Sensitivity analysis: Could better methods be used?

Andrea Saltelli
Institute for Systems, Informatics and Safety, Joint Research Centre of the European Commission, Italy

Abstract. Some recent articles are reviewed where sensitivity analysis (SA) is implemented via either an elementary “one factor at a time” (OAT) approach or via a derivative-based method. In these works, as customary, SA is used for mechanism identification and/or model selection. OAT and derivative-based methods have important limitations: (1) Only a reduced portion of the space of the input factors is explored, (2) the possibility that factors might interact is discounted, (3) the methods do not allow self-verification. Given that all models involved are highly nonlinear and potentially nonadditive, the adopted methods might fail to provide the full effect of any given factor on the output. This could deceive the analyst, unless the analysis were really meant to focus on a narrow range around the nominal value, where linearity may be assumed. Different methods are suggested, such as a rationalized OAT screening test, a regression-based method, and two implementations of global quantitative sensitivity analysis measures. Computational cost, efficiency, and limitations of the proposed strategies are discussed, and an example is offered.

1. Problem

Sensitivity analysis (SA) of model output is a valuable tool in the craftsmanship of modeling. It may help in verifying that the response of a model to its input conforms to theory. It may assist in the model calibration process, for example by optimizing the experimental conditions most suited to the determination of a given unknown factor. When testing different mechanistic hypotheses against available evidence, SA may help to decide to what extent the existing uncertainties allow a given mechanism to be unambiguously identified.

We argue that often the full potential of SA is not exploited and that, in some instances, SA is used improperly, especially when making statements about the relative importance of input factors. For the sake of illustration, we focus on some recent articles published in this journal.

Most investigators in the sample selected performed sensitivity analysis by changing “one factor at a time” (OAT, in the jargon of experimental and numerical design), and exploring what the model did with the new datum. In these analyses the baseline value was kept constant, that is the factors were moved away from the baseline only once (or twice) and the baseline was not changed throughout the analysis. Let us call this approach "elementary OAT", or EOAT.

While this approach is easy to implement, computationally inexpensive, and useful to provide a glimpse at the model behavior, it is limited. Any conclusion drawn on the relation between the output considered and the individual factor being varied is only legitimate around the baseline case.

An EOAT treatment is tried by Klonecki and Levy [1997], where the response of a model for the chemical ozone tendency (rate of change) is investigated by varying chemico-physical parameters one at a time. For example, uncertainty ranges found in the literature for kinetic constants are used (one constant and side of the range at a time) to investigate the sensitivity of the ozone tendency to uncertain reaction rates.

Clegg and Toumi [1997] test the sensitivity of sulphur dioxide oxidation in sea salt with respect to activity coefficients, relative humidity, dry sea-salt mass, and aerosol lifetime by changing one factor at a time. The same is done by Kermine and Wexler [1997], for a model of aerosol growth: factors are changed EOAT away from the base, case and the mean diameters of aerosol particles in the various modes (nuclei, Aitken, accumulation) as well as number concentration are computed. Kreidenweis and Zhang [1997], investigate the sensitivity of in the vertical profile of SO2 and CCN in a two-dimensional Eulerian cloud model. A set of nine simulations are performed for the purpose of SA, including two where the gridding of the model is changed.

In all the studies above the investigators were not planning a full-fledged SA, but were interested instead in a cursory appreciation of the overall effect of any given factor on the output, mostly in the context of mechanism identification.

In fact, all models considered (e.g. aerosol dynamics) are highly nonlinear, and the overall effect of a given factor could well escape or deceive the analyst when EOAT is used. This may easily happen because in some corner or edge of the input factors' space a different pattern of sensitivity exists, or because factors interact with each other.

A derivative-based approach is applied instead by Capaldo and Pandis [1997], where the effect of a given input x on the output y is assumed to be proportional to the derivative dy/dx. Derivative-based sensitivity analysis methods have been used extensively in chemistry, in a variety of applications such as the solution of inverse problems, for example computing kinetic constants from measured flow rates in a batch or flow reactors, (see Turanyi [1990], for a review), or relating variables at the molecular scale to those at the
macroscopic one [Rabitz, 1989]. Local sensitivity analysis allows the treatment of large systems of differential equations (see, e.g., the work on differential analysis and adjoint techniques by the Oak Ridge school [Cacuci, 1981a,b]). Present-day computational tools for local SA allow large numbers of sensitivity coefficients to be computed simultaneously, which provide a much more comprehensive picture than local EOAT. Local SA, used in this kind of problem setting, has vastly proven its worth.

We contend that neither an EOAT approach nor a derivative-based SA should be used to rank the impact of different uncertain (or variable) input factors in determining the variation of the output under examination, unless the model is known to be linear or the range of variation is small. The fact that such a use has a long record in the literature should not be taken as a foundation of its correctness.

For instance, when using a derivative-based approach, the model under examination is forcibly linearized: that is, one assumes that the effect of $x$ on $y$ is completely captured by $\frac{dy}{dx}$. Both in EOAT and in the derivative-based methods the existence of interactions (see below) is discounted. Mostly, the investigator using these approaches obtains a partial view of the behavior of the model’s response. Most nonlinear models are a complex function of the multidimensional space of the input factors. This complexity is lost when one focuses on the baseline point (of zero volume), as in the case of local derivative method. The same complexity is greatly simplified when one departs from that point only along the main axes, changing one factor at a time, as already mentioned.

Experience with environmental models (e.g., those involving mass transfer with chemical reaction) shows the following: (1) Different factors are affected by different ranges of variation / uncertainty. Different sensitivity patterns predominate in different regions of the space of the input, especially when models are nonlinear. (2) There are often non-negligible interactions (i.e., the effect of changing $x_j$ and $x_i$ is different from the sum of the individual effects). For these reasons, statements of the kind “$x_j$ is more important than $x_i$” using EOAT or local SA approaches can be unsustained (e.g., they might be true, but this is not inferable from the SA).

As an example, let us consider again the approach of Capaldo and Pandis [1997]. This particular study, among those reviewed above, is the most thorough from the SA perspective, and we are taking it here as the object of our criticism only for the sake of exemplification, as it is one of those most commonly seen in the literature. The target sensitivity measure considered by the authors is

$$S = \frac{x}{y} \frac{\partial y}{\partial x} \tag{1}$$

that is, the effect on the relative variation of $y$ of perturbing $x$ by a fixed fraction of $x$’s central value. $S$ is estimated via the computation of

$$S = \frac{\ln(y) - \ln(y_b)}{\ln(x) - \ln(x_b)} \tag{2}$$

where the subscript $b$ indicates a baseline value and $x$ is a generic input factor. Each of the $x$ is then given a different variation between the baseline and a “sensitivity test value”. These authors compare the prediction of five different models over nine different scenarios, varying a total of 26 factors (seven meteorological, nine physical and ten chemical). Baseline and sensitivity test values for the first three meteorological factors are shown in Table 1. Capaldo and Pandis acknowledge that different parameters exhibit different variation ranges, but their SA-based inference on the relative importance of different factors is unwarranted, as it suffers from the forced linearization of the model. Their approach assumes that $y$ responds linearly to $x$ in the range $[x, x_b]$ for all 26 factors. Further, each derivative is taken while the other factors are held at the central nominal value; the existence of edges and corners in the space of the input is neglected, and interactions are discounted.

Their conclusion may well hold; that is, it might indeed be the case that the ratio of SO$_2$ between the free troposphere (FT) and the marine boundary level (MBL) is 10 times more influential than any other parameter as far as nss-sulphate levels are concerned, as the authors state. Yet in a system with 26 factors with appreciable coefficients of variation (e.g. cloud frequency varying between 1 and 3), significant interaction may be expected, which may also include the SO$_2$ ratio in FT/MBL; these could considerably change the picture of the model sensitivities.

In the section entitled “improving the prediction”, the same authors implement a transversal calibration of all five models with respect to a subset of the factors. For this setting we would recommend as a possible alternative the Monte Carlo-based approach of Fedra et al. [1981]. In fact, if one wanted to effectively discriminate among competing models, the entire space of variation of the inputs should be explored. According to Fedra et al., rather than calibrating to a unique value of the factors, the calibration should more realistically lead to admissible ranges for the same factors.

The sensitivity analysis of Capaldo and Pandis is very much dependent upon the extremes selected for the differentiation if each factor, which increase the subjectivity of the analysis. This subjectivity would be reduced if one could perform SA via Monte Carlo by sampling from a nonuniform distribution (see below).

Finally, the $\hat{S}$ coefficients give no guarantee of correctness, as local SA methods do not allow a self-verification of their correctness. A sensitivity analysis method can be said to allow self-verification when the analyst can estimate or quantify the error of the analysis directly from its results. As discussed in the next section, this is often possible. In Monte Carlo based regression, hypothesis testing can be used; when using variance-based methods, one can verify the fraction of the output variance accounted for by the factors. The $\hat{S}$ coefficients instead do not offer an estimate of the error implicit in the linearization process.

2. What Should Then Be Used?

A useful review of sensitivity analysis methods is by Helton [1993], while recent material on SA is available in a series of recent special issues (Special Issue on Sensitivity Analysis of Model Output, Journal of Statistical Computation and Simulation, 57(1-4), 1997; Special Issue on Sensitivity

| Table 1. Baseline and Sensitivity Test Values From Capaldo and Pandis [1997] |
|--------------------------------|-----------------|--------------------|
| Factor                        | Baseline        | Sensitivity Test  |
| Temperature, K                | 290             | 300                |
| Mixing height, m              | 500             | 1,000              |
| Cloud frequency, 1/d          | 3               | 1                  |
2.1. First Approach

Use an established regression analysis method, such as the standardized regression coefficients SRC(y,x), and base the sensitivity analysis on the coefficients [Draper and Smith 1981, Iman et al. 1985, Iman and Helton 1988] (see Raes et al. [1992] for a recent example in the field of aerosol dynamics). This Monte Carlo based method allows self-verification, as the effectiveness of the SRC(y,x), as a measure of sensitivity is conditional upon the model coefficient of determination:

\[ R^2 = \frac{\sum (\tilde{y} - \bar{y})^2}{\sum (y - \bar{y})^2} \]  

where the \( y \) are the original model evaluations and the \( \tilde{y} \) are the ones for the regression model. R^2 can be computed together with the SRC(y,x), and gives the fraction of the model’s output variance accounted for by the linear regression model. If R^2 is close to one, the absolute value of the SRC(y,x) can be used to rank the relative influence of the input factors on the output. Otherwise, the analyst must be ready to try one of the methods below. This method (as all those based on Monte Carlo) also offers the advantage that input factors can be sampled from distributions, including nonuniform ones. Sampling, for example, from a normal distribution reduces the influence of the selected endpoints as compared to a uniform distribution with the same extremes. This could correspond to the realistic situation where one were to believe the central value more than the extreme ones.

The cost of this method in terms of model evaluations N is proportional to N^2, where the y are the original model evaluations and the \( \tilde{y} \) are the model evaluations in the Num model. If the randomization is feasible, this approach is applied to the original model evaluations and the \( \tilde{y} \) are used to evaluate the output’s variance (fractional contribution of factor x_j to the variance of y), which can be written as:

\[ S_j = \frac{V(E(y|x_j))}{V} \]  

where V is the unconditional variance of y, the inner expectation value in the numerator is taken over all the space of the x_j with x_i fixed to a given value, and the outer variance V is over all the possible values of x_i. This method works well even for nonlinear models as long as they are additive. In this case the sum of the S over the k factors is close to one. When this is not the case, the next method should be used. The cost of FAST is quadratically dependent on the number of factors k. Minimum sample sizes \( N_k \) for a few values of k are given below [from Cukier et al. 1975]:

<table>
<thead>
<tr>
<th>k</th>
<th>N_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>

2.2. Second Approach

Use the Morris method [Morris, 1991]. This is a screening method that belongs to the OAT class, but with Morris the baseline changes at each step; that is, this method wanders in the space of the input factors rather than oscillating around the baseline as in EOAT. Morris seeks to determine which factors may be considered to have effects that are (1) negligible, (2) linear and additive, or (3) nonlinear or involved in interactions with other parameters. Morris allows the input factors to be ranked in order of importance. This method is rather qualitative, and model-independent. It is in general cheaper than the regression-based ones: its cost is \( N_k = r(k+1) \), where k is the number of factors and r the sample size. Values of r in the range (4-10) have been reported [Morris, 1991; Campolongo and Saltelli, 1997]. Morris does not allow self-verification in the sense discussed above. It should be used when the number of factors being varied is high (e.g. tens), and also in combination with a more quantitative method. In this latter case the purpose of the screening via Morris is to fix to their midpoint the noninfluential factors.

2.3. Third Approach

Use the Fourier amplitude sensitivity test (FAST) [Cukier, et al. 1973; Schablich and Schulter, 1973; Cukier, et al., 1975; Koda et al., 1979; McRae, et al., 1982]. FAST represents one of the most elegant methods for SA and is based on determining fractional contributions of individual factors to the variance of the output. The method is suited for nonlinear models, and, although it does not explicitly compute interactions, it tells the analyst how much of the total output variance can be due to interactions. In this respect, FAST allows self-verification (i.e., the analyst is informed of how much variation remains unexplained). In the terminology of ANOVA (analysis of variance), FAST offers estimates of S_j, the “first-order” contributions of the input factors to the output’s variance (fractional contribution of factor x_j to the variance of y), which can be written as:

\[ S_j = \frac{V(E(y|x_j))}{V} \]  

where V is the unconditional variance of y, the inner expectation value in the numerator is taken over all the space of the x_j with x_i fixed to a given value, and the outer variance V is over all the possible values of x_i. This method works well even for nonlinear models as long as they are additive. In this case the sum of the S over the k factors is close to one. When this is not the case, the next method should be used. The cost of FAST is quadratically dependent on the number of factors k. Minimum sample sizes \( N_k \) for a few values of k are given below [from Cukier et al. 1975]:

<table>
<thead>
<tr>
<th>k</th>
<th>N_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>

2.4. Fourth Approach

Use Sobol’ sensitivity indices which allow a full decomposition of the output variance, including interactions [Sobol', 1990; Homma and Saltelli, 1996]. This method gives the same answer as FAST when used at the first order [Saltelli and Bolado, 1998]. It can be considered as a quantitative method, as no fraction of the output variance is left unaccounted for. For an application to atmospheric chemistry, see Saltelli and Hjorth [1995]. This method can give, for each factor x_j, the index SN, defined as:

\[ S_N = S_j + \sum_{j \neq k} S_{j+k} + \sum_{k \neq j} S_{j+k} + \ldots \]  

where the y are the original model evaluations and the \( \tilde{y} \) are the model evaluations in the Num model. If the randomization is feasible, this approach is applied to the original model evaluations and the \( \tilde{y} \) are used to evaluate the output’s variance (fractional contribution of factor x_j to the variance of y), which can be written as:

\[ S_j = \frac{V(E(y|x_j))}{V} \]  

where V is the unconditional variance of y, the inner expectation value in the numerator is taken over all the space of the x_j with x_i fixed to a given value, and the outer variance V is over all the possible values of x_i. This method works well even for nonlinear models as long as they are additive. In this case the sum of the S over the k factors is close to one. When this is not the case, the next method should be used. The cost of FAST is quadratically dependent on the number of factors k. Minimum sample sizes \( N_k \) for a few values of k are given below [from Cukier et al. 1975]:

<table>
<thead>
<tr>
<th>k</th>
<th>N_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>
that is, an index which includes all effects of $x_i$, (first order, second order, higher order). For a system with just three input factors, the total sensitivity of $x_i$ is:

$$S_{2i} = S_i + S_{ii} + S_{i1} + S_{i2}$$  

(6)

This method is model independent, it works even for non-linear non-additive models, but is computationally expensive, as one must evaluate the model $N_k$ times, where

$$N_k = (k + 1)N$$  

(7)

and $k$ is the number of factors one wishes to vary, and $N$ is a number in the range of 100 or higher. There is now a cheaper computational implementation of the same $S_n$ measure, which can be computed via an extension of the FAST method [Saltelli et al., 1998].

3. An example

It would be far too easy, for the purpose of illustrating the advantages of the methods discussed here, to select a model displaying nonmonotonic and nonadditive features which would render EAOT and local sensitivity totally inapplicable. Further, these models can be easily generated analytically [Saltelli and Sobol', 1995].

We have taken instead the latest model used in our research work [Campolongo et al., 1998] without changing its configuration. This model, KIM, with 68 uncertain input factors, is suitable for a screening test like Morris. KIM stands for kinetic model for the oxidation of DMS ($\text{CH}_3\text{SCH}_3$), and incorporates a description of the multiphase tropospheric reaction pathways for the formation of sulphur-containing molecules, such as sulphur dioxide ($\text{SO}_2$) and methane sulphonic acid ($\text{MSA, CH}_3\text{SOH}$), from DMS [Saltelli and Hjorth, 1995; Remedio et al., 1994; Campolongo et al., 1998].

KIM involves the numerical solution of about 40 differential equations describing chemical kinetics and air-to-drop exchanges.

The variable considered is the concentration ratio in marine aerosol between MSA and non-sea-salt sulphates (nss-sulphate, including $\text{SO}_2$ and $\text{H}_2\text{SO}_4$), that is

$$\alpha = \frac{\text{MSA}}{\text{SO}_2 + \text{H}_2\text{SO}_4}$$  

(8)

whose temperature dependency is the subject of the Campolongo et al., [1998] study.

Morris results are contrasted against local SA in Table 2. Derivatives of the output were computed by changing all inputs of a 1% around the nominal value. Only those factors that are displayed that are in the list of the top five for either method (Morris or derivative).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Ranking from Morris</th>
<th>Ranking from $\frac{\partial \alpha}{\partial x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WATLIQ</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>QW7</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>DEH2O2</td>
<td>3 (6)</td>
<td></td>
</tr>
<tr>
<td>RHL2O2</td>
<td>4 (8)</td>
<td></td>
</tr>
<tr>
<td>RHLDMS</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>RHL3</td>
<td>(6)</td>
<td>3</td>
</tr>
<tr>
<td>YOHRAD</td>
<td>(8)</td>
<td>4</td>
</tr>
<tr>
<td>W11</td>
<td>(14)</td>
<td>5</td>
</tr>
</tbody>
</table>

Ranking in parentheses are not among the top five factors for one of the methods.

Table 3. Results of Monte Carlo Analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>SRC Ranking from $R^2 = 0.74$</th>
<th>SRRC Ranking from $R^2 = 0.87$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q21</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>WATLIQ</td>
<td>(15)</td>
<td>3</td>
</tr>
<tr>
<td>QW7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>RHL2O2</td>
<td>(6)</td>
<td>(8)</td>
</tr>
<tr>
<td>RHLDMS</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>RHL3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>YOHRAD</td>
<td>...</td>
<td>4</td>
</tr>
<tr>
<td>W11</td>
<td>...</td>
<td>5</td>
</tr>
</tbody>
</table>

SRC denotes standardized regression coefficients, SRRC denotes standardized rank regression coefficients. Ranking in parentheses are not among the top five.
nonlinear than purely chemical ones. Superposition of transport to chemical reaction is also a potential source of complexity in the pattern of sensitivities.

4. Conclusions

We were motivated in the present note by the belief that long-established practices in modeling, especially in the field of chemistry, physics, and natural sciences, are sometimes used out of context. This is the use of EOAT or a local SA method to draw conclusions on the relative impact of uncertain or variable input factors on the prediction of a model, unless the variation of the factors is small. Even when the investigator does not desire to embark in a quantitative SA study, the Morris method, or even a plain Monte Carlo analysis at low sample size, should be preferred over elementary OAT. Similarly, the use of local methods should be confined to situations where they can perform effectively, as in the case of inverse problems or in the computation of large sets of sensitivity coefficients as function of time.

More generally, we advocate the use of global, possibly quantitative, sensitivity analysis methods for all problem settings where finite parameters variations are involved.

Acknowledgments. The author is grateful to Francesca Campolongo, who revised this manuscript and provided helpful suggestions.

References


A. Saltelli, European Commission / Joint Research Centre, Institute for Systems Informatics and Safety, TP 361, Ispra, 1-21020 Italy.

(Received May 11, 1998; revised September 28, 1998; accepted September 30, 1998.)