Global Sensitivity Analysis for Multiple Interpretive Models with Uncertain Parameters

A. Dell'Oca, M. Riva, A. Guadagnini

Dipartimento di Ingegneria Civile e Ambientale (DICA), Politecnico di Milano, Piazza L. Da Vinci, 32, 20133 Milano, Italy

Abstract

We propose a set of new indices to assist global sensitivity analysis (GSA) in the presence of data allowing for interpretations based on a collection of diverse models whose parameters could be affected by uncertainty. Our GSA metrics enable us to assess the sensitivity of various features (as rendered through statistical moments) of the probability density function of a quantity of interest with respect to imperfect knowledge of (*i*) the interpretive model employed to characterize the system behavior and (*ii*) the ensuing model parameters. We exemplify our methodology for the case of heavy metal sorption onto soil, for which we consider three broadly used (equilibrium isotherm) models. Our analyses consider (*a*) an unconstrained case, i.e., when no data are available to constrain parameter uncertainty and to evaluate the (relative) plausibility of each considered model, and (*b*) a constrained case, i.e., when the analysis is constrained against experimental observations. Our moment-based indices are structured according to two key components: (*a*) a model-choice contribution, associated with the possibility of analyzing the system of interest by taking advantage of multiple model conceptualizations (or mathematical renderings); and (*b*) a parameter-choice contribution, related to the uncertainty in the parameters of a selected model. Our results indicate that a given parameter can be associated with diverse degrees of importance, depending on the considered statistical moment of the target model output. The influence on the latter of parameter and model uncertainty evolves as a function of the available level of information about the modeled system behavior.

Plain Language Summary

The quality and amount of data available in many practical situations justify the interpretation of the system under investigation through a collection of alternative interpretative models. This is reflected by the observation that there is uncertainty about model structure/format. The situation is exacerbated by the observation that parameters associated with each model could also be affected by uncertainty. In this context, quantification of the influence of these multiple sources of uncertainties on environmental quantities of interest is key to increase our understanding and confidence on model(s) functioning and guide further actions (including, e.g., model calibration or collection of new data). We propose an original Global Sensitivity Analysis (GSA) approach that enables us to quantify the sensitivity of a target quantity with respect to each of the parameters stemming from situations where multiple interpretative models have been formulated. The proposed GSA allows (*i*) investigating the sensitivity of model outputs through diverse aspects of uncertainty (i.e., focusing on various statistical moments of the probability density function of the target output) as well as (*ii*) discriminating between contributions to sensitivity due to our lack of knowledge in (*a*) model format and (*b*) parameter values.

Highlights

- A novel Global Sensitivity Analysis in case of multiple alternative interpretive models is proposed.
- The approach allows discriminating between contributions to sensitivity due to our lack of knowledge in (*a*) model format and (*b*) parameter values.
- We analyze the evolution of the proposed sensitivity metrics as observations about the system under investigation become available.

1. Introduction

Sensitivity analysis is key to assist understanding (and eventually improvement) of models aiming at rendering the dynamics of environmental/hydrological systems. Challenges associated with a sensitivity analysis are exacerbated by the increasing complexity of conceptual models, in terms of model formulation and associated parametrization, which is in turn sustained by our increased knowledge of environmental dynamics and by the exponentially increasing computational power available for numerical model simulations (e.g., Paniconi and Putti, 2015; Förster et al., 2014; Herman et al., 2013; Wagener and Montanari, 2011; Koutsoyiannis, 2010). Sensitivity analysis techniques can be classified according to two categories, i.e., local and global methods (e.g., Gupta and Razavi, 2018; Pianosi et al. 2016; Borgonovo and Plischke, 2016; Razavi and Gupta, 2016a,b). According to the former approach, sensitivity (typically resting on the derivative concept) is evaluated in the neighborhood of a given combination of parameters of a model. Global sensitivity analysis (GSA) approaches rest on the evaluation of sensitivity (usually quantified through metrics entailing the variability of the model response) across the entire support within which model system parameters vary. As such, the typically encountered uncertainty about parameters driving the behavior of environmental/hydrological systems can be readily accommodated in this context (Saltelli et al., 2008). Our study is set within a GSA framework.

In this broad framework, it is important to note that the formal definition of a sensitivity metric must be linked to the nature of the question(s) the GSA is intended to address. Understanding of the way uncertainty associated with model parameters affects uncertainty of given modeling goals / results can be useful to address various research questions, such as: Which are the most important model parameters with respect to given model output(s) / response(s)? (e.g., Hill et al., 2016; Ruano et al., 2012; Wagner et al., 2009; Pappenberger et al., 2008; Gupta et al., 2008; Muleta and Nicklow, 2005); What are the relationships between key features of model output(s) uncertainty and model parameter(s)? (e.g., Pianosi and Wagner, 2015; Borgonovo, 2007; Liu et al., 2006); Could we set some parameter(s) (which are deemed as uninfluential) at prescribed value(s) without significantly affecting model results? (e.g., Chu et al., 2015; Punzo et al., 2015; Nossent et al., 2011; van Griensven et al., 2006; Degenring et al., 2004); At which space/time locations can one expect the highest sensitivity of model output(s) to model parameters and which parameter(s) data could be most beneficial for model calibration? (e.g., Hölter et al., 2018; Wang et al., 2018; Younes et al., 2016; Ciriello et al., 2015; Fajraoui et al., 2011; Yue et al., 2008).

Dell'Oca et al. (2017) focus on these research questions by proposing a moment-based GSA approach enabling quantification of the influence of uncertain model parameters on the (statistical) moments of a target model output. In this sense, these authors (*i*) define sensitivity in terms of the (average) variation of main statistical moments of the probability density function (*pdf*) of an output due to model parameter uncertainty and (*ii*) propose summary sensitivity indices (termed *AMA* indices after the Authors' initials) to quantify the concept. Here, we extend the *AMA* indices to embed the effect of uncertainties both in the system model conceptualization and in the model(s) parameters. Our study rests on the observation that physical processes and natural systems within which they take place are complex, making state variables amenable to a multiplicity of interpretations and (conceptual/mathematical) descriptions, including system parameterizations. As such, predictions and uncertainty analyses based on a unique model can yield statistical bias and underestimation of uncertainty, thus justifying the assessment of multiple (competing) model system conceptualizations (e.g., Clark et al., 2008; Wholing and Vrugt, 2008; Ye et al., 2008a; Beven 2006; Poeter and Anderson, 2005; Bredehoeft, 2005; Burnham and Anderson, 2002). When a collection of alternative models is available, one can then use various criteria to (*a*) rank such models and/or (*b*) weigh model-based predictions (e.g., Höge et al., 2019; Neuman, 2012; Ye et al., 2008b, 2004; Neuman et al., 2003; Kass and Raftery, 1995). When prior information is unavailable, one can assign equal prior probability (or a priori model weight) to each model. Otherwise, a posterior model probability (or a posterior model weight) can be linked to each model when observations on the state variables of interest are made available (e.g., Rodríguez-Escales et al., 2018; Höge et al., 2018; Liu et al., 2016; Knuth et al., 2015; Schöniger et al., 2014; Ye et al., 2008b).

In this context, the study of Dai and Ye (2015) highlights the importance of extending the classical variance-based GSA approach to include the possibility that a collection of models can be employed to assess a quantity of interest. In this multi-model framework, these authors introduce a definition of sensitivity that combines the model-averaging approach with the variance-based methodology to quantify an averaged (across the set of models) contribution of a parameter (including its interaction with other parameters) to a model-averaged variance of the output of interest. Dai et al. (2017) focus on the quantification (through a variance-based method) of the sensitivity of model prediction(s) to the uncertainty in the conceptualization of diverse model system processes, a setting corresponding to a lack of knowledge about the conceptualization of differing processes (and eventually their controlling parameters) contributing to a model structure. The GSA approach presented by Dai and Ye (2015) and Dai et al. (2017) is based on the rationale that the definition of sensitivity is that of uncertainty (as rendered through the variance) apportioning, i.e., the highest sensitivity is attributed to the factor providing the highest contribution to the uncertainty of the output of interest.

A key element distinguishing our study from those of Dai and Ye (2015) and Dai et al. (2017) is the definition of the metric employed to quantify sensitivity. Here, we identify sensitivity with the (average) variation of a set of statistical moments of the *pdf* of a desired output (thus considering various features of the uncertainty on the output) due to the uncertainty on (*a*) the conceptualization of the system functioning (as rendered through various alternative model formulations) and (*b*) model parameters. In this sense, our definition of sensitivity is more relevant for factor fixing aspects (i.e., with reference to questions of the kind "How does a parameter affect a given statistical moment of the output *pdf*?" and/or "Is it possible to fix its value arbitrarily?"), rather than for the uncertainty apportioning aspect. A comparison between the metrics at heart of our GSA approach and those proposed by Dai and Ye (2015) and Dai et al. (2017) is provided in Appendix B-C.

Following the idea of uncertainty apportioning, Baroni and Tarantola (2014) present a probabilistic framework for uncertainty and global sensitivity analysis focusing on hydrological studies. These authors distinguish five sources of uncertainty, respectively corresponding to model input (e.g., weather data), time-varying model parameters (e.g., crop height), scalar model parameters (e.g., soil hydraulic properties), available observations (e.g., soil moisture), and model structure. With reference to the latter, these authors identify model structure uncertainty with the use of diverse levels of (space-time) refinement and descriptive detail of an otherwise deterministically given mathematical model, i.e., they do not analyze the impact of considering a set of alternative conceptual/mathematical formulations to render the investigated system. A GSA is then performed upon relying on the Sobol indices. In a similar fashion, Schoups and Hopmans (2006) present a sensitivity analysis methodology aimed at assessing the relative importance of three common sources of uncertainty, i.e., parameters, observation error and the model structure, here encompassing both the level of model refinement/details and the format of the conceptual/mathematical formulation), and measurements error, by focusing on the fractional contribution of each of these sources to the total predictive error between observations and corresponding model outputs. As mentioned above, our multi-model GSA is guided by a diverse rationale.

We exemplify our approach by targeting a geochemical phenomenon associated with the process of sorption of single metal onto a soil and considering three differing models that are typically used in the literature, each associated with uncertain parameters. We structure our analyses across two stages. We start in the absence of data/observations on the target model output against which one would perform model calibration (i.e., an equal prior probability is assigned to each candidate model and uncertainty of each model parameters is rendered through uniform *pdf*s characterized by the same coefficient of variation). In this case, the purpose of a GSA is that of (*i*) improving our understanding of the model functioning, in terms of the relevance of each model parameter on the considered model output, and (*ii*) identifying parameters which might be uninfluential and whose values could be fixed (and excluded from a subsequent calibration procedure) without affecting the model output (e.g., Hutcheson and McAdams, 2010; Liu et al., 2006). We then perform our GSA after model parameters and weights are estimated through model calibration against system state observations. At this stage, the main purpose of the GSA is that of identifying parameters having the largest impact on the variation in the statistical moments of the output, thus guiding additional efforts for their characterization.

The rest of the work is organized as follows. Section 2.1 describes the sensitivity metrics/indices at core of the proposed GSA; Section 2.2 provides the expressions for multi-model statistical moments of a target quantity; and Section 2.3 details the computational approach employed. Section 3.1 and Section 3.2 exemplify the proposed methodology for the unconstrained and constrained cases, respectively. A discussion is given in Section 4, and Section 5 provides our major conclusions.

2. Methodology

2.1 Sensitivity Index

We consider a quantity Δ (which represents a given modeling goal, e.g., concentration of adsorbed metal onto soil or dissolved chemical in an aquifer, hydraulic head in a well, travel time of a solute in a well-field, or other quantities of interest in an environmental scenario) that could be rendered through a suite of N_M alternative (possibly competing) conceptual and mathematical models collected in a model set ***M***. The latter somehow represents our ability to interpret a given process or a set of processes contributing to characterize Δ.

In the presence of observations of Δ, one can rely on various criteria to (*a*) rank available models, and/or (*b*) weigh results rendered by these models (e.g., Höge et al., 2019; Neuman, et al. 2012; Clark et al., 2008; Ye et al., 2008a, 2005, 2004; Poeter and Hill, 2007; Neuman et al., 2003). In this context, one can evaluate prior (before new data/information about the system under analysis become available (e.g., Rodríguez-Escales et al., 2018; Ye et al., 2008b) and posterior (after data/information from the system are available (e.g., Bianchi Janetti et al., 2019, 2012; Ranee et al., 2016; Moghadesi et al., 2015; Ye et al., 2008a) weights associated with each model included in ***M***. Prior probability weights are usually taken to be equally apportioned amongst models or determined on the basis of expert opinion. Posterior weights depend on the model discrimination criterion of choice and on the prior weight associated with each model (e.g., Schöniger et al., 2014). Models included in ***M*** can be ranked according to such posterior weights or can be employed to provide model-averaged statistics of the quantity of choice Δ.

Each model M^j (*j* = 1, …, N_M) in ***M*** is characterized by a set of parameters, which we collect in vector $\boldsymbol{\theta}^j$. We treat each parameter of model *j*, i.e., θ_i^j (*i* = 1, …, N_j) as a random variable to account for our incomplete knowledge of its exact value. It is then possible to evaluate the impact of uncertainties associated with both the models and their parameters on a quantity of interest (Δ) in terms of moment-based sensitivity indices. We do so by extending the *AMA* sensitivity indices introduced by Dell'Oca et al. (2017), that considered a unique system model with uncertain parameters, to the case of possible multiple interpretative models, as rendered by the collection of models included in ***M***.

Within this multi-model context, we quantify the sensitivity of a given statistical moment (SM) to parameter θ_i^j through the following index

$$AMA\mathrm{SM}_{\theta_i^j} = \frac{w\left(M^j\right)}{\Im}\left\{\underbrace{\left|\mathrm{SM}\left[\Delta\right]-\mathrm{SM}_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]\right|}_{model\text{-}choice\ contribution}+\underbrace{E_{\theta_i^j}\left[\left|\mathrm{SM}_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]-\mathrm{SM}_{-\theta_i^j}\left[\Delta \mid M^j\right]\right|\right]}_{parameter\text{-}choice\ contribution}\right\} \tag{1}$$

with

$$\Im = \begin{cases} \left|\mathrm{SM}\left[\Delta\right]\right| & if \ \ \mathrm{SM}\left[\Delta\right] \neq 0 \\ 1 & if \ \ \mathrm{SM}\left[\Delta\right] = 0 \end{cases} \tag{2}$$

Here, $AMA\mathrm{SM}_{\theta_i^j}$ is the AMA index associated with a given statistical moment SM (*e.g.*, SM = E, V, γ, k when considering the expected value, variance, skewness and kurtosis, respectively) of Δ considering the *i*-th parameter of model M^j; $w\left(M^j\right)$ is the (prior or posterior) probability of model M^j; $E_{\theta_i^j}$ is the expectation operator in the space of variability of θ_i^j; $\mathrm{SM}_{-\theta_i^j}\left[\Delta \mid M^j\right]$ and $\mathrm{SM}_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]$ correspond to the value of the statistical moment of Δ considering uncertainty of (*a*) all parameters of model M^j except θ_i^j (i.e., the value of the statistical moment resulting from conditioning to θ_i^j) and (*b*) the whole set of parameters in $\boldsymbol{\theta}^j$, respectively. The quantity $\mathrm{SM}\left[\Delta\right]$ in (1) represents the value of the given statistical moment of Δ evaluated in a multi-model context, i.e., when both the model employed to interpret a given process and its associated parameters are uncertain. In Section 2.2 we provide further details about the evaluation of the statistical moment of Δ in a multi-model context. We note here that the first and the second terms in (1) are scalar quantities that can be consistently summed (see also discussion of Figure 1 below).

The rationale behind definition (1) is that the sensitivity of Δ, as grounded on a set of statistical moments, to the parameter θ_i^j is proportional to the induced variations of these statistical moments (with respect to their multi-model counterparts) due to the conditioning on the parameter θ_i^j. As such, the rationale at the core of (1) is different from that employed in GSA methodologies grounded on an uncertainty apportioning (e.g., Dai and Ye, 2015) or on a derivative based (e.g., Razavi and Gupta, 2019; Rakovec et al. 2014, within the single-model context) rationales.

In the context of the proposed GSA, we emphasize that conditioning to θ_i^j requires starting with selecting/choosing a model M^j. It then follows that two terms contribute to index $AMA\mathrm{SM}_{\theta_i^j}$ (1): (*i*) the first one is related to the choice of model M^j (within model vector $\boldsymbol{M}$) and is termed here as the *model-choice contribution* (note that this contribution is common to all parameters in M^j and is non-zero in the presence of multiple models, even if each model is characterized by deterministically known parameters); (*ii*) the second one is a *parameter-choice contribution* and represents the contribution to $AMA\mathrm{SM}_{\theta_i^j}$ due to uncertainty of parameter θ_i^j embedded in model M^j (note that this contribution does not vanish in the presence of uncertain model parameters even in cases where there is only one available model; see also our discussion to Figure 1 below). It is also important to remark that indices $AMA\mathrm{SM}_{\theta_i^j}$ are particularly suited for factor fixing studies because they allow highlighting those parameters that could potentially induce strong variations of the investigated statistical moments of the target model output.

Note that relying on multiple statistical moments explicitly recognizes that variance is not an exhaustive metric to quantify sensitivity (e.g., Dell'Oca et al., 2017; Pianosi and Wagner, 2015; Borgonovo, 2007; Liu et al., 2006).

Figure 1 depicts a sketch of the overall concept underpinning (1) when one considers two possible model choices, M^1 and M^2, each associated with a set of two parameters, i.e., (θ_1^1, θ_2^1) for M^1 and (θ_1^2, θ_2^2) for M^2. One can clearly visualize the nature of the diverse terms contributing to index $AMA\mathrm{SM}_{\theta_i^j}$ through the depiction of Figure 1. Let us consider, for example, index $AMA\mathrm{SM}_{\theta_i^1}$: (*i*) the distance between $\mathrm{SM}[\Delta]$ and $\mathrm{SM}_{\boldsymbol{\theta}^1}\left[\Delta \mid M^1\right]$ (dark blue double pointed arrow in Figure 1) corresponds to the model-choice contribution; (*ii*) the average (with respect to θ_i^1) distance between $\mathrm{SM}_{\boldsymbol{\theta}^1}\left[\Delta \mid M_1\right]$ and $\mathrm{SM}_{-\theta_i^1}\left[\Delta \mid M^1\right]$ (light blue and blue double pointed arrows for θ_1^1 and θ_2^1, respectively) corresponds to the parameter-choice contribution for parameter θ_i^1 (i = 1, 2); (*iii*) the sum of (*i*) and (*ii*), weighted by $w(M^1)$ and normalized by $\mathrm{SM}[\Delta]$, quantifies the influence of θ_i^1 on Δ, as expressed by index $AMA\mathrm{SM}_{\theta_i^1}$. The same line of reasoning can be employed to describe the contributions of the parameters of M^2 to our sensitivity indices. Note that, in the presence of a unique interpretive model, e.g., M^1, the distance between $\mathrm{SM}[\Delta]$ and $\mathrm{SM}_{\boldsymbol{\theta}^1}\left[\Delta \mid M_1\right]$ vanishes, reflecting the deterministic choice of the model.

We recall here that when one considers only one interpretive model, the first term in (1) vanishes and $AMA\mathrm{SM}_{\theta_i^j}$ coincides with index $AMA\mathrm{SM}_{\theta_i}$ defined by Dell'Oca et al. (2017), i.e.,

$$AMA\mathrm{SM}_{\theta_i^j} \equiv AMA\mathrm{SM}_{\theta_i} = \frac{1}{\mathfrak{I}} E_{\theta_i^j}\left[\left|\mathrm{SM}_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right] - \mathrm{SM}_{-\theta_i^j}\left[\Delta \mid M^j\right]\right|\right] \tag{3}$$

Comparison of (3) and (1) suggests that some similarities in the indices $AMA\mathrm{SM}_{\theta_i}$ and $AMA\mathrm{SM}_{\theta_i^j}$ are expected when the parameter-choice contribution in (1) is dominant.

2.2 Multi-model Statistical Moments

Here, we provide the details for the evaluation of the statistical moments of Δ in a multi-model context, i.e., $\mathrm{SM}[\Delta]$, leading to the evaluation of index (1). Introducing the probability density function of Δ, $p[\Delta]$, as (e.g., Ye et al., 2004)

$$p[\Delta] = \sum_{j=1}^{N_M} p\left[\Delta \mid M^j\right] w\left(M^j\right) \tag{4}$$

where $p\left[\Delta \mid M^j\right]$ is the *pdf* of Δ given model M^j, the expected value (first statistical moment) of Δ is defined as

$$E[\Delta] = \sum_{j=1}^{N_M} w\left(M^j\right) E_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right] \tag{5}$$

while the *n*-th order (central) moment can be evaluated as

$$SM^{(n)}[\Delta] = \sum_{j=1}^{N_M} w\left(M^j\right) \sum_{k=0}^{n} \binom{n}{k} SM_{\boldsymbol{\theta}^j}^{(k)}\left[\Delta \mid M^j\right]\left(E_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right] - E[\Delta]\right)^{n-k} \tag{6}$$

Variance, $V = SM^{(2)}$, skewness, $\gamma = SM^{(3)} / V^{3/2}$ and kurtosis, $k = SM^{(4)} / V^2$, of Δ can then be assessed from (5) as

$$V[\Delta] = \underbrace{\sum_{j=1}^{N_M} w\left(M^j\right) V_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]}_{within\text{-}model} + \underbrace{\sum_{j=1}^{N_M} w\left(M^j\right)\left(E[\Delta] - E_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]\right)^2}_{between\text{-}model} \tag{7}$$

$$\gamma[\Delta]=\underbrace{\sum_{j=1}^{N_M}\left(\frac{V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}{V[\Delta]}\right)^{3/2} w\left(M^j\right)\gamma_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}_{within-model}-\underbrace{\sum_{j=1}^{N_M} w\left(M^j\right)\frac{\left(E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)^3}{\left(V[\Delta]\right)^{3/2}}}_{between\text{-}model}$$

$$-\underbrace{3\sum_{j=1}^{N_M}\frac{V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}{\left(V[\Delta]\right)^{3/2}} w\left(M^j\right)\left(E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)}_{mixed} \tag{8}$$

$$k[\Delta]=\underbrace{\sum_{j=1}^{N_M}\left(\frac{V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}{V[\Delta]}\right)^{2} w\left(M^j\right)k_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}_{within\text{-}model}+\underbrace{\sum_{j=1}^{N_M} w\left(M^j\right)\frac{\left(E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)^4}{\left(V[\Delta]\right)^{2}}}_{between\text{-}model}$$

$$+\underbrace{4\sum_{j=1}^{N_M} w\left(M^j\right)\left(\frac{E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}{V[\Delta]}\right)^2\left\{\frac{3}{2}V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]-\frac{\gamma_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\left(V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)^{3/2}}{E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]}\right\}}_{mixed} \tag{9}$$

The first term on the right hand side of (7) is typically referred to as the within-model variance and coincides with the average (across the collection of models) of the variances associated with each model whereas the second term in (7) is denoted as the between-model variance and is proportional to the *off-set* between $E[\Delta]$ and $E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]$ (e.g., Draper, 1995). Along the same line, one recognizes that equations (8)-(9) show that, in a multi-model context, the shape of the distribution of Δ, as quantified by skewness and kurtosis, is affected by the following three terms: (*i*) a within-model component, which corresponds to the weighted average of the values (of skewness or kurtosis) rendered by each model, the weights being proportional to $w\left(M^j\right)$ and the ratio between $V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]$ and $V[\Delta]$; (*ii*) a between-model component, which is proportional to the *off-set* $\left(E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)$; and (*iii*) a mixed component, that takes into account the *off-set* $\left(E[\Delta]-E_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]\right)$ (i.e., the between-model variability) as well as the variability within each model, as quantified by $V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]$ in (7) or by $V_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]$ and $\gamma_{\boldsymbol{\theta}^j}\left[\Delta\,|\,M^j\right]$ in (8).

It is here important to note that (*a*) our definition of sensitivity in (1) is grounded on the assessment of the (average) variation of a SM due to conditioning on a given parameter in a given model while (*b*) formulations (7)-(9) serve uncertainty apportioning, i.e., (7)-(9) allow identifying the main factors (between model format and model parameters) contributing to the uncertainty of an output of interest.

2.3 Numerical Evaluation

Inspection of equation (1) reveals that one needs to evaluate $\mathrm{SM}_{-\theta_i^j}\left[\Delta\,|\,M^j\right]$, i.e., the statistical moment of Δ conditional to diverse values of θ_i^j, considering model M^j. Here, we do so through a straightforward Monte Carlo sampling scheme. We (*i*) discretize the support of each parameter by way of a given number of equal bins, i.e., N_{bins} (for simplicity we employ the same value of N_{bins} for all parameters) and (*ii*) evaluate the conditional statistical moment, $\mathrm{SM}_{-\theta_i^j}\left[\Delta \mid M^j\right]$, associated with each bin. We then proceed to evaluate the second term in (1) by taking expectation with respect to θ_i^j, i.e., $E_{\theta_i^j}$. The unconditional statistical moment, i.e., $\mathrm{SM}_{\boldsymbol{\theta}^j}\left[\Delta \mid M^j\right]$, is evaluated through the algebraic expressions (4)-(8). Convergence of the results with respect to the N_{bins} is assessed by increased the latter by regular increments of 10 until the relative variation of all investigated quantities is smaller than a fixed threshold. Here, for simplicity and ease of implementation we select the median value within each bin as representative and evaluate each model under investigation considering all of the combinations of the parameter values identified according to this, i.e., we run model *j* for a total of $\left(N_{bins}\right)^{N_j}$ times. As such, while more efficient sampling strategies can be employed, there are no particular constraints in the sampling scheme one can employ in our GSA.

3. Illustrative examples

As a showcase example to illustrate the application of the theoretical framework introduced in Section 2, we focus on the geochemical process of sorption of metals onto soil matrices. We do so by leveraging on the study of Bianchi Janetti et al. (2012) who analyze the ability of diverse isotherm models to interpret experimental observations documenting copper sorption onto Bet Dagan soil type (see Table 1 of Bianchi Janetti et al. (2012) for the description of the key characteristics of the soil type). For each experiment, 1g of soil was mixed with 40 mL of a solution containing copper.Seven diverse initial concentrations were considered. The same initial pH was set for all experiments. The resulting mixture was then shaken for 48 h to attain equilibrium. A 10-mL sample of the solution was then extracted and passed through a 0.22 μm filter. Copper concentration was evaluated by inductively coupled plasma-mass spectrometry. Three replicates were performed for each initial concentration, the resulting average and variance (assumed to represent measurement error variance) being employed in the calibration procedure. The complete description of the details of the experimental set-up and conditions are offered by Bianchi Janetti et al. (2012).

The following three commonly used interpretive isotherm models have been considered by Bianchi Janetti et al. (2012) to interpret the experimental results,

$$c_0 = \frac{K^F c^{n^F}}{V} + c \qquad \text{Freundlich } (F) \qquad (10a)$$

$$c_0 = \frac{1}{V}\frac{q_m^L K^L c}{1+K^L c} + c \qquad \text{Langmuir } (L) \qquad (10b)$$

$$c_0 = \frac{1}{V}\frac{K^R c}{1+\alpha^R c^{\beta^R}} + c \qquad \text{Redlich-Peterson } (R) \qquad (10c)$$

Here, c_0 (mmol L^{-1}) and c (mmol L^{-1}) are the concentration of solute initially dissolved in the fluid and adsorbed onto the soil, respectively; V (L g^{-1}) is the ratio between the solution volume and the soil mass considered in the batch experiments. The Freundlich model (Freundlich, 1906) is based on the assumption that sorption energies are characterized by an exponential distribution. The model is parameterized by the partition coefficient K^F (L^{n^F} $mmol^{1-n^F}$ g^{-1}) and a dimensionless exponent 0 < n^F < 1. The Langmuir model (Langmuir, 1918) assumes that sorption takes place at a fixed number of well-localized sites, all of which are characterized by uniformly distributed sorption energies, whereas q_m^L (mmol g^{-1}) corresponds to the saturation of all sorption sites and K^L (L $mmol^{-1}$) is the adsorption rate. The Redlich-Peterson model (Redlich-Peterson, 1959) is a combination of the Freundlich and Langmuir models is parameterized by K^R (L g^{-1}), α^R (

$L^{\beta^R}\,\mathrm{mmol}^{-\beta^R}$) and 0 < β^R <1. We follow Bianchi Janetti et al. (2012) and present our results in terms of mass per volume of solute, i.e., C (mg L^{-1}).

In the following we focus on to two diverse cases: unconstrained (Section 3.1; i.e., no data are employed to constrain model parameters and weights) and constrained (Section 3.2; i.e., model parameters and weights are evaluated on the basis of the experimental results).

3.1 Unconstrained case

Here, we perform the (moment-based) global sensitivity analysis detailed in Section 2 by considering the statistical moment of C before the parameters and weights of the candidate models are constrained through calibration against available observations. We term this case as unconstrained. In this context, we assign an equal weight, $w\left(M^j\right)=1/3$, to each of the models (10a)-(10c). We treat parameter uncertainties by assuming model parameters θ_i^j as independent identically distributed (iid) random variables, characterized by a uniform probability density function (reflecting the lack of prior knowledge about model parameters) and a constant coefficient of variation, $CV_{\theta_i^j}$ (to reduce biases in the analysis arising from considering diverse relative intervals of variation for each θ_i^j). Note that our proposed GSA approach is not limited to the specific choices we consider in this study, whereas the use of diverse formats of the prior distribution of θ_i^j and of diverse values of $CV_{\theta_i^j}$ are fully compatible with the approach. We set $CV_{\theta_i^j}=0.7$, to consider a relatively wide parameter space defined as $\left[1-\sqrt{3}CV_{\theta_i^j}, 1+\sqrt{3}CV_{\theta_i^j}\right]\times E\left[\theta_i^j\right]$, $E\left[\theta_i^j\right]$ being the expected value of θ_i^j listed in Table 1, taken to coincide with the corresponding parameter estimate evaluated by Bianchi Janetti et al. (2012). Regarding the convergence of the results, we verified that that considering N_{bins} = 100 (i.e., 10^4 or 10^6 evaluations of the Freundlich and Langmuir models or the Redlich-Peterson model, respectively) was sufficient to ensure a relative variation smaller than 1% for all of the considered results.

Figure 2 depicts the (a) expected value, $E_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$, (b) variance, $V_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$, (c) skewness, $\gamma_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$, and (d) kurtosis $k_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$, of the adsorbed concentration C versus the initial solute concentration, C_0, conditional to the Freundlich (black curve), the Langmuir (blue curves) or the Redlich-Peterson (red curves) model. The corresponding multi-model statistical moments are also depicted (purple curves). For all statistical moments of order larger than one we also depict the within-model (dashed purple curves) and the between-model (dotted purple curves) contributions. Figures 2c-d also include the mixed terms (dash-dotted curves, see definition in (8)-(9)).

Overall inspection of Figure 2 shows that in each model all statistical moments (hence the shape of the *pdf*) of C depends on C_0. All centered SMs (see Figure 2b-d) are well approximated by their within-model contributions, the between-model contribution for the variance being at least an order of magnitude smaller than the within-model terms, and the between-model and mixed terms for skewness and kurtosis being close to zero. As such, the uncertainty about the model (as imbued in the variability of the center of mass of the *pdf*s of C) to describe the adsorbed concentrations has a minor role on the ensuing uncertainty in C (as rendered through the centered SMs here analyzed) than our lack of knowledge about the model parameters.

Figure 3a depicts indices $AMAE_{K^F}$ (black continuous curve) and $AMAE_{n^F}$ (blue curve) together with the corresponding model-choice (red continuous curve) and parameter-choice (symbols) contributions as a function of C_0. These results are complemented by Figure 3b, depicting indices $AMAE_{K^F}$ (black curve) and $AMAE_{n^F}$ (blue curve), which are obtained by assuming that the Freundlich isotherm is associated with a unit weight (i.e., corresponding to a single model approach). Figures 3c, d and Figures 3e, f show the corresponding results performed for the Langmuir and Redlich-Peterson models, respectively.

Joint inspection of Figures 3a, c, e reveals that the parameter-choice contributions (linked to parameter variability given a selected model) practically coincide with indices $AMAE_{\theta_i^j}$, the model-choice contribution being at least one order of magnitude smaller than the parameter-choice contribution. Comparison of indices $AMAE_{K^F}$ and $AMAE_{n^F}$ in Figure 3a reveals that the influence of K^F on the expected value of *C* is larger than that of n^F for all values of C_0. This suggests that the uncertainty of K^F has, on average, a stronger impact than that of n^F on the mean of *C*. Figure 3c shows that K^L influences the expected value of *C* to a greater extent than q_m^L for low C_0, these two parameters tending to be equally influential as C_0 increases. Figure 3e shows that indices $AMAE_{\alpha^R}$ and $AMAE_{\beta^R}$ are significantly smaller than $AMAE_{K^R}$ and all three indices decrease with increasing C_0. One can also note that index $AMAE_{\theta_i^j}$ (Figure 3a, c, e) and its counterpart $AMAE_{\theta_i}$ (Figures 3b, d, f) display a similar trend when plotted versus C_0. This result suggests that for this unconstrained setting (where each model is associated with the same weight) the uncertainty in the model parameters induces very similar contributions to the relative change of the expected value of *C*, regardless of whether one relies on a multi- or a single-model approach.

Similar results have been obtained for indices $AMAV_{\theta_i^j}$ (see Figure 4), $AMA\gamma_{\theta_i^j}$ (see Figure A.1) and $AMAk_{\theta_i}$ (see Figure A.2).

3.2 Constrained case

Here, we perform our GSA after adsorption data become available and model parameters and weights are estimated by means of model calibration. We term this as constrained case and diagnose the influence of the residual (i.e., following model calibration) uncertainty associated with model parameters on the statistical moments of *C* in the presence of various models.

We rely on the model calibration results of Bianchi Janetti et al. (2012) who estimate model parameters through a Maximum Likelihood (ML) (see e.g., Carrera and Neuman, 1986) procedure and (posterior) model weights $w(M_j)$ on the bases of the Kashyap (1982) model discrimination criterion (KIC), allowing to consider (*i*) measurement error variance in the parameter estimation process as well as (*ii*) conceptual model uncertainty (see e.g., Höge et al., 2019; Moghadasi et al., 2015; Bianch Janetti et al., 2012; Ye et al., 2004). Parameter estimates and posterior weights are listed in Table 1. Consistent with the ML procedure adopted for model calibration, a Gaussian density is associated with each model parameter, which is then characterized by its (estimated) mean and standard deviation.

All our results are based on N_{bins} = 100, a value that is consistent with the unconstrained case and is sufficient to ensure a relative variation smaller than 1% for all of the considered results.

Table 1. Results of the ML model calibrations (from Bianchi Janetti et al. (2012)).

Model	Parameter	Estimated Mean Value	Standard Deviation	Coefficient of Variation	Model weight
Freundlich	K^F	0.99	0.04	0.04	0.8851

(10a)	n^F	0.26	0.05	0.19	
Langmuir	K^L	0.28	0.41	1.46	0.0369
(10b)	q_m^L	2.86	0.09	0.03	
Redlich-	K^R	3.89	0.14	0.04	
Peterson	α^R	3.28	0.16	0.05	0.0780
(10c)	β^R	0.79	0.01	0.01	

Figure 5 is the counterpart of Figure 2 based on the ML model calibration results and posterior model weights of Table 1. Figure 5a also includes the experimental data (black dots).

A striking feature of Figure 5a is the marked/modest relative differences observed for low/high C_0 between the expected values of C resulting from the diverse interpretive models considered. Note that such relative differences tend to be larger/smaller than those recorded for the unconstrained case (see Figure 2a) for the lowest/largest values of C_0. These features are consistent with the observation that only few data are available for model calibration at low C_0.

Comparison of $V_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$ and $V\left[C\right]$ in Figure 5b and in Figure 2b reveals that both quantities are smaller for the constrained than for the unconstrained scenario, with the exception of $V_{\boldsymbol{\theta}^L}\left[C\,|\,M^L\right]$. This result is consistent with the reduced relative uncertainty (as expressed through the coefficient of variation) associated with the parameter values estimated through model calibration (with the exception of CV_{K^L}) with respect to their prior (or unconstrained) counterparts (for which $CV_{\theta_i^j} = 0.7$). We further note that in the constrained case the between-model contribution to $V\left[C\right]$ can be of the same order of magnitude of (or even larger than) the corresponding within-model contribution. The latter result is a consequence of model calibration that causes a redistribution of the relative importance of the lack of knowledge about (*a*) model parameters and (*b*) model structure on the variability of *C*.

The comparison of $\gamma_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$ in Figure 5c and in Figure 2c reveals a reduction of the skewness of the *pdf*s of *C* as rendered by the Freundlich and Redlich-Peterson models ($\gamma_{\boldsymbol{\theta}^F}\left[C\,|\,M^F\right]$ and $\gamma_{\boldsymbol{\theta}^R}\left[C\,|\,M^R\right]$ almost vanished in the constrained scenario), and an increase of $\gamma_{\boldsymbol{\theta}^L}\left[C\,|\,L\right]$ for the constrained respect to the unconstrained scenario. On the other hand, $\gamma\left[C\right]$ (purple curve) is significantly larger (~ one order of magnitude) in the constrained than in the unconstrained scenario. Figure 5c also indicates that the within-model contributions (dashed purple curve) are dominant to $\gamma\left[C\right]$, followed by the mixed term (dot-dashed purple curve). The major factor causing the dominance of the within-model contribution (see the first term in (8)) is the term associated with the Langmuir model which, despite having a low model weight (= 0.0369), exhibits a high ratio between $V_{\boldsymbol{\theta}^L}\left[C\,|\,M^L\right]$ and $V\left[C\right]$. This observation is sustained by the observed similarity in the trend of $\gamma_{\boldsymbol{\theta}^L}\left[C\,|\,M^L\right]$ and $\gamma\left[C\right]$, as a function of C_0. A similar behavior emerges also from the comparison of $k_{\boldsymbol{\theta}^j}\left[C\,|\,M^j\right]$ and $k\left[C\right]$ (see Figure 5d and Figure 2d), and is again due to the dominance of the factor associated with the Langmuir model in the within-model contribution in (9).

Comparison of the sensitivity indices in Figure 6 with their counterparts in Figure 3 provides an indication of the way the influence of each model parameter on the expected value of *C*, as expressed in terms of $AMAE_{\theta_i^j}$, evolves as information about the system under investigation are acquired. It is possible to note that (*a*) the parameter-choice contributions to $AMAE_{\theta_i^j}$ in Figures 6a, c, e are smaller than in the uniformed case (Figure 3a, c, e), as a consequence of the model calibration procedure (with exception of the term associated with K^L); and (*b*) the model-choice contribution increases and becomes dominant for most of the parameters for low C_0, in agreement with the previously noted marked relative discrepancies between the $E_{\boldsymbol{\theta}^j}\left[C \mid M_j\right]$ for the three considered models (see Figure 5). Results included in Figure 7 complement those of Figure 4 for the constrained case. Comparison of these two sets of results clearly indicates that, following data acquisition, the major factor contributing to each index $AMAV_{\theta_i^j}$ tends to be the model-choice contribution rather than the component associated with the parameter-choice contributions, a sole exception being given by $AMAV_{K^L}$. Inspection of sensitivity indices grounded on the skewness, $AMA\gamma_{\theta_i^j}$ (Figure A.3), and the kurtosis, $AMAk_{\theta_i^j}$ (Figure A.4), provides results in line with those for the expected value and the variance (see Appendix A).

Comparison of $AMA\mathrm{SM}_{\theta_i^j}$ (Figure 6-7a, c, e) and their counterparts $AMA\mathrm{SM}_{\theta_i}$ (Figure 6-7b, d, f) reveals that for the constrained case the influence of *i*-th parameter on the mean and variance of *C* could either change markedly (when the model-choice contribution is the dominant one) or only marginally (when the parameter-choice contribution is dominant) as a consequence of considering only one or multiple models. These aspects are encapsulated, for example, in the results depicting the behavior of (*i*) the Freundlich and Redlich-Peterson models, where the model-choice contribution to $AMAV_{\theta_i^j}$ is the most relevant component, or (*ii*) the results for K^L in the Langmuir model, where the parameter-choice contribution to $AMAV_{K^L}$ is the dominant term. Similar results are detected for the sensitivity indices grounded on the skewness and the kurtosis of the *pdf*s of *C* (see Appendix A).

4. Discussion

The comparison of the importance of the model-choice and parameter-choice contributions for the unconstrained and constrained cases highlights that: (*a*) in the former case, where all models are associated with equal weight, the influence of the choice of the model on the statistical moments of *C* are negligible with respect to the impact of the uncertainty of the model parameters; (*b*) in the latter case, the variability of the first four statistical moments of *C* is more strongly controlled by the residual lack of knowledge about the adequacy of each of the candidate models to interpret the system rather than by the residual uncertainty about the estimated model parameters; (*c*) an exception to the observed decreased influence of the parameter-choice contributions is noted with reference to K^L, whereas the quality of the estimate of this parameter is quite modest (i.e., the corresponding coefficient of variation is higher than that considered for the unconstrained case, see Table 1), thus leading to a strong sensitivity of the SMs of *C* to K^L even as the Langmuir model is associated with a low posterior model weight. From a practical perspective, these observations suggest that the variability in the SMs of *C* here analyzed could be further constrained by: (*i*) collecting additional data with the aim of reducing uncertainty associated with estimates of K^L or (*ii*) excluding the Langmuir model from model set $\boldsymbol{M}$, in light of its low model weight.

We note that the results could be impacted by considering diverse formats of the *pdf*s of the parameters either in the unconstrained or in the constrained case (whereas we rely on a uniform and Gaussian distribution, respectively). Diverse studies, performed in a single-model context, highlight the relevance of the choice of the parameters' distributions on the results of a GSA (e.g., Paleari and Confalonieri, 2016; Shin et al., 2013; Wang et al., 2013; Kelleher et al., 2011). Furthermore, it has to be recognized that the results of the GSA here performed could also depend on the method employed to estimate model weights (here we resort to the Kashyap (1982) model discrimination criterion (KIC), as described in Section 3.2). It would then be of interest to analyze in future studies the relevance of diverse choices for these elements within the context of multi-model GSAs.

For the sake of clarity in the interpretation of the results, we recall here that (*a*) the sensitivity index in (1) corresponds to the (average) variation of a SM due to conditioning on a given parameter in a given model while (*b*) formulations (7)-(9) allow quantifying the contribution to the uncertainty of an output stemming from the model format and parameters. Thus, it is not surprising that, for the constrained case here examined, even as the behavior of the multi-model skewness and kurtosis of C (see Figure 5) (and, to a lesser extent, that of the multi-model variance) are clearly dictated by the terms associated with the Langmuir model in the within-model contributions in (8)-(9), it is yet possible that the major contributing factor to the sensitivity index for a parameter (e.g., β^R or q_m^L) is the model-choice contribution.

From a physical perspective, the results for the unconstrained case suggest an overall major relevance of parameter K^j (with j = F, L, R) on the modeling goals. While this parameter is associated with a different specific meaning depending on the model, it can be generally understood as a constant of proportionality driving the intensity of the sorption process. As mentioned in Section 1, we focus on the unconstrained case to gain insight about model functioning, and it is in such a set-up that we can mitigate possible bias about parameter relevance associated with diverse relative sizes of the *pdf*s' supports (i.e., we impose the same coefficient of variation for each parameter) and with diverse model weights (i.e., each model has the same weight). Otherwise, results for the constrained case are more suited to guide further system investigations.

With reference to the numerical sampling scheme, we point out that the formulation in (1) does not prevent the use of other, and possibly more efficient, sampling strategies, as compared to the one we detail in Section 2.3. We further note that considering complex and computationally expensive models might require reducing the computational burden by leveraging on the use of a surrogate model, given the ability of the latter to successfully render the required statistical moments (see e.g., Dell'Oca et al. 2017).

As a final remark, we note that it could be of interest to evaluate the way a (statistical) moment of Δ is influenced by (*a*) a parameter (synthesizing some properties/features of the investigated system) in cases where the latter appears in more than one model in $\boldsymbol{M}$; or (*b*) diverse conceptualizations and/or mathematical formulations of the various processes embedded in a model of the system under investigation. These aspects might assist in exploring answers to questions of the kind 'to which processes are the statistical moments of Δ most sensitive? Is such a sensitivity due to the uncertainty about process conceptualization or is it mainly due to parametric uncertainty employed in our conceptualization of each process?'. We present the mathematical formulation associated with these aspects in Appendix B and C, respectively. Comparisons of the present methodology with other multi-model GSAs (e.g., Dai et al., 2017; Dai and Ye, 2015) will be the subject of future studies, where we will consider the case where parameters appear in more than one model in $\boldsymbol{M}$ (in the showcase introduced in Section 3 each of the parameters considered was associated with a unique model), as well as the possibility that a model comprises processes for which alternative conceptualizations are available (here, we consider a unique process, i.e., adsorption).

5. Conclusions

Our work leads to the following major findings.

1) We develop and illustrate sensitivity indices providing metrics for global sensitivity across multiple interpretive models with uncertain parameters. We assess the sensitivity of the first four statistical moments of a target quantity of interest (Δ) with respect to its

variations related to the uncertainty in the model structure and associated parameters. Our illustrative example demonstrates that a given parameter can be associated with diverse degrees of importance, depending on the statistical moment of Δ considered.

2) Our (moment-based) sensitivity indices are structured according to two key components: (*a*) a model-choice contribution, which takes into account the possibility of analyzing the system of interest by taking advantage of various model conceptualizations (or mathematical renderings); (*b*) a parameter-choice contribution, associated with the uncertainty in the parameters of a selected model.
3) In the showcase example analyzed, involving three competing models to interpret concentration of metal sorption on soil samples, *C*, the values of the proposed sensitivity indices vary depending on whether we diagnose the system response prior to acquiring data on *C* or after observing the outcomes of some experiments and performing Maximum Likelihood models calibration with ensuing estimation of probability weights related to each model. Our moment-based indices resulting from the latter setting (here termed as constrained scenario) quantify the influence of the residual (i.e., following model calibration) uncertainty associated with model parameters on the statistical moments of *C* in the presence of multiple models.
4) Our results show that in the absence of conditioning on observations of *C* (here termed as unconstrained scenario) all sensitivity indices are dominated by the variability in the model parameters, contributions due to the various model conceptualizations being at least of an order of magnitude smaller. Otherwise, conditioning on acquired *C* data yields a decrease in the values of the sensitivity indices (a symptom of reduced relative variability in the ensuing statistical moments of *C*) and an increases of the model-choice contribution that could be of the same order of magnitude of (or even greater than) the parameter-choice term. An exception to the latter observation is noted for indices associated with parameters which estimates are characterized by poor quality.

Acknowledgements

The authors are grateful to the EU and MIUR for funding, in the frame of the collaborative international Consortium (WE-NEED) financed under the ERA-NET WaterWorks2014 Cofunded Call. This ERA-NET is an integral part of the 2015 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI). Prof. A. Guadagnini acknowledges funding from Région Grand-Est and Strasbourg-Eurométropole through the ‘Chair Gutenberg’. Data are available online (http://dx.doi.org/10.17632/55nmwbfpwg.1).

References

Baroni, G., & Tarantola, S. (2014). A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling & Software*, 51, 26-34.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, 18-36.

Bianchi Janetti, E., Dror, I., Riva, M., Guadagnini, A., & Berkowitz, B. (2012). Estimation of single-metal & competitive sorption isotherms through Maximum Likelihood and Model Quality Criteria. *Soil Science Society American Journal*, 76(4). 1229-1245. doi:10.2136/sssaj2012.0010.

Bianchi Janetti, E., Guadagnini, L., Riva, M., & Guadagnini, A. (2019). Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer. *Journal of Hydrology*, 574, 544-556. doi: 10.1016/j.jhydrol.2019.04.035.

Bredehoeft, J. D. (2005). The conceptualization model problem-surprise. *Hydrogeology Journal*, 13, 37-46.

Borgonovo, E., & Plischke, E. (2016). Sensitivity Analysis: a review of recent advances. *European Journal of Operational Research*, 248, 869-887. doi:10.1016/j.ejor.2015.06.032.

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering System Safety*, 92, 771-784.

Burnham, K. P., & Anderson, A. R. (2002). Model Selection and Multiple Model Inference: A Practical Information-Theoretical Approach. 2nd ed., *Springer*, New York.

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research*, 22 (2), 199-210.

Chu, J., Zhang, C., Fu, G., Li, Y., & Zhou, H. (2015). Improving multi-objective reservoir operation optimization with sensitivity constrained dimension reduction. *Hydrology Earth System Science*, 19, 3557-3570. doi.org/10.5194/hess-19-3557-2015.

Ciriello, V., Edery, Y., Guadagnini, A., & Berkowitz, B. (2015). Multimodel framework for characterization of transport in porous media. *Water Resources Research*, 51(5), 3384-3402. doi:10.1002/2015WR017047.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagner, T., & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44(12). https://doi.org/10.1029/2007WR006735.

Dai, H., & M. Ye (2015). Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation. *Journal of Hydrology.*, 528, 286-300. doi:10.1016/j.jhydrol.2015.06.034.

Dai, H., Ye, M., Walker, A. P., & Chen, X. (2017). A new process sensitivity index to identify important system processes under process model and parametric uncertainty. *Water Resources Research,* 53, 3476–3490. doi:10.1002/2016WR019715.

Degenring, D., Froemel, C., Dikta, G., & Takors, R. (2004). Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, 14(7), 729-745.

Dell'Oca, A., Riva, M., & Guadagnini, A. (2017). Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrology Earth System Science,* 21, 6219-6234. doi:10.5194/hess-21-6219-2017.

Draper, D. (1995). Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society, Series B, 57(1), 45-97.

Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T.A., Ackerer, P., & Guadagnini, A. (2011). Use of Global Sensitivity Analysis and Polynomial Chaos Expansion for Interpretation of Non-reactive Transport Experiments in Laboratory-Scale Porous Media. *Water Resources Research,* 47, W02521. doi:10.1029/2010WR009639.

Förster, K., Meon, G., Marke, T., & Strasser, U. (2014). Effect of meteorological forcing & snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany). *Hydrology Earth System Science*, 18, 4703-4720. doi:10.5194/hess-18-4703-2014.

Gupta, H. V., Wagener, T., & Liu, Y. Q. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrology Processes*, 22(18), 3802–3813.

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.

Hill, M. C., Foglia, L., Christensen, S., Rakovec, O., & Borgonovo, E. (2016). Model validation: Testing models using data and sensitivity analysis. *The Handbook of Groundwater Engineering*, edited by Cushman J. H. & Tartakovsky, D., 3rd ed., Taylor & Francis, Boca Raton, Fla. doi:10.1201/9781315371801-22.

Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrology Earth System Science*, 17, 5109–5125. doi:10.5194/hess-17-5109-2013.

Höge, M., Wöhling, T., & Nowak, W. (2018). A primer for model selection: the decisive role of model complexity. *Water Resources Research*, 54(3), 1688-1715. doi:10.1002/2017WR021902.

Höge, M., Guthke, A., & Nowak, W. (2019). The hydrologist's guide to Bayesian model selection, averaging and combination. *Journal of Hydrology*, 572, 96-107. https://doi.org/10.1016/j.jhydrol.2019.01.072.

Hölter, R., Zhao, C., Mahmoudia, E., Lavasan, A. A., Datcheva, M., Königa, M., & Schanza, T. (2018). Optimal measurements design for parameter identification in mechanized tunneling. *Underground Space*, 3(1), 34-44. doi:10.1016/j.undsp.2018.01.004.

Hutcheson, R. S., & McAdams, D. A. (2010). A hybrid sensitivity analysis for use in early design. *Journal of Mechanical Design*, 132(11). doi:10.1115/1.4001408.

Kashyap, R. L. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 4(2), 99–104.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American statistical association*, 90(430), 773-795.

Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K. & Marshall, L. (2012). Investigating controls on the thermal sensitivity of Pennsylvania streams. *Hydrological Processes*, 26, 771-785. doi:10.1002/hyp.8186.

Knuth, K. H., Habeck, M., Malakar, N. K., Mubeen, A. M., & Placek, B. (2015). Bayesian evidence and model selection. *Digital Signal Processes*, 47, 50–67.

Koutsoyiannis, D. (2010). HESS Opinions "A random walk on water". *Hydrology Earth System Science*, 14, 585-601. doi:10.5194/hess-14-585-2010.

Liu, H., Sudjianto, A., & Chen, W. (2006). Relative entropy based method for probabilistic sensitivity analysis in engineering design. *Journal of Mechanical Design*, 128, 326-336.

Liu, P., Elshall, A. S., Ye, M., Beerli, P., Zeng, X., Lu, D., & Tao, Y. (2016). Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods. *Water Resources Research,* 52, 734–758. doi:10.1002/2014WR016718.

Moghadasi, L., Guadagnini, A., Inzoli, F., & Bartosek, M. (2015). Interpretation of two-phase relative permeability curves through multiple formulations and model quality criteria. *Journal* of *Petroleum Science Engineering*, 126, 738-749. doi:10.1016/j.petrol.2015.10.027.

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306, 127-145.

Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models. *Stochastic Environmental Research and Risk Assessment*, 17 (5), 291-305. doi:10.1007/s00477-003-0151-7.

Neuman, S.P., Xue, L., Ye, M., & Lu, D. (2012). Bayesian analysis of data-worth considering model and parameter uncertainties. *Advances in Water Research,* 36, 75-85. doi:10.1016/j.advwatres.2011.02.007.

Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling Software*, 26(12), 1515–1525.

Paleari, L., & Confalonieri, R. (2016). Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions. *Ecological Modelling*, 340, 57–63. https://doi.org/10.1016/j.ecolmodel.2016.09.008.

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: survey and outlook. *Water Resources Research,* 51, 7090–7129. doi:10.1002/2015WR017780.

Pappenberger, F., Beven, K. J., Ratto, M., & Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances Water Research*, 31(1), 1–14.

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 & Software,* 79, 214-232.

Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67, 1-11.

Poeter, E. P., & Anderson, D. A. (2005). Multimodel ranking and inference in ground water modeling. *Ground Water,* 43 (4), 597–605.

Poeter, E. P., & Hill, M. C. (2007). MMA: A computer code for multi-model analysis. *United States Geological Survey,* Technology & Methods 6-E3, USGS, Reston, Va.

Punzo, V., Marcello, M., & Biagio, C. (2015). Do we really need to calibrate all the parameters? Variance-based sensitivity analysis to simplify microscopic traffic flow models. *IEEE Transaction on Intelligent Transport Systems*, 16, 184-193.

Rakovec, O., Hill, M. C., Clark, M., Weerts, A., Teuling, A., & Uijlenhoet, R. (2014). Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. *Water Resources Research*, 50(1), 409-426. https://doi.org/10.1002/2013WR014063.

Ranaee, E., Riva, M., Porta, G., & Guadagnini, A. (2016). Comparative Assessment of Three-Phase Oil Relative Permeability Models. *Water Resources Research,* 52, 1-16. doi:10.1002/2016WR018872.

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. Environmental Modelling & Software, 114, 1-11. https://doi.org/10.1016/j.envsoft.2018.12.002.

Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient globa sensitivity analysis: Part I – Theory. *Water Resources Research*, 52, 423-439. doi:10.1002/2015WR017558.

Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: Part II – Applications. *Water Resources Research*, 52, 440-455. doi:10.1002/2015WR017559.

Rodríguez-Escales, P., Canelles, A., Sanchez-Vila, S., Folch, A., Kurtzman, D., Rossetto, R., Fernández-Escalante, E., Lobo-Ferreira, J.-P., Sapiano, M., San-Sebastián, J., & Schüth, C. (2018). A risk assessment methodology to evaluate the risk failure of managed aquifer recharge in the Mediterranean Basin. *Hydrology Earth System Science*, 22, 3213-3227. doi:10.5194/hess-22-3213-2018.

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 & Software*, 37, 103–109.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). Global Sensitivity Analysis. The Primer, *Wiley*, 305 pp.

Schoniger, A., Wholing, T., Samaniego, L., & Nowak, W. (2014). Model selection on solid ground: rigorous comparison of nine ways to evaluate Bayesian model Evidence. *Water Resources Research,* 50. doi:10.1002/2014WR016062.

Schoups, G., & Hopmans, J. W. (2006). Evaluation of model complexity and input uncertainty of field-scale water flow and transport. *Vadose Zone Journal*, 5, 951-962. doi:10.2136/vzj2005.0130.

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. *Journal of Hydrology*, 503, 135-152. https://doi.org/10.1016/j.jhydrol.2013.08.047.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, 324(1–4), 10–23.

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, W02501. doi:10.1029/2008WR007347.

Wagener, T., & Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, 47, W060301. doi:10.1029/2010WR009469, 2011.

Wang, S., Ancell, B. C., Huang, G. H., & Baetz, B. W. (2018). Improving Robustness of Hydrologic Ensemble Predictions Through Probabilistic Pre- & Post-Processing in Sequential Data Assimilation. *Water Resources Research*, 54(3). 2129-2151. doi:10.1002/2018WR022546.

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.

Wöhling, T., & Vrugt, J. A. (2008). Combining multiobjective optimization & Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. *Water Resources Research*, 44, W12432. doi:10.1029/2008WR007154.

Younes, A., Delay, F., Fajraoui, N., Fahs, M., & Mara, T. A. (2016). Global sensitivity analysis and Bayesian parameter inference for solute transport in porous media colonized by biofilms. Journal of Contaminant Hydrology, 191, 1-18. doi:10.1016/j.jconhyd.2016.04.007.

Ye, M., Neuman, S. P., & Meyer, P. D. (2004). Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resources Research* 40, W05113. doi.org/10.1029/2003WR002557.

Ye, M., Neuman, S. P., Meyer, P. D., & Pohlmann, K. F. (2005). Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff. *Water Resources Research*, 41, W12429. doi:10.1029/2005WR004260.

Ye, M., Meyer, P. D., & Neuman, S. P. (2008). On model selection criteria in multimodel analysis. *Water Resources Research*, 44, W03428. doi:10.1029/2008WR006803.

Ye, M., Pohlmann, K. F., & Chapman, J. B. (2008). Expert elicitation of recharge model probabilities for the Death Valley regional flow system. *Journal of Hydrology*, 354, 102-115.

Yue, H., Brown, M., He, F., Jia, J., & Kell, D. B. (2008). Sensitivity analysis & robust experimental design of a signal transduction pathway system. *International Journal of Chemical Kinetics*, 40, 730-741. doi:10.1002/kin.20369.

Appendix A. Sensitivity Analysis grounded on the Skewness and the Kurtosis of model output, Unconstrained and Constrained Scenarios.

Figure A.1a depicts $AMA\gamma_{K^F}$ (black continuous curve) and $AMA\gamma_{n^F}$ (blue curve) together with the corresponding model-choice (red continuous curve) and parameter-choice (symbols) contributions as a function of C_0 for the unconstrained scenario. These results are complemented by Figure A.1b, depicting indices $AMA\gamma_{K^F}$ (black curve) and $AMAE_{n^F}$ (blue curve), which are obtained by assuming a unit weight (i.e., single model approach) for the Freundlich isotherm (10a). Figures A.1c-A1.d and Figures A.1e-f show the corresponding results obtained for the Langmuir (10b) and Redlich-Peterson (10c) models, respectively. Figure A.2 depicts the collection of companion results for $AMAk_{\theta_i^j}$ and $AMAk_{\theta_i}$.

Joint analysis of Figures A.1-A.2 and Figures 2-3 reveals an overall similarity in the sensitivity of the first four statistical moments of *C* with respect to (*i*) the parameters of each model, either in the multi- or single-model context, and (*ii*) the (less relevant) contributions associated with the model choice.

Figures A.3-A.4 are the counterparts of Figures A.1-A.2 for the constrained scenario, respectively. Comparison of Figures 3.A - 4.A with Figures 6 - 7 indicates an overall similarity of the sensitivity of the diverse SMs with respect to parameters- and the model-choice contributions.

Appendix B. Extension of the AMA sensitivity indices for the multi-model context when a parameter appears within multiple models.

One can assess the influence that a parameter associated with multiple models can have on a given statistical moment (SM) of an investigated quantity (Δ) in terms of the following indices

$$AMA\mathrm{SM}_{\theta_i^{\mathbf{M}}} = \sum_{M^j:\theta_i \subseteq \boldsymbol{\theta}|M^j} AMA\mathrm{SM}_{\theta_i^j} \tag{B.1}$$

where $AMA\mathrm{SM}_{\theta_i^j}$ is given by (1). The key difference between $AMA\mathrm{SM}_{\theta_i^j}$ and $AMA\mathrm{SM}_{\theta_i^{\mathbf{M}}}$ is that the former answers the question 'how does the variability in a given parameter influence $\mathrm{SM}\left[\Delta\right]$, when we look at such a parameter only in model M^j?' while the latter answers the question 'how does the variability in a given parameter influence $\mathrm{SM}\left[\Delta\right]$, when we look at such a parameter across the set of models within which it appears?'.

As such, index $AMA\mathrm{SM}_{\theta_i^j}$ allows quantifying the implications on SM of selecting model M^j (at the expenses of other models) and setting parameter θ_i^j to a given value. Otherwise, $AMA\mathrm{SM}_{\theta_i^{\mathbf{M}}}$ focuses on the diagnosis of the relevance of parameter θ_i across a collection of competing model alternatives within which θ_i appears. Note that $AMA\mathrm{SM}_{\theta_i^{\mathbf{M}}} \equiv AMA\mathrm{SM}_{\theta_i^j}$ if θ_i is associated with only one model of the collection.

When $\mathrm{SM}\left[\Delta\right]$ corresponds to the variance of Δ, it should be noted that $AMAV_{\boldsymbol{\theta}_i}$ differs from the index $S_{\boldsymbol{\theta}_i}$ proposed by Dai and Ye (2015) and defined as

$$S_{\boldsymbol{\theta}_i} = \frac{\sum_{M^j:\theta_i^j\subseteq\boldsymbol{\theta}^j} w\left(M^j\right)\left(E_{\theta_i^j}\left[V_{\boldsymbol{\theta}^j}\left[\Delta \,|\, M^j\right] - V_{-\theta_i^j}\left[\Delta \,|\, M^j\right]\right]\right)}{\sum_{M^j:\theta_i^j\subseteq\boldsymbol{\theta}^j} w\left(M_j\right) V_{\boldsymbol{\theta}^j}\left[\Delta \,|\, M^j\right]} \tag{2.B}$$

Index $S_{\boldsymbol{\theta}_i}$ quantifies the influence of θ_i^j to Δ as the ratio between: (*i*) the expected change of the variance of Δ for a given model, i.e., $V_{\boldsymbol{\theta}^j}\left[\Delta \,|\, M^j\right]$, due to the knowledge of parameter θ_i^j, as averaged over all of the models in which θ_i^j appears; and (*ii*) the weighted-average (over the diverse competing models in which θ_i^j appears) of $V_{\boldsymbol{\theta}^j}\left[\Delta \,|\, M^j\right]$. As such, $AMAV_{\boldsymbol{\theta}_i}$ and $S_{\boldsymbol{\theta}_i}$ convey different information about the nature of the sensitivity of Δ with respect to θ_i^j.

With reference to the showcase detailed in Section 3, it should be noted that index $S_{\boldsymbol{\theta}_i}$ in (2.B) coincides with the well-known Sobol index, given that each parameter θ_i^j appears only in a given model M^j and not across a set of models.

Appendix C. Embedding uncertainty of model processes within the AMA indices

The metric proposed in Section 2 (Eq. (1)) could be modified to evaluate the sensitivity of a given statistical moment with respect to the possibility of rendering a process (or multiple processes) involved in the model construction through a variety of alternative mathematical formulations. In this context, each model, M^j in the set $\boldsymbol{M}$ is viewed as a union of diverse processes, i.e.,

$$M^j = \left(\cup P^k\left(\boldsymbol{\theta}^{j,k}\right)\right) \tag{C.1}$$

Here, P^k is the *k*-th process and $\boldsymbol{\theta}^{j,k}$ is the vector of parameters relevant to the *k*-th process, as rendered through the mathematical formulation associated with model M^j. As an example, a subsurface solute transport model of an adsorbable compound could include a variety of processes (e.g., adsorption, advection, hydrodynamic dispersion, distributed recharge, or others), each characterized by its own set of parameters. All possible combinations of the identified model processes give rise to the set of models $\boldsymbol{M}$.

It is then possible to group such *k* mathematical formulations of a given process within vector $\boldsymbol{P}^k$. One can quantify the influence of these various mathematical formulations of the process on a target statistical moment (SM) of an investigated quantity (Δ) as

$$AMA\mathrm{SM}_{\boldsymbol{P}^k} = \underbrace{\frac{1}{\Im}\sum_{M^j:P^k\subseteq\boldsymbol{P}^k|M^j} w\left(M^j\right)\left\{\left|\mathrm{SM}[\Delta] - \mathrm{SM}_{\boldsymbol{\theta}^{j,k},-\boldsymbol{P}^k}\left[\Delta \,|\, M^j\right]\right|\right\}}_{model\text{-}choice\ contribution} +$$

$$+\underbrace{\frac{1}{\Im}\sum_{M^j:P^k\subseteq\boldsymbol{P}^k|M^j} w\left(M^j\right)\left\{\sum_{\theta_i^{j,k}|M^j} E_{\theta_i^{j,k}|M^j}\left[\left|\mathrm{SM}_{\boldsymbol{\theta}^{j,k},-\boldsymbol{P}^k}\left[\Delta \,|\, M^j\right] - \mathrm{SM}_{-\theta_i^{j,k},-\boldsymbol{P}^k}\left[\Delta \,|\, M^j\right]\right|\right]\right\}}_{parameter\text{-}choice\ contribution}$$

(C.2)

Here, $\mathrm{SM}_{\boldsymbol{\theta}^{j,k},-\boldsymbol{P}^k}\left[\Delta \,|\, M^j\right]$ is the selected statistic, which is evaluated considering variability in the parameters of the *k*-th process in model M^j, the variability associated with other processes components, i.e., $-\boldsymbol{P}^k$, having been averaged out; $\mathrm{SM}_{-\theta_i^{j,k},-\boldsymbol{P}^k}\left[\Delta \,|\, M^j\right]$ is the statistic conditioned to parameter $\theta_i^{j,k}$, of process P^k and given M^j, the variability associated with the remaining parameters of P^k given M^j having been averaged out (as well as the one linked to other processes, i.e., $-\boldsymbol{P}^k$, in model M^j).

Note that it is still possible to distinguish between (*a*) a contribution due to variability in the conceptualization (or mathematical rendering) of a process and (*b*) a contribution determined by the lack of knowledge about the parameters appearing in the mathematical formulation of the process. The former or the latter contribution vanishes if the mathematical model of the *k*-th process or its parameters are deterministically known, respectively.

Figure D.1 depicts a sketch of the overall concept underpinning index $AMA\mathrm{SM}_{P^k}$ when only two processes (i.e., k = 1, 2) are considered. Three conceptualizations of process P^1 are considered, each characterized by two parameters affected by uncertainty. Otherwise, only one conceptualization is considered for process P^2, whose parameters are taken to be deterministic. This gives rise to three distinct models. Note that values of $\mathrm{SM}_{\boldsymbol{\theta}^{j,k},-\boldsymbol{P}^k}\left[\Delta \mid M^j\right]$ for k = (1, 2) coincide in this example, because P^2 is deterministically known.

One can then visualize the nature of the various contributions to index $AMA\mathrm{SM}_{P^k}$ through the depiction of Figure C.1 When considering, for example, index $AMA\mathrm{SM}_{P^1}$, we observe that: (*i*) the sum of the distances between $\mathrm{SM}[\Delta]$ and $\mathrm{SM}_{\boldsymbol{\theta}^{j,1},-\boldsymbol{P}^1}\left[\Delta \mid M^j\right]$ (dark purple double pointed arrow) across models M^j corresponds to the model-choice contribution; (*ii*) the weighted sum over the models M^j of the averaged distances between $\mathrm{SM}_{\boldsymbol{\theta}^{j,1},-\boldsymbol{P}^1}\left[\Delta \mid M^j\right]$ and $\mathrm{SM}_{-\theta_i^{j,1},-\boldsymbol{P}^1}\left[\Delta \mid M^j\right]$ (dark and light blue, green and yellow double pointed arrows for parameters $\theta_1^{j,1}$ and $\theta_2^{j,1}$ in model M^1, M^2, and M^3, respectively) corresponds to the parameter-choice contribution. We remark that, even as we consider only one conceptualization (here with deterministic parameter(s)) for process P^2, there is still a model-choice contribution (as quantified by the black double pointed arrows in Figure 1.C) which reflects the observation that Δ is determined by the interaction of processes P^1 and P^2.

In case $\mathrm{SM}[\Delta]$ corresponds to the variance of Δ (i.e., $\mathrm{SM}[\Delta] \equiv V[\Delta]$), it should be noted that $AMA\mathrm{SM}_{P^k}$ differs from the index S_{P^k} proposed by *Dai et al*. [2017], which reads

$$PS_k = \sum_{M^j:P^k \subseteq \boldsymbol{P}^k \mid M^j} \frac{w\left(M^j\right)\left\{V[\Delta]-V_{\boldsymbol{\theta}^{j,k},-\boldsymbol{P}^k}\left[\Delta \mid M_j\right]\right\}}{V[\Delta]} \tag{C.3}$$

The latter identifies the most influential process as the one leadings to the highest reduction of the overall variance of Δ. As such, it is a metric focusing on apportionment of uncertainty (due to variability in model process conceptualization). Otherwise, index $AMA\mathrm{SM}_{P^k}$ is keyed to quantifying variability in a given statistical moment of Δ, highlighting the contribution due to uncertainty in the process conceptualization and in the process parameter(s). It is then clear that $AMA\mathrm{SM}_{P^k}$ and PS_k provide different information and could potentially be jointly used to assist comprehensive sensitivity analyses, a topic which is the subject of a future study.

With reference to the showcase detailed in Section 3, the evaluation of (C.2) and (C.3) would be of very limited interest, given that each model M^j encompasses only one process (i.e., adsorption).

Figures

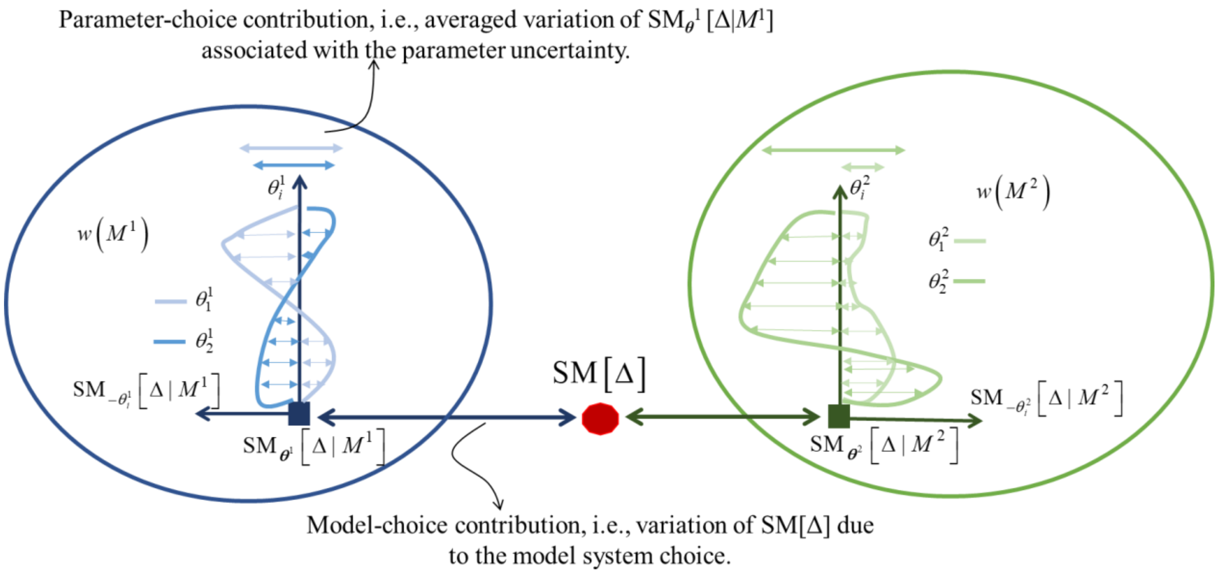

Figure 1. Sketch of the overall concept underpinning the sensitivity indices in (1), considering the generic statistical moment SM for the quantity of interest Δ. For ease of illustration we consider only two interpretative models, M^1 and M^2, with corresponding model weights $w(M^1)$ and $w(M^2)$. Each model has a set of two uncertain parameters, i.e., (θ_1^1, θ_2^1) for M_1 and (θ_1^2, θ_2^2) for M_2. $\mathrm{SM}[\Delta]$, $\mathrm{SM}_{\theta^j}\left[\Delta \mid M^j\right]$ and $\mathrm{SM}_{:\theta_i^j}\left[\Delta \mid M^j\right]$ with j =(1, 2) and i = (1, 2) correspond to the ensuing SM evaluated in a multi-model context, evaluated considering uncertainty in all the parameters of model M^j and evaluated considering uncertainty in all the parameters of model M^j except θ_i^j .

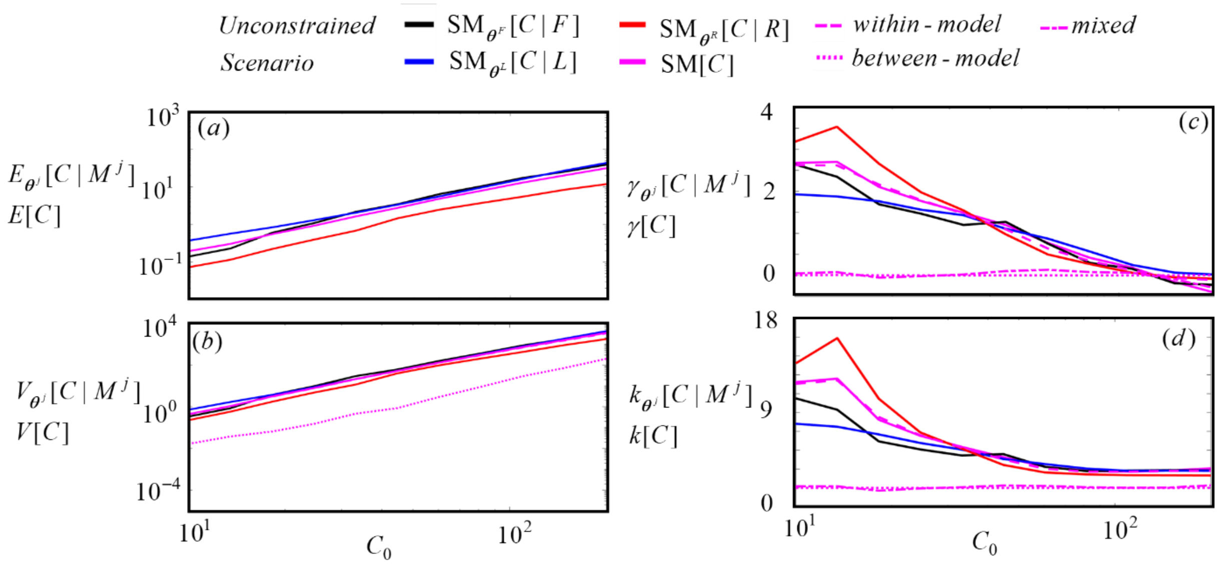

Figure 2. Unconstrained scenario, a) expected value, (b) variance, (c) skewness, and (d) kurtosis, of the adsorbed concentration C versus the initial solute concentration, C_0, conditional to the Freundlich (black curve), the Langmuir (blue curves) and the Redlich-Peterson (red curves) model. The corresponding multi-model statistical moments are also depicted (purple curves). For all central statistical moments the within-model (dashed purple curves) and the between-model (dotted purple) contributions are also depicted. Figures 2c-d also include the mixed terms (defined in (7)-(8)).

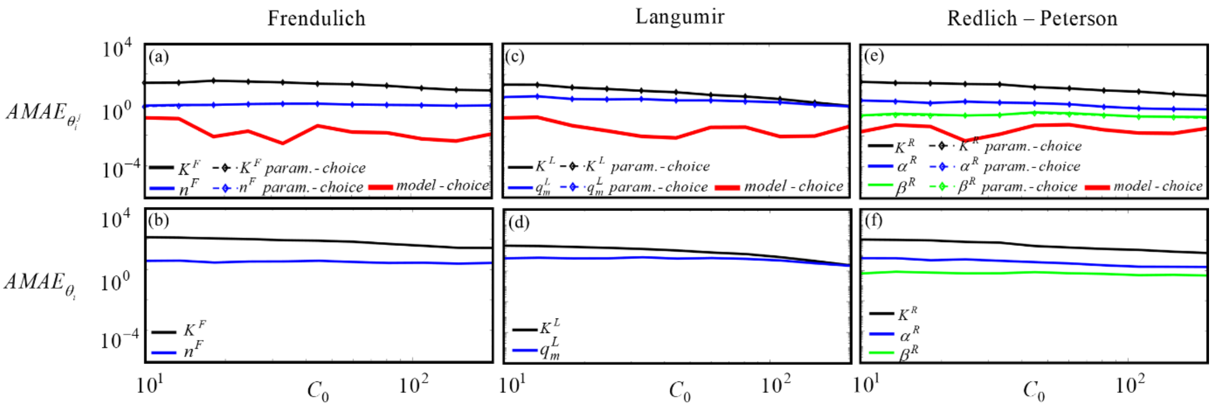

Figure 3. Unconstrained scenario. Sensitivity indices (1) associated with the expected value of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAE_{K^F}$, black continuous curve, and $AMAE_{n^F}$, blue continuous curve); (c) the Langumir ($AMAE_{K^L}$, black continuous curve, and $AMAE_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAE_{K^R}$, black continuous curve, $AMAE_{\alpha^R}$, blue continuous curve, and $AMAE_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

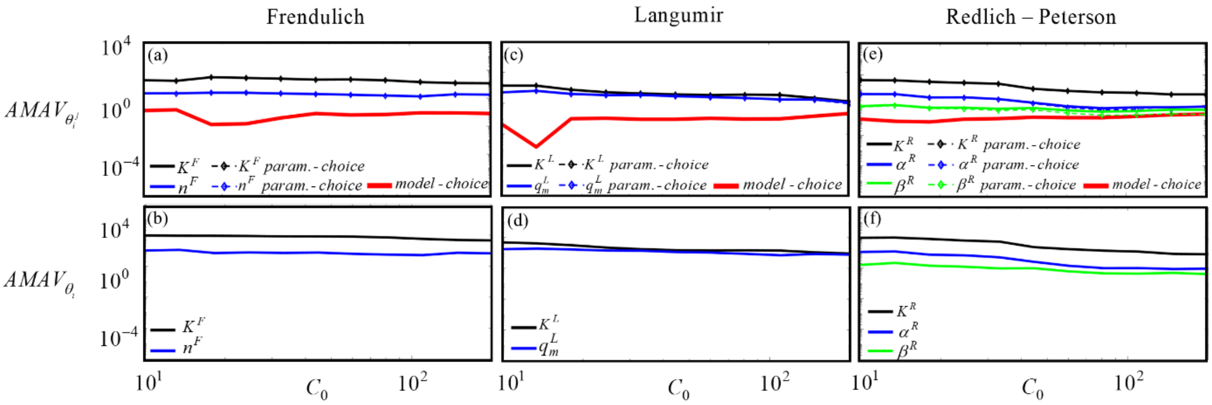

Figure 4. Unconstrained scenario. Sensitivity indices (1) associated with the Variance of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAE_{K^F}$, black continuous curve, and $AMAE_{n^F}$, blue continuous curve); (c) the Langumir ($AMAE_{K^L}$, black continuous curve, and $AMAE_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAE_{K^R}$, black continuous curve, $AMAE_{\alpha^R}$, blue continuous curve, and $AMAE_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

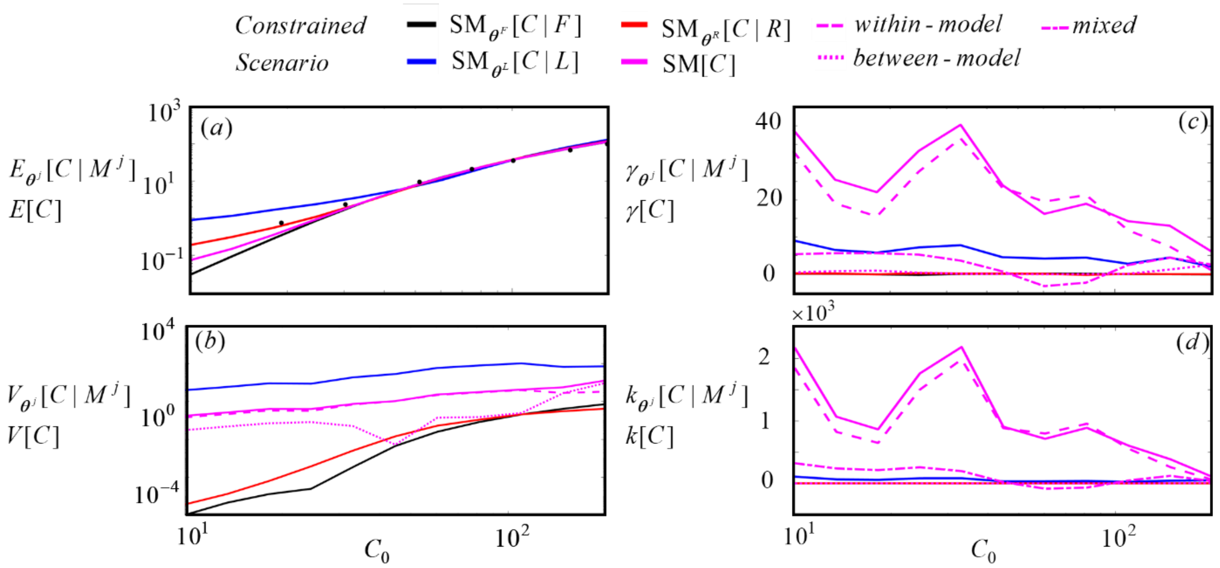

Figure 5. Constrained scenario, (a) expected value, (b) variance, (c) skewness, and (d) kurtosis, of the adsorbed concentration C versus the initial solute concentration, C_0, conditional to the Freundlich (black curve), the Langmuir (blue curves) and the Redlich-Peterson (red curves) model. The corresponding multi-model statistical moments are also depicted (purple curves). For all central statistical moments the within-model (dashed purple curves) and the between-model (dotted purple) contributions are also depicted. Figures 5c-d also include the mixed terms (defined in (7)-(8)).

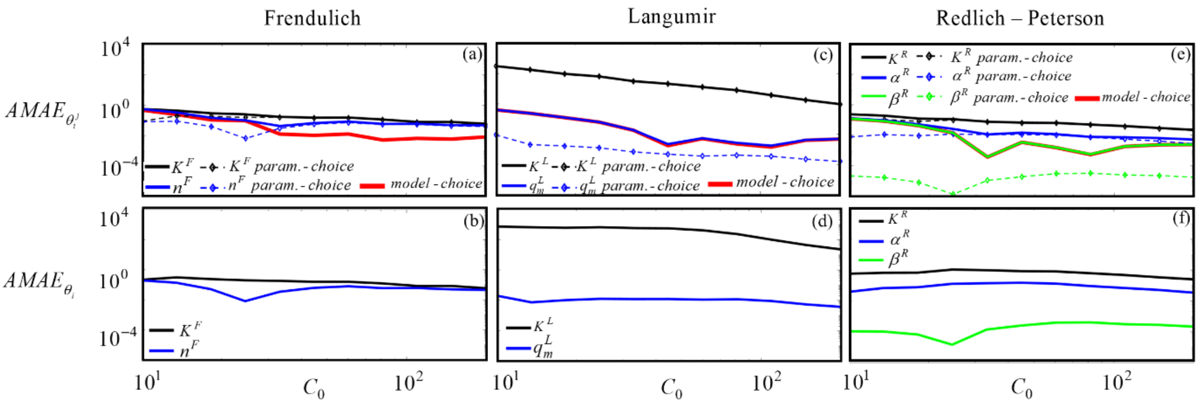

Figure 6. Constrained scenario. Sensitivity indices (1) associated with the expected value of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAE_{K^F}$, black continuous curve, and $AMAE_{n^F}$, blue continuous curve); (c) the Langumir ($AMAE_{K^L}$, black continuous curve, and $AMAE_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAE_{K^R}$, black continuous curve, $AMAE_{\alpha^R}$, blue continuous curve, and $AMAE_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

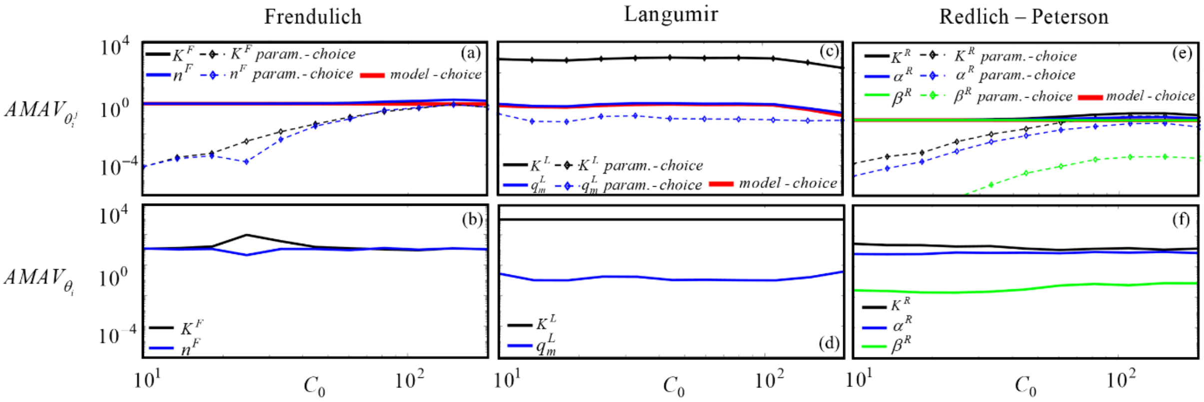

Figure 7. Constrained scenario. Sensitivity indices (1) associated with the Variance of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAE_{K^F}$, black continuous curve, and $AMAE_{n^F}$, blue continuous curve); (c) the Langumir ($AMAE_{K^L}$, black continuous curve, and $AMAE_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAE_{K^R}$, black continuous curve, $AMAE_{\alpha^R}$, blue continuous curve, and $AMAE_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

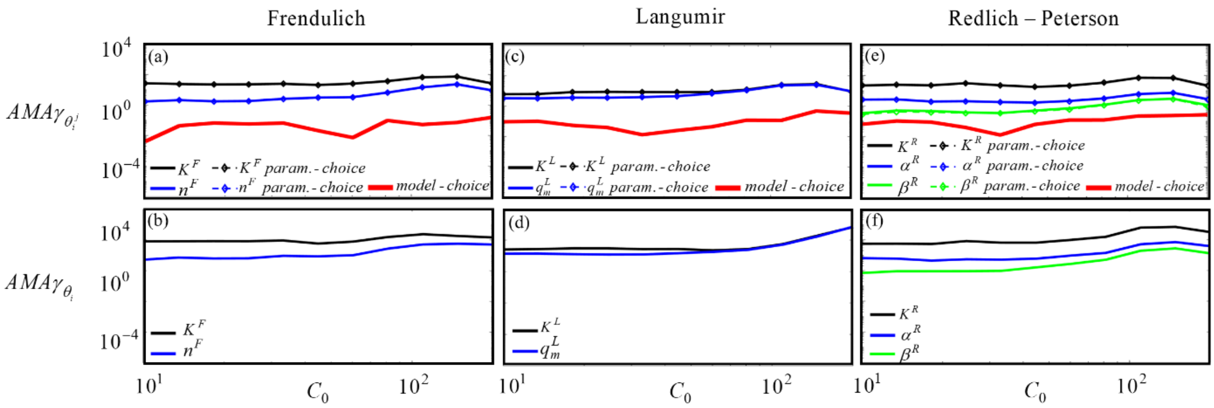

Figure A.1. Unconstrained scenario. Sensitivity indices in (1) associated with the skewness of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMA\gamma_{K^F}$, black continuous curve, and $AMA\gamma_{n^F}$, blue continuous curve); (c) the Langumir ($AMA\gamma_{K^L}$, black continuous curve, and $AMA\gamma_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMA\gamma_{K^R}$, black continuous curve, $AMA\gamma_{\alpha^R}$, blue continuous curve, and $AMA\gamma_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) of each model, are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

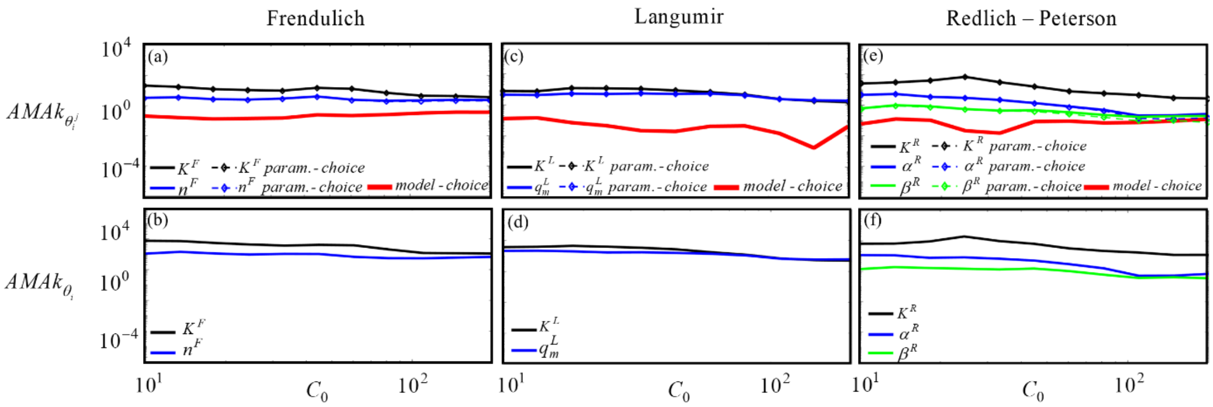

Figure A.2. Unconstrained scenario. Sensitivity indices in (1) associated with the kurtosis of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAk_{K^F}$, black continuous curve, and $AMAk_{n^F}$, blue continuous curve); (c) the Langumir ($AMAk_{K^L}$, black continuous curve, and $AMAk_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAk_{K^R}$, black continuous curve, $AMAk_{\alpha^R}$, blue continuous curve, and $AMAk_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) of each model, are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

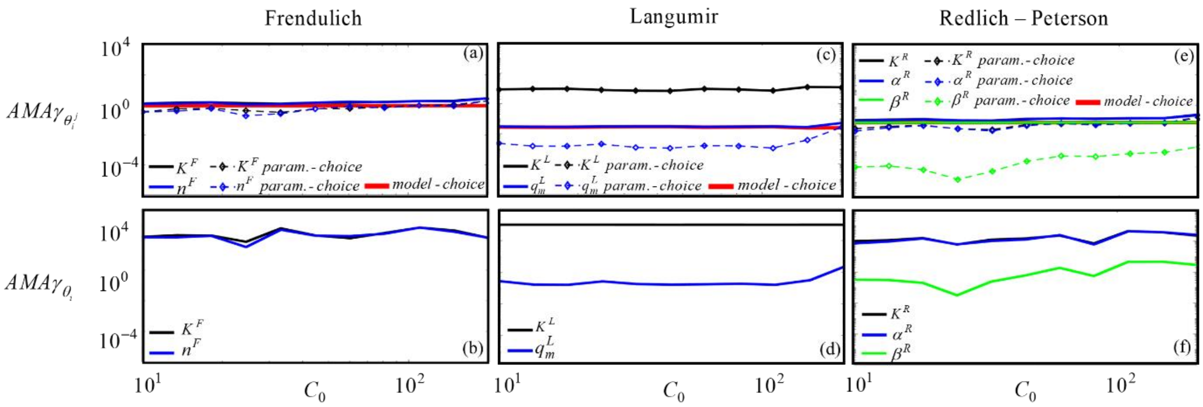

Figure A.3. Constrained scenario. Sensitivity indices in (1) associated with the skewness of C for the multi-model context versus C_0, considering (a) the Freundulich ($AMA\gamma_{K^F}$, black continuous curve, and $AMA\gamma_{n^F}$, blue continuous curve); (c) the Langumir ($AMA\gamma_{K^L}$, black continuous curve, and $AMA\gamma_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMA\gamma_{K^R}$, black continuous curve, $AMA\gamma_{\alpha^R}$, blue continuous curve, and $AMA\gamma_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) of each model, are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

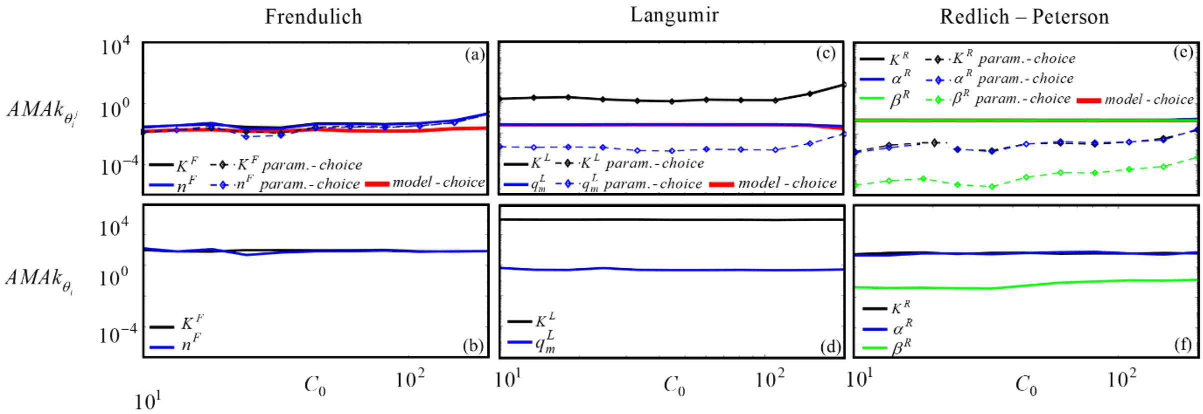

Figure A.4. Constrained scenario. Sensitivity indices in (1) associated with the kurtosis of C for the multi-model context versus C_0 considering (a) the Freundulich ($AMAk_{K^F}$, black continuous curve, and $AMAk_{n^F}$, blue continuous curve); (c) the Langumir ($AMAk_{K^L}$, black continuous curve, and $AMAk_{q_m^L}$, blue continuous curve); (e) the Redlich-Peterson ($AMAk_{K^R}$, black continuous curve, $AMAk_{\alpha^R}$, blue continuous curve, and $AMAk_{\beta^R}$, green continuous curve). For each index the corresponding (*i*) parameter-choice contribution (symbols) and (*ii*) the model-choice contribution (red curves) of each model, are also depicted. The counterparts for the single-model context are depicted for (b) the Freundulich, (d) the Langumir and (f) the Redlich-Peterson model.

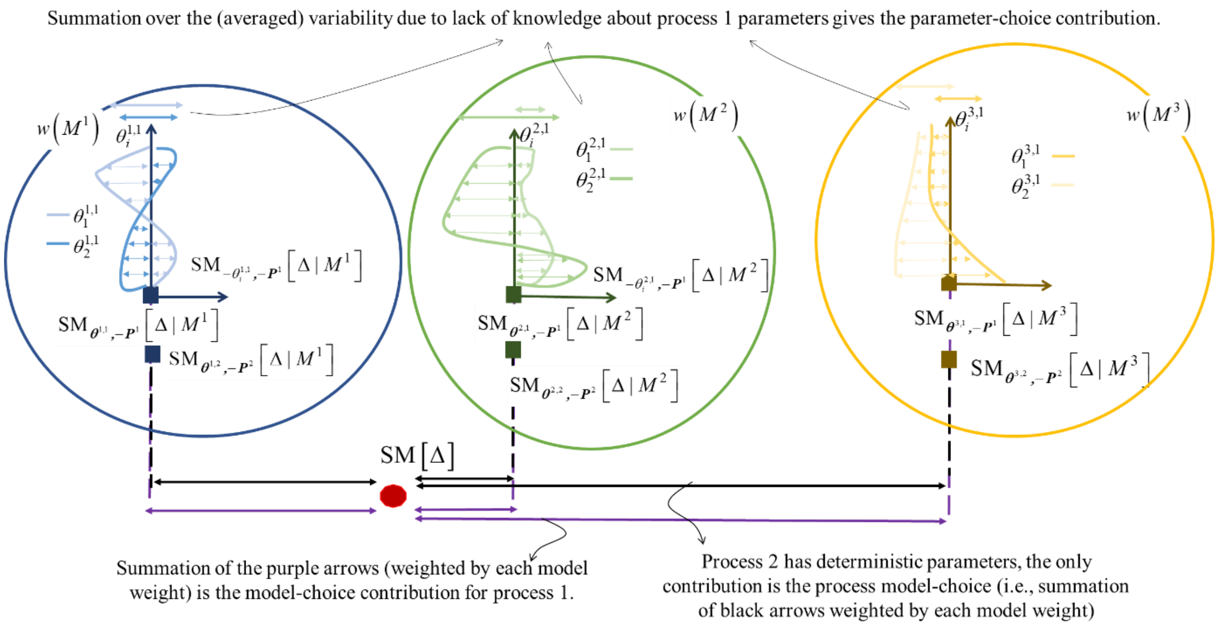

Figure D.1. Sketch of the overall concept underpinning the sensitivity indices $AMASM_{P^k}$ in (B.2), considering the generic statistical moment SM for the quantity of interest Δ. For ease of illustration we consider only two processes, i.e., P^k with k = (1, 2). Three distinct conceptualizations for process P^1 are considered, each characterized by two parameters, i.e., i = (1, 2), affected by uncertainty, i.e., $\theta_i^{j,1}$. Otherwise, only one conceptualization is considered for process P^2, whose parameters are taken to be deterministic. This gives rise to three distinct models, i.e., j = (1, 2, 3), each with its model-weight, i.e., $w(M^j)$. $\mathrm{SM}[\Delta]$ is the ensuing SM once uncertainty in all processes and associated parameters has been accommodated; $\mathrm{SM}_{\boldsymbol{\theta}^{j,k},: \mathbf{P}^{j,k}}\left[\Delta \mid M^j\right]$ is the target SM once uncertainty in the parameters of the k-th process in model M_j has been accommodated, the variability associated with other process components having been averaged out; $\mathrm{SM}_{:\theta_i^{j,k},: \mathbf{P}^{j,k}}\left[\Delta \mid M^j\right]$ is the target SM conditioned to $\theta_i^{j,k}$ of process P^k and given M^j, while the variability associated with the remaining parameter of P^k given M_j has been averaged out.