Lecture 26. Global Sensitivity Analysis (GSA) for Models

"In *global SA*, the aim is to apportion the uncertainty in the output variable to the uncertainty in each input factor."

— Campolongo et al. in Ch. 2 of [1]

1 What is global sensitivity analysis?

Global sensitivity analysis (GSA) tells us how (or whether) a variation in one thing (a model input) causes variation in another thing (a model output), but goes beyond a local derivative-based method to try and capture the response across a wide range of system behaviors.

Global sensitivity analysis is "the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input." [5]

1.1 Compared with simple, local derivative-based sensitivity analysis

The simplest method for obtaining a sensitivity index, as discussed in the previous lecture, is to obtain a local derivative. For modelers, this means fixing all factors except the one for which you are obtaining a sensitivity, then varying that factor of interest. This is called a one-at-a-time (OAT) or simple sensitivity analysis. Saltelli cautions: "Differential analysis [local derivative-based SA] can be used with caution when the range of variation of the factors is small. Some methods... perform poorly for nonlinear, non-additive models." [3]

A GSA goes beyond this by incorporating:

- Scale and shape
- Effect(s) of stochastic variation in the input(s) on the output(s)
- Range of possible values or probability distributions possible for the input(s)

If a local derivative is assumed to be representative in a model that is actually complex and nonlinear, or if uncertainty is neglected, this "might result in an under-estimation of predictive uncertainty and an incomplete understanding of the input-output relationship." [5]

1.2 Paired with uncertainty analysis

Saltelli cautions: "Local methods are less helpful when SA is used to compare the effect of various factors on the output, as in this case the relative uncertainty of each input should be weighted." [3]

2 Capturing more complexities

Let's briefly describe a few key terms that will be used in our discussions.

Sensitivity index A derivative-based valuation of how influential a factor, or input variable, is on an output of interest. [5]

Main effect A 'first order effect' that is based on the variance of the output of interest with the factor fixed at its (estimated) midpoint; i.e. "If we could eliminate the uncertainty in one of the [inputs], making it into a constant, how much would this reduce variance of [the output]?" [5]

Total effect A higher-order effect that is based on the variance of the output of interest over all possible values of the factor—an attempt to capture the total contribution to the variation in model output due to this factor.

For the energy modeler, it may well be the case that "factors in a model follow a very asymmetric distribution of importance, few factors accounting for most of the output uncertainty with most factors playing little or no role." [5]

2.1 Interactions (Complex and nonlinear models)

The sensitivity index gives us a metric for understanding the response of one variable (output) to another (input), as described in the previous lecture. When we refer to 'sensitivity index' without additional context, this is taken to mean a first-order effect.

The first-order effect is a number, scaled between 0 and 1, representing the sensitivity to a certain factor. If the input factors are independent (non-interacting), the sum of the first order effects (when normalized) should be 1. Unity represents the combined sensitivity to all features. Therefore, in that case we can think of the first order effect as being a fraction representing that particular factor's influence on the outcome.

If the input factors are not independent, the sum of the first order effects (when normalized) will *not* be 1. If the combination of the first-order sensitivity indices never reaches 1, that indicates we haven't completely captured the full picture of the influences of the relevant factors throughout the process. This could be an indication that we have interactions between the factors. The effect of an interaction between two factors can also be called a two-way or second-order effect.

The total effect is "the sum of all the sensitivity indices involving the factor in question." [3] It is a number that has been normalized in the same way as the first-order effect, but the total effects will not sum to 1. In fact, for the case of interacting factors where the first-order effects sum to less than 1, the total effects will be greater than 1. This is because the total effect captures sensitivity behaviors in the model related to variance (Scn. 7.2) in the factor and due to all interactions involving that particular factor of interest. The total effect can also be called a total sensitivity index [3].

For any individual factor, looking at its main effect and total effect estimates together can be informative or even illuminating. Without these calculations, the linear effects may be evident from the first-order indices, while the nonlinear effects may not be evident.

3 Do I need GSA methods?

Your time and computational resources may limit your choices on how much analysis you can do, but here are a few questions to ask that would indicate you should look further into GSA methods:

- Is the model nonlinear? Do you know or suspect that the factors are interacting with each other?
- Are your input factors "affected by uncertainties of different orders of magnitude?" [3]
- Have you performed a simple (local derivative or parametric) sensitivity analysis and found that the results are not as useful as you expected, or that they don't make sense based on your understanding of the system?

An in-depth sensitivity analysis of a system model can help us to find "where [we] should improve our knowledge in order to reduce the risk of failure." [5] This may drive model improvements and improved or additional data collection procedures as well as improving understanding of the model's behavior. Interestingly, a thorough SA may also help us to decide what model to use or how a system ought to be modeled: "Sensitivity analysis can be employed also when dealing with uncertain or competing model structures or scenarios, treating the choice of the model as one of the sources of uncertainty." [4]

4 Desirable properties for GSA

Some of the most important benefits of a high-quality sensitivity analysis are summarized in this list adapted from Saltelli [3]:

1. Ability to cope with influence of scale & shape
2. Evaluation of a factor's effects while all other factors are also varying
3. **Efficiency:** Accounting for the challenges above in a way that is manageable computationally.

4. **Model Independence:** Works with models that are not necessarily additive or linear
5. **Simplicity:** Independent of factor grouping (to help interpretability of the results, in additional to increasing efficiency)

5 Global properties for GSA

1. "**The inclusion of scale and shape**: The sensitivity estimates of individual factors incorporate the effect of the range and the shape of their probability density functions." [3]
2. "**Multidimensional averaging:** The sensitivity estimates of individual factors are evaluated varying all other factors as well." [3]

This is the most concise summary I have seen describing the benefits of, and the need for, GSA.

6 How to perform a (global) sensitivity analysis

The steps shown in the previous SA lecture (adapted from [3]) were actually entirely applicable to a more detailed SA, and are reproduced here.

1. Design the modeling experiment (identify what question the model should answer) and determine which of the input factors you are concerned with in the analysis.
2. Assign a range of variation to each input factor (and a probability density function, if you have a known or estimated distribution).
3. Evaluate the model, creating an output distribution for the response(s) of interest.
4. Assess the influence or relative importance of each input factor you are concerned with on the output variable(s).

Saltelli [3] also provides a diagram that illustrates this procedure, including sampling-based analysis and output distribution, as shown in Figure 1.

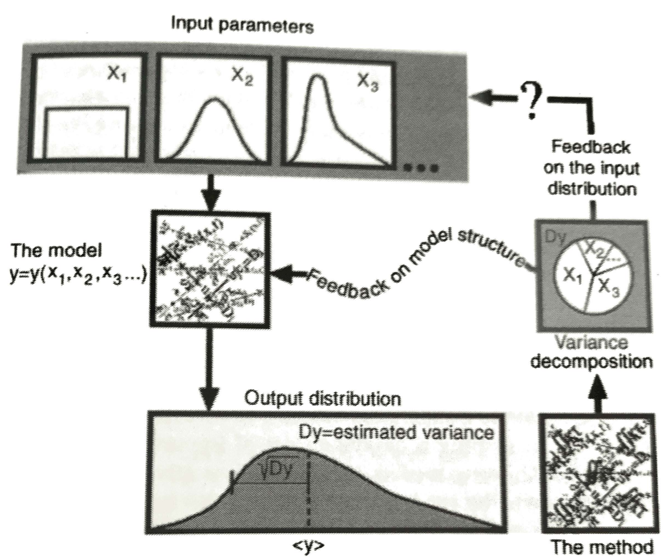

Figure 1: Schematic view of sampling-based sensitivity analysis [3]

7 Probability Distribution Functions (PDF)

As illustrated in Fig. 1, knowledge of the distribution (or likely distribution) of the input(s) is beneficial for the modeler using SA, as it can allow him or her to provide an expected distribution for the output(s).

One key thing to understand is that "for continuous probability distributions, PROBABILITY = AREA." [2]

7.1 Uniform distribution

A uniform distribution means there is an equal likelihood of any value along a specified range. When you do a typical parametric analysis, or when you perform sensitivity analysis without using probability densities, you are usually implicitly assuming a uniform distribution.

7.2 Gaussian or normal distribution

The Gaussian, or normal, distribution is frequently encountered in experiments and frequently used by modelers. It is identified by two key items: the mean, or center of the distribution, and the variance or standard deviation, both of which describe the "spread" of the data. Remember that the standard deviation is simply the square root of the variance.

The distribution is often normalized (to allow for the use of lookup tables, and comparisons across disparate applications) so that its total area is 1. If you want to find the probability of landing below or above a certain value, or landing between two values, in a normal distribution, you just integrate. Fig. 2 illustrates a more complex distribution where the analyst wants to find the probability of landing in between values a and b.

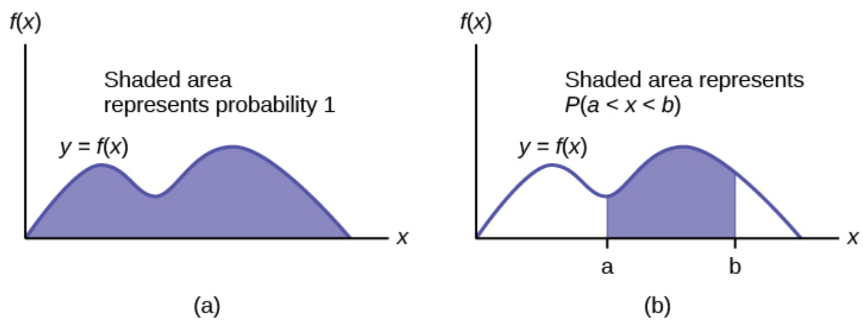

Figure 2: Integrating a PDF [2]

We will only deal with uniform or normal distributions in this class, but you certainly may run into other distributions with real-world data and in your energy system modeling efforts in practice.

References

[1] Francesca Campolongo et al. "Hitchhiker's guide to sensitivity analysis". In: *Sensitivity analysis*. IEEE Computer Society Press, 2000. URL: http://www.forskningsdatabasen.dk/en/catalog/2192873057.

[2] Barbara Illowsky and Susan Dean. *Introductory statistics*. cnx.org, 2014. URL: http://svetn.org/uploads/docs/Elementary_Statistics--FINAL_Verified.pdf.

[3] A Saltelli, K Chan, and E M Scott. *Sensitivity Analysis*. en. Wiley, 16 3 2009. ISBN: 9780470743829.

[4] A Saltelli, S Tarantola, and F Campolongo. "Sensitivity Analysis as an Ingredient of Modeling". In: *Stat. Sci.* 15.4 (2000), pp. 377–395. ISSN: 0883-4237. URL: http://www.jstor.org/stable/2676831.

[5] A Saltelli et al. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. en. John Wiley & Sons, 16 7 2004. ISBN: 9780470870945.