Received: 14 May 2020 | Revised: 6 July 2021 | Accepted: 7 October 2021

DOI: 10.1111/sjos.12560

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Interactions and computer experiments

Emanuele Borgonovo[1] | **Elmar Plischke**[2] | **Giovanni Rabitti**[3]

[1]Department of Decision Sciences and Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy

[2]Institute of Disposal Research, Clausthal University of Technology, Clausthal-Zellerfeld, Germany

[3]Department of Actuarial Mathematics and Statistics and Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, UK

Correspondence
Emanuele Borgonovo, Department of Decision Sciences and Bocconi Institute for Data Science and Analytics, Bocconi University, Via Roentgen 1, 20136, Milan, Italy.
Email: emanuele.borgonovo@unibocconi.it

[Correction added on 26 May 2022, after first online publication: CRUI funding statement has been added.]

Abstract

Identifying interactions and understanding the underlying generating mechanism is essential for interpreting the response of black-box models. We offer a systematic analysis of interaction types and corresponding sources, merging results of the broad statistical literature with findings developed within the computer experiment literature. Piecewise-definiteness emerges a self-standing interaction mechanism, alternative to the presence of interaction terms. We find that the scale of the analysis is essential for interpretation, and that no single method is capable of providing the correct identification of the underlying interaction generating mechanisms; conversely a combined approach involving indicators at difference scales is required. We propose a graphical tool called Mikado plot that exploits the link between interaction indicators at the finite scale and global scales to ease the regional visualization of two-factor interactions. The findings are illustrated via numerical experiments with three well-known computer models of different dimensionality and structure.

KEYWORDS

design and analysis of computer experiments, functional ANOVA, global sensitivity indices

1 | INTRODUCTION

The study of interactions is an integral part of statistical investigations (Cox, 1984; Vanderweele, 2015; Wu, 2015). However, the interpretation and the identification of the underlying generating mechanisms are delicate tasks.

Example 1. Consider an analyst working on a dataset generated by a black-box computer code via Monte Carlo simulation whose response Y depends on two uncertain inputs (or factors), X_1, X_2. The analyst wishes to gain information about the underling input–output mapping, supposing the relationship $Y = g(X_1, X_2)$. A classic approach to shed some light on g is to fit a linear model. Supposing that

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \tag{1}$$

the analyst obtain a coefficient of model determination $R^2 = 0.749$ with coefficient estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ all statistically significant (p-values are below 10^{-106}). (We used the *fitlm.m* subroutine in MATLAB.) To improve the regression fit, "*one may wish to consider fitting higher-order models such as a second-order response surface to the output. Such a model allows one to explore* […] *two-factor interaction (cross-product) effects.*" (Santner et al., 2018, p. 252). The analyst then employs a cubic response surface of the type

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{1,1} X_1^2 + \beta_{2,2} X_2^2 + \beta_{1,2} X_1 X_2 \\ & + \beta_{1,1,1} X_1^3 + \beta_{2,2,2} X_2^3 + \beta_{1,1,2} X_1^2 X_2 + \beta_{1,2,2} X_1 X_2^2 + \varepsilon, \end{aligned} \tag{2}$$

adding several product terms, commonly called interaction terms. With the model in (2), R^2 increases from 0.762 to 0.896, and all interaction coefficients are statistically significant, indicating that interactions matter in the input–output response.

However, attributing these interactions to the presence of product terms is misleading. The original input–output mapping (see Appendix C) neither contains any product term nor any higher-order effects. Interactions are, instead, due to another mechanism.

The literature has unveiled *spurious* (Friedman & Popescu, 2008), *removable* (Berrington de González & Cox, 2007), *context-specific* (Højsgaard, 2004) interactions. These interactions are associated with different mathematical structures, which have, however, only been sparsely explored. At the same time, the computer experiments literature has formulated a variety of methods for interactions quantification (Liu & Owen, 2006), but is has not yet benefited from the theoretical progress registered in the broader statistical literature. The purpose of this work is to bridge this gap by offering a systematic characterization of interaction types, interaction generating mechanisms, and methods for determining interactions. We link interaction types to corresponding mathematical structures present in the input–output mapping. We focus on piecewise-definiteness as a self-standing interaction generating mechanism, showing that it generalizes the notion of context-specific interactions of (Højsgaard, 2004). In particular, it is a mechanism that differs from the presence of interaction terms: If interactions are generated by piecewise-definiteness, they can never be removed, while interactions due to the presence of product terms can be removed under certain conditions.

We then explore the identification of interaction generating mechanisms as a way toward interpreting the interactions hidden in a black-box model. Here, an important element is the scale at which the analysis of interactions is carried out. We address methods that explore interactions at the infinitesimal, finite, and global scales. While there is a path that links several interaction indicators at the alternative scales, the analysis reveals that there is no single method that can universally identify the true interaction generating mechanism. The analyst needs to resort simultaneously to methods that act on different scales to identify the correct interaction type and the

corresponding generating mechanism (Section 8). In particular, methods that combine a finite scale with a global scale provide several insights.

Indeed, reinterpreting the Sobol' pick-and-freeze design as the random sampling of finite change sensitivity indices allows one to obtain information about the region where interactions are most active, and about whether interactions are synergistic (positive) or antagonistic (negative) at no additional cost. We propose to visualize this information in a new tool called Mikado plot (Section 7).

The analysis becomes even more challenging when inputs are dependent. We investigate a combination of methods based on generalized global sensitivity indices that has the potential to help the analyst in obtaining the correct insights; however, we highlight that this is a first approach leaving further exploration as a line of further research. We conclude the work with numerical experiments on three well-known simulators, illustrating for each of them the alternative combinations of methods necessary to correctly interpret the nature of the involved interactions.

The remainder of the work is organized as follows. Section 2 proposes a review on the treatment of interactions in the statistical and computer experiment literature. Section 3 proposes formal definitions of interaction types and mechanisms. Section Section 4 discusses quantification indices for structural interactions on local and global scales. Section 5 addresses spurious interactions. Section 6 highlights the importance of the scale at which an interaction analysis is carried out and links interaction indicators on a finite scale to indicators on a global scale. Section 7 introduces Mikado plots. Section 8 reviews critically the findings. Section 9 concludes the work with numerical experiments.

2 | LITERATURE REVIEW ON STATISTICAL INTERACTION QUANTIFICATION

The analysis of interactions has a long tradition in Statistics. Interaction quantification emerges in the study of contingency tables (e.g., Højsgaard, 2003; Lauritzen, 2012), in the analysis of variance (ANOVA) of multifactor experiments (Gelman, 2005; Landsheer & van den Wittenboer, 2015) and in the design and analysis of computer experiments (Box et al., 2005; Lewis & Dean, 2001; Wang, 2007; Wu, 2015). A full coverage of the subject is outside the reach of this section, limiting the attention to works more closely related to ours. We refer to the monograph of Vanderweele (2015) or to the surveys of (Cox, 1984; Wu, 2015) for broader overviews. In particular, Vanderweele (2015) evidences that the correct interpretation of interactions is crucial for a proper statistical inference.

In Cox's interpretation (Cox, 1984), interactions are deviations from additivity. (We recall that, instead, in the Lancaster–Streitberg interpretation (Chakraborty & Zhang, 2019; Lancaster, 1971; Streitberg, 1990), interactions are studied in association with the presence/absence of statistical dependence among the random variables of interest.) Within this interpretation, one finds *spurious* (Friedman & Popescu, 2008), *removable* (Berrington de González & Cox, 2007), and *context-specific* interactions (Højsgaard, 2003; Højsgaard, 2004).

Spurious interactions are interactions generated by the statistical dependence of the covariates and are not necessarily the reflection of the presence of product terms in the input–output mapping (Friedman & Popescu, 2008; Oakley & O'Hagan, 2004). Context-specific interactions are associated with the presence of product terms between a group of covariates (or inputs of the computer simulation), where the inclusion/exclusion of these terms depends on the value assumed

by covariates outside the group (Højsgaard, 2004). Finally, removable interactions are interactions that can be eliminated by application of a monotonic transformation of the input–output mapping (Berrington de González & Cox, 2007).

The literature on simulation has developed several methods for studying interactions. Here, we recall that in computer experiments the response of the system is obtained running a computer simulator instead of measuring field data (see Santner et al., 2018; Wu, 2015) for a detailed discussion about the differences between computer and field experiments.). This opens the possibility of using differentiation techniques and a first natural class of interaction measures is represented by higher-order partial derivatives. Their estimation requires the input–output mapping to be smooth and can be carried out via the application of automatic differentiation (Griewank & Walther, 2008). A different setup however applies when the analyst is concerned with the response of the simulator across two (or more) scenarios of interest, and the scale of input variations is finite rather than infinitesimal (see Marangoni et al., 2017 for a recent example in climate modeling). As Kleijnen (2015) underlines, the analyst can then borrow directly from design of experiments, using factorial designs (full or with some lower resolution) or designs based on orthogonal arrays (Morris et al., 2008). A scenario-based approach remains a local approach, providing an indication of interactions at a limited number of locations in the model input space. Conversely, global approaches based on the functional ANOVA expansion (Efron & Stein, 1981) allow a more thorough exploration of the model input space. These methods have been intensively investigated in the computer experiment literature (Owen, 2003; Saltelli & Tarantola, 2002). Sensitivity measures that convey the importance of interactions are the total order variance-based sensitivity indices (Homma & Saltelli, 1996) and the superset importance measures (Liu & Owen, 2006). The statistical properties of the corresponding estimators, including consistency and asymptotic normality, are studied in the works of Janon et al. (2014) and Gamboa et al. (2016). Pairwise interactions are studied in depth in Fruth et al. (2014), and in Roustant et al. (2014) where variance-based indices are connected to second-order mixed derivatives. However, under dependence, difficulties in the interpretation of interactions emerge (Oakley & O'Hagan, 2004). One has to distinguish between spurious interactions (Friedman & Popescu, 2008) and contributions due to interactions actually associated with the structure of input–output mapping implied by the computer code. An analysis of interaction in this case might benefit from the approaches to the generalized functional ANOVA expansion developed in works such as (Chastaing et al., 2012, 2015; Hooker, 2004, 2007; Li et al., 2010; Li & Rabitz, 2012; Rahman, 2014).

However, while interactions have been studied in the computer experiments literature, investigations are generally limited to their quantification, when this is performed (on this aspect, see the critique of Saltelli et al., 2019). The statistical interpretation of interactions is not yet an established step in computer experiments. A plausible reason is the lack of a formal framework connecting the two research streams. We propose to fill in this gap in the remainder of this article, and start with a formal definition of interactions.

3 | FORMALIZING INTERACTION TYPES AND GENERATING MECHANISMS

In this section, we formalize the notions discussed qualitatively in the previous section, with the goal of obtaining a unifying framework that characterizes interaction types and links them to the underlying generating mechanism. Let $\mathcal{X} \subset \mathbb{R}^n$, with $g : \mathcal{X} \to \mathbb{R}$, $\mathbf{x} = (x_1, \dots, x_n) \mapsto y = g(\mathbf{x})$, a generic mapping. In this work, we shall focus on deterministic responses, that is, we restrict

attention to the case in which there is no stochastic error or the analyst is interested in the expected response of the code. The formal considerations for purely deterministic simulators can be applied to stochastic simulators if one considers the expected value of the response or another deterministic function of the output distribution. A stochastic simulator is usually written as $y = g(\mathbf{x}) + \epsilon(\mathbf{x})$. Then, consider replacing the deterministic model output with $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$. We have $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = g(\mathbf{x}^0) + \epsilon(\mathbf{x}^0)$. Letting $\tilde{g}(\mathbf{x}^0) = g(\mathbf{x}^0) + \epsilon(\mathbf{x}^0)$, when $\mathbf{x}$ shifts from $\mathbf{x}^0$ to $\mathbf{x}^1$, it is $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^1] = g(\mathbf{x}^1) + \epsilon(\mathbf{x}^1) - g(\mathbf{x}^0) - \epsilon(\mathbf{x}^0) = \Delta\tilde{g}$. Thus, the definition of interactions holding for $g(\mathbf{X})$ holds for $\mathbb{E}[Y|\mathbf{X}] = \tilde{g}(\mathbf{X})$. —

Following the tradition of the computer experiment literature, we regard interactions as deviations from additivity (see Koehler & Owen, 1996, p. 262), in the sense of Cox (1984).

Let $\mathbf{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ and $\mathbf{x}^1 = (x_1^1, x_2^1, \dots, x_n^1)$ be two points in $\mathcal{X}$ differing in at least two coordinates, serving as arbitrarily called "low" and "high" values in a two-level design. For notation simplicity, we consider a deterministic mapping. Also, let $[n] = \{1, 2, \dots, n\}$ and $i_k \in [n]$ denote a generic index, let $z = \{i_1, i_2, \dots, i_k\}, z \subset [n]$, denote a subset of indices and $-z = [n] \setminus z$ its complement. Let $\overline{\mathbf{x}}^z = (\mathbf{x}_z^1 : \mathbf{x}_{-z}^0)$ denote the point in $\mathcal{X}$ obtained by considering the variates with indices in z at level 1 and the remaining at level 0. Then $\Delta_z g = g(\overline{\mathbf{x}}^z) - g(\mathbf{x}^0)$ and $\Delta_{-z} g = g(\overline{\mathbf{x}}^{-z}) - g(\mathbf{x}^0)$ denote the changes in g due to the shift of variables with indices in z and without indices in z, respectively.

Definition 1. Let $g : \mathcal{X} \to \mathbb{R}$, and consider the points defined above. We say that g is additive on $\mathcal{X}$ if for all changes $\mathbf{x}^0 \to \mathbf{x}^1$ with $\mathbf{x}^0, \mathbf{x}^1 \in \mathcal{X}$

$$\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0) = \sum_{i=1}^{n} \Delta_i g, \tag{3}$$

where $\Delta_i g = g(\overline{\mathbf{x}}^i) - g(\mathbf{x}^0) = g(x_i^1 : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0)$.

That is, g is additive if the effect of the change $\mathbf{x}^0 \to \mathbf{x}^1$ on g if the sum of the individual changes provoked by each x_i for all changes $\mathbf{x}^0 \to \mathbf{x}^1$.

Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ and let $x_i \in \mathcal{X}_i$. Then consider n univariate mappings $a_i : \mathcal{X}_i \to \mathbb{R}$. In mathematical analysis, one calls separately additive a mapping that can be written as

$$g(\mathbf{x}) = \sum_{i=1}^{n} a_i(x_i), \tag{4}$$

for all $\mathbf{x} \in \mathcal{X}$. We have the following (see Appendix B for proofs).

Proposition 1. *A mapping $g : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} \subset \mathbb{R}^n$ satisfies (3) if and only if it satisfies (4).*

Thus, an analyst is dealing with an interaction whenever the input–output mapping is not the sum of univariate mappings as in (4).

Definition 2. We say that a mapping g presents structural interactions if $\Delta g \neq \sum_{i=1}^{n} \Delta_i g$ holds for some change $\mathbf{x}^0 \to \mathbf{x}^1$.

Indeed, the inclusion of interaction terms of the form $\beta_{i,j,\dots,k}(x_i x_j \dots x_k)$ in a regression model makes us shift from an additive to a nonadditive surface (see Example 1). However, structural interactions may not be due solely to the presence of interaction terms. Piecewise-defined mappings appear frequently as test cases in computer experiments (see the studies of Kim et al., 2005;

Liu & Owen, 2006; Römisch, 2013; Roustant et al., 2018). The next definition formalizes the notion of piecewise-defined function.

Definition 3 (Borgonovo & Peccati, 2010; Herrera, 2007). Consider a finite partition of the domain $\mathcal{X}$, $\Pi_{\mathcal{X}} = \{\mathcal{X}_1^{\Pi}, \mathcal{X}_2^{\Pi}, \dots, \mathcal{X}_L^{\Pi}\}$. A mapping g is piecewise-defined if it can be written as:

$$g(\mathbf{x}) = h_\ell(\mathbf{x}) \quad \text{if} \quad \mathbf{x} \in \mathcal{X}_\ell^{\Pi}, \quad \ell = 1, 2, \dots, L, \tag{5}$$

with $h_\ell(\mathbf{x}) \neq h_m(\mathbf{x})$ almost everywhere on $\mathcal{X}_\ell^{\Pi}$ or $\mathcal{X}_m^{\Pi}$ for all $\ell \neq m$, with $\ell, m = 1, 2, \dots, L$.

Example 2. Given the following mapping on $\mathcal{X} = [-1, 1]^2$:

$$g(x_1, x_2) = \begin{cases} \sin(x_1) + \sin(x_2) & \text{if} \quad -1 \leq x_1 < 0 \\ \sin(x_1) + \sin(2x_2) & \text{if} \quad 0 \leq x_1 \leq 1 \end{cases}, \tag{6}$$

and consider the change $\mathbf{x}^0 = \left(-\frac{1}{2}, -\frac{1}{2}\right) \rightarrow \mathbf{x}^1 = \left(\frac{3}{4}, \frac{3}{4}\right)$. One obtains

$$\Delta g = 1.2818, \quad \Delta_1 g = 0.6737, \quad \Delta_2 g = 0.4830, \tag{7}$$

so that $\Delta g \neq \Delta_1 g + \Delta_2 g$. Thus, g is not additive.

The mapping in Example 2 is locally the sum of univariate functions. However, even if $g = \sum_{i=1}^{n} a_i(x_i)$ at every point of the domain, the response is not additive. Thus, piecewise-definiteness is, by itself, an interaction generating mechanism and structural interactions can be due both to piecewise-definiteness and to the presence of interaction terms. The two mechanisms can act simultaneously. This is exactly what happens in the case of Højsgaard (2004)'s context-specific interactions.

Definition 4. Given a piecewise-defined function $g : \mathcal{X} \rightarrow \mathbb{R}, \mathcal{X} \subset \mathbb{R}^n$, we say that g contains a context-specific interaction if at least one of the restrictions $h_m : \mathcal{X}_m^{\Pi} \rightarrow \mathbb{R}$ of g onto $\mathcal{X}_m^{\Pi}$ is not additive.

Example 3. The mapping $g : [0, 1]^4 \rightarrow \mathbb{R}$,

$$g(x_1, x_2, x_3, x_4) = \begin{cases} x_1 + x_2 + x_3 & \text{if} \quad x_4 \leq 0.5 \\ x_1 + x_2 + x_3 + x_1 x_2 & \text{if} \quad x_4 > 0.5 \end{cases}, \tag{8}$$

displays a context-specific interaction: the interaction term $x_1 x_2$ is present only for $x_4 > 0.5$.

To understand whether these two interaction mechanisms (e.g., piecewise-definiteness and presence of interaction terms) are of the same or of a different nature, let us consider the notion of removable interactions. Berrington de González & Cox (2007, p. 374) call an interaction removable *if a transformation of the outcome scale can be found that induces additivity*. To make this statement formal, we introduce the following definition.

Definition 5. We say that $g : \mathcal{X} \rightarrow \mathbb{R}$, $\mathcal{X} \subset \mathbb{R}^n$, presents removable interactions on $\mathcal{X}$ if there exists a monotonic transformation $\eta(\cdot) : g(\mathcal{X}) \rightarrow \mathbb{R}$ such that the transformed function $z = \eta \circ g$, $z : \eta(g(\mathcal{X})) \rightarrow \mathbb{R}$ is additive.

We can link the above definition with a result obtained by Scheffé in his monograph on the analysis of variance (Scheffé, 1959, p. 95, (4.1.12)) and generalized in Scheffé (1970). Scheffé

presents a necessary and sufficient condition for the existence of a transformation that makes the mapping additive: for the existence of univariate functions $a_1(x_1), \dots, a_n(x_n)$ and of a monotonic transformation of the output $\eta(\cdot)$ with $\eta'(g) > 0$ such that $\eta(g(\mathbf{x})) = \sum_{i=1}^{n} a_i(x_i)$ one requires that there exists a function $\omega(g)$ such that

$$g''_{i,j} - \omega(g) g'_i g'_j = 0, \tag{9}$$

for all pairs $i < j$, where $\omega(\cdot)$ is an integrable function. Equation (9) is part of the family of quasi-linear elliptic differential equations

$$A(x_i, x_j, g'_i, g'_j) g''_{i,j} + B(x_i, x_j, g, g'_i, g'_j) = 0.$$

The family admits, in general, no closed form solution and it is therefore not possible to characterize in closed from the complete set of mappings whose interactions are removable.

If such a function exists then the univariate functions $a_i(x_i)$ can be represented as (Scheffé, 1970)

$$a_i(x_i) = \int \psi_i(x_i) dx_i + c_i, \tag{10}$$

where $\psi_i(x_i) = c \frac{\partial g}{\partial x_i} \exp\left(-\int \omega(g) dg\right)$ depends only on x_i, and c and c_i are constants, for all $i = 1, 2, \dots, n$. Because the condition in (9) is necessary, we can use it to exclude the existence of transformations to additivity for piecewise-defined functions.

Theorem 1. *Let $n > 2$, $g : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} \subset \mathbb{R}^n$ be a piecewise-defined function. Then, there does not exist a monotone transformation $\eta(\cdot)$ such that $\eta \circ g$ is additive.*

Hence, interactions due to piecewise-definiteness are never removable. Thus, presence of interaction terms and piecewise-definiteness are distinct interaction generating mechanisms.

Example 4. The mappings $s = e^{x_1 + x_2}$ on $\mathcal{X} = \mathbb{R}^2$, and $t = \sin(x_1 + x_2)$ on $\mathcal{X} = [0, \pi] \times [0, \pi]$ are not additive. However, they satisfy Scheffé's conditions (9) with $\omega(s) = \frac{1}{s}$ and $\omega(t) = \frac{1}{1-t^2}$, respectively. Thus, they present removable interactions, with obvious transformations.

These observations imply that, we are reassured that interactions are due to the presence of interaction terms if g satisfies (9), and that they can be removed via a suitable transformation.

4 | QUANTIFYING STRUCTURAL INTERACTIONS: ANALYSES AT DIFFERENT SCALES

A crucial role toward understanding the interaction generating mechanism is played by the choice of the interaction quantification method. The literature proposes indicators that act on alternative scales. We analyze several indicators in this section, starting with indicators at a finite scale and moving to indicators at a global scale. In the analysis, the functional ANOVA decomposition plays an important role in identifying structural interactions. We discuss its relevance first for an analysis of interactions at the finite scale, Section 4.1, and then at a global scale, Section 4.2.

4.1 | Structural interactions at a finite scale

Definition 1 considers the variation of an input–output mapping across two points in the input space, $\mathbf{x}^0 \to \mathbf{x}^1$, which results in the change in model output $\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0)$. It has been shown that this change can be decomposed in the following finite-change ANOVA expansion (Borgonovo, 2010; Kuo et al., 2010; Rabitz & Alis, 1999):

$$\Delta g = \sum_{z \subset [n]} \tau_z^{0\to 1}, \tag{11}$$

where the finite-change effects $\tau_z^{0\to 1}$ are defined recursively as

$$\tau_z^{0\to 1} = \Delta_z g - \sum_{u \subset z, u \neq z} \tau_u^{0\to 1}, \tag{12}$$

with $\Delta_z g = g(\overline{\mathbf{x}}^z) - g(\mathbf{x}^0)$. The quantity $\tau_z^{0\to 1}$ represents the contribution to Δg of the residual interaction among the group of indices in z. Equation (11) is called cut-ANOVA (Rabitz & Alis, 1999) or anchored-ANOVA decomposition (Kuo et al., 2010).

For each index group z the finite change ANOVA (11) can be split into

$$\Delta g = \sum_{v \cap z \neq \emptyset} \tau_v^{0\to 1} + \sum_{v \subset -z} \tau_v^{0\to 1} = \overline{\tau}_z^{0\to 1} + \underline{\tau}_{-z}^{0\to 1}, \tag{13}$$

where

$$\overline{\tau}_z^{0\to 1} = \sum_{v \cap z \neq \emptyset} \tau_v^{0\to 1} \text{ and } \underline{\tau}_z^{0\to 1} = \sum_{v \subset z} \tau_v^{0\to 1},$$

are, respectively, the total and the subset finite change indices of z. For the case of individual factors, $z = \{i\}$, we define $\overline{\tau}_i^{0\to 1} = \overline{\tau}_{\{i\}}^{0\to 1}$, the total effect of factor i, $\underline{\tau}_i^{0\to 1}$ the individual effect of factor i and

$$\Upsilon_i^{0\to 1} = \sum_{z \ni i, z \neq \{i\}} \tau_z^{0\to 1}, \tag{14}$$

the interaction effect of factor i. Note that $\Upsilon_i^{0\to 1}$ equals the difference between the total and the finite change effect of $\{i\}$. These identities are the finite-change equivalents of corresponding ones obtained for the decomposition of the variance of g (see Liu & Owen, 2006; see also Section 4.2 for further details).

The link between Definition 1 and the finite-change indices defined above is immediate. Let $\{i\}$ be an individual index, and $\mathcal{Z}_2$ be the collection of all groups of indices z that contain two or more elements. If g is additive then $\tau_z^{0\to 1} = 0$ for all $z \in \mathcal{Z}_2$, $\Upsilon_i^{0\to 1} = \tau_i^{0\to 1}$; conversely, if structural interactions are present in g, then $\tau_z^{0\to 1} \neq 0$ for some z and $\Upsilon_i^{0\to 1} \neq \tau_i^{0\to 1}$.

Example 5. Consider the change in $g = \frac{x_1}{x_1 + x_2}$, as $\mathbf{x}$ varies from $\mathbf{x}^0 = (1, 1)$ to $\mathbf{x}^1 = (2, 5)$. One registers $\Delta g = -\frac{3}{14}$, $\Delta g = \tau_1^{0\to 1} + \tau_2^{0\to 1} + \tau_{1,2}^{0\to 1}$, with $\tau_1^{0\to 1} = \frac{1}{6}$, $\tau_2^{0\to 1} = -\frac{1}{3}$, and $\tau_{1,2}^{0\to 1} = -\frac{1}{21}$, so that $\Upsilon_i^{0\to 1} = \left(-\frac{1}{21}, -\frac{1}{21}\right)$, $\overline{\tau}_i^{0\to 1} = \left(\frac{5}{42}, -\frac{8}{21}\right)$.

We note that in design of experiments a second-order interaction effect is typically written as $A_{i,j}(\mathbf{x}^0, \mathbf{x}^1) = \frac{1}{2}[g(\overline{\mathbf{x}}^{i,j}) + g(\mathbf{x}^0) - g(\overline{\mathbf{x}}^i) - g(\overline{\mathbf{x}}^j)]$ (Wu, 2015, equation (3.1), p. 614). Thus, we have $\tau_{i,j}^{0\to 1} = 2A_{i,j}(\mathbf{x}^0, \mathbf{x}^1)$. Similar relationships can be obtained for higher-order effects. In terms of interaction interpretation, the finite-change effects $\tau_z^{0\to 1}, z \in \mathcal{Z}_2$ and $\Upsilon_i^{0\to 1}$ are indices that deliver information about the presence of structural interactions.

Calculating all the terms of the expansion in (11) requires to evaluate g on a full factorial design. In Appendix A, we connect this design to the determination of the discrete Laplace operator on all vertices of the hypercube with vertices $\mathbf{x^0}$ and $\mathbf{x^1}$. Because this hypercube has 2^n vertices, determining the complete orthogonalized decomposition of a finite change is a computationally intensive task. For instance, for $n = 20$, 2^n is larger than 10^6. However, the literature has introduced a computational shortcut that allows us to determine the triplets $\tau_i^{0\to 1}$, $\Upsilon_i^{0\to 1}$ and $\overline{\tau}_i^{0\to 1}$ at a linear instead of an exponential cost in n. One has

$$\tau_i^{1\to 0} = \Delta_{-i}g - \Delta g = \underline{\tau}_{-i}^{0\to 1} - \left(\underline{\tau}_{-i}^{0\to 1} + \overline{\tau}_i^{0\to 1}\right) = -\overline{\tau}_i^{0\to 1}, \tag{15}$$

which is the finite scale equivalent of the so-called pick-and-freeze design (Borgonovo, 2010; Gamboa et al., 2016). Thus, the computational cost for computing $\tau_i^{0\to 1}$, $\Upsilon_i^{0\to 1}$ and $\overline{\tau}_i^{0\to 1}$ is $2n + 2$ evaluations of g instead of 2^n. Moreover, we have another computational shortcut. If g is additive, then only two function evaluations are necessary to compute the discrete Laplace operator (see Proposition 3 in Appendix A).

4.2 | Structural interactions on a global scale

In the analysis of computer experiments, variance-based approaches that rely on Hoeffding's ANOVA decomposition of the model output variance are frequently used. These approaches yield information on interactions at a global scale, providing complementary and alternative insights to the finite-change indices that we have analyzed in the previous section.

The functional ANOVA expansion plays a fundamental role in the study of interactions on a global scale (Durrande et al., 2013; Liu & Owen, 2006; Saltelli et al., 2000). One regards the simulator inputs and output as a pair of a n dimensional random vector and a random variable $(\mathbf{X}, Y)$ on reference probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$. Let $F_{\mathbf{X}}$ denote the joint cumulative distribution functions of the inputs. Assuming that $F_{\mathbf{X}} = \prod_{i=1}^n dF_i$ and that $g \in \mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_{\mathbf{X}})$, the seminal results of (Efron & Stein, 1981; Hoeffding, 1948) allow us to write:

$$\mathbb{V}[Y] = \sum_{z\subset[n], z\neq\emptyset} V_z, \tag{16}$$

where $\mathbb{V}[Y]$ is the model output variance,

$$V_z = \int \left[g_z(\mathbf{x}^z)\right]^2 dF_z(\mathbf{x}^z) \text{ and } g_z(\mathbf{x}^z) = \int g(\mathbf{x}^z : \mathbf{x}^{-z}) dF_{-z}(\mathbf{x}^{-z}) - \sum_{v\subset z, v\neq z} g_v(\mathbf{x}^v). \tag{17}$$

In (16), the generic term V_z represents the residual contribution to $\mathbb{V}[Y]$ of the interaction among the simulator inputs whose indices are in z. When $(z = \{i\})$, the term V_i represents the individual contribution of X_i to $\mathbb{V}[Y]$. Liu and Owen (2006) define the sensitivity indices

$$\underline{\tau}_u^2 = \sum_{z \subset u} V_z \text{ and } \overline{\tau}_u^2 = \sum_{z \cap u \neq \emptyset} V_z. \tag{18}$$

The index $\overline{\tau}_u^2$ represents the total contribution to the variance of g of the simulator inputs with indices in u (see also Owen, 2014, p. 247). Hooker (2004) and Liu & Owen (2006) define the superset importance index of a group of simulator inputs u as

$$\Upsilon_u = \sum_{z \supset u} V_z. \tag{19}$$

The quantity Υ_u measures the contribution of all terms that include the indices in u in the ANOVA decomposition of g. The link between the functional ANOVA decomposition and Definition 1 is straightforward and formally similar to the one obtained for finite-change effects. Specifically, we have that $\overline{\tau}_i^2 = V_i$ and $\Upsilon_z = 0$ for all $z \in \mathcal{Z}_2$. This is illustrated in the next example.

Example 6. Consider the following input–output mappings: $Y_a = (X_1 - 1)X_2 + X_1$ and $Y_b = X_1 + X_2$, with X_1 and X_2 independently and uniformly distributed on $[-2, 2]$. For the first mapping, we find $\mathbb{V}[Y_a] = 4.44$, the ANOVA decomposition $V_1 = V_2 = 1.333$ and $V_{1,2} = 1.78$ that yields the following total and interaction indices $\overline{\tau}_1^2 = \overline{\tau}_2^2 = 3.11$ and $\Upsilon_{1,2} = V_{1,2} = 1.78$. For the second mapping, we have $\mathbb{V}[Y_b] = 2.67$, $V_1 = V_2 = 1.33$, and $V_{1,2} = 0$, so that $\overline{\tau}_1^2 = \overline{\tau}_2^2 = 1.33$ and $\Upsilon_{1,2} = V_{1,2} = 0$.

Further quantities that help to elegantly characterize the presence of interactions are the notions of dimension distribution and effective dimension introduced by Caflisch et al. (1997) and Owen (2003). The ratios $S_z = V_z/\mathbb{V}[Y]$ are called Sobol' indices (Durrande et al., 2013; Sobol', 1993) and can be regarded as the probability mass function of a random variable, say T, with support $2^{[n]}$, and such that $\mathbb{P}(T = z) = S_z$. The dimension distributions of g in the superimposition sense and truncation sense are then defined as the distribution of the cardinality of T and of $\max\{i : i \in z\}$, respectively. We shall restrict attention to the mean dimension in the superimposition sense (mean dimension, henceforth). Owen (2013) proves that the mean dimension (D_g) satisfies $D_g = \sum_z |z| S_z = \mathbb{V}[Y]^{-1} \sum_{i=1}^n \overline{\tau}_i^2$, and that for any additive square-integrable function $D_g = 1$. As a consequence, $D_g > 1$ implies that some nonnull interaction term V_z with $|z| > 1$ is present in the decomposition of $\mathbb{V}[Y]$.

Example 7. Consider the piecewise constant simulators $g_1 = \mathbb{I}_{E_1 \cup \cdots \cup E_n}$ and $g_2 = \mathbb{I}_{E_1} \cdots \mathbb{I}_{E_n}$, where E_i denotes the event $\{x_i \geq 0.5\}, i = 1, 2, \ldots, n$, $n \geq 2$. Provided with random inputs $X_i \sim U[0, 1]$ independent and identically distributed (i.i.d.), both simulators satisfy $S_z = \frac{1}{2^n - 1}$ for all multi-indices z. They have the same mean dimension in the superimposition sense equal to

$$D_g = \sum_z |z| S_z = \frac{\sum_{j=1}^n \binom{n}{j} j}{\sum_{j=1}^n \binom{n}{j}} = \frac{n \cdot 2^{n-1}}{2^n - 1}. \tag{20}$$

The mean dimension is greater than unity signaling the presence of interactions and increases with n. For instance, D_g equals 1.33, 1.71, and 2.13 for $n = 2, 3$, and 4, respectively; the value is well approximated by $\frac{n}{2}$ for large n.

One says that g has mean dimension s if $\sum_{|z| \leq s} S_z \geq 0.99$ (Caflisch et al., 1997; Owen, 2003). This means that if $s = 1$ then g has mean dimension 1 if we register $\sum_{i=1}^n S_i \geq 0.99$. Consider the

function in Example 5, with $X_1, X_2 \sim Gamma(10, 1)$ i.i.d.. One registers $S_1 + S_2 = 0.999 > 0.99$; however, g is not additive in this example. That is, a numerical value D_g close to unity does not necessarily mean that the input–output mapping is additive. Thus, an indication of additivity on a global scale needs to be carefully interpreted. In fact, while additivity on a finite scale implies additivity on a global scale, the above example shows that the converse is not necessarily true.

With the goal of offering a further link with established results on interactions in the broad statistical literature, we conclude this section formalizing the link between the global interaction indices with the discussion in Cox (1984) and Berrington de González and Cox (2007) about constancy of variance and absence of interactions. The motivation comes from the study of Coveney and Clayton (2019), who write in discussing their model for human atrial cardiac cells that fixing some inputs leads unexpectedly to an increase of the output uncertainty and hypothesize that this might be due to the presence of interactions. To analyse this fact, we start proposing a link between the classical ANOVA expansion and the statistical notion of heteroskedasticity.

Definition 6. A random variable Y is called homoskedastic with respect to a random variable X_i if the conditional variance $\mathbb{V}[Y|X_i]$ is constant, or heteroskedastic otherwise.

Proposition 2. *An input–output mapping g is heteroskedastic with respect to X_i if and only if there are nonvanishing interaction terms involving X_i in the ANOVA decomposition of $\mathbb{V}[Y]$.*

A sufficient condition for for this result is the following.

Corollary 1. *If there exits a subset $S \subseteq \mathcal{X}_i$ in the support of X_i with $\mathbb{P}_{\mathbf{X}}(S) \neq 0$ for which $\mathbb{V}[Y|X_i = x_i] > \mathbb{V}[Y]$ for almost all $x_i \in S$, then Y is heteroskedastic with respect to X_i.*

Example 8. Consider $g : \mathbb{R}^2 \to \mathbb{R}, (X_1, X_2) \mapsto e^{X_1}|\sin(X_2)|$. With X_1 standard normal and X_2 normal with mean and SD equal to 1, the simulator variance is $\mathbb{V}[Y] = 2.72$, while $\mathbb{V}[Y|X_2 = 1] = 3.31$.

Corollary 1 implies that an increase in variance after receiving perfect information on factor X_i can occur only in the presence of interactions. The above result provides the formal framework for the interpretation of Coveney and Clayton (2019).

5 | UNDERSTANDING THE INTERACTION TYPE: SPURIOUS INTERACTIONS

The interaction quantification measures we have discussed in the previous section detect structural interactions. That is, if one of those interaction indicators is different from zero then we know that interactions are either due to the presence of interaction terms or to the presence of discontinuities. However, when model inputs are dependent, interactions may emerge even if they are not present in the input–output mapping. To illustrate, Oakley and O'Hagan (2004) consider the function $g = x_1$ with a joint density $f_{X_1,X_2}(x_1, x_2)$. Applying (17), one finds the functional ANOVA terms (see Oakley and O'Hagan (2004), p. 753): $g_0 = \mathbb{E}[g] = \mathbb{E}[X_1]$, $g_1(X_1) = X_1 - \mathbb{E}[X_1]$, $g_2(X_2) = \mathbb{E}[X_1|X_2] - \mathbb{E}[X_1]$ and $g_{1,2}(X_1, X_2) = -g_2(X_2)$. If X_1 and X_2 are not independent, the interaction term $g_{1,2}(X_1, X_2)$ is neither null nor a structural part of g. Friedman and Popescu (2008) deem this a *spurious interaction*. In general, input dependence generates issues with the construction of the functional ANOVA expansion itself. The problem has been addressed in a series of

TABLE 1 Correlative and structural terms of the variance decomposition for the example

ρ	$\mathbb{V}[Y]$	V_1	V_1^c	V_2	V_2^c	$V_{1,2}$	$V_{1,2}^c$
0.7	21.4	4	12.81	9	10.36	0	−14.77
0	13	4	0	9	0	0	0
−0.7	4.6	4	−3.99	9	−6.44	0	2.03

works by Hooker (2007), Li and Rabitz (2012), Chastaing et al. (2012), Rahman (2014), Chastaing et al. (2015), that show that under dependence we can obtain a decomposition that satisfies

$$\mathbb{V}[Y] = \sum_{\emptyset \neq z \subset Z} \left[V_z + Cov\left(g_z, \sum_{\emptyset \neq v \subset Z, v \neq z} g_v \right) \right]. \tag{21}$$

Note that, when the inputs are independent, this formula reduces to Equation (16). By normalizing the covariance decomposition in (21), one obtains (Li & Rabitz, 2012)

$$\mathbb{V}[Y]^{-1} \sum_{\emptyset \neq z \subset Z} \left[V_z + Cov\left(g_z, \sum_{\emptyset \neq v \subset Z, v \neq z} g_v \right) \right] = \sum_{\emptyset \neq z \subset Z} \left[S_z + S_z^c \right] = 1. \tag{22}$$

In (22) the variance-based index S_z is the marginal contribution of the indices in z, while the index S_z^c reflects the contribution due to correlations. Hence, while S_z reflects the structural contribution of inputs in z, S_z^c is an indicator of the relevance of spurious interactions. For further understanding the meaning of spurious interactions, we recall the work of Lu et al. (2018), who provide several analytical expressions for Sobol' indices when the input output mapping is linear or generalized linear model. For simplicity, suppose that the input–output mapping can be represented as $Y = \mathbf{X}\beta$. If $\mathbf{X}$ is multivariate normal with mean $\mu \in \mathbb{R}^n$ and variance–covariance matrix Σ, Lu et al. (2018) show that for input $i = 1, 2, \ldots, n$, we have

$$\mathbb{V}[\mathbb{E}[Y|X_i]] = \left(\beta_i \sigma_i + \sum_{j \neq i} \beta_j \rho_{j,i} \sigma_j \right)^2. \tag{23}$$

Considering Equation (22), it is readily seen that, for a linear model with normally distributed inputs, we have the structural and correlative indices given by

$$V_i = \beta_i^2 \sigma_i^2 \quad \text{and} \quad V_i^c = \left(\sum_{j \neq i} \beta_j \rho_{ji} \sigma_j \right)^2 + 2\beta_i \sigma_i \sum_{j \neq i} \beta_j \rho_{ji} \sigma_j, \tag{24}$$

respectively. In Lu et al. (2018), one can find analytical expressions of the Sobol' indices also for log and logit link functions when the inputs follow a multivariate normal distribution.

Example 9. We revisit the example of Xu and Gertner (2007). They consider the linear model $Y = 2X_1 + 3X_2$ where the inputs have a standard multivariate normal distribution with correlation coefficient ρ. As in Xu and Gertner (2007), we set $\rho = 0.7$ and $\rho = -0.7$. They compute the variance-based contributions for X_1 and X_2, obtaining for the case $\rho = 0.7$: $\mathbb{V}[\mathbb{E}(Y|X_1)] = 16.81$, and $\mathbb{V}[\mathbb{E}(Y|X_2)] = 19.36$. We can obtain further insights on this result using (24). Direct

application yields the structural and correlative indices shown in Table 1. Table 1 reports the variance-based contributions also for the uncorrelated case $\rho = 0$. Consider X_1: We see that the sum of the structural $V_1 = 4$ and correlative contributions $V_1^c = 12.81$ equals $\mathbb{V}[\mathbb{E}(Y|X_1)] = 16.81$. The same holds for X_2. Table 1 also shows that the structural interaction effect is null for all values of ρ, as a consequence of the input–output additivity. This provides the right insights that interactions for this model are spurious, that is, due to the presence of correlations. In this respect, the correlative interaction effects change sign as ρ changes sign, signaling that the correlative interaction contributions (and the individual as well in this case) are highly affected by the type of correlation assumed by the analyst. In contrast, the values of the structural contributions remain unaffected by correlations.

However, in general no closed form is available for the determination of the structural and correlative contributions. Then, a numerical approach is needed. A successful approach is the D-MORPH regression of Li and Rabitz (2010). The method combines ridge regression with the projection of the functional ANOVA effects $g_z(\mathbf{x}_z)$ onto an appropriate basis formed by orthonormal polynomials. Fitting the resulting response surface on a dataset of input–output realizations of $\mathbf{X}$ and Y allows the determination of coefficients through which the structural and correlative variance contributions can be estimated—see (Li & Rabitz, 2012, 2017) for additional details. An interesting avenue of future research is the determination of a systematic approach that combines D-MORPH regression and standard statistical methods such as linear regression for allowing the analyst to distinguish spurious from structural interactions in applications in which g is known through a simulator or through a statistical model fitted on observed data. The authors have performed some preliminary experiments on a well-known dataset and the combination of D-MORPH plus linear regression seems promising. However, further analysis is needed and due to space limitations, we leave this important aspect as part of future research.

6 | LOCAL INTERACTIONS: MIND THE STEP

In this section, we study interaction indicators at the infinitesimal scale, to understand further their insights. We start considering the case of smooth input-output mappings, for which partial derivatives exist everywhere in the domain. In fact, a widely used approach for measuring interactions in computer experiment studies is by calculating mixed higher order derivatives (see, among others, Roustant et al., 2014). Specifically, the expression $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ is used to denote the absence of interactions (Friedman & Popescu, 2008). In the next result, we provide formal conditions for this intuition to hold.

Theorem 2. *Let $\mathcal{X} \subset \mathbb{R}^n$ and let $g : \mathcal{X} \to \mathbb{R}$. If g is additive and has second-order mixed derivatives everywhere in $\mathcal{X}$ then $g''_{i,j}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$. Conversely, if $g''_{i,j}(\mathbf{x}) = 0$ for all $x \in \mathcal{X}$ then g is additive.*

This is in accordance with the results of Roustant et al. (2014), who investigate the connection between partial derivatives and Sobol' indices assuming that g is twice differentiable everywhere and that all of its first-order and second-order partial derivatives are in $\mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_{\mathbf{X}})$. To proceed in this way, however, the assumption that g is smooth plays a crucial role, as the next (counter-)example shows.

Example 10 (Example 7 continued). Consider the two functions in Example 7, for the case $n \geq 2$. For these functions, we have $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ for all i, j. However, Equation (20) shows

that the mean dimension D_g increases with n, signaling a growing relevance of interactions as dimensionality increases.

Example 10 shows that if interactions are due to piecewise-definiteness then, even if g is almost everywhere differentiable (but not everywhere), the condition $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ fails, that is, we register $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ but we cannot conclude that the mapping is additive. This implies that a high-order differential sensitivity approach is not reliable in the presence of piecewise interactions. Let us then analyze whether, acting on a different scale, one obtains different insights. In this respect, an intermediate step between a local and a global sensitivity exercise is the calculation of interactions at different scales. Let us study whether this approach can help in revealing the presence of interactions in mappings of the type of Example 10. In particular, as interaction indicators, let us consider the finite-scale interaction indices $\tau_{\{u\}}^{\mathbf{x}^{(k)}\to\mathbf{x}^{(k+1)}}$.

Example 11 (Example 7 continued). Consider the two functions in Example 7, with $n = 5$. Evaluating the functions at $\mathbf{x}^0 = [0.4, 0.4, 0.4, 0.4, 0.4]$ and $\mathbf{x}^1 = [0.6, 0.6, 0.6, 0.6, 0.6]$, for the OR-type function $g_1 = \mathbb{I}_{E_1\cup\cdots\cup E_n}$, we find $\tau_{\{i,j\}}^{\mathbf{x}^0\to\mathbf{x}^1} = -1$, for $i, j = 1, 2, \ldots, 5, i \neq j$, signalling the presence of interactions already at the second order. For the AND-type function $g_2 = \mathbb{I}_{E_1}\cdots\mathbb{I}_{E_n}$, we find a unique nonnull interaction effect $\tau_{\{1,2,3,4,5\}}^{\mathbf{x}^0\to\mathbf{x}^1} = 1$. Then, by Definition 2, the two mappings present structural interactions.

The above example shows that shifting from an infinitesimal to a finite scale allows the analyst to correctly infer that the two input–output mappings of Example 7 are not additive, in spite of the fact that $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$. A limitation of the above analysis resides in the fact that the result depends on the selected points: Had the analyst evaluated the models at $\mathbf{x}^0 = [0.3, 0.3, 0.3, 0.3, 0.3]$ and $\mathbf{x}^1 = [0.4, 0.4, 0.4, 0.4, 0.4]$, she would have spotted no interaction effects.

To overcome this limitation, a further step is to move from a finite, but local, to a global scale. The link is given by sampling N realizations of $\mathbf{X}$ from the input distribution through a Monte Carlo or Quasi-Monte Carlo random number generator. Then, let $\mathbf{x}^k$ and $\mathbf{x}^{k+1}$ denote two realizations of $\mathbf{X}$ and $\tau_u^{\mathbf{x}^k\to\mathbf{x}^{k+1}}$ denote a corresponding finite change interaction index. Potentially, this procedure makes available to the analyst $N-1$ replicates of $\tau_u^{\mathbf{x}^k\to\mathbf{x}^{k+1}}$. These effects yield rich insights about the magnitude of interactions, their sign and at which location the highest interactions occur. Not only, but this randomization also makes it possible to estimate global sensitivity indices. Starting from the cut-ANOVA in (11) and averaging, we can write

$$g_z(\mathbf{x}^z) = \mathbb{E}\left[\tau_z^{\mathbf{X}^0\to\mathbf{x}}\right] = \int\left(g(\mathbf{x}_z : \mathbf{x}^0_{\sim z}) - \sum_{u\subset z, u\neq z}\tau_u^{\mathbf{x}^0\to\mathbf{x}} - g(\mathbf{x}^0)\right)dF_{\mathbf{X}}(\mathbf{x}^0), \tag{25}$$

that is, every functional ANOVA term can be obtained by averaging over the random cut-point $\mathbf{X}^0$ (Rabitz & Alis, 1999). Considering the variance of both sides of (25), we find for all $\mathbf{x} \in \mathcal{X}$

$$V_z = \mathbb{V}\left[g_z(\mathbf{X}^z)\right] = \mathbb{V}_z\left[\mathbb{E}\left[\tau_z^{\mathbf{X}^0\to\mathbf{x}}\right]\right]. \tag{26}$$

Equation (26) shows that every interaction term in the variance-decomposition contains as much information on interactions as the mean of finite-change interaction indices. Moreover, we can directly link finite change interaction indices and superset importance measures.

Theorem 3. *Let $g \in \mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), F_{\mathbf{X}})$ and let $\mathbf{X}, \mathbf{Z} \sim U[0,1]^n$ be independent and identically distributed. Then, for $u \subset [n]$, we have:*

$$\Upsilon_u = \frac{1}{2^{|u|}}\mathbb{E}\left[\tau_u^{\mathbf{X}\to\mathbf{Z}}\right]^2, \quad \underline{\tau}_u^2 = \mathbb{E}\left[g(\mathbf{X})\underline{\tau}_u^{\mathbf{Z}\to\mathbf{X}}\right], \quad \overline{\tau}_u^2 = \frac{1}{2}\mathbb{E}\left[\underline{\tau}_u^{\mathbf{X}\to\mathbf{Z}}\right]^2, \tag{27}$$

where $|u|$ denotes the cardinality of u. Moreover, we have

$$\sum_{i=1}^{n} V_i = \sum_{i=1}^{n}\mathbb{E}\left[g(\mathbf{X})\tau_i^{\mathbf{Z}\to\mathbf{X}}\right] \quad \text{and} \quad \sum_u |u| V_u = \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\tau_i^{\mathbf{X}\to\mathbf{Z}}\right]^2. \tag{28}$$

The next results presents the corresponding Monte Carlo estimates.

Corollary 2. *The global interaction indicators (27) and (28) can be estimated from*

$$\widehat{\Upsilon}_u = \frac{1}{2^{|u|}N}\sum_{k=1}^{N}\left[\tau_u^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}}\right]^2, \ \widehat{\underline{\tau}_u^2} = \frac{1}{N}\sum_{k=1}^{N} g(\mathbf{x}^{(k)})\underline{\tau}_u^{\mathbf{z}^{(k)}\to\mathbf{x}^{(k)}}, \ \widehat{\overline{\tau}_u^2} = \frac{1}{2N}\sum_{k=1}^{N}\left[\underline{\tau}_u^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}}\right]^2. \tag{29}$$

We also have

$$\sum_{i=1}^{n}\widehat{V}_i = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{n} g(\mathbf{x}^{(k)})\tau_i^{\mathbf{z}^{(k)}\to\mathbf{x}^{(k)}} \quad \text{and} \quad \sum_u |u|\widehat{V}_u = \frac{1}{2N}\sum_{k=1}^{N}\sum_{i=1}^{n}\left[\tau_i^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}}\right]^2. \tag{30}$$

With $u = \{i,j\}$ we have

$$\widehat{\Upsilon}_{i,j} = \frac{1}{4N}\sum_{k=1}^{N}\left[\tau_{i,j}^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}}\right]^2 = \frac{1}{N}\sum_{k=1}^{N}\left[A_{i,j}^{(k)}\right]^2 = \frac{1}{N}\sum_{k=1}^{N}[EI^{(k)}]^2, \tag{31}$$

where EI is the screening interaction effect introduced in (Campolongo et al., 2011).

The above results show that finite change effects give rise to alternative estimators of global sensitivity indices. Precisely, if we consider the indices $\tau_u^{\mathbf{x}^0\to\mathbf{x}^1}$ of the orthogonal decomposition of Δg, we recover Υ_u of Liu and Owen (2006). If we consider $\underline{\tau}_u^{\mathbf{x}^0\to\mathbf{x}^1}$, we obtain estimators for $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$. Moreover, when $u = \{i\}$, the estimator $\widehat{\overline{\tau}_i^2}$ in (29) is the Jansen estimator of the total effect of the ith factor (Campolongo et al., 2011). When $u = \{i,j\}$, then we find back the Monte Carlo estimator $\widehat{\Upsilon}_{i,j}$ in (29) studied in Fruth et al. (2014). The consistency of these estimators follows from the strong law of large numbers (see, among others, Janon et al., 2014). Also, it is easy to see that estimators of (27)–(28) in Theorem 3 coincide with pick-and-freeze estimators of Gamboa et al. (2016). This connection has the implication that if analysts implement a pick-and-freeze design and record the sign of the effects before squaring and averaging, they have available the signs of interactions at multiple locations of the simulator input space. This fact can be used to obtain a new visualization tool that we discuss in the next section.

7 | THE MIKADO PLOT

The link between finite change and global measures implied by Theorem 3 and Corollary 2 can be exploited to enrich information regarding interactions. Specifically, consider the pick-and-freeze

estimator in (31). This estimator is associated with replicates of the second-order interaction index,

$$\tau_{i,j}^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}} = g\left(z_i^{(k)}, z_j^{(k)}; \mathbf{x}_{\sim i,j}^{(k)}\right) - g\left(z_i^{(k)}; \mathbf{x}_{\sim i}^{(k)}\right) - g\left(z_j^{(k)}; \mathbf{x}_{\sim j}^{(k)}\right) + g\left(\mathbf{x}^{(k)}\right). \tag{32}$$

Considering $2N$ locations $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$ and $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(N)}$ the evaluation of $\mathbb{E}[\tau_{i,j}^{\mathbf{X}\to\mathbf{Z}}]$, requires $N \times (1 + n + \frac{n}{2}(n-1))$ model runs. We can visualize the information in the following format: A Mikado plot is a collection of segments whose extremes are the two-dimensional projections on the $X_i - X_j$ plane of the pair of points $\mathbf{x}^{(k)}$ and $\mathbf{z}^{(k)}$ used in (32) to compute the second-order indices. Thus, the position of a segment indicates the two points on the $X_i - X_j$ plane across which the finite change interaction is registered. The color of the segment is used to represent the sign of the interaction. Such sign is delivered by the second-order Newton ratio

$$\nu_{i,j}^{(k)} = \frac{\tau_{i,j}^{\mathbf{x}^{(k)}\to\mathbf{z}^{(k)}}}{\left(z_i^{(k)} - x_i^{(k)}\right)\left(z_j^{(k)} - x_j^{(k)}\right)}, \tag{33}$$

which is found by a simple renormalization of Equation (32). The goal is to visualize simultaneously the regions of the $X_i - X_j$ plane that contribute the most to interactions, as well as to display the sign of interactions: By the link with the pick-and-freeze design (Theorem 3), these are also the regions that contribute the most to global interactions. The pseudocode, presented in Algorithm 1, computes the quantities of Corollary 2 and produces a Mikado plot. In the algorithm, the sum

Algorithm 1. Pseudocode for Mikado plots, N sample size, $g(\cdot)$ simulation model, L number of lines.

1. Create two independent input samples x^A and x^B of size $N \times d$
2. Simulate outputs $y^A = g(x^A)$, $y^B = g(x^B)$
3. **for** each i in $1 \ldots n$:
4. Create mixture sample x^i by inserting the ith column from x^B into x^A
5. Simulate output $y^i = g(x^i)$
6. Set $\Delta y = y^i - y^A$
7. Estimate $\underline{\widehat{\tau}}_i^2 = \frac{1}{n}\left\langle y^B, \Delta y\right\rangle$ and $\widehat{\widetilde{\tau}}_i^2 = \frac{1}{2n}\langle \Delta y, \Delta y\rangle$
8. **for** each j in $1 \ldots n$ with $j < i$:
9. Create mixture sample x^{ij} by inserting the jth column from x^B into x^i
10. Simulate output $y^{ij} = g(x^{ij})$
11. Set $\Delta^2 y = y^{ij} - y^i - y^j + y^A = y^{ij} - y^j - \Delta y$
12. Estimate $\underline{\widehat{\tau}}_{i,j}^2 = \frac{1}{n}\left\langle y^B, y^{ij} - y^A\right\rangle$ and $\widehat{\widetilde{\tau}}_{i,j}^2 = \frac{1}{4n}\left\langle \Delta^2 y, \Delta^2 y\right\rangle$ for (i,j)
13. Select the indices I belonging to the L absolutely largest values in $\Delta^2 y$
14. Plot lines between $(x_{I,j}^A, x_{I,i}^A)$ and $(x_{I,j}^B, x_{I,i}^B)$
choosing a color based upon the sign of $\Delta^2 y / ((y_i^B - y_i^A)(y_j^B - y_j^A))$
15. Advance to the next subplot
16. Return $\widehat{\widetilde{\tau}}_i^2$, $\underline{\widehat{\tau}}_i^2$, $\widehat{\widetilde{\tau}}_{i,j}^2$, $\underline{\widehat{\tau}}_{i,j}^2$

over the product of output terms is written with inner product notation. In line 11, by construction both y^i and y^j are available and they need to be stored. At line 14, the plots are arranged in such way that the smaller index j is located on the x-axis. A MATLAB implementation can be found at https://github.com/emanueleborgonovo/MikadoPlots.

Example 12. Figure 1 reports the Mikado plot for the well-known Ishigami function $g(\mathbf{X}) = \sin(X_1)(1.0 + 0.1X_3^4) + 7.0\sin(X_2)^2$ where the three inputs are uniformly distributed on $[-\pi, \pi]$. The two graphs differ in the sample size used: in the left graph the input space was inspected at $N = 10{,}000$ locations, in the right graph at $N = 100{,}000$ location. The Ishigami example has been frequently studied in the computer experiment literature and the analytical value of the Sobol' interaction index $S_{1,3}$ is 0.2437 (while $S_{1,2}$ and $S_{2,3}$ are null). Using the proposed estimators, we obtain estimates of 0.2401 at $N = 10{,}000$ and of 0.2442 at $N = 100{,}000$. The design also allows one to obtain regional information about the behavior of interactions. The Mikado plots display the segments associated with the 20 locations where the highest absolute value of $\tau_{i,j}^{\mathbf{x}\to\mathbf{z}}$ is registered. The sign is determined via the corresponding second-order Newton ratios and we use blue/dash-dotted for positive and red/dashed for negative interactions. The Mikado plots show that interactions are both positive and negative with a division: no negative interactions are encountered for X_3 smaller than -2, at either sample sizes. Symmetrically, for $X_3 > 2$ no positive interaction is registered. As a simple product-term interaction yields an X-shaped Mikado plot, we expect a more complex interaction term in this example.

TABLE 2 Interaction types versus interaction measurement

Scale	Pw.-def.	Spurious	Sign	Discrete	Cost	Assumption
Infinites	No	No	Yes	No	$4nN$	$\mathcal{C}^1$
Finite	Yes	No	Yes	Yes	2^n	Existence
Global	Yes	Yes	No	Yes	$N^2 2^n$	$\mathcal{L}^2$

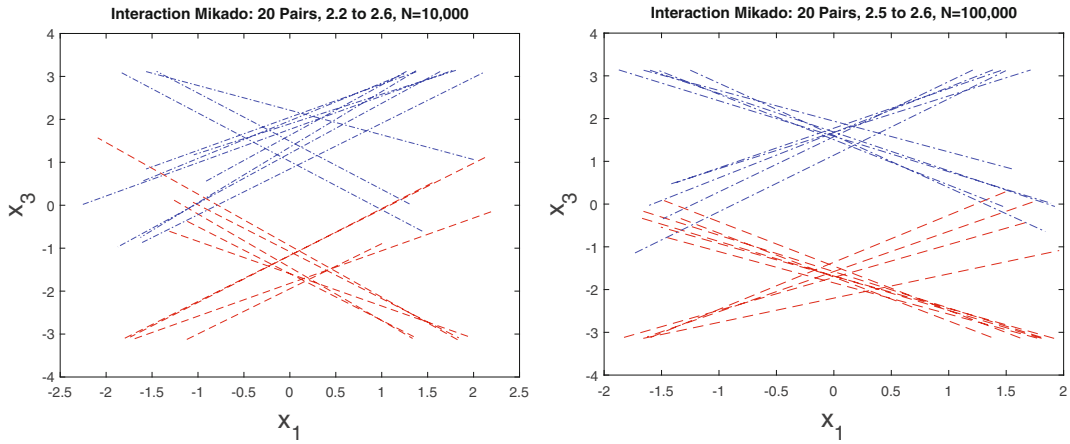


FIGURE 1 Mikado plot for the Ishigami example. Lines refer to the locations in the $X_1 - X_3$ plane associated with the 30 highest (in magnitude) second-order finite-change interaction effects, $|\tau_{i,j}^{\mathbf{x}\to\mathbf{z}}|$. blue-colored, dash-dotted lines (red-colored, dashed lines) denote a positive (negative) interaction. Sample sizes are $N = 10{,}000$ and $N = 100{,}000$ in the left and right graph, respectively

8 | DISCUSSION: CAN A SINGLE INDICATOR FIND THE INTERACTION MECHANISM?

The analysis conducted thus far shows that there is no single method or scale that is capable of delivering a complete picture about the presence of interactions in numerical simulators. Let us consider the triplet "interaction type, investigation scale and quantification method" (Table 2). Methods based on differentiation identify structural interactions and delivering consistent information about sign with dependent or independent inputs. At the same time, their use requires a smooth (twice-differentiable) input–output mappings and they may fail to detect interactions when the mapping is piecewise constant as Example 11 shows. Methods that act on a finite-scale unveil structural interactions due to piecewise-definiteness and to interaction terms as well, are immune to spurious interactions, and reveal the sign of interactions at each location. However, they may be computationally expensive. Global methods allow analysts to appreciate the overall relevance of interactions, to detect structural as well as spurious spurious. However, they do not provide information about the sign of interactions and they may be computationally expensive.

Regarding computational burden, at the infinitesimal scale, the analyst could resort to automatic differentiation (Griewank & Walther, 2008), obtaining quantitative information on partial derivatives while evaluating the computer code without requiring additional model runs. Alternatively, a brute force calculation of second order derivatives based on second order Newton's ratios demands $4n$ model runs. (This number is a lower limit, as one should repeat the calculation for a sequence of decreasing values of the increments used in Newton's ratio, till a limit is approached.) If the estimation of the second-order derivatives is randomized at N locations in $\mathcal{X}$, the estimation cost becomes $C^{\text{Newton}} = 4 \cdot n \cdot N$. At the finite scale, the computational cost associated with the determination of interactions up to order k is $C^k = \sum_{s=0}^{k} \binom{n}{s}$. If $k = n$ (e.g., in a full factorial design) the analyst incurs a cost of 2^n model runs. However, if analysts wish to achieve a less granular information, they can use computational shortcuts. For instance, it is possible to compute the indices $\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$, $\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ and $\overline{\tau}_i^{\mathbf{x}^0}$ at a cost of $2n + 2$ model runs using the shortcut in Equation (15). Moreover, several efficient schemes allow one to compute interaction effects up to a desired order, avoiding the cost of a full factorial design (see Schoen et al., 2017 and Zhou & Xu, 2017 for recent proposals). At the global scale, the modified pick-and-freeze method of Saltelli (2002) allows one to obtain estimates for the first-order, total-order and second-order superset indices at a cost of $N(n + 2)$ model runs. Alternatively, analysts proceed at a cost of N model runs by fitting an emulator over the input–output sample generated for uncertainty quantification (Lin et al., 2010). Depending on the emulation approach used in the analysis, analytical formulas for the estimators of variance-based sensitivity indices of all orders may be available. Then, if the fit is accurate, one has a computationally convenient way of calculating global sensitivity measures.

Given these considerations, we would suggest a four step procedure: (1) specify the interaction or interaction type of interest; (2) select the quantification scale; (3) select a design that allows to estimate such measure, and (4) select the appropriate method. To illustrate, suppose that the analyst is interested in the precise interaction between two specific inputs as they move across two scenarios $\mathbf{x}^0 \to \mathbf{x}^1$. Then, the analyst needs a factorial experiment with four simulator evaluations. If instead, the analyst is interested at the pairwise interaction between X_i and X_j precisely at location $\mathbf{x}^0$, then a second-order mixed derivative is needed. Or, the analyst may just be facing the question of whether there are interactions at all (in any form). Then, a design that allows the estimation of first-order variance-based indices could be the most convenient to apply; if their sum is lower than unity, we know that interactions are present. However, if the analyst is interested

TABLE 3 Case studies

Case study	Inputs	Interaction analysis
Wing-weight	10	Absence of global interactions does not imply the absence of interactions on a finite scale
STOCFOR3	23,541	Interactions analysis for a high dimensional simulator
Gopherus agassizii dynamics	8	Quantification of interactions for a nondifferentiable simulator

in knowing whether interactions are synergistic or antagonistic, she needs a design that allows to retain information on the sign of interactions (see Corollary 2 and the associated discussion with Mikado plots).

We note that the interaction generating mechanisms that we have defined and analyzed appear in computer simulations as well as in real-life experiments. There are, however, practical differences that emerge in the two types of investigations (Wu, 2015). Real-life experiments usually involve a limited number of factors. Computer-based experiments, on the other hand, may involve input–output mappings with a higher dimensionality. The parameter space exploration in real-life experiments might be limited by constraints and is usually performed trough orthogonal (if not factorial) designs. Simulation experiments allow analysts greater flexibility in choosing the input space exploration method (in the context of uncertainty quantification Latin hypercube designs are frequently used Owen, 2020) and the identification scale.

9 | CASE STUDIES

This section presents three case studies aimed at illustrating previous findings in alternative contexts that may emerge in applications. Table 3 summarizes the different interaction quantification settings in each case study. The analysis is carried out on a desktop with Intel(R) core i7-7700HQ CPU at 2.80GHz and 64GB RAM. All calculations are performed in Matlab R2017b.

9.1 | Global interactions versus infinitesimal and finite scale interactions

The Wing-weight model has been introduced in Forrester et al. (2008) to simulate the weight on a light aircraft wing. It is an example of a simulator characterized by a low dimensionality ($n = 10$) and fast running time (0.02 s). We assume throughout the section that the analyst does not know the analytical expression of the model but has access to a black-box computer code in which the calculations are implemented. Jiménez Rugama and Gilquin (2018) offer estimates of global sensitivity measures from an input–output sample of size $N = 65{,}536$. Their accurate estimates (Jiménez Rugama & Gilquin, 2018, Table 12, p. 736) lead to $\sum_{i=1}^{10} \hat{S}_i = 0.9814$. The sum of first-order variance-based indices is close to unity. Fitting an additive linear regression surface to the input–output sample leads to a highly significant fit with a coefficient of model determination $R^2 = 0.982$ (We used an $N = 65{,}536$ sample from the simulator and the `fitlm.m` subroutine). Thus, on a global scale, interactions do not play a major role. However, concluding that the

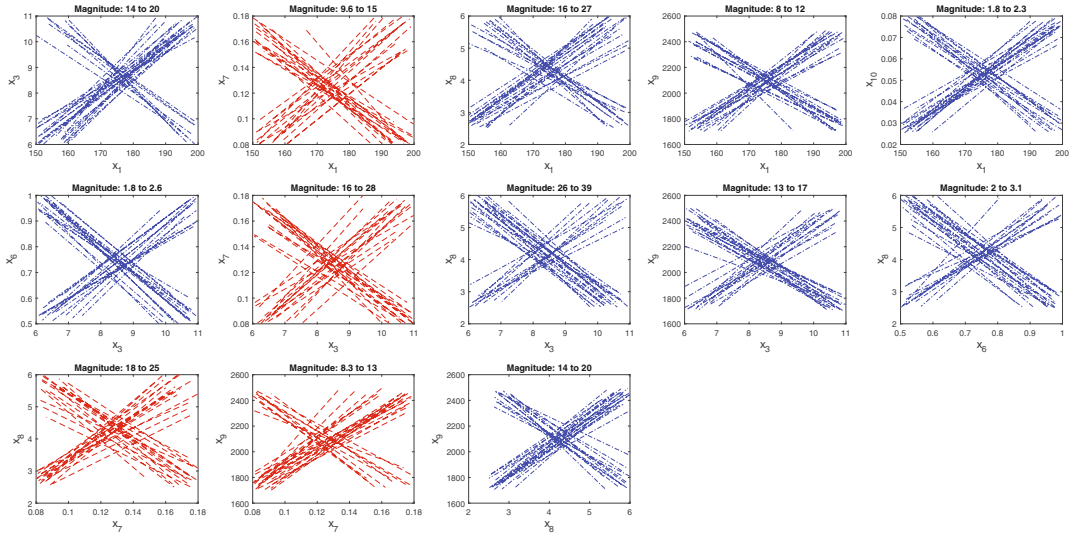


FIGURE 2 Mikado plot for the interaction analysis of the Wing-weight simulator on a finite scale. Top of each panel: smallest and highest magnitude of the corresponding interaction. blue means a positive interaction, red negative

input–output mapping is additive might be misleading and to ascertain this aspect, we conduct an interaction analysis on a finite scale.

We consider pairwise interactions; overall, we have 45 potentially relevant second-order interactions. To study finite-scale interactions, we randomly sample 8192 points in the input space, and consider second-order finite changes between each next pair of realizations; thus, we compute 8191 second-order interaction indices $\tau_{i,j}^{\mathbf{x}^k \to \mathbf{x}^k+1}$, $k = 1, 2, \ldots, 8191$ for each pair. Comparing the magnitudes of the interaction contribution to the total change in each scenario, 13 pairs turn out to be numerically significant. The corresponding Mikado plots are displayed in Figure 2. At the top of each panel in Figure 2 we can read the smallest and highest magnitude of the interactions. The six strongest interactions are the ones between X_8 and X_3 (second row, third panel), X_3 and X_7 (second row, second panel), X_8 and X_1 (first row, third panel), X_8 and X_7 (second row, fourth panel), X_3 and X_9 and X_8 and X_9. Of these, the interactions between X_7 and X_3 and between X_8 and X_7 are negative. Inputs X_2, X_4, X_5, X_6, and X_{10} do not appear to be involved in significant interactions with the remaining inputs. Regarding the interaction generating mechanism, because the input–output mapping is smooth, we can rule out piecewise-definiteness; thus, interactions are due to the presence of product terms (as one can also verify from the analytical expression). From a broader perspective, this result shows that the scale at which an interaction analysis is carried out matters in their interpretation: for this simulator fitting an additive linear model leads to good prediction accuracy, but hides the presence of interactions.

9.2 | Interpreting the response of a high-dimensional simulator

The Wing-weight simulator does not raise any dimensionality concern. We then consider a simulator of larger size, STOCFOR3, the largest linear program in the well-known Netlib repository (the input data of STOCFOR3 are freely available). Linear optimization appears in several

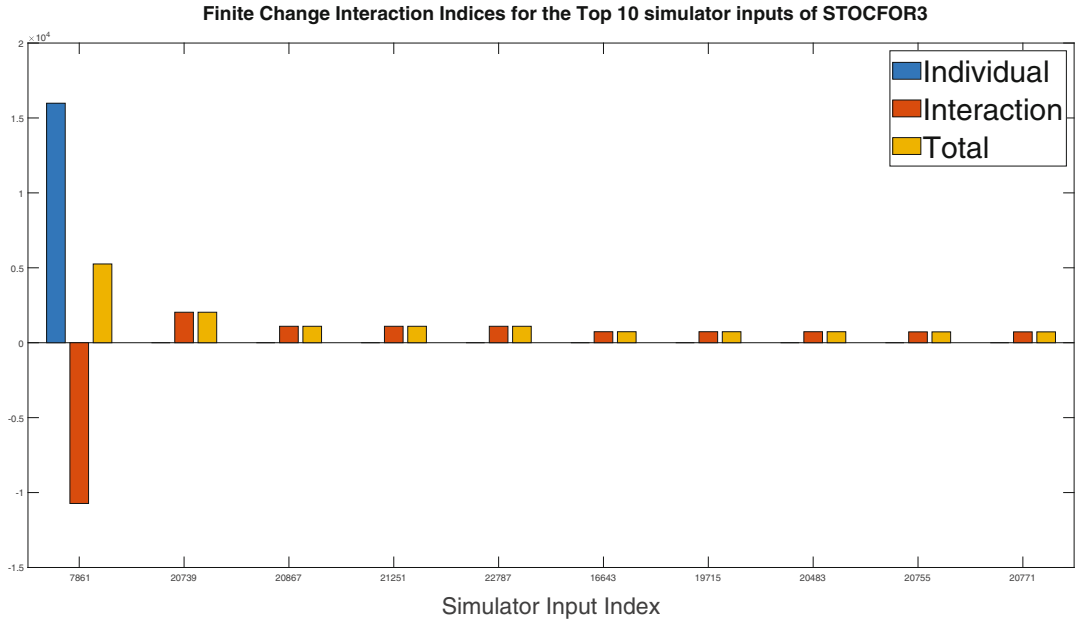


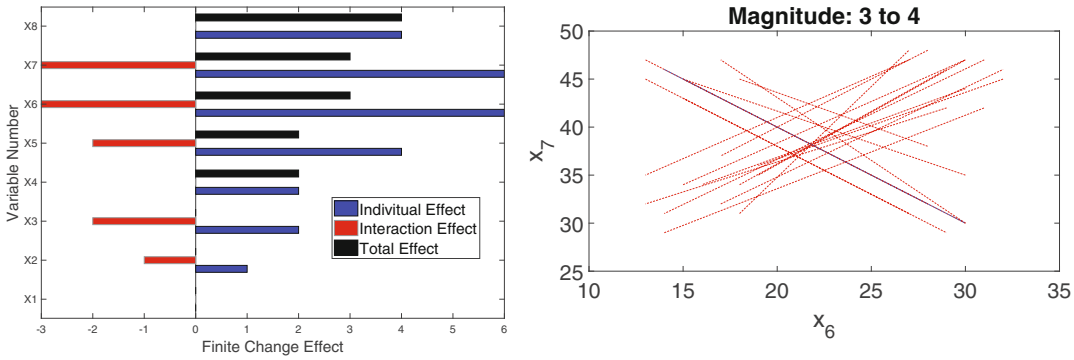
FIGURE 3 Interaction effects for the first 10 most important coefficients of the STOCFOR3 simulator

problems, comprising statistical analyses and regression, with the well-known Dantzig's selector of Candes and Tao (2007) as an outstanding example. We study the sensitivity of the optimal value of this linear program to changes in its 23,541 coefficients for variations of ±99% of their values. In other words, we consider graph (a) in Figure A1, in which $\mathbf{x}^0$ and $\mathbf{x}^1$ are the extremes of the 23,541-dimensional hyperbox. We observe that the output of this simulator is, at any value of the coefficients, a linearly additive map.

Given the simulator size, the shortcut in Equation (15) turns out essential to understand whether the response of the code is, indeed, additive—to illustrate, there are about $2.7 \cdot 10^8$ s order interactions. In fact, with 47,084 simulator evaluations we can obtain all first ($\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$), total ($\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$) and interaction effects ($\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$). Each simulator evaluation entails a new optimization. The analysis takes about 14 h on the above-mentioned computer. Figure 3 reports the finite change effects for the first 10 simulator inputs ranked using the magnitude of $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$. Figure 3 shows that, indeed, interaction effects are present in the simulator response. Not only, but we can understand the underlying cause. Note that the profit of a linear program is a piecewise linear additive map with respect to the coefficients of the objective functions. Second-order derivatives are null at any point $\mathbf{x} \in \mathcal{X}$. Thus, interactions are due to piecewise-definiteness and not to the presence of interaction terms (see the discussion in Section 8).

9.3 | When the output is discrete

In this section, we discuss the analysis of the *Gopherus agassizii* desert tortoise by Hogdson and Townley (2004). The simulator utilizes a Leslie matrix with eight size classes and both the inputs



(a) Finite change interaction indices. (b) Mikado plot for the pair X_6, X_7.

FIGURE 4 Interaction analysis for the *Gopherus agassizii* simulator of Hogdson and Townley (2004). (a) Finite change interaction indices; (b) Mikado plot for the pair X_6, X_7

and output of this model are discrete random variables. The spectral radius of the Leslie matrix is close to but lower than unity, so that the example population eventually dies out. The output value of the model is the number of time steps until at least two individuals remain alive. The model input is the vector of the initial class values, $\mathbf{x}^0 = (89{,}163, 62, 27, 16, 13, 29, 5)$.

Given the discrete nature of the problem, an interaction analysis can be carried out at a finite or at a global scale, but a differentiation approach is ruled out. On a global scale, we use discrete uniform distributions between $\mathbf{x}^0$ and $\mathbf{x}^1$, with $\mathbf{x}^1 = \mathbf{x}^0 + 20$, and independent inputs. At a sample size of $N = 50{,}000$, we register variance-based indices equal to $\hat{S}_1 = 0.0014$, $\hat{S}_2 = 0.0032$, $\hat{S}_3 = 0.0045$, $\hat{S}_4 = 0.0277$, $\hat{S}_5 = 0.0773$, $\hat{S}_6 = 0.3103$, $\hat{S}_7 = 0.3134$, $\hat{S}_8 = 0.2350$. The sum of first-order indices accounts for about 97% of the simulator output variance, showing that interactions do not play a major role on a global scale. An analysis of interactions on a finite scale for the change $\mathbf{x}^0 \to \mathbf{x}^1$ produces the results in Figure 4a. Figure 4a shows that the simulator responds additively to the change in X_8, but interaction effects are relevant and tend to be opposite to individual effects for all other model inputs. Figure 4b displays the Mikado plot for the interaction between X_6 and X_7, the most relevant one. The magnitude of this interaction varies from 3 to 4 and the interaction is mostly negative. This negative interaction evidences a bottleneck in the adolescence of the species, where only a limited amount of individuals reaches the more mature ages where survival and reproduction rates are higher. Only classes 6, 7, and 8 produce offsprings, with an increasing reproduction rate; the promotion rate from class 6 to class 7 is about 25% of individuals, from class 7 to class 8 is below 2%, while the survival rate in class 6 is 68%, and 86% in classes 7 and 8.

From a broader viewpoint, this case study evidences that combining methods on a global and finite scale is essential for the analyst to obtain insights on the nature of interactions, since a global approach alone indicates the simulator as almost additive.

ACKNOWLEDGMENTS

We wish to thank the editor, Professor Skaug, the anonymous associate editor and three anonymous reviewers for the precious suggestions and observations. Financial support from the Isaac Newton Institute for Mathematical Sciences in Cambridge is kindly acknowledged by Giovanni Rabitti. Elmar Plischke was visiting the Department of Decision Sciences at Bocconi University while this paper was written. Open Access Funding provided by Universita Bocconi within the CRUI-CARE Agreement.

ORCID

Emanuele Borgonovo https://orcid.org/0000-0001-8659-6017

REFERENCES

Berrington de González, A., & Cox, D. (2007). Interpretation of interaction: A review. *The Annals of Applied Statistics*, *1*(2), 371–385.

Borgonovo, E. (2010). Sensitivity analysis with finite changes: An application to modified EOQ models. *European Journal of Operational Research*, *200*(1), 127–138.

Borgonovo, E., & Peccati, L. (2010). Moment calculations for piecewise-defined functions: An application to stochastic optimization with coherent risk measures. *The Annals of Operations Research*, *176*(1), 235–238.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). Wiley-Interscience.

Caflisch, R. E., Morokoff, W., & Owen, A. B. (1997). Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, *1*, 27–46.

Campolongo, F., Saltelli, A., & Cariboni, J. (2011). From screening to quantitative sensitivity analysis: A unified approach. *Computer Physics Communications*, *182*(4), 978–988.

Candes, E., & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, *35*(6), 2313–2351.

Chakraborty, S., & Zhang, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association*, *114*(528), 1638–1650.

Chastaing, G., Gamboa, F., & Prieur, C. (2012). Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. *Electronic Journal of Statistics*, *6*, 2420–2448.

Chastaing, G., Gamboa, F., & Prieur, C. (2015). Generalized Sobol sensitivity indices for dependent variables: Numerical methods. *Journal of Statistical Computation and Simulation*, *85*(7), 1306–1333.

Coveney, S., & Clayton, R. H. (2019). Sensitivity and uncertainty analysis of two human atrial cardiac cell models using Gaussian process emulators. *bioRxiv*. https://doi.org/10.1101/818047.

Cox, D. R. (1984). Interaction. *The International Statistical Review*, *52*(1), 1–24.

Durrande, N., Ginsbourger, D., Roustant, O., & Carraro, L. (2013). ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. *The Journal of Multivariate Analysis*, *115*, 57–67.

Efron, B., & Stein, C. (1981). The jackknife estimate of variance. *Annals of Statistics*, *9*(3), 586–596.

Ferrari-Trecate, G., & Muselli, M. (2002). *A new learning method for piecewise linear regression*. In J. R. Dorronsoro (Ed.), *Artificial neural networks — ICANN 2002* (pp. 444–449). Springer.

Forrester, A., Sóbester, A., & Keane, A. (2008). *Engineering design via surrogate modeling*. John Wiley & Sons.

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3), 916–954.

Fruth, J., Roustant, O., & Kuhnt, S. (2014). Total interaction index: A variance-based sensitivity index for second-order interaction screening. *Journal of Statistical Planning and Inference*, *147*, 212–223.

Gamboa, F., Janon, A., Klein, T., Lagnoux, A., & Prieur, C. (2016). Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics*, *50*(4), 881–902.

Gelman, A. (2005). Analysis of variance: Why it is more important than ever. *Annals of Statistics*, *33*(1), 1–53.

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: Principles and techniques of algorithmic differentiation* (2nd ed.). SIAM.

Herrera, I. (2007). Theory of differential equations in discontinuous piecewise-defined functions. *Numerical Methods Partial Differential Equations*, *23*, 597–639.

Hoeffding, W. (1948). A non parametric test of independence. *Annals of Mathematical Statistics*, *90*(4), 546–557.

Hogdson, D., & Townley, S. (2004). Linking management changes to population dynamic responses: The transfer function of a projection matrix perturbation. *Journal of Applied Ecology*, *41*, 1155–1161.

Højsgaard, S. (2003). Split models for contingency tables. *Computational Statistics & Data Analysis*, *42*(4), 621–645.

Højsgaard, S. (2004). Statistical inference in context specific interaction models for contingency tables. *Scandinavian Journal of Statistics*, *31*(1), 143–158.

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, *52*(1), 1–17.

Hooker, G. (2004). Discovering additive structure in black box functions. *Proceedings of KDD2004, ACMDL*, pp. 575–580.

Hooker, G. (2007). Generalized functional ANOVA diagnostics for high dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, *16*(3), 709–732.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., & Prieur, C. (2014). Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, *18*, 342–364.

Jiménez Rugama, L. A., & Gilquin, L. (2018). Reliable error estimation for Sobol' indices. *Statistics and Computing*, *28*(4), 725–738.

Kim, H.-M., Mallick, B. K., & Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, *100*(470), 653–668.

Kirichenko, A., & van Zanten, H. (2017). Estimating a smooth function on a large graph by Bayesian Laplacian regularisation. *Electronic Journal of Statistics*, *11*(1), 891–915.

Kleijnen, J. P. C. (2015). *Design and analysis of simulation experiments* (2nd ed.). Springer.

Koehler, J., & Owen, A. (1996). *Computer experiments*. In S. Ghosh & C. Rao (Eds.), *Handbook of statistics* (pp. 261–308). Elsevier Science.

Kuo, F., Sloan, I., Wasilkowski, G., & Woźniakowski, H. (2010). On decompositions of multivariate functions. *Mathematics of Computation*, *79*(270), 953–966.

Lancaster, H. O. (1971). The multiplicative definition of interaction. *The Australian Journal of Statistics*, *13*(1), 36–44.

Landsheer, J. A., & van den Wittenboer, G. (2015). Unbalanced 2 x 2 factorial designs and the interaction effect: A troublesome combination. *PLoS One*, *10*(3), e0121412. https://doi.org/10.1371/journal.pone.0121412

Lauritzen, S. L. (2012). *Interaction models*. In S. Dudoit (Ed.), *Selected works of terry speed* (pp. 91–94). Springer-Verlag.

Lewis, S. M., & Dean, A. M. (2001). Detection of interaction in experiments on large numbers of factors. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *63*(4), 633–672.

Li, G., & Rabitz, H. (2010). D-MORPH regression: application to modeling with unknown parameters more than observation data. *The Journal of Mathematical Chemistry*, *48*(4), 1010–1035.

Li, G., & Rabitz, H. (2012). General formulation of HDMR component functions with independent and correlated variables. *The Journal of Mathematical Chemistry*, *50*, 99–130.

Li, G., & Rabitz, H. (2017). Relationship between sensitivity indices defined by variance- and covariance-based methods. *Reliability Engineering and System Safety*, *167*, 136–157.

Li, G., Rabitz, H., Yelvington, P., Oluwole, O., Bacon, F., Kolb, C., & Schoendorf, J. (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *The Journal of Physical Chemistry*, *114*, 6022–6032.

Lin, C. D., Bingham, D., Sitter, R. R., & Tang, B. (2010). A new and flexible method for constructing designs for computer experiments. *Annals of Statistics*, *38*(3), 1460–1477.

Liu, R., & Owen, A. B. (2006). Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, *101*(474), 712–721.

Lu, R., Wang, D., Wang, M., & Rempala, G. A. (2018). Estimation of Sobol's sensitivity indices under generalized linear models. *Communications in Statistics - Theory and Methods*, *47*(21), 5163–5195.

Marangoni, G., Tavoni, M., Bosetti, V., Borgonovo, E., Capros, P., Fricko, O., Gernaat, D., Guivarch, C., Havlik, P., Huppmann, D., Johnson, N., Karkatsoulis, P., Keppo, I., Krey, V., Broin, E. Ó., Price, J., & Van Vuuren, D. P. (2017). Sensitivity of projected long-term CO2 emissions across the shared socioeconomic pathways. *Nature Climate Change*, *7*(2), 113–117.

Morris, M. D., Moore, L. M., & McKay, M. D. (2008). Using orthogonal arrays in the sensitivity analysis of computer models. *Technometrics*, *50*(2), 205–215.

Oakley, J. E., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *66*(3), 751–769.

Owen, A. B. (2003). The dimension distribution and quadrature test functions. *Statistica Sinica*, *13*, 1–17.

Owen, A. B. (2013). Variance components and generalized Sobol indices. *SIAM/ASA Journal on Uncertainty Quantification*, *1*, 19–41.

Owen, A. B. (2014). Sobol' indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, *2*(1), 245–251.

A. B. Owen. *A first coure in experimental design*. 2020. https://statweb.stanford.edu/\protect\LY1\textbraceleft~\protect\LY1\textbracerightowen/courses/363/

Rabitz, H., & Alis, O. (1999). General foundations of high-dimensional model representations. *The Journal of Mathematical Chemistry*, *25*(2-3), 197–233.

Rahman, S. (2014). A generalized ANOVA dimensional decomposition for dependent probability measures. *arXiv:1408.0722*.

Römisch, W. (2013). *ANOVA decomposition of convex piecewise linear functions*. In J. Dick, F. Y. Kuo, G. W. Peters, & I. H. Sloan (Eds.), *Monte Carlo and Quasi-Monte Carlo methods* (Vol. *2012*, pp. 581–596). Springer.

Roustant, O., Fruth, J., Iooss, B., & Kuhnt, S. (2014). Crossed-derivative based sensitivity measures for interaction screening. *Mathematics and Computers in Simulation*, *105*, 105–118.

Roustant, O., Padonou, E., Deville, Y., Clement, A., Perrin, G., Giorla, J., & Wynn, H. (2018). Group kernels for Gaussian process metamodels with categorical inputs. *ArXiv:1802.02368*, 1–33.

Saltelli, A. (2002). Making best use of model valuations to compute sensitivity indices. *Computer Physics Communications*, *145*, 280–297.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39.

Saltelli, A., & Tarantola, S. (2002). On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, *97*(459), 702–709.

Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modelling. *Statistical Science*, *19*(4), 377–395.

Santner, T. J., Williams, B., & Notz, W. (2018). *The design and analysis of computer experiments* (2nd ed.). Springer.

Scheffé, H. (1959). *The analysis of variance*. John Wiley & Sons.

Scheffé, H. (1970). Note on separation of variables. *Technometrics*, *12*(2), 388–393.

Schoen, E. D., Vo-Thanh, N., & Goos, P. (2017). Two-level orthogonal screening designs with 24, 28, 32, and 36 runs. *Journal of the American Statistical Association*, *112*(519), 1354–1369.

Sobol, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computer Experiment*, *1*, 407–414.

Streitberg, B. (1990). Lancaster interactions revisited. *Annals of Statistics*, *18*(4), 1878–1885.

Vanderweele, T. J. (2015). *Explanation in causal inference*. Oxford University Press.

Wang, P. C. (2007). Planning experiments when some specified interactions are investigated. *Computational Statistics & Data Analysis*, *51*(9), 4143–4151.

Wardetzky, M., Mathur, S., Kälberer, F., & Grinspun, E. (2007). Discrete Laplace operators: No free lunch. *Proceedings of the 5th Eurographics Symposium on Geometry Processing* (pp. 33–37). Eurographics Association.

Wu, J. (2015). Post-Fisherian experimentation: From physical to virtual. *Journal of the American Statistical Association*, *110*(510), 610–620.

Xu, C., & Gertner, G. (2007). Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics & Data Analysis*, *51*(12), 5579–5590.

Zhou, Y.-D., & Xu, H. (2017). Composite designs based on orthogonal arrays and definitive screening designs. *Journal of the American Statistical Association*, *112*(520), 1675–1683.

How to cite this article: Borgonovo, E., Plischke, E., & Rabitti, G. (2022). Interactions and computer experiments. *Scandinavian Journal of Statistics*, *49*(3), 1274–1303. https://doi.org/10.1111/sjos.12560

APPENDIX A. CONNECTION TO THE DISCRETE LAPLACE OPERATOR

The discrete Laplace operator has found many applications in Statistics, including Bayesian function estimation on graphs (Kirichenko & van Zanten, 2017; Wardetzky et al., 2007). Consider the hypercube having $\mathbf{x}^0$ and $\mathbf{x}^1$ on opposite vertices, that is, the set of points connected to $\mathbf{x}^0$ by variation in one or more coordinate. This is the set of nodes in graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$:

$$\mathcal{V} = \{\mathbf{x}^0, \overline{\mathbf{x}}^i, \overline{\mathbf{x}}^{\{i,j\}}, \dots, \overline{\mathbf{x}}^{\{1,2,\dots,n-1\}}, \mathbf{x}^1\}, \tag{A1}$$

where the superscripts indicate which coordinates of the base case $\mathbf{x}^0$ are shifted to those of $\mathbf{x}^1$ with $\overline{\mathbf{x}}^z = (\mathbf{x}^0_{-z} : \mathbf{x}^1_z)$. These points are reached from $\mathbf{x}^0$ by changing single coordinates, pairs, triplets, and so on, including the change in all coordinates. The set of edges $\mathcal{E}$ is formed connecting all points in $\mathcal{V}$ to $\mathbf{x}_0$ via one edge. The graph on the left in Figure A1 offers a visualization in three dimensions.

Then, let g be a function of the vertices, that is, $g : \mathcal{V} \to \mathbb{R}$. Given a reference vertex $v \in \mathcal{V}$, one writes the discrete Laplace operator as

$$(\mathcal{L}g)(v) = \sum_{u \sim v} (g(v) - g(u)), \tag{A2}$$

where $u \sim v$ denote the edges of the graph, $(u, v) \in \mathcal{E}$. Then, for the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ specified above one finds

$$-(\mathcal{L}g)(\mathbf{x}^0) = \sum_{i=1}^{n} \Delta_i g + \sum_{j>i} \Delta_{i,j} g + \cdots + \sum_{i=1}^{n} \Delta_{-i} g + \Delta g. \tag{A3}$$

Proposition 3. *Given g and $\mathcal{G}(\mathcal{V}, \mathcal{E})$ defined above, if g is additive then*

$$(\mathcal{L}g)(\mathbf{x}^0) = -\Delta g \cdot n \cdot (2^n - 1), \tag{A4}$$

and

$$\sum_{\mathbf{x} \sim \mathbf{x}^0, \mathbf{x} \in \mathcal{V}} g(\mathbf{x}) = \left(2^n - 1\right) \left[n \cdot \Delta g + g(\mathbf{x}^0)\right]. \tag{A5}$$

Proof of Proposition 3. As $\Delta_z g = g(\overline{\mathbf{x}}^z) - g(\mathbf{x}^0) = \sum_{j=1}^{k} \left(g(x_{i_j}^1) - g(x_{i_j}^0)\right)$, (A4) holds. Then, noting that (A2) can be written as

$$-(\mathcal{L}g)(\mathbf{x}^0) = \sum_{\mathbf{x} \sim \mathbf{x}^0} \left(g(\mathbf{x}) - g(\mathbf{x}^0)\right) = \left(\sum_{\mathbf{x} \sim \mathbf{x}^0} g(\mathbf{x})\right) - \left(2^n - 1\right) g(\mathbf{x}^0),$$

we obtain (A5). ∎

Equations (A4) and (A5) show that, if g is additive, the discrete Laplace operator on $\mathcal{G}(\mathcal{V}, \mathcal{E})$ can be computed by just two evaluations of g, at $\mathbf{x}^0$ and $\mathbf{x}^1$, respectively. Additionally, if $g(\mathbf{x}^0) = 0$, a simple rewriting of (A5) shows that

$$\frac{1}{n} \sum_{\mathbf{x} \sim \mathbf{x}^0, \mathbf{x} \in \mathcal{V}} g(\mathbf{x}) = \left(2^n - 1\right) \Delta g, \tag{A6}$$

that is the mean effect of g can be computed using only these two evaluations.

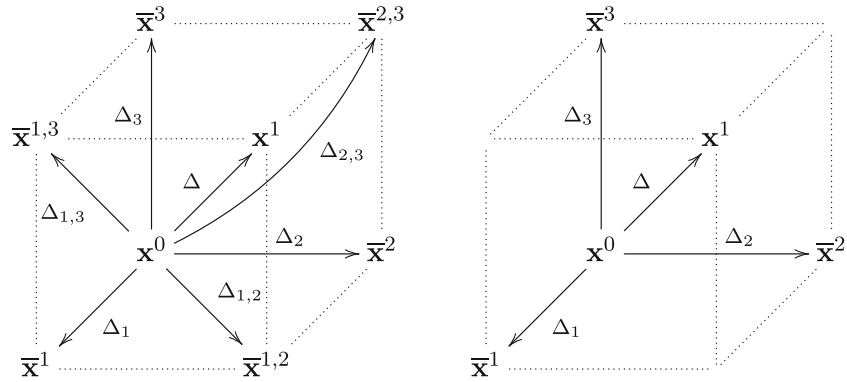


FIGURE A1 Three-dimensional visualization. Left: full factorial design $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Right: one at a time design $\mathcal{G}(\mathcal{V}', \mathcal{E}')$

Consider now a subgraph of $\mathcal{G}(\mathcal{V}, \mathcal{E})$, in which the set of nodes contains $n + 2$ vertices determined by $\mathbf{x}^0$, $\mathbf{x}^1$ and all the evaluation points $\overline{\mathbf{x}}^i$, for all $i \in [n]$. We denote this graph by $\mathcal{V}'$ (main text, Figure A1b). Then, each one of these vertices is connected to $\mathbf{x}^0$ by one edge, so that the set of edges has cardinality $n + 1$.

Proposition 4. *Given $\mathcal{G}(\mathcal{V}', \mathcal{E}')$, and $g : \mathcal{V}' \mapsto \mathbb{R}$, if g is additive then*

$$(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g = 0. \tag{A7}$$

Proof of Proposition 4. Rewriting (A7), one finds

$$-(\mathcal{L}g)(\mathbf{x}^0) = \sum_{\{\mathbf{x}^1\} \cup \left\{\overline{\mathbf{x}}^i, i=1,\ldots,n.\right\}} \left(g(u) - g(\mathbf{x}^0)\right) = g(\mathbf{x}^1) - g(\mathbf{x}^0) + \sum_{i=1}^{n} \Delta_i g,$$

because all the $n + 1$ edges incident to $\mathbf{x}^0$ must be included. Then, $g(\mathbf{x}^1) - g(\mathbf{x}^0) = \Delta g$ by definition, and $\sum_{i=1}^{n} \Delta_i g$ is also equal to Δg by Definition 3. ∎

On the other hand, let us consider $\mathcal{G}(\mathcal{V}', \mathcal{E}')$ and a generic mapping g which is not necessarily additive. Then, the application of the discrete Laplace operator leads to $(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g \neq 0$, and, specifically,

$$(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g = \Delta g - \sum_{i=1}^{n} \Delta_i g. \tag{A8}$$

APPENDIX B. PROOFS

Proof of Proposition 1. We start with the if. If g can be represented in the form of (4), we can write:

$$\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0) = \sum_{i=1}^{d} a_i(x_i^1) - \sum_{i=1}^{d} a_i(x_i^0) = \sum_{i=1}^{d} [a_i(x_i^1) - a_i(x_i^0)].$$

Then, we have

$$\Delta_i g = g(x_i^1; \mathbf{x}_{\sim i}^0) - g(\mathbf{x}^0) = a_i(x_i^1) + \sum_{s=1, s\neq i}^{d} a_s(x_s^0) - a_i(x_i^0) - \sum_{s=1, s\neq i}^{d} a_s(x_s^0) = a_i(x_i^1) - a_i(x_i^0).$$

Therefore, combining this last equality with the previous equality, we have $\Delta g = \sum_{i=1}^{d} \Delta_i g$. Let us now consider the converse implication. We assume that $g(\mathbf{x}) - g(\mathbf{x}^0) = \sum_{i=1}^{d} [g(x_i; \mathbf{x}_{\sim i}^0) - g(\mathbf{x}^0)]$, for all $\mathbf{x}, \mathbf{x}^0$ in $\mathcal{X}$. Then, fix any $\mathbf{x}^0 \in \mathcal{X}$ and define $a_i(x_i) = g(x_i; \mathbf{x}_{\sim i}^0) - (1 - \frac{1}{d} g(\mathbf{x}^0))$. Then $\sum_{i=1}^{d} a_i(x_i) = g(\mathbf{x}^0) + \sum_{i=1}^{d} \Delta_i g = g(\mathbf{x}^0) + \Delta g = g(\mathbf{x})$, as we wished to prove. ∎

Proof of Theorem 1. We prove this result showing that the Scheffé's condition (9) is not satisfied. Because this condition is necessary, it does not exist a transformation to additivity. We distinguish two cases. If the mapping g is not differentiable, clearly it cannot satisfy Scheffé's condition. Consider now the case of g differentiable. By contradiction, assume that there exists a transformation $\eta(g)$ that makes g additive. From the result of Scheffé (1970), it follows that it exists a function $\omega(g)$ satisfying condition (9) and that the univariate functions $a_i(x_i)$ can then be represented as in (10). Consider the function $\Psi(x_i, x_j) = \frac{\psi_i(x_i)}{\psi_j(x_j)} = \frac{\partial g}{\partial x_i} \Big/ \frac{\partial g}{\partial x_j}$. This function $\Psi(x_i, x_j)$ is a function of x_i and x_j only, for all $i < j$. By assumption, g is piecewise-defined and consequently it holds that $\frac{\partial g(\mathbf{x})}{\partial x_i} = \frac{\partial h_\ell(\mathbf{x})}{\partial x_i}$ with $\mathbf{x} \in \mathcal{X}_\ell^\Pi$, $\ell = 1, 2, \ldots, L$ (see Definition 3). Thus we have that $\Psi(\mathbf{x}) = \frac{\partial h_\ell(\mathbf{x})}{\partial x_i} \Big/ \frac{\partial h_\ell(\mathbf{x})}{\partial x_j}$ for $\mathbf{x} \in \mathcal{X}_\ell^\Pi$, $\ell = 1, 2, \ldots, L$. However, $\Psi(\mathbf{x})$ is not a function of x_i and x_j only, since the values of these derivatives depend on which set of the partition $\mathcal{X}_\ell^\Pi$ they are calculated. Indeed, since the partition sets $\mathcal{X}_\ell^\Pi$, $\ell = 1, 2, \ldots, L$, are defined in terms of $\mathbf{x}$, the values of the partial derivatives depend on the remaining variables $\mathbf{x}_{-ij}$ as well. This is a contradiction, as the constancy of $\Psi(x_i, x_j)$ with respect to the other variables for all $i < j$ is a necessary condition for the existence of a transformation to additivity (see Scheffé, 1970, p. 392). ∎

Proof of Theorem 2. If g is separately additive (see Equation (4)), then $g''_{i,j}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$. Conversely, suppose that the second derivatives are null everywhere. Then, consider the change the function between two points in $\mathcal{X}$. We can expand such change using a Taylor series:

$$g(\mathbf{x}^1) = g(\mathbf{x}^0) + \sum_{i=1}^{n} g'_i(\mathbf{x}^0)(x_i^1 - x_i^0) + \sum_{i,j=1}^{n} \frac{1}{2} g''_{i,j}(\mathbf{x}^0)(x_i^1 - x_i^0)(x_j^1 - x_j^0) + \ldots \tag{B1}$$

Then, note that the assumption $g''_{i,j}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$ implies that the function $g''_{i,j}(\mathbf{x})$ is a constant on $\mathcal{X}$ and infinitely many times differentiable, with null derivatives of all orders. This implies that all mixed derivatives from order 2 onwards are null. Therefore, we have $g(\mathbf{x}) = g(\mathbf{x}^0) + \sum_{i=1}^{n} g'_i(\mathbf{x}^0)(x_i - x_i^0)$. Then, one can write $g(x^1) - g(\mathbf{x}^0) = \sum_{i=1}^{n} g'_i(\mathbf{x}^0)(x_i - x_i^0)$. Now, observe that $g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0) = g'_i(\mathbf{x}^0)(x_i - x_i^0) + \sum_{i=1}^{n} \frac{1}{2} g''_i(\mathbf{x}^0)(x_i - x_i^0)^2 + \ldots$. Again, by the nullity of the second and higher-order derivatives, we obtain $g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0) = g'_i(\mathbf{x}^0)(x_i - x_i^0)$, and therefore

$$g(\mathbf{x}^1) - g(\mathbf{x}^0) = \sum_{i=1}^{n} g'_i(\mathbf{x}^0)(x_i - x_i^0) = \sum_{i=1}^{n} g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0),$$

∎

Proof of Proposition 2. By definition,

$$\mathbb{V}[Y|X_i] = \mathbb{E}\left[(Y - \mathbb{E}[Y|X_i])^2|X_i\right]. \tag{B2}$$

Now, $\mathbb{E}[Y|X_i = x_i]$ is $g_0 + g_i(x_i)$ so that $Y - \mathbb{E}[Y|X_i = x_i] = \sum_{j\neq i} g_j(x_j) + \sum_{|\alpha|\geq 2} g_\alpha(x_\alpha)$. Hence,

$$\mathbb{V}[Y|X_i] = \mathbb{E}\left[\left(\sum_{j\neq i} g_j(x_j) + \sum_{|\alpha|\geq 2} g_\alpha(x_\alpha)\right)^2 \Big| X_i\right]. \tag{B3}$$

Then, due to the orthogonality of the functions in (B3), we can write

$$\mathbb{V}[Y|X_i] = \sum_{j\neq i}\mathbb{E}\left[g_j^2(x_j)\right] + \sum_{|\alpha|\geq 2}\mathbb{E}\left[g_\alpha(x_\alpha)^2\Big|X_i\right]. \tag{B4}$$

Now, the first sum, $\sum_{j\neq i}\mathbb{E}[g_j^2(x_j)]$, does not depend on X_i. Thus, if only these terms were present in the decomposition, we would not have heteroskedasticity. Under heteroskedasticity, at least one of the terms $\mathbb{E}[g_\alpha(x_\alpha)^2|X_i]$ with $i \in \alpha$ has to be different from zero, because $\mathbb{V}[Y|X_i]$ is not constant. Thus, we have interaction terms in the function and these terms involve X_i. Conversely, if there is at least one multi-index α with $|\alpha| \geq 2$ and $i \in \alpha$, we have that $\mathbb{V}[Y|X_i]$ varies with X_i and, therefore, there is heteroskedasticity. ∎

Proof of Corollary 1. From the variance decomposition formula, $\mathbb{E}[\mathbb{V}[Y|X_i]] \leq \mathbb{V}[Y]$ follows. Then, if there is a point $x_i \in S$ where $\mathbb{V}[Y|X_i = x_i] > \mathbb{V}[Y]$, the conditional variance cannot be a constant. ∎

Proof of Theorem 3. By theorem 1 in Liu and Owen (2006), the superset importance can be written as

$$\Upsilon_u = \frac{1}{2^{|u|}}\int\left(\sum_{v\subset u}(-1)^{|u-v|}g(\mathbf{x}_v, \mathbf{z}_{-v})\right)^2 d\mathbf{x}^u d\mathbf{z}.$$

Rewriting the expression of $\tau_u^{\mathbf{x}\to\mathbf{z}}$, we find that $\tau_u^{\mathbf{x}\to\mathbf{z}} = \sum_{v\subset u}(-1)^{|u-v|}g(\mathbf{x}_v, \mathbf{z}_{-v})$. Using this equality, it follows that $\Upsilon_u = \frac{1}{2^{|u|}}\int(\tau_u^{\mathbf{x}\to\mathbf{z}})^2 d\mathbf{x}^u d\mathbf{z}$, proving the first statement in (27). The second statement can be proved rewriting the estimator of the Sobol index

$$\underline{\tau}_u^2 = \mathbb{E}[g(\mathbf{X})(g(\mathbf{X}_u : \mathbf{Z}_{-u}) - g(\mathbf{Z}))] = \sum_{v\subset u}\mathbb{E}\left[g(\mathbf{X})\tau_v^{\mathbf{Z}\to\mathbf{X}}\right].$$

Similarly, for the third item, we have: $\overline{\tau}_u^2 = \frac{1}{2}\mathbb{E}[(g(\mathbf{Z}_u : \mathbf{X}_{-u}) - g(\mathbf{X}))]^2 = \frac{1}{2}\mathbb{E}\left[\sum_{v\subset u}\tau_v^{\mathbf{X}\to\mathbf{Z}}\right]^2$.

In order to prove (28), by theorem 2.2 in Owen (2013) it holds that $\sum_u |u|V_u = \sum_{i=1}^n \overline{\tau}_i^2$. By (27), we find $\sum_u |u|V_u = \frac{1}{2}\sum_{i=1}^n\mathbb{E}\left[\tau_i^{\mathbf{X}\to\mathbf{Z}}\right]^2$. For the other equality in (28), we consider the equation $\sum_{i=1}^n V_i = \sum_{i=1}^n \underline{\tau}_i^2$ at p. 32 in Owen (2013) and plug in the estimator for $\underline{\tau}_i^2$ in (27). ∎

APPENDIX C. FERRARI–TRECATE AND MUSELLI'S FUNCTION IN EXAMPLE 1

The piecewise linear function used in Ferrari-Trecate and Muselli (2002) is:

$$g(x_1, x_2) = \begin{cases} 3 + 4x_1 + 2x_2, & \text{if } 0.5x_1 + 0.29x_2 \geq 0 \text{ and } x_2 \geq 0 \\ -5 - 6x_1 + 6x_2 & \text{if } 0.5x_1 + 0.29x_2 < 0 \text{ and } 0.5x_1 - 0.29x_2 < 0, \\ -2 + 4x_1 - 2x_2 & \text{if } 0.5x_1 - 0.29x_2 \geq 0 \text{ and } x_2 < 0 \end{cases}$$

where the input space is $[-1, 1] \times [-1, 1]$.