Statistics and Computing manuscript No.
(will be inserted by the editor)

Nonparametric estimation of probabilistic sensitivity measures

Isadora Antoniano-Villalobos · Emanuele Borgonovo · Xuefei Lu

This is a post-peer-review, pre-copyedit version of an article published in Statistics and Computing. *The final authenticated version is available online at:* https://doi.org/10.1007/s11222-019-09887-9

Received: date / Accepted: date

Abstract Computer experiments are becoming increasingly important in scientific investigations. In the presence of uncertainty, analysts employ probabilistic sensitivity methods to identify the key-drivers of change in the quantities of interest. Simulation complexity, large dimensionality and long running times may force analysts to make statistical inference at small sample sizes. Methods designed to estimate probabilistic sensitivity measures at relatively low computational costs are attracting increasing interest. We first, propose new estimators based on a one-sample design and building on the idea of placing piecewise constant Bayesian priors on the conditional distributions of the output given each input, after partitioning the input space. We then present two alternatives, based on Bayesian nonparametric density estimation, which bypass the need for predefined partitions. Quantification of uncertainty in the estimation process through is possible without requiring additional simulator evaluations via Bootstrap in the simplest proposal, or from the posterior distribution over the sensitivity measures, when the entire inferential procedure is Bayesian. The performance of the proposed methods is compared to that of traditional point estimators in a series of numerical experiments comprising synthetic but challenging simulators, as well as a realistic application.

Keywords Bayesian nonparametrics · Density estimation · Density regression · Design and analysis of computer experiments · Probabilistic sensitivity analysis

Isadora Antoniano-Villalobos
E-mail: isadora.antoniano@unive.it
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy and Bocconi Institute for Data Science and Analytics (BIDSA), Bocconi University, Milan, Italy.

Emanuele Borgonovo
E-mail: emanuele.borgonovo@unibocconi.it
Department of Decision Sciences and BIDSA, Bocconi University, Milan, Italy.

Xuefei Lu
E-mail: xuefei.lu@unibocconi.it
Department of Energy, Politecnico di Milano and Department of Decision Sciences, Bocconi University, Milan, Italy.

1 Introduction

The use of computer simulations is becoming increasingly important in broad areas of science (Lin et al. 2010; Wong et al. 2017). High-fidelity mathematical models allow analysts to perform virtual (or *in silico*) experiments on complex natural or societal phenomena of interest (see Smith et al. 2009, among others). However, the level of sophistication of the models is often too high for analytical solutions to be available and the input-output mapping remains a black-box to the analyst. It then becomes important to carefully design and execute the computer experiment. The design and analysis of computer experiments (DACE) has entered the statistical literature with the seminal work of Sacks et al. (1989) (see also the monographs of Santner et al. (2003); Kleijnen (2008)). Since then, researchers have studied the creation of space-filling designs (Pronzato and Müller 2012; He 2017), the calibration of computer codes with real data (Tuo and Wu 2015), their emulation (Marrel et al. 2012; Le Gratiet et al. 2014), the quantification of uncertainty in their output (Oakley and O'Hagan 2002; Ghanem et al. 2016) and their sensitivity analysis (Oakley and O'Hagan 2004; Borgonovo et al. 2014). These areas are intertwined. A given design may allow, for instance, not only an uncertainty quan-

tification, but also the creation of an emulator and the analysis of sensitivity.

Probabilistic (or global) sensitivity measures are an indispensable component of uncertainty quantification, as they highlight which areas should be given priority when planning data collection or further modelling efforts. International agencies such as the U.S. Environmental Protection Agency (U.S. Environmental Protection Agency 2009) or the European Commission (2009) have issued guidelines recommending them as the gold standard for ensuring reliability and transparency when using the output of a computer code for decision-making under uncertainty. Over the years, several probabilistic sensitivity measures have been proposed. Alternative measures possess different properties making them preferable in alternative contexts and for different purposes. We recall regression-based (Helton and Sallaberry 2009), variance-based (Saltelli and Tarantola 2002; Jiménez Rugama and Gilquin 2018) and moment-independent measures (Borgonovo et al. 2014; Borgonovo and Iooss 2017), all of which offer alternative ways to quantify the degree of statistical dependence between the simulator inputs and the output. A transversal issue in realistic applications is that analytical expressions of these measures are unavailable and analysts must resort to estimation.

Recent works (e.g. Strong et al. 2012; Plischke et al. 2013; Borgonovo et al. 2016) evidence the *one-sample* (or *given-data*) approach as an attractive design for estimating global sensitivity measures from a single probabilistic sensitivity analysis sample, but leave some research questions open. One-sample estimation procedures are closely related to scatter-plot smoothing, where partitioning of the covariate space plays a central role (Hastie and Tibshirani 1990). Strong and Oakley (2013, Fig. 1, p. 759) show that the choice of partition size affects estimation, especially when the sample size is small. Some heuristics for determining a partition selection strategy which is optimal in some sense have been proposed, but finding a universally valid heuristic seems out of reach (see Appendix A.2 for numerical experiments illustrating this issue). Moreover, because most one-sample estimators are consistent (in the frequentist sense), an accurate estimation of the error is often overlooked. However, especially at small sample sizes, it is essential for transparency that uncertainty in the estimates is part of result communication (see Le Gratiet et al. (2014); Janon et al. (2014b); Le Gratiet et al. (2014) among others).

This work proposes to enrich the one-sample design through the use of Bayesian non-parametric (BNP) methods, aiming to reduce and even eliminate the partition selection problem, while making uncertainty in the estimates a natural ingredient. We proceed as follows. We first study a direct extension of currently employed one-sample estimators using non-parametric models to augment the output sample within each partition set. The non-parametric models allow one to adequately generate additional synthetic data according to two alternative schemes. The intuition supporting these designs can be interpreted as setting a prior on the conditional distribution of the output, given that the input realization falls within a given set of the partition. We build estimators based on this intuition for variance-based, density (pdf)-based and cumulative distribution function (cdf)-based global sensitivity measures. Although not fully Bayesian, this approach provides a way to quantify estimation uncertainty. We compare the results with given-data estimators currently in use at low sample sizes, through numerical experiments. The results show that, while the proposed estimators recover the correct ordinal ranking of the inputs, the estimates are strongly influenced by the partition choice, leading to biased estimates, if the partition is not optimal. However, as we show in Appendix A.2, there is no universal rule to determine such optimal partition. Therefore, we investigate two additional classes of estimators based on Bayesian non-parametric joint and conditional density estimation methods. These estimators eliminate the partition selection problem and, at the same time, enable error quantification. We illustrate the use of these methods for the estimation of some variance-based, density-based and distribution-based sensitivity measures. Numerical experiments are carried out first for a 21-input simulator with correlated normal inputs and for a 2-input non-normal example with independent inputs. An application to the global sensitivity analysis of the benchmark computer code known as LevelE (Saltelli and Tarantola 2002), follows. Results show that the estimators correctly identify the most important inputs, with respect to each sensitivity measure considered. Additionally, the analyst obtains a quantification of the uncertainty in the estimates in the form of a posterior distribution, which can be used to assess the reliability of the ranking of inputs given the available input-output sample.

The remainder of the paper is organized as follows. Section 2 introduces the framework of global sensitivity analysis and the one-sample estimation approach. Section 3 combines non-parametric methods and the one-sample approach to create two new partition-dependent estimators. Section 4 derives two classes of Bayesian partition-free estimators by adopting a non-parametric density estimation approach. Section 5 presents numerical results for the LevelE code. Section 6 offers discussion and conclusions. The appendices illustrate the

algorithms of the proposed estimators and discuss additional implementation details.

2 Probabilistic sensitivity measures and their one-sample estimation

Formally, the sensitivity analysis framework considers a multivariate mapping $g : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X} \subseteq \mathbb{R}^k$ and $\mathcal{Y} \subseteq \mathbb{R}^d$. In the DACE set-up, g represents the set of operations performed by a computer code on a vector of inputs $\mathbf{x}$ to produce a vector of outputs $\mathbf{y} = g(\mathbf{x})$ of interest. If the response of the simulator is stochastic, an error term appears in the model response ($\mathbf{y} = g(\mathbf{x}) + \epsilon(\mathbf{x})$ and the output space is possibly affected). For simplicity, we focus on deterministic univariate responses, with $d = 1$. When information is not sufficient to fix the values of the inputs, it is common to "assume to have information about the factors' probability distribution "(Saltelli and Tarantola 2002, p. 704). The input and output probability spaces are denoted $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_{\mathbf{X}})$, and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \mathbb{P}_Y)$, respectively, where $\mathcal{B}$ denotes the Borel sigma-algebra and $\mathbb{P}_Y$ is the probability measure induced by $\mathcal{X}$ through g.

It has been recently shown that several probabilistic sensitivity measures can be expressed as expectations of measures of discrepancy between $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X^i}$. In particular, this work focuses on probabilistic sensitivity measures of the form:

$$\xi_i := \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X^i})], \tag{1}$$

where the expectation is calculated with respect to the marginal distribution of X^i and ζ is a pre-metric on the set of probability measures over $\mathcal{Y}$. The quantity ξ_i is called the *probabilistic sensitivity measure* of X^i with *inner operator* ζ (Borgonovo et al. 2014).

Table 1 reports three probabilistic sensitivity measures encompassed by this construction, namely, the *variance-based* sensitivity measure η_i (Iman and Hora 1990; Sobol' 1993), the *density-based* importance measure δ_i and the *cdf-based* importance measure β_i (Pearson 1905; Oakley and O'Hagan 2004; Borgonovo 2007; Baucells and Borgonovo 2013). Over the years, variance-based sensitivity measures have become the reference measures for global sensitivity analysis. A reason for this is that, when the simulator inputs are independent, the variance of the model output can be orthogonally decomposed into additive components which contain information about the expected percentage of output variance reduction associated with a given model input or model input group. Thus, η_i represents the expected fractional reduction in the simulator output variance attained by fixing input X^i. Originally introduced by Pearson (1905) with the name of correlation ratio, under independence, η_i coincides with the Sobol' first order sensitivity index (Homma and Saltelli 1996). Under dependence, the notion of expected variance reduction and Sobol' indices become distinct (see Mara and Tarantola 2012; Chastaing et al. 2012; Li and Rabitz 2012, for further details). In several applications, global sensitivity measures are used to quantify the degree of statistical dependence between Y on X^i. In this respect, the *nullity-implies-independence property* of Rényi's's postulate D for measures of statistical dependence (Rényi 1959) becomes relevant. In fact, it is reassuring for the analyst that a zero value of a sensitivity measures implies that Y is independent of X_i. A class of sensitivity measures that satisfy such postulate are *distribution-based* (or *moment-independent*) measures (see e.g. Gamboa et al. 2018; Da Veiga 2015; Rahman 2016). As representatives, we select the sensitivity measures δ_i and β_i (Table 1), which quantify the expected separation between $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X^i}$ through the L^1-norm between densities and the Kolmogorov-Smirnov distance between cumulative distribution functions, respectively. We select these sensitivity measures because their estimation is generally considered a challenging task (for further details see e.g. Borgonovo et al. 2014).

A brute force strategy for the estimation of the sensitivity measures in the form of Eq. (1) requires a number of model evaluations equal to $C = kn^2$, where n denotes the sample size required for a Monte Carlo uncertainty quantification. This becomes rapidly infeasible when either the simulator dimension or the sample size increase. However, notable advances in the literature have contributed in abating the computational burden of estimating sensitivity measures. For instance, pick-and-freeze designs produce first and total order sensitivity indices at a cost proportional to $n(k + 1)$ model runs (Saltelli 2002). Designs based on replicates (e.g. Mara and Joseph 2008; Tissot and Prieur 2015) further diminish this cost. Specifically, the design of Tissot and Prieur (2015) lowers the cost to $2n$ model evaluations for the calculation of first order variance-based sensitivity indices. The works of Röhlig et al. (2009); Strong et al. (2012); Plischke et al. (2013) and Borgonovo et al. (2016) show that sensitivity measures in the form of Eq. (1) can be estimated from a single Monte Carlo sample, $\{(\mathbf{x}_j, y_j) : j = 1, \dots n\}$, thus lowering the computational cost to $C = n$ simulator runs. These approaches receive the common name of *one-sample* or *given-data* estimation methods. Once such a sample is available, an alternative two-stage estimation strategy is possible involving the use of emulators. In

Table 1 Some probabilistic sensitivity measures

Measure	$\zeta(\mathbb{P}_Y, \mathbb{P}_{Y\mid X^i})$	ξ_i
η_i	$(\mathbb{E}[Y\|X^i] - \mathbb{E}[Y])^2/\mathbb{V}[Y]$	$\mathbb{V}[\mathbb{E}(Y\|X^i)]/\mathbb{V}[Y]$
δ_i	$\frac{1}{2}\int_{\mathcal{Y}} \lvert f_{Y\mid X^i}(y\|X^i) - f_Y(y)\rvert \mathrm{d}y$	$\frac{1}{2}\mathbb{E}\left[\int_{\mathcal{Y}} \lvert f_{Y\mid X^i}(y\|X^i) - f_Y(y)\rvert \mathrm{d}y\right]$
β_i	$\sup_{y\in\mathcal{Y}} \lvert F_{Y\mid X^i}(y\|X^i) - F_Y(y)\rvert$	$\mathbb{E}\left[\sup_{y\in\mathcal{Y}} \lvert F_{Y\mid X^i}(y\|X^i) - F_Y(y)\rvert\right]$

this case, the fitting of the emulator is performed at a first step. Because the emulator is fast to execute, at a second step sensitivity measures can then be estimated with a potentially unlimited number of simulated observations with emulated output. Several types of emulators have been developed, based on smoothing methods (Da Veiga et al. 2009; Storlie and Helton 2008), polynomial chaos expansion (Blatman and Sudret 2010), Gaussian process regression or kriging (Oakley and O'Hagan 2004; Marrell et al. 2009; Le Gratiet et al. 2014) to name a few. We refer to the handbook chapter of Le Gratiet et al. (2017) and the recent monograph of Santner et al. (2018) for further details.

In this work, we focus on one-sample methods which do not rely on an emulator. One-sample methods can be seen as generalizations of the intuition developed by Pearson (1905) for estimating the correlation ratio. If X^i has a finite support then an input-output sample of (sufficiently large) size n contains repeated observations of $Y = g(x^i, X^{-i})$, for each fixed value $X^i = x^i$, while the other factors, $X^{-i} = (X^1, \ldots, X^{i-1}, X^{i+1}, \ldots, X^k)$, remain random. This allows the estimation of $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y\mid X^i = x^i})$ directly from the given sample. For a continuous X^i, a similar result may be achieved by partitioning the support $\mathcal{X}^i$ of X^i into M bins $\{\mathcal{X}^i_m\}_{m=1}^M$. The point condition $(X^i = x^i)$ is then replaced by the bin condition $(X^i \in \mathcal{X}^i_m)$. Then, for any sensitivity measure encompassed by Eq. (1), a one-sample estimator is given by:

$$\widehat{\xi}_i = \sum_{m=1}^{M} \mathbb{P}_{X^i}(\mathcal{X}^i_m)\, \hat{\zeta}^i_m, \tag{2}$$

where $\hat{\zeta}^i_m$ may be any estimator of $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y\mid X^i \in \mathcal{X}^i_m})$. Note that by using equiprobable partition sets, $\mathbb{P}_{X^i}(\mathcal{X}^i_m)$ reduces to $1/M$ asymptotically. In practice, this partition probability is estimated by the sample proportion, n^i_m/n, where n^i_m denotes the number of realizations for which the i-th input falls within the m-th partition set of its support. Borgonovo et al. (2016, Theorem 2) show that, under mild conditions on the inner operator ζ, a consistent version of the estimator in Eq. (2) can be obtained, if the size M of the partition is chosen as a monotonically increasing function of the sample size n, such that $\lim_{n\to\infty} \frac{n}{M(n)} = \infty$.

Appendix A.1 details the one-sample estimators of the sensitivity measures in Table 1 that will be used for comparison in the remainder. These, as most of the one-sample estimators found in the literature, are constructed either as deterministic approximations or as (frequentist) point estimators. In general, finding asymptotic distributions of the estimators of global sensitivity measures to quantify estimation uncertainty, is not straight forward. Variance-based sensitivity measures, which have have received the most attention in recent years, are a notable exception. Gamboa et al. (2016); Janon et al. (2014a) and Tissot and Prieur (2015), for example, propose pick-and-freeze and replicated Latin hypercube design-based estimators which are asymptotically normal. Error bounds for Sobol' indices were also recently derived by Jiménez Rugama and Gilquin (2018), using pick-and-freeze or replicates, and by Le Gratiet et al. (2014), building on the work of Oakley and O'Hagan (2004), which we discuss below. However, similar results are not available for other sensitivity measures. An alternative solution is the use of bootstrap (see e.g. Archer et al. (1997); Storlie et al. (2009) for variance-based, and Plischke et al. (2013) for moment-independent sensitivity measures, respectively).

A second issue of partition-based one-sample methods is the sample size bias induced by the partition. Quantities related to the marginal distribution of Y are estimated using the full sample size n, but those related to the within bin distribution of $Y|X^i \in \mathcal{X}^i_m$ are estimated using a smaller sample size $n^i_m \approx n/M$. While a sample size correction is implicit in the estimation of variances, the same is not true for the pdf and cdf estimates (see Eqs. (17)–(19) in Appendix). In other words, there is a different granularity when estimating the conditional and the unconditional distributions. In Section 3, we propose two partition-dependent estimators which mitigate the sample size bias, while providing a natural way to quantify the estimation error, allowing interval estimation.

Within the Bayesian paradigm, unknown objects are treated as random, and assigned a prior probability measure which reflects the analyst's uncertainty about their values. In this context, Oakley and O'Hagan (2004) treat the input-output mapping g as unknown

(at least before evaluation). Thus, they define a semi-parametric regression model with a Gaussian process prior, which allows posterior inference on variance-based sensitivity measures for uncorrelated input simulators. In fact, it is possible to calculate posterior means for the conditional and unconditional variance of Y and $Y|X^i$ respectively, either analytically or via numerical integration. The approach eliminates the need for a partition of the covariate space, thus solving the second issue mentioned above. However, the posterior distributions of the variance-based measures (e.g. η_i) are not available analytically, and quantification of the estimation error is not treated. Le Gratiet et al. (2014) extend the approach and propose a multifidelity co-kriging emulator. The Bayesian estimation of the emulator parameters makes it possible to find interval estimates for first order sensitivity measures. This has been applied to the estimation of sensitivity measures based on derivatives and Shapley values by De Lozzo and Marrel (2016) and Benoumechiara and Elie-Dit-Cosaque (2018), respectively. Both Le Gratiet et al. (2014) and Oakley and O'Hagan (2004) define a Gaussian process prior on the ensemble of possible simulator input-output mappings $g(\mathbf{X})$. The proposed estimates are based on the output of the fitted Gaussian emulator, assuming independent model inputs.

In the following sections, we present alternative schemes that allow interval estimation for sensitivity measures, and that are applicable, in principle, to any sensitivity measure. The proposed estimators do not require the fitting of an emulator, nor do they rely on an assumption of independence of the simulator inputs. Opposite to Le Gratiet et al. (2014) and Oakley and O'Hagan (2004) the prior is not set on the unknown g, but on appropriately chosen conditional or joint distribution of the output X and single input X^i. For illustrative purposes, we focus on the estimation of the three sensitivity measures present in Table 1.

3 Non-parametric partition-dependent estimation

Our first proposal can be interpreted as a refinement of the estimators introduced in the previous section and discussed in Appendix A.1, which rely on a partitioning of the input space. In practice, this is tantamount to assuming that the conditional distribution $\mathbb{P}_{Y|X^i=x_i}$ can be well approximated, within each partition, i.e. for every $x^i \in \mathcal{X}^i_m$, by a single distribution $\mathbb{P}^i_m$. Inspired by the Bayesian paradigm, uncertainty about the collection $\{\mathbb{P}^i_m\}_{m=1}^M$ can be expressed through a prior (see e.g. Hjort et al. 2010, for an extensive discussion about flexible priors on families of distributions, their properties and their use). Notice that, since $\mathbb{P}_{X^i}$ is assumed known, the marginal distribution of Y, $\mathbb{P}_Y(y) = \int_{\mathcal{X}^i} \mathbb{P}_{Y|X^i=x^i}(y|x^i)\mathrm{d}\mathbb{P}_{X^i}(x^i)$, is fully determined by $\mathbb{P}_{Y|X^i}$, so no additional uncertainty remains.

For simplicity, we assume that such distributions are independent and identically distributed (i.i.d.), so the problem becomes that of finding a prior which assigns probability 1 to a large enough set of probability distributions supported on $\mathcal{Y}$. We focus our attention on the *Dirichlet Process* (DP), first introduced by Ferguson (1973) and widely studied in the BNP literature (see e.g. Hjort et al. 2010, Chapter 2, for a discussion on its properties). We therefore define, for each $i = 1, \dots k$ the following Bayesian non-parametric model:

$$Y|(\mathbb{P}^i_m, X^i \in \mathcal{X}^i_m) \sim \mathbb{P}^i_m; \quad \mathbb{P}^i_m \overset{iid}{\sim} \mathcal{DP}(\alpha G), \tag{3}$$

where $\mathcal{DP}(\alpha G)$ denotes a Dirichlet process with base measure G and concentration parameter α. The Dirichlet process could be replaced by a more general stick-breaking process, achieving greater flexibility at a similar computational cost (see e.g. Ishwaran and James 2001; Pitman and Yor 1997; Lijoi et al. 2007). In this case, the algorithms and proposed estimators would maintain a similar structure so we focus on the Dirichlet process, without loss of generality, in order to use a notation more familiar to a wider audience. With regards to the hypothesis of independence between the $\mathbb{P}^i_m$, it could be removed through the application of recent developments in BNP methods (see Wood et al. (2011); Teh et al. (2006); Teh and Jordan (2010); Camerlenghi et al. (2017) and Camerlenghi et al. (2019)). This, however, would lead to a complication of the estimation algorithms which goes beyond the scope of this paper.

The posterior for this model, given the simulator input-output realizations, $Data = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, can be written as follows:

$$\begin{aligned} &Y|(X^i \in \mathcal{X}^i_m, \mathbb{P}^i_m) \sim \mathbb{P}^i_m; \\ &\mathbb{P}^i_m|Data \overset{ind}{\sim} \mathcal{DP}\left((\alpha + n^i_m)\widetilde{G}^i_m\right), \end{aligned} \tag{4}$$

where

$$\begin{aligned} \widetilde{G}^i_m &= \mathbb{E}[\mathbb{P}^i_m|Data] \\ &= \frac{\alpha}{\alpha + n^i_m} G + \frac{n^i_m}{\alpha + n^i_m} \sum_{y \in \mathbf{y}^i_m} \frac{1}{n^i_m} \delta^{Dirac}(y). \end{aligned} \tag{5}$$

The sensitivity measures we aim to estimate are functionals of the conditional and marginal distributions. Close form expressions for the posterior distributions of such functionals are not available, making a full Bayesian analysis infeasible. One may alternatively, consider an estimator of ξ_i based on the posterior means

in Eqs. (5) and (24), respectively, taking the form:

$$\widetilde{\xi}_i = \sum_{m=1}^{M} \frac{n_m^i}{n} \zeta(\widetilde{G}, \widetilde{G}_m^i). \tag{6}$$

Unfortunately, the direct calculation of $\widetilde{\xi}_i$ is impractical. Moreover, our purpose is to provide interval estimation, so as to quantify the uncertainty associated to point estimates. A way out is to sample observations (i.e., predicted realizations of the output) from $\widetilde{G}$ and $\widetilde{G}_m^i$, in order to enrich the sample. More specifically, we have a vector $\mathbf{y}$ of n observations from the original simulator used to estimate $\mathbb{P}_Y$, but only n_m^i of these belong to $\mathbf{y}_m^i$ and are therefore used to estimate $\mathbb{P}_m^i$. Because $n_m^i < n$, a sample bias emerges affects the empirical estimation of ξ_i (details in Appendix A.2). By re-sampling from $\widetilde{G}$ and $\widetilde{G}_m^i$ we can enlarge both vectors, making them of the same size and, potentially, arbitrarily large. Our proposal here is simply to sample $n - n_m^i$ observations from $\widetilde{G}_m^i$, thus obtaining two vectors of size n. This procedure is known in the literature as the non-parametric *Bayesian bootstrap* (Bb) (Hjort 1985, 1991). In our case, for each m a sample $\widetilde{\mathbf{y}}_m^i = \{\widetilde{y}_{n_m^i+1}^i, \ldots, \widetilde{y}_n^i\}$ of size $n - n_m^i$ is obtained from $\widetilde{G}_m^i$. A value of $\widehat{\xi}_i^{Bb,s}$ in Eq. (2) can be calculated through any of the methods discussed in Section 2, using $\mathbf{y}$ to estimate all quantities related to the marginal distribution of Y and the extended vector $\mathbf{y}_m^{Bb,i,s} = (\mathbf{y}_m^i, \widetilde{\mathbf{y}}_m^{i,s})$ to estimate all quantities related to the conditional distribution of $Y|X^i \in \mathcal{X}_m^i$. Informally, the weighted average over m can be seen as approximately simulated from the posterior distribution of ξ_i. By repeating this procedure S times, we obtain a Bb sample $\{\widehat{\xi}_i^{Bb,s} : s = 1, 2, ..., S\}$. We propose the Monte Carlo average:

$$\widehat{\xi}_i^{Bb} = \frac{1}{S} \sum_{s=1}^{S} \widehat{\xi}_i^{Bb,s},$$

as a point estimator of ξ_i. The empirical quantiles can provide error bands, to quantify the estimation error. Note that, because each $\widetilde{y}_j^i$ is simulated from a single distribution, $\widetilde{G}_m^i$, the sampling process can be done in parallel and the method is computationally fast. However, the uncertainty is underestimated because the additional variability captured by the posterior distribution of Eq. (4) is ignored.

A more accurate alternative is to sample $\widetilde{\mathbf{y}}_m^i$ jointly from the Dirichlet process posterior distribution (4), instead of sampling each $\widetilde{y}_j^i$ from the posterior mean. This can be done via the *Pólya Urn scheme* (Pu) of Blackwell and MacQueen (1973). Specifically, $\widetilde{\mathbf{y}}_m^i$ is generated as a realization of the Pólya sequence:

$$\widetilde{Y}_{j+1}^i | \left(\widetilde{y}_{n_m^i+1:j}^i, D^i \right) \sim \frac{\alpha}{\alpha + j} G + \frac{j}{\alpha + j} \widehat{\mathbb{P}}_j \quad \forall\, j \geq n_m^i.$$

Once again, the extended samples $\mathbf{y}_m^{Pu,i,s} = (\mathbf{y}_m^i, \widetilde{\mathbf{y}}_m^{i,s})$ can be used to obtain a value $\widehat{\xi}_i^{Pu,s}$ by any available method to calculate the expression in Eq. (2). We use $\widehat{\xi}_i^{Pu}$ to denote the Monte Carlo average of a sample of size S generated in this way. Note that $\widehat{\xi}_i^{Pu}$ has the same expectation as $\widehat{\xi}_i^{Bb}$. However, a greater variability which accounts for the uncertainty on $\mathbb{P}_m^i$ a posteriori, results in wider credibility intervals. The sampling procedure is now sequential for $s = 1, \ldots, S$, so the price for greater accuracy in uncertainty estimation is a slightly higher computational time. Additional technical details for Bb and Pu estimators are presented in Appendix A.3.

We illustrate the performance of these estimators via a toy simulator with correlated inputs for which the sensitivity measures can be calculated analytically (see Table 2). This 21-input simulator is given by

$$Y = \sum_{i=1}^{21} a_i X^i, \tag{7}$$

where $X^i \sim \text{Normal}(1, 1)$, with $a_1 = \cdots = a_7 = -4$, $a_8 = \cdots = a_{14} = 2$, and $a_{15} = \cdots = a_{21} = 1$. The 21 inputs are correlated with $\rho_{i,j} = 0.5$. Therefore, Model inputs with indices in $1, 2, ..., 7$ are the most important, followed by inputs with indices in $15, 16, ..., 21$, followed by inputs with indices in $8, 9, ..., 14$. Table 2 displays the analytical values of the sensitivity measures.

We then study the results of numerical experiments carried out to test the performance of the frequentist estimators and the non-parametric estimators discussed above. Details regarding hyperparameter choices are provided in Appendix A.3 and results are reported in Fig. 1. The graphs in the first, second and third rows display estimates of η_i, δ_i and β_i, respectively, at samples of sizes $n = \{300, 600, 900\}$. Each graph is divided into three blocks displaying estimates obtained with alternative choices of M. For instance, for $n = 300$, we discuss $M = 3, 7, 9$, for $n = 900$ $M = 4, 10, 29$. The dotted lines display the analytical values. For clarity, instead of reporting seven sensitivity measures per group, we show numerical values for a representative of each input group, namely X^3, X^{10} and X^{18}.

Fig. 1 shows that estimates tend to be downward biased at small partition sizes, and upward biased as

Table 2 Analytical values of η_i, δ_i and β_i for the two test simulators used in this section

	21-input			2-input	
Measure	$X_1 - X_7$	$X_8 - X_{14}$	$X_{15} - X_{21}$	X_1	X_2
η_i	0.309	0.064	0.092	0.496	0.496
δ_i	0.212	0.084	0.102	0.315	0.315
β_i	0.205	0.083	0.101	0.289	0.289

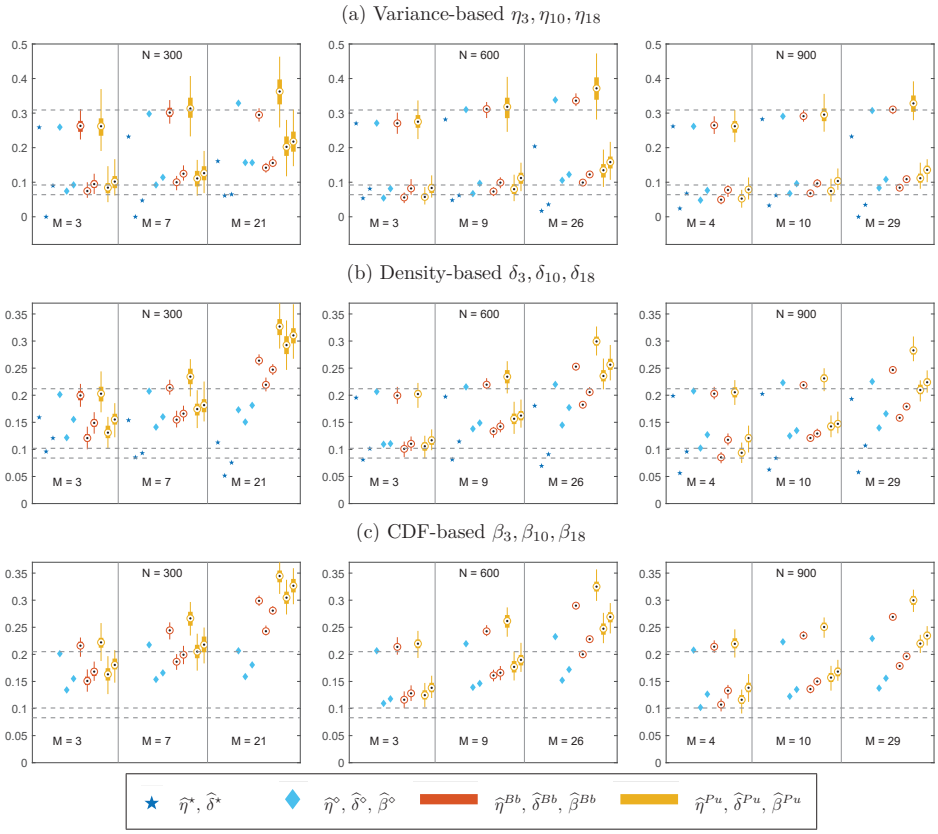


Fig. 1 Results for the correlated 21−input simulator in Eq. (7): Comparison of sensitivity measures estimates using frequentist pdf/cdf-based estimators and partition-dependent non-parametric estimators. The *Bb* and *Pu* estimates are displayed together with the 95% error bands corresponding to empirical quantiles.

the partition size is increased. Also, the change M is higher at small sample sizes and at $n = 300$ we register the highest sensibility of the estimates to the partition size. However, the bias is systematic, that is, it affects all estimates similarly. Consider that in realistic applications, an analyst might not know the true values of the sensitivity measures, and the main interest would be on the ordinal ranking of the inputs. In this respect, for variance-based sensitivity measures (Fig. 1(a)) for all n and M considered, $\widehat{\eta}_i^{\star}$, $\widehat{\eta}_i^{\diamond}$, $\widehat{\eta}_i^{\mathrm{Bb}}$ and $\widehat{\eta}_i^{\mathrm{Pu}}$ are able to correctly identify X^3 as the most relevant variable. Regarding δ_i and β_i (Fig. 1(b) and Fig. 1(c), respectively), one identifies X_3 as the most important input in almost all combinations of sample sizes and partition selections. The ranking becomes unclear for $n = 300$ and $M = 21$. However, this choice would leave about 9 realizations per partition, a number too small to be reasonably chosen by the analyst. For the remaining group of inputs, the overlapping error bands for the *Bb* and *Pu* estimates would not allow us to deem X^{10} more relevant than X^{18}, with either η_i,δ_i or β_i. Thus higher sample sizes would be needed for neatly ranking the second and third most important groups of simulator inputs.

Overall, Fig. 1 suggests that the results display a strong dependence on the partition size M. While *i)* as observed in Strong and Oakley (2013) (see their Fig. 1, at p. 759), the importance of selecting an optimal partition size diminishes as the sample size increases and *ii)* a suboptimal partition selection has in most cases an identical impact on the sensitivity measures (i.e., the sensitivity measures of all inputs are simultaneously upward or downward biased), the analyst is still left with the question of what is the optimal partition size for a given sample. Unfortunately, there seems to be no universally optimal selection rule (see Appendix A.2). Clearly, the problem would be solved if partition-independent estimators were available. In the next section, we study two proposals of Bayesian estimators that avoid the partition choice problem.

4 Bayesian non-parametric partition-free estimation

In this section, we propose two classes of Bayesian partition-free estimators. The first is based on the use of an infinite mixture model to estimate the joint density of Y and X^i. The second, uses a Bayesian non-parametric regression model to estimate the conditional density of Y given X^i.

4.1 Joint density-based estimation

The intuition is that all sensitivity measures under consideration can be recovered from the joint distribution of Y and X^i. Therefore, in order to do Bayesian inference on ξ_i it suffices to place a prior on the joint density $f_{X^i,Y}$. We propose to do so by means of a non-parametric mixture model (see, e.g. Ferguson (1983); Lo (1984)). In other words, we consider $f_{X^i,Y}$ to be defined as a mixture:

$$f_{X^i,Y}(\cdot,\cdot)|P = \int \mathcal{K}(\cdot,\cdot|\theta)\mathrm{d}P(\theta), \tag{8}$$

where $\mathcal{K}$ is a parametric bivariate density and the mixing measure P is a probability distribution over an appropriate space of parameters. The model is completed by assigning a non-parametric prior, Π, on P. Most common choices of Π assign probability one to discrete distributions of the form

$$P = \sum_{\ell=1}^{\infty} w_\ell \, \delta^{Dirac}_{\theta_\ell}, \tag{9}$$

placing mass w_ℓ on locations (θ_ℓ). In the literature, particular attention has been paid to nonparametric priors admitting a stick-breaking construction (Pitman 1996; Sethuraman 1994) where the weights $\underline{w} = (w_1, w_2, ...)$ are defined as realization of random variables satisfying

$$W_1 = V_1, \quad W_\ell = V_\ell \prod_{\ell'=1}^{\ell-1}(1 - V_{\ell'}) \tag{10}$$

and independent of $\underline{\theta} = (\theta_1, \theta_2, ...) \overset{iid}{\sim} G$. Rich families of stick-breaking priors can be defined via different distributional assignments for the sequence $(V_1, V_2, \ldots)$ (Favaro et al. 2012; Ishwaran and James 2001, see, e.g.,). The main advantage over other types of construction is that the stick-breaking representation of the random weights allows for efficient simulation algorithms, specially in the context of nonparametric mixture models (Ishwaran and James 2001; Papaspiliopoulos and Roberts 2008; Kalli et al. 2011; Yau et al. 2011). However, the most popular stickbreaking prior remains the Dirichlet process, well known even outside the specialized community of Bayesian nonparametrics. For this reason, we will focus our analysis on DP mixtures, thus letting $P \sim \Pi = \mathcal{DP}(\alpha G)$. Additionally, for simplicity, we choose $\mathcal{K}$ to be a bivariate normal density, following the density estimation scheme of Escobar and West (1995). In this case, $\theta_\ell = (\mu_\ell, \Sigma_\ell)$ and, to simplify calculations, we select G as a conjugate Normal inverse-Wishart distribution. Thus, the integral in Eq. (8) reduces to a sum and the joint density can be written as:

$$f_{X^i,Y}(\cdot,\cdot)|P = \sum_{\ell=1}^{\infty} w_\ell \cdot \mathcal{N}(\cdot,\cdot|\mu_\ell, \Sigma_\ell), \tag{11}$$

where the weights follow Eq. (10), with $V_i \overset{iid}{\sim} \mathrm{Beta}(1, \alpha)$.

Inference on this model is usually achieved via an MCMC scheme resulting in a sample from the posterior distribution of $f_{X^i,Y}$ given the data (D). In the case of the DP-mixture, the function `DPdensity` from the `R` package `DPpackage` provides an off-the-rack solution. In practice, the MCMC scheme generates, at each iteration $s = 1, \ldots, S$, values $(\underline{w}^s, \underline{\mu}^s, \underline{\Sigma}^s)$ which, substituted in expression (11), produce a density function, $f^{BNJ,s}_{X^i,Y}$. Analytical expressions for the marginal and conditional densities, $f^{BNJ,s}_{Y}$ and $f^{BNJ,s}_{Y|X^i}$ as mixtures of normal distributions are made easily available by the choice of the Gaussian kernel. Clearly, it is also possible to evaluate the corresponding cumulative distribution functions. Thus, it is possible compute the global sensitivity measures of interest, $\eta^{BNJ,s}_i$, $\delta^{BNJ,s}_i$, $\beta^{BNJ,s}_i$ from their definitions (Table 1), obtaining a posterior sample of each. We denote the sample means by $\widehat{\eta}^{BNJ}_i$, $\widehat{\delta}^{BNJ}_i$ and $\widehat{\beta}^{BNJ}_i$, respectively. Approximate credibility intervals can be obtained from the empirical quantiles of the samples. The procedure is summarized Appendix A.3, to which we refer for further details.

It is important to observe that the known marginal distribution for X does not, in general, coincide with the marginal distribution for X derived from each $f^{BNJ,s}_{X^i,Y}$. Thus, by using only the joint density $f_{X^i,Y}$ to estimate the sensitivity measures, important information, standard in global sensitivity analysis is wasted. In fact, inference for conditional densities based in the joint model is known to be approximate (see e.g. Müller and Quintana 2004). In the next section, we present an alternative estimation method which avoids this problem through a recent Bayesian approach to conditional density estimation.

4.2 Conditional density-based estimation

We now propose to use a Bayesian non-parametric regression model to do inference directly on the conditional density of $Y|X^i$, thus using all of the information contained in the data to estimate the relationship between the variables and exploiting the knowledge of the marginal distribution of X to obtain the marginal distribution of Y. The idea is to transform the non-parametric mixture of Eq. (11) into a mixture of conditional densities:

$$f_{Y|X}(y|x) = \int \mathcal{K}(y|x,\theta)\mathrm{d}P_x(\theta), \tag{12}$$

This time a non-parametric prior, Π, is placed on the family, $\{P_x\}_{x\in\mathcal{X}}$ of mixing distributions indexed by x. Analogous to the DP mixture model of the previous section, a dependent DP mixture model or DDP mixture (MacEachern 1999, 2000) is obtained when P_x follows a DP prior, marginally for every x, so that:

$$\mathbb{P}_x(\theta) = \sum_{\ell=1}^{\infty} w_\ell(x)\delta_{\theta_\ell(x)}. \tag{13}$$

The random covariate-dependent weights $W_\ell(x)$ follow the stick-breaking construction of Eq. (10), for i.i.d. random processes $\{V_\ell(x) : x \in \mathcal{X}\}$. In other words, $\mathbf{V}(x) \sim \mathcal{DP}$ for every x. It has been proved sufficient flexibility is achieved through models in which only the particles θ_ℓ or the weights w_ℓ depend on the covariate x (Barrientos et al. 2012), the second option being favoured due to better predictive capabilities. Several proposals have been studied in the literature, focusing on alternative definitions of the random functional weights $w_\ell(x)$ (e.g. Dunson and Park 2008; Griffin and Steel 2006; Dunson and Rodriguez 2011).

The stick-breaking structure of the weights, which imposes a geometric decay, may be bypassed through an alternative construction allowing further flexibility:

$$w_\ell(x) = \frac{\omega_\ell \mathcal{K}(x|\psi_\ell)}{\sum_{\ell'=1}^{\infty}\omega_{\ell'}\mathcal{K}(x|\psi_{\ell'})}. \tag{14}$$

The denominator of this expression is, again, an infinite mixture of parametric kernels, $\mathcal{K}$, this time with support $\mathcal{X}$. Each ω_ℓ can be interpreted as the probability that a realization of Y comes from the ℓ-th regression component regardless of the value of X, just as ω_ℓ is the conditional probability given $X = x$. Such density regression model, where the weights w_ℓ in Eq. (14) follow the stick-breaking representation of Eq. (10) and the extended parameters (θ_ℓ, ψ_ℓ) are i.i.d. from some adequate base measure, G, was proposed by Antoniano-Villalobos et al. (2014), to which we refer the reader for additional details on the role and choice of hyper parameters, as well as the algorithm used for inference.

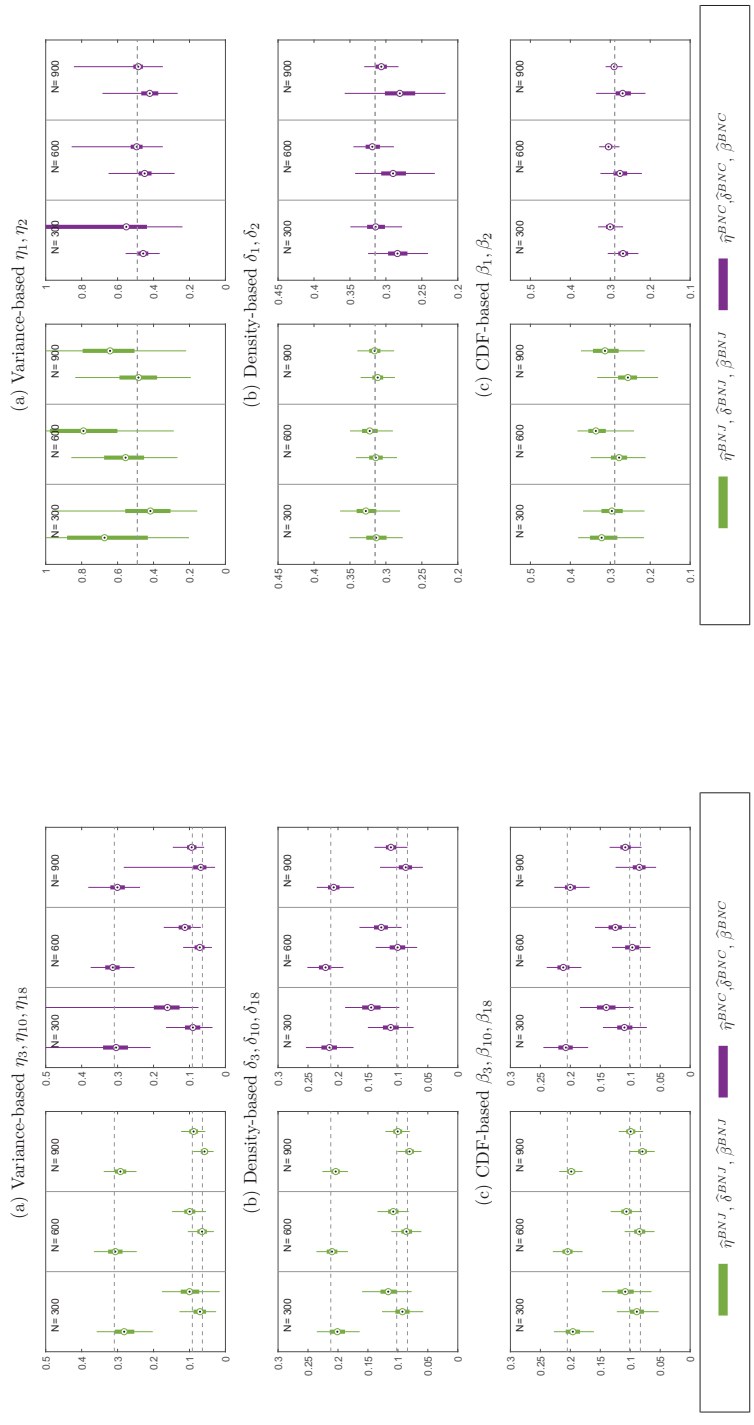
We adopt this construction to estimate the conditional density $f_{Y|X^i}(y|x_i)$ as a mixture of linear regression models:

$$f_{Y|X^i}(y|x^i) = \sum_{\ell=1}^{\infty} w_\ell(x^i)\mathcal{N}(y|a_\ell + b_\ell x^i, \sigma_\ell), \tag{15}$$

where $w_\ell(x_i)$ is given by Eq. (14), with a DP prior. Once again, a MCMC approach is used to generate a sample, this time from the posterior distribution of $f_{Y|X^i}$. Each $f^{BNC,s}_{Y|X^i}(y|x_i), s = 1,\ldots,S$, together with the known marginal for X^i can be used to calculate (e.g. by numerical integration) a corresponding marginal for Y. As discussed in Section 4.1, this is all that is needed to compute the global sensitivity measures of interest, $\eta_i^{BNC,s}$, $\delta_i^{BNC,s}$ and $\beta_i^{BNC,s}$. These, again allow point estimation, e.g. via the Monte Carlo averages, which we denote by $\widehat{\eta}_i^{BNC}$, $\widehat{\delta}_i^{BNC}$ and $\widehat{\beta}_i^{BNC}$, and interval estimation, via empirical quantiles. Section Appendix A.3 summarizes the procedure and offers additional technical details.

4.3 Simulation study

We examine the performance of the classes of partition-independent estimators proposed in Sections 4.1 and 4.2, first via the 21-input simulators with correlated inputs introduced in Section 3. We use the `DPdensity` function from the `R` package `DPpackage` to generate an MCMC posterior sample for BNJ estimators, and the MATLAB© subroutine provided by Antoniano-Villalobos et al. (2014) for BNC estimators. For both joint and conditional density-based estimation, we set a burn-in period as $10n$ and the stored MCMC samples size $S = 1000$. Results are illustrated in Fig. 2(i). The Bayesian non-parametric joint estimators $\widehat{\eta}_i^{BNJ}$, $\widehat{\delta}_i^{BNJ}$ and $\widehat{\beta}_i^{BNJ}$ correctly recover the true values of the parameters and, as the sample size increases from $n = 300$ to $n = 600$, the credibility intervals become narrower. At $n = 900$, there is no more overlap among the three groups of sensitivity measures, allowing the analyst to rank the inputs neatly. Regarding the Bayesian non-parametric conditional estimates, we observe that X_3 is correctly identified as the most relevant input at all sample sizes. The values $\eta_{10}^{BNC}, \delta_{10}^{BNC}$, β_{10}^{BNC} as well $\eta_{18}^{BNC}, \delta_{18}^{BNC}$, β_{18}^{BNC} are overestimated by the BNC estimators at $n = 300$. However, the bias is reduced as n increases and at $n = 900$ the credibility intervals appear centred around the analytical values of the sensitivity measures. We also observe that for both classes of estimates the



(i) Results for input X^3, X^{10}, X^{18} of the 21−input simulator in Eq. (7)

(ii) Results for input X^1, X^2 the 2−input simulator in Eq. (16)

Fig. 2 Comparison of sensitivity measures estimates with 95% credibility intervals using Bayesian non-parametric partition-free joint/conditional estimators. The dashed lines are analytical values of sensitivity measures in Table 2.

analytical value of the sensitivity measures falls within the 95% credible intervals.

For this example joint Bayesian estimators seem to outperform their conditional counterpart. This is to be expected, because the joint Gaussian structure of the data is more easily recovered by the joint model in this case, so the loss due to ignoring the true distribution of X^i has a lesser effect on the results. However, we can appreciate a reassuring improvement of the BNC estimates as the sample size increases. One may argue that, in a situation in which the true conditional distribution of Y given X^i is unknown and may be complex, estimation based on the conditional density model may be preferred, as more robust; the price to pay is that a larger sample size may be required, specially in high-dimensional situations. We then challenge these results for an analytical test case in which the distributions are not normal.

Let us consider the following 2−input simulator:

$$Y = \frac{X^1}{X^1 + X^2}, \tag{16}$$

with X^1,$X^2 \overset{iid}{\sim}$ Gamma$(3, 1)$, so that the output Y follows a Beta distribution.

Assume for the moment that the analyst does not know the true value of the sensitivity measures. In terms of ordinal ranking, Fig. 2(ii) suggests that the two simulator inputs are equally important. The credibility intervals of X_1 and X_2 obtained with both the BNJ and BNC estimators are overlapping at the all sample sizes and for all sensitivity measures, so that the analyst cannot deem one of them more important than the other. The performance of the two estimators is similar. However, note that the credible intervals of the joint model (BNJ) are wider than those for the conditional model especially for variance-based sensitivity measures. As expected, for a non-normal distribution, the joint model presents from the wrongful introduction of the marginal distribution for X_i. We analyze this behavior further in addressing results for the LevelE case study.

5 Case study: LevelE simulator

In this section, we assess the performance of the proposed estimators through the benchmark simulator of sensitivity analysis, LevelE. The setup is that the analyst is studying an available and given sample from this simulator. We recall that the LevelE code simulates the release of radiological dose from a nuclear waste disposal site to humans over geological eras. The code has been developed in an international exercise launched by the Nuclear Energy Agency (NEA) in the mid 1980's Nuclear Energy Agency (1989). Goal of the exercise was the realization of a reference simulator for the prediction of flow and transport of radionuclides in actual geologic formations against which to compare other simulators developed internationally. Since then, LevelE has become a benchmark simulator of sensitivity analysis (Saltelli et al. 2000; Saltelli and Tarantola 2002). During the international exercise, distributions for the uncertain simulator inputs were assessed (Table 3), and have become the reference for analysis on this code. From a technical viewpoint, the LevelE code solves a set of nested partial differential equations that compute the released radiological dose in Sievert/year over a time range of $t = 10,000$ to 2×10^9 years. The detailed equations of the code are reported in Saltelli and Tarantola (2002).

Previous works have discussed the sensitivity analysis of this simulator using alternative sampling methods and sizes. For instance, Saltelli et al. (2000) employ $3,084$ simulator evaluations to obtain point estimates of the first and total order variance-based sensitivity indices. Saltelli and Tarantola (2002) employ $10,000$ simulator runs for the point estimation of first-order variance-based sensitivity indices, a second experiment with $16,384$ runs for the point of the first and total order sensitivity indices according to the design in Saltelli (2002) (no uncertainty in the estimates is provided). In Ratto et al. (2007), stable patterns for the estimation of variance-based sensitivity measures are obtained at a cost of about $1,024$, after the input-output dataset has been used to train an emulator. In Castaings et al. (2012), design based on substituted columns sampling and permuted columns sampling are used, with convergence at about 10^4 runs. Wei et al. (2014) propose a copula-based estimation methods that reduces the cost to about $1,000$ runs for point estimates, with 20 replicates for obtaining confidence intervals. Plischke and Borgonovo (2017) apply a given-data design for the point estimators $\widehat{\eta}_i^{\diamond}$, $\widehat{\delta}_i^{\diamond}$ and $\widehat{\beta}_i^{\diamond}$ using a sample up to size $n = 5,000$, with estimates becoming stable for $n > 1,000$ runs. Thus, a sample of size $n = 1,000$ can be considered state of art for the identification of the key-uncertainty drivers of LevelE. We consider that the analyst has available a sample of size $n = 900$, and studies estimates also at a smaller sample size, namely $n = 600$. Figures 3 and 4 display the results.

The graphs in Fig. 3 report the Bayesian bootstrap and Pólya urn estimators, vis-á-vis the point estimators for variance-based (graphs in row *a*), density-based (row *b*) and cdf-based (row *c*) sensitivity measures. The results show that already at $n = 600$ the two most important simulator inputs are correctly identified. However, the estimates are sensitive to the partition size.

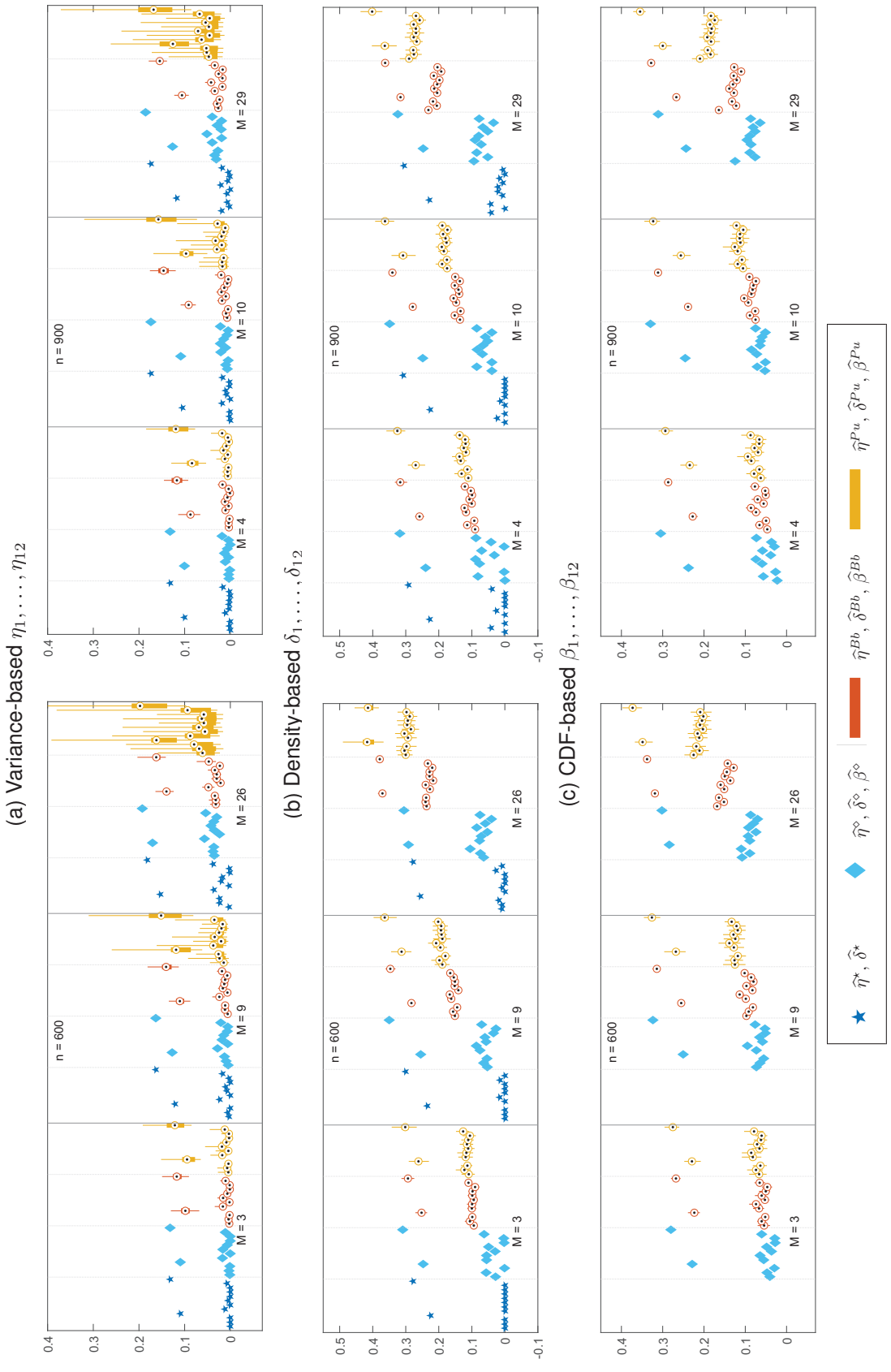


Fig. 3 Results for the LevelE code: Comparison of sensitivity measures estimates using frequintist pdf/cdf-based estimators and Partition-dependent Bayesian non-parametric estimators. Bayesian estimates include 95% credibility intervals.

Table 3 Simulator inputs for the LevelE code. $U(\cdot,\cdot)$ and $LU(\cdot,\cdot)$ stand for the uniform and log-uniform distributions respectively

Input	Meaning	Distribution
X_1	Containment time	U(100,1000)
X_2	Iodine Leach rate	LU($10^{-3}, 10^{-2}$)
X_3	Neptunium chain Leach rate	LU($10^{-6}, 10^{-5}$)
X_4	Iodine retention factor (1st layer)	LU($10^{-3}, 10^{-1}$)
X_5	Geosphere water velocity 1st layer	U(100,500)
X_6	Geosphere Length 1st layer	U(1,5)
X_7	Factor to compute Neptunium retention coefficients Layer 1	U(3,30)
X_8	water velocity in geosphere's 2nd layer	LU($10^{-2}, 10^{-1}$)
X_9	Length of geosphere's 2nd layer	U(50,200)
X_{10}	Retention factor for I (2nd layer)	U(1,5)
X_{11}	Factor to compute Neptunium retention coefficients Layer 2	U(3,30)
X_{12}	Stream flow rate	LU($10^5, 10^7$)

Consider the right-hand side of row *(a)*. The credibility intervals of the variance-based Pólya urn estimators with $M = 26$ are completely overlapping. This signals that, had the analyst chosen such partition size, the estimates would not be meaningful. The separation becomes, instead, clearer at smaller partition sizes with $M = 9$ being possibly the optimal choice. Note that the estimates tend to be upward biased as the partition size increases, in agreement with our previous experiments and also with previous literature findings.

We then come to the joint and conditional partition-independent Bayesian density estimators (Fig. 4). The two graphs in row *(a)* display the estimates and credibility intervals for variance-based sensitivity measures ($\widehat{\eta}_i^{BNJ}$,$\widehat{\eta}_i^{BNC}$), the two graphs in row *(b)* for density-based sensitivity measures ($\widehat{\delta}_i^{BNJ}$, $\widehat{\delta}_i^{BNC}$) and the two graphs in row *(c)* for cdf-based ($\widehat{\beta}_i^{BNJ}$, $\widehat{\beta}_i^{BNC}$) sensitivity measures. Figure 4 shows that the two key-uncertainty drivers are correctly identified already at $n = 600$, by $\widehat{\eta}_i^{BNC}$, $\widehat{\delta}_i^{BNC}$ and $\widehat{\beta}_i^{BNC}$, as the credibility intervals of the associated sensitivity measures separate from the credibility intervals of the remaining simulator inputs. The $\widehat{\delta}_i^{BNJ}$ and $\widehat{\beta}_i^{BNJ}$ correctly identify the two most influential simulator inputs. However, $\widehat{\eta}_i^{BNJ}$ fails to produce meaningful results for variance-based sensitivity measures at either sample sizes. This confirms the results of Section 4.3. The deviation from normality strongly affects the ability of BNJ to capture the conditional density of Y given X_i, since much of the information contained in the data goes into the estimation of unnecessary components of the density mixture of the marginal distribution of X_i. This reduces the estimation precision, leading to the wider confidence intervals.

Let us consider the perspective of an analyst interpreting the results overall. From the available data, the analyst is able to obtain alternative estimators for representatives of three categories of sensitivity measures, with display of credibility intervals. With the exception of $\widehat{\eta}_i^{BNJ}$, the estimators communicate that uncertainty in the simulator response is mostly driven by two simulator inputs, with the remaining ones being of lower significance. Thus, the analyst is allowed to confidently report the key-uncertainty drivers to the decision-maker even if the sample size is limited. At the same time, Fig.s 3 and 4 communicate that the sample is not sufficient to rank the medium and low-important simulator inputs with confidence. If the decision-maker (modeler) wished sharper estimates of the sensitivity measures of these inputs, the analyst would need a larger sample size.

6 Discussion and Conclusions

This work has presented new alternative approaches to the estimation of probabilistic sensitivity measures from a single sample produced for uncertainty quantification. We have studied four classes of estimators. The first two are akin to frequentist partition-based methods currently in use. The estimators are computationally simple to calculate, but they leave the analyst with the problem of choosing the optimal partition. We then introduced two classes of Bayesian estimation procedures based on Bayesian nonparametric density estimation (BNJ and BNC, for short). Their numerical implementation requires a combination of numerical integration and MCMC. However, they eliminate the partition selection problem and provide a way to quantify the estimation uncertainty. Numerical results have shown that the BNJ estimator displays a very good performance when the model input and output distributions are normal or close to normal, while BNC estimators are less sensitive to a normality assumption.

The approaches are based on the intuition of using the available dataset of input-output realizations to obtain a posterior quantification of the uncertainty in the

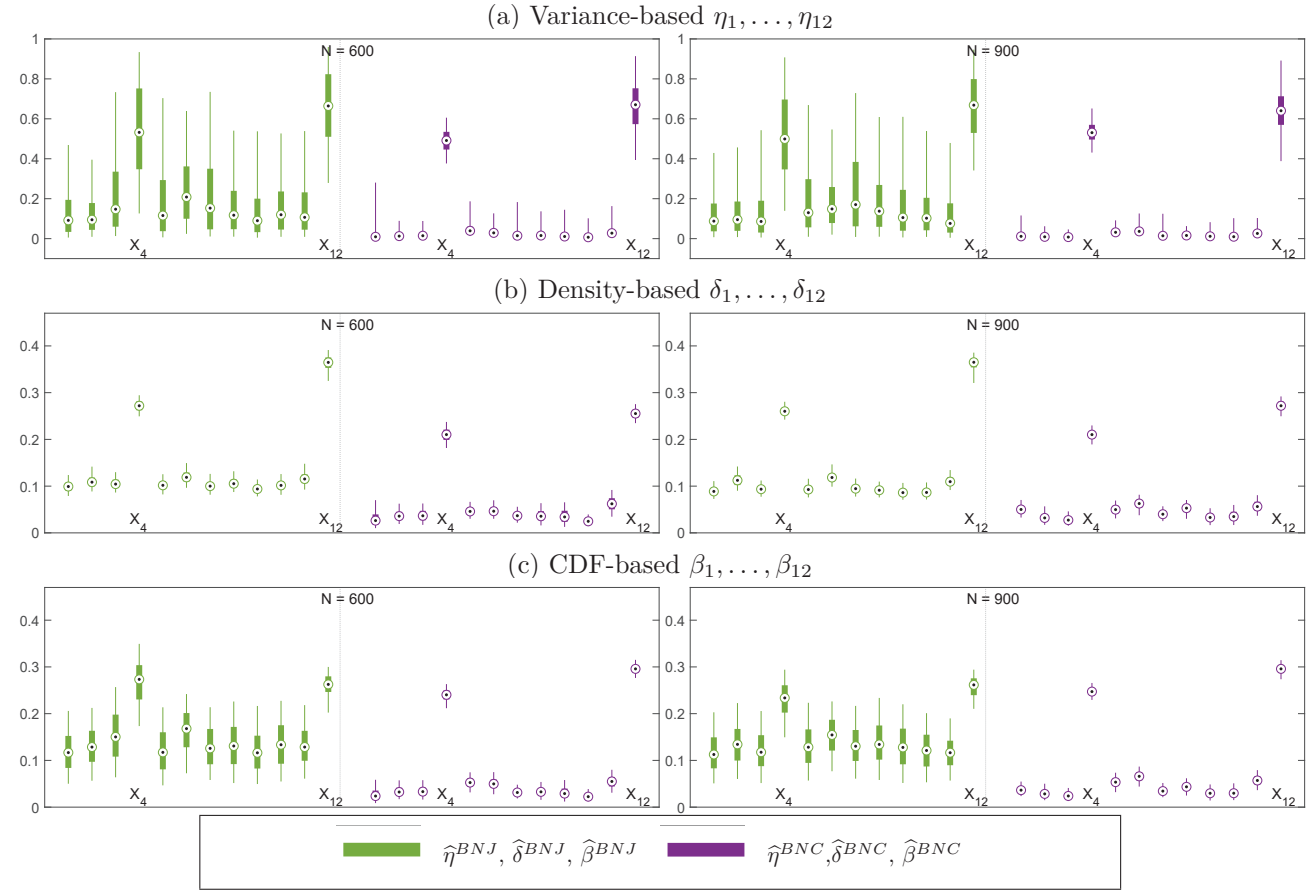


Fig. 4 Results for the LevelE code: Comparison of sensitivity measures estimates with 95% credibility intervals using Bayesian non-parametric partition-free joint/conditional estimators.

estimates of global sensitivity measures. The intervals around the estimates can be used to obtain guidance on whether the sample size is sufficient to infer the simulator input rankings or if additional model runs are needed. One criterion is, for instance, the separation between the error bands or credible intervals of the sensitivity measures for different inputs. If the intervals show a partial or complete overlapping, then the sample size is not sufficient to confidently rank the inputs and a larger sample is needed. Conversely, if some of the inputs stand out, then further runs would not change their ranking but would yield sharper intervals. If the goal of the analysis is to identify the minimum sample size at which inputs can be confidently ranked, one can think of a sequential version of the approach, as a first line of future research following the present work.

The estimators proposed here augment the palette of existing estimators, because they do not require a particular design or additional model runs with respect to several currently available approaches. Thus, an analyst can use them in conjunction with other estimators available in the literature. In particular, estimators for variance-based sensitivity measures have been intensively studied in the literature. We expect that specialized estimation approaches proposed for the case of independent inputs (e.g Le Gratiet et al. 2014, 2017), may outperform the ones proposed here. Indeed, such comparison with two-stage approaches in which an emulator is first fit to the available dataset seems an interesting a second line of future research. However, we emphasized that the estimation methods proposed in this work do not require the independence assumption and allow the calculation of density- and cdf-based sensitivity measures, whose quantification is usually considered challenging. While we have focused on estimators for three well-known global sensitivity measures, the paradigm presented here can be applied to the estimation of any global sensitivity measure, including, among others, value of information, and sensitivity measures based on other discrepancies between densities or cumulative distribution functions.

From a more general perspective, the work shows that combining recent advances in Bayesian non-parametric

density estimation with probabilistic sensitivity analysis in DACE may lead to improvements in the estimation of global sensitivity measures. Research in Bayesian non-parametric density estimation is active in Statistics and Machine Learning, but the advances in this discipline are not directly known to the DACE community. This work represents a first systematic bridge between these two closely related areas of Statistics, and we hope it could favour further research for transferring findings in Bayesian-non parametric estimation to the field of computer experiments. At the same time, exposing Bayesian estimation to the demands coming from probabilistic sensitivity analysis of realistic simulators may challenge state of the art and stimulate further research in Bayesian-non parametric estimation.

Acknowledgements We thank the Associate Editor and two anonymous reviewers for their valuable contributions.

Note The code can be downloaded from: `https://github.com/LuXuefei/Nonparametric-estimation-of-probabilistic-sensitivity-measures`, along with the simulated data to reproduce results.

A Appendix

A.1 Given Data Estimators for the sensitivity measures in Table 1

The one-sample estimator of η_i used here relies on a plug-in estimator of the inner statistic, based on the output sample mean and variance, $\bar{y}$ and s_y^2 respectively, to estimate the marginal mean and variance of Y. The within cluster sample mean $\bar{y}_m^i = \frac{1}{n_m^i}\sum_{y\in\mathbf{y}_m^i} y$ with $\mathbf{y}_m^i = \{y_j : x_j^i \in \mathcal{X}_m^i, j = 1,2,...,n\}$ is used to estimate the conditional mean of $Y|X^i \in \mathcal{X}_m^i$. The final expression (see e.g. Strong et al. 2012) takes the form of Eq. (2) with:

$$\widehat{\eta}_i^\star = \sum_{m=1}^{M} \frac{n_m^i}{n} \frac{(\bar{y}_m^i - \bar{y})^2}{s_y^2}. \tag{17}$$

The one-sample estimator for the $\delta-$importance introduced by Plischke et al. (2013) can be written as:

$$\widehat{\delta}_i^\star = \sum_{m=1}^{M} \frac{n_m^i}{n} \int_{\mathcal{Y}} |\hat{f}_Y^\star(y) - \hat{f}_m^i(y)|\mathrm{d}y, \tag{18}$$

where $\hat{f}_Y^\star$ and $\hat{f}_m^i$ denote kernel-smoothed histograms of the full output vector $\mathbf{y} = (y_1, \dots y_n)$ and the within cluster output vector $\mathbf{y}_m^i$, respectively. The authors propose a quadrature method for the numerical integration required by the L^1-norm in the inner operator, but other solutions could be used, producing similar estimators. Because estimates of this type rely on the approximation or estimation of probability density functions, we refer to them as *pdf-based estimators*.

Plischke and Borgonovo (2017) observe that the kernel-smoothing methods commonly involved in the calculation of pdf-based estimators may induce bias, even at high sample sizes, for simulators with a sparse output. Therefore, they introduce alternative *cdf-based estimators* which rely on the properties of empirical cumulative distribution functions. Scheffé's theorem allows one to write the L^1-distance between two probability density functions in terms of the associated probability functions, as $\int_{\mathcal{Y}} |f_1(y) - f_2(y)|\mathrm{d}y = 2(\mathbb{P}_1(Y \in B) - \mathbb{P}_2(Y \in B))$, where B is the set of values for which $f_1(y) > f_2(y)$. Since B can be written as a union of intervals $(a(t), b(t))_{t=1}^T$, these probabilities can be calculated from the corresponding cumulative distribution functions. Thus, a cdf-based estimator of δ_i can be obtained as:

$$\widehat{\delta}_i^\diamond = \sum_{m=1}^{M} \frac{n_m^i}{n} \sum_{t=1}^{T_m^i} \left(\hat{F}_m^i(\hat{b}_m^i(t)) - \hat{F}_m^i(\hat{a}_m^i(t))\right) - \left(\hat{F}_Y(\hat{b}_m^i(t)) - \hat{F}_Y(\hat{a}_m^i(t))\right). \tag{19}$$

For further details on the estimation of the intervals $(\hat{a}_m^i(t), \hat{b}_m^i(t))$, we refer to Plischke and Borgonovo (2017).

Since β_i is itself a cdf-based sensitivity measure, the definition of a one-sample cdf-based estimator is straightforward:

$$\widehat{\beta}_i^\diamond = \sum_{m=1}^{M} \frac{n_m^i}{n} \max_{j\in\{1,\dots,n\}} \left|\hat{F}_Y(y_j) - \hat{F}_m^i(y_j)\right|, \tag{20}$$

where $\hat{F}_Y$, and $\hat{F}_m^i$ are the empirical cdf's of $\mathbf{y}$ and $\mathbf{y}_m^i$, respectively, i.e.:

$$\hat{F}_Y(y) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(-\infty, y_j]}(y);$$
$$\hat{F}_m^i(y) = \frac{1}{n_m^i} \sum_{y_j\in\mathbf{y}_m^i} \mathbb{1}_{(-\infty, y_j]}(y), \tag{21}$$

and $\mathbb{1}_A(y)$ denotes the indicator function, taking the value 1 if $y \in A$ and 0 otherwise.

Recalling that the expected value of a random variable Y can be calculated as the integral of its survival function, $\mathbb{E}[Y] = \int_{\mathcal{Y}}(1 - F_Y(y))\mathrm{d}y$, a cdf-based one-sample estimator of the variance-based sensitivity measure, η_i is given by:

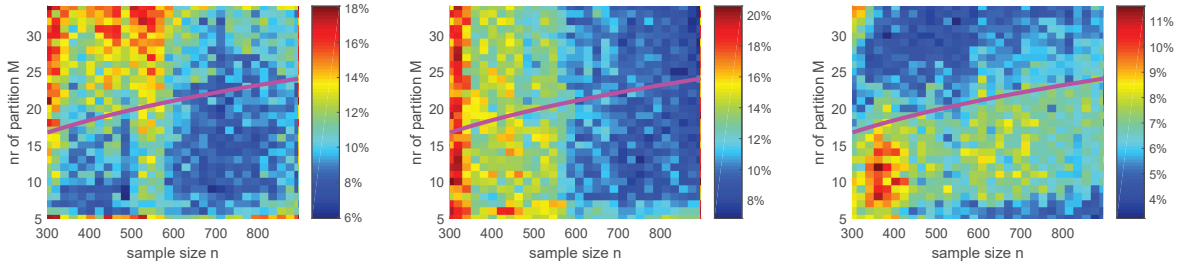
$$\widehat{\eta}_i^\diamond = \sum_{m=1}^{M} \frac{n_m^i}{n} \frac{\left(\int_{\mathcal{Y}} \hat{F}_m^i(y) - \hat{F}_Y(y)\mathrm{d}y\right)^2}{\widehat{\sigma}_Y^2}. \tag{22}$$

Notice that, since the empirical distribution functions are piecewise constant, the integral in the above expression reduces to a sum. Plischke and Borgonovo (2017) propose an efficient way to calculate this integral.

A.2 Numerical experiments for the partition selection problem

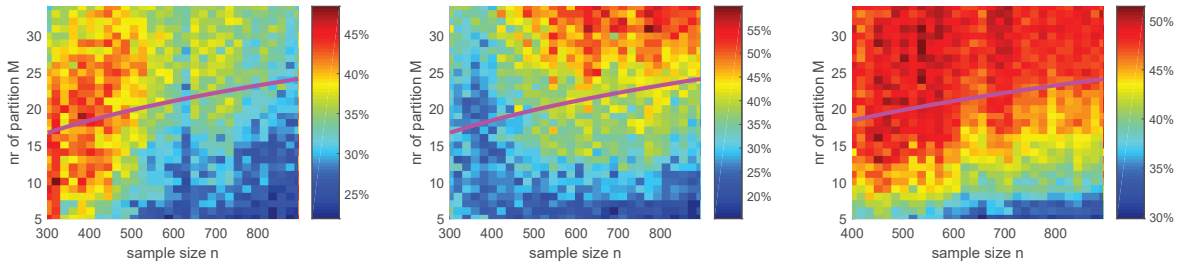
The authors performed several thought experiments on test cases. The results show the difficulty, maybe impossibility, of finding a universally valid rule for linking the partition size M to the sample size n. We report some experiments results.

Assume the analyst wants to find an "optimal "(in some sense) partition refining strategy, i.e., a relationship that produces the partition size M that minimizes the estimation error at sample size n for the pdf-based point estimators $\widehat{\eta}_i^\star$, $\widehat{\delta}_i^\star$ and



(i) Variance-based $\widehat{\eta}_1^\star$ (ii) Density-based $\widehat{\delta}_1^\star$ (iii) CDF-based $\widehat{\beta}_1^\diamond$

Fig. 5 RMSE of sensitivity measures estimates for X^1 of the 2-input simulator in Eq. (16). Magenta lines correspond to $M = 2.5\sqrt[3]{n}$; $n \in [300, 900]$, $M \in [5, 34]$



(i) Variance-based $\widehat{\eta}_3^\star$ (ii) Density-based $\widehat{\delta}_3^\star$ (iii) CDF-based $\widehat{\beta}_3^\diamond$

Fig. 6 RMSE of sensitivity measures estimates for X^3 of the 21-input simulator in Eq. (7). Magenta lines correspond to $M = 2.5\sqrt[3]{n}$; $n \in [300, 900]$, $M \in [5, 34]$

cdf-based point estimator $\widehat{\beta}_i^\diamond$ [Eqs. (17), (18) and (20)]. We focus on one estimator type for simplicity and also because Borgonovo et al. (2016) propose an heuristic inspired by the rule of histogram partitioning of Freedman-Diaconis (Freedman and Diaconis 1981), in which $M \sim \sqrt[3]{n}$.

To evaluate the estimators' performance at fixed values of M and n, we use the Root Mean Square Error (RMSE):

$$\text{RMSE}_i(n) \approx \sqrt{\frac{\sum_{s=1}^{S}\left(\widehat{\xi}_s^i(n) - \xi_i\right)^2}{S}}$$

where S is the number of bootstrap replicates. $\hat{\xi}_{i,l}$ is the l-th bootstrap replicates of ξ_i.

We estimate the sensitivity measures with sample sizes varying from 300 to 900, and partition sizes covering the natural numbers between 5 and 35. Then we calculate the RMSEs with $S = 100$ bootstrap replicates. Figures 5 and 6 present the heatplot of RMSEs in percentage ($\text{RMSE}_i/\xi_i \cdot 100\%$). The horizontal axis indicates the sample size, and the vertical axis the partition size. The darker the color of a region in the plot, the lower the estimation error. For example, in Fig. 5(a), dark (blue) refers to low RMSE (less than 10 percent), and light (red) to relative high RMSE (higher than 14 percent). The magenta line maps n into M using the previously mentioned heuristic function. Fig. 5 shows that the proposed heuristic works well on the 2-input simulator (Eq. (16)), with the magenta line falling mainly into dark coloured regions. However, for the 21-input simulator (Eq. (7)) we would incur in high errors at small sample sizes. For instance consider graph a) in Fig. 6. The graph reports the error in the estimates of δ_3 for the second model. The heuristic would propose values of M at about 20 for all values of n as optimal partition sizes. However, the partition size that minimizes the error is at about $M = 10$ or lower. The different behavior here could also be related to the differences in structure and dimensionality of the models. However, even for the same model, the heatplots differ significantly across the sensitivity measures. For the first simulator (Eq. (16)), the ideal partition size for the variance-based estimator is between 10 to 15 (Fig. 5 (b)), while for mutual information, if falls between 20 to 25 (Fig. 5 (c)).
These results show that aiming at postulating a universally valid heuristic might be a cumbersome task.

A.3 Details on the non-parametric estimators

We present further details regarding the implementation of the Bayesian non-parametric estimation methods in Sections 3 and 4. Inference on the three selected sensitivity measures η_i, β_i and δ_i is performed independently for each $i = 1, \ldots, k$. Therefore, in order to simplify the notation, we will leave out the index i throughout this appendix, considering its value fixed. Throughout this section, all the integrals are approximated numerically using trapezoidal rule, and all the supremes are approximated by the maximum on a predetermined grid over $\mathcal{Y}$.

Partition-dependent Bu and Pu estimation. Note that model 3 is coherent, in the sense that it induces a unique prior over the unconditional distribution of Y, whenever the partitions are equiprobable, that is when $\mathbb{P}(X^i \in \mathcal{X}_m^i) = \frac{1}{M}$ for all $i = 1, 2, ..., k$ and $m = 1, 2, ..., M$. In fact,

$$\mathbb{P}_Y(\cdot|\mathbb{P}_{1:M}^i) = \sum_{m=1}^{M} \mathbb{P}_m^i(\cdot)\mathbb{P}(X^i \in \mathcal{X}_m^i) = \frac{1}{M}\sum_{m=1}^{M} \mathbb{P}_m^i(\cdot).$$

Then, by marginalizing, we obtain

$$\mathbb{P}_Y(\cdot|\alpha G) = \frac{1}{M}\sum_{m=1}^{M}\int \mathbb{P}_m^i(\cdot)\mathrm{d}\mathcal{DP}(\mathbb{P}_m^i|\alpha G) = \int \mathbb{P}(\cdot)\mathrm{d}\mathcal{DP}(\mathbb{P}|\alpha G),$$

because $\int \mathbb{P}_m^i(\cdot)\mathrm{d}\mathcal{DP}(\mathbb{P}_m^i|\alpha G)$ does not depend on i or m. In other words, a priori, $\mathbb{P}_Y \sim \mathcal{DP}(\alpha G)$, so that the prior for

the marginal simulator distribution is also a Dirichlet process. This statement alone, however, provides no information on the probabilistic dependence of Y on X^i. Thus, it is not meaningful, by itself, for a sensitivity analysis.

The posterior of the marginal for Y can be obtained as:

$$\mathbb{P}_Y(\cdot|\alpha G, D^i) = \frac{1}{M}\sum_{m=1}^{M}\int \mathbb{P}^i_m(\cdot)\mathrm{d}\mathcal{DP}(\mathbb{P}^i_m|(\alpha+n^i_m)\widetilde{G}^i_m), \tag{23}$$

which may depend both on i and m. However, the marginal coherence of the model still holds, at least asymptotically. Informally, for an equiprobable partition, $\mathbb{P}(X^i \in \mathcal{X}^i_m) = 1/M$, $n^i_m \simeq n/M$ when the sample size n is sufficiently large, so $\alpha/(\alpha+n^i_m) \simeq M\alpha/(M\alpha+n)$ and $n^i_m/(\alpha+n^i_m) \simeq n/(M\alpha+n)$. Furthermore, $\sum_m (1/n^i_m)\delta_y^{Dirac} \simeq \sum_m (M/n)\delta_y^{Dirac}$. Thus, asymptotically, $\mathbb{P}_Y(\cdot|\alpha G, D^i)$ does not depend on m or i and $\mathbb{P}_Y(\cdot|\alpha G, D^i) \sim \mathcal{DP}((\alpha+n)\widetilde{G})$, where

$$\widetilde{G} = \frac{\alpha}{\alpha+n}G + \frac{n}{\alpha+n}\widehat{\mathbb{P}}_n, \tag{24}$$

and $\widehat{\mathbb{P}}_n$ denotes the empirical distribution of Y based on the full set of observations, $(y_1, \ldots, y_n)$. Note that this is the usual posterior corresponding to the DP prior on $\mathbb{P}_Y$.

For the numerical experiments in Sections 3 and 4.3, the input data, $\mathbf{x}$, are obtained by transforming a 21-dimensional standard Gaussian sample generated via a Halton Quasi random sequence via the Cholesky decomposition of the covariance matrix. Specifically, MATLAB© functions `haltonset` and `chol` are used. The mass parameter, α, for the DP prior is set equal to $0.1n/M$ throughout. The base measure, G, is a Normal distribution with hyper-parameters fixed via an empirical approach, based on the available sample $\mathbf{y}$. $\mathbf{y}^i_m$. Overall, these choices centre the prior distribution for $Y|X^i \in \mathcal{X}^i_m$ roughly around the marginal distribution of Y, thus favouring, a priori, independence between the Y and X^i, with a precision proportional to the number of observations in each partition set. In practical applications, prior information elicited from experts may be expressed through different choices of α and G. Algorithm 1 details the inferential procedure. Note that the calculations of $\eta^{C,s}$, $\delta^{C,s}$ and $\beta^{C,s}$ are equivalent to the pdf-based estimators in Eqs. (17), (18), (20) but with the enriched samples. Alternatively, the cdf-based estimators in Eqs. (22) and (19) could be used for $\eta^{C,s}$ and $\delta^{C,s}$.

Algorithm 1: Partition-dependent Bb and Pu estimation

Input: $C \in \{Bb, Pu\}$, $\{(\mathbf{x}_j, y_j)\}_{j=1}^n$, α,prior distribution G, number of partition sets M, bootstrap size S

Output: $\widehat{\eta}^C, \widehat{\delta}^C, \widehat{\beta}^C, \eta^{C,s}, \delta^{C,s}, \beta^{C,s}$, $s = 1, .., S$

1 Obtain the partition $\{\mathcal{X}_m\}_{m=1}^M$ of $\mathcal{X}$ according to the sample proportion and corresponding $\{\mathbf{y}_m\}$.

2 **for** $s \in \{1, \ldots, S\}$ **do**

3 **for** $m \in \{1, \ldots, M\}$ **do**

4 **if** $C = Bb$ **then**

5 Generate $n - n_m$ new points $\widetilde{\mathbf{y}}^s_m$ from the posterior mean distribution in Eq. (5).

6 **if** $C = Pu$ **then**

7 Generate $n - n_m$ new points $\widetilde{\mathbf{y}}^s_m$ through Pólya urn scheme in Eq. (6).

8 Obtain the extended vector $\mathbf{y}^{C,s}_m = (\mathbf{y}_m, \widetilde{\mathbf{y}}^s_m)$.

9 Calculate

$$\eta^{C,s} = \sum_{m=1}^{M}\frac{n_m}{N}\frac{\left(\bar{y}^{C,s}_m - \bar{y}\right)^2}{s^2_y},$$

where $\bar{y}^{C,s}_m$ is the sample mean of $\mathbf{y}^{C,s}_m$; $\bar{y}$ and s^2_y are the sample mean and variance of $\mathbf{y}$.

10 Calculate

$$\delta^{C,s} = \sum_{m=1}^{M}\frac{n_m}{N}\int_{\mathcal{Y}}|\hat{f}^{\star}_Y(y) - \hat{f}^{C,s}_m(y)|\mathrm{d}y,$$

$$\beta^{C,s} = \sum_{m=1}^{M}\frac{n_m}{N}\sup_{y\in\mathbf{y}^{C,s}_m}\left|\hat{F}_Y(y) - \hat{F}^{C,s}_m(y)\right|,$$

where $\hat{f}^{\star}_Y$ and $\hat{f}^{C,s}_m$ are kernel smoothing functions of $\mathbf{y}$ and $\mathbf{y}^{C,s}_m$, respectively; $\hat{F}_Y$, and $\hat{F}^{C,s}_m$ are the empirical cdf's of $\mathbf{y}$ and $\mathbf{y}^{C,s}_m$, respectively.

11 Calculate the partition-depended estimator of η, δ and β using

$$\widehat{\eta}^C = \frac{1}{S}\sum_{s=1}^{S}\eta^{C,s}, \qquad \widehat{\delta}^C = \frac{1}{S}\sum_{s=1}^{S}\delta^{C,s}, \quad \widehat{\beta}^C = \frac{1}{S}\sum_{s=1}^{S}\beta^{C,s}.$$

Partition-free BNJ estimation. Following the proposal in Jara et al. (2011), we choose G to be a Normal-Inverse Wishart distribution

$$(\mu_\ell, \Sigma_\ell)|(m_1, \gamma, \psi_1) \overset{iid}{\sim} \mathcal{N}(\mu_\ell|m_1, \frac{1}{\gamma}\Sigma) IW(\Sigma_\ell|4, \psi_1), \ell = 1, 2, \ldots$$

where $\mathcal{N}(\cdot|m, A)$ denotes a bivariate normal distribution with mean m and covariance matrix A, and $IW(\cdot|4, \psi)$ denotes an Inverse-Wishart distribution with mean ψ^{-1}. A hyper-prior is assigned to the parameters of the base measure, with hyperparameters determined empirically:

$$\gamma \sim \text{Gamma}\left(\cdot|0.5, 0.5\right),\ m_1|(m_2, s_2) \sim \mathcal{N}(\cdot|m_2, s_2),$$
$$\psi_1|(s_2) \sim IW(\cdot|4, s_2^{-1}),$$

where $\text{Gamma}(\cdot|a_1, a_2)$ denotes the Gamma distribution with mean a_1/a_2. Inference is achieved through the function `DPdensity` from the `DPpackage` in `R`. The output is a MCMC posterior sample $\underline{\theta}^s = (\underline{w}^s, \underline{\mu}^s, \underline{\Sigma}^s)$, $s = 1, \ldots, S$. In practice, the number J_s of components with non-zero weights is finite, thus we have

$$\underline{w}^s = (w_1^s, \ldots, w_{J_s}^s),$$
$$\underline{\mu}^s = (\mu_1^s, \ldots, \mu_{J_s}^s), \quad \underline{\Sigma}^s = (\Sigma_1^s, \ldots, \Sigma_{J_s}^s),$$
$$\text{with} \qquad \mu_\ell^s = \begin{bmatrix} \mu_{1,\ell}^s \\ \mu_{2,\ell}^s \end{bmatrix}, \quad \Sigma_\ell^s = \begin{bmatrix} \sigma_{1,\ell}^s & \sigma_{3,\ell}^s \\ \sigma_{3,\ell}^s & \sigma_{2,\ell}^s \end{bmatrix}. \tag{25}$$

Given the posterior realizations, the corresponding joint density can be obtained:

$$f_{X,Y}^{BNJ,s}(x, y|\underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}(x, y|\mu_\ell^s, \Sigma_\ell^s). \tag{26}$$

By the properties of the bivariate Normal distribution, the marginal and conditional distributions, $f_Y^{BNJ,s}$ and $f_{Y|X^i}^{BNJ,s}$ recpectively, are also mixtures of Normal distributions:

$$f_Y^{BNJ,s}(y|\underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}(y|\mu_{2,\ell}^s, \sigma_{2,\ell}^s),$$
$$f_{Y|x}^{BNJ,s}(y|x, \underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}\left(\cdot|\nu_{2,\ell}^s, \tau_{2,\ell}^s\right) \tag{27}$$

where $\nu_{2,\ell}^s = \mu_{2,\ell}^s + \sigma_{3,\ell}^s(x - \mu_{1,\ell}^s)/\sigma_{1,\ell}^s$ and $\tau_{2,\ell}^s = \sigma_{2,\ell}^s - (\sigma_{3,\ell}^s)^2/\sigma_{1,\ell}^s$. Clearly, the corresponding cdfs, $F_Y^{BNJ,s}$ and $F_{Y|X}^{BNJ,s}$, as well as the marginal mean and variance can be calculated trivially. In particular,

$$\mu_Y^s \coloneqq \mathbb{E}[Y|\underline{\theta}^s] = \sum_{\ell=1}^{J_s} w_\ell^s \mu_{2,\ell}^s,$$
$$V_Y^s \coloneqq \mathbb{V}[Y|\underline{\theta}^s] = \sum_{\ell=1}^{J_s} w_\ell^s \left(\sigma_{2,\ell}^s + \left(\mu_Y^s - \mu_{2,\ell}^s\right)^2\right). \tag{28}$$

For the numerical experiments in Sections 3 and 4.3, the mass parameter is fixed to $\alpha = 1$, and the remaining hyperparameters are set to $m_2 = (\mu_X, \bar{y})$ and $s_2 = \text{diag}(\sigma_X^2, s_y^2)$. The Pseudo-code of BNJ estimation for η, δ and β is illustrated in Algorithm 2.

Algorithm 2: Partition-free joint density-based estimation

Input: $\{(\mathbf{x}_j, y_j)\}_{j=1}^n$, the input distribution f_X, MCMC posterior sample size S

Output: $\widehat{\eta}^{BNJ}, \widehat{\delta}^{BNJ}, \widehat{\beta}^{BNJ}, \eta^{BNJ,s}, \delta^{BNJ,s}, \beta^{BNJ,s}$, $s = 1, .., S$

1 Generate an MCMC posterior sample $\underline{\theta}^s = (\underline{w}^s, \underline{\mu}^s, \underline{\Sigma}^s), s = 1, \ldots, S$ of DP mixture in Eq. (11) through the function `DPdensity` from the `R` package `DPpackage`, where the number J_s of components with non-zero weights is finite, and

$$\underline{w}^s = (w_1^s, \ldots, w_{J_s}^s),$$
$$\underline{\mu}^s = (\mu_1^s, \ldots, \mu_{J_s}^s), \quad \underline{\Sigma}^s = (\Sigma_1^s, \ldots, \Sigma_{J_s}^s),$$
$$\text{with} \qquad \mu_\ell^s = \begin{bmatrix} \mu_{1,\ell}^s \\ \mu_{2,\ell}^s \end{bmatrix}, \quad \Sigma_\ell^s = \begin{bmatrix} \sigma_{1,\ell}^s & \sigma_{3,\ell}^s \\ \sigma_{3,\ell}^s & \sigma_{2,\ell}^s \end{bmatrix}.$$

2 Calculate posterior densities $f_{X,Y}^{BNJ,s}$,the corresponding marginal and conditional distributions $f_Y^{BNJ,s}, f_{Y|X}^{BNJ,s}$ using the following equations:

$$f_{X,Y}^{BNJ,s}(x, y|\underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}(x, y|\mu_\ell^s, \Sigma_\ell^s),$$
$$f_Y^{BNJ,s}(y|\underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}(y|\mu_{2,\ell}^s, \sigma_{2,\ell}^s),$$
$$f_{Y|x}^{BNJ,s}(y|x, \underline{\theta}^s) = \sum_{\ell=1}^{J_s} w_\ell^s \cdot \mathcal{N}\left(\cdot|\nu_{2,\ell}^s, \tau_{2,\ell}^s\right).$$

The corresponding marginal and conditional cdfs $F_Y^{BNJ,s}$ and $F_{Y|X}^{BNJ,s}$ can be obtained trivially.

3 The MCMC samples of η, δ and β can be obtained as follows:

$$\eta^{BNJ,s} \approx \frac{V^s}{V_Y^s};$$
$$\delta^{BNJ,s} \approx \frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{Y}} \left| f_{X,Y}^{BNJ,s} - f_X \cdot f_Y^{BNJ,s} \right| \mathrm{d}y\mathrm{d}x;$$
$$\beta^{BNJ,s} \approx \int_{\mathcal{X}} \sup_{\mathcal{Y}} \left| F_Y^{BNJ,s} - F_{Y|X}^{BNJ,s} \right| f_X \mathrm{d}x, \quad s = 1 \ldots S$$

where

$$V^s = \int_{\mathcal{X}} \left(\sum_{\ell=1}^{J_s} w_\ell^s \frac{\sigma_{3,\ell}^s}{\sigma_{1,\ell}^s} \left(x - \mu_{1,\ell}^s\right) \right)^2 f_X \mathrm{d}x,$$
$$V_Y^s = \sum_{\ell=1}^{J_s} w_\ell^s \left(\sigma_{2,\ell}^s + \left(\mu_Y^s - \mu_{2,\ell}^s\right)^2 \right).$$

4 Point estimates of η, δ and β are obtained as MC averages:

$$\widehat{\eta}^{BNJ} = \frac{1}{S}\sum_{s=1}^{S} \eta^{BNJ,s}, \qquad \widehat{\delta}^{BNJ} = \frac{1}{S}\sum_{s=1}^{S} \delta^{BNJ,s},$$
$$\widehat{\beta}^{BNJ} = \frac{1}{S}\sum_{s=1}^{S} \beta^{BNJ,s}.$$

Partition-free BNC estimation. Following the proposal of Antoniano-Villalobos et al. (2014), we fix $\alpha = 1$ and choose $\mathcal{K}(x|\psi_\ell)$ to be a Normal kernel, with $\psi_\ell = (\mu_\ell, \tau)$. The base measure G is given by:

$$(\mathbf{b}_\ell, \sigma_\ell, \mu_\ell) \overset{iid}{\sim} \mathcal{N}\left(\mathbf{b}_\ell \mid \mathbf{b}_0, \sigma_\ell C^{-1}\right) \text{Gamma}(\sigma_\ell^{-1} \mid 1, 1) \mathcal{N}\left(\mu_\ell \mid \mu_0, (\tau/10)^{-1}\right),$$

where $\mathbf{b}_\ell = (a_\ell, b_\ell)$ and $\tau \sim \text{Gamma}(\cdot \mid 1, 1)$.The hyperparameters are chosen empirically. As an illustration, consider the 21-input simulator. Fig. 7 shows the scatter-plot of $(\mathbf{x}^3, \mathbf{y})$ and the corresponding convex hull, i.e. the smallest convex set containing all points. In this case, we fix $\mathbf{b}_0 = (-1.5, -5.5)$ and $C^{-1} = \text{diag}(43^2, 11^2)$, in order to allow each local linear component to lie between the blue and red lines in the figure, which represent the main behaviour of the data. We use the MATLAB© subroutine provided by Antoniano-Villalobos et al. (2014) to generate an MCMC posterior sample $(\underline{\theta}^s, \underline{\psi}^s) = (\underline{a}^s, \underline{b}^s, \underline{\sigma}^s, \underline{\omega}^s, \underline{\mu}^s, \tau^s)$, $s = 1 \dots S$, where

$$\underline{a}^s = (a_1^s, \dots, a_{J_s}^s), \quad \underline{b}^s = (b_1^s, \dots, b_{J_s}^s), \quad \underline{\sigma}^s = (\sigma_1^s, \dots, \sigma_{J_s}^s),$$
$$\underline{\omega}^s = (\omega_1^s, \dots, \omega_{J_s}^s), \quad \underline{\mu}^s = (\mu_1^s, \dots, \mu_{J_s}^s). \tag{29}$$

Given the a posterior realization $(\underline{\theta}^s, \underline{\psi}^s))$, a conditional density can be obtained from Eqs. (14) and (15):

$$f_{Y|X}^{BNC,s}(y|x, \underline{\theta}^s, \underline{\psi}^s) = \sum_{\ell=1}^{J_s} w_\ell^s(x) \mathcal{N}\left(y|a_\ell^s + b_\ell^s x, \sigma_\ell^s\right). \tag{30}$$

The corresponding marginal pdf $f_Y^{BNC,s}$ of Y is obtained by integrating with respect to the true f_X:

$$f_Y^{BNC,s}(y|\underline{\theta}^s) \approx \int_{\mathcal{X}} f_{Y|X}^{BNC,s} f_X \mathrm{d}x. \tag{31}$$

Clearly, the corresponding marginal and conditional cdfs, $F_{Y|X}^{BNC,s}$ and $F_Y^{BNC,s}$, respectively can be obtained trivially. In particular, posterior realizations of the marginal mean and variance of Y are given by

$$\mu_Y^s := \mathbb{E}[Y|\underline{\theta}^s, \underline{\psi}^s] \approx \int_{\mathcal{Y}} y f_Y^{BNC,s} \mathrm{d}y,$$
$$V_Y^s := \mathbb{V}[Y|\underline{\theta}^s, \underline{\psi}^s] \approx \int_{\mathcal{Y}} (y - \mu_Y^s)^2 f_Y^{BNC,s} \mathrm{d}y$$

The Pseudo-code of BNC estimation for η, δ and β is illustrated in Algorithm 3.

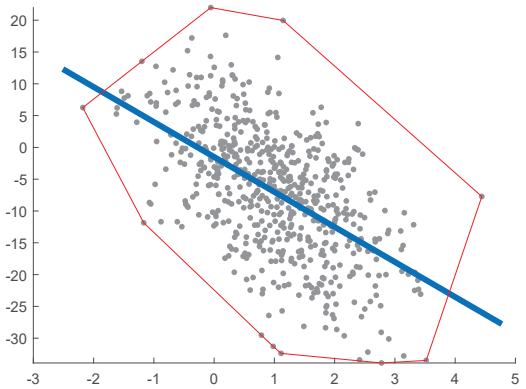


Fig. 7 Scatter plot of Y and X^3 for the 21-input simulator. Red lines constitute the convex hull of $\{(x_1^3, y_1), \dots, (x_n^3, y_n)\}$. The bold blue and red lines are used for prior specification.

Algorithm 3: Partition-free conditional density-based estimation

Input: $\{(\mathbf{x}_j, y_j)\}_{j=1}^n$, the input distribution f_X, MCMC posterior sample size S

Output: $\widehat{\eta}^{BNC}, \widehat{\delta}^{BNC}, \widehat{\beta}^{BNC}, \eta^{BNC,s}, \delta^{BNC,s}, \beta^{BNC,s}$, $s = 1, .., S$

1 Generate an MCMC posterior sample $(\underline{\theta}^s, \underline{\psi}^s) = (\underline{a}^s, \underline{b}^s, \underline{\sigma}^s, \underline{\omega}^s, \underline{\mu}^s, \tau^s), s = 1 \dots S$ using the MATLAB© subroutine provided by Antoniano-Villalobos et al. (2014), where

$$\underline{a}^s = (a_1^s, \dots, a_{J_s}^s), \quad \underline{b}^s = (b_1^s, \dots, b_{J_s}^s), \quad \underline{\sigma}^s = (\sigma_1^s, \dots, \sigma_{J_s}^s),$$
$$\underline{\omega}^s = (\omega_1^s, \dots, \omega_{J_s}^s), \quad \underline{\mu}^s = (\mu_1^s, \dots, \mu_{J_s}^s).$$

2 Calculate the conditional densities $f_{Y|X}^{BNC,s}$, the corresponding marginal pdfs $f_Y^{BNC,s}$ using the following equations:

$$f_{Y|X}^{BNC,s}(y|x, \underline{\theta}^s, \underline{\psi}^s) = \sum_{\ell=1}^{J_s} w_\ell^s(x) \mathcal{N}\left(y|a_\ell^s + b_\ell^s x, \sigma_\ell^s\right),$$
$$f_Y^{BNC,s}(y|\underline{\theta}^s) \approx \int_{\mathcal{X}} f_{Y|X}^{BNC,s} f_X \mathrm{d}x.$$

The corresponding conditional cdfs $F_{Y|X}^{BNC,s}$ and $F_Y^{BNC,s}$ can be obtained trivially.

3 The MCMC samples of the sensitivity measures of interest can be obtained as follows:

$$\eta^{BNC,s} \approx \frac{V^s}{V_Y^s};$$
$$\delta^{BNC,s} \approx \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{Y}} \left| ; f_Y^{BNC,s} - f_{Y|X}^{BNC,s} \right| \mathrm{d}y f_X \mathrm{d}x;$$
$$\beta^{BNC,s} \approx \int_{\mathcal{X}} \sup_{\mathcal{Y}} \left| F_Y^{BNC,s} - F_{Y|X}^{BNC,s} \right| f_X \mathrm{d}x, \quad s = 1 \dots S$$

where

$$\mu_Y^s(x) := \mathbb{E}[Y|x, \underline{\theta}^s, \underline{\psi}^s] = \sum_{\ell=1}^{J_s} \omega_\ell^s(x) \left(a_\ell + b_\ell x\right),$$
$$\widetilde{\mu}_Y^s := \mathbb{E}[\mu_Y^s(X)] \approx \int_{\mathcal{X}} \mu_Y^s(x) f_X \mathrm{d}x,$$
$$V^s = \mathbb{V}[\mu_Y^s(X)] \approx \int_{\mathcal{X}} (\mu_Y^s(x) - \widetilde{\mu}_Y^s)^2 f_X \mathrm{d}x.$$

4 The point estimators of interest are obtained as MC averages:

$$\widehat{\eta}^{BNC} = \frac{1}{S} \sum_{s=1}^{S} \eta^{BNC,s}, \qquad \widehat{\delta}^{BNC} = \frac{1}{S} \sum_{s=1}^{S} \delta^{BNC,s},$$
$$\widehat{\beta}^{BNC} = \frac{1}{S} \sum_{s=1}^{S} \beta^{BNC,s}.$$

References

Antoniano-Villalobos, I., Wade, S., Walker, S.G.: A Bayesian nonparametric regression model with, normalized weights: a study of hippocampal atrophy in in Alzheimer's disease. Journal of the American Statistical Association **109**(506), 477–490 (2014)

Archer, G., Saltelli, A., Sobol, I.: Sensitivity measures, ANOVA-like techniques and the use of bootstrap. Journal of Statistical Computation and Simulation **58**(2), 99–120 (1997)

Barrientos, A.F., Jara, A., Quintana, F.A.: On the support of MacEachern's dependent Dirichlet processes and extensions. Bayesian Analysis **7**, 277–310 (2012)

Baucells, M., Borgonovo, E.: Invariant probabilistic sensitivity analysis. Management Science **59**(11), 2536–2549 (2013)

Benoumechiara, N., Elie-Dit-Cosaque, K.: Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. arXiv preprint arXiv:1801.03300 (2018)

Blackwell, D., MacQueen, J.B.: Ferguson distributions via Pólya urn schemes. The Annals of Statistics **1**(2), 353–355 (1973)

Blatman, G., Sudret, B.: Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. Reliability Engineering & System Safety **95**(11), 1216–1229 (2010)

Borgonovo, E.: A new uncertainty importance measure. Reliability Engineering & System Safety **92**(6), 771–784 (2007)

Borgonovo, E., Hazen, G., Plischke, E.: A common rationale for global sensitivity measures and their estimation. Risk Analysis **36**(10), 1871–1895 (2016)

Borgonovo, E., Iooss, B.: Moment-independent and reliability-based importance measures. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) Handbook of Uncertainty Quantification, pp. 1265–1287. Springer (2017)

Borgonovo, E., Tarantola, S., Plischke, E., Morris, M.D.: Transformation and invariance in the sensitivity analysis of computer experiments. Journal of the Royal Statistical Society Series B **76**(5), 925–947 (2014)

Camerlenghi, F., Lijoi, A., Orbanz, P., Prünster, I.: Distribution theory for hierarchical processes. (2019)

Camerlenghi, F., Lijoi, A., Prünster, I.: Bayesian prediction with multiple-samples information. Journal of Multivariate Analysis **156**, 18–28 (2017)

Castaings, W., Borgonovo, E., Tarantola, S., Morris, M.D.: Sampling strategies in density-based sensitivity analysis. Environmental Modelling & Software **38**, 13–26 (2012)

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

Da Veiga, S.: Global sensitivity analysis with dependence measures. Journal of Statistical Computation and Simulation **85**(7), 1283–1305 (2015)

Da Veiga, S., Wahl, F., F., G.: Local polynomial estimation for sensitivity analysis on models with correlated inputs. Technometrics **51**(4), 452–463 (2009)

De Lozzo, M., Marrel, A.: Estimation of the derivative-based global sensitivity measures using a Gaussian process metamodel. SIAM-ASA Journal on Uncertainty Quantification **4**(1), 708–738 (2016)

Dunson, D.B., Park, J.H.: Kernel stick-breaking processes. Biometrika **95**(2), 307–323 (2008)

Dunson, D.B., Rodriguez, A.: Nonparametric Bayesian models through probit stick-breaking processes. Bayesian Analysis **6**, 145–178 (2011)

Escobar, M.D., West, M.: Bayesian density estimation and inference using mixtures. Journal of the Americal Statistical Association **90**(430), 577–588 (1995)

European Commission: Impact Assessment Guidelines (2009)

Favaro, S., Lijoi, A., Prünster, I.: On the stick-breaking representation of normalized inverse Gaussian priors. Biometrika **99**(3), 663–674 (2012)

Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. The Annals of Statistics **1**(2), 209–230 (1973)

Ferguson, T.S.: Bayesian density estimation by mixtures of normal distributions. In: Rizvi, M.H., Rustagi, J.S., Siegmund, D. (eds.) Recent Advances in Statistics, pp. 287–302. Academic Press (1983)

Freedman, D., Diaconis, P.: On the histogram as a density estimator: L2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **57**(4), 453–476 (1981)

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

Gamboa, F., Klein, T., Lagnoux, A.: Sensitivity analysis based on Cramér von Mises distance. SIAM/ASA Journal on Uncertainty Quantification **6**(2), 522–548 (2018)

Ghanem, R., Higdon, D., Owhadi, H. (eds.): Handbook of Uncertainty Quantification. Springer International Publishing (2016)

Ghosal, S., van der Vaart, A.: Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (2017)

Griffin, J.E., Steel, M.F.J.: Order-based dependent Dirichlet processes. Journal of the American Statistical Association **101**(473), 179–194 (2006)

Hastie, T.J., Tibshirani, R.J.: Generalized Additive Models. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Chapman and Hall/CRC (1990)

He, X.: Rotated sphere packing designs. Journal of the American Statistical Association **112**(520), 1612–1622 (2017)

Helton, J.C., Sallaberry, C.J.: Computational implementation of sampling-based approaches to the calculation of expected dose in performance assessments for the proposed high-level radioactive waste repository at Yucca Mountain, Nevada. Reliability Engineering & System Safety **94**(3), 699–721 (2009)

Hjort, N.L.: Bayesian nonparametric bootstrap confidence intervals. Tech. rep., Stanford University Laboratory for Computational Statistics (1985)

Hjort, N.L.: Bayesian and empirical Bayesian bootstrapping. Tech. rep., Matematisk Institutt, Universitetet i Oslo (1991)

Hjort, N.L., Holmes, C., Müller, P., Walker, S.G. (eds.): Bayesian Nonparametrics. Cambridge University Press (2010)

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

Iman, R.L., Hora, S.C.: A robust measure of uncertainty importance for use in fault tree system analysis. Risk Analysis **10**, 401–406 (1990)

Ishwaran, H., James, L.F.: Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association **96**(453), 161–173 (2001)

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

Janon, A., Nodet, M., Prieur, C.: Uncertainties assessment in global sensitivity indices estimation from metamodels. International Journal for Uncertainty Quantification **4**(1), 21–36 (2014b)

Jara, A., Hanson, T., Quintana, F., Müller, P., Rosner, G.: DPpackage : Bayesian semi- and nonparametric modeling in R. Journal of Statistical Software **40**(5), 1–30 (2011)

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

Kalli, M., Griffin, J.E., Walker, S.G.: Slice sampling mixture models. Statistics and Computing **21**(1), 93–105 (2011)

Kleijnen, J.P.C.: Design and Analysis of Simulation Experiments. Springer, Berlin (2008)

Le Gratiet, L., Cannamela, C., Iooss, B.: A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes. SIAM/ASA Journal on Uncertainty Quantification **2**, 336–363 (2014)

Le Gratiet, L., Marelli, S., Sudret, B.: Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) Handbook of Uncertainty Quantification, pp. 1289–1325. Springer (2017)

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

Lijoi, A., Mena, R.H., Prünster, I.: Controlling the reinforcement in Bayesian non-parametric mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**(4), 715–740 (2007)

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

Lo, A.Y.: On a class of Bayesian nonparametric estimates: I. Density estimates. The Annals of Statistics **12**(1), 351–357 (1984)

MacEachern, S.N.: Dependent nonparametric processes. In: ASA Proceedings-Section on Bayesian Statistical Science American Statistical Association, pp. 50–55. American Statistical Association, Baltimore (1999)

MacEachern, S.N.: Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University (2000)

Mara, T., Tarantola, S.: Variance-based sensitivity indices for models with dependent inputs. Reliability Engineering & System Safety **107**, 115–121 (2012)

Mara, T.A., Joseph, O.R.: Comparison of some efficient methods to evaluate the main effect of computer model factors. Journal of Statistical Computation and Simulation **78**(2), 167–178 (2008)

Marrel, A., Iooss, B., Veiga, S., Ribatet, M.: Global sensitivity analysis of stochastic computer models with joint metamodels. Statistics and Computing **22**(3), 833–847 (2012)

Marrell, A., Iooss, B., Laurent, B., Roustant, O.: Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering & System Safety **94**, 742–751 (2009)

Müller, P., Quintana, F.A.: Nonparametric Bayesian data analysis. Statistical Science **19**(1), 95–110 (2004)

Nuclear Energy Agency: PSACOIN Level E Intercomparison. Tech. rep., Organisation for Economic Co-operation and Development, Paris (France) (1989)

Oakley, J., O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika **89**(4), 769–784 (2002)

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

Papaspiliopoulos, O., Roberts, G.O.: Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. Biometrika **95**(1), 169–186 (2008)

Pearson, K.: On the General Theory of Skew Correlation and Non-linear Regression, vol. XIV of *Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs*. Dulau & Co., London (1905)

Pitman, J.: Random discrete distributions invariant under size-biased permutation. Advances in Applied Probability **28**(2), 525–539 (1996)

Pitman, J., Yor, M.: The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Annals of Probability **25**(2) (1997)

Plischke, E., Borgonovo, E.: Probabilistic sensitivity measures from empirical cumulative distribution functions. Work in Progress (2017)

Plischke, E., Borgonovo, E., Smith, C.L.: Global sensitivity measures from given data. European Journal of Operational Research **226**(3), 536–550 (2013)

Pronzato, L., Müller, W.G.: Design of computer experiments: space filling and beyond. Statistics and Computing **22**(3), 681–701 (2012)

Rahman, S.: The f-sensitivity index. SIAM/ASA Journal on Uncertainty Quantification **4**(1), 130–162 (2016)

Ratto, M., Pagano, A., Young, P.: State dependent parameter metamodelling and sensitivity analysis. Computer Physics Communications **177**(11), 863–876 (2007)

Rényi, A.: On measures of statistical dependence. Acta Mathematica Academiae Scientiarum Hungarica **10**, 441–451 (1959)

Röhlig, K.J., Plischke, E., Bolado Lavin, R., Becker, D.A., Ekstroem, P.A., Hotzel, S.: Lessons learnt from studies on sensitivity analysis techniques in the EU project PAMINA: a benchmark study. In: Reliability, Risk and Safety. Theory and Applications. Proceedings of the ESREL 2009 Annual Conference, vol. 3, pp. 1769–1775. CRC Press, Boca Raton (2009)

Sacks, J., Welch, W., Mitchell, T., Wynn, H.: Design and analysis of computer experiments. Statistical Science **4**(4), 409–423 (1989)

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

Saltelli, A., Tarantola, S.: 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 (2002)

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

Santner, T., Williams, B., Notz, W.: The Design and Analysis of Computer Experiments. Springer New York (2003)

Santner, T.J., Williams, B., Notz, W.: The Design and Analysis of Computer Experiments. Springer, New York, NY, second ed. (2018)

Sethuraman, J.: A constructive definition of Dirichlet priors. Statistica Sinica **4**(2), 639–650 (1994)

Smith, R.L., Tebaldi, C., Nychka, D., Mearns, L.O.: Bayesian modeling of uncertainty in ensembles of climate models. Journal of the American Statistical Association **104**(485), 97–116 (2009)

Sobol', I.M.: Sensitivity estimates for nonlinear mathematical models. Mathematical Modelling & Computational Experiments **1**, 407–414 (1993)

Storlie, C., Swiler, L., Helton, J., Sallaberry, C.: Implementation and evaluation of nonparametric regression procedures

for sensitivity analysis of computationally demanding models. Reliability Engineering & System Safety **94**(11), 1735–1763 (2009)

Storlie, C.B., Helton, J.C.: Multiple predictor smoothing methods for sensitivity analysis: description of techniques. Reliability Engineering & System Safety **93**(1), 28–54 (2008)

Strong, M., Oakley, J.E.: An efficient method for computing partial expected value of perfect information for correlated inputs. Medical Decision-Making **33**(6), 755–766 (2013)

Strong, M., Oakley, J.E., Chilcott, J.: Managing structural uncertainty in health economic decision models: a discrepancy approach. Journal of the Royal Statistical Society: Series C (Applied Statistics) **61**(1), 25–45 (2012)

Teh, Y.W., Jordan, M.I.: Hierarchical Bayesian nonparametric models with applications. In: Hjort, N., Holmes, C., Müller, P., Walker, S. (eds.) Bayesian Nonparametrics: Principles and Practice. Cambridge University Press (2010)

Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical Dirichlet processes. Journal of the American Statistical Association **101**(476), 1566–1581 (2006)

Tissot, J.Y., Prieur, C.: A randomized orthogonal array-based procedure for the estimation of first- and second-order Sobol' indices. Journal of Statistical Computation and Simulation **85**(7), 1358–1381 (2015)

Tuo, R., Wu, J.: Efficient calibration of imperfect computer codes. Annals of Statistics **43**(6), 2331–2352 (2015)

U.S. Environmental Protection Agency: Guidance on the Development, Evaluation, and Application of Environmental Models (2009)

Wei, P., Lu, Z., Song, J.: Moment-independent sensitivity analysis using copula. Risk Analysis **34**(2), 210–222 (2014)

Wong, R.K.W., Storlie, C.B., Lee, T.C.M.: A frequentist approach to computer model calibration. Journal of the Royal Statistical Society. Series B: Statistical Methodology **79**(2), 635–648 (2017)

Wood, F., Gasthaus, J., Archambeau, C., James, L., Teh, Y.W.: The sequence memoizer. Communications of the Association for Computing Machines **54**(2), 91–98 (2011)

Yau, C., Papaspiliopoulos, O., Roberts, G.O., Holmes, C.: Bayesian non-parametric hidden Markov models with applications in genomics. Journal of the Royal Statistical Society. Series B (Statistical Methodology) **73**(1), 37–57 (2011)