Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software 22 (2007) 1509–1518
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

An effective screening design for sensitivity analysis of large models

Francesca Campolongo [a], Jessica Cariboni [a,b,*], Andrea Saltelli [a]

[a] European Commission, Joint Research Centre, IPSC, Via Fermi 1, 21020 Ispra, Varese, Italy
[b] K.U.Leuven, U.C.S., W. De Croylaan 54, B-3001 Leuven, Belgium

Received 17 January 2006; received in revised form 11 October 2006; accepted 15 October 2006
Available online 18 January 2007

Abstract

In 1991 Morris proposed an effective screening sensitivity measure to identify the few important factors in models with many factors. The method is based on computing for each input a number of incremental ratios, namely elementary effects, which are then averaged to assess the overall importance of the input. Despite its value, the method is still rarely used and instead local analyses varying one factor at a time around a baseline point are usually employed.

In this piece of work we propose a revised version of the elementary effects method, improved in terms of both the definition of the measure and the sampling strategy. In the present form the method shares many of the positive qualities of the variance-based techniques, having the advantage of a lower computational cost, as demonstrated by the analytical examples.

The method is employed to assess the sensitivity of a chemical reaction model for dimethylsulphide (DMS), a gas involved in climate change. Results of the sensitivity analysis open up the ground for model reconsideration: some model components may need a more thorough modelling effort while some others may need to be simplified.

Keywords: Sensitivity analysis; Screening problem; Model-free methods; Effective sampling strategy; Dimethylsulphide (DMS)

1. Introduction

The use of models to inform and support the decision-making process is becoming extremely important nowadays. Whatever the type of modelling process, certain common steps must be followed if the goal is to obtain credible results and valuable information (see, e.g. Jakeman et al., 2006 for a list of recommended steps in the development of environmental models). Uncertainty and sensitivity analysis are important steps in the model building process. Different approaches can be followed to test the sensitivity of a model (Saltelli et al., 2004, 2005; Cacuci and Ionesco-Bujor, 2004). Sensitivity methods range from the quantitative variance-based methods, defined from the decomposition of the total output variance into the contributions of the input factors, to other forms of global sensitivity with regional properties (Pappenberger et al., 2006), down to the simplest class of the One Factor At a Time (OAT) screening techniques, which simply vary one factor at a time and measure the variation in the output. In contrast to the OAT, global methods require a high number of model evaluations, increasing with the number of factors. In 1991 Morris proposed a method, which is particularly well-suited when the number of uncertain factors is high and/or the model is expensive to compute. The method is based on calculating for each input a number of incremental ratios, called Elementary Effects (EE), from which basic statistics are computed to derive sensitivity information. While the EE method was proven to be a very good compromise between accuracy and efficiency, especially for sensitivity

* Corresponding author. Institute for the Protection and Security of the Citizen (IPSC), ESAF European Commission, Joint Research Centre, TP 361, 21020 Ispra, VA, Italy. Tel.: +39 0332 789372; fax: +39 0332 785733.
E-mail address: jessica.cariboni@jrc.it (J. Cariboni).

1364-8152/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2006.10.004

analysis of large models (see, for instance, Campolongo and Braddock, 1999), it is still not extensively used. In this paper we aim to recall the attention of the modelling community to the effectiveness of this method and to enhance its efficiency. We propose a revised measure which increases its interpretability in the case of complex models with multiple inputs and outputs and which allows for its applicability to groups of factors. We propose a refined sampling strategy that allows for a better exploration of the space of the input factors. We then employ the method to assess the sensitivity of a large chemical model to its input factors.

Section 2 describes the EE screening sensitivity measure as originally proposed. In Section 3, we present the revised version of the EE measure, show how to apply it to groups of factors and describe the improved sampling strategy. Section 4 investigates by means of analytical examples the relationship between the revised EE method and the variance-based measures. Section 5 illustrates the sensitivity of a model relevant to climate change studies, which describes the tropospheric air and droplet chemistry for dimethylsulphide (DMS). Section 6 contains our conclusions.

2. The elementary effects method

The guiding philosophy of the original EE method (Morris, 1991) is to determine which input factors may be considered to have effects which are (a) negligible, (b) linear and additive, or (c) non-linear or involved in interactions with other factors. For each input, two sensitivity measures are computed: μ, which assesses the overall influence of the factor on the output, and σ, which estimates the ensemble of the factor's higher order effects, i.e. non-linear and/or due to interactions with other factors. The experimental plan is composed of individually randomised OAT experiments. Each model input $X_i, i = 1,\ldots,k$, is assumed to vary across p selected levels in the space of the input factors. The region of experimentation Ω is thus a k-dimensional p-level grid.

Following a standard practice in sensitivity analysis, factors are assumed to be uniformly distributed in [0,1] and then transformed from the unit hypercube to their actual distributions.

For a given value of $\mathbf{X}$, the elementary effect of the ith input factor is defined as:

$$d_i(\mathbf{X}) = \left(\frac{y(X_1,\ldots,X_{i-1},X_i+\Delta,X_{i+1},\ldots,X_k) - y(\mathbf{X})}{\Delta}\right)$$

where Δ is a value in $\{1/(p-1),\ldots,1-1/(p-1)\}$, p is the number of levels, $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ is any selected value in Ω such that the transformed point $(\mathbf{X} + \mathbf{e}_i\Delta)$ is still in Ω for each index $i = 1,\ldots,k$, and $\mathbf{e}_i$ is a vector of zeros, but with a unit as its ith component. The finite distribution of elementary effects associated with the ith input factor, is obtained by randomly sampling different $\mathbf{X}$ from Ω, and is denoted by F_i, i.e. $d_i(\mathbf{X}) \sim F_i$. The number of elements of each F_i is $p^{k-1}[p-\Delta(p-1)]$. A convenient choice for the parameters p and Δ is p even and Δ equal to $p/[2(p-1)]$. This choice has the advantage that, although the design sampling strategy does not guarantee equal-probability sampling from each F_i, at least a certain symmetric treatment of inputs is ensured, which may be desirable (for details see Morris, 1991).

The sensitivity measures proposed by Morris (1991), μ and σ, are respectively the mean and the standard deviation of the distribution F_i. To estimate these quantities, Morris suggests sampling r elementary effects from each F_i via an efficient design that constructs r trajectories of $(k+1)$ points in the input space, each providing k elementary effects, one per input factor. The total cost of the experiment is thus $r(k+1)$.

A comparison between the performances of the EE measures and variance-based ones was presented in Campolongo and Saltelli (1997). The authors demonstrated that the overall performance of the measure μ in ranking the factors in order of importance is satisfactory. However, in some instances μ could be prone to Type II error, i.e. failing in the identification of a factor of considerable influence on the model, being instead more resilient to Type I error, i.e. the erroneous identification of a factor as influential when it is not.[1] If the distribution F_i contains both positive and negative elements, i.e. if the model is non-monotonic, some effects may cancel each other out, thus producing a low μ value even for an important factor. To properly rank factors one must consider the values of both μ and σ at the same time, since a factor with elementary effects of different signs (i.e. which cancel each other out) would have a low value of μ but a considerable value of σ.

3. Enhancing the method

3.1. Improving the sampling strategy

The EE method is based on the construction of r trajectories in the input space, typically between 10 and 50. The design is based on generating a random starting point for each trajectory and then completing it by moving one factor at a time in a random order. This strategy could lead to a non-optimal coverage of the input space, especially for models with a large number of input factors.

We propose an improvement of the sampling strategy, which aims at a better scanning of the input domain without increasing the number of model executions needed. The idea is to select the r trajectories in such a way as to maximise their dispersion in the input space. We start by generating a high number of different Morris trajectories, $M \sim 500-1000$, and then we choose, for example, $r = 10$ with the highest 'spread', where the concept of 'spread' is based on the following definition of 'distance.' The distance d_{ml} between a couple of trajectories m and l is defined as:

$$d_{ml} = \begin{cases} \displaystyle\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\sqrt{\sum_{z=1}^{k}\left[X_i^m(z) - X_j^l(z)\right]^2} & \text{for } m \neq l \\ 0 & \text{otherwise} \end{cases}$$

[1] The null hypothesis is that a factor is non-influential, since the goal of Morris design is the screening of unimportant factors.

where k is the number of input factors and $X_i^m(z)$ indicates the zth coordinate of the ith point of the mth Morris trajectory. In other words d_{ml} is the sum of the geometric distances between all the couples of points of the two fixed trajectories.[2] Note that the use of the Euclidean distance d_{ml} is totally arbitrary. As suggested by an anonymous reviewer, the use of alternative metrics, such as for instance the Manhattan one, should be investigated. This issue is left for further research.

The best r trajectories out of M are selected by maximising the distance d_{ml} among them. A brute force approach is used.[3] First we consider for each possible combination of r trajectories out of M the quantity D, which is the sum of all the distances d_{ml} between couples of trajectories belonging to the combination. For instance, if we select the combination 4, 6, 7, and 9 (i.e. $r=4$) out of the possible $M=\{1,2,3,4,5,6,7,8,9,10\}$, we define $D_{4,6,7,9}$ as $D_{4,6,7,9} = \sqrt{d_{4,6}^2 + d_{4,7}^2 + d_{4,9}^2 + d_{6,7}^2 + d_{6,9}^2 + d_{7,9}^2}$. Then we consider the combination with the highest value of D.

Fig. 1 compares the empirical distributions (sampled values) for $k=4$ factors drawn respectively via the original sampling strategy (a), or via the revised strategy (b) from theoretical distributions that are discrete uniform with $p=4$ levels, i.e. DU{0,1/3,2/3,1}. $r=20$ trajectories are employed. The empirical distributions shown in Fig. 1b are manifestly closer than in Fig. 1a to the theoretical shape of the distributions (discrete uniform).

Note that the original sampling strategy proposed by Morris can be enhanced by using as starting points for the trajectories a group of points generated with Latin Hypercube Sampling (LHS) rather than with random sampling. However, the advantage of using LHS with the new strategy is nullified by the optimisation procedure, which by itself is sufficient to ensure a good scan of the inputs' space.

The sampling strategy proposed in this section is always to be preferred to the original one proposed by Morris (1991), as it offers a better scan of the input space without increasing the number of model evaluations required. Unless specified differently, this new strategy will be employed in all the examples discussed in the following.

3.2. Refining the measure μ

When a model is complex and presents several inputs and several outputs (as with the KIM model analysed in this work), the simultaneous use of the two sensitivity measures μ and σ may be inefficient. We propose here a revised version of the measure μ, that we call μ^*, which on its own is sufficient to provide a reliable ranking. μ^* is the estimate of the mean of the distribution of the absolute values of the elementary effects that we denote with G_i, e.g. $|d_i(\mathbf{X})| \sim G_i$. The use of μ^* solves the problem of the effects of opposite signs which occurs when the model is non-monotonic. The drawback is the loss of information on the sign of the effect. Nevertheless, this information can be recovered by the simultaneous examination of μ and μ^*, as an estimate of μ comes at no extra computational cost. If μ and μ^* are both high, the sign of the effect is always the same, i.e. the output function is monotonic with

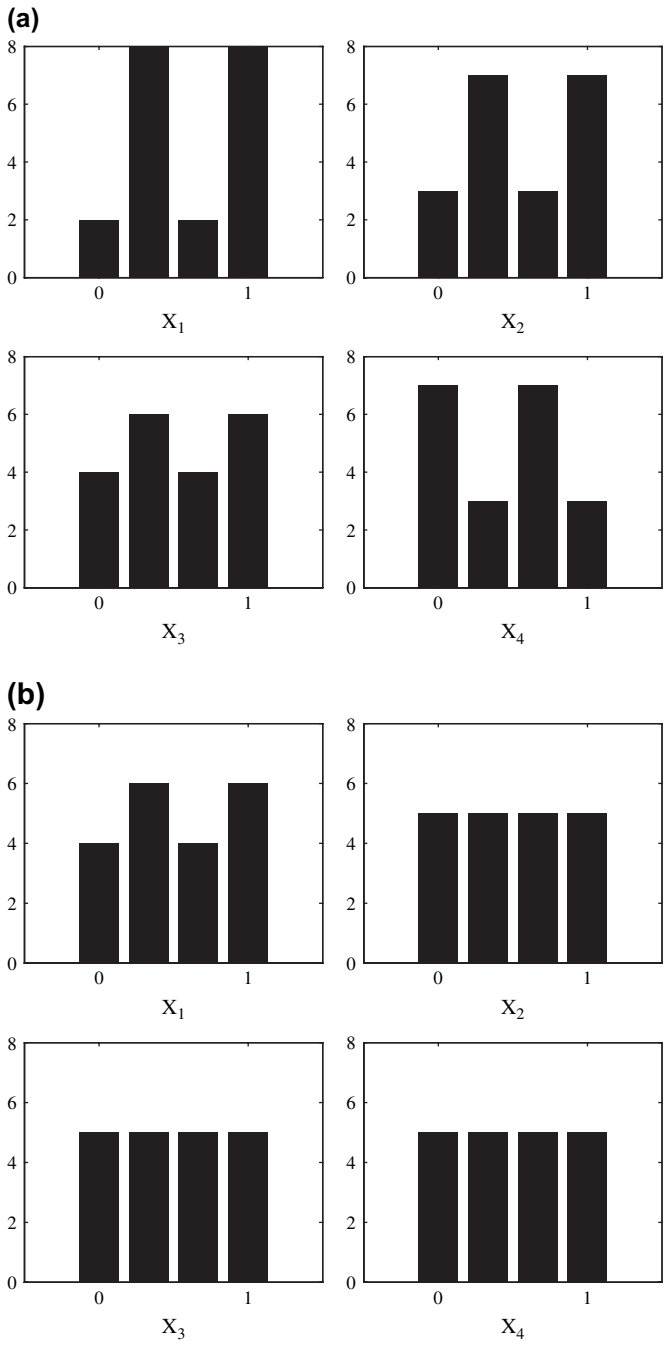

Fig. 1. Empirical distributions (sampled values) for 4 factors, X_1, X_2, X_3, X_4 whose theoretical distributions are uniform discrete with 4 levels, i.e. UD{0,1/3,2/3,1}. In (a) the samples are extracted using the original EE sampling strategy. In (b) the revised sampling strategy is employed. The size of the sample is $r=20$.

[2] As the M trajectories that we use as a starting sample are all different, the distance between two of them is always different from 0. Nevertheless, we have included the concept of "zero" distance to force d to fulfil all the properties of a metric. The space of all possible trajectories with the metric d is thus a metric space.

[3] Future research could be devoted to solving this combinatorial optimisation problem and to identifying more appropriate techniques for generating the optimal set of trajectories.

respect to that factor. If, in contrast, μ is low while μ^* is high, the factor carries the effects of different signs, depending on the values assumed by the other factors.[4]

Note that, alternatively to the absolute value, we could also consider the squared effects. Nevertheless, it has been proved (e.g. Huber, 1981) that the robustness of the mean absolute deviation is stronger than that of the mean square deviation.

To examine the effects due to interactions we still use the σ measure proposed by Morris and consider the standard deviation of the distribution F_i.

3.3. Working with groups

Another advantage of μ^* is that, in contrast to μ, it can be adjusted to work with a group of factors, i.e. to produce an overall sensitivity measure relative to a group. Fitness to work with groups, which is typical of the variance-based measures (Saltelli et al., 2000), is very advantageous for models containing a high number of factors (hundreds or thousands), as it allows for the reduction of the number of model executions required, at the cost of losing information on the relative strength of the inputs belonging to the same group.

When working with groups, the idea is to move all factors of the same group simultaneously. In the original definition given by Morris, the elementary effect is obtained by subtracting the function evaluated at $\mathbf{X}$ from that evaluated after incrementing one factor. This definition could not be extended straightforwardly to cases in which more than one factor are moved at the same time, as two factors may have been changed in opposite directions, i.e. one increased and one decreased by Δ. In contrast, if using μ^*, this problem is overcome, as the focus is not on the elementary effect itself but on its absolute value. For a group $\mathbf{u} = (X_{i_1}, X_{i_2})$, the absolute elementary effect in point $\mathbf{X}$ is:

$$|d_{\mathbf{u}}(\mathbf{X})| = \frac{\left| y\left(\tilde{\mathbf{X}}\right) - y(\mathbf{X}) \right|}{\Delta}$$

where $\mathbf{X}$ is any selected value in Ω such that the transformed point $\tilde{\mathbf{X}}$ is still in Ω, and each of the components $\tilde{X}_{i_1}, \tilde{X}_{i_2}$ have been either increased or decreased by Δ with respect to X_{i_1}, X_{i_2}.

While the conceptual design is unchanged (we move one group at a time), from a computational point of view the strategy has been slightly modified to allow factors of the same group to be changed simultaneously in opposite directions.

The application of μ^* to groups has been tested on analytical functions commonly used as a benchmark in sensitivity analysis (see, for instance, Pappenberger et al., 2006): the g-function due to Sobol' (1993). In the k-dimensional unit cube (i.e. $0 \leq X_i \leq 1$, all i) the g-function is defined as:

$$g = \prod_{i=1}^{k} g_i(X_i), \text{ where } g_i(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i},$$

and the a_i are parameters, such that $a_i \geq 0$. The values of the a_i determine the relative importance of the X_i: the smaller a_i, the more important X_i. For this function the partial variances, and consequently the sensitivity indices, can be computed analytically (Saltelli and Sobol', 1995).

We consider a g-function with 9 input factors, all uniformly distributed in the range [0,1], and build three test cases where groups have different features. Table 1 shows for each test case the g-function parameters' values, the groups, and the results of the sensitivity experiment, i.e. the analytical values (in the form of the variance-based total sensitivity S_T index defined in Section 4), and the results of the EE exercise with a sample size of $r = 10$.

Results confirm the fitness of this method to handle groups of factors, regardless of the features of the groups' composition. Even in the more complex case study, in which each of the three groups contains a mix of important and unimportant

Table 1
g-Function parameters, groups, and sensitivity analysis results for the three test cases

Test case 1
a_i Parameters' values
$a_1 = 0.02$, $a_2 = 0.03$, $a_3 = 0.05$, $a_4 = 11$, $a_5 = 12.5$, $a_6 = 13$, $a_7 = 34$, $a_8 = 35$, $a_9 = 37$
Groups
$\mathbf{u} = \{X_1, X_2, X_3\}$, $\mathbf{v} = \{X_4, X_5, X_6\}$, $\mathbf{w} = \{X_7, X_8, X_9\}$,
Morris group μ^*
$\mu^*(\mathbf{u}) = 7.948$, $\mu^*(\mathbf{v}) = 1.058$, $\mu^*(\mathbf{w}) = 0.708$
S_T group analytical
$S_T(\mathbf{u}) = 0.995$, $S_T(\mathbf{v}) = 0.010$, $S_T(\mathbf{w}) = 0.001$
Test case 2
a_i Parameters' values
$a_1 = 0.02$, $a_2 = 0.03$, $a_3 = 0.04$, $a_4 = 0.05$, $a_5 = 0.06$, $a_6 = 0.07$, $a_7 = 34$, $a_8 = 35$, $a_9 = 37$
Groups
$\mathbf{u} = \{X_1, X_3, X_5\}$, $\mathbf{v} = \{X_2, X_4, X_6\}$, $\mathbf{w} = \{X_7, X_8, X_9\}$,
Morris group μ^*
$\mu^*(\mathbf{u}) = 42.339$, $\mu^*(\mathbf{v}) = 32.656$, $\mu^*(\mathbf{w}) = 2.735$
S_T group analytical
$S_T(\mathbf{u}) = 0.694$, $S_T(\mathbf{v}) = 0.686$, $S_T(\mathbf{w}) = 0.001$
Test case 3
a_i Parameters' values
$a_1 = 0.02$, $a_2 = 0.03$, $a_3 = 0.05$, $a_4 = 11$, $a_5 = 12.5$, $a_6 = 13$, $a_7 = 34$, $a_8 = 35$, $a_9 = 37$
Groups
$\mathbf{u} = \{X_1, X_4, X_8\}$, $\mathbf{v} = \{X_3, X_5, X_9\}$, $\mathbf{w} = \{X_2, X_6, X_7\}$,
Morris group μ^*
$\mu^*(\mathbf{u}) = 8.108$, $\mu^*(\mathbf{v}) = 7.083$, $\mu^*(\mathbf{w}) = 6.364$
S_T group analytical
$S_T(\mathbf{u}) = 0.436$, $S_T(\mathbf{v}) = 0.393$, $S_T(\mathbf{w}) = 0.429$

The number of trajectories considered is $r = 10$.

[4] The same information on signs could be derived by considering the σ measure proposed by Morris. A high σ associated with a low μ indicates a factor with effects of oscillating signs.

factors, while the estimated ranking is not exact, a similar level of importance among the three groups has been successfully ascertained.

4. μ^* versus the variance-based measures

Let us compare the sensitivity measures μ^* and σ, discussed in the previous sections, with a class of well established sensitivity measures, the variance-based measures (Saltelli et al., 1999), which can be regarded as good practice in sensitivity analysis (Saltelli et al., 2004; Santner et al., 2003). The idea of variance-based measures is that the total output variance $V(Y)$ for a model with k input factors can be decomposed as: $V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \ldots + V_{12\ldots k}$ where $V_i = V(E(Y|X_i))$, $V_{ij} = V(E(Y|X_i, X_j)) - V_i - V_j$, and so on.

The sensitivity measure $S_i = V_i/V(Y)$ estimates the expected fraction of output variance that could be removed if we learnt the true value of X_i. S_i identifies factors that can lead to the greatest reduction of the output variance (Saltelli and Tarantola, 2002). Nevertheless it does not capture the effects due to interactions among factors. The total effect measure $S_{T_i} = E(V(Y|\mathbf{X}_{-i}))/V_Y$(Homma and Saltelli, 1996), where $-i$ indicates all indices but i, accounts instead for all the effects involving factor X_i, i.e. its main effect plus its interactions of any order with all the other factors. S_T recognises 'unimportant' factors, i.e. factors that can be fixed at any given value over their ranges of uncertainty without affecting significantly the output (Saltelli and Tarantola, 2002). The variance decomposition can also be written by taking the factors into subsets, thus facilitating the extension of the measures defined above into groups (Sobol', 1993).

The variance-based indices can be computed using the original method of Sobol' (1993) or the Fourier Amplitude Sensitivity test (Cukier et al., 1973; Cukier et al., 1978) developed in Saltelli et al. (1999). For this work we use the extension of the Sobol' method described in Saltelli (2002), which allows for the computation of S_i and S_{T_i} at a computational cost of $N = W^*(k+1)$, where k is the number of factors and W is related to the Monte Carlo sample size used to estimate each integral in the definition of the indices (see Saltelli, 2002).[5]

To compare μ^* and σ with the variance-based indices, we run some tests on analytical functions and compare the factors' ranking as obtained via μ^* and via S_T.[6] As the computational cost of the two experiments has to be the same, we are obliged to compute μ^* at sample sizes much higher than usual, so as to avoid unreliable estimates of S_T. With such high sample sizes we employ the original EE sampling strategy, since the benefits of the improved one vanish as the sample size increases,[7] while the computational cost for the preparation of the new design may become non-negligible. Afterwards the implications of using the EE method at the usual low sample size, e.g. $r = (10 \sim 20)$, are discussed, and the results obtained with the new more efficient sampling strategy are presented.

4.1. The test case of Morris

The test function proposed by Morris (1991) contains 20 input factors and has the following form:

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{i,j} w_i w_j + \sum_{i<j<l}^{20} \beta_{i,j,l} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{i,j,l,s} w_i w_j w_l w_s$$

where $w_i = 2 \times (X_i - 1/2)$ except for $i = 3,5$, and 7, where $w_i = 2 \times (1.1X_i/(X_i + 0.1) - 1/2)$. Each input factor X_i is supposed to be uniformly distributed in [0,1]. Coefficients with relatively large values are:

$$\beta_i = +20 \;\; i = 1, \ldots 10; \quad \beta_{i,j} = -15 \;\; i,j = 1, \ldots 6;$$

$$\beta_{i,j,l} = -10 \;\; i,j,l = 1, \ldots 5; \quad \beta_{i,j,l,s} = +5 \;\; i,j,l,s = 1, \ldots 4.$$

The remaining first- and second-order coefficients are independently generated from a standard normal distribution. The remaining third- and fourth-order coefficients are set to zero. In Fig. 2 μ^* values are plotted versus S_T. S_T is computed at sample size $N = 21504$, while μ^* is computed at $N = 21504$ (via the original sampling strategy, white diamonds in the

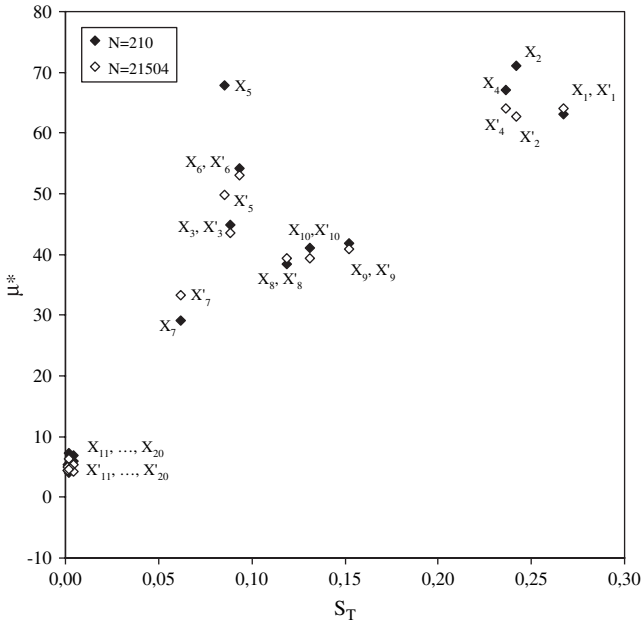

Fig. 2. Plots of μ^* versus S_T for the Morris test case. The Sobol' indices are estimated at the computational cost of $N = 21504$, while the EE values are computed at two different computational costs ($N = 21504$, apostrophed indices for the inputs, and $N = 210$, simple indices for the inputs).

[5] W is at least 100 and can be as high as 10,000 for particularly slow converging models.

[6] Both measures address 'unimportant' factors.

[7] When the number of trajectories generated for the experiment is high enough, the scanning of the input factors' space is already satisfactory, and the benefit achieved by substituting a 'better' combination of trajectories is irrelevant.

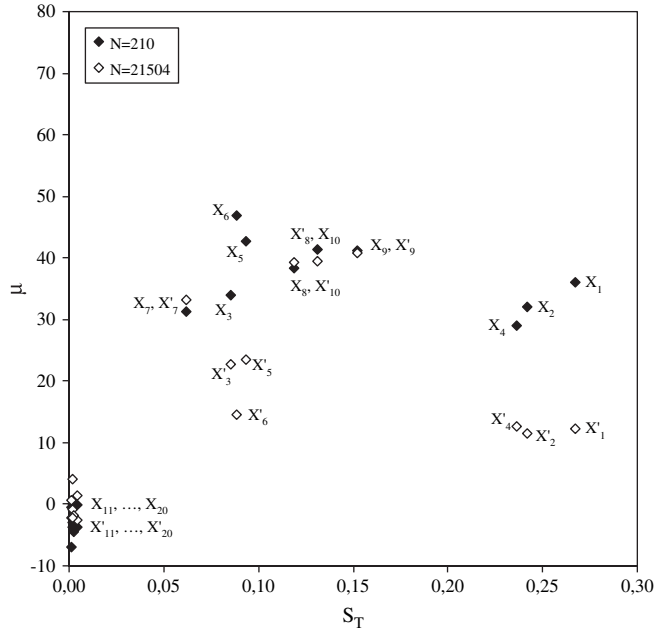

Fig. 3. Plots of μ versus S_T for the Morris test case. The Sobol' indices are estimated at the computational cost of $N = 21504$, while the EE values are computed at two different computational costs ($N = 21504$, apostrophed indices for the inputs, and $N = 210$, simple indices for the inputs).

figure) for comparison purposes only, as well as at $N = 210$ (via the improved sampling strategy, black diamonds in the figure) to reproduce a typical screening experiment. Although the results do not lie on a straight line, the proportionality between the two measures is evident. The low sample size results are encouraging.

Fig. 3 shows that the agreement with the S_T indices worsens if the original version of the EE method is used (measure μ and original sampling design). In this case, the simultaneous use of μ and σ is essential (Fig. 4) to distinguish the most important factors from non-influential ones. However, when

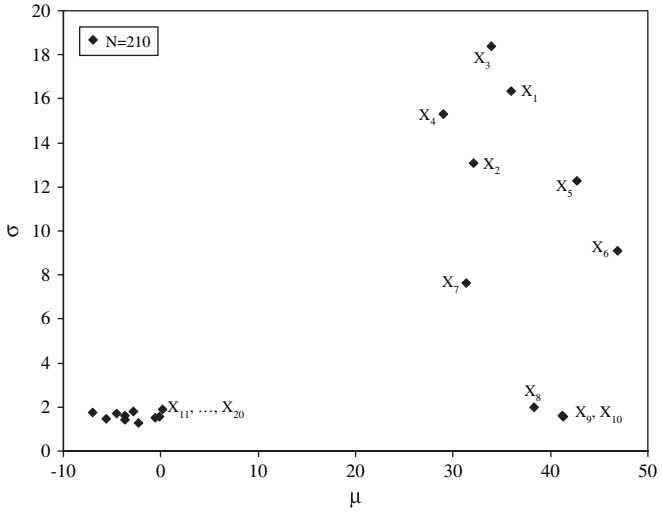

Fig. 4. EE sensitivity test on the Morris function at the computational cost of $N = 210$. Plots of σ versus μ.

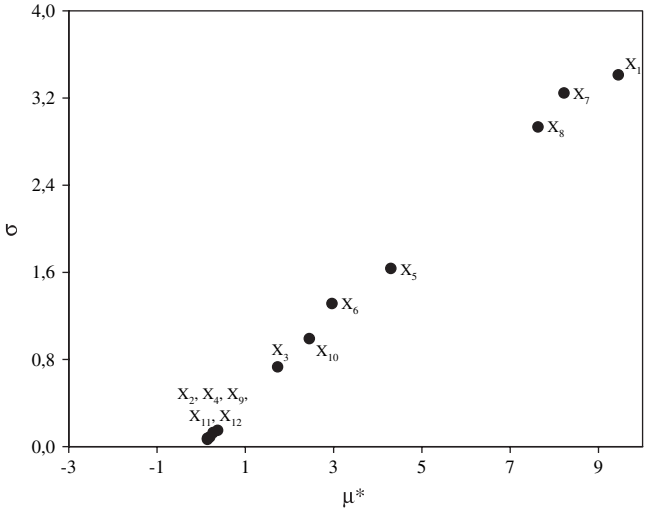

Fig. 5. EE sensitivity test on the g-function at a computational cost of $N = 130$. Plots of σ versus μ^*.

a model has multiple outputs, and several graphs have to be read simultaneously, it may be difficult to draw overall conclusions. The chemical model in Section 5 represents an example of these models, in which the use of μ^* allows the analysis to be simplified.

4.2. The g-function

Let us consider a g-function with 12 input factors all uniformly distributed in [0,1]. The values of the a_i parameters are reported in Table 2. Results are reported in Figs. 5 and 6, which show the improvement obtained in reading results when substituting μ^* for μ. Using the latter, most of the factors are squeezed on μ equal to zero, due to the perfect cancelling of the positive and negative effects. To detect the importance of these factors it is essential to consider the high value of σ, which reveals the deficiencies of μ.

Fig. 7 plots μ^* at sample sizes $N = 6656$ (old sampling strategy, white diamonds) and $N = 130$ (new sampling strategy, black diamonds) versus the analytic values of the S_{T_i}. Despite the non-linearity and non-monotonicity of the function, the EE revised measure performs very well even at very low cost ($N = 130$).

The above comparisons confirm that the EE method, and in particular its refined version with the modulus and the new sampling strategy, effectively substitutes for the calculation of the more expensive S_T. It is confirmed that the EE method is a sound practice and always to be preferred to the many local analyses still found in the literature (see, for instance, Murphy et al., 2004; or Saltelli et al., 2006, for a review). The use of local analyses rather than global may be attributed

Table 2
Values of the parameters used in the computations of the g-function

Factor	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
a_i	0.001	89.9	5.54	42.10	0.78	1.26	0.04	0.79	74.51	4.32	82.51	41.62

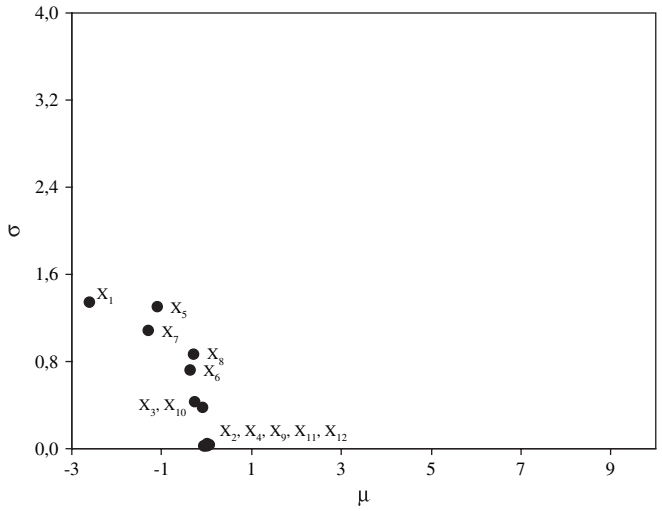

Fig. 6. EE sensitivity test on the g-function at a computational cost of $N = 130$. Plots of σ versus μ.

to a lack of knowledge of better techniques, to the lower computational cost, and sometimes even to the simplicity of their sampling design, which allows for the handling of unstable models. In fact, when a model is unstable, it may crash if executed on a set of input values rather different from the nominal values on which it is calibrated. A local design will permit the substitution of the sampling point causing the crash, whereas the more rigid variance-based design will not. The EE design shares with the local method not only the merit of a low computational cost, but also this ease of correction in the case of a crash, as it is based on trajectories independent of one another.

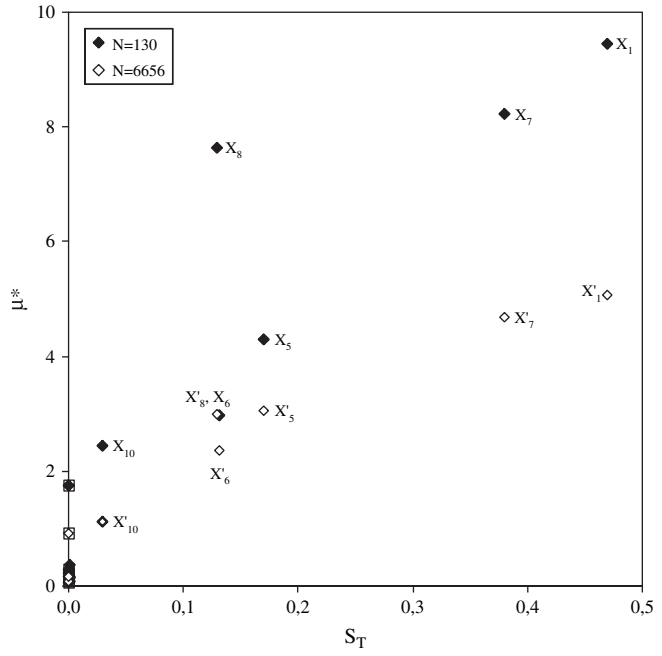

Fig. 7. Plots of μ^* versus S_T for the g-function test case. The Sobol' indices are estimated at the computational cost of $N = 6656$, while the EE values are computed at two different computational costs ($N = 6656$, apostrophed indices for the inputs, and $N = 130$, simple indices for the inputs).

5. The kim model

Dimethylsulphide (DMS, CH_3SCH_3) is the major biogenic sulphur gas emitted into the atmosphere from phytoplankton in the oceans. DMS has been tentatively identified as the major precursor of condensation nuclei and eventually cloud condensation nuclei in remote marine regions. DMS could thus have a significant influence on the earth's radiation budget and possibly on climate regulation. The effect of DMS on the climate is critically dependent on the production of gas-phase sulphuric acid and new particles, and the yield of H_2SO_4 from DMS depends on a very complex mechanism including gas-phase and multi-phase processes.

KIM is a multi-phase (gas-phase, liquid-phase, and gas–liquid exchange) chemical model of the DMS oxidation, which also includes temperature dependencies (Saltelli and Hjorth, 1995; Campolongo et al., 1999). KIM has been employed with the aim of quantifying the relative importance of the oxidation pathways of DMS, which ultimately determine the H_2SO_4 yield, and also with the aim of calculating the oxidation products of DMS under different atmospheric conditions (EUR report No. EUR 19569 EN, 2000).

In general the process of building environmental models is rather complex. The large uncertainties associated with environmental processes (atmospheric, hydrological, ecological, etc.) make necessary the use of uncertainty and sensitivity analysis (see, for instance, Hill et al., 2006; León et al., 2007), sometimes in the form of Monte Carlo simulation (see Raaflaub and Collins, 2006). The KIM model was not an exception to this. The reaction scheme adopted in KIM was affected by large uncertainty (structural uncertainty); moreover, the error bars associated with the rate constants governing DMS oxidation kinetics were also uncertain, and in some instances almost arbitrary (factor uncertainty). These concerns, together with the scarcity of observed data for a proper model calibration, led to a model implementation in which uncertainty and sensitivity analysis played a central role.

KIM is a non-linear dynamic model, which solves a system of approximately 40 ordinary differential equations involving about 60 species, and estimates concentrations of different sulphur compounds at several points in time. KIM is encoded into a computer programme and run in a Monte Carlo fashion, transmitting the uncertainty in the input parameters into the output variables. For the purpose of our analysis we treat the model as deterministic and take as a model output the average of a set of Monte Carlo simulations (1000). On a Pentium(R) 4 with CPU 3.00 GHz and 1.50 GB of RAM each model evaluation takes between 1 and 5 min, depending on the values chosen for the model parameters.

The EE method is employed here to assess the sensitivity of KIM to its uncertain inputs. The focus is on 56 inputs and their effects on 5 different model outputs. Uncertain input factors describe the starting concentrations, kinetic constants and

thermodynamics parameters. The outputs quantify the concentrations of sulphur (S) in five different compounds.

For this model, which has over 50 uncertain inputs, several outputs, and a non-negligible evaluation time, the EE method, and in particular its revised version, offers the best solution to perform a sound sensitivity analysis in a reasonable time. In our experiment, the values of the three sensitivity measures, μ, μ^* and σ, are obtained by performing 570 model executions ($k = 56$ and $r = 10$) and using the sampling strategy illustrated in Section 3.1. The number of levels chosen for each factor is $p = 4$. In Fig. 8, the graphical representation on the (μ, σ) plane as introduced by Morris (1991) is proposed as an example to represent results for one of the 5 outputs, the percentage of S as SO_2. σ is plotted versus μ and μ^*. Fig. 8 confirms that, if following the original Morris approach, it is essential to consider together the values of both μ and σ to appreciate correctly the relative influence of the inputs. The ranking provided by μ alone would not be sufficient to detect factors' importance, since some inputs have a negative effect, and therefore a negative value of μ, which would push them to the end of the list. Yet, in the case of many inputs and multiple outputs, the (μ, σ) representation becomes rather impracticable (Fig. 8 should be replicated for all the outputs of interest).

Instead, μ^* permits the detection of the most important factors overall, i.e. for all the model outputs under analysis. The ranking of the KIM inputs according to μ^* points out the overall importance of factors such as the intensity of rain, ranked

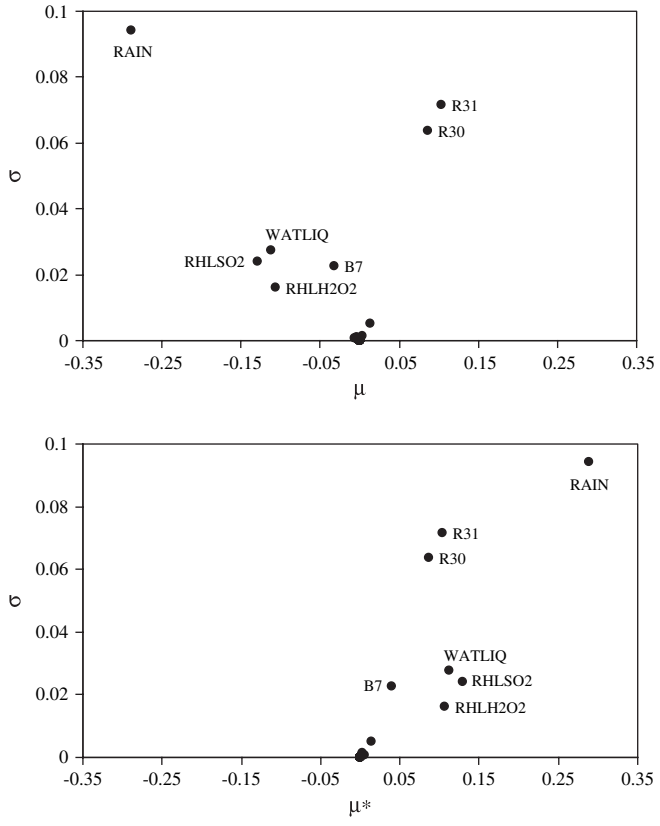

Fig. 8. Plots of σ respectively versus μ and μ^* for the KIM output percentage of S in SO_2.

either first or second for all outputs, and the water concentration in air, always recognised among the first 10 factors.

Concerning the intensity of rain, this result may call for a more accurate modelling of the effect of the rain on the system. At the moment rain is assumed to fall once every 10 days and is modelled in the form of a variable which follows a discrete uniform distribution in the set of integers (1,…,10). The effect of rain is then simulated by wet removal of all sulphur species. Future work on KIM may include a more sophisticated approach for including the effect of rain. The water concentration in air is a parameter which was introduced in KIM in order to compensate for the non-inclusion of the dynamic behaviour of clouds. The behaviour of clouds is thought to have a strong influence on the behaviour of the system but was not included, since the coupling of the full chemistry of KIM with a cloud dynamics module would have increased the complexity of the system by too great a degree. Instead, the cloud dynamics within the space of a day are simulated acting on the parameter ''water liquid content'' via a Monte Carlo approach. The parameter is kept constant within a run but is varied from run to run in a Monte Carlo fashion. The average of a Monte Carlo simulation is taken as representative of the variable behaviour of clouds. The sensitivity of the model to the water concentration in air can be considered an indication of the need for a more appropriate modelling of the dynamic behaviour of clouds.

Other overall important factors in KIM include the OH radical concentration in the liquid phase, and the Henry law constants for DMS and O_3. In general, the results highlight the importance of the aqueous-phase chemistry and the need to produce accurate estimates of the physical parameters involved.

μ^* also provides an answer to the screening problem by identifying the subset of factors which, being on the whole non-influential, can be fixed to any value within their ranges of uncertainty without significantly affecting the model outcomes (see Table 3). A detailed description of the meaning of these factors and of their role in the model can be found in the EUR report No. EUR 19569 EN (2000). Although the ranking of the inputs varies from output to output, results confirm that the modeller may easily select a subset of, for example, 15–20 factors which play a small role in the model, i.e. which are responsible for only a small percentage of the total output variance, thus preparing the ground for model simplification. Sensitivity analysis highlights, for instance, the fact that the enthalpies of formation (indicated as DEH in Table 3) are almost non-influential for most of the compounds in the model. The modeller may consider either eliminating or simplifying parts of the model where they are involved or revising the model structure so as to increase their importance, the choice depending on the modeller's preferences.

To corroborate the screening results obtained with the EE method, we perform a sensitivity analysis on groups of factors using the variance-based techniques. The aim of this second experiment is to confirm that the non-significant parameters identified through the EE method are actually responsible for a negligible fraction of the total output variance. Since

Table 3
List of the 20 less important inputs for each of the 5 KIM outputs

S in DMS (%)	S in DMSO (%)	S in $DMSO_2$ (%)	S in MSA (%)	S in SO_2 (%)
RHL(H_2O_2)	W16	RHL(CH_3SOOH)	RHL(CH_3SOOH)	R20
RHL(MSA)	RHL(CH_3SOOH)	R13	W11	R19
R17	R11	R17	RHL($DMSO_2$)	RHL($DMSO_2$)
R19	R18	W16	R9	W10
W11	R19	R10	RHL(SO_2)	RHL(CH_3SOOH)
R34	R31	R11	W8	DEH(O_3)
R11	R37	R37	R17	R34
DEH(SO_2)	W8	RHL(MSA)	DEH(DMSO)	W14
R37	W15	R25	DEH(H_2O_2)	R37
DEH(DMSO)	RHL(CO_2)	DEH(H_2O_2)	DEH(CH_3SOOH)	DEH(SO_2)
DEH(O_3)	RHL(NH_3)	DEH(DMSO)	R19	W11
DEH(CH_3SOOH)	RHL(MSA)	DEH(CH_3SOOH)	R37	DEH(DMSO)
DEH(H_2O_2)	DEH(CO_2)	RHL(NH_3)	DEH(SO_2)	DEH(H_2O_2)
R10	DEH(SO_2)	DEH(CO_2)	DEH(CO_2)	DEH(DMS)
DEH(CO_2)	DEH(H_2O_2)	DEH(SO_2)	DEH(DMS)	DEH(CH_3SOOH)
DEH(DMS)	DEH(O_3)	DEH(O_3)	DEH(O_3)	DEH(CO_2)
RHL(NH_3)	DEH(NH_3)	DEH(NH_3)	RHL(NH_3)	RHL(NH_3)
DEH(NH_3)	DEH(DMS)	DEH(DMS)	DEH(NH_3)	DEH(NH_3)
DEH($DMSO_2$)	DEH($DMSO_2$)	DEH($DMSO_2$)	DEH($DMSO_2$)	DEH($DMSO_2$)
DEH(MSA)	DEH(MSA)	DEH(MSA)	DEH(MSA)	DEH(MSA)

The ranking is obtained according to descending values of μ^*. RHL(X) and DEH(Y) indicate respectively the Henry low constant for gas X and the enthalpy of formation for compound Y while R and W label constant rates for reactions in gas or liquid phase. This table allows for fixing as many as 14 factors if the model is used to produce only the 5 outputs above.

the model is computationally expensive, the only way to apply variance-based techniques is to group factors into subsets. Starting from the outcomes of the screening, we group the 56 factors into two subsets, the first containing the least influential 14 factors, the second the remaining 42. To identify these least influential factors, we compute for each output variable the Savage scores (Iman and Conover, 1987) and order factors according to the sum of their scores. This strategy produces a single ranking starting from multiple outputs, thus allowing for the grouping of factors into two sets of 14 and 42, respectively. The variance-based analysis fully confirms the EE results, since for all the outputs the first group of factors accounts for less than 1% of the total variance. This corroborates the fitness of the EE method for use as a screening technique in models with multiple outputs and many input factors.

6. Conclusions

Although 14 years have passed since the EE sensitivity method was published (Morris, 1991), examples of its application are still rare in the literature. Too often the simplest local analyses, varying one factor at the time around a baseline point, are employed.

In this piece of work, we have recalled the efficacy of the EE design and proposed some improvements for it. We have revised the definition of its sensitivity measures in order to extend its utility to models with multiple outputs and to allow for analysis by groups of factors. We have proposed a more effective sampling strategy, which allows a better exploration of the space of the uncertain input factors. We have empirically demonstrated that the use of the EE sensitivity measure μ^* as a proxy of the variance-based total index is acceptable and convenient.

The sensitivity analysis performed on KIM, a model of the tropospheric chemistry of DMS, highlights the advantage of this revised measure when several outputs are to be analysed simultaneously. When applied to KIM, the method detected a subset of important factors that call for more accurate estimates or even for additional modelling efforts, and a subset of almost non-influential factors which open up the ground for a model simplification process.

References

Cacuci, D.G., Ionesco-Bujor, M., 2004. A comparative review of sensitivity and uncertainty analysis of large-scale systems-II: statistical methods. Nuclear Science and Engineering 147, 204–217.

Campolongo, F., Braddock, R., 1999. Sensitivity analysis of the IMAGE Greenhouse Model. Environmental Modelling and Software 14, 275–282.

Campolongo, F., Saltelli, A., 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. Reliability Engineering and System Safety 57 (1), 49–69.

Campolongo, F., Saltelli, A., Jensen, N.R., Wilson, J., Hjorth, J., 1999. The role of multiphase chemistry in the oxidation of dimethylsulphide (DMS). A latitude dependent analysis. Journal of Atmospheric Chemistry 32, 327–356.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, 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.

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

EUR report No. EUR 19569 EN. 2000 Dimethyl sulphide (DMS): oxidation mechanism in relation to aerosol and climate (DOMAC). CNRS-Orleans (F), University of Wuppertal (D), University of Crete (GR), EC-JRC-Ispra (I), FHG-ITA Hannover (D), University of Castilla La Mancha (SP).

Hill, M.J., Roxburgh, S.H., McKeon, G.M., Carter, J.O., Barrett, D.J., 2006. Analysis of soil carbon outcomes from interaction between climate and grazing pressure in Australian rangelands using Range-ASSESS. Environmental Modelling and Software 21 (6), 779–801.

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

Huber, P.J., 1981. Robust Statistics. Wiley Series in Probability and Mathematical Statistics (Chapter 1).

Iman, R.L., Conover, W.J., 1987. A measure of top-down correlation. Technometrics 293, 401–406.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21 (5), 602–614.

León, L.F., Lam, D.C.L., Schertzer, W.M., Swayne, D.A., Imberger, J., 2007. Towards coupling 3D hydrodynamic lake model with the Canadian Region Climate Model: Simulation on Great Slave Lake. Environmental Modelling & Software 22 (6), 787–796.

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

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Letters to Nature 430, 768–772.

Pappenberger, F., Iorgulesscu, I., Beven, K.J., 2006. Sensitivity analysis based on regional splits and regression trees (SARS-RT). Environmental Modelling and Software 21 (7), 976–990.

Raaflaub, L.D., Collins, M.J., 2006. The effect of error in gridded digital elevation models on the estimation of topographic parameters. Environmental Modelling & Software 21 (5), 710–732.

Saltelli, A., Hjorth, J., 1995. Uncertainty and sensitivity analysis of OH-initiated dimethyl sulphide (DMS) oxidation kinetics. Journal of Atmospheric Chemistry 21, 187–221.

Saltelli, A., Sobol', I.M., 1995. About the use of rank transformation in sensitivity analysis of model output. Reliability Engineering and System Safety 50, 225–239.

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

Saltelli, A., Chan, K., Scott, M. (Eds.), 2000. Handbook of Sensitivity Analysis. Probability and Statistics Series. John Wiley & Sons Publishers.

Saltelli, A., 2002. Making best use of model valuations to compute sensitivity indices. Computer Physics Communications 145, 280–297.

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.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. In: Probability and Statistics Series. John Wiley & Sons Publishers.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models. Chemical Review 105 (7), 2811–2828.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2006. Sensitivity analysis practices. Better strategies for model-based inference. Reliability Engineering and System Safety 91, 1109–1125.

Santner, T.J., Williams, B.J., Notz, W.I., 2003. Design and Analysis of Computer Experiments. Springer-Verlag.

Sobol', I.M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical Modelling and Computational Experiments 1 (4), 407–414.