



Global Sensitivity Analysis of environmental models: Convergence and validation



Fanny Sarrazin ^{a,*}, Francesca Pianosi ^a, Thorsten Wagener ^{a, b}

^a Department of Civil Engineering University of Bristol, University Walk, BS81TR, Bristol, UK

^b Cabot Institute, Royal Fort House, University of Bristol, BS8 1UJ, Bristol, UK

ARTICLE INFO

Article history:

Received 12 June 2015

Received in revised form

11 January 2016

Accepted 1 February 2016

Available online 13 February 2016

Keywords:

Sensitivity Analysis

Screening

Ranking

Convergence

Validation

SWAT

ABSTRACT

We address two critical choices in Global Sensitivity Analysis (GSA): the choice of the sample size and of the threshold for the identification of insensitive input factors. Guidance to assist users with those two choices is still insufficient. We aim at filling this gap. Firstly, we define criteria to quantify the convergence of sensitivity indices, of ranking and of screening, based on a bootstrap approach. Secondly, we investigate the screening threshold with a quantitative validation procedure for screening results. We apply the proposed methodologies to three hydrological models with varying complexity utilizing three widely-used GSA methods (RSA, Morris, Sobol'). We demonstrate that convergence of screening and ranking can be reached before sensitivity estimates stabilize. Convergence dynamics appear to be case-dependent, which suggests that “fit-for-all” rules for sample sizes should not be used. Other modellers can easily adopt our criteria and procedures for a wide range of GSA methods and cases.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Sensitivity Analysis (SA) aims to characterize the impact that changes in the model input factors (e.g. parameters, initial states, input data, time/spatial resolution grid etc.) have on the model output (e.g. a statistic of the simulated time series, such as the average simulated streamflow, or an objective function, like the Root Mean Squared Error). SA is a diagnostic tool that can guide model calibration and verification, support the prioritization of efforts for uncertainty reduction, or help with model-based decision-making (Norton, 2015; Pianosi et al., 2016; Song et al., 2015). Such purposes are generally implemented as four different objectives of GSA: screening (or Factor Fixing), ranking (or Factor Prioritization), Variance Cutting, and Factor Mapping (Saltelli et al., 2008). Screening refers to the identification of those input factors, if any, which have no influence on the model output and therefore can be fixed to any value within their feasible range with negligible implications on the output. For instance, in Kannan et al. (2007) and in Vanuytrecht et al. (2014), screening of model parameters is applied as a preliminary step to inform a subsequent calibration, which is tailored to the subset of influential parameters. Ranking

describes the ordering of the input factors according to their relative influence on the model output. It is typically used to enhance our understanding of the model and to identify dominant controls of the model's behaviour (e.g. Van Werkhoven et al., 2008), as well as to prioritize efforts for uncertainty reduction (e.g. Sin et al., 2011), or to support model development (Hartmann et al., 2013). The Variance Cutting setting is used for the reduction of the output variance to a value below a user chosen tolerance. It aims at obtaining specific sensitivities for the different input factors and is, for example, applied in reliability and risk assessment (e.g. Saltelli and Tarantola, 2002). Finally, Factor Mapping aims at identifying those conditions (e.g. sub-ranges of input factors like parameters or forcing inputs) that produce critical values of the output. It can be used to enhance model understanding (e.g. Spear and Hornberger, 1980) or to support robust decision-making (Singh et al., 2014).

Unlike Local Sensitivity Analysis (LSA), where the variability of the model output is explored around some reference input factor setting (e.g. Ljung, 1999 for a general link between LSA and model calibration; Hill and Tiedeman, 2007, for an example application to groundwater models), Global Sensitivity Analysis (GSA) rather attempts to explore the entire space of the input factors. It therefore typically requires larger computational resources than LSA. Generally, the implementation of GSA methods is sampling-based and the value of the sensitivity indices are approximated using

* Corresponding author.

E-mail address: fanny.sarrazin@bristol.ac.uk (F. Sarrazin).

Monte Carlo (MC) simulations. A critical step of sampling-based GSA is therefore the choice of the sample size to run the MC experiment. If the sample is too small to adequately cover the input space, the analysis may not provide robust results. On the other hand, for very large sample sizes the computational cost may become very high while not improving the precision of the results significantly. In environmental applications, where models are often complex and simulations expensive, an acceptable trade-off has to be found between the need to obtain robust results and the need to limit computational cost.

The total number of model evaluations (N) used in GSA typically increases with the number of model input factors (M). For some GSA methods, depending on the methodology used to derive the estimates of the sensitivity indices, N is expressed as a function of M and of a base sample size (n) that must be specified by the user (i.e. $N = f(n, M)$). Thus, choosing the value of the total number of model evaluations (N) comes down to choosing the value of the base sample size (n). For other methods, no explicit expression relates N to M and therefore N is directly chosen by the GSA user ($N = n$). Suggestions for the choice of n can be found in the literature for several GSA methods. For instance, Saltelli et al. (2008, Table 6.9) report typical values of n for the Elementary Effect Test (EET, or method of Morris (Morris, 1991)), for Regional Sensitivity Analysis (RSA; Young et al., 1978; Spear and Hornberger, 1980) and for Variance-Based Sensitivity Analysis (VB-SA; Sobol', 1990; Saltelli, 2002). However, a relatively limited number of studies actually focus on a rigorous assessment of the convergence of GSA results. Fig. 1 reports several examples taken from the literature regarding the relationship between N and M for EET, RSA and VB-SA. From these studies, we make three observations:

1. Previous convergence studies assessed different types of convergence, namely convergence of the sensitivity indices, of the screening results (identification of the non-influential input factors), and of the ranking (ordering of input factors according to their relative importance). This lack of uniformity in the definition of 'convergence' makes it difficult to consistently compare the results obtained for models of different complexities when using different GSA methods. However, a preliminary conclusion that seems to emerge from these studies is that different sample sizes are required for different types of convergence. For instance, in the case of EET, Vanuytrecht et al. (2014) highlight that while a low sample size ($n = 25$) can be suitable for screening, it can be insufficient for factor ranking. Nossent et al. (2011) find that a base sample size of 12,000 is needed to ensure the convergence of Variance-Based sensitivity indices in their specific case study, however, a much smaller sample size ($n < 2000$) is sufficient if one is only interested in ranking the most important input factors.
2. Within a given type of convergence, different values of the base sample size are found for the same method when applied to different models. For instance to ensure convergence of the value of Variance-Based sensitivity indices (Fig. 1 bottom left panel), Tang et al. (2007) use a base sample size n of 8192 (for a case study with 18 input factors), while Yang (2011) uses n equal to 3000 (for a case study with 5 input factors). This suggests that the base sample size may also be a function of the number of input factors or of other characteristics of the model or of the case study. It is also worth noting that these studies show that convergence is often reached using a base sample size significantly larger than the values suggested in Saltelli et al. (2008).
3. Convergence is generally assessed based on a visual analysis of the stability of the results for increasing sample size. Some authors use the confidence intervals of the sensitivity indices for a more quantitative assessment of their convergence (e.g.

Campolongo and Saltelli, 1997; Nossent et al., 2011). However, they do not explicitly define a convergence criterion. Herman et al. (2013) and Vanrolleghem et al. (2015) both introduce a quantitative criterion to measure the convergence of the sensitivity indices values (that will be discussed in Section 2.1), but they do not consider the convergence of ranking or screening.

Another issue for GSA is the choice of the screening threshold i.e. a threshold value for the sensitivity indices below which factors are classified as insensitive (more details in Section 2.1). In this respect, the following can be learned from existing studies:

1. For Variance-Based SA, the input factors that have a sensitivity index below 0.01 are often considered non-influential (Tang et al., 2007; Sin et al., 2011; Cosenza et al., 2013; Vanrolleghem et al., 2015). The adequacy of this screening threshold is tested in Tang et al. (2007), however the validation strategy used in that work (based on a visual approach introduced by Andres (1997)) has some limitations that we discuss and overcome here (more details in Section 2.2). Nossent et al. (2011) consider a screening threshold value of zero. They identify as statistically significant any input factor for which the lower bound of the confidence interval on the sensitivity index is positive. This method is quite conservative since, in our experience, a sensitivity index could have positive confidence bounds, and therefore a non-zero value, even if the input factor has negligible effect on the output.
2. EET, which is widely used for screening purpose, provides a relative measure of sensitivity that has a different meaning and range of variation depending on the model output definition in the particular case under study. Therefore, case-specific threshold values are usually taken (Vanuytrecht et al., 2014) and little guidance exists in the literature on this topic. Cosenza et al. (2013) and Vanrolleghem et al. (2015) present an attempt at defining an absolute value for the screening threshold for EET. However, they do not validate the adequacy of their proposed threshold values.

Based on this literature review, we believe that there is a lack of guidance to support GSA users in the choice of an adequate sample size and in the definition of a screening threshold, while there is an opportunity for improving current approaches to the validation of GSA results. Thus, the objectives of the present study are:

1. To define quantitative criteria to assess different types of convergence of GSA results, i.e. convergence of sensitivity indices, ranking and screening.
2. Based on these quantitative convergence measures, to investigate the convergence of three widely used GSA methods and to assess whether it is possible to give general guidelines for an adequate choice of the base sample size.
3. To develop a methodology to quantitatively validate screening results and therefore to formally investigate the adequacy of different choices for the screening threshold.

Here, we consider three widely used GSA methods, the Elementary Effect Test (EET), Regional Sensitivity Analysis (RSA) and Variance-Based Sensitivity Analysis (VB-SA), implemented in the Sensitivity Analysis For Everybody (SAFE) toolbox (Pianosi et al., 2015). We apply GSA to three hydrological models of increasing complexity (HyMod, HBV and SWAT). The input factors are the model parameters and the output is the model accuracy. However, our approach could equally be applied to other GSA methods or models, and with different experimental set-ups, i.e. different

Download English Version:

<https://daneshyari.com/en/article/6962559>

Download Persian Version:

<https://daneshyari.com/article/6962559>

[Daneshyari.com](https://daneshyari.com)