

Assessing convergence in global sensitivity analysis: a review of methods for assessing and monitoring convergence

Xifu Sun^{1,2,3,*}, Anthony J. Jakeman⁴, Barry F.W. Croke^{4,5}, Stephen G. Roberts⁵, and John D. Jakeman⁶

¹Key Laboratory of Virtual Geographic Environment (Ministry of Education of PR China), Nanjing Normal University, Nanjing, China

²Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

³State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing, China

⁴Institute for Water Futures, Fenner School of Environment and Society, Australian National University, Canberra, Australia

⁵Mathematical Sciences Institute, Australian National University, Canberra, Australia

⁶Optimization and Uncertainty Quantification Department, Sandia National laboratories, Albuquerque, NM, USA

Abstract

In global sensitivity analysis (GSA) of a model, a proper convergence analysis of metrics is essential for ensuring a level of confidence or trustworthiness in sensitivity results obtained, yet is somewhat deficient in practice. The level of confidence in sensitivity measures, particularly in relation to their influence and support for decisions from scientific, social and policy perspectives, is heavily reliant on the convergence of GSA. We review the literature and summarize the available methods for monitoring and assessing convergence of sensitivity measures based on application purposes. The aim is to expose the various choices for convergence assessment and encourage further testing of available methods to clarify their level of robustness. Furthermore, the review identifies a pressing need for comparative studies on convergence assessment methods to establish a clear hierarchy of effectiveness and encourages the adoption of systematic approaches for enhanced robustness in sensitivity analysis.

Keywords

Global sensitivity analysis; Convergence; Good modeling practice

1. Introduction

Good modeling practice (GMP) is essential for the development, evaluation and ultimate utility of environmental models (Jakeman et al., 2006). Sensitivity analysis is an indispensable tool (Razavi et al., 2021; Saltelli et al., 2021) with a long history (Tarantola et al., 2024), and is vital for supporting GMP including understanding model behavior and diagnosing problems such as parameter interactions and

Correspondence:

Contact X. Sun at xifu.sun@outlook.com

Cite this article as:

Sun, X., Jakeman, A.J., Croke, B.F.W., Roberts, S.G., & Jakeman, J.D.
Assessing convergence in global sensitivity analysis: a review of methods for assessing and monitoring convergence
Socio-Environmental Systems Modelling, vol. 6, 18678, 2024, doi:10.18174/sesmo.18678

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).



Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
<http://www.sesmo.org>

implausible input-output relationships (Guillaume et al., 2019). A classical definition of sensitivity analysis is a pertinent foundation here and is provided by Saltelli et al. (2000) as: “*sensitivity analysis (SA) is the study of how variations in the output of a model (numerical or otherwise) can be apportioned, qualitatively or quantitatively, to different sources of variations, and of how the given model depends upon the information fed into it*”.

SA has comprehensive and diverse literature across many domains and has been fertile ground for the development of new methods, principally for global sensitivity analysis (GSA) whose convergence this paper is concerned with. In the environmental modeling literature, the broad range of uses of GSA has made it a subject of increasing attention, such as in Castaings et al. (2012), Estrada & Diaz (2010), Nossent et al. (2011), Pianosi et al. (2016), Ravalico et al. (2010), Saltelli & Annoni (2010), and Yang (2011). This attention to numerical GSA methods is required because it is usually unrealistic to consider analytical SA methods, in particular for environmental models due to their uncertainty, complexity and non-linear nature (Wang & Solomatine, 2019). Moreover, any method devised for GSA has its own purposes, strengths and weaknesses depending on the model and its problem context. Therefore, computational GSA methods have become a popular means of investigating the influences of model parameters and inputs (both often referred to as factors) on model responses (Khorashadi Zadeh et al., 2017). In the hydrological domain alone, for example, GSA has already been applied widely to well-known models like MODFLOW, VIC, Noah-MP and SWAT (Mai & Tolson, 2019).

In terms of software for computational GSA methods, there is now much available. Douglas-Smith et al. (2020) have summarized the recent nature and trend of mainstream software tools and techniques that have been developed. In regards to development of convergence assessment in software packages or libraries, Pianosi et al. (2015) created the SAFE toolbox, initially for MATLAB/Octave and now extended to R and Python, to support robustness assessment and convergence of sensitivity indices, along with extensive visualization tools. Designed for both non-specialists and experienced users, it features fully commented code. Moreover, a unique survey (Pianosi et al., 2020) was conducted to measure its success in adoption. Hsieh et al. (2020) developed the *pk sensi* package in R to include a convergence assessment method for sensitivity indices in Sarrazin et al. (2016). Razavi et al. (2019) built a software toolbox *VARS-TOOL* that includes VARS (Razavi & Gupta, 2016b, 2016a), progressive Latin Hypercube sampling (Sheikholeslami & Razavi, 2017), convergence testing to give the “Reliability” convergence measurement for the purpose of ranking in Razavi and Gupta (2016b), and the grouping method for screening in Sheikholeslami et al. (2019a). Furthermore, Sun (2021) implemented several convergence assessment methods for ranking purposes from the literature in a GitHub library *SAConvergenceAnalysis* with open access, while Jakeman (2023) provides metrics for assessing convergence in regard to surrogate-based estimates of sensitivity rankings.

There are several ways of categorizing the use of GSA, which is a fundamental consideration in selecting a GSA approach and assessing the convergence of its results. The commonly applied one separates sensitivity analysis into four application categories: screening (or factor fixing), ranking (or factor prioritization), variance cutting, and factor mapping (Saltelli et al., 2007; Sarrazin et al., 2016). Recently, a modern list of possible SA purposes is summarized as: (a) scientific discovery; (b) dimensionality reduction; (c) data worth assessment; and (d) decision support (Razavi et al., 2021). Depending on the purpose of using GSA for the model of interest, one should choose methods applicable to the appropriate category. Furthermore, by assessing convergence, we mean a practice that involves monitoring and measuring convergence, indicating stability of model factors and confidence against sample size, and reporting on and justifying assumptions/choices that were made in the process.

In any GSA exercise, there are several methodological steps that must be undertaken, and these constitute a workflow as enunciated by Pianosi et al. (2016). Most fundamentally, a series of samples must be taken from the model parameter space and, along with a given set of model inputs or forcings, the model is run forward for each sample to generate a response surface of the model. In the socio-environmental domain, inputs will typically be defined temporally and often spatially as well. Metrics or indices may be chosen to calculate quantities of interest of the model response, which can involve either some function of the model outputs alone or some error measure of the outputs with respect to observations (Shin et al., 2013). Thus, performance metrics of actual model responses is one way of

doing GSA. Razavi and Gupta (Gupta & Razavi, 2018; Razavi & Gupta, 2019) concluded that a specifically targeted aspect of the quantities of interest, a compressed set of properties that characterize the quantities of interest, and the spatio-temporally varying quantities of interest themselves are also types of responses in addition to the performance metrics. For example, Berezowski et al. (2015) utilize spatial distribution rather than metrics to show how the WetSpa model is sensitive to spatial input data. A fundamental consideration is to decide when adequate sampling of parameter space has been achieved that indicates sufficient or acceptable confidence in the metrics calculated (Khorashadi Zadeh et al., 2017).

Attempts to identify the factors that may affect the results of an SA exercise have been tailored to each problem at hand. The number of samples is the most obvious factor impacting the performance of SA. In addition to sample size, various authors have concluded that not only is the model used a key factor but collectively they have suggested the following crucial ones influence the outcomes of an SA exercise: complexity, computational budget, model input related factors (boundary conditions, parameters chosen, prior distribution, parameter ranges, correlation, natural properties, definition, scale, and measurement errors), the defined objective function, the objective(s) of the SA and associated definition of them, the selection of appropriate GSA methods, monitoring of convergence, and estimation of the uncertainty in SA measures (Crosetto et al., 2000; Devak & Dhanya, 2017; Qian & Mahdi, 2020; Song et al., 2015; Wang et al., 2013; Yang, 2011; Yang et al., 2012).

Therefore, many considerations potentially affect the results, and ultimately the convergence, of a GSA. These are largely recognized in Pianosi et al. (2016) in their proposed complete workflow on the application of sensitivity analysis methods. They characterized the performance of a GSA exercise as being impacted by eight influences: experimental set-up, the GSA method (often called the estimator of specified sensitivity indices), the input/factor variability space, the sampling strategy, the sample size, the robustness and convergence assessment, visualization of results, and assessment of credibility with respect to the SA results in the sense of matching underlying assumptions. They also mentioned issues with respect to observational errors in the forcing and model response data, the potential for model emulation, and dealing with unsatisfactory model behavior. Furthermore, the authors extended their workflow depiction to provide a systematic review of SA methods and linked SA with other fields including uncertainty analysis, model calibration, model diagnostic evaluation, dominant controls analysis, decision making, and model emulation.

In GSA, the approach to assessing convergence markedly differs between using quantitative and qualitative measures, reflecting their intrinsic methodological disparities. Quantitative measures offer numerical estimations of the impacts that parameters have on model outputs, facilitating precise comparisons and rankings. In contrast, qualitative measures provide insights into the existence and patterns of sensitivities without assigning explicit numerical values, thereby prioritizing the identification of influential parameters rather than quantifying their impact. Consequently, convergence in quantitative measures is typically assessed through numerical thresholds or criteria, whereas qualitative measures rely on the consistency of observed patterns across successive model simulations.

This divergence in qualitative and quantitative approaches underscores the complexity of establishing a unified framework for convergence assessment of GSA methods. The challenges are further compounded by the findings of Sarrazin et al. (2016). They highlighted existing gaps in the practice of GSA convergence, noting that there is lack of uniformity in how convergence is defined, convergence criteria and thresholds are often not clearly established, and the number of samples required can vary significantly for different models even when applying the same GSA method. These observations suggest that although the GSA community has made significant advancement in the GSA field, the application of convergence analysis exhibits variability that could impact the reliability of GSA measures.

It is important to acknowledge that sampling strategies significantly influence the outcomes of various sensitivity analysis methods. While they do not directly measure convergence, sampling strategies significantly contribute to the efficiency of convergence assessment by optimizing the number of samples needed. Numerous studies have reported how different sampling methods affect the rates of convergence, underscoring that more effective sampling strategies (in consort with convergence

assessment) can enhance the exploration of the parameter space (Anstett-Collin et al., 2015; Janssen, 2013). This is because the performance of sampling methods, along with the quality of the generated samples, directly determines the efficiency and robustness of any sampling-based analysis (Sheikholeslami & Razavi, 2017). However, as the focus of this paper is not on sampling methods, we will not delve into this topic in depth.

In this paper, we aim to summarize the contributions of different convergence assessment methods from previous convergence studies of GSA. A comprehensive historical perspective, specifying what has been contributed by a large number of authors in the environmental literature with respect to the choice of methods and models investigated, is undertaken so that readers can easily identify where they might search for particular information. But the historical perspective also serves to indicate that so-called findings tend to be case-dependent in regard to models, methods and choices therein. This case dependency reinforces the need for decision choices in the GSA procedure to be transparent and, wherever possible, alternatives at least to be discussed if not investigated for appreciating the robustness of outcomes with respect to decision choices. Nevertheless, we suggest some prime choices to consider when undertaking a convergence analysis for its assessment. Next, convergence assessment methods based on different application purposes are introduced and discussed.

2. Review of convergence assessment

This section brings together literature on the techniques that have been developed for assessing the convergence of GSA methods. The connections and relationships among the convergence assessment methods discussed in this section are depicted in Figure 1, and a comprehensive list of corresponding references is provided in Table 1. Although the selection of a threshold for convergence reached can be a somewhat arbitrary factor or depend on computational budget, it should still be selected based on thoughtful and careful considerations. The threshold selected affects not only the final sample size attained but also the speed in reaching convergence. In simple terms, convergence is essentially reached when the results for sensitivity measures of interest and pertinence do not change within a certain tolerance by adding more model runs.

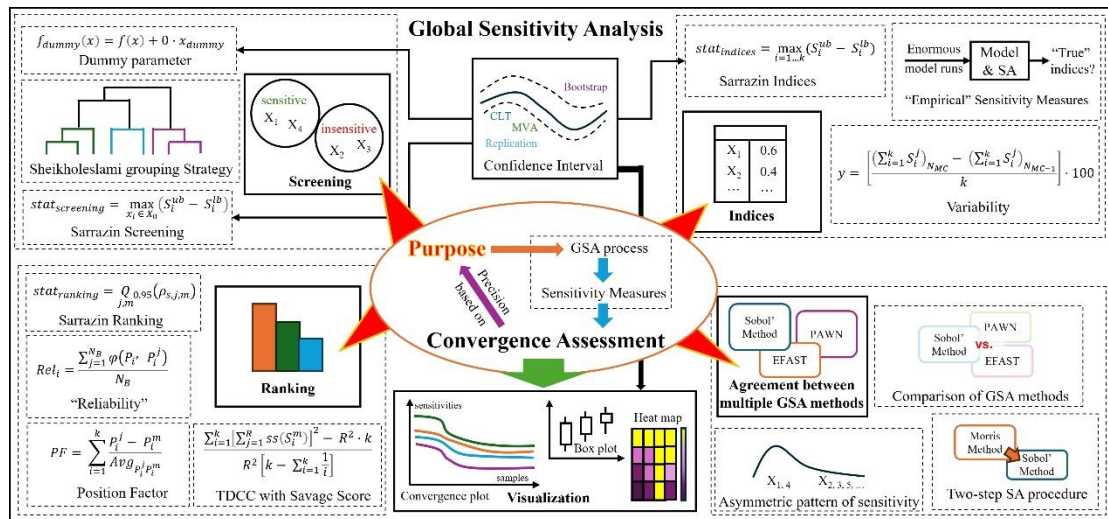


Figure 1: The connection and relationships among the convergence assessment methods discussed in this paper are illustrated. The process of GSA initiates with identifying the purpose and acquiring sensitivity measures. The choice of convergence assessment method is then precisely aligned to main accuracy. Based on this purpose, the methodologies for monitoring convergence can be broadly categorized into screening, ranking, employing indices, and evaluating the agreement between multiple GSA methods. The confidence interval plays a critical component in various methods, enhancing the robustness of the analysis. The culmination of this process is the visualization, which facilitates the interpretation and communication of the findings.

Depending on the purpose of the sensitivity analysis study, the actual measurement of convergence may differ. Sarrazin et al. (2016) and Awad et al. (2019) both argued that the convergence of sensitivity quantities should be measured differently based on the required precision for indices, ranking, or screening. In the literature, the purpose “screening” is often defined as the procedure of classifying model parameters into separate groups according to their sensitivity levels, guided by specific criteria. On the other hand, the purpose “ranking” refers to the act of organizing the model parameters in the order based on their relative significance in terms of sensitivity measures. Finally, the purpose “indices” pertains to the process of obtaining accurate sensitivity measures within a specified level of confidence. In this section, we review and comment on the existing methods for monitoring the convergence of GSA results.

Table 1: References related to methods for assessing convergence rates (Note *: unclear as to who proposed the method first; **: paper that proposed the method first; ***: paper that applied the method but not the one which proposed the method)

CONVERGENCE ASSESSMENT	References
Precision of convergence based on purpose	(Awad et al., 2019; Sarrazin et al., 2016)
The sum of sensitivities of insensitive parameters may be significant	(Hart & Gremaud, 2019; Hartmann et al., 2018; Mailier et al., 2011; Touzani & Busby, 2014; Zhang et al., 2013)
Sheikholeslami grouping strategy	(Sheikholeslami et al., 2019a)** (Huo et al., 2019; Khan & Kaklis, 2021; Sheikholeslami et al., 2021; Sheikholeslami et al., 2019b)
Dummy parameter*	(Khorashadi Zadeh et al., 2017; Liu et al., 2019a; Peeters et al., 2018; Peng et al., 2020; Sun et al., 2022; Upreti et al., 2020)
• Discussion on dummy parameter	(Castaings et al., 2012; Khorashadi Zadeh et al., 2017; Mai & Tolson, 2019; Marino et al., 2008)
Confidence Interval (CI)	
• Bootstrap	(Archer et al., 1997; Isaksson et al., 2008)
- Suggestions on CI threshold	(Baroni et al., 2018; Ghasemizade et al., 2017; Herman et al., 2013; Zhan & Zhang, 2013; Zhang et al., 2013)
• Central Limit Theorem*	(Yang, 2011)
• Replication*	(Sun et al., 2021, 2022; Tarantola et al., 2012)
• Model Variable Augmentation (MVA)	(Mai & Tolson, 2019)
"Empirical" Sensitivity Measures*	(Dai & Ye, 2015; Mai & Tolson, 2019; Sheikholeslami et al., 2021; Sheikholeslami & Razavi, 2017; Wang et al., 2020)
Agreement between multiple GSA methods	
• Set reference method	(Moreau et al., 2013; Sarrazin et al., 2016; Uliana et al., 2019; Zhan et al., 2013)
• Two-step SA procedure	(Faggianelli et al., 2017; Moreau et al., 2013; Song et al., 2013a, b; Specka et al., 2019)
• Small group of parameters dominant	(Saltelli et al., 2004, 2007; Uliana et al., 2019; Wagener & Pianosi, 2019)
• Asymmetric pattern of sensitivity	(Saltelli et al., 2004)
• Agreement on dominant parameters	(Confalonieri et al., 2010a)
• No agreement on nondominant parameters	(Cosenza et al., 2013; Sheikholeslami et al., 2017; Song et al., 2012)
• Contradicting results between multiple GSA	(Cloke et al., 2008; Kavetski & Clark, 2010; Razavi & Gupta, 2019; Reusser et al., 2011; Saltelli & Bolado, 1998; Tang et al., 2007)
• Comparison of GSA methods	(Ciric et al., 2012; Confalonieri et al., 2010a, b; Gupta & Razavi, 2018; Kroll et al., 2016; Medina & Muñoz, 2020a; Paleari et al., 2021; Sheikholeslami et al., 2017; Sun et al., 2022)
Sarrazin et al. Convergence Formula	(Sarrazin et al., 2016)
• Proposed similar formula	(Ghasemizade et al., 2017; Gilquin et al., 2021; Leolini et al., 2018; Reinhart et al., 2020)

(Table continued on next page)

Table 1 (Continued)

CONVERGENCE ASSESSMENT	References
"Variability"	(Vanrolleghem et al., 2015)
"Reliability"	(Razavi & Gupta, 2016b)
• Extended "Reliability"	(Sheikholeslami et al., 2019a)
Position Factor	(Ruano et al., 2012)
• Suggested threshold	(Branger et al., 2015; Likhachev, 2019)
• Cosenza version	(Cosenza et al., 2013)
• Robles version	(Robles et al., 2014a, 2014b)
• Liu version	(Liu et al., 2019b)
Top-down Coefficient of Concordance (TDCC)	(Iman & Conover, 1987)
• Quantification of Plasticity	(Confalonieri et al., 2012)** (Paleari & Confalonieri, 2016; Ravasi et al., 2020; Silvestro et al., 2017; Touhami et al., 2013)
• Other comparable methods	
- Calibrated visual criterion	(Garcia et al., 2019)
- Savage score	(Gilardelli et al., 2018)***
- Kendall's coefficient of concordance	(Nguyen & Reiter, 2015)***
- Vigna correlation index	(Vigna, 2015)
- Other rank coefficient of concordance	(Borroni, 2013; Coolen-Maturi, 2014)

2.1 Arbitrary thresholds and grouping

In defining a threshold for convergence of a measure, one should be careful to recognize potential risks. Unfortunately, it is still quite common for researchers to set an arbitrary threshold for identifying sensitive or insensitive parameters.

It is risky to seek convergence by removing non-sensitive parameters from the model due to possible correlations with sensitive parameters (Mailier et al., 2011), and this may also result in model variations that are not explained in the lower dimensional space (Hart & Gremaud, 2019). Moreover, the sum of small sensitivity index values of certain parameters may constitute a significant proportion of the output variance (Touzani & Busby, 2014) to be considered. One example in Zhang et al. (2013) is that 10% of the total-order effect (combination of the first-order and total-effect) was set as the threshold for sensitivity, but the sum of index values of insensitive parameters was actually higher than 10% in a few cases. In the study proposed by Hartmann et al. (2018), 0.3 was chosen to be the threshold for neglecting parameters, but those parameters in fact had larger sensitivities in total than the sensitive ones. Additionally, Wang et al. (2022) proposed a proof-of-concept adaptive method. They investigated the impact of excluding those factors that have no effect, or negligible effect, on the quantities of interest. They warned that fixing should be based on its impact on the question being asked, and that the default values chosen for factor fixing can considerably influence the anticipated error.

For screening, invoking an assumed threshold seems to be the most prevalent approach in literature. Some examples here bear testimony. Neumann (2012) used a value for the first-order index greater than 0.1 as sensitive and 0.05 for the total-effect. A first-order index value greater than 0.01 has been commonly used as a sensitivity threshold for EFAST (Cosenza et al., 2014; Peng et al., 2020; Vanrolleghem et al., 2015), and for the total-effect index value of less than 0.1 has been used to define non-influential parameters (Cosenza et al., 2014).

Zhan and Zhang (2013) used 0.05 to separate sensitive parameters, whereas sensitive parameters slightly larger than 0.05 were set as a separate group deemed slightly sensitive for both the Sobol' method and Importance Measurement. Tang et al. (2007) used 1% of variance as the threshold for sensitive parameters, and also set a highly sensitive group where each parameter contributed more than 10% of the variance. In addition, various approaches have been taken to group parameters into

three or more categories. For example, for the extended-FAST method, Zhao et al. (2019) partitioned the sensitivity of parameters into low (0.05 to 0.1), medium (0.1 to 0.2) and high (> 0.2) groups. In general, 0.1 and 0.05 are the most common choices for variance-based sensitivity analysis methods when choosing a threshold for screening/grouping purposes, a behavior to mimic the 90% or 95% confidence interval. However, the choice of arbitrary threshold(s) is not recommended, as it is unlikely to avoid Type I or II errors from the ensuing classifications. Additionally, the sensitivity measures may have confidence intervals themselves, and the lower bound and the upper bound of the confidence intervals could sit on different sides of the threshold; thus, these sensitivity measures representing particular parameters may be recognized as a different sensitive group in a different sensitivity analysis experiment. Building on this, it has been demonstrated that parameters with higher sensitivities can achieve their final ranking or indices outcomes more quickly, even with limited sample sizes. This is because less sensitive parameters continue to shift their rankings and introduce fluctuations in their indices even with minor changes in sample size (Khorashadi Zadeh et al., 2017; Nossent et al., 2011).

In advancing the use of grouping, Sheikholeslami et al. (2019a) reviewed existing grouping strategies for GSA methods and proposed a new factor grouping strategy with a detailed flowchart, intended especially for high-dimensional models. This newly developed factor grouping strategy utilizes both bootstrap and agglomerative hierarchical clustering to group factors, and it also uses two different strategies (an elbow method or a minimum robustness-based method) to find the optimal number of groups. This grouping method was tested on the Sobol' G-function and a highly-parameterized MESH model to achieve a noticeable reduction in computational effort.

Although many studies have mentioned the latter grouping method (Gabriele & Francke, 2020; Do & Razavi, 2020; Gupta & Razavi, 2018; KC et al., 2020; Koo et al., 2020; Şalap-Ayça et al., 2021), only four studies other than the original in Sheikholeslami et al. (2019a) seem to have implemented this grouping method for actual applications. Thus, Huo et al. (2019) used it and found it required less parameter sampling. To test the performance of the data-driven VISCOUS method, Sheikholeslami et al. (2021) implemented the grouping method to identify crucial processes. In Sheikholeslami et al. (2019b), this grouping method was also applied to the study of the STAR-VARS and Sobol' methods on the HBV-SASK rainfall-runoff model and the MESH land surface-hydrology model. It was also applied in Khan and Kaklis (2021) to a 104-parameter computer-aided ship design based on Free-Form Deformation.

2.2 Dummy parameters

Dummy parameters are sometimes used to assess convergence for some GSA exercises. The approach consists of adding a dummy parameter in the calculation of sensitivities without modifying the actual model. An easy set-up of the dummy parameter for any particular model output $f(x)$ is

$$f_{dummy}(x) = f(x) + 0 \cdot x_{dummy}$$

The sensitivity of the dummy parameter is directly estimated through the fundamental principles of GSA using different independent sample sets. By design, the variability of dummy parameters does not theoretically influence the model outputs nor the sensitivity estimates for the other parameters. Moreover, the calculation of the sensitivity measures for the dummy parameters does not increase the intended number of model runs, and the sensitivity of a dummy parameter provides a potential threshold for approximation of the error for the sensitivity analysis (Khorashadi Zadeh et al., 2017). Several attempts have been made to utilize the dummy parameter approach in GSA: as the influential threshold (Khorashadi Zadeh et al., 2017; Upreti et al., 2020); approximation of random noise (Peeters et al., 2018); judgement of whether the sensitivity indexes of certain parameters are significantly different from zero (Liu et al., 2019a); the accuracy of activity scores based on different gradient approximation methods (Sun et al., 2022); and validation of the effectiveness of selected screening thresholds (Peng et al., 2020).

The effectiveness of the dummy parameter strategy is seriously impacted by the sample size as large errors in dummy parameter sensitivity can be observed at low sample size (Castaings et al., 2012). In addition, Castaings et al. (2012) found for density-based sensitivity analysis applied to the Ishigami-Homma function that the error in sensitivity of the dummy parameter significantly reduces when the

number of replicates grows toward the number of base samples. Khorashadi Zadeh et al. (2017) stated that, in theory, the index value of dummy parameters should ultimately converge to zero with a large enough number of samples. However, Mai and Tolson (2019) obtained a non-zero index value for their dummy parameter even with a sufficiently large number of samples. Marino et al. (2008) also pointed out that the index value of dummy parameters would be small but non-zero due to aliasing and interference effects. Thus, the non-zero approximation error yielded by dummy parameters makes it unrealistic to obtain an extremely precise index value for sample-based sensitivity analysis methods for complex models when faced with a limited computational budget.

The dummy parameter approach does not function for certain GSA methods, such as the Morris method, and indeed the variance-based Sobol' method with specific estimators. The reason is that dummy parameters can only be used when permitted by the sampling method. If only one parameter is being varied at a time, then dummy parameters are not effective. For example, the Morris method aggregates elementary effects by permuting a single model parameter each time to provide a global view of sensitivity, and this renders the dummy parameter to always be zero. In regard to the variance-based Sobol' method, Khorashadi Zadeh et al. (2017) used the Sobol' 1993 estimator (Sobol, 1993) for the first-order sensitivity index and the Homma 1996 estimator (Homma & Saltelli, 1996) for the total-effect sensitivity index. These two estimators require the evaluation of the expected value of the model output f_0 , but f_0 is the only term where the approximation errors of the dummy parameter derive. With the use of other estimators for the first-order and total-effect indices, the expected value f_0 is not needed, and the dummy parameter will always have zero sensitivity. Furthermore, it has been argued that the Sobol' 1993 and Homma 1996 estimators not be recommended anymore because these two estimators are inefficient compared to other estimators, such as the Formula (b) and Jansen 1999 proposed in Saltelli et al. (2010). Additionally, employing a single dummy parameter may be insufficient for accurately assessing approximation errors; therefore, utilizing multiple dummy parameters is recommended to comprehensively evaluate their sensitivity bounds without requiring significant extra computational effort.

2.3 Sequential approach

The sequential approach, or multi-stage sampling (Sheikholeslami & Razavi, 2017), involves increasing the number of samples required for a sensitivity analysis study step by step, allowing the stability of index values to be examined, often visually (Vanrolleghem et al., 2015), till the difference between two consecutive steps is within a certain tolerance (Benedetti et al., 2011; Cosenza et al., 2014). In GSA studies that do not set the sample size *a priori*, there is reasonably widespread usage of the sequential approach since the plots involved can easily display whether or not the trend in the sensitivity measure has converged to a specific value.

The sequential approach can be used alone or coupled with other convergence monitoring methods such as bootstrapping. To inspect the convergence of estimated indices, Zhan and Zhang (2013), for example, gradually increased the base sample size in their study by 100 uniform steps. Tang et al. (2007) also examined statistical convergence as a function of increasing sample size to find the sufficient sample size for LHS in Regional SA in evaluating the lumped Sacramento soil moisture account model, which is in contrast to directly following the sample factor approach suggested by Sieber and Uhlenbrook (2005). Hart and Gremaud (2019) coupled the sequential approach with replication of Sobol' indices to understand the sampling variability in the Sobol' G-function.

In the application of the sequential approach, visualization plays a crucial role in elucidating the convergence process and the stability of sensitivity indices. Various visualization techniques, such as convergence plots, offer intuitive insights into the stepwise refinement inherent in the sequential approach. These visualization tools not only support the interpretation of the incremental results but also provide a graphical representation of the convergence status, underscoring the synergy between these methods.

2.4 Confidence interval

A common way of monitoring convergence in sensitivity analysis and quantifying the confidence interval (CI) of measures is by using bootstrapping. There are many different settings for the bootstrap used in the SA literature in regard to confidence interval level and number of resamples from the original set, such as 95% CI with 100, 1000 or 2000 resamples, 95% CI along with 25th and 75th percentiles for 500 resamples, and 300 resamples without indicating the percentage of CI required (e.g. Garcia et al., 2019; Huo et al., 2019; Nossent et al., 2011; Tang et al., 2007; Wang et al., 2018). Even with so many different choices of the number of resamples for bootstrapping, 95% seems to have been the most universal choice of CI width. Of course, the choice of width of a confidence interval can be arbitrary and should reflect the purpose and context of a study. Essentially, it is a question of whether a 95% CI is too small or too large for certain model studies. For example, Baroni et al. (2018) set convergence as being reached by increasing the sample size of the Sobol' method until the upper bound of the 95% CI is less than the threshold of 0.1. The 95% CI has also been chosen in many other studies (Ghasemizade et al., 2017; Zhan & Zhang, 2013; Zhang et al., 2013), whilst Herman et al. (2013) stated that “convergence was considered acceptable if the 95% confidence interval represented less than 10% of the sensitivity index value for the most sensitive parameters”.

Despite the common use of bootstrapping, it has limitations and is not suitable in every case. Archer et al. (1997) pointed out that reliable percentiles need a large number of original samples, and a skewed bootstrap distribution can also impact the performance. Isaksson et al. (2008) found that bootstrapping behaves poorly with limited samples in the application of SA for computationally intensive models. In other words, the performance and reliability of bootstrapping are significantly impacted by the number of samples. Furthermore, the choice of bootstrapping with replacement (Nossent et al., 2011; Yang, 2011) or without replacement (Pappenberger et al., 2008) can significantly impact the confidence interval. As stated in the supplementary file of Khorashadi Zadeh et al. (2017), bootstrapping with replacement would produce a biased and overestimated confidence interval due to “small vertical jumps in the empirical distributions” caused by the multiple presences of the same samples, and this is resolved by using bootstrapping without replacement.

Other than bootstrapping, there have been other choices for calculating the confidence interval. The Central limit theorem (CLT) with 95% CI was tested by Yang (2011) along with the bootstrap; however, Yang found that CLT performs poorly for complicated models and requires more model runs than bootstrapping for a 5-parameter model. The use of replications, where independent sets of samples are taken and the model output re-evaluated with each new sample set, has also been applied in several studies (Sun et al., 2021, 2022; Tarantola et al., 2012). Thus, the standard errors of all the replications can be used to estimate the chosen confidence interval. The choice between bootstrapping and replication depends on computational budget, but replication has a better chance of exploring previously missed parts of the parameter space, whereas bootstrapping can only extract information from a set group of sampled parameter values. Indeed, methods that use existing samples (e.g., bootstrapping) by definition are limited to the existing sample set. If there is part of parameter space that has not been explored in the existing sample set, then the analysis results will be biased. Only replication has the capacity to include points in areas that were underrepresented in the original sample set.

Recently, Mai and Tolson (2019) proposed a new method to compete with bootstrapping called Model Variable Augmentation (MVA), which is able to operate at low sample sizes where bootstrapping is known to be unreliable; however, the implementation of MVA for measuring confidence intervals has to be decided a priori. MVA is defined as

$$y_{MVA} = 0 \cdot z_0 + z_1 f(x) - z_2 f(x) + c f(x),$$

where $f(x)$ is the original model output, z_0 , z_1 , and z_2 are augmented variables, and c is a constant to keep the variance of the new model output y_{MVA} the same as the original one.

Furthermore, MVA uses the concept of the dummy parameter, shown as the term $0 \cdot z_0$, which may render this method inappropriate under certain circumstances (see Section 2.2 Dummy Parameter).

Unfortunately, there has been no further testing or application of MVA, even among the eight documents that cited the work in Mai and Tolson (2019).

2.5 “Empirical” sensitivity measures

Even though many models, especially environmental models, can be complex, computationally heavy, and not amenable to analytic calculations of exact sensitivity indices, “true” sensitivity indices have been calculated by taking a huge number of samples and, accordingly, model runs. Although the term “true” has been utilized in the literature to describe sensitivity indices achieved through extensive sampling, we have elected to change this terminology to “empirical”, to reflect the nature of these measures more appropriately as based on observed data and analysis, acknowledging the inherent approximations and methodological choices involved in sensitivity analysis. For example, Wang et al. (2020) used 10^6 Monte Carlo simulations to calculate the “empirical” Sobol’ sensitivity indices for the 10-dimensional HBV-SASK hydrological model. Dai and Ye (2015) obtained the so-called reference values of indices from 2×10^6 quasi-Monte Carlo samples and associated model simulations (8×10^6 of model runs) for calculating the absolute errors in total-effect sensitivity index. Sheikholeslami et al. (2017) took 5×10^5 LHS parameter sets to generate the “empirical” CDF of model responses. In Mai and Tolson (2019), “the true rankings of the inputs were assigned based on the rankings from a 10^5 model run SA”. Sheikholeslami et al. (2021) used approximately 1 million model runs to obtain the “empirical” sensitivity indices of the Sobol’ method for the HBV and VIC hydrological models. This calculation of “empirical” sensitivity indices can be a good way of confirming the sensitivity results at relatively low sample size but may not be practical for environmental model studies that are constrained by the computational budget.

2.6 Agreement between multiple SA methods

Assessing the agreement between different sensitivity analysis methods is also used to check the robustness of a sensitivity analysis and, of course, such assessment can be used in conjunction with convergence analysis. Due to fundamental differences in the basis of GSA methods from various categories, it can be reassuring if the importance of a set of parameters or inferences is, in general, confirmed across different methods. A simple approach is to set one method as the reference method and check if other methods agree with this it.

Campolongo et al. (2007) proposed the Morris method as a sensible screening tool to use in sequence with more computationally demanding methods. This approach is corroborated by several studies (Moreau et al., 2013; Sarrazin et al., 2016; Song et al., 2013a; Zhan et al., 2013) as stated by Uliana et al. (2019). In fact, there are several papers that have used the Morris method first to reduce the number of parameters and then hand the reduced model to another GSA method (usually variance-based SA methods) for further analysis. Such a strategy has been called a two-step sensitivity analysis procedure. Faggianelli et al. (2017), for example, used the Morris method to screen a 112-parameter office building model first, then applied the Sobol’ method to the 10 most important parameters identified by the Morris method. Specka et al. (2019) also used the two-step SA, with the Morris method first then EFAST, applied to a 200-parameter Agro-ecosystem MONICA model. Song et al. (2013b) applied the Morris method first then EFAST on a forest growth model 3-PG. Similarly, Moreau et al. (2013) used Morris to screen first, leaving the 6 most important parameters for the analysis of variance (ANOVA) with fractional factorial design involved.

Various studies have shown that only a small group of parameters are dominant in regard to sensitivity no matter how complex the model is (Saltelli et al., 2004, 2007; Uliana et al., 2019; Wagener & Pianosi, 2019), which indicates an asymmetric pattern in the distribution of sensitivity for model parameters (Saltelli et al., 2004). Additionally, for a Water Accounting Rice Model, Confalonieri et al. (2010a) found agreement between GSA methods in terms of the similar importance of parameters. For insensitive parameters, however, this agreement may not exist (Cosenza et al., 2013; Sheikholeslami et al., 2017; Song et al., 2012). This phenomenon explains the intention of using different GSA methods for the same study for the purpose of comparison in terms of efficiency, computational cost and, of course, the sensitivity results. Among them, Tang et al. and Yang (Tang et al., 2007; Yang, 2011) both pointed out the superiority of the Sobol’ method for nonlinear models with strong interactions, though Herman et

al. (2013) questioned the efficiency of the Sobol' method in a spatially-distributed hydrological case.

It has been shown that EFAST performs better and converges earlier than the Sobol' method (Zhao et al., 2014), and other studies have made a similar observation (Gómez-Delgado & Tarantola, 2006; Wang & Solomatine, 2019). However, Wang and Solomatine (2019) used negative sensitivity index values in supporting this argument, which means that the sampling was insufficient. Later on, Upreti et al. (2020) stated that PAWN is more computationally efficient compared to EFAST. Additionally, KC et al. (2021) compared the Morris, Sobol', EFAST and PAWN methods for estimating sensitivity measures for empirical fire spread models, and they concluded that PAWN converged fastest with EFAST being the slowest. On the other hand, Khorashadi Zadeh et al. (2017) found no difference in terms of convergence rate between the Sobol' method and PAWN. Of course, the above results are an indication that no GSA method is ideal in all circumstances, including in regard to convergence rate (Song et al., 2015) but rather depends on the purpose of the SA exercise, the quantitative (be able to quantify the sensitivity estimates) or qualitative (be able to rank the parameters or set a threshold in the order of importance) aspects of the SA method, and characteristics of the model in question.

There are various methods to measure the agreement between multiple GSA methods. Spearman's rank correlation has been used for comparing the ranking of Morris and Sobol' sensitivity results in several studies (Cosenza et al., 2013; Herman et al., 2013; Song et al., 2012). Additionally, Cosenza et al. (2013) listed many methods comparing the agreement of multiple GSA methods such as relevance, number of simulations, visual comparisons of scatter plots of sensitivity indices, and Venn diagrams to visualize classification into important or non-influential parameters. However, neither Spearman's rank correlation nor Venn diagrams have seen much discussion in terms of their strengths but rather purely employment. Position Factor and Top-down coefficient of concordance (TDCC) are also used for comparing the agreement between GSA methods, and they will be introduced in following subsections.

Nevertheless, various studies (Cloke et al., 2008; Tang et al., 2007) have reported that different GSA methods can yield contradictory sensitivity results for the same model application. To explain these observations, Razavi and Gupta (2019) argued that the differences in the fundamental principles and philosophies of the different GSA approaches cause the variations in behavior. Reusser et al. (2011) reasoned four possible sources of this being: a response surface that is rough (Kavetski & Clark, 2010), interference errors (Saltelli & Bolado, 1998), sampling methods used, and algorithms computing the partial variance. There are many other studies (Ciric et al., 2012; Confalonieri et al., 2010a; Gupta & Razavi, 2018; Kroll et al., 2016; Medina & Muñoz, 2020a; Paleari et al., 2021; Sheikholeslami et al., 2017; Sun et al., 2022) that have compared multiple SA methods, but their results will not be applicable in all contexts. Future studies should carefully consider variations in choices among their own experimental set-ups and other possible influences on results so that conclusions are more conditional and transparent by stipulating the assumptions and context relevant to the study.

2.7 Sarrazin et al. convergence formulas

Sarrazin et al. (2016) developed three convergence criteria illustrated through empirical studies (a 5-parameter HyMod model, a 13-parameter HBV model, and a 50-parameter SWAT model). One criterion is the adjusted and weighted rank correlation coefficient coupled with the bootstrap for ranking, whereas the other two criteria assess the convergence of the Morris and Sobol' methods by keeping the maximum difference of the upper bound and lower bound of sensitivity results in a certain range for screening and indices. These convergence criteria are also employed in subsequent application to a vegetation-recharge model (Sarrazin et al., 2018). For assessing the convergence of a sensitivity index value, they proposed a criterion called $stat_{indices}$, and it is defined as

$$stat_{indices} = \max_{i=1\dots k} (S_i^{ub} - S_i^{lb}),$$

where S_i^{ub} and S_i^{lb} are the upper and lower bounds of the sensitivity index value of the i -th parameter, and k is the number of parameters. Sarrazin et al. use the width of the 95% confidence interval obtained by bootstrapping and found $stat_{indices} = 0.05$ to be a reasonable threshold for the criterion to indicate convergence.

For ranking of sensitivities, Sarrazin et al. (2016) modified and weighted a rank correlation coefficient as the criterion called $stat_{ranking}$, and it is defined as

$$\rho_{s,j,m} = \sum_{i=1}^k |R_i^j - R_i^m| \frac{\max(S_i^j, S_i^m)^2}{\sum_{i=1}^k \max(S_i^j, S_i^m)^2},$$

where S_i^j and S_i^k are the index values and R_i^j and R_i^k are the ranks of i -th parameter using j -th and m -th samples accordingly. In addition, a single scalar was developed to aggregate ρ of all possible pairs with

$$stat_{ranking} = Q_{0.95}(\rho_{s,j,m}),$$

and convergence is considered reached when it is below 1. The authors also proposed a similar measure $stat_{screening}$ for screening, and it requires a pre-defined threshold T to be set in order to form a subset X_0 of less sensitive model parameters:

$$X_0 = \{x_i \text{ when } S_i < T\},$$

and the indicator using the subset X_0 is defined as

$$stat_{screening} = \max_{x_i \in X_0} (S_i^{ub} - S_i^{lb}),$$

where x_i is the i -th model parameter, S_i is the sensitivity measure of x_i , and S_i^{ub} and S_i^{lb} are the upper bound and lower bound of the 95% confidence interval of S_i correspondingly. The threshold T is set as 0.05 by Sarrazin et al. to take the subset as lower-sensitivity rather than non-sensitive. The convergence of the screening result is considered reached when $stat_{screening}$ is below 0.05.

Numerous studies have mentioned, implemented, or adopted the idea of these criteria for assessing convergence, while other studies (Asensio-Sevilla et al., 2020; Chaney et al., 2016; Medina & Muñoz, 2020b) have invoked some of their arguments about convergence or have taken their advice on suggested sample size for a direct application. Among them, Wate et al. (2020) adopted the Sarrazin indices and convergence criteria in using the variance-based Sobol' method for a stochastic building performance simulator (S-BPS) and found them useful for assessing error in estimating indices with a sequential approach. Gschwend et al. (2017) applied the Morris method to investigate different liquid fuels through a thermodynamic engine model, assessing convergence of the absolute elementary effect using the Sarrazin indices. In the context of a physiologically-based pharmacokinetic model, Hsieh et al. (2018) utilized the Sarrazin et al. convergence criteria by checking the range of the 95% confidence interval for both Sobol' indices and Morris indices but with a threshold of 0.1 rather than 0.05. Later, Hsieh et al. (2020) coded an R package *pkensi* to include the $stat_{indices}$ criterion. For a hydrogen predictive model, Seo et al. (2021) used the Sarrazin indices convergence criteria with the Sobol' method, although they obtained negative index values for certain parameters even when the criteria indicated convergence. KC et al. (2021) compared four GSA methods for empirical fire spread models (Dry Eucalypt and Rothermel), but they interpreted the index criteria of Sarrazin as checking if the maximum difference between consecutive index values is less than 0.05. Similarly, use of the maximum difference in indices of two consecutive runs as a stopping criterion for estimating the Sobol' index can also be seen in other studies (Gilquin et al., 2021; Leolini et al., 2018), though the threshold can be different. Reinhart et al. (2020) used the width of the confidence interval to monitor the convergence of Sobol' indices for a drainage water recycling model, but they visually evaluated the convergence of ranking rather than applying the ranking criteria of Sarrazin et al. (2016). Ghasemizade et al. (2017) considered a width of the confidence interval below 0.14 rather than the default 0.05 as acceptable using the Sobol' method for the HydroGeoSphere model.

A few studies (Chisari et al., 2018; Cruz May et al., 2021; Hsieh et al., 2021) have claimed or implied that their results satisfied the convergence criteria of Sarrazin et al. (2016) but did not show any results related to the criteria. KC et al. (2020) stated that they applied the three criteria from Sarrazin et al. to wildfire models with the Morris, Sobol' and EFAST methods, but did not show much detail in terms of results from the criteria. Awad et al. (2019) compared the Morris extension method and Sobol' method

on a bilinear theoretical model and a civil engineering non-linear model with the use of three criteria.

In terms of criticisms, Garcia et al. (2019), in the context of the Morris method followed by the Sobol' method in a two-step SA, argued that the convergence criteria in Sarrazin et al. (2016) could lead to a computational surcharge when the goal is only to ensure convergence of the most important parameters. Gokarakonda et al. (2019) pointed out that Nguyen and Reiter (2015) did not obtain satisfactory results by using the rank correlation coefficients recommended in Sarrazin et al. (2016), but in fact those authors used Kendall's coefficient of concordance rather than the Sarrazin et al. (2016) criteria.

2.8 "Variability"

To identify if convergence is reached, Vanrolleghem et al. (2015) set a cut-off threshold (CT) for a normalized sum of the sensitivity indices, and the "Variability" of this normalized sum was examined through increasing model runs.

$$\text{Variability } y = \left[\frac{(\sum_{i=1}^k S_i^j)_{N_{MC}} - (\sum_{i=1}^k S_i^j)_{N_{MC-1}}}{k} \right] \cdot 100,$$

where S_i^j is the sensitivity index of i -th input of j -th model output, k is the number of parameters, and N_{MC} and N_{MC-1} are two sets of samples.

Of the 50 plus documents that have cited the "Variability" work of Vanrolleghem et al. (2015) so far, only three papers seem to have applied "Variability" for monitoring convergence. Likhachev (2019) applied the Morris method to the titanium nitride (TiN) B-spline dispersion model and assessed the convergence of μ^* using "Variability" with a threshold of 3% by increasing trajectories. Kroll et al. (2016) used "Variability" to monitor the convergence for the Morris and EFAST methods applied to a wastewater treatment plant model (WWTP). Finally, Salviano et al. (2021) employed "Variability" to check the convergence of the smoothing spline ANOVA method of GSA. However, only the study of Likhachev (2019) presented the values of "Variability" through a plot (cumulative variability of μ^* versus number of trajectories r) to obtain a sufficient number of trajectories, and no value of "Variability" was found in the other two studies.

2.9 "Reliability"

In order to utilize the information provided by bootstrap resamples, Razavi and Gupta (2016b) proposed a measure they termed Reliability for assessing ranking. The "Reliability" indicates the number of resamples in bootstrapping required to provide the same rank as the original sample set, and it is defined as

$$\text{Rel}_i = \frac{\sum_{j=1}^{N_B} \varphi(P_i, P_i^j)}{N_B}$$

where N_B is the number of bootstrap resamples, P_i and P_i^j are the rank of i -th parameter obtained from the original sample set and the j -th bootstrap resample accordingly. The function $\varphi(P_i, P_i^j)$ is defined as

$$\varphi(P_i, P_i^j) = \begin{cases} 0 & P_i \neq P_i^j \\ 1 & P_i = P_i^j \end{cases}$$

This "Reliability" measure was also called a "Robustness" measure in Razavi et al. (2019). Of the 60 plus documents that have cited Razavi and Gupta (2016b), only five studies seem to have used "Reliability" for the assessment of convergence, whereas a few studies (Akomeah et al., 2019; Razavi & Gupta, 2019) have mentioned the "Reliability" measure but did not employ it. Using VARS-50 for a hydrological model,

Bajracharya et al. (2020) compared different model evaluation metrics based on “Reliability”. While the error in the slope of the flow duration curve (SFDC) was found to have low reliability, it was still recommended as SFDC can identify the sensitivity of parameters overlooked by conventional error metrics. For the Noah-MP land surface model, Huo et al. (2019) measured the “robustness” of multivariate adaptive regression splines (MARS), though this measure is basically the originally-termed Reliability measure. To complement a grouping method, Sheikholeslami et al. (2019a) extended use of the Reliability measure to take into account factor grouping. Additionally, Sheikholeslami et al. (2017) assessed the ranking of the Regional SA (RSA) (Hornberger & Spear, 1981) and VARS methods using Reliability for the 10-parameter RIVICE model.

2.10 Position Factor

A measure called Position Factor (PF) has been used in some studies to “evaluate the convergence of ranking (stability in ranking) numerically”, where “parameter ranking is considered to be stable when the value of the position factor is low” (Ruano et al., 2012). Although 66 documents have cited Ruano et al. (2012), only a limited number of studies have actually employed the Position Factor. On the other hand, the Position Factor concept has received a lot of development in the process, as indicated below.

The original Position Factor is defined as

$$PF = \sum_{i=1}^k \frac{P_i^j - P_i^m}{Avg_{P_i^j P_i^m}},$$

where k is the number of parameters, P_i^j and P_i^m are the rank of the i -th parameter obtained from the j -th resample and the m -th resample accordingly, and $Avg_{P_i^j P_i^m}$ is the average of P_i^j and P_i^m .

Likhachev (2019) used the Position Factor to measure the difference in ranking obtained with different numbers of trajectories, considering the threshold of reliable ranking to be when $PF < 2$. Branger et al. (2015) used the Position Factor to monitor the ranking by the Morris method for an energy-economy model Res-IRF. Similarly, they also considered $PF = 2$ as the threshold, which was claimed to be robust to an increase in samples. In addition, they considered the convergence of PF over a range of sample sizes with low values rather than stopping at the first low PF value. Sreedevi et al. (2019) applied the Position Factor for the SHETRAN model to measure the differences in ranking by the Morris method obtained with different trajectories; however, the threshold invoked for the Position Factor is not stated, and the final PF was larger than 2.

A modified position factor PF_{abs} was proposed by Cosenza et al. (2013) to use the absolute value of the ranked difference of parameters between resamples.

$$PF_{abs} = \sum_{i=1}^k \frac{|P_i^j - P_i^m|}{Avg_{P_i^j P_i^m}}$$

Later, Robles et al. (2014a, 2014b) normalized PF_{abs} by dividing by the maximum of the Position Factor PF_{max} . Furthermore, it was considered that the convergence criteria are to obtain two consecutive PF_{norm} values of less than 0.3.

$$PF_{norm} = \sum_{i=1}^k \frac{|P_i^j - P_i^m|}{Avg_{P_i^j P_i^m}} \cdot \frac{1}{PF_{max}}$$

Liu et al. (2019b) modified the Position Factor by changing the numerator to a varied weight based on the sensitivity measures, and this new Position Factor was also compared with the original Position Factor:

$$PF_{weighted} = \sum_{i=1}^k \frac{|\mu_i^{*(j)} - \mu_i^{*(m)}| / \mu_i^{*(j)}}{Avg_{P_i^j P_i^m}}$$

Along with previous studies (Ge & Menendez, 2014; Zhan et al., 2013), Liu et al. (2019b) argued that the significance of a specific parameter depends on the magnitude of the sensitivity measure (say μ^*) rather than the rank. Moreover, a change in rank does not reflect a great change in the sensitivity measure. Thus, it was suggested that the judgment of the original Position Factor purely based on rankings is not wise and that an unnecessarily large number of trajectories would be required to reach convergence.

2.11 Top-down Coefficient of Concordance with Savage score

The Top-down Coefficient of Concordance (TDCC) (Iman & Conover, 1987) measures the correlation between parameter ranks and weights the top-ranking parameters more. It is usually coupled with the Savage score (Savage, 1956), which is the expected value of the order statistic using a sample from a set distribution to form an increasing sequence, to quantify the agreement between multiple sensitivity measures obtained from different sample sets or configurations, especially for the purpose of ranking. The TDCC is defined as

$$TDCC = \frac{\sum_{i=1}^k [\sum_{j=1}^R ss(S_i^m)]^2 - R^2 \cdot k}{R^2 \left[k - \sum_{i=1}^k \frac{1}{i} \right]}$$

where k is the number of parameters, R is the number of resamples, S_i^m is the index value of i -th parameter obtained by m -th resample, and $ss(S_i^m)$ is the savage score of S_i^m defined as

$$ss(S_i^m) = \sum_{i=P_i^m}^k \frac{1}{i}$$

Here P_i^m is the rank of i -th parameter obtained by m -th resample.

Studies (Ciric et al., 2012; Confalonieri et al., 2010a; Confalonieri et al., 2010b; Marino et al., 2008; Paleari et al., 2021) have implemented the TDCC with Savage score for measuring several items: the ranking between multiple SA methods; the adequacy of sample size (Brembilla et al., 2015; Krishnan et al., 2021; Krishnan & Aggarwal, 2018; Lebedeva et al., 2012); for different combinations of inputs (Jabloun et al., 2018); for different simulation days (Specka et al., 2019); for different parameter variation ranges (Tan et al., 2017); or for the difference between pairs of species (Locatelli et al., 2017). For a specific crop model, Confalonieri et al. (2012) proposed a quantification measure of L (of plasticity) that combined TDCC with the standard deviation of a normalized agrometeorological indicator (SAM), and this indicator has been mentioned or used in several studies (Paleari & Confalonieri, 2016; Ravasi et al., 2020; Silvestro et al., 2017; Touhami et al., 2013).

There have been attempts to use the Savage score as a standalone convergence measurement. Gilardelli et al. (2018) used the Savage score to compare the ranking obtained by the Morris and EFAST methods, whilst Campolongo et al. (2007) used it to identify the least sensitive parameters by ordering the sum of scores from each parameter for a revised Morris method. However, Garcia et al. (2019) pointed out that the Savage score assigns a low score to parameters that are only important to one model output in favor of parameters important to multiple model outputs.

TDCC is not the only measure for assessing a coefficient of concordance. Kendall's coefficient of concordance (KCC) was also used in several studies for the purpose of ranking (Nguyen & Reiter, 2015). However, KCC gives equal weight to all of the parameters regardless of their ranking, and Helton et al. (Helton et al., 2005; Locatelli et al., 2017) deemed KCC inappropriate for finding important parameters because KCC serves as a bad indicator when insignificant parameters are the majority of the model

parameters. Although TDCC has been widely adopted in many SA studies, there are other correlation indices warranting consideration in the future. The correlation index proposed by Vigna (2015) considers if there is a tie between the measurements, as ties are not dealt with in the current setting of TDCC. If a model has multiple parameters with the same sensitivity analytically, a correlation index, such as TDCC, is unable to provide a correct ranking indicator, as TDCC gives a much higher value than the threshold even if the sample size is adequate for other ranking measurements. In addition, there are several new weighted rank coefficients of concordance measures (Borroni, 2013; Coolen-Maturi, 2014) that could also be considered and tested in future ranking studies.

2.12 Visualization

The utility of visualization techniques in GSA is particularly pronounced when applied in conjunction with the sequential approach, detailed in Section 2.3. The sequential approach's step-by-step refinement of sample sizes offers rich data for visual analysis, enabling researchers to graphically assess convergence and the impact of incremental sample addition. This interplay between visualization and the sequential approach, or visualization alone, enhances the reliability and robustness of GSA. It not only supports the interpretation of the GSA results (Nossent et al., 2011) but also helps in identifying certain patterns and small details not captured by a single or small number of measures.

Convergence plots graphically represent the trend of sensitivity indices as the number of model evaluations increases (Yang, 2011), allowing researchers to visually assess when additional simulations cease to significantly alter the outcome, or to view the sensitivity indices of multiple model parameters together, thereby visually identifying general patterns. Similarly, plots displaying confidence limits around sensitivity indices can be instrumental in illustrating the uncertainty associated with the estimated sensitivities, thereby offering another layer of insight into the convergence and reliability of the analysis. These plots, as employed in numerous studies such as Sun et al. (2021), make it evident when the confidence intervals around sensitivity indices stabilize, indicating convergence. Khorashadi Zadeh et al. (2017) used probability density function plots of error measures (NSE and ME) as convergence criteria, wherein complete convergence is achieved when the slope of the plot is considered flat. Pianosi et al. (2016) summarized 13 examples of useful visualization tools for exhibiting GSA results in their appendix, including the most commonly seen plots for the convergence purpose, such as the convergence plot and pattern plots. Qian and Mahdi (2020) illustrate several common graphical methods, including the boxplot, confidence plot and convergence plot. Among them, the slope of the convergence plot is used as an indicator to check the status of convergence in many studies, but this depends on the context, how plots are configured, and how the overlap of multiple lines is carefully adjusted to increase visibility. While a variety of visualization techniques in GSA are discussed here, it is crucial to emphasize that not all directly serve convergence analysis. Techniques such as different formats of heat maps (Baroni et al., 2018; Herman et al., 2013; Li & Ren, 2019; Van Werkhoven et al., 2008; Wagener et al., 2009) and 3D plots (Mailier et al., 2011; Matthews et al., 2007; Wang et al., 2013), although valuable for presenting sensitivity analysis results and identifying patterns, may not inherently provide information regarding convergence. Nevertheless, when used in conjunction with sequential approaches, they can complement the convergence assessment by highlighting patterns or discrepancies that warrant further investigation. For example, Likhachev (2019) used 3D plots to show changes in the Morris method's sensitivity measure μ^* with respect to the changes in the number of trajectories and parameters, along with the cross-sections as 2D plots on the side to exhibit the more hidden part.

In order to summarize useful suggestions for avoiding common mistakes and enhancing the usefulness of visualization, Kelleher and Wagener (2011) proposed ten guidelines for scientific visualization. These ten guidelines list comprehensive recommendations including: the data selection, graphic encoding and attributes, the purpose of plotting, axis ranges and aspect ratio, solving overlapping of points in scatter plots, using line connection smartly, aggregate datasets, and the selection of color schemes. Additionally, they emphasize the importance of scientific visualization and give examples of visualization to illustrate their guidelines.

2.13 Other convergence approaches

Other less common convergence approaches include that of Touzani and Busby (2014) who presented an error criterion to assess the convergence for derivative-based global sensitivity measures (DGSM) and emphasized the impact of influential parameters. The criterion uses an Euclidean norm of DGSM indices instead of individual indices. The stopping threshold of this criterion is set as 0.05, but the DGSM indices still showed a large fluctuation in the convergence plots past the threshold. Thus, this error criterion may need a much lower threshold or some other modifications to be more informative.

Garcia et al. (2019) proposed a calibrated visual criterion in the R platform to mimic the process of selecting important parameters by the Morris method in preparing for a two-step SA. In addition, the convergence of the calibrated visual criterion was assessed by using bootstrap resamples, similar to the concept of “Reliability”. It was tested on the FLBEIA bio-economic fisheries simulation model and found to be cheaper computationally than the *stat_{screening}* measure, which is based on the width of confidence interval (Sarrazin et al., 2016).

3. Discussion and future directions

As has been argued here and elsewhere, there are many influences that impact the results of a GSA exercise. Thus, the practice of adopting recommendations from previous studies to select a sample size for GSA a priori or finding simple relations between the number of model parameters and sample size is largely fraught with uncertainty, especially in terms of realizing trustworthy convergence of SA results. Therefore, almost whatever the context of a new study, there are many reasons for the embracing of a sequential approach, as the basis in employing convergence analysis whereby convergence is monitored with increasing sample size.

For the purpose of screening, setting an arbitrary threshold for sensitivity is not recommended as it may cause errors of Type I (i.e., sensitive parameters classified as insensitive) or II (i.e., insensitive parameters classified as sensitive), and it may also underestimate the impact of the sum of less sensitive parameters (Touzani & Busby, 2014). The dummy parameter approach (Khorashadi Zadeh et al., 2017) constitutes use of a more advanced threshold, but it may not be suitable for specific GSA methods such as those of Morris (due to the sampling used) or Sobol’ (for select estimators) methods. However, it can still be implemented with other methods, such as Active Subspaces (Sun et al., 2022), but one should take care implementing it with small sample sizes (Castaings et al., 2012). Many methods, such as the Kendall rank correlation (Nguyen & Reiter, 2015), do not seem to have been implemented much, so there needs to be more investigation of these methods. The grouping method of Sheikholeslami et al. (2019a) can be a suitable choice for multi-group screening, but it still requires more dedicated tests. Similarly, the screening criteria of Sarrazin et al. (2016) is easy and straightforward to use, but it also requires more comprehensive testing against other methods.

For the purpose of ranking, the Position Factor approach has seen several modifications in form, and the modified Position Factor with weighting (Liu et al., 2019b) seems a valuable change as it considers the magnitude of the sensitivity measures in evaluating the ranking. In terms of weighting, the ranking criterion of Sarrazin et al. (2016) uses bootstrap resamples additionally, and this is more stringent than the original Position Factor for detecting convergence. Sun et al. (2022) employed 8 test functions to compare the ranking of Activity Scoring (based on the Active Subspace concept) with the variance-based Sobol’ method and the Morris method. They used four ranking convergence measurements (Position Factor, *stat_{ranking}*, “Reliability”, and TDCC), and found that the different approaches to convergence reflect different perspectives.

The notion of “Reliability” seeks to inspect the convergence status of each individual model parameter, and this gives a different perspective than the Sarrazin et al. (2016) ranking criterion *stat_{ranking}* which provides the convergence status in a general view. “Reliability” and *stat_{ranking}* act complementarily, whereas the original Position Factor does not provide additional information compared to visually identifying the ranking except if the number of model parameters is huge and visual inspection is not practical. The TDCC measure with Savage score is less informative in providing an indicator of

convergence status than the other three ranking measurements because it may require more model evaluations than other ranking measurements. Ideally, all these ranking convergence measurements should be applied and compared under different circumstances with various model characteristics and GSA methods. Fortunately, such comprehensive comparisons can often be achieved by post-processing results from the same sampling and model runs.

The Confidence Interval approach is the most obvious option to assess convergence, especially for ranking or estimating sensitivity indices. Among the choices for calculating confidence intervals, bootstrapping does not require extra model runs compared to replication-based CI, but it also comes with the limitation of being biased under certain circumstances. Model Variable Augmentation (MVA) has not been tested much, other than in Mai and Tolson (2019), and it utilizes the dummy parameter concept, which would have the same drawbacks as that of the dummy parameter alone. The “Variability” approach is quite straightforward in that its formula is simple to implement, but it cannot identify changes in each individual model parameter when the sum of sensitivity measures does not change. The usefulness of a “true”, or more correctly “empirical”, sensitivity measure really depends on how one sets up the sensitivity analysis, including which methods are chosen. “Empirical” sensitivity analysis works well for simple analytical functions, but for real-world problems, it is not always practical. Running a lot of samples does not necessarily mean you will find the absolute true sensitivity; it just moves one closer to some converged value, typically with no mathematical proof in support. So, for complex models or when the computational budget is too restrictive, the “empirical” sensitivity might not suffice though the general trends against sample size may be informative. Last but not least, the indices criteria of Sarrazin et al. (2016), as mentioned earlier, cannot directly be used to ensure the convergence of individual model parameters, and it also relies on the quality of bootstrap resamples.

Looking to the future, GSA stands on the cusp of significant advancements. Opportunities abound for applied studies focused on the convergence of sensitivity analysis that are currently poorly represented in the literature. Future GSA studies should fully recognize all potential choices that impact the inferences from a GSA exercise and, in the interests of transparency, note the assumptions made and hence the conditional nature of the outcomes, including with respect to the convergence attained. Secondly, multi-step GSA approaches deserve more attention around applying possible combinations of GSA methods at each step, as the current studies have largely but not totally been limited to the Morris method followed by one variance-based GSA method.

In addition, while some studies, such as Sarrazin et al. (2016), have begun to address the rate of convergence by extracting and discussing convergence rates from various studies, this area remains significantly underexplored. Figure 1 in Sarrazin et al. (2016) exemplifies the kind of empirical insights that can inform our understanding of convergence rates in GSA. This acknowledgment highlights that, although some work has been done, comprehensive testing, investigation and analysis remain paramount for advancing our decisions and ensuring more reliable GSA results. While such studies lay a foundation, enhancing GSA software tools is essential for advancing the field and ensuring reliable results. Future efforts should focus on detailed convergence indicators across varying sample sizes to refine GSA methods.

If the response surface of the model in question is smooth, emulation has much potential to be a more computationally efficient approach (Budamala & Baburao Mahindrakar, 2020; Wang et al., 2014), as it can simplify and reduce the dimension of models through the development of model response surfaces (Yang et al., 2018) and reduce sampling and the associated number of model runs (Wang et al., 2020). Furthermore, emulation methods such as Polynomial Chaos (Hu et al., 2015; Shin et al., 2015) and Artificial Neural Networks (Sudheer et al., 2011; Wu et al., 2021) yield sensitivities as a by-product as they are smooth approximations and can be differentiated. However, the choice of appropriate technique from all existing emulation techniques, the conditions required for building accurate response surfaces within a certain tolerance, and the impact of emulation on the convergence of GSA exercises are all worth investigating. With respect to making transparent assumptions in GSA exercises, an example approach is given by Page et al. (2023) who introduce the CREDIBLE Uncertainty Estimation (CURE) toolbox, offering a comprehensive framework for uncertainty estimation. This includes explicit documentation of assumptions and choices through a condition tree implementation, aiding in creating

an audit trail for model decisions. In line with the CURE toolbox, a template for listing the context, assumptions, and choices in a GSA exercise and its convergence is advocated here as a way forward for summarizing and listing choices made, their justification and comparison among alternatives. Such a template could also provide guidance at each step of the GSA application and support the transparency of the process and its learnings more widely for future convergence studies. Additionally, this template would be refined iteratively through future GSA applications in shaping an optimal pathway for achieving and communicating a level of assurance of GSA applications within the available resources.

4. Conclusions

With foremost attention given here to progress and applications in the field of environmental modeling, we have listed and commented on the available convergence assessment methods and covered a wealth of applications to be found in the literature. If the purpose of a GSA exercise is somewhat flexible and the computational budget is limited, then screening/ranking can take fewer samples to converge than estimating specific indices. For assessing convergence, a sequential approach, with incremental increases in sample size and visualization of convergence rates, should generally be easy and worthwhile to adopt as it allows one to assess, or appreciate by visualization, whether and to what extent convergence is being reached.

There is a straightforward gap to be filled in that there is a lack of studies carefully testing the various convergence analysis methods together. Hence, it is not possible to give a definitive answer on which convergence analysis method is superior, though any method will have pros and cons. In concluding our examination of GSA in environmental modeling, we underscore the importance of a methodical approach to the convergence assessment of sensitivity analysis. To achieve this, we propose the following:

1. Identify the Purpose of Sensitivity Analysis
 - Ranking: If the goal is to rank parameters by their influence on the output.
 - Screening: If the goal is to identify non-influential parameters to reduce model complexity.
 - Indices: If the goal is to quantify the sensitivity indices for in-depth analysis.
2. Consider Model Complexity and Computational Resources
 - For Ranking:
 - i. Limited resources: Use the Position Factor (PF) or Top-down Coefficient of Concordance (TDCC) with Savage scores to assess the stability of parameter rankings efficiently.
 - ii. Adequate resources: Apply the Sarrazin $stat_{ranking}$ criterion to assess ranking stability through weighted rank correlation, enhancing the reliability of parameter influence hierarchy.
 - iii. Abundant resources: Assess agreement between multiple SA methods to ensure robustness in rankings.
 - For Screening:
 - i. High-dimensional models: Implement the Sheikholeslami grouping method for efficient parameter grouping, reducing computational demands.
 - ii. General cases: Implement the Sarrazin $stat_{screening}$ criterion to manage the maximum difference in sensitivity measures for non-sensitive parameters, ensuring efficient screening convergence.
 - iii. Consider the Dummy parameter approach with caution, ensuring the method is compatible with chosen SA techniques (e.g., avoid when using certain Sobol' estimators).
 - For Indices:
 - i. Seeking confidence in results: Utilize bootstrapping for confidence interval estimation but consider sample size limitations. If bootstrapping is unreliable, Model Variable Augmentation (MVA) can be an alternative for low sample sizes.

- ii. Desiring detailed convergence assessment: Apply Sarrazin *stat_{indices}* to monitor the maximum difference in sensitivity index bounds, ensuring accurate quantification of parameter impacts.
3. Address Specific Method Limitations
 - Avoid arbitrary thresholds: Recognize the high risk of misclassification due to lack of empirical support and the potential for Type I or II errors.
 - Dummy parameter caution: Acknowledge its case-specific nature and incompatibility with certain GSA methods, necessitating careful consideration of its applicability.
 4. Advanced Considerations for Refinement and Verification
 - “Variability” and “Reliability”: Employ “Variability” and “Reliability” metrics to further verify the stability and reliability of the sensitivity analysis results, especially when computational resources allow for extensive resampling or when using bootstrapping to assess ranking robustness.
 - Final Verification: For all purposes, especially when results are uncertain or further validation is needed, consider a detailed comparison by exploring agreement between multiple SA methods or convergence assessment methods to confirm the reliability and robustness of the sensitivity analysis results.
 5. Visualization and Communication
 - Utilize advanced visualization techniques for clear interpretation and communication of SA results, supporting empirical validation and convergence assessment.

In conclusion, this study has focused on the detailed aspects of GSA within environmental modeling, pointing out the essential need for customized approaches to convergence assessment. By thoroughly examining various methods and their applications, we highlight the wide range of important factors needed for effective convergence assessment, from starting with the goals of the sensitivity analysis to balancing model complexity and computational constraints. The steps outlined for detailed convergence assessment, along with straightforward advice for overcoming method limitations and making improvement, hopefully provide a clear guide for researchers to increase the precision and trustworthiness of their GSA applications. Ultimately, this work has aimed to point out the critical importance of careful convergence assessment in making sensitivity analyses sound and reliable, thus acting as a pillar in helping to move any environmental model development and evaluation forward, and in supporting good modeling practice in general.

Acknowledgement

Xifu Sun’s research was funded by Jiangsu Funding Program for Excellent Postdoctoral Talent. John Jakeman’s work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan. The authors would like to thank the two anonymous examiners for reviewing this article as a part of a chapter of Sun’s PhD thesis. They provided invaluable comments and suggestions as did the reviewers for the associated journal paper.

References

- Akomeah, E., Lindenschmidt, K. E., & Chapra, S. C. (2019). Comparison of aquatic ecosystem functioning between eutrophic and hypereutrophic cold-region river-lake systems. *Ecological Modelling*, 393, 25–36. <https://doi.org/10.1016/j.ecolmodel.2018.12.004>
- Anstett-Collin, F., Goffart, J., Mara, T., & Denis-Vidal, L. (2015). Sensitivity analysis of complex models: Coping with dynamic and static inputs. *Reliability Engineering and System Safety*, 134, 268–275. <https://doi.org/10.1016/j.res.2014.08.010>
- Archer, G. E. B., Saltelli, A., & Sobol, I. M. (1997). Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2), 99–120. <https://doi.org/10.1080/00949659708811825>
- Asensio-Sevilla, M. I., Santos-Martín, M. T., Álvarez-León, D., & Ferragut-Canals, L. (2020). Global sensitivity analysis of fuel-type-dependent input variables of a simplified physical fire spread model. *Mathematics and Computers in Simulation*, 172, 33–44. <https://doi.org/10.1016/j.matcom.2020.01.001>
- Awad, M., Senga Kiese, T., Assaghir, Z., & Ventura, A. (2019). Convergence of sensitivity analysis methods for evaluating combined influences of model inputs. *Reliability Engineering and System Safety*, 189, 109–122. <https://doi.org/10.1016/j.res.2019.03.050>
- Bajracharya, A., Awoye, H., Stadnyk, T., & Asadzadeh, M. (2020). Time variant sensitivity analysis of hydrological model parameters in a cold region using flow signatures. *Water*, 12(4), 961. <https://doi.org/10.3390/W12040961>
- Baroni, G., Scheiffele, L. M., Schrön, M., Ingwersen, J., & Oswald, S. E. (2018). Uncertainty, sensitivity and improvements in soil moisture estimation with cosmic-ray neutron sensing. *Journal of Hydrology*, 564, 873–887. <https://doi.org/10.1016/j.jhydrol.2018.07.053>
- Baroni, G., & Francke, T. (2020). An effective strategy for combining variance- and distribution-based global sensitivity analysis. *Environmental Modelling and Software*, 134, 104851. <https://doi.org/10.1016/j.envsoft.2020.104851>
- Benedetti, L., Claeys, F., Nopens, I., & Vanrolleghem, P. A. (2011). Assessing the convergence of LHS Monte Carlo simulations of wastewater treatment models. *Water Science and Technology*, 63(10), 2219–2224. <https://doi.org/10.2166/wst.2011.453>
- Berezowski, T., Nossent, J., Chormański, J., & Batelaan, O. (2015). Spatial sensitivity analysis of snow cover data in a distributed rainfall-runoff model. *Hydrology and Earth System Sciences*, 19(4), 1887–1904. <https://doi.org/10.5194/hess-19-1887-2015>
- Borroni, C. G. (2013). A new rank correlation measure. *Statistical Papers*, 54(2), 255–270. <https://doi.org/10.1007/s00362-011-0423-0>
- Branger, F., Giraudet, L. G., Guivarch, C., & Quirion, P. (2015). Global sensitivity analysis of an energy-economy model of the residential building sector. *Environmental Modelling and Software*, 70, 45–54. <https://doi.org/10.1016/j.envsoft.2015.03.021>
- Brembilla, E., Mardaljevic, J., & Hopfe, C. J. (2015). Sensitivity analysis studying the impact of reflectance values assigned in climate-based daylight modelling. *14th International Conference of IBPSA - Building Simulation 2015, BS 2015, Conference Proceedings, September*, 1197–1204.
- Budamala, V., & Baburao Mahindrakar, A. (2020). Integration of Adaptive Emulators and Sensitivity Analysis for Enhancement of Complex Hydrological Models. *Environmental Processes*, 7(4), 1235–1253. <https://doi.org/10.1007/s40710-020-00468-x>
- Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22(10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>
- Castaigns, W., Borgonovo, E., Morris, M. D., & Tarantola, S. (2012). Sampling strategies in density-based sensitivity analysis. *Environmental Modelling and Software*, 38, 13–26. <https://doi.org/10.1016/j.envsoft.2012.04.017>
- Chaney, N. W., Herman, J. D., Ek, M. B., & Wood, E. F. (2016). Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *Journal of Geophysical Research*, 121(22), 13,218–13,235. <https://doi.org/10.1002/2016JD024821>
- Chisari, C., Rizzano, G., Amadio, C., & Galdi, V. (2018). Sensitivity analysis and calibration of phenomenological models for seismic analyses. *Soil Dynamics and Earthquake Engineering*, 109, 10–22. <https://doi.org/10.1016/j.soildyn.2018.02.024>
- Ciric, C., Ciffroy, P., & Charles, S. (2012). Use of sensitivity analysis to identify influential and non-influential parameters within an aquatic ecosystem model. *Ecological Modelling*, 246(C), 119–130. <https://doi.org/10.1016/j.ecolmodel.2012.06.024>
- Cloke, H. L., Pappenberger, F., & Renaud, J. P. (2008). Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes. *Hydrological Processes*, 22(11), 1660–1674. <https://doi.org/10.1002/hyp.6734>
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010a). Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling*, 221(16), 1897–1906. <https://doi.org/10.1016/j.ecolmodel.2010.04.021>

- Confalonieri, R., Bregaglio, S., & Acutis, M. (2012). Quantifying plasticity in simulation models. *Ecological Modelling*, 225, 159–166. <https://doi.org/10.1016/j.ecolmodel.2011.11.022>
- Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M., & Genovese, G. (2010b). Sensitivity analysis of the rice model WARM in Europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling & Software*, 25(4), 479–488. <https://doi.org/10.1016/j.envsoft.2009.10.005>
- Coolen-Maturi, T. (2014). A new weighted rank coefficient of concordance. *Journal of Applied Statistics*, 41(8), 1721–1745. <https://doi.org/10.1080/02664763.2014.889664>
- Cosenza, A., Mannina, G., Vanrolleghem, P. A., & Neumann, M. B. (2013). Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods. *Environmental Modelling and Software*, 49, 40–52. <https://doi.org/10.1016/j.envsoft.2013.07.009>
- Cosenza, A., Mannina, G., Vanrolleghem, P. A., & Neumann, M. B. (2014). Variance-based sensitivity analysis for wastewater treatment plant modelling. *Science of the Total Environment*, 470–471, 1068–1077. <https://doi.org/10.1016/j.scitotenv.2013.10.069>
- Crosetto, M., Tarantola, S., & Saltelli, A. (2000). Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment*, 81(1), 71–79. [https://doi.org/10.1016/S0167-8809\(00\)00169-9](https://doi.org/10.1016/S0167-8809(00)00169-9)
- Cruz May, E., Bassam, A., Ricalde, L. J., Escalante Soberanis, M. A., Oubram, O., May Tzuc, O., Alanis, A. Y., & Livas-García, A. (2021). Global sensitivity analysis for a real-time electricity market forecast by a machine learning approach: A case study of Mexico. *International Journal of Electrical Power and Energy Systems*, 135, 107505. <https://doi.org/10.1016/j.ijepes.2021.107505>
- Dai, H., & Ye, M. (2015). Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation. *Journal of Hydrology*, 528, 286–300. <https://doi.org/10.1016/j.jhydrol.2015.06.034>
- Devak, M., & Dhanya, C. T. (2017). Sensitivity analysis of hydrological models: Review and way forward. *Journal of Water and Climate Change*, 8(4), 557–575. <https://doi.org/10.2166/wcc.2017.149>
- Do, N. C., & Razavi, S. (2020). Correlation Effects? A Major but Often Neglected Component in Sensitivity and Uncertainty Analysis. *Water Resources Research*, 56(3). <https://doi.org/10.1029/2019WR025436>
- Douglas-Smith, D., Iwanaga, T., Croke, B. F. W., & Jakeman, A. J. (2020). Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. *Environmental Modelling and Software*, 124, 104588. <https://doi.org/10.1016/j.envsoft.2019.104588>
- Estrada, V., & Diaz, M. S. (2010). Global sensitivity analysis in the development of first principle-based eutrophication models. *Environmental Modelling and Software*, 25(12), 1539–1551. <https://doi.org/10.1016/j.envsoft.2010.06.009>
- Faggianelli, G. A., Mora, L., & Merheb, R. (2017). Uncertainty quantification for Energy Savings Performance Contracting: Application to an office building. *Energy and Buildings*, 152, 61–72. <https://doi.org/10.1016/j.enbuild.2017.07.022>
- Garcia, D., Arostegui, I., & Prellezo, R. (2019). Robust combination of the Morris and Sobol methods in complex multidimensional models. *Environmental Modelling and Software*, 122, 104517. <https://doi.org/10.1016/j.envsoft.2019.104517>
- Ge, Q., & Menendez, M. (2014). An Efficient Sensitivity Analysis Approach for Computationally Expensive Microscopic Traffic Simulation Models. *International Journal of Transportation*, 2(2), 49–64. <https://doi.org/10.14257/ijt.2014.2.2.04>
- Ghasemzade, M., Baroni, G., Abbaspour, K., & Schirmer, M. (2017). Combined analysis of time-varying sensitivity and identifiability indices to diagnose the response of a complex environmental model. *Environmental Modelling and Software*, 88, 22–34. <https://doi.org/10.1016/j.envsoft.2016.10.011>
- Gilardelli, C., Confalonieri, R., Cappelli, G. A., & Bellocchi, G. (2018). Sensitivity of WOFOST-based modelling solutions to crop parameters under climate change. *Ecological Modelling*, 368, 1–14. <https://doi.org/10.1016/j.ecolmodel.2017.11.003>
- Gilquin, L., Prieur, C., Arnaud, E., & Monod, H. (2021). Iterative estimation of Sobol' indices based on replicated designs. *Computational and Applied Mathematics*, 40(1), 1–23. <https://doi.org/10.1007/s40314-020-01402-5>
- Gokarakonda, S., van Treeck, C., & Rawal, R. (2019). Influence of building design and control parameters on the potential of mixed-mode buildings in India. *Building and Environment*, 148, 157–172. <https://doi.org/10.1016/j.buildenv.2018.10.043>
- Gómez-Delgado, M., & Tarantola, S. (2006). GLOBAL sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science*, 20(4), 449–466. <https://doi.org/10.1080/13658810600607709>
- Gschwend, D., Soltic, P., Edinger, P., Wokaun, A., & Vogel, F. (2017). Performance evaluation of gasoline alternatives using a thermodynamic spark-ignition engine model. *Sustainable Energy and Fuels*, 1(9), 1991–2005. <https://doi.org/10.1039/C7SE00276A>
- Guillaume, J. H. A., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M. C., Jakeman, A. J., Keesman, K. J., Razavi, S., & Stigter, J. D. (2019). Introductory overview of identifiability analysis: A guide to

- evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling and Software*, 119, 418–432. <https://doi.org/10.1016/j.envsoft.2019.07.007>
- Gupta, H. V., & Razavi, S. (2018). Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models. *Water Resources Research*, 54(11), 8692–8717. <https://doi.org/10.1029/2018WR022668>
- Hart, J., & Gremaud, P. (2019). Robustness of the Sobol' indices to distributional uncertainty. *International Journal for Uncertainty Quantification*, 9(5), 453–469. <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019030553>
- Hartmann, A., Šimůnek, J., Aidoo, M. K., Seidel, S. J., & Lazarovitch, N. (2018). Implementation and Application of a Root Growth Module in HYDRUS. *Vadose Zone Journal*, 17(1), 170040. <https://doi.org/10.2136/vzj2017.02.0040>
- Helton, J. C., Davis, F. J., & Johnson, J. D. (2005). A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliability Engineering and System Safety*, 89(3), 305–330. <https://doi.org/10.1016/j.res.2004.09.006>
- Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrology and Earth System Sciences*, 17(7), 2893–2903. <https://doi.org/10.5194/hess-17-2893-2013>
- Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1), 1–17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6)
- Hornberger, G., & Spear, R. (1981). Approach to the preliminary analysis of environmental systems. *J. Environ. Manage.; (United States)*, 12:1, 7–18. <https://www.osti.gov/biblio/6396608>
- Hsieh, N. H., Bois, F. Y., Tsakalozou, E., Ni, Z., Yoon, M., Sun, W., Klein, M., Reifeld, B., & Chiu, W. A. (2021). A Bayesian population physiologically based pharmacokinetic absorption modeling approach to support generic drug development: application to bupropion hydrochloride oral dosage forms. *Journal of Pharmacokinetics and Pharmacodynamics*, 5. <https://doi.org/10.1007/s10928-021-09778-5>
- Hsieh, N. H., Reifeld, B., Bois, F. Y., & Chiu, W. A. (2018). Applying a global sensitivity analysis workflow to improve the computational efficiencies in physiologically-based pharmacokinetic modeling. *Frontiers in Pharmacology*, 9, 1–17. <https://doi.org/10.3389/fphar.2018.00588>
- Hsieh, N. H., Reifeld, B., & Chiu, W. A. (2020). pksensi: An R package to apply global sensitivity analysis in physiologically based kinetic modeling. *SoftwareX*, 12, 100609. <https://doi.org/10.1016/j.softx.2020.100609>
- Hu, Y., Garcia-Cabrejo, O., Cai, X., Valocchi, A. J., & DuPont, B. (2015). Global sensitivity analysis for large-scale socio-hydrological models using Hadoop. *Environmental Modelling & Software*, 73, 231–243. <https://doi.org/10.1016/j.envsoft.2015.08.015>
- Huo, X., Gupta, H., Niu, G. Y., Gong, W., & Duan, Q. (2019). Parameter Sensitivity Analysis for Computationally Intensive Spatially Distributed Dynamical Environmental Systems Models. *Journal of Advances in Modeling Earth Systems*, 11(9), 2896–2909. <https://doi.org/10.1029/2018MS001573>
- Iman, R. L., & Conover, W. J. (1987). A Measure of Top-Down Correlation. *Technometrics*, 29(3), 351. <https://doi.org/10.2307/1269344>
- Isaksson, A., Wallman, M., Göransson, H., & Gustafsson, M. G. (2008). Cross-validation and bootstrapping are unreliable in small sample classification. *Pattern Recognition Letters*, 29(14), 1960–1965. <https://doi.org/10.1016/j.patrec.2008.06.018>
- Jabloun, M., Li, X., Zhang, X., Tao, F., Hu, C., & Olesen, J. E. (2018). Sensitivity of simulated crop yield and nitrate leaching of the wheat-maize cropping system in the North China Plain to model parameters. *Agricultural and Forest Meteorology*, 263, 25–40. <https://doi.org/10.1016/j.agrformet.2018.08.002>
- 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. <https://doi.org/10.1016/j.envsoft.2006.01.004>
- Jakeman, J. (2023). PyApprox: A software package for sensitivity analysis, Bayesian inference, optimal experimental design, and multi-fidelity uncertainty quantification and surrogate modeling. *Environmental Modelling and Software*, 170, 105825. <https://doi.org/10.1016/j.envsoft.2023.105825>
- Janssen, H. (2013). Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. *Reliability Engineering and System Safety*, 109, 123–132. <https://doi.org/10.1016/j.res.2012.08.003>
- Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resources Research*, 46(10), 2009WR008896. <https://doi.org/10.1029/2009WR008896>
- KC, U., Aryal, J., Garg, S., & Hilton, J. (2021). Global sensitivity analysis for uncertainty quantification in fire spread models. *Environmental Modelling and Software*, 143, 105110. <https://doi.org/10.1016/j.envsoft.2021.105110>
- KC, U., Garg, S., Hilton, J., & Aryal, J. (2020). A cloud-based framework for sensitivity analysis of natural hazard models. *Environmental Modelling and Software*, 134, 104800. <https://doi.org/10.1016/j.envsoft.2020.104800>
- Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, 26(6), 822–827. <https://doi.org/10.1016/j.envsoft.2010.12.006>

- Khan, S., & Kaklis, P. (2021). From regional sensitivity to intra-sensitivity for parametric analysis of free-form shapes: Application to ship design. *Advanced Engineering Informatics*, 49, 101314. <https://doi.org/10.1016/j.aei.2021.101314>
- Khorashadi Zadeh, F., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., & Bauwens, W. (2017). Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environmental Modelling and Software*, 91, 210–222. <https://doi.org/10.1016/j.envsoft.2017.02.001>
- Koo, H., Iwanaga, T., Croke, B. F. W., Jakeman, A. J., Yang, J., Wang, H. H., Sun, X., Lü, G., Li, X., Yue, T., Yuan, W., Liu, X., & Chen, M. (2020). Position paper: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources. *Environmental Modelling and Software*, 134, 104857. <https://doi.org/10.1016/j.envsoft.2020.104857>
- Krishnan, P., & Aggarwal, P. (2018). Global sensitivity and uncertainty analyses of a web based crop simulation model (web InfoCrop wheat) for soil parameters. *Plant and Soil*, 423(1–2), 443–463. <https://doi.org/10.1007/s11104-017-3498-0>
- Krishnan, P., Maity, P. P., & Kundu, M. (2021). Sensitivity analysis of cultivar parameters to simulate wheat crop growth and yield under moisture and temperature stress conditions. *Heliyon*, 7(7), e07602. <https://doi.org/10.1016/j.heliyon.2021.e07602>
- Kroll, S., Dirckx, G., Donckels, B. M. R., Van Dorpe, M., Weemaes, M., & Willems, P. (2016). Modelling real-time control of WWTP influent flow under data scarcity. *Water Science and Technology*, 73(7), 1637–1643. <https://doi.org/10.2166/wst.2015.641>
- Lebedeva, G., Sorokin, A., Faratian, D., Mullen, P., Goltsov, A., Langdon, S. P., Harrison, D. J., & Goryanin, I. (2012). Model-based global sensitivity analysis as applied to identification of anti-cancer drug targets and biomarkers of drug resistance in the ErbB2/3 network. *European Journal of Pharmaceutical Sciences*, 46(4), 244–258. <https://doi.org/10.1016/j.ejps.2011.10.026>
- Leolini, L., Bregaglio, S., Moriondo, M., Ramos, M. C., Bindi, M., & Ginaldi, F. (2018). A model library to simulate grapevine growth and development: software implementation, sensitivity analysis and field level application. *European Journal of Agronomy*, 99, 92–105. <https://doi.org/10.1016/j.eja.2018.06.006>
- Li, P., & Ren, L. (2019). Evaluating the effects of limited irrigation on crop water productivity and reducing deep groundwater exploitation in the North China Plain using an agro-hydrological model: I. Parameter sensitivity analysis, calibration and model validation. *Journal of Hydrology*, 574(2), 497–516. <https://doi.org/10.1016/j.jhydrol.2019.04.053>
- Likhachev, D. V. (2019). Parametric sensitivity analysis as an essential ingredient of spectroscopic ellipsometry data modeling: An application of the Morris screening method. *Journal of Applied Physics*, 126(18). <https://doi.org/10.1063/1.5126074>
- Liu, J., Liu, Z., Zhu, A. X., Shen, F., Lei, Q., & Duan, Z. (2019a). Global sensitivity analysis of the APSIM-Oryza rice growth model under different environmental conditions. *Science of the Total Environment*, 651, 953–968. <https://doi.org/10.1016/j.scitotenv.2018.09.254>
- Liu, K., Xin, X., Ma, J., Zhang, J., & Yu, Q. (2019b). Sensitivity analysis of ship traffic in restricted two-way waterways considering the impact of LNG carriers. *Ocean Engineering*, 192, 106556. <https://doi.org/10.1016/j.oceaneng.2019.106556>
- Locatelli, T., Tarantola, S., Gardiner, B., & Patenaude, G. (2017). Variance-based sensitivity analysis of a wind risk model - Model behaviour and lessons for forest modelling. *Environmental Modelling and Software*, 87, 84–109. <https://doi.org/10.1016/j.envsoft.2016.10.010>
- Mai, J., & Tolson, B. A. (2019). Model Variable Augmentation (MVA) for Diagnostic Assessment of Sensitivity Analysis Results. *Water Resources Research*, 55(4), 2631–2651. <https://doi.org/10.1029/2018WR023382>
- Mailier, J., Delmotte, A., Cloutier, M., Jolicoeur, M., & VandeWouwer, A. (2011). Sensitivity analysis for parameter identification of a plant growth model. In *IFAC Proceedings Volumes (IFAC-PapersOnline)* (Vol. 44, Issue 1 PART 1). IFAC. <https://doi.org/10.3182/20110828-6-IT-1002.00315>
- Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1), 178–196. <https://doi.org/10.1016/j.jtbi.2008.04.011>
- Matthews, C. J., Newton, D. B., Braddock, R. D., & Yu, B. (2007). Analysing the sensitivity behaviour of two hydrology models. *Environmental Modeling and Assessment*, 12(1), 27–41. <https://doi.org/10.1007/s10666-006-9049-3>
- Medina, Y., & Muñoz, E. (2020a). A simple time-varying sensitivity analysis (TVSA) for assessment of temporal variability of hydrological processes. *Water*, 12(9), 2463. <https://doi.org/10.3390/w12092463>
- Medina, Y., & Muñoz, E. (2020b). Analysis of the relative importance of model parameters in watersheds with different hydrological regimes. *Water*, 12(9), 2376. <https://doi.org/10.3390/w12092376>
- Moreau, P., Viaud, V., Parnaudeau, V., Salmon-Monviola, J., & Durand, P. (2013). An approach for global sensitivity analysis of a complex environmental model to spatial inputs and parameters: A case study of an agro-hydrological model. *Environmental Modelling and Software*, 47, 74–87. <https://doi.org/10.1016/j.envsoft.2013.04.006>
- Neumann, M. B. (2012). Comparison of sensitivity analysis methods for pollutant degradation modelling: A case

- study from drinking water treatment. *Science of the Total Environment*, 433, 530–537. <https://doi.org/10.1016/j.scitotenv.2012.06.026>
- Nguyen, A. T., & Reiter, S. (2015). A performance comparison of sensitivity analysis methods for building energy models. *Building Simulation*, 8(6), 651–664. <https://doi.org/10.1007/s12273-015-0245-4>
- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling and Software*, 26(12), 1515–1525. <https://doi.org/10.1016/j.envsoft.2011.08.010>
- Page, T., Smith, P., Beven, K., Pianosi, F., Sarrazin, F., Almeida, S., Holcombe, L., Freer, J., Chappell, N., & Wagener, T. (2023). Technical note: The CREDIBLE Uncertainty Estimation (CURE) toolbox: facilitating the communication of epistemic uncertainty. *Hydrology and Earth System Sciences*, 27(13), 2523–2534. <https://doi.org/10.5194/hess-27-2523-2023>
- Paleari, L., & Confalonieri, R. (2016). Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions. *Ecological Modelling*. <https://doi.org/10.1016/j.ecolmodel.2016.09.008>
- Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? *Ecological Modelling*, 455, 109648. <https://doi.org/10.1016/j.ecolmodel.2021.109648>
- Pappenberger, F., Beven, K. J., Ratto, M., & Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, 31(1), 1–14. <https://doi.org/10.1016/j.advwatres.2007.04.009>
- Peeters, L. J. M., Pagendam, D. E., Crosbie, R. S., Rachakonda, P. K., Dawes, W. R., Gao, L., Marvanek, S. P., Zhang, Y. Q., & McVicar, T. R. (2018). Determining the initial spatial extent of an environmental impact assessment with a probabilistic screening methodology. *Environmental Modelling and Software*, 109, 353–367. <https://doi.org/10.1016/j.envsoft.2018.08.020>
- Peng, X., Adamowski, J., Inam, A., Alizadeh, M. R., & Albano, R. (2020). Development of a behaviour-pattern based global sensitivity analysis procedure for coupled socioeconomic and environmental models. *Journal of Hydrology*, 585, 124745. <https://doi.org/10.1016/j.jhydrol.2020.124745>
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling and Software*, 79, 214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software*, 70, 80–85. <https://doi.org/10.1016/j.envsoft.2015.04.009>
- Pianosi, F., Sarrazin, F., & Wagener, T. (2020). How successfully is open-source research software adopted? Results and implications of surveying the users of a sensitivity analysis toolbox. *Environmental Modelling and Software*, 124. <https://doi.org/10.1016/j.envsoft.2019.104579>
- Qian, G., & Mahdi, A. (2020). Sensitivity analysis methods in the biomedical sciences. *Mathematical Biosciences*, 323, 108306. <https://doi.org/10.1016/j.mbs.2020.108306>
- Ravalico, J. K., Dandy, G. C., & Maier, H. R. (2010). Management Option Rank Equivalence (MORE) - A new method of sensitivity analysis for decision-making. *Environmental Modelling and Software*, 25(2), 171–181. <https://doi.org/10.1016/j.envsoft.2009.06.012>
- Ravasi, R. A., Paleari, L., Vesely, F. M., Movedi, E., Thoenke, W., & Confalonieri, R. (2020). Ideotype definition to adapt legumes to climate change: A case study for field pea in Northern Italy. *Agricultural and Forest Meteorology*, 291. <https://doi.org/10.1016/j.agrformet.2020.108081>
- Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, 52(1), 423–439. <https://doi.org/10.1002/2015WR017558>
- Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, 52(1), 440–455. <https://doi.org/10.1002/2015WR017559>
- Razavi, S., & Gupta, H. V. (2019). A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. *Environmental Modelling and Software*, 114, 1–11. <https://doi.org/10.1016/j.envsoft.2018.12.002>
- Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., ... Maier, H. R. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, 137. <https://doi.org/10.1016/j.envsoft.2020.104954>
- Razavi, S., Shekholeslami, R., Gupta, H. V., & Haghnegahdar, A. (2019). VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environmental Modelling and Software*, 112, 95–107. <https://doi.org/10.1016/j.envsoft.2018.10.005>
- Reinhart, B. D., Frankenberger, J. R., Hay, C. H., Bowling, L. C., & Hancock, B. G. (2020). Development and Sensitivity Analysis of an Online Tool for Evaluating Drainage Water Recycling Decisions. *Transactions of the ASABE*, 63(6), 1991–2002. <https://doi.org/10.13031/trans.13900>
- Reusser, D. E., Buytaert, W., & Zehe, E. (2011). Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. *Water Resources Research*, 47(7). <https://doi.org/10.1029/2010WR009947>

- Robles, A., Ruano, M. V., Ribes, J., Seco, A., & Ferrer, J. (2014a). Global sensitivity analysis of a filtration model for submerged anaerobic membrane bioreactors (AnMBR). *Bioresource Technology*, *158*, 365–373. <https://doi.org/10.1016/j.biortech.2014.02.087>
- Robles, A., Ruano, M. V., Ribes, J., Seco, A., & Ferrer, J. (2014b). Model-based automatic tuning of a filtration control system for submerged anaerobic membrane bioreactors (AnMBR). *Journal of Membrane Science*, *465*, 14–26. <https://doi.org/10.1016/j.memsci.2014.04.012>
- Ruano, M. V., Ribes, J., Seco, A., & Ferrer, J. (2012). An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors. *Environmental Modelling and Software*, *37*, 103–109. <https://doi.org/10.1016/j.envsoft.2012.03.008>
- Şalap-Ayça, S., Jankowski, P., Clarke, K. C., & Nara, A. (2021). Is less more? Experimenting with visual stacking of coincident maps for spatial global sensitivity analysis in urban land-use change modeling. *Environmental Modelling and Software*, *145*. <https://doi.org/10.1016/j.envsoft.2021.105181>
- Saltelli, A., Chan, K., & Scott, E. (2000). *Sensitivity Analysis: Wiley Series in Probability and Statistics*. Wiley.
- Saltelli, Andrea, & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, *25*(12), 1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
- Saltelli, A., & Bolado, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics and Data Analysis*, *26*(4), 445–460. [https://doi.org/10.1016/S0167-9473\(97\)00043-1](https://doi.org/10.1016/S0167-9473(97)00043-1)
- Saltelli, A., Jakeman, A., Razavi, S., & Wu, Q. (2021). Sensitivity analysis: A discipline coming of age. *Environmental Modelling & Software*, *146*, 105226. <https://doi.org/10.1016/j.envsoft.2021.105226>
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2007). Global Sensitivity Analysis. The Primer. In *International Statistical Review*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470725184>
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*.
- Salviano, L. O., Gasparin, E. E., Mattos, V. C. N., Barbizan, B., Saltara, F., de Mello, P. E. B., Dezan, D. J., & Yanagihara, J. I. (2021). Sensitivity analysis and optimization of a CO₂ centrifugal compressor impeller with a vaneless diffuser. *Structural and Multidisciplinary Optimization*, *64*(3), 1607–1627. <https://doi.org/10.1007/s00158-021-02914-2>
- Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., & Wagener, T. (2018). V2Karst V1.1: A parsimonious large-scale integrated vegetation-recharge model to simulate the impact of climate and land cover change in karst regions. In *Geoscientific Model Development* (Vol. 11, Issue 12). <https://doi.org/10.5194/gmd-11-4933-2018>
- Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, *79*, 135–152. <https://doi.org/10.1016/j.envsoft.2016.02.005>
- Savage, I. R. (1956). Contributions to the Theory of Rank Order Statistics-the Two-Sample Case. *The Annals of Mathematical Statistics*, *27*(3), 590–615. <https://www.jstor.org/stable/2237370>
- Seo, S. B., Duchnowski, E. M., O'Neal, M., Motta, A. T., Passelaigue, F., Kang, S., Pastore, G., Manera, A., Petrov, V., Huang, P. H., & Brown, N. R. (2021). Sensitivity analysis of BISON model for characterization of impact of experimental parameters on hydrogen migration and redistribution in zirconium-based alloys. *Journal of Nuclear Materials*, *550*, 152941. <https://doi.org/10.1016/j.jnucmat.2021.152941>
- Sheikholeslami, R., Gharari, S., Papalexou, S. M., & Clark, M. P. (2021). VISCOUS: A Variance-Based Sensitivity Analysis Using Copulas for Efficient Identification of Dominant Hydrological Processes. *Water Resources Research*, *57*(7), 1–24. <https://doi.org/10.1029/2020WR028435>
- Sheikholeslami, R., & Razavi, S. (2017). Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models. *Environmental Modelling and Software*, *93*, 109–126. <https://doi.org/10.1016/j.envsoft.2017.03.010>
- Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., & Haghnegahdar, A. (2019a). Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. *Environmental Modelling and Software*, *111*, 282–299. <https://doi.org/10.1016/j.envsoft.2018.09.002>
- Sheikholeslami, R., Razavi, S., & Haghnegahdar, A. (2019b). What should we do when a model crashes? Recommendations for global sensitivity analysis of Earth and environmental systems models. *Geoscientific Model Development*, *12*(10), 4275–4296. <https://doi.org/10.5194/gmd-12-4275-2019>
- Sheikholeslami, R., Yassin, F., Lindenschmidt, K. E., & Razavi, S. (2017). Improved understanding of river ice processes using global sensitivity analysis approaches. *Journal of Hydrologic Engineering*, *22*(11), 04017048. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001574](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001574)
- Shin, M. J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152.

- <https://doi.org/10.1016/j.jhydrol.2013.08.047>
- Shin, M. J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, *520*, 1–16. <https://doi.org/10.1016/j.jhydrol.2014.11.040>
- Sieber, A., & Uhlenbrook, S. (2005). Sensitivity analyses of a distributed catchment model to verify the model structure. *Journal of Hydrology*, *310*(1–4), 216–235. <https://doi.org/10.1016/j.jhydrol.2005.01.004>
- Silvestro, P. C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., & Casa, R. (2017). Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. *PLoS ONE*, *12*(11), 1–30. <https://doi.org/10.1371/journal.pone.0187485>
- Sobol, M. (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modeling and Computational Experiment*, *1*(4), 407–414. [https://doi.org/1061-7590/93/04407-008\\$9.00](https://doi.org/1061-7590/93/04407-008$9.00)
- Song, X. M., Kong, F. Z., Zhan, C. S., Han, J. W., & Zhang, X. H. (2013a). Parameter identification and global sensitivity analysis of Xin'anjiang model using meta-modeling approach. *Water Science and Engineering*, *6*(1), 1–17. <https://doi.org/10.3882/j.issn.1674-2370.2013.01.001>
- Song, X., Bryan, B. A., Almeida, A. C., Paul, K. I., Zhao, G., & Ren, Y. (2013b). Time-dependent sensitivity of a process-based ecological model. *Ecological Modelling*, *265*, 114–123. <https://doi.org/10.1016/j.ecolmodel.2013.06.013>
- Song, X., Bryan, B. A., Paul, K. I., & Zhao, G. (2012). Variance-based sensitivity analysis of a forest growth model. *Ecological Modelling*, *247*, 135–143. <https://doi.org/10.1016/j.ecolmodel.2012.08.005>
- Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, *523*(225), 739–757. <https://doi.org/10.1016/j.jhydrol.2015.02.013>
- Specka, X., Nendel, C., & Wieland, R. (2019). Temporal sensitivity analysis of the MONICA model: Application of two global approaches to analyze the dynamics of parameter sensitivity. *Agriculture (Switzerland)*, *9*(2), 37. <https://doi.org/10.3390/agriculture9020037>
- Sreedevi, S., Eldho, T. I., Madhusoodhanan, C. G., & Jayasankar, T. (2019). Multiobjective sensitivity analysis and model parameterization approach for coupled streamflow and groundwater table depth simulations using SHETRAN in a wet humid tropical catchment. *Journal of Hydrology*, *579*, 124217. <https://doi.org/10.1016/j.jhydrol.2019.124217>
- Sudheer, K. P., Lakshmi, G., & Chaubey, I. (2011). Application of a pseudo simulator to evaluate the sensitivity of parameters in complex watershed models. *Environmental Modelling and Software*, *26*(2), 135–143. <https://doi.org/10.1016/j.envsoft.2010.07.007>
- Sun, X. (2021). *SACConvergenceAnalysis*. GitHub. <https://github.com/Xifus/SACConvergenceAnalysis>
- Sun, X., Croke, B., Jakeman, A., & Roberts, S. (2022). Benchmarking Active Subspace methods of global sensitivity analysis against variance-based Sobol' and Morris methods with established test functions. *Environmental Modelling & Software*, *149*, 105310. <https://doi.org/10.1016/j.envsoft.2022.105310>
- Sun, X., Croke, B., Roberts, S., & Jakeman, A. (2021). Comparing methods of randomizing Sobol' sequences for improving uncertainty of metrics in variance-based global sensitivity estimation. *Reliability Engineering & System Safety*, *210*, 107499. <https://doi.org/10.1016/j.res.2021.107499>
- Tan, J., Cui, Y., & Luo, Y. (2017). Assessment of uncertainty and sensitivity analyses for ORYZA model under different ranges of parameter variation. *European Journal of Agronomy*, *91*, 54–62. <https://doi.org/10.1016/j.eja.2017.09.001>
- Tang, Y., Reed, P., Wagener, T., & Van Werkhoven, K. (2007). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences*, *11*(2), 793–817. <https://doi.org/10.5194/hess-11-793-2007>
- Tarantola, S., Becker, W., & Zeitz, D. (2012). A comparison of two sampling methods for global sensitivity analysis. *Computer Physics Communications*, *183*(5), 1061–1072. <https://doi.org/10.1016/j.cpc.2011.12.015>
- Tarantola, S., Ferretti, F., Lo Piano, S., Kozlova, M., Lachi, A., Rosati, R., Puy, A., Roy, P., Vannucci, G., Kuc-Czarnecka, M., & Saltelli, A. (2024). An annotated timeline of sensitivity analysis. *Environmental Modelling & Software*, *174*, 105977. <https://doi.org/10.1016/j.envsoft.2024.105977>
- Touhami, H. B., Lardy, R., Barra, V., & Bellocchi, G. (2013). Screening parameters in the Pasture Simulation model using the Morris method. *Ecological Modelling*, *266*(1), 42–57. <https://doi.org/10.1016/j.ecolmodel.2013.07.005>
- Touzani, S., & Busby, D. (2014). Screening Method Using the Derivative-based Global Sensitivity Indices with Application to Reservoir Simulator. *Oil & Gas Science and Technology – Revue d'IFP Energies Nouvelles*, *69*(4), 619–632. <https://doi.org/10.2516/ogst/2013195>
- Uliana, E. M., da Silva, D. D., Moreira, M. C., & Pereira, D. dos R. (2019). Global sensitivity analysis methods applied to hydrologic modeling with the SAC-SMA model. *Engenharia Agricola*, *39*(1), 65–74. <https://doi.org/10.1590/1809-4430-Eng.Agric.v39n1p65-74/2019>
- Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Li, Z., Huang, W., & Casa, R. (2020). A Comparison of Moment-Independent and Variance-Based Global Sensitivity Analysis Approaches for Wheat Yield Estimation with the Aquacrop-OS Model. *Agronomy*, *10*(4), 607. <https://doi.org/10.3390/agronomy10040607>
- Van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across

- a hydroclimatic gradient. *Water Resources Research*, 44(1), 1–16. <https://doi.org/10.1029/2007WR006271>
- Vanrolleghem, P. A., Mannina, G., Cosenza, A., & Neumann, M. B. (2015). Global sensitivity analysis for urban water quality modelling: Terminology, convergence and comparison of different methods. *Journal of Hydrology*, 522, 339–352. <https://doi.org/10.1016/j.jhydrol.2014.12.056>
- Vigna, S. (2015). A weighted correlation index for rankings with ties. *WWW 2015 - Proceedings of the 24th International Conference on World Wide Web*, 1166–1176. <https://doi.org/10.1145/2736277.2741088>
- Wagener, T., & Pianosi, F. (2019). What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. *Earth-Science Reviews*, 194, 1–18. <https://doi.org/10.1016/j.earscirev.2019.04.006>
- Wagener, T., Van Werkhoven, K., Reed, P., & Tang, Y. (2009). Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling. *Water Resources Research*, 45(2), 1–5. <https://doi.org/10.1029/2008WR007347>
- Wang, A., & Solomatine, D. P. (2019). Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. *Water*, 11(5), 1062, 1–26. <https://doi.org/10.3390/w11051062>
- Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environmental Modelling and Software*, 60, 167–179. <https://doi.org/10.1016/j.envsoft.2014.05.026>
- Wang, H., Gong, W., Duan, Q., & Di, Z. (2020). Evaluation of parameter interaction effect of hydrological models using the sparse polynomial chaos (SPC) method. *Environmental Modelling and Software*, 125. <https://doi.org/10.1016/j.envsoft.2019.104612>
- Wang, J., Li, X., Lu, L., & Fang, F. (2013). Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. *Environmental Modelling and Software*, 48, 171–182. <https://doi.org/10.1016/j.envsoft.2013.06.007>
- Wang, Q., Guillaume, J. H. A., Jakeman, J. D., Yang, T., Iwanaga, T., Croke, B., & Jakeman, A. J. (2022). Assessing the predictive impact of factor fixing with an adaptive uncertainty-based approach. *Environmental Modelling and Software*, 148, 105290. <https://doi.org/10.1016/j.envsoft.2021.105290>
- Wang, S., Flipo, N., & Romary, T. (2018). Time-dependent global sensitivity analysis of the C-RIVE biogeochemical model in contrasted hydrological and trophic contexts. *Water Research*, 144, 341–355. <https://doi.org/10.1016/j.watres.2018.07.033>
- Wate, P., Iglesias, M., Coors, V., & Robinson, D. (2020). Framework for emulation and uncertainty quantification of a stochastic building performance simulator. *Applied Energy*, 258, 113759. <https://doi.org/10.1016/j.apenergy.2019.113759>
- Wu, Z., Ma, B., Wang, H., Hu, C., Lv, H., & Zhang, X. (2021). Identification of Sensitive Parameters of Urban Flood Model Based on Artificial Neural Network. *Water Resources Management*, 35(7), 2115–2128. <https://doi.org/10.1007/s11269-021-02825-3>
- Yang, J. (2011). Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling and Software*, 26(4), 444–457. <https://doi.org/10.1016/j.envsoft.2010.10.007>
- Yang, J., Jakeman, A., Fang, G., & Chen, X. (2018). Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. *Environmental Modelling and Software*, 101, 289–300. <https://doi.org/10.1016/j.envsoft.2017.11.037>
- Yang, J., Liu, Y., Yang, W., & Chen, Y. (2012). Multi-Objective Sensitivity Analysis of a Fully Distributed Hydrologic Model WetSpa. *Water Resources Management*, 26(1), 109–128. <https://doi.org/10.1007/s11269-011-9908-9>
- Zhan, C., Song, X., Xia, J., & Tong, C. (2013). An efficient integrated approach for global sensitivity analysis of hydrological model parameters. *Environmental Modelling and Software*, 41, 39–52. <https://doi.org/10.1016/j.envsoft.2012.10.009>
- Zhan, Y., & Zhang, M. (2013). Application of a combined sensitivity analysis approach on a pesticide environmental risk indicator. *Environmental Modelling and Software*, 49, 129–140. <https://doi.org/10.1016/j.envsoft.2013.08.005>
- Zhang, C., Chu, J., & Fu, G. (2013). Sobol’'s sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. *Journal of Hydrology*, 480, 58–68. <https://doi.org/10.1016/j.jhydrol.2012.12.005>
- Zhao, G., Bryan, B. A., & Song, X. (2014). Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters. *Ecological Modelling*, 279, 1–11. <https://doi.org/10.1016/j.ecolmodel.2014.02.003>
- Zhao, Y., Nan, Z., Yu, W., & Zhang, L. (2019). Calibrating a hydrological model by stratifying frozen ground types and seasons in a cold alpine basin. *Water*, 11(5), 985. <https://doi.org/10.3390/w11050985>