Blackwell Publishing

Martha Lovvoll
Production Editor
Phone: (781) 388-8409
Fax: (781) 388-8257
mlovvoll@bos.blackwellpublishing.com

Risk Analysis

To the contributor:

PDF proofs of your article to appear in the forthcoming issue of the journal are attached. Be aware that a professional proofreader is reading these pages at the same time you are and will correct any remaining spelling and grammatical errors, typesetting errors, and discrepancies in style. If you are unable to access this file, please contact me immediately and a hard copy can be mailed to you.

Please note essential corrections only and return them within 48 hours of receipt by express mail to:

Deborah Willow
Willowhut Publishing Services, Inc.
Montana Shipping Outlet
107 Dewey Avenue, Box 580
Eureka, MT 59917

You may also fax corrections to her at **(406) 882-4583**.

If you only have a few corrections or would like to email your corrections, compose a detailed list and send to **willowhut@interbel.net**

PLEASE NOTE: It is important that your proofs are returned to the proofreader within 48 hours of receipt to ensure the issue is produced on schedule.

Please also note that only corrections to actual typesetting errors are allowed. Rewriting of material originally submitted will be disallowed at the editor's discretion, and the editor reserves the right to charge back the cost of such changes to authors.

Thank you,

Martha

OFFPRINT ORDER FORM
for Risk Analysis

Blackwell Publishing

All first-named authors receive 25 free offprints of their article. If you wish to order additional offprints of your published article at your own expense, you must complete the form below and return it (with the correct remittance*) to:

Sheridan Reprints
Chris Jones
450 Fame Avenue
Hanover, PA 17331
Phone: 1-800-352-2210, ext. 8128
Fax: 1-717-633-8929
Email: cjones@tsp.sheridan.com

Please return this form within **5** days of receiving it. Required fields must be completed correctly before the issue goes to press in order for this pricing to apply.

Authors may also obtain a PDF version of their article through Synergy, our online content site http://www.blackwell-synergy.com. If your library has a subscription to the journal, you can obtain a PDF without charge. If your library does not have a subscription, you may download a copy via our pay-per-view service. Please refer to the copyright assignment form submitted with the article for terms and conditions of use.

Please fill out the following information: (PLEASE PRINT)

Name: ____

Address: ____

____

Phone: __________________ Email: ________________________

Article Title:____

Authors:____

Number of pages	Charge per 50 copies***	Charge per 100 copies	Additional 100s
1-8	$95.00	$115.00	$55.00
9-16	$160.00	$195.00	$90.00
17-24	$210.00	$265.00	$130.00
25-32	$285.00	$355.00	$165.00
Over 32	$10.00/page	$10.00/page	$10.00/page

Payment Information:

❏ Check enclosed
❏ Purchase order enclosed
❏ Credit card: ❏ MasterCard ❏ VISA Exp. date: ______/______

Number: ___________ - ___________ - ___________ - ___________

Signature: ____

*Payment must be in U.S. funds drawn on a U.S. bank.
**If paying with credit card, you may fax your order to the number listed above.
***Minimum order is 50 copies.
Notes: This pricing is for black and white only. Orders are shipped 3-4 weeks after the issue is published; please note that delivery may take 6-10 weeks.

Actual Page Range of Article*	Volume/Issue*	Qty.*	Total Price
–	/		$ .00
Cover (*optional*, add $80)			$ .00
Domestic Shipping (add $5)			$ .00
International Shipping (add $15)**			$ .00
TOTAL (*please do not forget to include charges for covers and shipping*)			$ .00

*Required fields. **Shipping is not traceable, for other shipping options please contact Chris Jones.

Blackwell Publishing

350 Main Street, Malden, MA 02148-9933 USA
Phone: (781) 388-8200; Fax: (781) 388-8257

PO Box 1354, 9600 Garsington Road, Oxford, OX4 2XG UK
Phone: +44 (0) 1865 778315; Fax: +44 (0) 1865 471775

www.blackwellpublishing.com

Recommend Your Library Subscribe

Dear Contributor:

Together with the Editorial office, our goal is to ensure that this journal is accessible to all researchers and students around the world. Please help us by recommending that your library subscribe to this journal. We enclose a library recommendation form below, which can be completed and forwarded to your librarian or library committee member.

Online Access: Subscription to most of our journals entitles all members of the institution to full online access. For details: www.blackwellpublishing.com/journals

E-mail Alerts: Register today to receive E-mail alerts with complete Table of Contents as future issues publish. For details: www.blackwellpublishing.com

Thank you for contributing to this journal.

Yours sincerely,

The Management, Blackwell Publishing

P.S. If you do not currently have a personal subscription to this journal, and would like to find out more about personal subscriptions, please call us toll-free at 800-835-6770 (outside of the U.S., dial 781-388-8200) or visit our website at http://www.blackwellpublishing.com

✂---

Library Recommendation Form

To: Librarian / Library Acquisition Committee / ________________________________

From: ____

Department: ____

Please include the **journal**: ___________________________________ **ISSN** ______________
in your next serials review meeting with my recommendation to subscribe.

I recommend the journal for the following reasons: (1 = very important; 2 = important)

1 2 **REFERENCE**: I will refer to this journal frequently for new research articles related to my work.
1 2 **STUDENT READINGS**: I will be referring my students to this journal regularly to assist their studies.
1 2 **BENEFIT FOR LIBRARY**: My evaluation of the journal's content and direction is very high. Its acquisition will add to the library's success in fulfilling department, faculty, and student needs.
1 2 **OWN AFFILIATION**: I am a member of the journal's advisory board. I therefore support it strongly and use it regularly in my work. I will regularly recommend articles to colleagues and students.

Signature:____Date:__________________________

This journal is published by Blackwell Publishing, 350 Main St., Malden, MA 02148, USA.
Tel: 800-835-6770 in the U.S., outside the U.S. dial (781) 388-8200; Fax: (781) 388-8232, Email: subscrip@bos.blackwellpublishing.com

10DM2

 DOI: 10.1111/j.1539-6924.2006.00806.x

Measuring Uncertainty Importance: Investigation and Comparison of Alternative Approaches

Emanuele Borgonovo*

Q1

Uncertainty importance measures are quantitative tools aiming at identifying the contribution of uncertain inputs to output uncertainty. Their application ranges from food safety [Frey and Patil (2002) (*Risk Analysis, 22*(3), 553–571)] to hurricane losses [Iman *et al*. (2005)]. Results and indications an analyst derives depend on the method selected for the study. In this work, we investigate the assumptions at the basis of various indicator families to discuss the information they convey to the analyst/decision maker. We start with nonparametric techniques, and then present variance-based methods. By means of an example we show that output variance does not always reflect a decision maker state of knowledge of the inputs. We then examine the use of moment-independent approaches to global sensitivity analysis, i.e., techniques that look at the entire output distribution without a specific reference to its moments. Numerical results demonstrate that both moment-independent and variance-based indicators agree in identifying non-influential parameters. However, differences in the ranking of the most relevant factors show that inputs that influence variance the most are not necessarily the ones that influence the output uncertainty distribution the most.

KEY WORDS: global sensitivity analysis; Importance measures; probabilistic risk assessment; uncertainty analysis; uncertainty importance measures

1. INTRODUCTION

Several recent works demonstrate that uncertainty and sensitivity analyses have become an essential part of the modelling and risk assessment of complex systems (Saltelli, 2002); Patil & Frey 2004; Iman *et al.*, 2005a, 2005b). For example, as reported in Iman *et al.* (2005a), the Florida Commission on Hurricane Loss Projection Methodology underlines that "*an important part of the auditing process requires uncertainty and sensitivity analyses to be performed with the applicant's proprietary model*." Apostolakis (1995, 2005) highlight that the treatment of uncertainties can play a critical role in probabilistic risk assessment.

Saltelli (2002) defines sensitivity analysis (SA) as the determination 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).*" The origin of global SA can be traced back to Cuckier *et al.* (1973), and since then various indicators have been introduced to address uncertainty importance. Without the claim of being exhaustive, we summarize the uncertainty indicators in the categories of nonparametric techniques (Saltelli & Marivoet, 1990; Iman *et al.* 2005a), screening methods[1] (Morris, 1991), variance-based methods (Sobol, 1993, 2001, and 2003; Rabitz *et al.*, 1998; Rabitz & Alis, 1999; Saltelli *et al.*, 2000; Alis & Rabitz, 2001; Frey & Patil, 2002; Saltelli, 2002; Patil & Frey 2004)

* Bocconi University, Quantitative Methods, Viale Isonzo 25, Milan, 20135, Italy; tel: 0039-02-5836-5608; emanuele.borgonovo@unibocconi.it

[1] Although of interest with respect to the problem of the curse of model dimensionality (Alis and Rabitz, 2001), screening methods will not be dealt with in this work since not in the scope of the present investigation. We limit ourselves to recall the quite successful approach proposed by Morris (1991) and utilized, for example, in Campolongo and Saltelli (1997).

Table I. The Uncertainty Importance Measures Dealt with in This Work

Importance Measure	Acronym	Family	Uncertainty Importance Measured Referring at
Andsten-Vaurio (1992)	AV	Variance based	Output variance
Pearson correlation coefficient	PEAR	Nonparametric	Input–output correlation
Standardized correlation coefficient	SCC	Nonparametric	Input–output correlation
Partial correlation coefficient	PCC	Nonparametric	Input-output correlation
Helton (1993)	H	Variance based	Output variance
Iman (1987)	I	Variance based	Output variance
Sobol (1993)	$S_{i1,i_2,\ldots,i_r}$	Variance based	Output variance
FAST (Saltelli *et al.*, 1999)	S1, ST	Variance based	Output variance
Borgonovo (2005)	δ	Moment independent	Entire output distribution

and moment-independent approaches (Park & Ahn, (1994), Chun *et al.*, 2000; Borgonovo, 2005). Table I lists the indicators discussed in this work.

We investigate how these families of global SA techniques address the problem of identifying uncertainty drivers. We examine common features and differences in their definitions and compare the information they convey to analysts/decision makers. In fact, Frey and Patil (2002) demonstrate that the information a decision maker derives from the analysis depends on the choice of the technique.

As Saltelli (2002) and Frey (2002) underline, the selected methods ought to satisfy the requirements of being "*global, quantitative, and model free*." The works of Sobol (1993, 2001, 2003), McKay (1996), Saltelli *et al.* (1999), Rabitz and Alis (1999), and Alis and Rabitz (2001) have solved the problem of achieving a complete output variance decomposition orig-
Q2 inated by Cuckier (1973), and variance-based methods have established themselves as the preferred way of measuring uncertainty importance. In addition, as Sobol (2001) and Alis and Rabitz (2001) have proven, variance decomposition reflects the underlying function decomposition in the absence of correlations among the parameters. Therefore, a variance-based analysis provides analysts with information not only on parameter contribution to the output variance, but also on model structure when parameters are independent.

The use of variance-based techniques has sometimes been given a broader interpretation, assuming that "*this moment is sufficient to describe output variability.* (Saltelli, 2002)." For example, Homma and Saltelli (1995) highlight the use of variance-based SA as a way of determining how "*the total uncertainty in model prediction is apportioned to uncertainty in the model input parameters*." Following the same inter-
Q3 pretation, Iman *et al.* (2005) refers to variance-based indictors as "*the expected percentage reduction in the uncertainty . . . that is attributable to each of the input variables.*" As Oakley and O'Hagan (2004) underline, Q4
in correspondence of the realized value of the uncertain parameter, variance can increase, in contrast with the expectations. By means of an example, we show that, in this case, relying on the sole variance as an indicator of uncertainty would lead the decision maker to non-informative conclusions.

We then examine the use of moment-independent uncertainty importance measures (Chun *et al.* 2000; Borgonovo, 2005), *i.e.*, uncertainty importance measures that look at the entire output distribution without referring to one of its moments[2]. We focus on an importance measure recently introduced in Borgonovo (2005). We discuss its properties and compare this technique to nonparametric and variance-based uncertainty importance measures.

We deal with the comparison of the numerical results provided for by the above discussed families of indicators. The numerical analysis is carried out by means of the probabilistic risk assessment model utilized by Iman (1987). On this model uncertainty importance measures were introduced for the first time, and the model has then been utilized as a test case in several works on uncertainty importance measures (see also Chun *et al.,* 2000). We examine whether the indicators agree in identifying the top-ranked and/or the low-ranked parameters by computing the correlation coefficients on the ranking and on the corresponding Savage scores[3].

[2] Most of the works at the basis of Bayesian statistics maintain that it is the entire distribution to reflect the decision maker state-of-knowledge on an uncertain quantity (de Finetti, 1937; Savage, 1972).

[3] Savage scores have been originally introduced by Iman and Conover (1987). For illustrations on their application we also refer the reader to Saltelli and Marivoet (1990) and Campolongo and Saltelli (1997).

Results show that parameters that are nonrelevant on the entire distribution—low value of the moment independent indicator—tend also not to be influential on input–output correlation and on output variance. A higher disagreement is registered among the ranking of the most influential factors. This reveals that inputs influencing the entire distribution the most are not necessarily the ones that influence variance the most. We show that the difference is explained by the scope and meaning of the indicators and can be exploited by the decision maker to gain full information on uncertainty propagation results. In fact, by utilizing variance-based techniques, an analyst is capable of gaining insights on the model structure—besides obtaining the parameter contribution to the output variance—and by using moment-independent indicators she/he obtains insights on the influence the
Q5 uncertain inputs have on the output distribution.

The remainder of the article is organized as follows. We start with a decision maker/analyst who chooses to measure uncertainty importance as result of input–output correlation making use of nonparametric techniques (Section 2). After showing that the mixed indicators introduced with the differential analysis of Helton (1993) and nonparametric techniques would produce equivalent ranking under the assumption that a linear regression fits the model output, in Section 3 we present the definition of variance-based importance measures (Iman, 1987) and the global sensitivity indices (Sobol, 1993, 2001, and 2003; Rabitz *et al.,* 1998; Rabitz & Alis 1999; Saltelli *et al.*, 1999, Saltelli *et al.*, 2000; Alis & Rabitz, 2001). Through an example we show that non-informative results can be obtained when a decision maker relies on variance as the sole representative of the output uncertainty. Section 4 deals with moment-independent sensitivity indicators. Section 5 presents a quantitative comparison among the results of the techniques by application to the Iman (1987) model. Ranking agreement is analyzed by the computation of the correlation coefficients on both ranks and Savage scores. Conclusions are offered in Section 6.

2. MEASURING UNCERTAINTY IMPORTANCE AS INPUT–OUTPUT CORRELATION

In this section, we deal with a decision maker/analyst who decides to utilize correlation as a measure of the influence of uncertain inputs on output uncertainty. The analyst then ranks as more influential the parameter X_i with the highest correlation coefficient with respect to (w.r.t.) the output (Y). Please refer to Tables I and II for acronyms and notation utilized throughout this work respectively.

Methods explored in the early literature to assess parameter uncertainty importance based on correlation and regression are classified under the name of nonparametric techniques. Saltelli and Marivoet (1990) provide a thorough description of these techniques, their properties, and applicability.

As Saltelli and Marivoet (1990) discuss, the Pearson correlation coefficient (PEAR) belongs to the family of nonparametric techniques defined as follows Q6

$$\mathrm{PEAR}_l = \frac{\mathrm{Cov}(Y, X_l)}{\sigma_l \sigma_Y}. \tag{1}$$

Suppose now that the model output can be fitted by a linear regression on the inputs, i.e., (Helton, 1993 or Frey & Patil, 2002)

$$Y = g(\underline{X}) \simeq b_0 + \sum_{i=1}^{n} b_i X_i. \tag{2}$$

If the approximation in Equation (2) works then a natural measure of the sensitivity of Y on X_l is the standardized regression coefficient (SRC) (Saltelli and Marivoet, 1990)

$$\mathrm{SRC}_l = \frac{b_l \sigma_l}{\sigma_Y}, \tag{3}$$

where σ_l and σ_Y are the standard deviations of x_l and Y, respectively, and b_l is the coefficient of the multivariate linear regression of Y on $\underline{X}$. Note that, if parameters are independent, since $b_l = \frac{\mathrm{Cov}(Y, X_l)}{\sigma_l^2}$, then SRC_l and PEAR_l actually coincide: $\mathrm{SRC}_l = \frac{\mathrm{Cov}(Y, X_l)}{\sigma_l \sigma_Y}$.

Other representatives of this family of methods are discussed in Saltelli and Marivoet (1990). In this work, we make use of the partial correlation coefficient (PCC_l) and the Spearman correlation coefficient (SPEAR_l). As Saltelli and Marivoet (1990) demonstrate, the performance of nonparametric techniques in explaining model variance is synthesized by the model coefficient of determination, R_Y^2. Let $\underline{x}^i$ be the ith realization of the input, $i = 1, 2, \ldots, N$,[4] y_i the output values registered in correspondence of the input

[4] N denotes sample size (Table II).

realizations. R_Y^2 is defined as

$$R_Y^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \quad (4)$$

where $\hat{y}_i$ is the estimate of the output obtained when each of the x^i is substituted in the linear regression, $\bar{y}$ is the average of the y_i, the effectively registered output values.

R_Y^2 represents the fraction of the model variance explained by the regression. A low value of R_Y^2 would signal "*a poor regression model* (Campolongo and Saltelli, 1997)". In that case, the ability of nonparametric techniques in capturing output variability is low, and "*it is hence unrealistic to assess influence of the input variables based on SRC's* (Campolongo and Saltelli, 1997)." As Frey and Patil (2002) underline and, as Saltelli and Marivoet (1990) discuss, this happens for nonlinear models and also when interactions among parameters emerge. In the next section, we discuss how variance-based methods can be utilized to integrate and improve the information obtained by nonparametric techniques.

3. MEASURING UNCERTAINTY IMPORTANCE USING VARIANCE

After the works of Sobol (1993, 2001 and 2003), Q7 Rabitz *et al.*, (1999), Alis and Rabitz (2001), Saltelli *et al.* (2000), Saltelli *et al.* (1999), and Saltelli (2002) variance-based techniques have been recognized as the preferred methods for assessing uncertainty importance. Before these works, uncertainty importance measures based on output variance were introduced in the works of Nakashima and Yamato (1982), Bier (1983), Iman (1987), and Iman and Hora (1990).

Let us consider the differential approach to sensitivity analysis utilized by Helton (1993). In this case the model output is written as

$$Y \cong g(\underline{X}^0) + \sum_{i=1}^{n} \frac{\partial g(\underline{X}^0)}{\partial X_i}(X_i - X_i^0) + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 g(\underline{X}^0)}{\partial X_i \partial X_k}(X_i - X_i^0)(X_k - X_k^0). \quad (5)$$

Helton (1993) shows that if the first-order terms of the above approximation are used, one can estimate the model variance as

$$V[Y] \cong \sum_{i=1}^{n}\left[\frac{\partial g(\underline{X}^0)}{\partial X_i}\right]^2 V[X_i]. \quad (6)$$

Hence, Helton (1993) envisions as natural sensitivity measures of the importance of a parameter on the output variance the quantity (see also Chun *et al.*, 2000) Q8

$$H_l = \sqrt{\left[\frac{\partial g(\underline{X}^0)}{\partial X_l}\right]^2 \frac{V[X_l]}{V[Y]}} = \frac{\partial g(\underline{X}^0)}{\partial X_l}\frac{\sigma_l}{\sigma_Y}, \quad (7)$$

which represents the uncertainty importance of X_l measured as the partial derivative of the output w.r.t. input $X_l(\frac{\partial Y}{\partial X_l})$ normalized through the ratio of the standard deviation of $X_l(\sigma_j)$ and Y (σ_Y). A similar measure is proposed in Andsten and Vaurio (1992), and defined as

$$\mathrm{AV}_l = \frac{\left[\frac{\partial g(\underline{X}^0)}{\partial X_l}\right]^2 \sigma_l^2}{\sum_{s=1}^{n}\left[\frac{\partial g(\underline{X}^0)}{\partial X_s}\right]^2 \sigma_s^2}. \quad (8)$$

It is easy to prove the following relationship between AV_l and H_l

$$\mathrm{AV}_l = \frac{\mathrm{H}_l^2}{\sum_{s=1}^{n}\mathrm{H}_s^2}. \quad (9)$$

It is worth noting that, if a model is linear or well approximated by a linear regression, i.e., Equation (2) holds, then one can write

$$\frac{\partial g(\underline{X}^0)}{\partial X_l} = b_l, \quad (10)$$

and therefore the Helton indicator for parameter X_l coincides with SRC_l [5]:

$$\mathrm{H}_l = \frac{\partial Y}{\partial X_l}\frac{\sigma_l}{\sigma_Y} = \frac{b_l\sigma_l}{\sigma_Y} = \mathrm{SCR}_l. \quad (11)$$

As in the case of a linear regression, the accuracy of the variance approximation in Equation (6) deteriorates rapidly with the nonlinearity and nonadditivity of the model. Hence, assessing the relevance of a parameter on the model variance utilizing H_l (Equation (7)) or SRC_l (Equation (3)) could lead to nonefficient results.

[5] Thanks to Equation (9) the ranking induced by AV_l would be the same as those of H_l, SRC_l for a linear model.

To overcome this problem, i.e., to find a variance-based indicator independent of the model linearity, an alternative class of importance measures have been developed. They have been referred to as *uncertainty importance measures* (Iman, 1987 and also Iman and Hora, 1990) and determine the influence of X_l as the reduction in output variance that follows a reduction in the uncertainty in X_l (Iman, 1987, see also Chun *et al.*, 2000 and Saltelli *et al.*, 2000, or Saltelli, 2002). We report their definition as given in Saltelli *et al.* (2000)

$$\mathrm{I}_l = V[Y] - E\{V[Y \mid X_l]\} = V\{E[Y \mid X_l]\}. \tag{12}$$

where $V[Y]$ is the variance of the model output Y, and $E\{V[Y \mid X_l]\}$ is the conditional expected value of $V[Y]$ given X_l and the expectation is taken over the possible values of X_l, weighted by the appropriate density.

It can be proven that the Iman uncertainty importance measure (I_l from now on) is the expected reduction in output variance that can be achieved if uncertainty in X_l is eliminated (Saltelli *et al.*, 2000; Saltelli, 2002). I_l produces the importance of individual parameter effects on the output variance, but neglects interaction effects. It has been generalized by global importance measures introduced in the works of Sobol (1993), Homma and Saltelli (1995, 1996). The model output variance is decomposed as follows (Efron and Stein, 1981; Sobol' (1993, 2001), and 2003)

$$V[Y] = \sum_{l=1}^{n} V_l + \sum_{l<j} V_{l,j} + \sum_{l<j<m} V_{l,j,m} + \cdots + V_{1,2,\ldots,n} \tag{13}$$

In Equation (13), V_l represents the contribution to the output variance provided for by parameter X_l individually. It turns out that V_l is the "*expected amount of variance reduction that would be achieved for* Y, *if we were able to specify* X_l *exactly* (Bedford, 1998)" and, therefore, coincides with the I_l indicator (Equation (12)). $V_{l,j}$ is the contribution to the variance of the interaction[6] between parameter X_l and X_j. $V_{l,j,m}$ is the contribution to the output variance of the interaction between parameters X_l, X_j, and X_m. Finally, $V_{1,2,\ldots,n}$ is the residual portion of the variance that can only be explained as effect of the interaction among all the parameters.

Sobol (1993, 2001, and 2003), Homma and Saltelli (1995, 1996) introduce the sensitivity indices of order r as the the ratios of the interaction terms of order r, $V_{i1,i_2,\ldots,i_r}$ in Equation (13), and $V[Y]$

$$S_{i_1,i_2,\ldots,i_r} \equiv \frac{V_{i_1,i_2,\ldots,i_r}}{V[Y]}. \tag{14}$$

Particular interest is deserved by the interpretation of the first-order (here denoted as $\mathrm{S1}_l$) and total order sensitivity indices—or "*total effects*" in Saltelli (2002)—(ST_l). The first-order indices, are defined as

$$S1_l \equiv \frac{V_l}{V[Y]} \quad l = 1, 2, \ldots, n, \tag{15}$$

represent the expected percent reduction in $V[Y]$ which is obtained when uncertainty in X_l is eliminated (Saltelli *et al.*, 2000; Saltelli, 2002). Note that if one selects $\mathrm{S1}_l$ as uncertainty importance measure of X_l, one would obtain the same ranking as with I_l.

The total effects

$$\mathrm{ST}_l \equiv \sum_{r=1}^{n} \sum_{i_1,i_2,\ldots,i_r} S_{i_1,i_2,\ldots,i_r}\ (i_1 = l), \tag{16}$$

represent the expected percentage of variance that remains if all parameters but X_l were known (Saltelli, 2002). By selecting ST_l as an 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, 1996; Saltelli *et al.*, 1999; Saltelli *et al.*, 2000). Q9

The variance decomposition terms in Equation (13) can be computed from the function decomposition proven by Sobol (1993)—see also Rabitz and Alis (1999) and Alis and Rabitz (2001)

$$g(\underline{X}) = g_0 + \sum_{l=1}^{n} g_l(X_l) + \sum_{l<j} g_{l,j}(X_l, X_j) + \cdots + g_{1,2,\ldots,n}(X_1, X_2, \ldots, X_n), \tag{17}$$

where

$$\begin{aligned} g_0 = E[Y] &= \int \cdots \int g(\underline{x}) \prod_{k} f_{X_k}(x_k)\, dx_k \\ g_0 + g_l(X_l) &= \int \cdots \int g(\underline{x}) \prod_{k \neq l} f_{X_k}(x_k)\, dx_k \\ g_0 + g_l(X_l) + g_j(X_j) + g_{l,j}(X_l, X_j) &= \int \cdots \int g(\underline{x}) \prod_{k \neq l,j} f_{X_k}(x_k)\, dx_k. \\ &\ldots \end{aligned} \tag{18}$$

Note that each of the summands in the variance decomposition of Equation (13) is the average of the square of the corresponding term in Sobol function decomposition (Sobol, 2001)

[6] Rabitz and Alis (1999) also use the term "cooperation" as a synonym for "interaction."

$$V_{l,j\ldots,m} = \int \cdots \int [g_{l,j,\ldots,m}(x_l, x_j, \ldots, x_m)]^2 \times \prod_{k=l,j,\ldots,m} f_{X_k}(x_k)\, dx_k. \qquad (19)$$

Equation (19) implies that variance analysis leads direct information on the model structure.

Sobol function decomposition rests on the following assumptions:

Assumption 1. $g(\underline{X})$ is measurable (Sobol 1993, 2001, and 2003).

Assumption 2. $f_{\underline{X}}(\underline{x}) = \prod_{l=1}^{n} f_{X_i}(x_i)$, where $f_{X_i}(x_i)$ is the density of each of the parameters (Rabitz and Alis, 1999).

Assumption 1 is of a technical nature, insofar it is required to perform the integrations. Assumption 2, however, shares two interpretations. In Rabitz and Alis (1999) it is broadly underlined that the decomposition of a measurable function can be performed without adding to the integration operation a state-of-knowledge meaning. That is, suppose one is given a multivariate measurable function $g(\underline{X})$ and that X_1, $X_2, \ldots, X_n$ vary in ranges $[-\infty \leq a_i, b_i \leq \infty]$; one can compute its decomposition by mechanically performing nested integrations as stated in Equation (17). In this sense, Sobol/Rabitz function decomposition would not differ conceptually from a decomposition in Taylor polynomials. However, if the probability distribution ought to reflect the uncertainty of a decision maker on the input parameters, then Assumption 2 is equivalent to state that the inputs are statistically independent. In case parameters are not independent and correlations emerge, Bedford (1998) shows that the function decomposition is no more unique, and "*the values taken on by the indices depend on the ordering of the variables*." In terms of model structure, Oakley and O'Hagan (2004) 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 input–output dependence when correlations emerge.

The problem of variance-based SA under correlations, has then been addressed by Saltelli and Tarantola (2002) (see also Saltelli *et al.*, 2004). Saltelli and Tarantola (2002) establish two lottery settings. The first setting consists in identifying the factor that, "*if determined, (i.e., fixed to its true value) would lead to the greatest reduction in the variance of Y.*" They show that the factors associated with the lowest $E_{X_i}\{V[Y \mid X_i]\}$ are the more effective in reducing output variance (Saltelli *et al.*, 2000) both in the presence and in the absence of correlations. Note that Equation (12) states that parameters with low value of $E_{X_i}\{V[Y \mid X_i]\}$ are the ones with a high value of the Iman importance.

The second lottery setting of Saltelli and Tarantola (2002) parallels the first, and consists in identifying the smallest number of factors that lead to a predetermined variance reduction. In this case, the terms $V[Y \mid X_i, X_j, \ldots, X_m]$ matter, and by extension of Equation (12), the sensitivity measures become $V_{i,j,\ldots,m} = V\{E(Y \mid X_i, X_j, \ldots, X_m)\}$.

As the above discussion illustrates, variance-based importance measures are built in such a way that the parameter with the highest value of the indicator is the most effective in reducing output variance $V[Y]$. The interpretation is, however, somewhat stretched so that they are deemed to measure "*the relative importance of each input in driving the uncertainty* (Oakley and O'Hagan, 2004)." This implication is related to the traditional choice of $V[Y]$ as the privileged indicator of uncertainty. However, Saltelli (2002) observes that focusing on variance as the sole measure of uncertainty, is equivalent to assume "*that this moment is sufficient to describe output variability.*" More specifically, the conditions under which $V[Y]$ is in a one-to-one correspondence with the decision maker uncertainty are (Huang & Litzenberger, 1988), Ch. 3, pp. 61–62): (1) the decision maker possesses a **Q10**
quadratic utility function, independently of the distribution of Y; (2) Y is normally distributed, independently of the form of the utility function.

We illustrate the effect of points (1) and (2) through an example. Consider the following model output

$$Y = g(X_1, X_2) = e^{X_1} |\sin X_2|, \qquad (20)$$

and, for the moment, suppose that the decision maker characterizes his/her uncertainty in the inputs as follows: X_1 is distributed according to a standard normal and X_2 according to a normal distribution with mean and standard deviation equal to 1. Given this state of knowledge, the output variance induced by uncertainty in both parameters is

$$V_{X_1 X_2}[Y] = 2.72. \qquad (21)$$

Note that since both terms in the product of Equation (20) are greater than zero, Y is not characterized by a normal distribution. Now, the decision maker is offered to pay a very small amount[7] in order to get perfect information on X_2 ($X_2 = 1$.) He/she bases his/her decision on $V[Y]$. In other words, the decision

[7] More precisely, the decision maker is offered to pay much less than the expected value of perfect information on X_2.

maker accepts the offer if variance is reduced by the new information; he/she refuses the offer if variance is increased, because in this second case he/she would consider himself more uncertain. Note that there is an expected variance reduction associated with X_2 equal to 10%. Before accepting the offer, the decision maker recomputes the model output variance and gets

$$V_{X_1}[Y] = (e^2 - e)(\sin^2(1)) = 3.31, \qquad (22)$$

where we have used the subscript X_1 in V_{X_1} $[Y]$ to denote that V_{X_1} $[Y]$ is provoked by uncertainty in X_1 alone. One can then understand the decision maker disappointment: he/she expected a reduction in variance, while the new information is provoking an uncertainty (variance) increase. He would then refuse the perfect information offer, judging it misleading. However, since he/she was asked to pay less than the expected value of perfect information, refusing the offer would mean to go against a basic decision analysis rule. Paraphrasing Saltelli (2002), the assumption that variance is sufficient to describe uncertainty does not hold in this case. In fact, what reflects a decision maker state of knowledge on an uncertain quantity is the distribution (Fig. 1) (de Finetti, 1937; Savage, 1972). In the next section, we discuss methods for global SA that consider the entire input and output distributions, without reference to a particular moment.

4. MEASURING UNCERTAINTY IMPORTANCE LOOKING AT THE ENTIRE DISTRIBUTION

Uncertainty importance measures that do not refer to a particular moment of the output but look at the entire distribution (Fig. 1) have been introduced in the works of Park and Ahn (1994), Chun *et al.* (2000), and Borgonovo (2005).

For the sake of brevity, while we refer to Borgonovo (2005) and Chun *et al.* (2000) for a comparison of the indicators, we focus on the following definition of moment-independent importance measure (Borgonovo, 2005)

$$\delta_l = \frac{1}{2} E_{X_l}[s(X_l)], \qquad (23)$$

where:

- δ_l is the importance of uncertain input X_l;
- $$s(X_l) = \int |f_Y(y) - f_{Y|X_l}(y)|\, dy, \qquad (24)$$

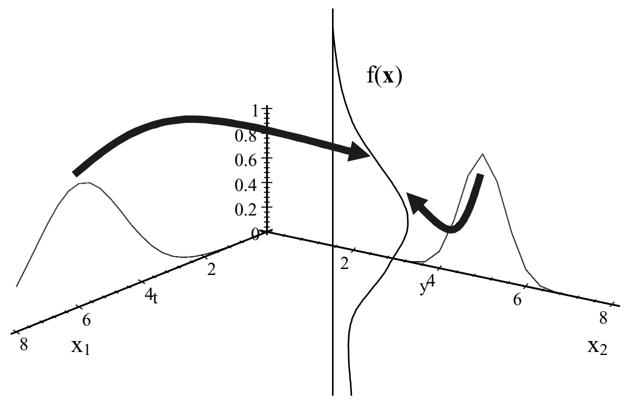

Fig. 1. A visual representation of global SA, in whch the relevance of an input is determined by considering the entire input and output distributions.

is the area between the output density $f_Y(y)$ (continuous line in Fig. 2) and the conditional density of Y given X_l, $f_{Y|X_l=x_l}(y)$ (dashed line in Fig. 2).

- E_{X_l} $[s(X_l)]$ is found from

$$E_{X_l}[s(X_l)] = \int f_{X_l}(x_l)\left[\int |f_Y(y) - f_{Y|X_l}(y)|\, dy\right] dx_l, \qquad (25)$$

where $f_{X_l}(x_l)$ is the density of X_l (Table II).

Let us now illustrate the meaning of the above definition (Equation (23)). $s(X_l)$ (Equation (24)) measures the shift which is provoked in the output distribution when X_l is fixed at one of its possible

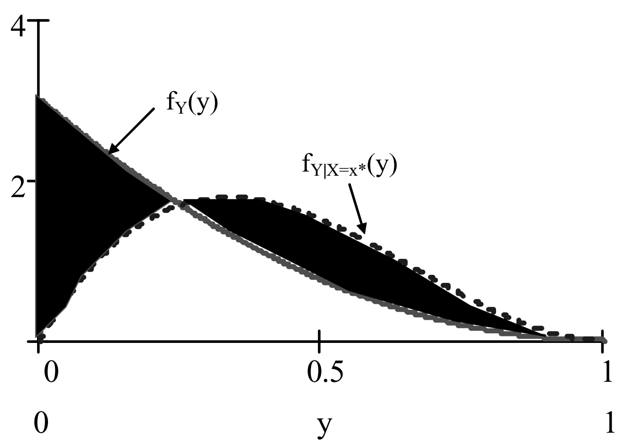

Fig. 2. $f_Y(y)$ (continuous) and $f_{Y|X_j=x_j^*}(y)$ (dashed). The shift between the two densities is measured by the shaded regions.

Table II. Mathematical Notation Utilized in This Work

Symbol	Meaning
Y	Model output
$g(\underline{X})$	Input-output functional relationship
$\underline{X} = (X_1, X_2, \ldots, X_n)$	Vector of uncertain input parameters
$\underline{x} = (x_1, x_2, \ldots, x_n)$	A realization of the input vector
n	Number of parameters
N	Sample size
$F_{\underline{X}}(\underline{x})$	Joint distribution of the input vector
$f_{\underline{X}}(\underline{x})$	Joint density of the input
$f_{X_l}(x_l) = \int \ldots \int f_{\underline{X}}(\underline{x}) \prod_{s \neq l} dx_s$	Marginal density of input X_l
$F_Y(y)$	Distribution of the output
$f_Y(y)$	Density of the output
$f_{Y\mid X_l}(y)$	Conditional density of Y given X_l

values. If one lets uncertainty coincide with the output distribution, then $s(X_l)$ provides an indication of the change in the decision maker view of the output provoked by knowledge of X_l. If the knowledge of X_l changes the output distribution significantly, then the area between the conditional and unconditional distributions of Y will be significant and the input will be registered among the relevant ones. If a parameter is nonrelevant to the decision maker view of the model output, then little or no change in the distribution of Y will be registered and the parameter will be associated with a low importance. Indeed, in the extreme case in which Y is independent of X_l, then $f_Y(y) = f_{Y\mid X_l}(y)$ and δ_l equals 0. Now, since $s(X_l)$ is dependent on X_l, $s(X_l)$ is a function of random variable. Taking the expectation based on the marginal distribution of X_l, namely $E_{X_l}[s(X_l)]$ (Equation (25)), one measures the average shift in the decision maker view on the output provoked by X_l.

The definition of δ can be extended to any group of inputs, $\underline{R} = (X_{i_1}, X_{i_2}, \ldots, X_{i_r})$, as follows (Borgonovo, 2005)

$$\delta_{i_1, i_2, \ldots, i_r} = \frac{1}{2} E_{\underline{R}}[s(\underline{R})], \tag{26}$$

$$= \int f_{X_{i_1}, X_{i_2}, \ldots, X_{i_r}}(x_{i_1}, x_{i_2}, \ldots, x_{i_r}) \times \left[\int \left| f_Y(y) - f_{Y\mid X_{i_1}, X_{i_2}, \ldots, X_{i_r}}(y) \right| dy \right] \times dx_{i_1} dx_{i_2} \ldots dx_{i_r}, \tag{27}$$

where

$$f_{X_{i_1}, X_{i_2}, \ldots, X_{i_r}}(x_{i_1}, x_{i_2}, \ldots, x_{i_r}) = \int \ldots \int f_{\underline{X}}(\underline{x}) \prod_{k \neq i_1, i_2, \ldots, i_r} dx_k. \tag{28}$$

Table III. Properties of the Delta Uncertainty Importance Measure

No.	Property
1	$0 \leq \delta_i \leq 1$
2	$\delta_i = 0$ if Y is independent of X_i
3	$\delta_{1,2,\ldots,n} = 1$
4	$\delta_{ij} = \delta_i$ if Y is dependent on X_i but independent of X_j
5	$\delta_i \leq \delta_{ij} \leq \delta_i + \delta_{j\mid i}$

Given the definitions in Equations (23) and (26), δ shares the properties reported in Table III (the proofs can be found in Borgonovo (2005)). One can summarize these properties as follows. Property 1 implies that the δ of an individual parameter or of a group can only assume values between 0 and 1. Property 2 suggests that a parameter/group has null importance when the model output Y is independent of that parameter/group. Property 3 states that the joint importance of all inputs equals 1. Properties 4 and 5 refer to the joint importance of two (or more) parameters. Property 4 says that if Y is dependent on X_i but independent of X_j then $\delta_{ij} = \delta_i$. Property 5 states that the joint importance of two parameters is greater than the importance of an individual parameter, but limited by the sum of such importance and the conditional term $\delta_{j\mid i}$ given by

$$\delta_{j\mid i} = \frac{1}{2} E_{X_i X_j} \left[\int \left| f_{Y\mid X_i}(y) - f_{Y\mid X_i X_j}(y) \right| dy \right]. \tag{29}$$

$\delta_{j\mid i}$ represents the expected shift from having fixed X_i and then fixing X_j. Let us clarify (see also Borgonovo, 2005). Based on the meaning of δ_i, δ_{ij} represents the expected shift provoked by fixing both X_i and X_j. One can think of such a shift as obtained in a two-step process. Starting with $f_Y(y)$ (Fig. 3), the first step is fixing X_i at x_i. This step leads to $f_{Y\mid X_i = i}(y)$. The second step then consists of fixing of X_j at x_j, which leads to $f_{Y\mid X_i X_j}(y)$. Now, δ shares the properties of a distance (see Malliavin, 1995 for definition of distance). Thus, the first step is measured by δ_i, which represents the distance between the density of Y and the conditional density of Y given X_i. The second step is measured by $\delta_{j\mid i}$, which is the distance between the density of Y given X_i, $f_{Y\mid X_i}(y)$, and the conditional density of Y given both X_i and X_j, $f_{Y\mid X_i X_j}(y)$. Now, going from $f_Y(y)$ to $f_{Y\mid X_i X_j}(y)$ through $f_{Y\mid X_i}(y)$ one is following a longer path than going directly from $f_Y(y)$ to $f_{Y\mid X_i X_j}(y)$. Therefore, one finds: $\delta_{ij} \leq \delta_i + \delta_{j\mid i}$.

Two remarks. The definitions of the δ uncertainty importance for individual parameters (Equation (23)) and for groups (Equation (26)) hold independently of

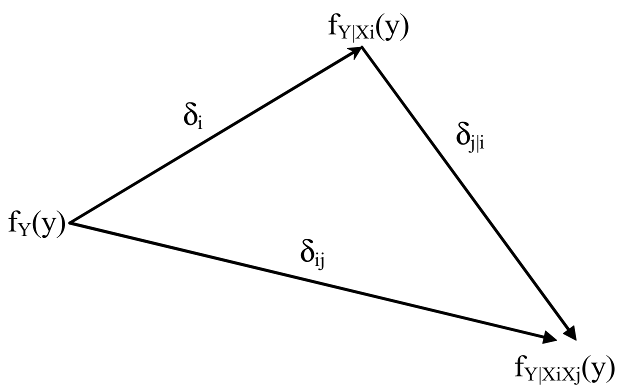

Fig. 3. Geometric interpretation of Property 5.

whether the parameters are correlated. In fact, Equations (23) and (27) require the specification of the joint density, $f_{\underline{X}}(\underline{x})$, without reference to the eventual independence of the parameters (i.e., it is not necessary that $f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$). As far as Saltelli and Tarantola's (2002) settings discussed in Section 3 are concerned, δ_l (Equation (23)) could be seen as corresponding to a bet regarding the identification of which of the uncertain inputs change the view of the decision maker on the problem the most. We recall that Setting 1 in Saltelli and Tarantola (2002) concerned the parameters that lead to the highest expected variance reduction. Similarly, $\delta_{i,j,\ldots,k}$ enables to identify the group of parameters that shifts the decision maker view the most, while Setting 2 in Saltelli and Tarantola (2002) concerns the group expected to reduce variance the most.

Let us now summarize the discussion of Sections 2–3. A decision maker can look at uncertainty importance in different ways (Table I, Column 4.) Choosing nonparametric techniques, one measures uncertainty importance looking at input–output correlation; using variance-based methods, one measures uncertainty importance based on contribution to output variance; making use of a moment-independent approach one measures uncertainty importance considering the whole output distribution. Since the indicators differ both in their conceptual and mathematical formulation, one can expect that, in general, different rankings are obtained with the measures. In the next section, we discuss the comparison of the numerical results for the Iman (1987) model.

5. QUANTITATIVE COMPARISON

This section details the quantitative comparison of the uncertainty importance measures presented in Sections 2, 3, and 4 by means of their application to the Iman (1987) model. The model calculates the risk of a system failure measured by the top event frequency as a function of seven parameters, $\underline{X} = (X_1, X_2, \ldots, X_7)$, where X_1 and X_2 are initiating event frequencies and the remaining parameters are system failure probabilities.

Table IV. Distributions for the Parameters of Iman (1987) Model

Parameter	Distribution	Mean	Error Factor I	Error Factor II
X_1	lognormal	2	2	6
X_2	lognormal	3	2	6
X_3	lognormal	1×10^{-3}	2	6
X_4	lognormal	2×10^{-3}	2	6
X_5	lognormal	4×10^{-3}	2	6
X_6	lognormal	5×10^{-3}	2	6
X_7	lognormal	3×10^{-3}	2	6

We present two cases, a reference case which is the same as in Chun *et al.* (2000) and a case of increased uncertainty. We denote the two cases with the symbols I and II, respectively. Table IV reports the input distributions[8]

The result of uncertainty propagation with a sample of size $N = 1000$ for Case I are displayed in Fig. 4. Fig. 4 shows that $f_Y(y)$ is lognormal, with mean equal to 2.32×10^{-6} and error factor equal to 2.32. The sample size has been increased to 10,000 in Case II, to cope with the increased uncertainty. The corresponding output density is again fitted by a lognormal shape, but with mean equal to 2.12×10^{-6} and error factor equal to 9.94.

Table V reports the values of the nonparametric uncertainty importance indicators in the two cases and the corresponding ranking (in parenthesis) obtained with the software SIMLAB (see, Reference 41) The most important parameter is X_2 in all cases

[8] The chosen distributions coincide with the ones selected in Chun *et al.* (2000). As often happens in risk analysis, the lognormal distribution

$$f_X(x) = \frac{1}{\xi X \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(X)-\eta}{\xi}\right]^2},$$

is expressed through its mean and error factor. The error factor is given by the ratio between the median and the 5th percentile of the distribution, which, in the lognormal case, also equals the ratio of the 95th percentile to the mean. From knowledge of the mean and error factor, one can then derive the two parameters (η, ξ) of the lognormal distribution from

$$\begin{cases} \xi = \frac{\ln(\text{Error Factor})}{1.645} \\ \eta = \ln(\text{Mean}) - \frac{\xi^2}{2} \end{cases}.$$

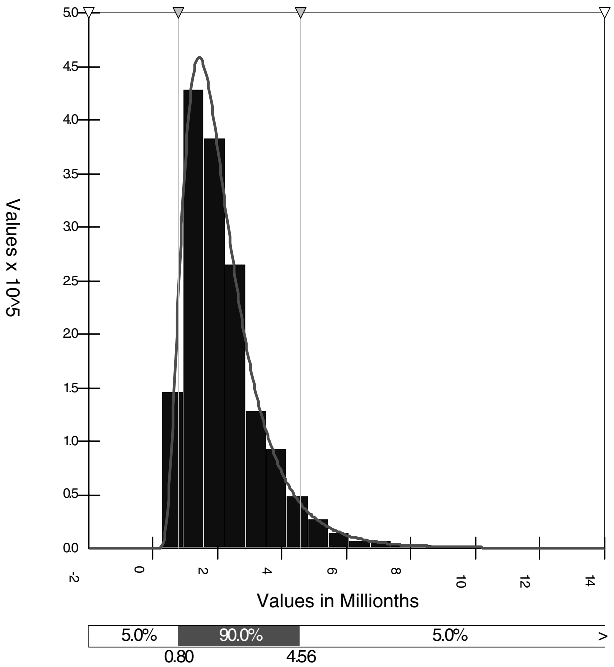

Fig. 4. Case I. Density of the top event frequency, $f_Y(y)$, as a result of uncertainty analysis for the Iman (1987) model.

and for all indicators, followed by X_6 and X_5, which rank second and third with all measures in all cases. X_4 ranks fourth in all cases. X_1, X_3, and X_7 are the low-ranked parameters.

The coefficient of model determination, R_Y^2, is equal to 0.92 in Case I and to 0.46 in Case II. This implies that the model is well fitted by a linear regression in Case I, while the fit is less accurate in Case II. As a consequence, a lower portion of model variance is explained by the liner regression in Case II. One also ought to expect that interaction terms play a more relevant role in this case. Let us now examine how the results obtained using $\mathrm{I}_l/\mathrm{S1}_l$ and ST_l confirm these observations. Table VI reports importance

Table V. Nonparametric Uncertainty Importance Measures for the Iman (1987) Model

	PEAR_l I	PCC_l I	SPEA_l I	PEAR_l II	PCC_l II	SPEA_l II
X_1	0.19 (6)	0.51 (5)	0.22 (5)	0.14 (5)	0.18 (6)	0.23 (5)
X_2	0.59 (1)	0.90 (1)	0.60 (1)	0.41 (1)	0.49 (1)	0.52 (1)
X_3	0.09 (7)	0.39 (6)	0.10 (7)	0.11 (6)	0.15 (7)	0.14 (7)
X_4	0.35 (4)	0.72 (4)	0.35 (4)	0.20 (4)	0.27 (4)	0.33 (4)
X_5	0.35 (3)	0.79 (3)	0.38 (3)	0.30 (3)	0.38 (3)	0.40 (3)
X_6	0.50 (2)	0.86 (2)	0.49 (2)	0.35 (2)	0.43 (2)	0.43 (2)
X_7	0.21 (5)	0.63 (7)	0.20 (6)	0.10 (7)	0.15 (5)	0.21 (6)

Table VI. Results for the Iman and Sobol Uncertainty Importance Measures

	$\mathrm{I}_l/\mathrm{S1}_l$ I	ST_l I	$\mathrm{I}_l/\mathrm{S1}_l$ II	ST_l II
X_1	0.03 (5)	0.09 (5)	0.02 (6)	0.11 (5)
X_2	0.37 (1)	0.55 (1)	0.25 (1)	0.50 (1)
X_3	0.02 (7)	0.06 (7)	0.01 (7)	0.09 (6)
X_4	0.07 (4)	0.13 (4)	0.03 (5)	0.19 (4)
X_5	0.14 (3)	0.23 (3)	0.06 (3)	0.36 (2)
X_6	0.18 (2)	0.27 (2)	0.15 (2)	0.26 (3)
X_7	0.03 (6)	0.07 (6)	0.05 (4)	0.09 (7)

measures and ranking (in parenthesis) obtained with $\mathrm{I}_l/\mathrm{S1}_l$ and ST_l.

Table VI shows that I_l and ST_l produce the same ranking in Case I, while their ranking differs in Case II. In Case I, the sum of the first-order indices is around 94%, indicating that almost all of the model variance is explained by individual effects. This is in agreement with the estimated high value of the coefficient of model determination testifying validity of the linear regression. More precisely, we recall that in this case Equation (6) holds and variance is the sum of individual contributions. This means that interaction terms play a minor role, and their inclusion in $\mathrm{S1}_l$ does not affect raking. The ranking between $\mathrm{S1}_l$ and ST_l, however, differs in Case II. In this case, $\sum_{i=1}^{7} \mathrm{S1}_i = 57\%$ signalling that a higher portion of the model variance is explained by interaction terms. We also recall the low value of the model coefficient of determination in this case ($R_Y^2 = 0.46$). This means that with increased uncertainty interaction terms start playing a relevant role for this model, and ought not to be neglected.

We now come to the comparison of the ranking obtained with variance-based indicators to the ranking obtained with nonparametric techniques. We do so by computing two measures of ranking agreement: the correlation coefficients on the (raw) input ranks and on their savage scores—see Campolongo and Saltelli (1997) or Saltelli and Marivoet (1990) for the definition Savage scores. The difference between the two correlations coefficients lies in the fact that, as Campolongo and Saltelli (1997) underline, Savage scores place emphasis on the agreement of top-ranked variables, while "*disagreements on the exact ranking of low-ranked variables*" are not revealed by correlations on Savage scores. Fig. 5 displays the correlation coefficients on the ranking (left tables) and on the corresponding Savage scores (right tables) in Cases I (above) and II (below).

Ranks	*PEAR I*	*PCC I*	*SPEA I*	*S1*	*ST I*
PEAR I	1				
PCC I	0.857	1			
SPEA I	0.929	0.964	1		
S1 I	0.929	0.964	1	1	
ST I	0.929	0.964	1	1	1

Savage	*PEAR I*	*PCC I*	*SPEA I*	*S1 I*	*ST I*
PEAR I	1				
PCC I	0.952	1			
SPEA I	0.966	0.994	1		
S1 I	0.966	0.994	1	1	
ST I	0.966	0.994	1	1	1

Ranks	*PEAR II*	*PCC II*	*SPEA II*	*S1 II*	*ST II*
PEAR II	1				
PCC II	0.893	1			
SPEA II	0.964	0.964	1		
S1 II	0.786	0.964	0.893	1	
ST II	0.964	0.857	0.929	0.75	1

Savage	*PEAR II*	*PCC II*	*SPEA II*	*S1 II*	*ST II*
PEAR II	1				
PCC II	0.977	1			
SPEA II	0.994	0.991	1		
S1 II	0.942	0.986	0.965	1	
ST II	0.943	0.920	0.937	0.885	1

Fig. 5. Correlation coefficients on input ranks (left) and Savage scores (right) for variance based and nonparametric uncertainty importance measures in Cases I and II.

Fig. 5 (left tables) shows that the overall ranking agreement is high in Case I. There is one case in which ranking agreement is perfect, namely between SPEA_I and S1_I and ST_I. The high values of the Savage scores correlation coefficients signal agreement between nonparametric and variance-based techniques in identifying the relevant parameters. Ranking discrepancies increase in Case II, when the model becomes less linear. As we mentioned, the mathematical reason of the higher disagreement is the worsening of the linear regression approximation in Case II. However, nonparametric and variance-based technique indications need not coincide since the indicators are constructed so as to look at different output properties, as discussed at the end of Section 4.

We are now left with the exploration of the results obtained using δ_I. We recall that making use of a moment-independent approach is equivalent to measuring uncertainty importance with reference to the entire output distribution (Table I). Table VII reports the numerical value of δ_I and the corresponding parameter ranking (in parenthesis) in the two cases.

Table VII shows that X_6 is the parameter associated with the greatest value of δ_I both in Cases I and II, followed by X_5, X_2, and X_4. X_3 is the least relevant parameter in both cases. It is worth pointing out that the sum of the deltas practically equals unity in both cases: $\sum_{i=1}^{7} \delta_i \simeq 1$. We recall that, since Property 3 (Table III) states that the joint importance of all inputs equals 1, there follows that, in this case, $\delta_{1,2,\ldots,7} \simeq \delta_1 + \delta_2 + \cdots + \delta_7$. In terms of Property 5 this means that $\delta_{ij} = \delta_i + \delta_j$, or $\delta_{j|i} = \delta_j$, i.e., the effect on Y of the uncertainty in X_j is independent of the effect of X_i. Borgonovo (2005) refers to this effect as "separability" of the uncertainty in X_i and X_j w.r.t. the uncertainty in Y.[9]

Table VII. Results for the Delta Importance Measure in Cases I and II

Parameter	δ_I I	δ_I II
X_1	0.11 (6)	0.11 (5)
X_2	0.17 (3)	0.16 (3)
X_3	0.09 (7)	0.09 (7)
X_4	0.13 (4)	0.14 (4)
X_5	0.18 (2)	0.17 (2)
X_6	0.20 (1)	0.18 (1)
X_7	0.11 (5)	0.10 (6)

We can then come to the comparison of the results of δ to the results of nonparametric and variance-based indicators. Fig. 6 (left) shows that the correlation coefficients on ranks vary from 0.82 to 0.93 in Case I and from 0.79 to 0.96, an indication of an overall ranking agreement, with better agreement in the case of low uncertainty. Let us then analyze this result further starting with low-ranked inputs. X_3 is ranked 7th or 6th by the measures reported in Table VII; X_1 is ranked either 5th or 6th; X_7 is ranked 5th or 6th, with the exception of Case II, in which it ranks 4th according to I_I; X_4 ranks either 4th or 5th; by all methods. One can then conclude that the methods agree in identifying the nonrelevant parameters. Let us now turn to the top-ranked parameters. The values of the correlation coefficients on Savage scores vary from 0.59 to 0.97 in Case I and from 0.48 to 0.94 in Case II (Fig. 6, right). By the meaning of Savage scores, these **Q11**
values signal that lower agreement is registered for the ranking of the most relevant inputs than for the ranking of the less significant ones.

We finally highlight the information an analyst would derive from a joint utilization of the techniques: (1) X_3, X_1, X_7, and X_4 can be considered as nonrelevant on output uncertainty, since they are the least influential when their effect on the entire output distribution (δ_I) or on its variance (I_I, ST_i) or input/output correlation (PEAR_I) are considered; (2) the model is well fitted by a linear regression in Case I, while interaction terms are relevant in Case II; (3) the most

[9] Exploring whether there are systematc conditions that assure that separability holds (i.e., $\delta_{1,2,\ldots,n} = \delta_1 + \delta_2 + \cdots + \delta_n$) and whether, under the same conditions, additivity holds (i.e., ($V = \sum_{i=1}^{n} V_i$) is subject of further investigations by the author.

Ranks	Delta I	S1 I	ST I	PEAR I
Delta I	1			
S1 I	0.86	1		
ST I	0.86	1	1	
PEAR I	0.82	0.93	0.93	1

Ranks	Delta II	S1 II	ST II	PEAR II
Delta II	1			
S1 II	0.79	1		
ST II	0.82	0.75	1	
PEAR II	0.86	0.79	0.96	1

Savage	Delta I	S1 I	ST I	PEAR I
Delta I	1			
S1 I	0.59	1		
ST I	0.59	1	1	
PEAR I	0.54	0.97	0.97	1

Savage	Delta II	S1 II	ST II	PEAR II
Delta II	1			
S1 II	0.57	1		
ST II	0.48	0.89	1	
PEAR II	0.60	0.94	0.94	1

Fig. 6. Correlations among the parameter ranking and Savage scores obtained with δ, I/S1, ST, and PEAR.

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. This result implies that the parameter influencing variance the most is not necessarily the parameter that influences the entire output distribution the most; and (4) The ranking difference ought to be attributed to the different meaning of the importance indicators, as explained at the end of Section 4.

6. CONCLUSIONS

In this work, we have investigated the role of global SA in indicating which among a set of uncertain inputs influences output uncertainty the most. To do so, we have compared different methods available in the literature by studying their definitions and properties. We have started with nonparametric techniques, i.e., SA methods that measure uncertainty importance based on input–output correlation. We have analyzed the assumptions on the basis of their utilization. We have illustrated how the linear approximation of variance utilized in Helton (1993) links the two approaches. We have then discussed the properties and meaning of variance-based techniques highlighting their formal definition and the related mathematical assumptions. We have seen that they convey a twofold information: (i) in the absence of correlations among the parameters, structural information is gained due to the fact that output variance decomposition is a direct reflection of the model function decomposition; (ii) uncertainty information, in the sense that parameters which influence output variance the most are identified. This information is obtained both in the presence or absence of correlations.

With this respect, we have investigated the validity of the assumption that variance "*is sufficient to describe output variability* (Saltelli, 2002)" After presenting standard results of utility theory, we have utilized a simple example to highlight that variance alone does not always measure the overall decision maker uncertainty on the model output.

We have then examined the use of moment-independent uncertainty importance measures, i.e., indicators that do not refer to a particular moment of the output distribution. We have made us of an indicator introduced in Borgonovo (2005), denoted as δ. After providing the formal definition, we have summarized its properties and interpretation. We have seen that its definition is well posed in the presence of correlations among the parameters, since one needs just to specify a joint input distribution without requiring independence.

We have obtained numerical results and quantitatively discussed the comparison of the various techniques by means of the risk assessment model utilized in Iman (1987), where uncertainty importance measures were introduced for the first time. We have analyzed two cases, a case of low and high uncertainty.

Results of both cases show that the indicators agree in identifying the less relevant factors. Discrepancies in the top-ranked inputs reveal that parameters influencing variance the most are not necessarily the ones that influence the entire output distribution the most.

The analysis has also shown that a modeler would mostly benefit from a joint utilization of the techniques. In fact, using nonparametric techniques and measuring the value of the model coefficient of determination, one gains information on the linearity of the model. If it holds, then one knows that variance-based indicators will produce results close to those of nonparametric techniques. Computing Sobol global sensitivity indices would lead the decision maker to obtain information on the model structure in the absence of correlations among the inputs. Finally, utilizing a moment-independent approach (δ) enables analysts to measure the influence of an uncertain input on the decision maker uncertainty.

ACKNOWLEDGMENTS

The author whishes to thank the anonymous referees for the very insightful and perceptive comments. A special thanks also to Marco Bonetti and Piero Veronese. Financial Support from Bocconi University is gratefully acknowledged.

REFERENCES

Alis, O. F., & Rabitz, H. (2001). Efficient computation of high dimensional model representations. *Journal of Mathematical Chemistry*, *29*(2), 127–142.

Andsten, R. S., & Vaurio, J. K. (1992). Sensitivity, uncertainty and importance analysis of a risk assessment. *Nuclear Technology*, *98*, 160–170.

Apostolakis, G. E. (1995). A Commentary on model uncertainty, *Proceedings of the Workshop on Model Uncertainty: Its Characterization and Quantification*, Center for Reliability Engineering, University of Maryland, College Park, Maryland, USA.

Apostolakis, G. E. (2005). How useful is quantitative risk assessment. *Risk Analysis*, *24*(3), 515–520.

Bedford, T. (1998). Sensitivity indices for (tree)- dependent variables. *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output*, Venice (Italy), 17–20.

Bier, V. M. (1983). A measure of uncertainty importance for components in fault trees, *Transactions of the American Nuclear Society*, *45*(1), 384–5.

Q12 Borgonovo, E. (2005). A new uncertainty importance measure. *Reliability Engineering and System Safety* (In press) 2006.

Campolongo, F., & Saltelli, A. (1997). Sensitivity analysis of an environmental model: An application of different analysis methods. *Reliability Engineering and System Safety*, *57*, 49–69.

Chun, M.-H., Han, S.-J., & 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*, 313–321.

Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petsheck, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. *Journal of Chemical Physics*, *59*, 3873–3878.

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives, *Annales de l'Institute Henri Poincarè*, Vol. 7, 1–68. Translated to English by H.E. Kyburg and reprinted in Kyburg and Smokler (1964).

Efron, B., & Stein, C. (1981). The Jackknife estimate of variance, *The Annals of Statistics*, *9*(3), 586–596.

Frey, C. H. (2002). Introduction to special section on sensitivity analysis and summary of NCSU/USDA Workshop on Sensitivity Analysis. *Risk Analysis*, *22*(3), 539–545.

Frey, C. H., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, *22*(3), 553–571.

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, *42*, 327–367.

Homma, T., & Saltelli, A. (1995). Use of Sobol quasirandom sequence generator for integration of modified uncertainty importance measure, *Journal of Nuclear Science and Technology*, *32*(11), 1164–1173.

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, *52*, 1–17.

Huang, C.-F., & Litzenberger, R. H. (1988). Foundations for financial economics, Prentice Hall, Upper Saddle River, NJ, USA, 365 pages, ISBN 0-13-500653-8.

Iman, R. L. (1987). A matrix-based approach to uncertainty and sensitivity analysis for fault trees. *Risk Analysis*, *7*(1), 21–33.

Iman, R. L., & Conover, W. J. (1987). A measure of top-down correlation. *Technometrics*, *29*(3), 351–357.

Iman, R. L., & Hora, S. C. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, *10*(3), 401–406.

Iman, R. L., Johnson, M. E., & Watson, C. C. Jr. (2005a). Sensitivity analysis for computer model projections of hurricane losses. *Risk Analysis*, *25*(5), 1277–1297.

Iman, R. L., Johnson, M. E., & Watson, C. C. Jr. (2005b). Uncertainty analysis for computer model projections of hurricane losses. *Risk Analysis*, *25*(5), 1299–1312.

Malliavin, P. (1995). *Integration and Probability*, New York: Springer Verlag Editions, ISBN 0-387-94409-5.

McKay, M. D. (1996). Variance-based methods for assessing uncertainty importance, in NUREG-1150 Analyses, LA-UR-96-2695, 1–27.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments, *Technometrics*, *33*(2), 161–174.

Nakashima, K., & Yamato, K. (1982). Variance-importance of system components. *IEEE Transactions on Reliability*; *R–31*(1). Q13

Oakley, J. E., & 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), 751–769.

Park, C. K., & Ahn, K. I. (1994). A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliability Engineering and System Safety*, *46*, 253–261.

Patil, S. R., & Frey, C. H. (2004). Comparison of sensitivity analysis methods based on application to a food safety risk assessment model. *Risk Analysis*, *24*(3), 573–585.

Rabitz, H., Alis, O. F., Shorter, J., & Shim, K. (1998). Efficient input output model representation. *Computer Physics Communications*, *117*, 11–20.

Rabitz, H., & Alis, Ö. F. (1999). General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, *25*(2-3), 197–233.

Saltelli, A., & 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.

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*, 1–4. Q14

Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, *41*(1), 39–56.

Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modelling. *Statistical Science*, *19*(4), 377–395.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, *22*(3), p. 579.

Saltelli, A., & 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), 702–709. Q15

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley, & Sons publishers. Q16

Savage, L. (1972). *The Foundations of Statistics*, Dover Publications, 2nd ed. Q17

SIMLAB Reference Manual, POLIS- JRC ISIS, Joint Research Center of the European Cummunity. Q18

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. Q19

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1), 271–280.

Sobol, I. M. (2003). Theorem and examples on high-dimensional model representation. *Reliability Engineering and System Safety*, *79*, 187–193.

QUERIES

Q1 Author: Please specify whether Iman et al. 2005a or 2005b is being referred to in the sentence “Their application range . . . ” in Abstract. We need to put journal name along with volume and page no as per the style.

Q2 Author: Reference “Cuckier (1973)” has not been included in the reference list. Please check.

Q3 Author: Please specify whether Iman *et al.* 2005a or 2005b is intended in the sentence “Following the name interpretation . . . ”

Q4 Author: Please check whether the edit in the sentence “As Oakley and O’Hagan (2004). . .” retains the intended sense.

Q5 Author: Please check whether the edit the sentence “In fact by utilizing . . . ” retains the intended sense.

Q6 Author: Please check whether the edit in the sentence “As Saltelli and Marivoet (1990). . . ” retains the intended sense.

Q7 Author: Reference “Rabitz et al. (1999)” has not been included in the reference list. Please check.

Q8 Author: Please check the sentence “Hence, Helton (1993) . . . ” for clarity.

Q9 Author: Please check whether the edit in the sentence “ . . . represent the expected percentage. . . ” retains the intended sense.

Q10 Author: Can the details (chapter, page range) of reference “Huang & Litzenberger (1988)” be included in the reference list.

Q11 Author: Please check the sentence “By the meaning. . . ” for clarity.

Q12 Author: Please update reference “Borgonovo (2005)”.

Q13 Author: Please provide the page range if applicable in reference “Nakashima and Yamato (1982)”.

Q14 Author: Reference “Saltelli (1997)” has not been cited in the text. Please check.

Q15 Author: Please provide the page range if applicable in reference “Saltelli (2002)”.

Q16 Author: Please provide the location details (city, state) for publisher in reference “Saltelli et al. (2004)”.

Q17 Author: Please provide the location (city, state) details of reference Savage (1972).

Q18 Author: Please provide the location (city, state) details in reference SIMLAB.

Q19 Author: Please provide the fullname of journal in reference “Sobol (1993)”.