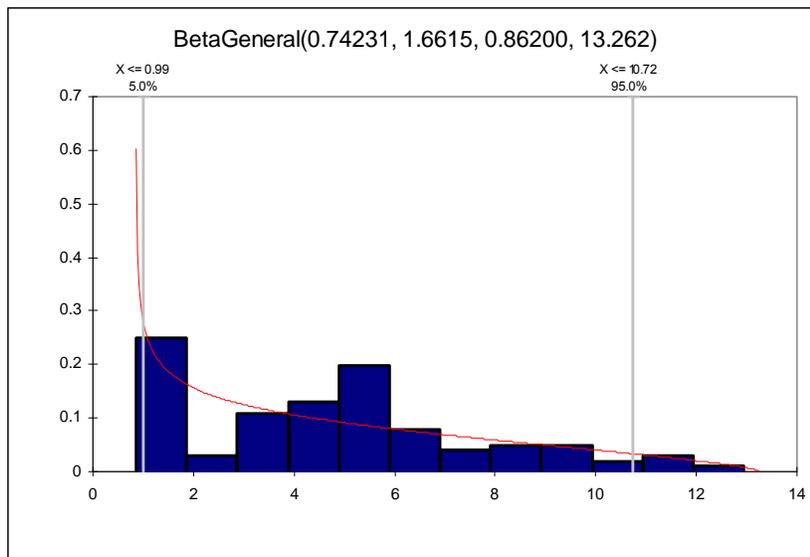


Figure 7: Conditional density given  $X_1=1.0219$ ,  $f_{Y|X_1=1.029}(y)$ .

<b>Parameter</b>	$\delta_i$	$S1_i$	$ST_i$	$CHT_i$
$X_1$	0.33	0.39	0.62	0.66
$X_2$	0.39	0.29	0.34	0.98
$X_3$	0.28	0.02	0.28	0.67

Table 2: Parameter global importance measured with delta.

out to be equal to 0.58. Repeating this steps for a 1000 times,  $\delta_i$  is then estimated to be  $\delta_i = 0.33$ .

Proceeding in a similar fashion for  $X_2$  and  $X_3$ , the moment independent indicators for  $X_2$  and  $X_3$  are computed. The results are reported in Table 2.

Table 2 shows that  $X_2$  is the most influential parameter, followed by  $X_1$  and  $X_3$ . As far as interactions are concerned, one can observe also that  $\delta_1 + \delta_2 + \delta_3 \cong 1$ , in this case. Recalling Property 3 in Table 1, then it holds that  $\delta_{123} = \delta_1 + \delta_2 + \delta_3$ , *i.e.*, using the terminology introduced in Section 3, eq. (34), the effects of uncertainty in the parameters on the uncertainty in  $Y$  are separable — for this model and for the given input distributions.

We now discuss the comparison of the above results to the ones obtained for the  $CHT_i$  indicator and for variance-based techniques,  $S1_i$  and  $ST_i$  (Table 2).

The fourth column of Table 2 shows the total sensitivity indices of the three parameters estimated with the Sobol' method, utilizing the software SIMLAB [44]. The third column shows the result for the  $S1_i$  indicator [eq. (1)] and the fifth column for the  $CHT$  indicator [eq. (13)], as reported in Chun *et al* (2000)<sup>3</sup> [10]. We note that  $ST_i$  and  $S1_i$  produce the same ranking, while  $\delta_i$  and  $CHT_i$  produce different ranking w.r.t. the other indicators (see also Figure 8).

The different ranking between  $ST_i/S1_i$  and  $\delta_i/CHT_i$  is explained by the fact that  $ST_i$  and  $S1_i$  are variance-based, while  $\delta_i$  and  $CHT_i$  are moment independent. This result confirms

<sup>3</sup>Note that  $IH_i = S1_i \cdot V[Y]$ ; hence the ranking obtained with  $S1_i$  is the same as the ranking obtained with  $IH_i$ . Such ranking is reported in Chun et al (2000).

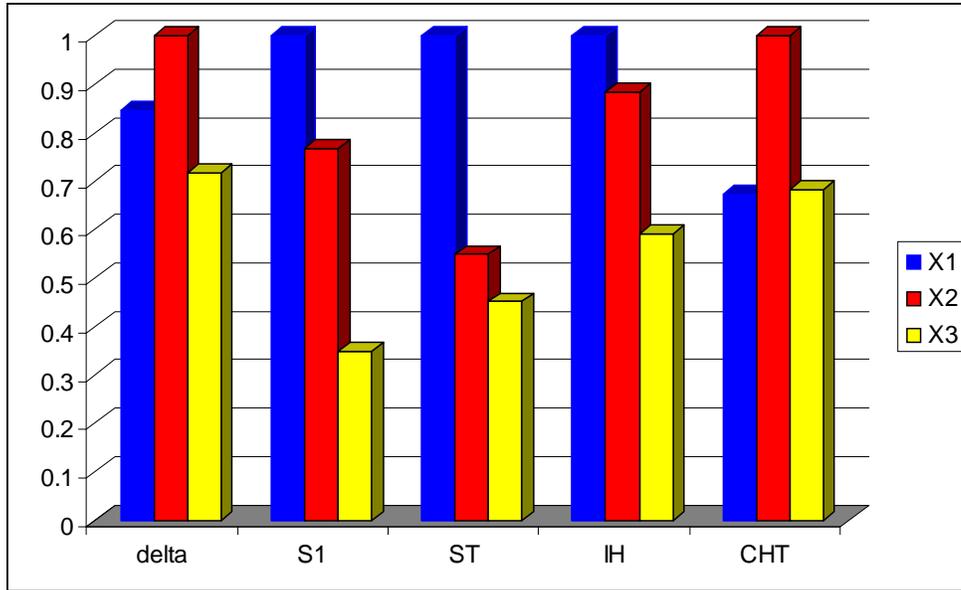


Figure 8: Parameter ranking comparison.

that a parameter which influences variance the most is not necessarily the parameter that influences the output distribution the most. The difference between the ranking produced by  $\delta_i$  and  $CHT_i$  can be explained as follows.  $CHT_i$  results in Table 2 are the importance of the parameters when uncertainty in each of them, one at a time, is completely eliminated (Table 3). Hence,  $CHT_i$  ranks inputs given that an hypothesized change in the state of knowledge of the decision maker happens. On the other hand,  $\delta_i$  represents the importance of the entire distribution of  $X_i$  w.r.t. the entire distribution of  $Y$ , given the current state of knowledge and without considering an artificially hypothesized change.

Finally, given the above discussion, comparing the ranks obtained with the four indicators in Table 2 enables one to conclude that  $X_2$  is the most important parameter when the entire output distribution is considered (it ranks first with both  $\delta$  and  $CHT$ ), while  $X_1$  is the most important parameter in explaining the variance of the model output.

Parameter	Distribution	Mean	Error Factor
$X_1$	lognormal	2	2
$X_2$	lognormal	3	2
$X_3$	lognormal	$1 \times 10^{-3}$	2
$X_4$	lognormal	$2 \times 10^{-3}$	2
$X_5$	lognormal	$4 \times 10^{-3}$	2
$X_6$	lognormal	$5 \times 10^{-3}$	2
$X_7$	lognormal	$3 \times 10^{-3}$	2

Table 3: Distributions for the parameters of the PRA model of Iman (1987).

## 5 Application to a Probabilistic Risk Assessment Model

The purpose of this Section is to illustrate the application of  $\delta$  to the probabilistic risk assessment (PRA) model utilized in Iman (1987) [22] where uncertainty importance measures were first introduced. The model has also been utilized in Chun *et al* (2000) (see [10] for a comparison of the *CHT* indicator with the Iman-Hora importance measure.) Besides computing the  $\delta$  for the model parameters, we also estimate the first and total order global sensitivity indices, to highlight the differences between Sobol' interactions and  $\delta$ -interactions ( $\delta_{j|i\cdot}$ )

The probability of the top event is written as (Iman (1987) [22]):

$$\begin{aligned}
Y_{Top} = & X_1X_3X_5 + X_1X_3X_6 + X_1X_4X_5 + X_1X_4X_6 + X_2X_3X_4 + X_2X_3X_5 + \\
& + X_2X_4X_5 + X_2X_5X_6 + X_2X_4X_7 + X_2X_6X_7
\end{aligned} \tag{36}$$

The numerical values of the input distributions utilized in this analysis are the same as the ones used in Chun *et al* (2000) [10] and are presented in Table 3.

The result of uncertainty propagation with a sample of size  $N = 1000$  are displayed in Figure 9. Figure 9 shows that  $f_{Y_{Top}}(y)$  is lognormal, with mean equal to  $2E - 6$  and error factor equal to 2.4. The calculation of the  $\delta$  importance measure for the parameters has been

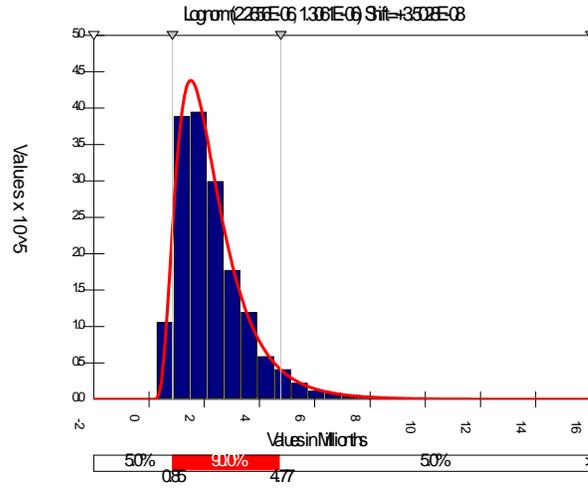


Figure 9:  $f_{Y_{Top}}(y)$  as a result of uncertainty analysis.

Parameter	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
$\delta_i$	0.11	0.17	0.09	0.13	0.18	0.20	0.11

Table 4: Delta for the parameters of the PRA model of Iman (1987).

performed in accordance with the computation algorithm proposed in Section 3. We have found the results of Table 4.

Table 4 shows that  $X_6$  is the most relevant parameter, followed by  $X_5$ ,  $X_2$ ,  $X_4$ ,  $X_7$ ,  $X_1$  and  $X_3$ . Thus, we can say that  $X_6$  is the most influential parameter on the top event while  $X_3$  is the least influential.

We then compare the above results to the ones obtained by making use of the Sobol' total effects, the  $CHT_i$  and  $S1_i$  indicators. Table 5 shows the results. The ranking obtained with  $CHT_i$  have been computed in Chun *et al* (2000) [10] and are as such reported in Table 5, while  $S1_i$  and  $ST_i$  have been computed with the Software SIMLAB [44].

To analyze the agreement among the ranking obtained with the four importance measures, we computed the Savage Score Correlation Coefficients (SSCC) (first introduced in Iman and Conover (1987) [23]; for an illustration see also Campolongo and Saltelli (1997) [8]) on the ranking in Table 5. The result is reported in Table 6.

<b>Parameter</b>	$\delta_i$	$S1_i$	$ST_i$	$CHT_i$
$X_1$	6	6	5	6
$X_2$	3	1	1	1
$X_3$	7	7	7	7
$X_4$	4	4	4	4
$X_5$	2	2	3	2
$X_6$	1	3	2	3
$X_7$	5	5	6	5

Table 5: Parameter Ranking with the four uncertainty importance measures.

	$\delta_i$	$S1_i$	$ST_i$	$CHT_i$
$\delta_i$	1	0.49	0.59	0.49
$ST_i$	0.59	0.93	1	0.93
$CHT_i$	0.49	1	0.93	1
$S1_i$	0.49	1	0.93	1

Table 6: SSCC matrix for the four importance measures used in this Section.

One notes that the ranking of  $CHT_i$  coincides with that of  $S1_i$ , while it differs from both the  $\delta_i$  and the  $ST_i$  ranking. Table 6 also shows that  $\delta_i$  results are in a higher agreement with  $ST_i$  results then with  $CHT_i$  and  $S1_i$ .

We now turn our attention to a more detailed comparison of the global sensitivity indices and  $\delta$ . We start with the ranking obtained with  $ST_i$  and  $\delta_i$ . One notes that both indicators agree in the identification of the least relevant parameters:  $X_3$  ranks 7<sup>th</sup> according to both indicators,  $X_1$  6<sup>th</sup>,  $X_7$  5<sup>th</sup> and  $X_4$  4<sup>th</sup>. Some difference in the agreement is found on the ranking of the most relevant parameters:  $X_6$  ranks first according to  $\delta_i$ , while it ranks second according to  $ST_i$ ;  $X_5$  ranks 2<sup>nd</sup> according to  $\delta_i$ , while it ranks 3<sup>rd</sup> according to  $ST_i$ ;  $X_2$  ranks 3<sup>nd</sup> according to  $\delta_i$ , while it ranks 1<sup>st</sup> according to  $ST_i$ . Since the ranking difference ought to be attributed to the different meaning of the importance indicators, the following summary of the result becomes natural: *i*)  $X_3$ ,  $X_1$ ,  $X_7$  and  $X_4$  are non-relevant on the model uncertainty, both when the entire distribution ( $\delta_i$ ) or its variance ( $ST_i$ ) are considered; *ii*) the most relevant parameter w.r.t. the entire output distribution is  $X_6$ , while the most relevant w.r.t. the output variance is  $X_2$ .

We then discuss how  $\delta$  and global sensitivity indices interpret interactions. From Table 4 one notes that  $\sum_{i=1}^7 \delta_i = 0.99$ . Recalling that  $\delta_{1,2,\dots,7} = 1$ , we have that  $\delta_{1,2,\dots,7} \simeq \sum_{i=1}^7 \delta_i$ , *i.e.* interactions play a minor role according to  $\delta$ . Using the terminology introduced in Section 3, we could say that the effect of uncertainty in the  $X_i$  on uncertainty in the top event are separable. Let us now examine the relevance of interactions that is revealed by Sobol' variance decomposition. Figure 10 shows the comparison between the total effects of each of the parameters and the first order indices.

From Figure 10 one notes that the percentage of interaction terms in each of the parameter importance is not as relevant. In fact,  $\sum_{i=1}^7 S1_i = 94\%$ , indicating that almost all of the model variance is explained by individual effects. Hence, results of Sobol' indices show that the model responds additively to the input uncertainty. Thus, in this case both separability and additivity play a role. We then performed additional calculations to verify whether the

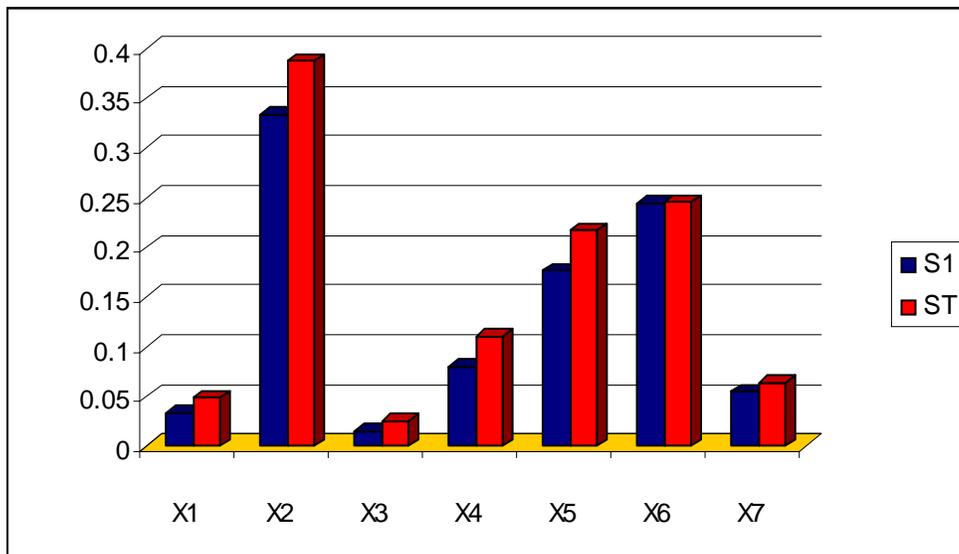


Figure 10: The comparison of first order (S1) and total order (ST) sensitivity indices shows the low relevance of interactions for the model at hand when the error factor equals 2.

above conclusion on interactions was robust w.r.t. the choice of the input distributions. We increased all error factors in Table 3 from 2 to 6. Results now show that given the new distributions  $\sum_{i=1}^7 S1_i = 57\%$ , signalling that a significant portion of the model variance is now explained by interaction terms. On the other hand,  $\sum_{i=1}^7 \delta_i = 86\%$ , implying that the output uncertainty is mainly attributable to individual parameter contributions, although it is not completely separable, as in the previous case.

## 6 Computational Aspects: an overview of current methods and opportunity for future research

Although the primary purpose of this work is to introduce the definition, properties and meaning of  $\delta$ , let us touch upon computational aspects of  $\delta$ . The computational cost of a technique is defined in terms of number of model runs necessary to estimate the sensitivity measure.

The estimation of  $\delta$  for the above mentioned models did not pose any particular numerical issues. However, the cost for estimating  $\delta_i$  using the algorithm proposed in Section 3 is equal

to  $nN_oN_i$  where  $n$  is the number of parameters,  $N_o$  is the number of runs necessary for the outer integration and  $N_i$  the number of integrals necessary for the internal integration. Thus for computationally intensive models or models requiring sample sizes of  $N > 1000$ <sup>4</sup>, the estimation of  $\delta_i$  can raise the “*curse of dimensionality* (Rabitz and Alis (1999) [35])” problem which affects most of global SA techniques (as an example, see the investigations of Frey and Patil (2002) [11] and Patil and Frey (2004) [32]). The trade-off between computational complexity and amount of uncertainty information delivered by an SA technique is being and has been extensively dealt with in the literature. For example Rabitz and Alis (1999) report that the cost of computing all the sensitivity indices in Sobol’ decomposition is equal to  $N \times \sum_{i=0}^n \frac{n!}{(n-i)!i!}$ , which grows exponentially with the number of parameters. After the works of Homma and Saltelli (1996) [18], at least three approaches have been envisioned to ease the estimation of the global sensitivity indices: *i*) Saltelli *et al* (1999) [39] proposed the Extended FAST, which enables one to compute the first order and the total order indices at a cost equal to  $N$ , with a gain of  $\sum_{i=0}^n \frac{n!}{(n-i)!i!}$  model runs; *ii*) Rabitz and Alis (1999) [35] and Alis and Rabitz (2001) [1] propose a two step approach based on finite difference decomposition; *iii*) Oakley and O’Hagan (2004) [31] demonstrate that further savings can be obtained if one adopts a Bayesian approach.

Besides the estimation of variance based indicators, several authors have dealt with the problem of increasing the efficacy of sampling methods in uncertainty propagation for computationally intensive models. Some examples are: *i*) Sobol’ quasi-random sequence generator (Sobol’ (1990) [45]) applied by Homma and Saltelli (1995) [17] in the computation of the Iman-Hora importance measure of eq. (1); *ii*) Latin Hypercube sampling, first introduced in McKay *et al* (1979) [28], and thoroughly discussed in Helton and Davis (2000) [15] and Helton and Davis (2003) [16].

Another way of circumventing the curse of dimensionality is to make use of screening methods. Screening methods are sensitivity analysis tools that enable to identify non-relevant

---

<sup>4</sup>This sample size is also used in Rabitz and Alis (1999) as a threshold for numerical complexity.

parameters and therefore to eliminate from the analysis variables that do not deserve further attention. We refer the reader to the methods of Morris (1991) [29] and the ones described in Kleijnen (2005) [21].

As far as the estimation of  $\delta$  is concerned, one can think of utilizing combinations of techniques to reduce either  $n$  or  $N_i/N_o$  or both. A first way is utilizing the algorithm used in this work together with a sampling method (Latin Hypercube Sampling or Sobol' quasi-random  $lp\tau$ ); this should allow to maintain  $N_i$  and  $N_o$  at their lowest size. A second way is applying a two-step method *a la* Alis and Rabitz (2001) [1]; this would reduce  $n$ . A third way is adopting a Bayesian approach *a la* Oakley and O'Hagan (2004) [31]; this would be effective in reducing  $N_iN_o$ . A fourth way is utilizing screening methods first to screen out non relevant inputs and then applying a full-fledged estimation of delta, eventually with an appropriate sampling method: in this case both  $n$  and  $N_i$  and  $N_o$  would be reduced. The refinement of computational strategies for  $\delta$  shall be the subject of future research by the author.

## 7 Conclusions

When uncertainty in model parameters is present, the problem of assessing which of the inputs influences output uncertainty the most is properly addressed by global SA.

The most recent literature development has assisted to the refinement and establishment of the theoretical and computational framework of variance-based techniques. We have seen that variance decomposition reflects model structure when the inputs are uncorrelated and provides guidance in data collection when an analyst wants to achieve a pre-determined variance reduction (even when parameters are correlated). However, in terms of uncertainty analysis, a limitation of a variance-based global SA is the fact that variance is just one of the moments of the output distribution and, as such, cannot be elected as representative of the whole decision-maker state of knowledge. In addition, when parameters are correlated, the direct relationship between variance and model structure does not hold.

In this work, we have addressed these issues by introducing a moment independent uncer-

tainty indicator ( $\delta$ ) that looks at the entire input/output distribution and whose definition is well posed also in the presence of correlations among the parameters. We have discussed the mathematical properties of  $\delta$ . We have seen that it is always between 0 and 1, it equals 0 if the output is not dependent upon an input, it is readily defined for parameter groups and it equals unity if the group of all inputs is considered. We have seen that its definition is well posed in the presence of correlations among the parameters, since one needs to specify a joint distribution of the inputs without requiring independence. We have also shown that the indicator does not presuppose a sensitivity case, *i.e.* a change in the decision-maker uncertainty, but reflects the current analyst/decision-maker state of knowledge.

We have illustrated the numerical aspects of the computation of  $\delta$ . We have compared its results to those of first and total order sensitivity indices, the Chun-Han-Tak and the Iman-Hora indicators by studying the application of these techniques to the Ishigami test function.

We have then discussed the application of the above techniques to the PRA model analyzed in Chun *et al* (2000) [10] and introduced in Iman (1987) [22].

Results of both applications show that variance-based indicators and  $\delta$  agree in identifying the less relevant parameters w.r.t. the output uncertainty. Discrepancies in ranking between the relevant parameters reveal that factors influencing variance the most are not necessarily the ones that influence the entire output distribution the most.

In summary, the analysis has shown that if one utilizes the moment independent importance measure introduced in this work one gains insights on which of the parameters influence uncertainty the most. Utilizing the new measure jointly with variance-based indicators would also enable the analyst to obtain insights on the parameters that achieve the greatest variance reduction and, when parameters are independent, on the model structure and interactions.

The work also paves the way to further research. The first line of research concerns the selection of the appropriate computational algorithm in the estimation of  $\delta$  for numerically

intensive models (see the discussion on alternative approaches illustrated in Section 6.) A second line of research is represented by exploring the conditions on model structure and input distributions under which separability holds (*i.e.*  $\delta_{1,2,\dots,n} = \delta_1 + \delta_2 + \dots + \delta_n$ ) and whether, under the same conditions, additivity holds (*i.e.*  $(V = \sum_{i=1}^n V_i)$ ).

## 8 Appendices

### 8.1 Proof of Property 1

**Proof.** We prove first that  $\delta_i \leq 1$ . By the triangle inequality,

$$|f_Y(y) - f_{Y|X_i}(y)| \leq |f_Y(y)| + |f_{Y|X_i}(y)| \quad (37)$$

Hence

$$\int |f_Y(y) - f_{Y|X_i}(y)| dy \leq \int |f_Y(y)| dy + \int |f_{Y|X_i}(y)| dy \quad (38)$$

Since:  $\int |f_Y(y)| dy = 1$  and  $\int |f_{Y|X_i}(y)| dy = 1$ , we have:

$$\int |f_Y(y) - f_{Y|X_i}(y)| dy \leq 2 \quad (39)$$

Hence:

$$E_{X_i}[s(X_i)] = E_{X_i}[\int |f_Y(y) - f_{Y|X_i}(y)| dy] \leq E_{X_i}[2] = 2 \quad (40)$$

Applying definition 2 [eq. (19)] there follows that  $\delta_i \leq 1$ .

The fact that  $\delta_i \geq 0$  follows from the monotonicity property of integrals, since  $|f_Y(y) - f_{Y|X_i}(y)| \geq 0$ . ■

### 8.2 Proof of Property 3

**Proof.** We prove that  $\delta_{1,2,\dots,n} = 1$ . By definition, when  $\underline{X}$  is fixed at  $\underline{X}^*$ ,  $y^* = g(\underline{X}^*)$  and  $P(Y = y^*) = 1$ . That is  $f_{Y|\underline{X}=\underline{X}^*}(y)$  is a *delta - Dirac* measure on  $y^*$ . Consider then a finite but small interval around  $y^*$  and write (Figure 11):

$$\tilde{f}_{Y|\underline{X}=\underline{X}^*}(y; y_1, y_2) = \begin{cases} \frac{1}{y_2 - y_1} & \text{if } y_1 < y < y_2 \\ 0 & \text{otherwise} \end{cases} \quad (41)$$

Note that  $f_{Y|\underline{X}=\underline{X}^*}(y) = \lim_{y_1 \rightarrow y_2} \tilde{f}_{Y|\underline{X}=\underline{X}^*}(y; y_1, y_2)$ . Now, consider still a small interval

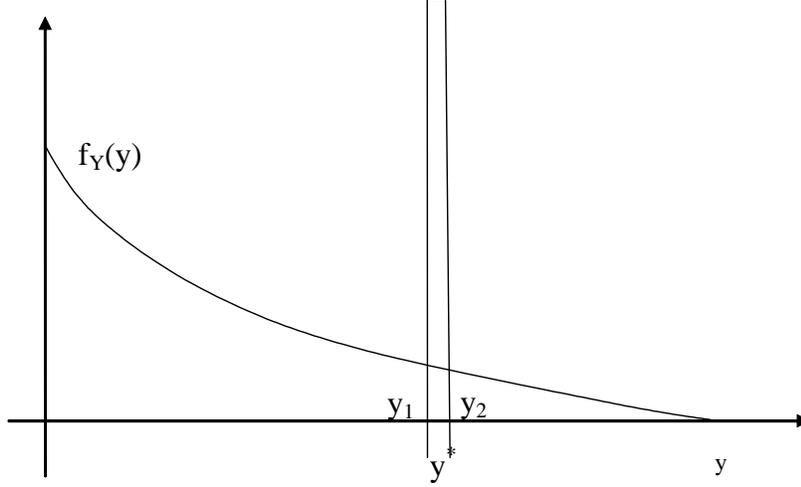


Figure 11: Unconditional density and conditional density when all the parameters are in a small interval around  $\underline{X}^*$ .

around  $y^*$  and compute the total shift when  $y$  belongs to such small interval around  $y^*$ :

$$s(\underline{X}) = \lim_{y_1 \rightarrow y_2} \int \left| f_Y(y) - \tilde{f}_{Y|\underline{X}}(y; y_1, y_2) \right| dy \quad (42)$$

Applying eq. (41) one gets:

$$s(\underline{X}) = \lim_{y_1 \rightarrow y_2} \int_{-\infty}^{y_1} |f_Y(y)| dy + \int_{y_1}^{y_2} \left| f_Y(y) - \frac{1}{y_1 - y_2} \right| dy + \int_{y_2}^{\infty} |f_Y(y)| dy \quad (43)$$

which is equivalent to:

$$s(\underline{X}) = \lim_{y_1 \rightarrow y_2} \int_{-\infty}^{y_1} |f_Y(y)| dy + \int_{y_2}^{\infty} |f_Y(y)| dy + \int_{y_1}^{y_2} \left| f_Y(y) - \frac{1}{y_1 - y_2} \right| dy \quad (44)$$

Noting that:

$$\lim_{y_1 \rightarrow y_2} \int_{-\infty}^{y_1} |f_Y(y)| dy + \int_{y_2}^{\infty} |f_Y(y)| dy = \int_{-\infty}^{\infty} f_Y(y) dy = 1 \quad (45)$$

and that  $f_Y(y)dy \rightarrow 0$  as  $y_1 \rightarrow y_2$ , we have:

$$\lim_{y_1 \rightarrow y_2} \int_{y_1}^{y_2} \left| f_Y(y) - \frac{1}{y_1 - y_2} \right| dy = \lim_{y_1 \rightarrow y_2} \int_{y_1}^{y_2} \left| \frac{1}{y_2 - y_1} \right| dy = 1 \quad (46)$$

Substituting back into eq. (44) one finds:

$$s(\underline{X}) = 1 + 1 = 2 \quad (47)$$

There follows that

$$\delta_{1,2,\dots,n} = \frac{1}{2}E_{\underline{X}}[s(\underline{X})] = \frac{1}{2}E_{\underline{X}}[2] = 1 \quad (48)$$

q.e.d.. ■

### 8.3 Proof of Property 5

**Proof.** Note that:

$$\left| f_Y(y) - f_{Y|X_i X_j}(y) \right| \leq \left| f_Y(y) - f_{Y|X_i}(y) \right| + \left| f_{Y|X_i}(y) - f_{Y|X_i X_j}(y) \right| \quad (49)$$

Taking the integral of both sides:

$$\int \left| f_Y(y) - f_{Y|X_i X_j}(y) \right| dy \leq \int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy + \int \left| f_{Y|X_i}(y) - f_{Y|X_i X_j}(y) \right| dy \quad (50)$$

Now, one can take the expectation, to get:

$$E_{X_i X_j} \left[ \int \left| f_Y(y) - f_{Y|X_i X_j}(y) \right| dy \right] \leq E_{X_i X_j} \left[ \int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy \right] + E_{X_i X_j} \left[ \int \left| f_{Y|X_i}(y) - f_{Y|X_i X_j}(y) \right| dy \right] \quad (51)$$

Since  $f_{Y|X_i}(y)$  depends on  $X_i$  and not on  $X_j$ , it is true that:

$$E_{X_i X_j} \left[ \int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy \right] = E_{X_i} \left[ \int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy \right] = 2\delta(X_i) \quad (52)$$

On the other side,

$$E_{X_i X_j} \left[ \int \left| f_{Y|X_i}(y) - f_{Y|X_j X_i}(y) \right| dy \right] \quad (53)$$

is a positive term, representing the expected shift between the distribution of  $Y$  given  $X_i$  and the distribution of  $Y$  conditional on  $X_i$ . We denote this term as

$$E_{X_i X_j} \left[ \int \left| f_{Y|X_i}(y) - f_{Y|X_j X_i}(y) \right| dy \right] = 2\delta_{j|i} \quad (54)$$

Again note that  $E_{X_i X_j} \left[ \int \left| f_{Y|X_i}(y) - f_{Y|X_i X_j}(y) \right| dy \right] \geq 0$  and that  $E_{X_i X_j} \left[ \int \left| f_{Y|X_i}(y) - f_{Y|X_i X_j}(y) \right| dy \right] = 0$  if  $Y$  is independent of  $X_j$ . Combining this facts gives:

$$\delta_{ij} \leq \delta_i + \delta_{j|i} \quad (55)$$

Note that combining eqs. (55) and (30) one can rewrite eq. (55) equivalently as:

$$\delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j|i} \quad (56)$$

since  $\delta_{j|i} \geq 0$ , which is the thesis. ■

## References

- [1] ALIS O. F. AND RABITZ H., 2001: “Efficient Computation of High Dimensional Model Representations,” *Journal of Mathematical Chemistry*, 29 (2), pp. 127-142.
- [2] APOSTOLAKIS G.E., 1995: “A Commentary on Model Uncertainty,” *Proceedings of the Workshop on Model Uncertainty: its Characterization and Quantification*, published by Center for Reliability Engineering, University of Maryland, College Park, Maryland, USA.
- [3] APOSTOLAKIS G. E., 2005: “How useful is quantitative risk assessment”, *Risk Analysis*, 24 (3), pp. 515-520.
- [4] BEDFORD T., 1998: “Sensitivity Indices for (Tree)- Dependent Variables,” *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output*, Venice (Italy), 1998, pp.17-20.
- [5] BIER V.M., 1983: “A measure of uncertainty importance for components in fault trees,” *Transactions of the American Nuclear Society*,45 (1), pp. 384-5.
- [6] BORGONOVO E. AND APOSTOLAKIS G.E., 2001: “A New Importance Measure for Risk-Informed Decision-Making,” *Reliability Engineering and System Safety*, 72 (2), 2001, pp. 193-212.
- [7] BORGONOVO E., APOSTOLAKIS G.E., TARANTOLA S. AND SALTELLI A., 2003: “Comparison of Local and Global Sensitivity Analysis Techniques in Probabilistic Safety Assessment,” *Reliability Engineering and System Safety*, 79, pp. 175-185.
- [8] CAMPOLONGO F. AND SALTELLI A., 1997: “Sensitivity Analysis of an Environmental Model: an Application of Different Analysis Methods,” *Reliability Engineering and System Safety*, 57, 49-69.
- [9] CHEOK M.C., PARRY G.W. AND SHERRY R.R., 1998: “Use of Importance Measures in Risk-Informed Regulatory Applications,” *Reliability Engineering and System Safety*, 60, pp. 213-226.
- [10] CHUN M-H., HAN S-J. AND TAK N-IL., 2000: “An uncertainty importance measure

- using a distance metric for the change in a cumulative distribution function,” *Reliability Engineering and System Safety*, 70, pp. 313-321.
- [11] FREY C. H. AND PATIL S.R., 2002: “Identification and Review of Sensitivity Analysis Methods,” *Risk Analysis*, 22 (3), pp. 553-571.
- [12] HAMMIT J.K. AND SHIYAKHTER I., 1999: “The Expected Value of Information and the Probability of Surprise,” *Risk Analysis*, 19 (1), 135-152.
- [13] HUANG C.-F. AND LITZENBERGER R.H., 1988: “Foundations for Financial Economics,” Prentice Hall, Upper Saddle River, NJ, USA, 365 pages, ISBN 0-13-500653-8.
- [14] HELTON J.C., 1993: “Uncertainty and Sensitivity Analysis Techniques for Use in Performance Assessment for Radioactive Waste Disposal,” *Reliability Engineering and System Safety*, 42, pp. 327-367.
- [15] HELTON J.C. AND DAVIS F.J., 2000: “Sampling-Based Methods,” in Sensitivity Analysis, A. Saltelli, K. Chan, and E.M. Scott, Eds. New York, NY: Wiley. pp. 101-153.
- [16] HELTON J.C. AND DAVIS F.J., 2003: "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering and System Safety*, 81, pp. 23-69.
- [17] HOMMA T. AND SALTELLI A., 1995: “Use of Sobol’ Quasirandom Sequence Generator for Integration of Modified Uncertainty Importance Measure,” *Journal of Nuclear Science and Technology*, 32 (11), pp. 1164-1173.
- [18] HOMMA T. AND SALTELLI A., 1996: “Importance Measures in Global Sensitivity Analysis of Nonlinear Models,” *Reliability Engineering and System Safety*, 52, 1-17.
- [19] HORA S.C. AND IMAN R.L., 1986: “A comparison of Maximus/Bounding and Bayes/Monte Carlo for Fault Tree Uncertainty Analysis,” Technical Report, SAND85-2839, Sandia National Laboratories, Albuquerque, New Mexico.
- [20] HORA S.C. AND HELTON J.C., 2003: “A Distribution-Free Test for the Relationship Between Model Input and Output when Using Latin Hypercube Sampling,” *Reliability Engineering and System Safety*, 79, pp. 333-339.

- [21] KLEIJNEN P.C., 2005: “An Overview of the Design and Analysis of Simulation Experiments for Sensitivity Analysis,” *European Journal of Operational Research*, 164, pp 287-300.
- [22] IMAN R.L., 1987: “A matrix-based approach to uncertainty and sensitivity analysis for fault trees,” *Risk Analysis*, 7 (1), pp. 21-33.
- [23] IMAN R.L. AND CONOVER W.J., 1987: “A Measure of Top-down Correlation,” *Technometrics*, 29 (3), pp. 351-357.
- [24] IMAN R.L. AND HORA S.C., 1990: “A Robust Measure of Uncertainty Importance for Use in Fault Tree System Analysis,” *Risk Analysis*, 10 (3), pp. 401-406.
- [25] IMAN R.L., JOHNSON M. E. AND WATSON C.C. JR., 2005: “Uncertainty Analysis for Computer Model Projections of Hurricane Losses,” *Risk Analysis*, 25 (5), pp.1299-1312.
- [26] MALLIAVIN P., 1995: “Integration and Probability,” Springer Verlag Editions, New York, NY, USA, ISBN 0-387-94409-5.
- [27] MCKAY M.D., 1996: “Variance-Based Methods for Assessing Uncertainty Importance,” in NUREG-1150 Analyses, LA-UR-96-2695, 1-27.
- [28] MCKAY MD, BECKMAN RJ, CONOVER WJ., 1979: “A comparison of three methods for selecting values of input variables in the analysis of output from a computer code,” *Technometrics*, 21(2), pp. 239–245.
- [29] MORRIS M.D., 1991: “Factorial Sampling Plans for Preliminary Computational Experiments,” *Technometrics*, 33 (2), pp. 161-174.
- [30] NAKASHIMA K. AND YAMATO K., 1982: “Variance-importance of system components,” *IEEE Transactions on Reliability* ;R-31(1).
- [31] OAKLEY J.E. AND O’HAGAN A., 2004: “Probabilistic sensitivity analysis of complex models: a Bayesian approach,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66 (3), pp. 751-769.
- [32] PATIL S.R. AND FREY C.H., 2004: “Comparison of Sensitivity Analysis Methods Based on Application to a Food Safety Risk Assessment Model,” *Risk Analysis*, 24 (3),

pp. 573-585.

- [33] PARK C.K. AND AHN K.I., 1994: “A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment,” *Reliability Engineering and System Safety*, 46, pp. 253-61.
- [34] RABITZ H., ALIS O. F., SHORTER J. AND SHIM K., 1998: ‘Efficient Input Output Model Representation’, *Computer Physics Communications*, 117, pp. 11-20.
- [35] RABITZ H. AND ALIS Ö. F., 1999: “General foundations of high-dimensional model representations,” *Journal of Mathematical Chemistry*, 25, no. 2-3, pp. 197–233.
- [36] RATTO M., TARANTOLA S. AND SALTELLI A., 2001: ‘Estimating Importance Indicators for Correlated Inputs’, *Proceeding of ESREL 2001*, Torino, Italy, pp. 157-164.
- [37] SALTELLI A. AND MARIVOET J., 1990: “Non-parametric Statistics in Sensitivity Analysis for Model Output: a Comparison of Selected Techniques,” *Reliability Engineering and System Safety*, 28, 229-253.
- [38] SALTELLI A., 1997: ‘The Role of Sensitivity Analysis in the Corroboration of Models and its Link to Model Structural and Parametric Uncertainty’, *Reliability Engineering and System Safety*, 57, pp. 1-4.
- [39] SALTELLI A., TARANTOLA S. AND CHAN K. P.-S., 1999: ‘A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output’, *Technometrics*, 41, 1, 39-56.
- [40] SALTELLI A., TARANTOLA S. AND CAMPOLONGO F., 2000: “Sensitivity Analysis as an Ingredient of Modelling”, *Statistical Science*, 19 (4), pp. 377-395.
- [41] SALTELLI A. AND TARANTOLA S., 2002: “On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal,” *Journal of the American Statistical Association*, 97 (459), p. 702-709.
- [42] SALTELLI A., 2002: ‘Sensitivity Analysis for Importance Assessment’, *Risk Analysis*, 22 (3), p. 579.
- [43] SAVAGE L., 1972: *The Foundations of Statistics*, Dover Publications, Second Edition.

- [44] JOINT RESEARCH CENTER OF THE EUROPEAN COMMUNITY: “SIMLAB Reference Manual,” POLIS- JRC ISIS, Joint Research Center of the European Community.
- [45] SOBOL I.M., 1990: “Quasi-Monte Carlo Methods,” *Progress in Nuclear Energy*, 24, pp. 55-61.
- [46] SOBOL I.M., 1993: “Sensitivity estimates for nonlinear mathematical models,” *Matem. Modelirovanie*, 2(1) (1990) 112-118 (in Russian). English Transl.: MMCE, 1(4) (1993) 407-414.
- [47] SOBOL’ I.M., 2001: “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates,” *Mathematics and Computers in Simulation*, 55(1), pp. 271-280.
- [48] SOBOL’ I.M., 2003: “Theorem and Examples on High-Dimensional Model Representation,” *Reliability Engineering and System Safety*, 79, 187-193.
- [49] TARANTOLA S., 2000: “Quantifying uncertainty importance when inputs are correlated,” in *Foresight and Precaution*, Cottam Harvey Pape and Tait Editors, Balkema Editions, Rotterdam, ISBN 9058091406.