International Journal for Uncertainty Quantificatio , 11(6):1–30 (2021)

GLOBAL SENSITIVITY ANALYSIS MEASURES BASED ON STATISTICAL DISTANCES

Souransu Nandi & Tarunraj Singh*

Department of Mechanical & Aerospace Engineering, University at Buffalo, Buffalo, 14260, New York, USA

*Address all correspondence to: Tarunraj Singh, Department of Mechanical & Aerospace Engineering, University at Buffalo, Buffalo, 14260, New York, USA, E-mail: tsingh@buffalo.edu

Original Manuscript Submitted: 7/13/2020; Final Draft Received: 1/31/2021

Global sensitivity analysis aims at quantifying and ranking the relative contribution of all the uncertain inputs of a mathematical model that impact the uncertainty in the output of that model, for any input-output mapping. Motivated by the limitations of the well-established Sobol' indices which are variance-based, there has been an interest in the development of non-moment-based global sensitivity metrics. This paper presents two complementary classes of metrics (one of which is a generalization of an already existing metric in the literature) which are based on the statistical distances between probability distributions rather than statistical moments. To alleviate the large computational cost associated with Monte Carlo sampling of the input-output model to estimate probability distributions, polynomial chaos surrogate models are proposed to be used. The surrogate models in conjunction with sparse quadrature-based rules, such as conjugate unscented transforms, permit efficient calculation of the proposed global sensitivity measures. Three benchmark sensitivity analysis examples are used to illustrate the proposed approach.

KEY WORDS: *global sensitivity analysis, statistical distances, moment-independent sensitivity analysis, polynomial chaos*

1. INTRODUCTION

In various applications of science, engineering, finance and management we frequently come across functional relationships of the form

$$Y = g(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^n$ is a vector of n inputs (i.e., $\boldsymbol{X} = [X_1, X_2, \cdots, X_n]^T$), $Y \in \mathbb{R}$ is a scalar output, and $g : \boldsymbol{X} \rightarrow Y$ is a function which maps the inputs to the output. The nature of g largely varies based on the fiel of study and implementation of the physics of the problem. For example, g could be the output from a dynamic system simulation, the output from a finit element model (FEM) model, a weather model simulation or even the output of a computer code emulating complex systems. It is also interesting to note that the inputs $\boldsymbol{X}$ could represent a wide range of factors. For example, $\boldsymbol{X}$ could represent a simple parameter of the model, an independent input variable, a *trigger* indicating the selection of a particular submodel from a finit dictionary of models, or could even decide on the structure of a model [1]. For any of the above-mentioned cases, it is often desirable to know which of the input parameters ($\boldsymbol{X}$) are most important, or which parameters can be said to be most influentia (with respect to some quantitative metric) in determining the nature of the output Y. All analyses performed to help answer that particular question fall under the domain of a branch of study referred to as sensitivity analysis (SA).

Referring to George Box's famous quote "All models are wrong but some are useful," it is prudent to realize that models are only an approximation of the true physical process and are, hence, invariably plagued by uncertainties. These uncertainties could be due to variations in the parameters of the model, variations in the variables that characterize a stochastic input or disturbance, or variations due to measurement errors. It is assumed in our studies that these uncertainties are reflecte in the inputs ($\boldsymbol{X}$). It is common practice in SA to consider the inputs as random variables or variables with finit support [1]. With the said available information about the inputs, the variation in

the output due to the stochastic nature of the inputs is investigated. A critical assumption is made about the nature of interdependence of the inputs. For the work in this article, it is universally assumed that all contributing inputs are independent of each other. The objective is to quantitatively map the variation in output to the contributing inputs.

An SA study of models can provide valuable information to users. For example, an investment portfolio manager could determine the magnitude of risk of return from individual investments based on the uncertainties and volatility of the portfolio inputs and make decisions accordingly. A weather station could determine which uncertain variable contributes most towards uncertainties in outputs leading to extreme weather events (such as storms or severe precipitation) and instruct more data collection to reduce uncertainties in that particular variable. Both the aforementioned scenarios are examples of reliability-oriented SA which are related to events occurring in the tails of the PDFs of the outputs of interest. In this paper, we look at SA in the context of uncertainty contribution of inputs to the global uncertainty in the output (instead of just the tails). For example, an operations research engineer could determine which input parameters are least significan for a particular cost function, and fi them at nominal values while solving a multivariate optimization problem with reduced model dimensionality (and hence save computational time). Generally, the benefit of SA are manifold and SA has therefore been an active area of research for many years (see [1–7] and references therein).

Methods for SA have been largely classifie into two main categories: local sensitivity analysis (LSA) and global sensitivity analysis (GSA). LSA techniques provide information regarding the effect of inputs on the output only at a specifi location in the input space. These methods are primarily based on the partial derivatives of the output with respect to the inputs [i.e., $(\partial Y/\partial X_i)(\boldsymbol{X} = \boldsymbol{x})$]. However, LSA does not provide adequate information about the nature of variation of Y when the input location is changed or when the inputs are varied simultaneously; i.e., the LSA metrics do not consider the effect of a joint variation of the inputs over the entire input space. This shortcoming has been noted in the literature (see [8]) and has thus motivated the development of GSA techniques. GSA methods observe outputs after considering the variation of inputs over the entire input space to determine quantitative measures. These measures are assigned to each input. The input factors can then be ranked in terms of importance based on these measures to determine which factors are most relevant or are consequently irrelevant. A brief review of the available GSA methods and metrics can be found in a recent review article (see [5]).

Usually determining global sensitivity analysis (GSA) metrics for large number of inputs becomes computationally very expensive. Therefore, as a preliminary test, screening methods such as the method of Morris (see [9]) are adopted to reduce the dimensionality of the relevant inputs. The methods of GSA are then subsequently applied to the lower-dimensional space. Consistent with that paradigm, the methods proposed in this article are also addressed at estimating relative importance of input variables in the lower-dimensional space.

One of the most popular methods for GSA are so-called variance-based methods. In these methods, the idea is to determine the total variance of the output due to the uncertainty in the inputs and to determine the fractional contribution of that variance from each input factor and the combination of the input factors. To this end, quantitative measures such as the Sobol' indices (S_i) and total Sobol' indices (S_{Ti}) have been defined Details about these indices can be found in [8]. However, it has also been recognized in the literature that ranking importance of input variables based on their variance contribution relies on the fact that the variability in the output Y is completely define via the second moment which is often untrue for nonlinear transformations (see [10]). Therefore, it might be inadequate to use variance-based methods to determine relative importance of inputs if moments higher than the second-order ones are significan in the output. Hence, it is recommended that a metric which is dependent on the output distribution, instead of particular moments of the output, would better serve the decision maker [11]. Note that the failure to capture the higher-order influenc between the inputs and the output is not the only identifie limitation of the Sobol' indices. The indices do not consider the stochastic dependence of inputs and evaluating the indices is computationally expensive. Although addressing stochastic dependence of the inputs in the proposed metrics of the paper lies beyond the scope of this work, an efficien process to evaluate them is presented.

With respect to non-moment-based GSA metrics, some work can already be found in the literature. Chun et al. in [12] described a metric which was based on the output cumulative distribution functions (CDFs) from a base case and a sensitivity case where the base case considered the standard output CDF while the sensitivity case considered the output CDF when either the input distribution was changed, uncertainty in the inputs were eliminated, or the uncertain input domain was changed. The statistical distance between the two CDFs was define as the metric. Another GSA

measure was proposed by Park and Ahn in [13] where the Kullback-Leibler (KL) divergence between the standard output probability density function (PDF) and the sensitivity case output PDF was used as the metric. However, it was shown that the KL divergence metric was not define under circumstances where the PDFs had different finit supports [14]. Borgonovo in [11] presented a new measure δ_i based on the PDF of the output Y and the conditional output $Y|X_i$. The metric did not require a sensitivity case like [12,13] and was standalone. Finally, motivated by [11], Gamboa et al. in [15] and Da Veiga in [6] presented a similar GSA metrics based on the Cramér–von Mises distance and other dissimilarity measures, respectively.

Da Veiga [6] and Rahman [7] synthesize sensitivity indices based on the Csiszár f-divergence which subsumes many standard dissimilarity measures such as Kullback-Leibler, Hellinger, and total variation distance. Da Veiga [6] notes that the f-divergence-based sensitivity indices are invariant under any invertible transformation of the random input variables and are conducive to sensitivity analysis for multi-input multi-output stochastic models. Da Veiga [6] introduces a new sensitivity index based on distance correlation which easily generalizes for the multivariate case and another based on the Hilbert-Schmidt independence criterion. Numerous benchmark problems are used to compare the ranking ability of the proposed metrics in relation to traditional ones such as the Sobol' first-orde , Sobol' total, and total variation.

Recognizing that the computational cost of evaluating the f-sensitivity metric is high, Rahman [7] proposed using surrogate approximation using polynomial dimensional decomposition (PDD). The ability of PDD to develop surrogate models for dependent inputs is the motivation for its use in the development of surrogate models relative to polynomial chaos, or stochastic collocation.

Inspired by the above-mentioned articles, this paper presents two types of non-moment-based GSA measures dependent on the PDF of the output and the conditional output as a generalization of [6,11,15]. The two types of measures that are presented in this paper are complementary to each other. For each type of measure, several subtypes of metrics are also define depending on the choice of statistical distances. All of these metrics are independent of hypothesizing a sensitivity case and are standalone (similar to [6,11,15]). The paper also investigates the efficien evaluation of the newly define metrics. The efficien y is improved in two stages. Since Monte Carlo (MC) sampling is used to determine output PDFs, a high-fidelit surrogate model is developed to increase sampling speed and function evaluation time. The surrogate model is developed using polynomial chaos (PC), a well established probabilistic modeling technique. The idea of using PC to do GSA is not new and has been investigated in, e.g., [16–18]. However, those articles investigated variance-based methods (and not moment-independent algorithms). Other modeling tools to increase efficien y in evaluating moment-independent GSA have also been studied (see [19]), but to the authors' best knowledge, it is the firs time PC is being used to evaluate non-moment-based GSA. To determine the said metrics, it is also required to evaluate certain numerical integrals. These integrals are calculated using efficien numerical integration techniques such as Gauss quadrature rules and conjugate unscented transform rules [20]. The proposed approach is numerically illustrated on three benchmark GSA problems (two stationary problems and one time-dependent problem). Comparisons are also made to the total Sobol' indices (S_{Ti}).

The rest of the paper is organized in the following way. In Section 2, we review measures that characterize disparity between probability density functions and present the definition of the proposed class of metrics. In Section 3, we summarize the basics of polynomial chaos and sigma-point-based numerical integration. In Section 4, we present the results of the proposed methods on three illustrative examples before finishin with concluding remarks in Section 5.

2. MOMENT-INDEPENDENT METRICS

Two different classes of metrics are proposed in this section. Both these classes are based on observing the disparity between the output PDF of the function $g(\boldsymbol{X})$ and the conditional output PDF evaluated at certain specifi locations over the input subspace over which the output PDF is conditioned. These disparities are then averaged over that subspace, to derive the fina value of the metrics.

2.1 Statistical Distance Measures

In probability theory, the disparity between two probability measures is often determined using certain metrics called statistical distances. These distances present a quantitative value as an answer to questions such as "How different are

any two probability distributions?" The larger the values of the distance, the more distinct are the PDFs (i.e., the PDFs are further away from each other) and vice versa. To measure disparity, there are several types of distances available, and each type of distance has its advantages and disadvantages. The quantitative value (of distance) changes with the choice of statistical distance used. However, they retain certain basic properties such as non-negativity, symmetry, and positive definiteness A list of important or popular statistical distances ($\mathcal{D}$) is presented in Table 1. The firs and second columns present the name and a convenient abbreviation (used in the rest of the paper) of the different statistical distances, respectively. The fina column presents the mathematical expression, where y is used to represent a realization of the random variable Y, $p_Y(y)$, and $q_Y(y)$ are two distinct PDFs over Y, $P_Y(y)$, and $Q_Y(y)$ are the corresponding CDFs, and Ω_Y is the support of Y (i.e., $Y \in \Omega_Y$). See [21] for details regarding the Wasserstein, Hellinger, total variation, and Kolmogorov distance, [22] for details regarding the Bhattacharya distance, and [23] for details regarding the Cramér–von Mises distance.

Henceforth in this article, $\mathcal{D}_A(p, q)$ is used to represent the statistical distance with abbreviation A. For example, $\mathcal{D}_W$ would refer to the Wasserstein distance.

To illustrate the use of a statistical distance, consider three distinct PDFs $p_Y(y)$, $q_Y(y)$, and $r_Y(y)$ over the same variable Y in Fig. 1. It is intuitive to us that $q_Y(y)$ is closer to $p_Y(y)$ as compared to $r_Y(y)$. This fact is quantifie via $\mathcal{D}$s. For example, by calculating the Wasserstein distance between the PDFs we get

$$\mathcal{D}_W(p, q) = 0.2000 \quad \text{and} \quad \mathcal{D}_W(p, r) = 0.9991. \tag{2}$$

TABLE 1: Statistical distances and their mathematical expressions

Statistical distance ($\mathcal{D}$)	Abbreviation	Description
Wasserstein	W	$\int_{\Omega_y} \lvert P_Y(y) - Q_Y(y)\rvert dy$
Hellinger	H	$\left[2\left(1 - \int_{\Omega_y} \sqrt{p_Y(y)q_Y(y)}dy\right)\right]^{(1/2)}$
Total Variation	T	$0.5\int_{\Omega_y} \lvert p_Y(y) - q_Y(y)\rvert dy$
Kolmogorov	K	$\sup_{y\in\Omega_y} \lvert P_Y(y) - Q_Y(y)\rvert$
Bhattacharya	B	$-\log\left(\int_{\Omega_y} \sqrt{p_Y(y)q_Y(y)}dy\right)$
Cramér–von Mises	C	$\left[\int_{\Omega_y} \lvert P_Y(y) - Q_Y(y)\rvert^2 dy\right]^{(1/2)}$

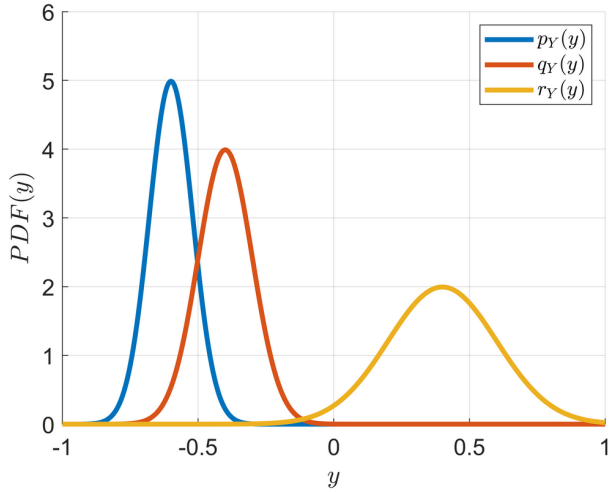


FIG. 1: Probability density functions $p_Y(y)$, $q_Y(y)$, and $r_Y(y)$

In order to understand the variation among the different statistical measures, consider the following exercise. Figure 2 shows two identical uniform distributions [$p_Y(y)$ and $q_Y(y)$] with their supports separated by a value of α. For each statistical distance $\mathcal{D}$, we investigate the dependence on the parameter α. The results are shown in Fig. 3.

Figure 3 helps highlight some of the properties of these measures. We see that the metrics $\mathcal{D}_H$, $\mathcal{D}_T$, and $\mathcal{D}_K$ saturate for values of α greater than 1. This is the region in the y space where $p_Y(y)$ and $q_Y(y)$ have domains distinctly separate with no overlap. Although moving the PDFs further apart increases the disparity between them the effect is not reflecte in the Hellinger distance $\mathcal{D}_H$, the total variation distance $\mathcal{D}_T$, and the Kolmogorov distance $\mathcal{D}_K$. Consequently, beyond a certain point, distinguishing between PDFs using $\mathcal{D}_H$, $\mathcal{D}_T$, and $\mathcal{D}_K$ is infeasible. For this particular case, we see a linear growth for $\mathcal{D}_T$ and $\mathcal{D}_K$ before saturation since the example considers only uniform distributions. However, since $\mathcal{D}_H$ involves nonlinear mapping of the PDFs, we get a nonlinear growth of the distance with α. The Bhattacharya distance $\mathcal{D}_B$ is an interesting metric since it is only define for PDFs which have overlapping domains and does not exist if the domains do not intersect (since the product of the PDFs yields 0 and evaluation of the metric requires taking a logarithm of the product). However, it increases exponentially with α until the domains separate. A feature of this metric is that it has a higher penalty for PDFs which are further apart than the other measures. Hence, the $\mathcal{D}_B$ is useful for distinguishing two PDFs which are almost equally far from a third PDF (unlike $\mathcal{D}_H$, $\mathcal{D}_T$, and $\mathcal{D}_K$). It should also be pointed out that in this paper, since the distances are being used

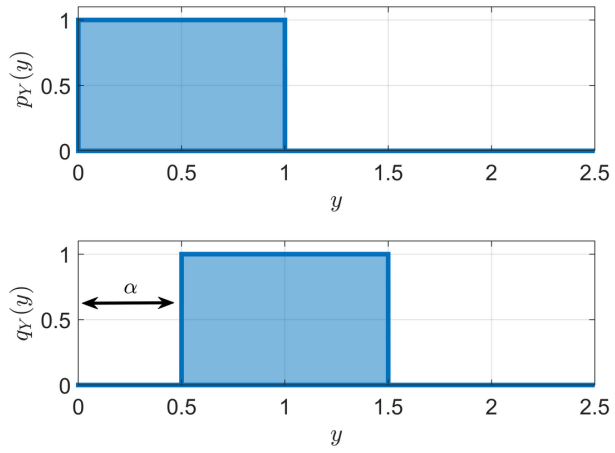


FIG. 2: Probability density functions $p_Y(y)$ and $q_Y(y)$ (uniform distributions)

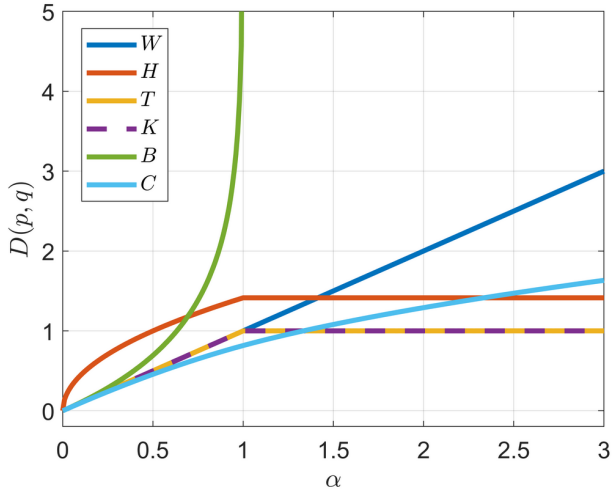


FIG. 3: Variation of $\mathcal{D}$ with α for uniform distributions

to quantify the disparity between output PDFs and conditional output PDFs, there is always an overlap in domain between them. Hence, the $\mathcal{D}_B$ always remains relevant and $\mathcal{D}_H$, $\mathcal{D}_T$, and $\mathcal{D}_K$ face the saturation issue only in the limit. The metrics $\mathcal{D}_W$ and $\mathcal{D}_C$ belong to a class of $\mathcal{D}$s of the form

$$\mathcal{D}(p,q) = \left[\int_{\Omega_y} |P_Y(y) - Q_Y(y)|^r dy\right]^{1/r}, \tag{3}$$

which is the rth norm of the difference between the CDFs. The choice $r = 1$ yields $\mathcal{D}_W$ while $r = 2$ yields $\mathcal{D}_C$. The infinit norm (i.e., $r \to \infty$) leads to $\mathcal{D}_K$. Values of r between 2 and ∞ yield distances with properties ranging from $\mathcal{D}_K$ to $\mathcal{D}_C$.

Although it appears that $\mathcal{D}_W$ varies in a manner identical to $\mathcal{D}_T$ and $\mathcal{D}_K$ until $\alpha = 1$, the result is only an artifact of the chosen example of uniform distributions. Hence, we present a similar exercise with two Gaussian distributions [$p_Y(y)$ and $q_Y(y)$] in Fig. 4 whose means are now separated by α. The variation of $\mathcal{D}$ with α is shown in Fig. 5.

It is evident from Fig. 5 that $\mathcal{D}_W$ is no longer identical to $\mathcal{D}_T$ and $\mathcal{D}_K$ and is therefore dependent on the distributions. $\mathcal{D}_H$, $\mathcal{D}_T$, and $\mathcal{D}_K$, however, still show saturation in the limit; i.e., sensitivity to change in the mean distance goes to zero asymptotically. The growth of $\mathcal{D}_W$ is linear in both the exercises since we do not consider a change in

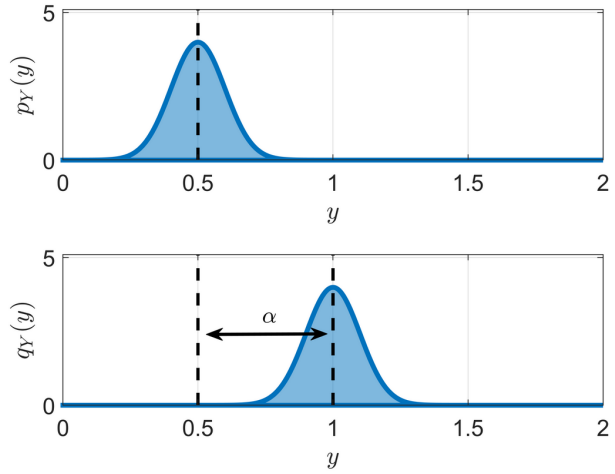


FIG. 4: Probability density functions $p_Y(y)$ and $q_Y(y)$ (normal distributions)

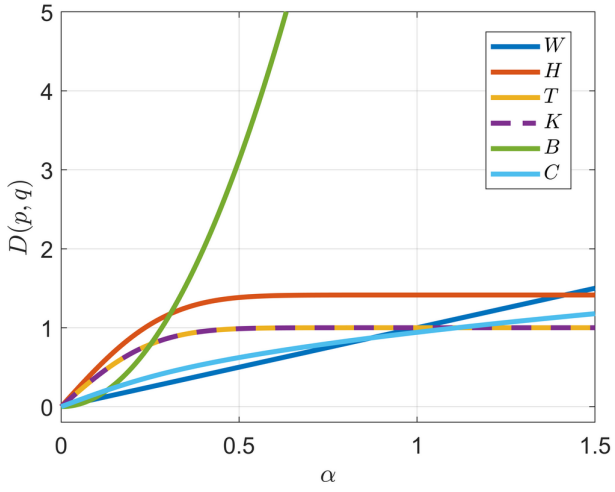


FIG. 5: Variation of $\mathcal{D}$ with α for normal distributions

the shape of the distributions. If the shape of the distributions changes with α, $\mathcal{D}_W$ would no longer grow linearly. $\mathcal{D}_B$ once again increases exponentially and returns very high values of disparity once $p_Y(y)$ and $q_Y(y)$ move apart. Finally, similar to the previous example, note that varying r from 1 to ∞ moves the distance curves from $\mathcal{D}_W$ (dark blue) to $\mathcal{D}_K$ (violet).

It is interesting to observe that for both the aforementioned examples, $\mathcal{D}_T$ and $\mathcal{D}_K$ yield identical results. This is once again an artifact of the chosen distributions and is not a property. This fact is illustrated with the following example (adopted from [24]): Consider two more distributions,

$$p_Y(y) = \mathcal{N}(0,1) \text{ and} \tag{4}$$

$$q_Y(y) = 0.7\mathcal{N}(0,1) + 0.15\mathcal{N}(2.35,1) + 0.15\mathcal{N}(-2.35,1), \tag{5}$$

where $\mathcal{N}(a,b)$ represents the Gaussian distribution with mean a and variance b. It can be shown that for these two PDFs illustrated in Fig. 6, the metrics are

$$\mathcal{D}_T(p,q) = 0.2 \text{ and } \mathcal{D}_K(p,q) = 0.1. \tag{6}$$

Although this paper presents some of the basic properties of $\mathcal{D}$s, a more detailed description about their properties, merits, and demerits can be found in [21–25]. We present at least six different measures for GSA. The measures are different based on the nature of their penalty with disparity, whether the measures saturate and computational effort is required to compute them. The user is provided a choice to select any one of the listed measures as per requirement and convenience.

The $\mathcal{D}$s can now be used to defin certain global sensitivity metrics by comparing disparities between output PDFs and conditional output PDFs. Depending on the nature of conditional PDFs, the metrics are divided into two classes. Class 1 considers the conditional output PDF $f_{Y|X_i}(y,x_i)$ which is the PDF of the output Y for a given fi ed input value of $X_i = x_i$. Class 2, on the other hand, considers the complementary conditional PDF $f_{Y|\tilde{\boldsymbol{X}}_i}(y,\tilde{\boldsymbol{X}}_i)$ which is the PDF of the output Y given all fi ed values of $X_j = x_j$, for all $j = 1,...,n$ with $j \neq i$, and where we used the notation $\tilde{\boldsymbol{X}}_i = [X_1,...,X_j,...,X_n]^T$ for all $j \neq i$. Details about the metrics will be elaborated in the following subsections.

2.2 Class 1 Metrics

Consider the averaged statistical distance over the input space Ω_{X_i}, define as

$$^1\overline{\mathcal{D}}_i := \int_{\Omega_{X_i}} \mathcal{D}(f_Y(y), f_{Y|X_i}(y,x_i)) f_{X_i}(x_i) dx_i, \tag{7}$$

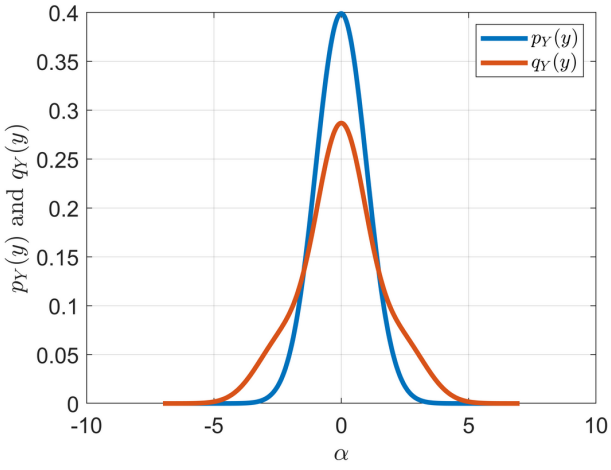


FIG. 6: Probability density functions $p_Y(y)$ and $q_Y(y)$ to show $\mathcal{D}_T$-$\mathcal{D}_K$ disparity

where Ω_{X_i} represents the sample space of the input random variable X_i. In Eq. (7), the output PDF f_Y is define as

$$f_Y(y) := \int_{\Omega_{\boldsymbol{X}}} f_{\boldsymbol{X},Y}(\boldsymbol{x}, y) d\boldsymbol{x}, \tag{8}$$

with $\Omega_{\boldsymbol{X}}$ the sample space of $\boldsymbol{X}$ and $f_{\boldsymbol{X},Y}(\boldsymbol{x}, y)$ the joint input-output PDF, and the conditional PDF $f_{Y|X_i}$ is define as

$$f_{Y|X_i}(y, x_i) := \int_{\Omega_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}} f_{Y,\tilde{\boldsymbol{X}}_{\boldsymbol{i}}|X_i}(\boldsymbol{x}, y) d\tilde{\boldsymbol{x}}_{\boldsymbol{i}}, \tag{9}$$

with $\Omega_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}$ the sample space of $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}$ and $f_{Y,\tilde{\boldsymbol{X}}_{\boldsymbol{i}}|X_i}(\boldsymbol{x}, y)$ the joint input-output PDF for a given fi ed value of X_i.

The distance $\mathcal{D}(f_Y(y), f_{Y|X_i}(y, x_i))$ in Eq. (7) quantifie the disparity between the output PDF and the output PDF conditioned on a single input parameter. If the particular input X_i is unimportant, it would contribute minimally to the marginalized output PDF, which would mean $f_Y(y)$ and $f_{Y|X_i}(y, x_i)$ are in close proximity, resulting in low values of $\mathcal{D}$. In contrast, if the input parameter X_i is indeed influential it would contribute significantl to the output PDF. This means that the conditioned PDF $f_{Y|X_i}(y, x_i)$ would be far from the marginalized PDF $f_Y(y)$ leading to higher values of $\mathcal{D}$. Hence, on observing what the values of $\mathcal{D}$ are on average, one can estimate the relative importance of the inputs.

Considering that the values of $\mathcal{D}$ can vary largely depending on the chosen distance measure, a normalization of ${}^1\overline{\mathcal{D}}_i$ is done to derive the fina definitio of the moment-independent sensitivity index of Class 1, denoted by ${}^1NS$, and we defin

$${}^1NS_i := \frac{{}^1\overline{\mathcal{D}}_i}{\sum_{j=1}^{n} {}^1\overline{\mathcal{D}}_j}. \tag{10}$$

The quantity ${}^1NS_i$ varies inside the interval $[0, 1]$ and can now be used to rank the input parameters. The larger the values of ${}^1NS_i$, the more significan is the input variable X_i. It should be noted that the Borgonovo metric δ_i from [11] is a specifi instance of ${}^1\overline{\mathcal{D}}_i$ where the statistical distance is chosen as the total variation distance $\mathcal{D}_T$. Similarly, the Gamboa distance in [15] is a specifi instance of ${}^1\overline{\mathcal{D}}_i$ where the statistical distance is the Cramér–von Mises distance $\mathcal{D}_C$.

A value of ${}^1NS_i = 0$ implies ${}^1\overline{\mathcal{D}}_i = 0$. Observing Eq. (7), it is evident that this is possible only when the product of $\mathcal{D}$ and f_{X_i} is 0 everywhere in Ω_{X_i}, since $\mathcal{D}$ and f_{X_i} are both non-negative quantities. For the subspace where $f_{X_i} = 0$, it is trivial to note that the output is independent of X_i since X_i does not instance. For the subspace where $\mathcal{D} = 0$, the implication is that the PDFs $f_Y(y)$ and $f_{Y|X_i}(y, x_i)$ are identical. Identical output and conditional output PDFs point to the fact that fixin X_i has no impact on the output, suggesting $g(\boldsymbol{X})$ is independent of X_i. Hence, a value of ${}^1NS_i = 0$ or ${}^1\overline{\mathcal{D}}_i = 0$ immediately suggests that Y is independent of X_i. Similarly, a value of ${}^1NS_i = 1$ implies ${}^1\overline{\mathcal{D}}_j = 0$ for all $j \neq i$. This indicates that $g(\boldsymbol{X})$ is independent of all X_j for $j \neq i$ and is solely dependent on X_i.

2.3 Class 2 Metrics

Similar to Class 1 metrics, consider the averaged statistical distance over the space $\Omega_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}$, define as

$${}^2\overline{\mathcal{D}}_i := \int_{\Omega_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}} \mathcal{D}(f_Y(y), f_{Y|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(y, \tilde{\boldsymbol{x}}_{\boldsymbol{i}})) f_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(\tilde{\boldsymbol{x}}_{\boldsymbol{i}}) d\tilde{\boldsymbol{x}}_{\boldsymbol{i}}, \tag{11}$$

where $f_{Y|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(y, \tilde{\boldsymbol{x}}_{\boldsymbol{i}})$ is the output PDF for given fi ed values for $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}$, define as

$$f_{Y|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(y, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}) = \int_{\Omega_{X_i}} f_{Y,X_i|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(\boldsymbol{x}, y) dx_i, \tag{12}$$

with $f_{Y,X_i|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(\boldsymbol{x}, y)$ the joint input-output PDF for given fi ed values for $\boldsymbol{X}_i$.

If X_i is an important input variable, then the conditioned PDF $f_{Y|\tilde{\boldsymbol{X}}_i}(y, \tilde{\boldsymbol{x}}_i)$ would be close to the output PDF $f_Y(y)$ since most of the statistics of Y has been contributed by X_i. This means that $\mathcal{D}$ would predict a low disparity value for the said PDFs. Averaging all the $\mathcal{D}$s over the input subspace $\Omega_{\tilde{\boldsymbol{X}}_i}$ would now indicate how close the output and the conditioned output PDFs are on average. Similarly, for irrelevant inputs, ${}^2\overline{\mathcal{D}}_i$ would give a higher number. Once again, the ${}^2\overline{\mathcal{D}}_i$ are normalized to obtain the moment-independent sensitivity index of Class 2, denoted by ${}^2NS$, define as

$$ {}^2NS_i := \frac{{}^2\overline{\mathcal{D}}_i}{\sum_{j=1}^{n} {}^2\overline{\mathcal{D}}_j}. \tag{13} $$

The quantity ${}^2NS_i$ also varies inside the interval $[0, 1]$. However, contrary to ${}^1NS_i$, important variables have values close to 0 while irrelevant variables have values closer to 1.

Using arguments similar to the previous index, a value of ${}^2NS_i = 0$ implies that the output and the conditional output PDFs $f_Y(y)$ and $f_{Y|\tilde{\boldsymbol{X}}_i}(y, \tilde{\boldsymbol{x}}_i)$ are identical. This is possible only if the entire variation in the output is derived from the variation in X_i and fixin $\tilde{\boldsymbol{X}}_i$ has no impact on Y. Hence, a value of ${}^2NS_i = 0$ or ${}^2\overline{\mathcal{D}}_i = 0$ immediately suggests that Y is only dependent on X_i. A value of ${}^2NS_i = 1$ indicates that the sum of all ${}^2\overline{\mathcal{D}}_j$ for $j \neq i$ is 0. This is only possible if $g(\boldsymbol{X})$ is only dependent on X_j where $j \neq i$. As a result, ${}^2NS_i = 1$ implies that the output is independent of X_i.

2.4 Specifi Relations between $\overline{\mathcal{D}}_i^T$, $\overline{\mathcal{D}}_i^K$, and $\overline{\mathcal{D}}_i^H$

It is well known that the total variation distance, the Kolmogorov distance, and the Hellinger distance can be described by the following linear relations; see [21]:

$$ \mathcal{D}_K(P, Q) \leq \mathcal{D}_T(P, Q), \tag{14} $$

$$ \mathcal{D}_T(P, Q) \leq \sqrt{2}\mathcal{D}_H(P, Q). \tag{15} $$

Based on these inequalities, we can establish similar inequalities for the $\overline{\mathcal{D}}_i$ metric. Following the definitio of $\overline{\mathcal{D}}_i$, it is evident that

$$ \overline{\mathcal{D}}_i^K \leq \overline{\mathcal{D}}_i^T, \tag{16} $$

and

$$ \overline{\mathcal{D}}_i^T \leq \sqrt{2}\overline{\mathcal{D}}_i^H. \tag{17} $$

Note that the relationships hold for both classes of metrics. Results obtained in the numerical examples (presented later) are seen to be consistent with the inequalities (16) and (17).

3. EFFICIENT EVALUATION OF MOMENT-INDEPENDENT SENSITIVITY INDICES

We often encounter problems in engineering where a single function evaluation is itself computationally very expensive. For those functions, deriving the aforementioned GSA measures can become impractical, especially when output PDFs and conditioned output PDFs need evaluation. Either analytical expressions for these PDFs do not exist, or they cannot be evaluated, and hence they must be approximated from a large number of sample realizations. This makes the calculation of the NS metrics through traditional techniques, such as the Monte Carlo (MC) method, extremely difficult This section presents methods to reduce the computational burden and provide tractable alternatives to closely approximate the NS metrics.

3.1 Polynomial Chaos Based Surrogate Model

Polynomial chaos is a probabilistic modeling tool to approximate a stochastic function with a polynomial function in terms of random variables. First introduced by Norbert Wiener in [26] to expand a Gaussian process in terms of an infinit series involving Hermite polynomials, PC methods have subsequently been investigated in [27–29]. As

an uncertainty quantificatio tool, PC has been used extensively in the literature to determine statistics of random processes as well as to develop surrogate models for complex stochastic systems (static as well as dynamic). In this paper, we determine a surrogate model $\hat{Y}$ for the true system $Y := g(\boldsymbol{X})$ such that instead of sampling Y directly, which can be expensive, we can sample $\hat{Y}$, relatively cheaply.

From PC theory, it is well known that a stochastic function Y can be written as an infinit polynomial series expansion in the form

$$Y = \sum_{i=0}^{\infty} \rho_i \Phi_i(\boldsymbol{X}), \tag{18}$$

where Φ_i is the ith element of a specifi set of orthogonal basis functions in terms of $\boldsymbol{X}$, and $\rho_i \in \mathbb{R}$ are the corresponding coefficients The shape of the basis functions Φ_i is determined by the probability measure of $\boldsymbol{X}$. The orthogonal bases required for some of the popular random variable types can be found in the Wiener-Askey scheme; see [29]. For all other distributions, Gram-Schmidt orthogonalization can be used to determine a set of orthogonal polynomial functions corresponding to the custom distribution. Equation (18) is typically truncated to a finit number of terms as an approximation to yield the surrogate model

$$Y \approx \hat{Y} = \sum_{i=1}^{N} \hat{\rho}_i \Phi_i(\boldsymbol{X}). \tag{19}$$

The objective is to determine the coefficient $\hat{\rho}_i$ of the surrogate model $\hat{Y}$ so that one can have a simple model of the true stochastic system as a polynomial function of the input stochastic variables.

Traditionally, there have been two broad category of methods to fin those coefficients namely, intrusive methods and nonintrusive methods; see [30]. Below, we briefl review a method from each of those categories.

3.1.1 Intrusive PC

In intrusive PC, we look for coefficient which minimize the mean value of the square of the model error, i.e., the cost function is

$$J = \int_{\Omega_{\boldsymbol{X}}} (Y - \hat{Y})^2 f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{20}$$

Using basic calculus of variation, recognizing that the necessary condition for the minima is given by the equation

$$-2 \int_{\Omega_{\boldsymbol{X}}} (Y - \hat{Y}) \frac{\partial \hat{Y}}{\partial \hat{\rho}_i} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 0, \tag{21}$$

where i varies from 1 to N, and citing the orthogonality property of $\partial \hat{Y} / \partial \hat{\rho}_i = \Phi_i$ we get the following closed form expression for the coefficients

$$\hat{\rho}_i = \frac{\int_{\Omega_{\boldsymbol{X}}} g(\boldsymbol{x}) \Phi_i f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int_{\Omega_{\boldsymbol{X}}} \Phi_i^2 f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}. \tag{22}$$

This method of evaluating the coefficient is also popularly known as the Galerkin projection method.

Although calculating the denominator of Eq. (22) is trivial, the numerator can be extremely difficul to evaluate for generic nonlinear functions and places where g is merely a computer code. Hence, in spite of obtaining a clean expression for the coefficient in Eq. (22), the need for the evaluation of a multivariate integral forms the most profound limitation of this method.

3.1.2 Nonintrusive PC

In order to circumvent multidimensional integrals, the coefficient can also be determined using function evaluations and least-squares regression. Consider the expression for the approximation error,

$$\boldsymbol{e} := \underbrace{\begin{bmatrix} \Phi_1(\boldsymbol{x}^{(1)}) & \Phi_2(\boldsymbol{x}^{(1)}) & \cdots & \Phi_N(\boldsymbol{x}^{(1)}) \\ \Phi_1(\boldsymbol{x}^{(2)}) & \Phi_2(\boldsymbol{x}^{(2)}) & \cdots & \Phi_N(\boldsymbol{x}^{(2)}) \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_1(\boldsymbol{x}^{(m)}) & \Phi_2(\boldsymbol{x}^{(m)}) & \cdots & \Phi_N(\boldsymbol{x}^{(m)}) \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_N \end{bmatrix}}_{\boldsymbol{Y_{PC}}} - \underbrace{\begin{bmatrix} y(\boldsymbol{x}^{(1)}) \\ y(\boldsymbol{x}^{(2)}) \\ \vdots \\ y(\boldsymbol{x}^{(m)}) \end{bmatrix}}_{\boldsymbol{y}}, \tag{23}$$

where $\boldsymbol{x}^{(i)}$ represents the ith sample point from a total of m samples in the input space, $A \in \mathbb{R}^{m \times N}$ is a matrix whose rows represent the basis functions evaluated at $\boldsymbol{x}^{(i)}$, $\boldsymbol{y}$ is a vector assimilated using the transformation g over $\boldsymbol{x}^{(i)}$ samples, and $\boldsymbol{e}$ is a vector of the approximation errors at each sample point. The coefficient $\boldsymbol{Y_{PC}}$ can be determined by minimizing the 2-norm of $\boldsymbol{e}$, resulting in

$$\boldsymbol{Y_{PC}} = (A^T A)^{-1} A^T \boldsymbol{y}, \tag{24}$$

see [30]. Equation (24) now provides a simple expression to determine the coefficient of the PC surrogate model from system realizations of the original function. The choice of the sample points $\boldsymbol{x}^{(i)}$ has been studied in the literature at great length. Choices include randomly sampling the input space, using Gauss quadrature points or smart sampling methods to judiciously cover the input space with fewer samples; see [31]. Moreover, if the true output realizations are corrupted by noise or disturbances, regularization methods can also be used while solving the least-squares problem [30]. It should be noted that for linear systems, it has been shown in [32] that a nonintrusive approach can result in the exact evaluation of the coefficient given in Eq. (22).

Irrespective of the manner in which the coefficient $\hat{\rho}_i$ are determined, the derivation of a surrogate model that emulates the true function greatly reduces computational requirements on sampling the function for subsequent analysis. The function needs to be sampled numerous times to get accurate estimates of the output PDFs and conditional PDFs. Since sampling the surrogate is much cheaper than sampling the true system, we now have a tractable way to determine the computationally expensive PDFs $f_Y(y)$, $f_{Y|X_i}(y, x_i)$ and $f_{Y|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(y, \tilde{\boldsymbol{x}}_{\boldsymbol{i}})$ using MC methods.

3.2 Efficien Evaluation of the Expectation Integral

Expectation integrals of the form

$$I := \int_{\Omega_{\boldsymbol{X}}} k(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \tag{25}$$

turn up in numerous applications of basic and applied sciences, ranging from quantum mechanics [33] to filterin [20,34]. As a result, many researchers over the years have endeavored to efficientl evaluate this integral. Some of the most popular methods used have been Monte Carlo sampling methods [35], quasi Monte Carlo methods [36], Gauss quadrature rules [37], sparse quadrature rules [38], and conjugate unscented transform (CUT) rules [20]. In this paper, we highlight some of these approaches and provide commentary on which method to use for the NS metrics.

In all of these methods, the integral I is approximated by a weighted sum of function evaluations,

$$I \approx \hat{I} = \sum_{i=1}^{N_{\text{method}}} w_i k(\boldsymbol{x}^{(i)}), \tag{26}$$

where $\boldsymbol{x}^{(i)}$ are certain samples from the input space, N_{method} denotes the number of samples, and w_i are specifi weights. The aforementioned methods primarily differ in the manner by which the location of the sample points $\boldsymbol{x}^{(i)}$ and their corresponding weights w_i are determined. The benefi of representing an integral with function evaluations lies in the fact that sample realizations are independent of each other and parallel computing techniques can be adopted to evaluate the system realizations simultaneously.

The objective of discussing expectation integrals has been to highlight the fact that in order to determine the values of NS, we firs need to evaluate Eqs. (7) and (11) which are expectation integrals such as Eq. (25). The

metric ${}^1\overline{\mathcal{D}}_i$ is a univariate integral while ${}^2\overline{\mathcal{D}}_i$ is an $(n-1)$-dimensional multivariate integral. These integrals can be approximated as weighted sums of the integrands as

$$ {}^1\overline{\mathcal{D}}_i \approx \sum_{j=1}^{N_{\text{method}}} w_j \mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|X_i}(\hat{y}, x_i^{(j)})\Big) \tag{27} $$

and

$$ {}^2\overline{\mathcal{D}}_i \approx \sum_{j=1}^{N_{\text{method}}} w_j \mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|\tilde{\boldsymbol{X}}_i}(\hat{y}, \tilde{\boldsymbol{x}}_i^{(j)})\Big), \tag{28} $$

where $x_i^{(j)}$ is the jth sample point out of a total of N_{method} samples in the Ω_{X_i} space and $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}$ is the jth sample point out of a total of N_{method} samples in the $\Omega_{\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}$ space. Note that y has been replaced by $\hat{y}$ in the equations to represent the surrogate model instead of the true system. In the remainder of this section, we will give a brief overview of techniques for efficientl computing the integral in Eq. (25).

3.2.1 The Monte Carlo Sampling Method

In this method, the samples $\boldsymbol{x}^{(i)}$ are randomly drawn from the input space $\Omega_{\boldsymbol{X}}$ based on the joint density function $f_{\boldsymbol{X}}(\boldsymbol{x})$. Each sample point is weighed equally leading to $w_i = 1/N_{\text{MC}}$, where N_{MC} is the number of samples drawn. The fina approximation is given by

$$ \hat{I}_{\text{MC}} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} k\Big(\boldsymbol{x}^{(i)}\Big). \tag{29} $$

Almost sure convergence of $\hat{I}_{\text{MC}}$ to I can be shown when N_{MC} tends to infinit . Although the implementation of MC is simple, convergence is rather slow, as the root-mean-square error decreases as $1/\sqrt{N_{\text{MC}}}$. However, it has the advantage of being independent of the number of variables (i.e., the dimension of $\boldsymbol{X}$). Hence, MC has been popular for evaluating high-dimensional multivariate integrals, where other quadrature rules encounter the curse of dimensionality [39].

3.2.2 Gauss Quadrature Rules

Gauss quadrature (GQ) is a method which relies on a polynomial approximation of the integrand in Eq. (25). For a univariate integral, an N_{GQ}-sample point GQ rule can accurately evaluate the integral I when the integrand can be well approximated by a polynomial of order $2N_{\text{GQ}} - 1$. Since polynomials can approximate any continuous function over a finit domain in the limit, increasing the value of N_{GQ} allows GQ to approximate nonlinear integrands. The location of the sample points and their weights are dependent on the order of the GQ rule and the weighing function $f_{\boldsymbol{X}}(\boldsymbol{x})$. The locations are simply given by the roots of the N_{GQ}th-order polynomial orthogonal to $f_{\boldsymbol{X}}(\boldsymbol{x})$ while the weights are calculated from the coefficient of the orthogonal polynomials.

GQ can also be easily extended to the multivariate case. This is done by taking the tensor product of the sample point locations as well as the weights in each univariate direction to form an n-dimensional grid. As a result, although it is simple to determine the grid locations and weights of GQ, it has the shortcoming of suffering from the curse of dimensionality; i.e., when the number of integration variables become high, the total number of grid points increases exponentially. A list of the most common distributions, their associated orthogonal polynomials, sample points, and the weights can be found in [40].

3.2.3 Conjugate Unscented Transform

Conjugate unscented transform is a recently developed set of rules to evaluate multivariate integrals where the weight function is either a uniform distribution or a Gaussian distribution. Similar to GQ, it is based on a polynomial approximation of the integrand in Eq. (25). It leverages the multidimensional symmetry in the uniform and the Gaussian

distribution to come up with a total number of sample points that is less than the number of sample points in a conventional tensor product GQ grid.

Since the integral of a polynomial is simply a weighted sum of the moments of the input variables, the CUT algorithm establishes constraints to capture a certain number of finit moments while solving for the location of the sample points and weights. These constraints are also referred to as the moment constraint equations and solving them yields the desired set of points and weights for varying dimensions. Depending on the order of moments to be captured, CUT presents three algorithms: CUT4, CUT6, and CUT8, which are capable of capturing moments up to orders 5, 7, and 9 respectively. Details about the algorithm can be found in [20]. A brief summary of the key concepts of CUT have also been included in this paper; see Appendix A.

3.3 Remarks on the Use of MC, GQ, and CUT

Since the convergence of MC is slow and requires an enormous number of samples to evaluate Eqs. (27) and (28), it is never used to determine $\overline{\mathcal{D}}_i$. GQ is always used to evaluate ${}^1\overline{\mathcal{D}}_i$ since it is a univariate integral and GQ provides the minimal set of points and weights to integrate a polynomial of any order for univariate integrals. For input vectors $\boldsymbol{X}$ which have a uniform distribution or a Gaussian distribution, CUT is preferred to evaluate ${}^2\overline{\mathcal{D}}_i$, considering it requires fewer number of points than GQ. However, if the input variables have distributions which are not uniform or Gaussian, GQ/quasi-MC/randomized QMC can be employed to evaluate ${}^2\overline{\mathcal{D}}_i$. Although other sparse quadrature rules have not been discussed in this article, they can also be used instead of GQ to calculate ${}^2\overline{\mathcal{D}}_i$.

3.4 Evaluation of the Moment-Independent Sensitivity Indices

This subsection now elaborates the step by step process needed from start to finis to yield the desired NS metrics using results from all the previous sections. A fl wchart in Fig. 7 is also presented for illustration.

The firs step involves developing the surrogate model $\hat{Y}$ from the model equation $Y = g(\boldsymbol{X})$ using PC as described in Section 3.1.

The second step is to determine the output PDF $f_{\hat{Y}}(\hat{y})$. This is done by sampling the input space $\Omega_{\boldsymbol{X}}$ and evaluating the surrogate function at each of those samples. For all examples in this paper, kernel density estimation (KDE) is used to estimate the univariate PDF. The estimate is based on the normal kernel function and is evaluated at equally spaced $\hat{y}$ values in the range of the output PDF $f_{\hat{Y}}(\hat{y})$.

The quantities ${}^1\overline{\mathcal{D}}_i$ as well as ${}^2\overline{\mathcal{D}}_i$ are approximated by a weighted sum of $\mathcal{D}$s evaluated at strategic points as represented by Eqs. (27) and (28). The jth sample point for ${}^1\overline{\mathcal{D}}_i$ is denoted by $x_i^{(j)}$. For each $x_i^{(j)}$ we need $\mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|X_i}(\hat{y}, x_i^{(j)})\Big)$ and for each $\mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|X_i}(\hat{y}, x_i^{(j)})\Big)$ we need the PDF $f_{\hat{Y}|X_i}\Big(\hat{y}, x_i^{(j)}\Big)$. This PDF is approximated by sampling the $\Omega_{\tilde{\boldsymbol{X}}_i}$ space, evaluating the surrogate model at the samples by keeping X_i fi ed at $x_i^{(j)}$ and finall plotting the histogram of $\hat{y}$. Similar to the output PDF, KDE is used to estimate $f_{\hat{Y}|X_i}\Big(\hat{y}, x_i^{(j)}\Big)$ from the histogram. Each value of $\mathcal{D}$ obtained from $f_{\hat{Y}|X_i}\Big(\hat{y}, x_i^{(j)}\Big)$ is then stored for assimilation later.

If the user makes the choice of determining ${}^2NS_i$, then we need the evaluation of ${}^2\overline{\mathcal{D}}_i$. Note that the jth sample point for ${}^2\overline{\mathcal{D}}_i$ is given by $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}$. For each $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}$ we need $\mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(\hat{y}, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)})\Big)$ and for each $\mathcal{D}\Big(f_{\hat{Y}}(\hat{y}), f_{\hat{Y}|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}(\hat{y}, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)})\Big)$ we need the PDF: $f_{\hat{Y}|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}\Big(\hat{y}, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}\Big)$. This PDF, again, is approximated by sampling the Ω_{X_i} space, evaluating the surrogate model at the samples by keeping $\tilde{\boldsymbol{X}}_{\boldsymbol{i}} = \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}$ and using KDE to obtain $f_{\hat{Y}|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}\Big(\hat{y}, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}\Big)$. Once again, similar to the previous case, each value of $\mathcal{D}$ obtained from $f_{\hat{Y}|\tilde{\boldsymbol{X}}_{\boldsymbol{i}}}\Big(\hat{y}, \tilde{\boldsymbol{x}}_{\boldsymbol{i}}^{(j)}\Big)$ is stored for assimilation later.

The penultimate step involves obtaining the weighted sum of all the stored $\mathcal{D}$s to yield ${}^1\overline{\mathcal{D}}_i$ and ${}^2\overline{\mathcal{D}}_i$. Once ${}^1\overline{\mathcal{D}}_i$ and ${}^2\overline{\mathcal{D}}_i$ have been determined for all i, ${}^1NS_i$ and ${}^2NS_i$ are finall determined using Eqs. (10) and (13), respectively.

At this point, it is prudent to mention that the user need not evaluate both Class 1 as well as Class 2 metrics since they are complements of each other and provide similar information. Both classes of metrics have been developed

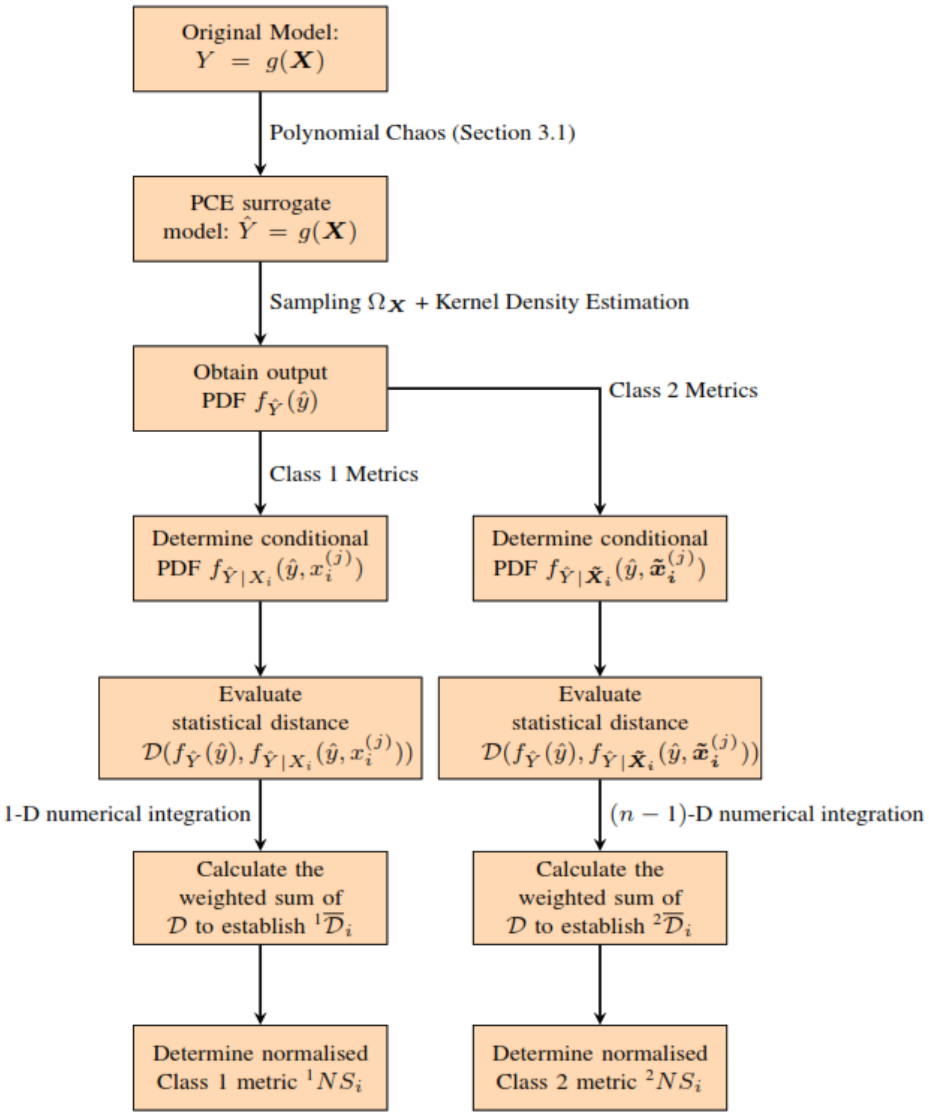


FIG. 7: Flowchart depicting the procedure for evaluating NS metrics

and presented in this article for academic purposes. Either one of the metrics can be evaluated. If the choice is made to evaluate Class 1 metrics, then ${}^1\overline{\mathcal{D}}_i$ needs to be calculated. Otherwise, for Class 2 metrics we only need ${}^2\overline{\mathcal{D}}_i$.

Note that there exists a trade-off in the computational effort required to evaluate ${}^1\overline{\mathcal{D}}_i$ and ${}^2\overline{\mathcal{D}}_i$. For evaluating ${}^1\overline{\mathcal{D}}_i$, the expectation integral is a 1-D integral. However, while determining the statistical distances for each fi ed value of X_i, the evaluation of the conditional PDF $f_{Y|X_i}$ requires sampling an $(n-1)$-D space to estimate it. For evaluating ${}^2\overline{\mathcal{D}}_i$, on the other hand, the expectation integral is an $(n-1)$-D integral while the estimation of the conditional PDF $f_{Y|\tilde{\boldsymbol{X}}_i}$ requires a 1-D sampling. In the opinion of the authors, the exact computational merit of one class over another is nontrivial and depends on a few factors. One factor is the computational effort required for function evaluation, a second factor is the number of MC samples required for convergence of the estimation of the conditional PDFs, a third factor is the number of quadrature points necessary to evaluate the expectation integral, and finall the number of input parameters being ranked. Since sampling a surrogate model is rather inexpensive, we could potentially consider the impact of function evaluations negligible. With an increase in the number of input parameters n, the quadrature

points required to evaluate a $(n-1)$-D expectation integral increases for calculating ${}^2\overline{\mathcal{D}}_i$. However, the growth in computation is compensated by the fact that the derivation of the conditional PDFs would not require a higher number of samples since the PDFs are conditioned on $\tilde{\boldsymbol{X}}_i$ and require a 1-D sampling. For ${}^1\overline{\mathcal{D}}_i$, with an increase in n, the computation required to calculate the expectation integral sees no change. However, there is an increase in the computation required to derive the conditional PDFs which are conditioned on X_i. Consequently, the efficien y of evaluating ${}^1\overline{\mathcal{D}}_i$ and ${}^2\overline{\mathcal{D}}_i$ entirely depends on the rates of convergence for the expectation integral and the PDF estimation, both of which are model dependent.

4. NUMERICAL RESULTS

This section presents an illustration of the GSA measures NS on three numerical examples. The firs example is the Ishigami function. The second example is the Sobol' G-function with four variables and the fina example is that of a dynamic epidemic model. For each example, NS_i values are determined and compared to rank the importance of the variables. The ranking is subsequently also compared with the variance-based measure S_{Ti}.

All numerical examples were coded and simulated in MATLAB (version 9.9). Added dependencies include the Symbolic Math Toolbox (version 8.6) and the Statistics and Machine Learning Toolbox (version 12.0).

4.1 The Ishigami Function

The Ishigami function

$$Y = g(\boldsymbol{X}) = \sin(X_1) + a\sin^2(X_2) + b(X_3)^4\sin(X_1) \tag{30}$$

is a benchmark problem for global SA algorithmsl see [2,11,12,41]. The input space of the Ishigami function is define by the hypercube: $\Omega_{\boldsymbol{X}} = [-\pi, \pi]^3$. It is also assumed that the input variables are independent and are uniformly distributed [i.e., $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3)$ where $f_{X_i}(x_i) \sim \mathcal{U}([-\pi, \pi])$]. Parameter values for the function are chosen to be $a = 7$ and $b = 0.1$, similar to [41]. The objective is to calculate estimates of ${}^1NS_i$ and ${}^2NS_i$.

A PC surrogate model is developed using the Galerkin projection method. The symbolic integrals in Eq. (22) are solved using MATLAB and are not intractable for the Ishigami function. The basis functions chosen are that of multivariate Legendre polynomials (as recommended by the Wiener-Askey scheme for uniformly distributed inputs) where the multivariate bases are derived from the tensor product of univariate Legendre polynomials. The PC order in each univariate direction is chosen to be $N_{X_i} = 7$. Since the fina set of bases is derived from a tensor product of the univariate bases set, the total number of bases becomes $N = N_{X_1}N_{X_2}N_{X_3} = 7^3 = 343$. A Galerkin projection therefore yields 343 coefficient Y_1–Y_{343} to complete the surrogate model:

$$\hat{Y} = \sum_{i=1}^{343} \hat{\rho}_i \Phi_i(X_1, X_2, X_3). \tag{31}$$

To check the fidelit of the surrogate model, the output PDFs from the true model and the surrogate model [i.e., $f_Y(y)$ and $f_{\hat{Y}}(\hat{y})$] are derived and compared. This is done by sampling $\Omega_{\boldsymbol{X}}$ 10^5 times, evaluating the true as well as the surrogate model at those sample locations, and plotting their histograms/PDFs. Such a plot is shown in Fig. 8. It is evident from the proximity of the PDFs in the plot that the surrogate model closely approximates the true system (note the high fidelit of approximation at the tails of the PDF). Note that the use of a surrogate model in this example is only for illustrative purposes since the computational cost of sampling the original Ishigami function is relatively cheap. The benefi of using the surrogate is more applicable for examples where sampling the original model is expensive.

In order to calculate the values of ${}^1\overline{\mathcal{D}}_i$, a 30-point GQ rule is employed. Since each input random variable is uniformly distributed, Gauss–Legendre quadrature rules are used for ${}^1\overline{\mathcal{D}}_1$, ${}^1\overline{\mathcal{D}}_2$, and ${}^1\overline{\mathcal{D}}_3$. However, prior to determining the statistical distances, the PDFs $f_{\hat{Y}|X_i}(\hat{y}, x_i^{(j)})$ are calculated at the GQ points $x_i^{(j)}, j = 1, 2, ..., N_{\text{GQ}}$. This is done by sampling the input subspace ($\Omega_{\tilde{\boldsymbol{X}}_i}$) 10^5 times while keeping $X_i = x_i^{(j)}$ where j represents the jth GQ node. Figures 9(a)–9(c) present the density functions $f_{\hat{Y}|X_1}\left(\hat{y}, x_1^{(j)}\right)$, $f_{\hat{Y}|X_2}\left(\hat{y}, x_2^{(j)}\right)$, and $f_{\hat{Y}|X_3}\left(\hat{y}, x_3^{(j)}\right)$ for each of the

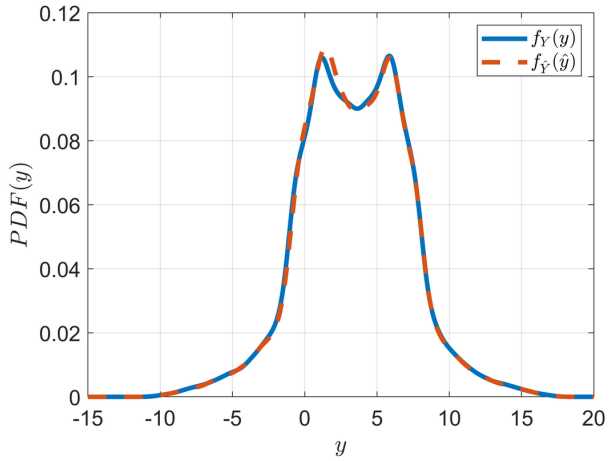


FIG. 8: Comparison of output PDFs from true and surrogate models

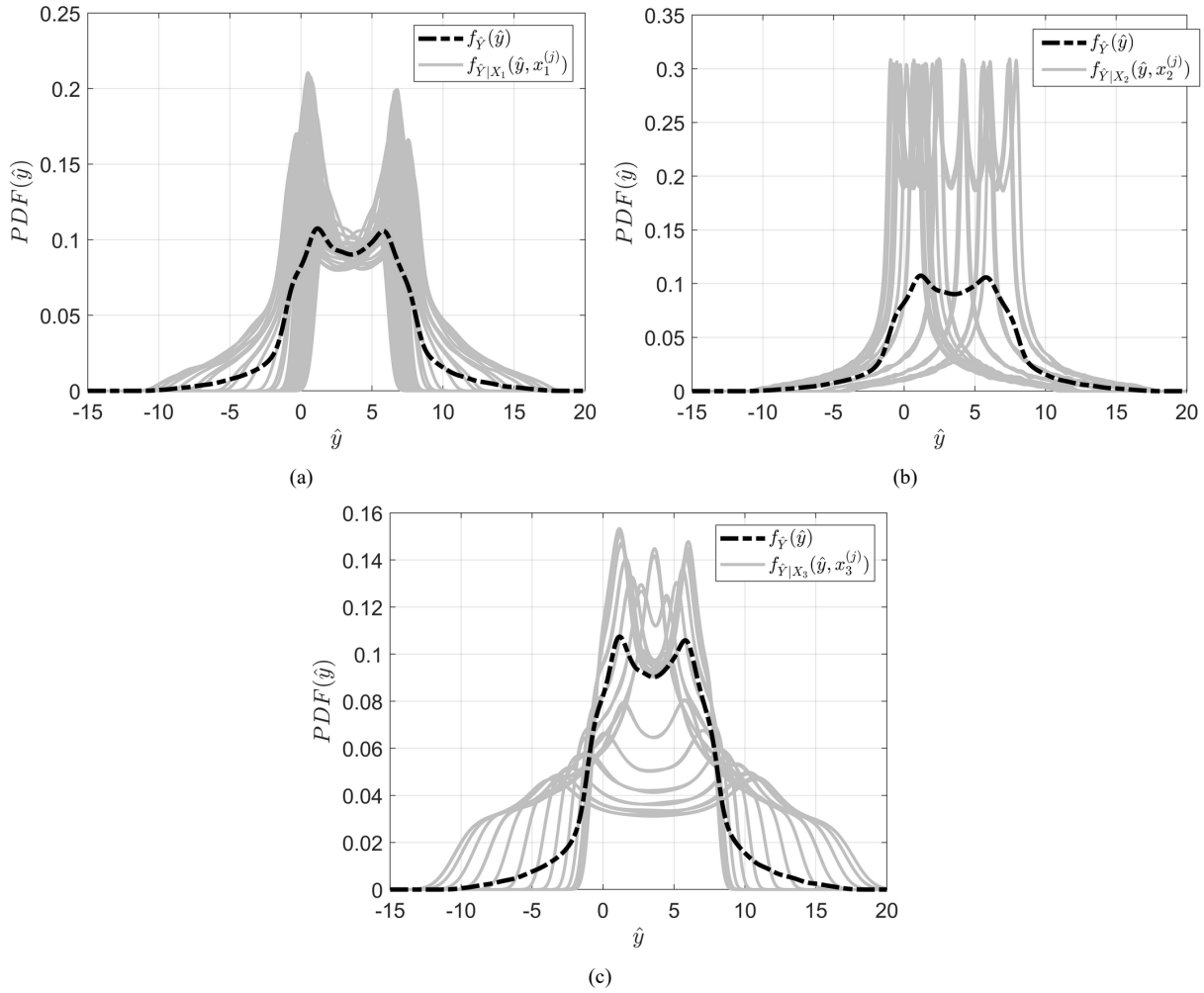


FIG. 9: Plots showing the variation of (a) $f_{\hat{Y}|X_1}\left(\hat{y}, x_1^{(j)}\right)$, (b) $f_{\hat{Y}|X_2}\left(\hat{y}, x_2^{(j)}\right)$, and (c) $f_{\hat{Y}|X_3}\left(\hat{y}, x_3^{(j)}\right)$ for the 30 Gauss–Legendre nodes (Ishigami)

30 Gauss–Legendre points, respectively. The output PDF $f_{\hat{Y}}(\hat{y})$ has also been shown in these plots for comparison since statistical distances between the output PDF and each of the gray plots in the figure are evaluated to determine ${}^1\overline{\mathcal{D}}_i$.

The value of the moment-independent sensitivity indices ${}^1\overline{\mathcal{D}}_i$ for each distance measured from Table 1 can be found in Table 2. The higher the value of the statistical distance, the further away is the conditional PDF from the output PDF. Hence, just by observing the magnitude of ${}^1\overline{\mathcal{D}}_i$ we can determine the order of importance of the variables. With that in mind, all the $\mathcal{D}$ measures are consistent among themselves in ranking the variables as X_2, X_1, and finall X_3. In order to more readily make comparisons between the measures, the ${}^1\overline{\mathcal{D}}$ values are normalized to determine the ${}^1NS$ metrics. These measures are listed in Table 3 along with the total Sobol' indices S_{Ti}. The total Sobol' indices S_{Ti} were calculated using analytical expressions derived in [16].

It is now easier to compare the relative importance predicted by the ${}^1NS$ metrics. It is interesting to note that the Sobol' indices present a different ranking (i.e., a ranking in the order of X_1, followed by X_2 and X_3). This can be attributed to the fact that Sobol' does not consider any moment higher than the second moment while the ${}^1NS$ metrics are based on the complete PDFs. Hence, the Ishigami function presents a convincing case for adopting a PDF-based GSA measure over a variance-based one (which could predict a misleading ranking of importance). Similar results have also been presented previously in [11] where it was shown that importance ranking via δ_i was different than S_{Ti}. It should be pointed out that the metrics ${}^1\overline{\mathcal{D}}_i^T$ or ${}^1NS_i^T$ are analogous to the Borgonovo metric δ_i [11].

For the Ishigami function, a study was done for the choice of a 30-point Gauss quadrature rule for the expectation integration calculation. Experiments were repeated for 5-, 10-, 15-, 20-, 25-, 30-, and 35-point Gauss–Legendre quadratures. For each experiment ${}^1NS$ was evaluated. The evolution of the moment-independent sensitivity indices with increasing quadrature points is illustrated in Fig. 10(a). It is evident from the plots that the NS metrics have converged well prior to the 30-point mark motivating the choice for a 30-point GQ rule. The choice of the number of MC samples was also determined after successive experiments. The convergence of the sensitivity indices for the Ishigami function with respect to the number of MC samples is presented in Fig. 10(b).

Next, we present results from the evaluation of the Class 2 metrics. In order to evaluate ${}^2\overline{\mathcal{D}}_i$, the CUT6 uniform algorithm was used to generate the sample points in the two-dimensional input subspace $\Omega_{\tilde{\boldsymbol{X}}_i}$. The algorithm yields 13 sample points and corresponding weights. For the convenience of the reader, sample points and weights for the evaluation of ${}^2\overline{\mathcal{D}}_1$ have been plotted in Fig. 11. The locations of the points are shown by the circles while the sizes of the circles reflec the associated weight.

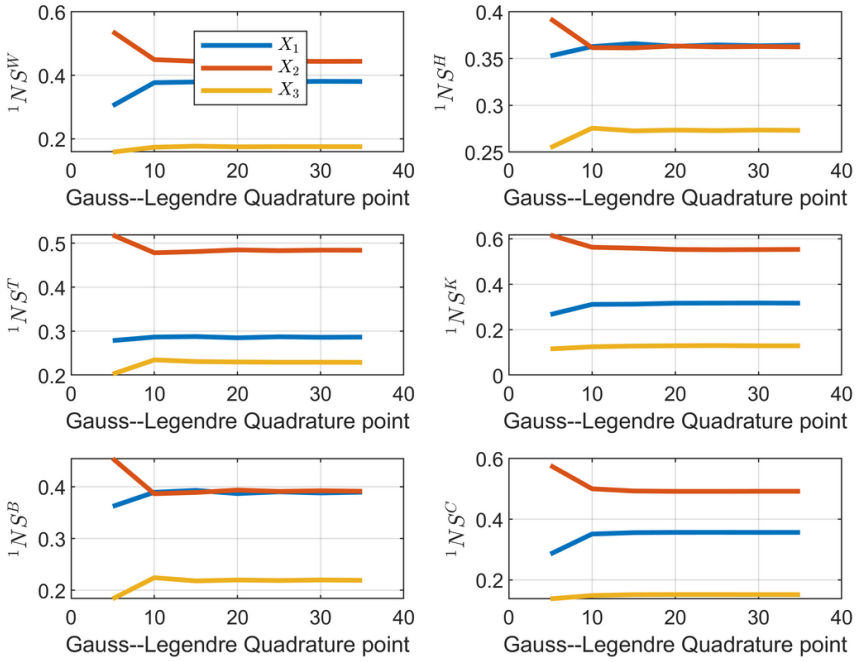
At each of these sample point locations, Ω_{X_1} is sampled 10^5 times to determine $f_{\hat{Y}|\tilde{\boldsymbol{X}}_1}\left(\hat{y}, \tilde{\boldsymbol{x}}_1^{(j)}\right)$. The PDFs $f_{\hat{Y}|\tilde{\boldsymbol{X}}_2}\left(\hat{y}, \tilde{\boldsymbol{x}}_2^{(j)}\right)$ and $f_{\hat{Y}|\tilde{\boldsymbol{X}}_3}\left(\hat{y}, \tilde{\boldsymbol{x}}_3^{(j)}\right)$ are evaluated in a similar manner and have been shown in Figs. 12(a)–12(c), respectively.

TABLE 2: ${}^1\overline{\mathcal{D}}_i$ values for different $\mathcal{D}$ measures (Ishigami)

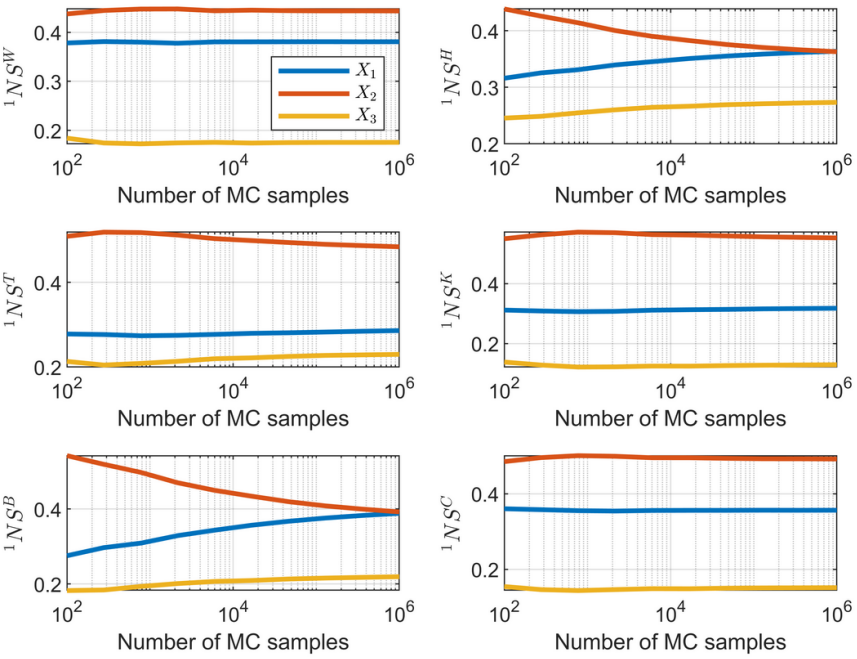
$\mathcal{D}$ measure	${}^1\overline{\mathcal{D}}_1$	${}^1\overline{\mathcal{D}}_2$	${}^1\overline{\mathcal{D}}_3$
W	2.0114	2.3476	0.9272
H	0.4722	0.4898	0.3567
T	0.2418	0.4206	0.1938
K	0.2179	0.3858	0.0887
B	0.1191	0.1306	0.0685
C	0.5485	0.7595	0.2324

TABLE 3: ${}^1NS_i$ values for different $\mathcal{D}$ measures (Ishigami)

Variable	${}^1\boldsymbol{NS}^{\boldsymbol{W}}$	${}^1\boldsymbol{NS}^{\boldsymbol{H}}$	${}^1\boldsymbol{NS}^{\boldsymbol{T}}$	${}^1\boldsymbol{NS}^{\boldsymbol{K}}$	${}^1\boldsymbol{NS}^{\boldsymbol{B}}$	${}^1\boldsymbol{NS}^{\boldsymbol{C}}$	$\boldsymbol{S_T}$
X_1	0.3805	0.3581	0.2824	0.3147	0.3742	0.3561	0.5576
X_2	0.4441	0.3714	0.4913	0.5572	0.4105	0.4930	0.4424
X_3	0.1754	0.2705	0.2264	0.1281	0.2152	0.1509	0.2437



(a)



(b)

FIG. 10: Evolution of the Class 1 NS metrics with increasing (a) quadrature points and (b) number of MC samples

The dotted curve in Figs. 12(a)–12(c) is the output PDF $f_{\hat{Y}}(\hat{y})$. The weighted average of the statistical distances between the dotted curve and the grey curves lead us to desired values of ${}^2\overline{\mathcal{D}}_i$. Depending on the type of $\mathcal{D}$ measure used, we get different values of ${}^2\overline{\mathcal{D}}_i$ as listed in Table 4.

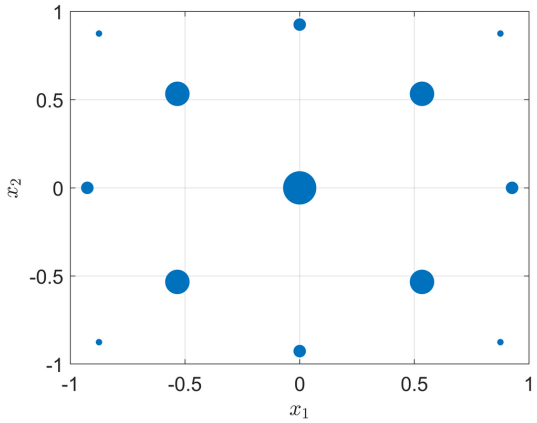


FIG. 11: 2D CUT6 points and weights for uniform distributions

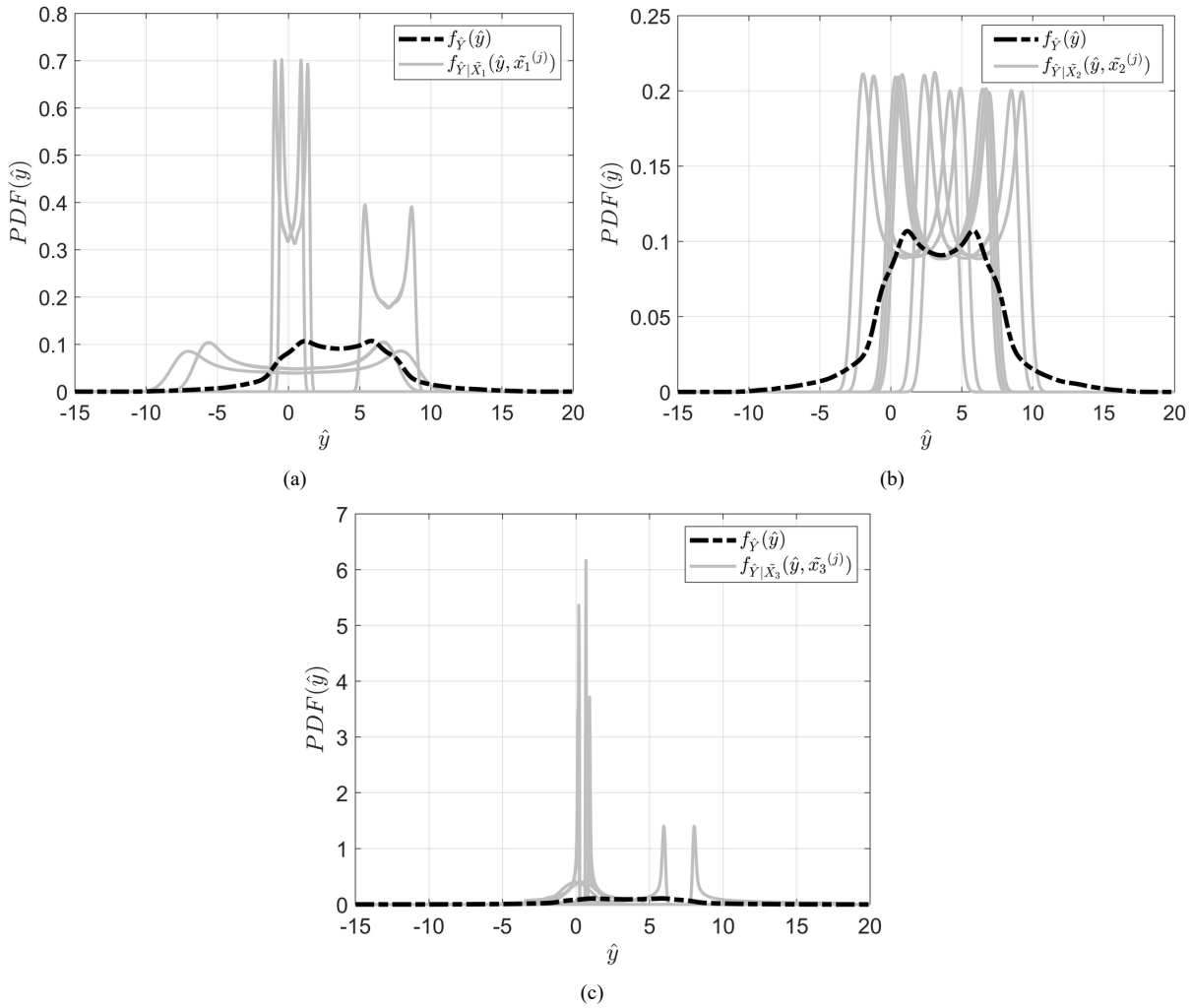


FIG. 12: Plots showing the variation of (a) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_1}\left(\hat{y}, \tilde{\boldsymbol{x}}_1^{(j)}\right)$, (b) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_2}\left(\hat{y}, \tilde{\boldsymbol{x}}_2^{(j)}\right)$, and (c) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_3}\left(\hat{y}, \tilde{\boldsymbol{x}}_3^{(j)}\right)$ for the 13 CUT6 nodes (Ishigami)

TABLE 4: ${}^2\overline{\mathcal{D}}_i$ values for different $\mathcal{D}$ measures (Ishigami)

$\mathcal{D}$ measure	${}^2\overline{\mathcal{D}}_1$	${}^2\overline{\mathcal{D}}_2$	${}^2\overline{\mathcal{D}}_3$
W	3.5855	1.6159	3.8016
H	0.8700	0.5733	0.9583
T	0.6369	0.3161	0.6972
K	0.5777	0.2322	0.5947
B	0.5377	0.1849	0.7053
C	1.1347	0.4688	1.1803

Since Class 2 metrics are complementary to the Class 1 metrics, one can again simply determine ranking of importance by observing the individual values of ${}^2\overline{\mathcal{D}}_i$. The closer the output PDF to the averaged output PDF conditioned on the inputs $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}$, the larger the influenc of the input parameter X_i. This is in contrast to metrics of Class 1, where an output PDF further away from the averaged output PDF conditioned on the input X_i means a large influenc of the parameter X_i. Hence, low values of the moment-independent sensitivity index ${}^2\overline{\mathcal{D}}_i$ indicate a high importance of the input parameter X_i.

The results for ${}^2\overline{\mathcal{D}}_i$ in Table 4 therefore are consistent with the results of ${}^1\overline{\mathcal{D}}_i$. All the $\mathcal{D}$ measures predict the same ranking in terms of importance: i.e., X_2 followed by X_1 and X_3. In fact, even visually inspecting Figs. 12(a)–12(c) we can see that the lighter shaded curves are generally most distant from the dashed curve in Figs. 12(a) and 12(c) as compared to Fig. 12(b), which is quantitatively verifie by the ${}^2\overline{\mathcal{D}}_i$ values.

For easier comparison between $\mathcal{D}$ measures, normalized ${}^2NS_i$ metrics are presented in Table 5. S_{Ti} have also been included for reference and were calculated using analytical expressions derived in [16].

4.2 The G-Sobol' Function

The second example chosen is another popular benchmark problem recognized in the literature as the Sobol' G-function [2,17,42,43] and is given by

$$Y = g(\boldsymbol{X}) = \prod_{i=1}^{n} \frac{|4X_i - 2| + a_i}{1 + a_i}, \tag{32}$$

where n represents the number of input variables, X_i are independent and uniformly distributed input variables, and a_i are certain constants. The G-function has an input domain $\Omega_{\boldsymbol{X}} = [0, 1]^n$. n is chosen to be 4 and a_i is chosen in a manner similar to [17], where

$$a_i = \frac{i - 1}{2}. \tag{33}$$

The values of a_i are directly related to the relative importance of the corresponding inputs X_i. The lower the value of a_i, the more important is X_i. This function is chosen to illustrate the reliability of the proposed metrics owing to the complexity shown by the function in terms of nonlinearity and nonsmoothness as well as its nonmonotonous nature. In fact, it has also been documented to be a challenging test for PC as an approximation tool [17] in terms of convergence. Hence, successfully predicting the relative importance of the input variables via the NS_i metrics is a legitimate test.

TABLE 5: ${}^2NS_i$ values for different $\mathcal{D}$ measures (Ishigami)

Variable	${}^2\boldsymbol{NS}^{\boldsymbol{W}}$	${}^2\boldsymbol{NS}^{\boldsymbol{H}}$	${}^2\boldsymbol{NS}^{\boldsymbol{T}}$	${}^2\boldsymbol{NS}^{\boldsymbol{K}}$	${}^2\boldsymbol{NS}^{\boldsymbol{B}}$	${}^2\boldsymbol{NS}^{\boldsymbol{C}}$	$\boldsymbol{S_T}$
X_1	0.3983	0.3623	0.3860	0.4123	0.3766	0.4076	0.5576
X_2	0.1795	0.2387	0.1915	0.1653	0.1295	0.1684	0.4424
X_3	0.4223	0.3990	0.4225	0.4234	0.4940	0.4240	0.2437

For this example as well, a PC surrogate model is developed using the Galerkin projection method since the multivariate integrals indicated by Eq. (22) are readily solvable. Since the input variables are once again uniformly distributed, the basis functions are dictated to be Legendre polynomials. The PC order in each univariate direction is chosen to be $N_{X_i} = 7$. As the fina set of bases is derived from a tensor product of the individual univariate bases, the total number of bases becomes $N = N_{X_1}N_{X_2}N_{X_3}N_{X_4} = 2401$. Similar to the previous example, the surrogate model is obtained as

$$\hat{Y} = \sum_{i=1}^{2401} \hat{\rho}_i \Phi_i(X_1, X_2, X_3, X_4), \tag{34}$$

where the coefficient Y_1–Y_{2401} are determined using Galerkin projection.

Following the derivation of the surrogate model, a 30-point Gauss–Legendre quadrature rule is used to evaluate ${}^1\overline{\mathcal{D}}_i$. The output PDF $f_{\hat{Y}}(\hat{y})$ is determined by taking 10^4 samples from $\Omega_{\boldsymbol{X}}$, and evaluating the surrogate model $\hat{Y}$ for each sample. The conditional PDFs $[f_{\hat{Y}|X_i}(\hat{y}, x_i^{(j)})]$ required at each of the GQ node points are also determined by randomly sampling the input subspace ($\Omega_{\tilde{\boldsymbol{X}}_i}$) 10^4 times and transforming them through the surrogate model. The conditional PDFs are presented in Figs. 13(a)–13(d).

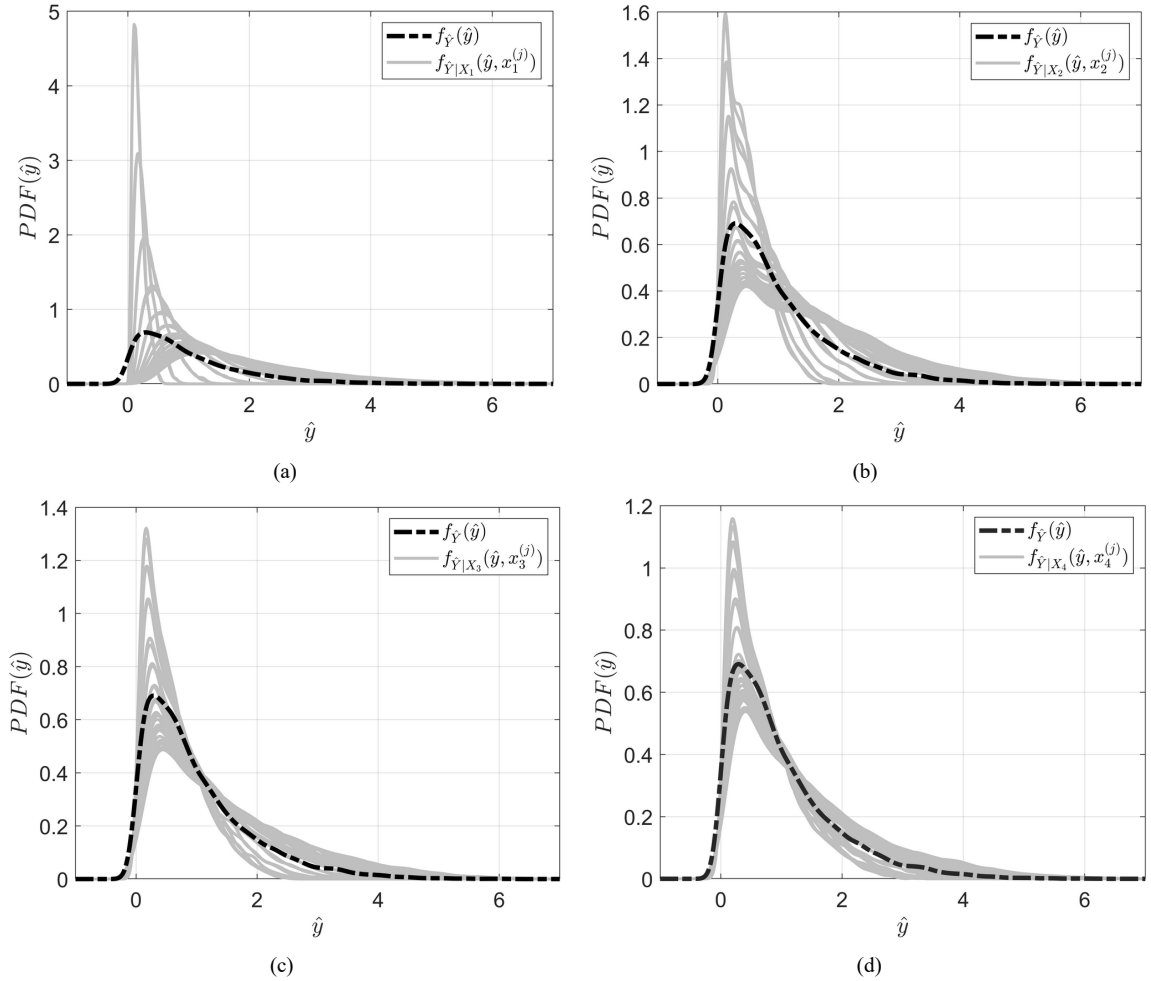


FIG. 13: Plots showing the variation of (a) $f_{\hat{Y}|X_1}\left(\hat{y}, x_1^{(j)}\right)$, (b) $f_{\hat{Y}|X_2}\left(\hat{y}, x_2^{(j)}\right)$, (c) $f_{\hat{Y}|X_3}\left(\hat{y}, x_3^{(j)}\right)$, and (d) $f_{\hat{Y}|X_4}\left(\hat{y}, x_4^{(j)}\right)$ for the 30 Gauss–Legendre nodes (G-function)

Once the statistical distances are measured between the output PDF [shown in blue in Figs. 13(a)–13(d)] and the conditional PDFs (shown in gray), they are averaged and the respective values of ${}^1\overline{\mathcal{D}}_i$ are calculated (listed in Table 6).

Since a_i was chosen to be increasing with i, the importance of the input variables wanes with i. It is visually somewhat observable from Figs. 13(a)–13(d) where we see that the gray lines keep getting closer to the blue curve for higher values of X_i. This is exactly what is seen from the ${}^1\overline{\mathcal{D}}_i$ measures in Table 6 as well as normalized ${}^1NS_i$ metrics in Table 7. The total Sobol' indices have also been included in the terminal column of Table 7. For this particular example, we see that the Sobol' indices are also able to predict the order of importance.

For determining ${}^2\overline{\mathcal{D}}_i$, once again we use the CUT6 algorithm to generate sample points in the $n-1=3$ dimensional space $\Omega_{\tilde{\boldsymbol{X}}_i}$. CUT6 yields a total of 35 points and weights. At each of these points, Ω_{X_i} is sampled 10^4 times to determine the PDFs $f_{\hat{Y}|\tilde{\boldsymbol{X}}_i}(\hat{y},\tilde{\boldsymbol{x}}_i)$. Figures 14(a)–14(d) present these PDFs in gray along with the output PDF $f_{\hat{Y}}(\hat{y})$ in blue. It is evident from these figure that the gray PDFs keep getting further away from the output PDF when conditioned over $\tilde{\boldsymbol{x}}_i$ as i increases.

On measuring the distances between the output PDFs and the conditional output PDFs, $\mathcal{D}$ is averaged to obtain the metrics ${}^2\overline{\mathcal{D}}_i$. Their values have been listed in Table 8. We see that the distances keep ascending with X_i for all the measures consistently, thereby successfully ranking the inputs in descending order of importance. The corresponding normalized metrics ${}^2NS_i$ are presented in Table 9.

4.3 The SIR Model

The third example illustrated is that of a dynamical system. The system represents a four-parameter SIR epidemic model and has also been considered for GSA analysis before in the literature [44]. The model is characterized by a set of three differential equations given by

$$\frac{dS}{dt} = \delta N - \delta S - \gamma k I S, \tag{35}$$

$$\frac{dI}{dt} = \gamma k I S - (r+\delta) I, \tag{36}$$

$$\frac{dR}{dt} = rI - \delta R, \tag{37}$$

where S, I, and R represent the number of susceptible, infected, and recovered people from a disease in a population of size N. The parameter δ represents the birth and death rate (assumed to be equal), γ is the infection coefficient k is

TABLE 6: ${}^1\overline{\mathcal{D}}_i$ values for different $\mathcal{D}$ measures (G-function)

$\mathcal{D}$ measure	${}^1\overline{\mathcal{D}}_1$	${}^1\overline{\mathcal{D}}_2$	${}^1\overline{\mathcal{D}}_3$	${}^1\overline{\mathcal{D}}_4$
W	0.5451	0.3513	0.2582	0.2075
H	0.5474	0.2730	0.1898	0.1475
T	0.3818	0.1857	0.1187	0.0941
K	0.3512	0.1783	0.1149	0.0911
B	0.1977	0.0474	0.0224	0.0132
C	0.3507	0.2084	0.1464	0.1153

TABLE 7: ${}^1NS_i$ values for different $\mathcal{D}$ measures (G-function)

Variable	${}^1NS^W$	${}^1NS^H$	${}^1NS^T$	${}^1NS^K$	${}^1NS^B$	${}^1NS^C$	S_T
X_1	0.4002	0.4728	0.4893	0.4775	0.7043	0.4273	0.5847
X_2	0.2579	0.2358	0.2380	0.2425	0.1689	0.2538	0.3018
X_3	0.1896	0.1640	0.1522	0.1562	0.0799	0.1784	0.1799
X_4	0.1523	0.1274	0.1206	0.1238	0.0469	0.1405	0.1184

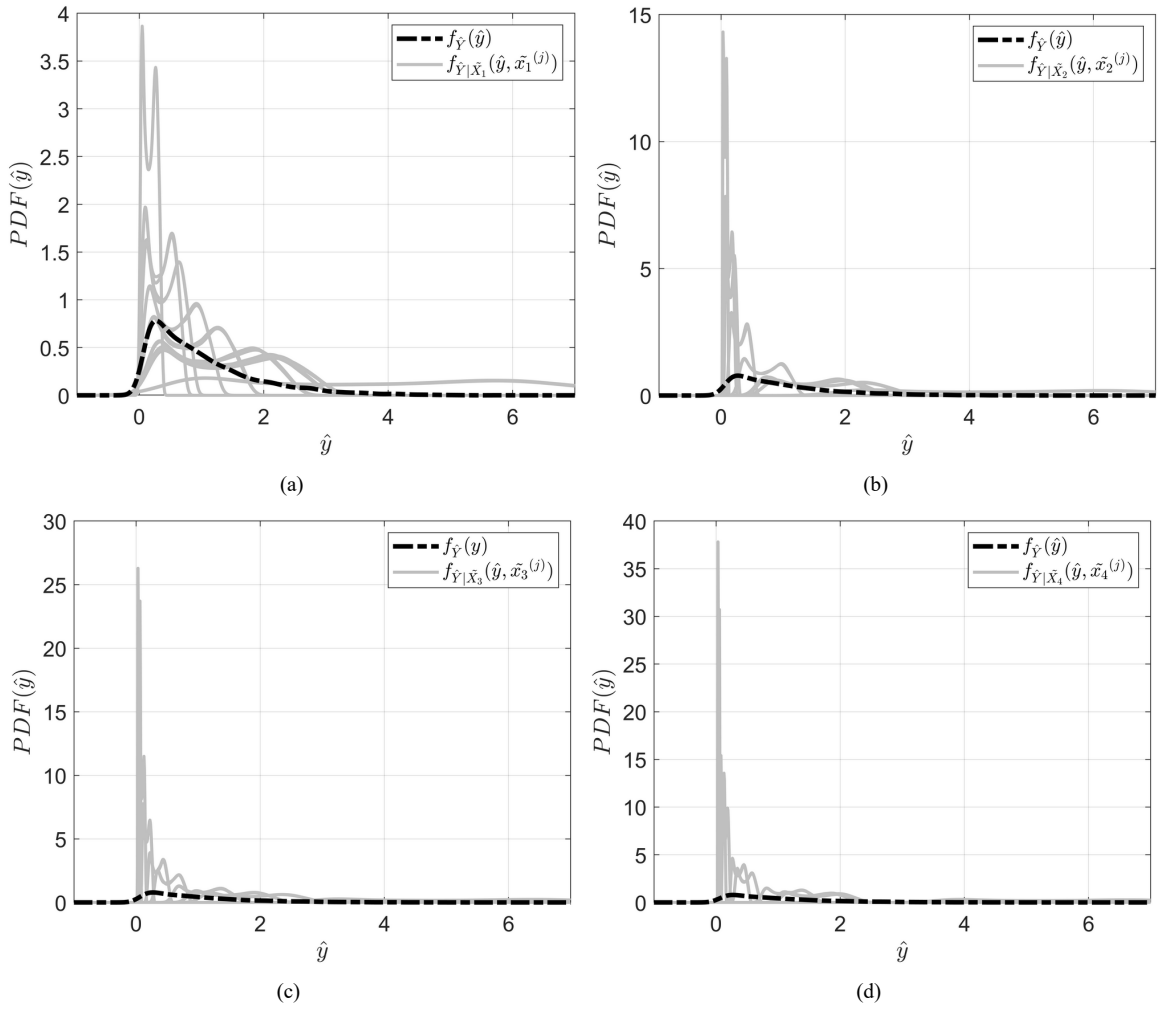


FIG. 14: Plots showing the variation of (a) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_1}\left(\hat{y}, \tilde{\boldsymbol{x}}_1^{(j)}\right)$, (b) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_2}\left(\hat{y}, \tilde{\boldsymbol{x}}_2^{(j)}\right)$, (c) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_3}\left(\hat{y}, \tilde{\boldsymbol{x}}_3^{(j)}\right)$, and (d) $f_{\hat{Y}|\tilde{\boldsymbol{X}}_4}\left(\hat{y}, \tilde{\boldsymbol{x}}_4^{(j)}\right)$ for the 35 CUT6 nodes (G-function)

TABLE 8: ${}^2\overline{\mathcal{D}}_i$ values for different $\mathcal{D}$ measures (G-function)

$\mathcal{D}$ measure	${}^2\overline{\mathcal{D}}_1$	${}^2\overline{\mathcal{D}}_2$	${}^2\overline{\mathcal{D}}_3$	${}^2\overline{\mathcal{D}}_4$
W	0.5314	0.7069	0.7731	0.8070
H	0.5136	0.8221	0.9493	1.0173
T	0.3677	0.5765	0.6882	0.7529
K	0.3424	0.5249	0.6164	0.6675
B	0.1931	0.5397	0.7379	0.8733
C	0.3416	0.4755	0.5297	0.5604

the interaction coefficien quantifying the probability of interaction between individuals, and r is the rate of recovery. The reason γk is not considered a single variable even though it always appears as a product is because γ is a property of the disease and varies from disease to disease, while k is a parameter which influence contact between people and can be controlled through policies and arrangements such as quarantine or isolation [44].

All the parameters are assumed to be uncertain and independent, and can be represented as

TABLE 9: ${}^2NS_i$ values for different $\mathcal{D}$ measures (G-function)

Variable	${}^2\boldsymbol{NS^W}$	${}^2\boldsymbol{NS^H}$	${}^2\boldsymbol{NS^T}$	${}^2\boldsymbol{NS^K}$	${}^2\boldsymbol{NS^B}$	${}^2\boldsymbol{NS^C}$	$\boldsymbol{S_T}$
X_1	0.1885	0.1555	0.1541	0.1592	0.0824	0.1791	0.5847
X_2	0.2508	0.2490	0.2417	0.2440	0.2302	0.2493	0.3018
X_3	0.2743	0.2875	0.2885	0.2866	0.3148	0.2777	0.1799
X_4	0.2863	0.3081	0.3156	0.3103	0.3726	0.2938	0.1184

$$\boldsymbol{X} = [X_1, X_2, X_3, X_4]^T = [\gamma, k, r, \delta]^T, \tag{38}$$

with their respective distributions given by

$$\gamma \sim \mathcal{U}([0,1]), \quad k \sim \text{Beta}(2,7), \quad r \sim \mathcal{U}([0,1]), \quad \text{and} \quad \delta \sim \mathcal{U}([0,1]). \tag{39}$$

The output of interest is the number of infected individuals, i.e., $I(t)$. The initial conditions are taken to be $S(0) = 900$, $I(0) = 100$, and $R(t) = 0$ with a total population size of $N = 1000$. A more detailed description of the model and the parameters can be found in [44].

In this example, since the distributions of the parameters are not the same, the PC basis functions are derived from a tensor product of univariate Legendre polynomials (in X_1, X_3, and X_4) and univariate Jacobi polynomials (in X_2) as suggested by the Wiener-Askey scheme. The number of bases considered in each direction was $N_{X_i} = 5$, leading to a total of $N = 625$ bases. The coefficient $[\hat{\rho}_i(t)]$ of the surrogate model

$$\hat{I}(t) = \hat{Y}(t) = \sum_{i=1}^{625} \hat{\rho}_i(t)\Phi_i(X_1, X_2, X_3, X_4) \tag{40}$$

are determined using the least-squares nonintrusive technique presented in Section 3.1.2. For this case the number of samples was considered to be $m = 10^4$. It should be noted that the coefficient are a function of time and need to be solved for, at each time instant. However, the pseudo-inverse operation of the matrix A needs only to be done once since it is time independent. Therefore, solving for $\hat{\rho}_i(t)$ at different times essentially becomes weighing the m realizations of the model appropriately.

For this example, a 15-point GQ rule is implemented to calculate the sensitivity indices ${}^1NS_i(t)$. A Gauss–Legendre set is used to determine ${}^1NS_1(t)$, ${}^1NS_3(t)$, and ${}^1NS_4(t)$ while a Gauss–Jacobi set is used to determine ${}^1NS_2(t)$ due to the nature of their distributions. The output PDFs and the class 1 conditional PDFs are evaluated at every time instant from 10^4 MC samples from $\Omega_{\tilde{\boldsymbol{X}}_i}$. This results in ${}^1\mathcal{D}_i(t)$ and correspondingly ${}^1NS_i(t)$ at every time instant. ${}^1NS_i(t)$ values have been plotted in Figs. 15(a)–15(f) for various $\mathcal{D}$s. The total time of simulation has been considered to be f ve units. We observe consistent results from all the $\mathcal{D}$s. We see that, initially, the variables can be ordered as X_1, X_2, X_3, X_4. At time $t = 0.5$ the variables are ordered as X_3, X_1, X_2, X_4 before eventually settling at X_3, X_4, X_1, X_2 after $t = 2$. It is evident from the plots that X_2 is least significan while X_3 becomes most significan after the transients are over.

The determination of Class 2 metrics for the epidemic model is slightly different than the previous examples. This is because the PDFs of the input variables are all not the same, and not all the PDFs are uniform or Gaussian. Hence in order to evaluate ${}^2\overline{\mathcal{D}}_i$ and ${}^2NS_i$, we require a set of points other than CUT and rely on GQ to average $\mathcal{D}$ over the $\Omega_{\tilde{\boldsymbol{X}}_i}$ space. For the moment-independent sensitivity indices ${}^2\overline{\mathcal{D}}_1$, ${}^2\overline{\mathcal{D}}_3$, and ${}^2\overline{\mathcal{D}}_4$, the variables over which the average is evaluated have one Beta distributed variable and two uniformly distributed ones. Hence, a GQ set of sigma points are derived for their evaluation from a tensor product of two uniform GQ set of points and one Beta GQ set of points. For this example, a 5-point GQ rule is used in each direction leading to a total of $5^3 = 125$ GQ points. However, for ${}^2\overline{\mathcal{D}}_2$, we need to calculate the average of $\mathcal{D}$ over $\Omega_{\tilde{\boldsymbol{X}}_2}$ which has three uniformly distributed variables. Hence, for this particular average, CUT6 is adopted to derive a set of 35 sigma points in the 3D $\Omega_{\tilde{\boldsymbol{X}}_2}$ subspace. Using these sigma points ${}^2\overline{\mathcal{D}}_i$ and consequently ${}^2NS_i$ at each time instant is calculated and presented in Figs. 16(a)–16(f) (where each figur corresponds to a distinct measure).

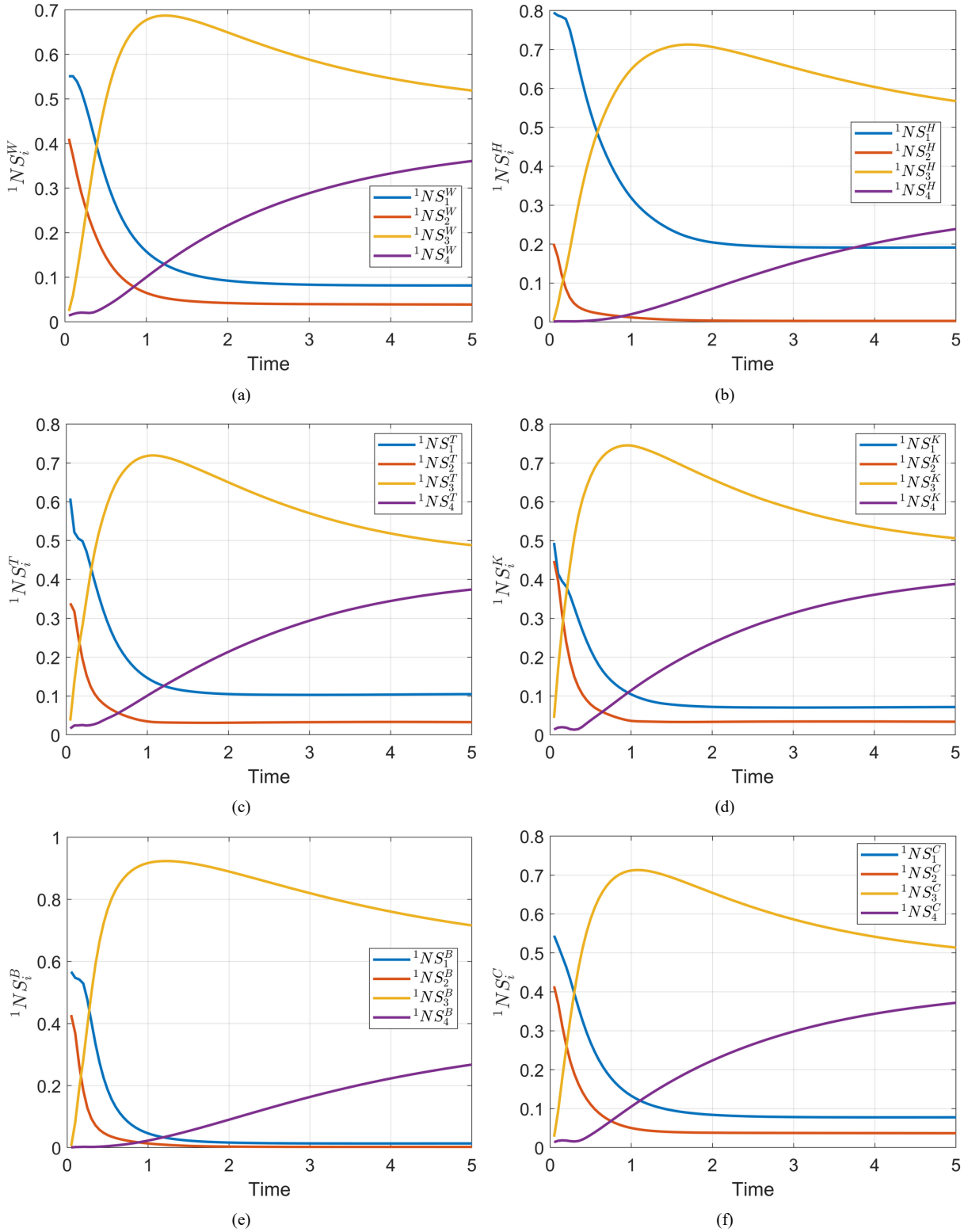


FIG. 15: Variation of (a) ${}^1NS_i^W$, (b) ${}^1NS_i^H$, (c) ${}^1NS_i^T$, (d) ${}^1NS_i^K$, (e) ${}^1NS_i^B$, and (f) ${}^1NS_i^C$ with time

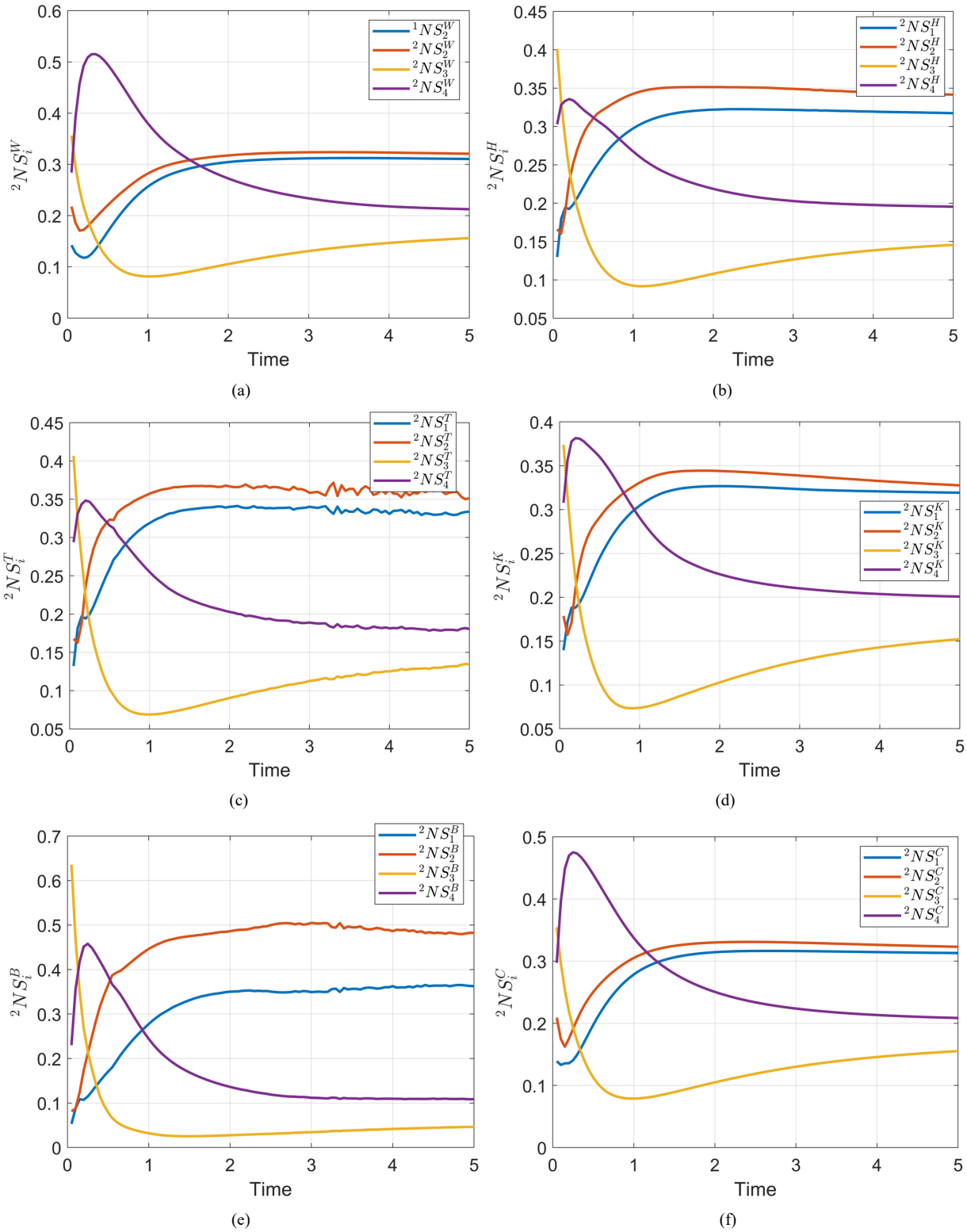


FIG. 16: Variation of (a) $^2NS_i^W$, (b) $^2NS_i^H$, (c) $^2NS_i^T$, (d) $^2NS_i^K$, (e) $^2NS_i^B$, and (f) $^2NS_i^C$ with time

Consistent with the variation of the Class 1 metrics, we observe from Figs. 16(a)–16(f) that, with time the order of ranking changes. Once again, we see that, initially the ranking starts in the order of X_1, X_2, X_3, X_4, followed by a transition to the order X_3, X_1, X_2, X_4, before eventually settling at X_3, X_4, X_1, X_2 at fina time. It should be noted that the order of importance is measured in a reverse manner for Class 2 metrics in comparison to Class 1 metrics. This is why we see a reversal in the magnitude of the curves in Figs. 16(a)–16(f) when compared to Figs. 15(a)–15(f).

To compare the results with the traditional Sobol' indices, we present the variation of the total Sobol' indices (S_{Ti}) with time for each input variable in Fig. 17. The indices are efficientl calculated using the PC coefficient as proposed by Sudret in [16]. It is interesting to note that for this example as well, we observe the same ranking of influenc from Sobol' as estimated by the non-moment-based NS metrics. This can be attributed to the fact that a majority of the uncertainty in the output quantity of interest can be quantifie by the second moment (variance). Hence, Sobol' indices perform reasonably well.

It should be pointed out here that Figs. 15(a)–15(f) are plotted without data at $t = 0$. This is because the initial value of the output of interest [i.e., $I(t)$] is already given and is a constant irrespective of the value of the parameters. Hence, the uncertainty in the parameters does not influenc the value of the output at $t = 0$.

5. CONCLUSION

The main motivation behind moment-independent sensitivity analysis measures stems from the fact that variance by itself is not sufficien to characterize all the uncertainty associated with an output of interest. To this end, the research community has endeavored to develop methods which are more dependent on the entire probability distribution of the outputs and the conditional outputs since that comprehensively quantifie the entire uncertainty.

In this article, we propose a generalization of the Borgonovo metric as a class of metrics where the statistical distance between the output and the conditional output could be the choice of the user depending on requirement and convenience. Furthermore, we also present a complementary class of measure (also based on the statistical distances between output and conditional distributions). In order to compute the measures efficientl , a surrogate model using PC is proposed to ease the sampling time. It has been identifie in the literature that PC could become prohibitive when the number of inputs increases due to the curse of dimensionality. However, as mentioned previously, if a screening method (such as the method of Morris) is used to reduce dimensionality of the problem prior to the application of PC, PC still remains relevant. Moreover, there is a report in the literature on PC [18] which can deal with higher uncertainties by smartly selecting its basis functions commonly known as sparse PC.

The proposed class of metrics eventually requires the computation of an average statistical distance. This average is evaluated using numerical integration schemes such as Gauss quadrature and conjugate unscented transform

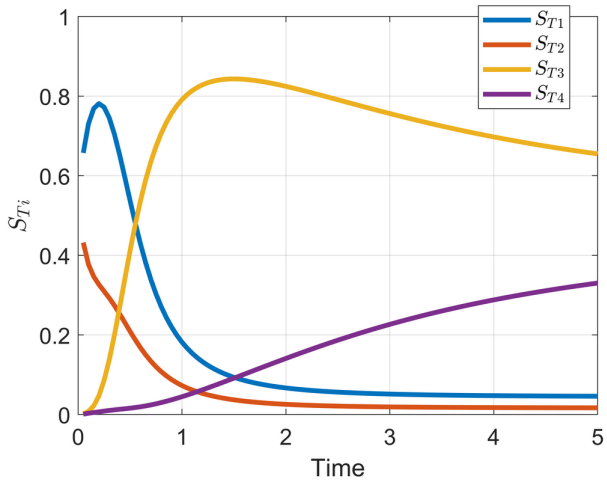


FIG. 17: Variation of the total Sobol' indices with time

to further reduce computation. Both sets of classes are then illustrated on numerical examples. Results present consistent rankings between the classes of measures and, for the Ishigami function case, also highlight the drawbacks of variance-based methods.

ACKNOWLEDGMENT

This material is based upon work supported through National Science Foundation (NSF) under Award Nos. CMMI-1537210 and CMMI-2021710.

REFERENCES

1. Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., *Sensitivity Analysis in Practice: A Guide to Assessing Scientifi Models*, New York: John Wiley & Sons, 2004.
2. Saltelli, A., Chan, K., and Scott, E.M., *Sensitivity Analysis*, Vol. 1, New York: Wiley, 2000.
3. Campolongo, F., Cariboni, J., and Saltelli, A., An Effective Screening Design for Sensitivity Analysis of Large Models, *Environ. Model. Software*, **22**(10):1509–1518, 2007.
4. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S., Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index, *Comput. Phys. Commun.*, **181**(2):259–270, 2010.
5. Borgonovo, E. and Plischke, E., Sensitivity Analysis: A Review of Recent Advances, *Eur. J. Oper. Res.*, **248**(3):869–887, 2016.
6. Da Veiga, S., Global Sensitivity Analysis with Dependence Measures, *J. Stat. Comput. Simul.*, **85**(7):1283–1305, 2015.
7. Rahman, S., The F-Sensitivity Index, *SIAM/ASA J. Uncertainty Quantif.*, **4**(1):130–162, 2016.
8. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global Sensitivity Analysis: The Primer*, New York: John Wiley & Sons, 2008.
9. Morris, M.D., Factorial Sampling Plans for Preliminary Computational Experiments, *Technometrics*, **33**(2):161–174, 1991.
10. Saltelli, A., Sensitivity Analysis for Importance Assessment, *Risk Anal.*, **22**(3):579–590, 2002.
11. Borgonovo, E., A New Uncertainty Importance Measure, *Reliab. Eng. Syst. Saf.*, **92**(6):771–784, 2007.
12. Chun, M.H., Han, S.J., and Tak, N.I., An Uncertainty Importance Measure Using a Distance Metric for the Change in a Cumulative Distribution Function, *Reliab. Eng. Syst. Saf.*, **70**(3):313–321, 2000.
13. Park, C.K. and Ahn, K.I., A New Approach for Measuring Uncertainty Importance and Distributional Sensitivity in Probabilistic Safety Assessment, *Reliab. Eng. Syst. Saf.*, **46**(3):253–261, 1994.
14. Borgonovo, E., Castaings, W., and Tarantola, S., Moment Independent Importance Measures: New Results and Analytical Test Cases, *Risk Anal.*, **31**(3):404–428, 2011.
15. Gamboa, F., Klein, T., and Lagnoux, A., Sensitivity Analysis Based on Cramér–von Mises Distance, *SIAM/ASA J. Uncertainty Quantif.*, **6**(2):522–548, 2018.
16. Sudret, B., Global Sensitivity Analysis Using Polynomial Chaos Expansions, *Reliab. Eng. Syst. Saf.*, **93**(7):964–979, 2008.
17. Crestaux, T., Le Maître, O., and Martinez, J.M., Polynomial Chaos Expansion for Sensitivity Analysis, *Reliab. Eng. Syst. Saf.*, **94**(7):1161–1172, 2009.
18. Blatman, G. and Sudret, B., Efficien Computation of Global Sensitivity Indices Using Sparse Polynomial Chaos Expansions, *Reliab. Eng. Syst. Saf.*, **95**(11):1216–1229, 2010.
19. Borgonovo, E., Castaings, W., and Tarantola, S., Model Emulation and Moment-Independent Sensitivity Analysis: An Application to Environmental Modelling, *Env. Model. Software*, **34**:105–115, 2012.
20. Adurthi, N., Singla, P., and Singh, T., Conjugate Unscented Transformation: Applications to Estimation and Control, *J. Dyn. Syst. Meas. Control*, **140**(3):030907, 2018.
21. Gibbs, A.L. and Su, F.E., On Choosing and Bounding Probability Metrics, *Int. Stat. Rev.*, **70**(3):419–435, 2002.
22. Bhattacharyya, A., On a Measure of Divergence between Two Statistical Populations Define by Their Probability Distributions, *Bull. Calcutta Math. Soc.*, **35**:99–109, 1943.
23. Baringhaus, L. and Henze, N., Cramér–von Mises Distance: Probabilistic Interpretation, Confidenc Intervals, and Neighbourhood-of-Model Validation, *J. Nonparametric Stat.*, **29**(2):167–188, 2017.

24. Alvarez-Esteban, P.C., Del Barrio, E., Cuesta-Albertos, J.A., and Matrán, C., Similarity of Samples and Trimming, *Bernoulli*, **18**(2):606–634, 2012.
25. Rachev, S.T., Klebanov, L., Stoyanov, S.V., and Fabozzi, F., *The Methods of Distances in the Theory of Probability and Statistics*, New York: Springer Science & Business Media, 2013.
26. Wiener, N., The Homogeneous Chaos, *Am. J. Math.*, **60**(4):897–936, 1938.
27. Ghanem, R.G. and Spanos, P.D., *Stochastic Finite Elements: A Spectral Approach*, New York: Springer New York, 1991.
28. Cameron, R.H. and Martin, W.T., The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals, *Ann. Math.*, **48**(2):385–392, 1947.
29. Xiu, D. and Karniadakis, G.E., The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations, *SIAM J. Sci. Comput.*, **24**(2):619–644, 2002.
30. Kim, K.K.K., Shen, D.E., Nagy, Z.K., and Braatz, R.D., Wiener's Polynomial Chaos for the Analysis and Control of Nonlinear Dynamical Systems with Probabilistic Uncertainties [Historical Perspectives], *IEEE Control Syst.*, **33**(5):58–67, 2013.
31. Hadigol, M. and Doostan, A., Least Squares Polynomial Chaos Expansion: A Review of Sampling Strategies, *Comput. Methods Appl. Mech. Eng.*, **332**:382–407, 2018.
32. Nandi, S. and Singh, T., Nonintrusive Global Sensitivity Analysis for Linear Systems with Process Noise, *J. Comput. Nonlinear Dyn.*, **14**(2):021003, 2019.
33. Feynman, R.P., Hibbs, A.R., and Styer, D.F., *Quantum Mechanics and Path Integrals*, North Chelmsford, MA: Courier Corporation, 2010.
34. Doucet, A., Godsill, S., and Andrieu, C., On Sequential Monte Carlo Sampling Methods for Bayesian Filtering, *Stat. Comput.*, **10**(3):197–208, 2000.
35. Stroud, A.H., *Approximate Calculation of Multiple Integrals*, Upper Saddle River, NJ: Prentice-Hall, 1971.
36. Caflisch R.E., Monte Carlo and Quasi-Monte Carlo Methods, *Acta Numer.*, **1998**:1–49, 1998.
37. Stroud, A.H. and Secrest, D., *Gaussian Quadrature Formulas*, Upper Saddle River, NJ: Prentice-Hall, 1966.
38. Gerstner, T. and Griebel, M., Numerical Integration Using Sparse Grids, *Numer. Algorithms*, **18**(3-4):209, 1998.
39. Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N., Monte Carlo Evaluation of Derivative-Based Global Sensitivity Measures, *Reliab. Eng. Syst. Saf.*, **94**(7):1135–1148, 2009.
40. Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Vol. 55, North Chelmsford, MA: Courier Corporation, 1965.
41. Plischke, E., An Effective Algorithm for Computing Global Sensitivity Indices (EASI), *Reliab. Eng. Syst. Saf.*, **95**(4):354–360, 2010.
42. Marrel, A., Iooss, B., Laurent, B., and Roustant, O., Calculations of Sobol' Indices for the Gaussian Process Metamodel, *Reliab. Eng. Syst. Saf.*, **94**(3):742–751, 2009.
43. Saltelli, A. and Sobol', I.M., About the Use of Rank Transformation in Sensitivity Analysis of Model Output, *Reliab. Eng. Syst. Saf.*, **50**(3):225–239, 1995.
44. Smith, R.C., *Uncertainty Quantification Theory, Implementation, and Applications*, Vol. 12, Philadelphia, PA: SIAM, 2013.

APPENDIX A. CONJUGATE UNSCENTED TRANSFORM

Consider the multivariate expectation integral:

$$E[g(\boldsymbol{X})] = \int \int \dots \int g(\boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{x}) \quad dx_1\, dx_1 \dots dx_n, \tag{A.1}$$

where we assume without loss of generality that the mean of the pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$ is 0. The discrete approximation of the expectation integral is

$$E[g(\boldsymbol{X})] \approx \sum_{i=1}^{N} w_i g(\boldsymbol{x}_i), \tag{A.2}$$

where $\boldsymbol{x}_i = \left[x_{(i,1)}, x_{(i,2)}, \cdots, x_{(i,n)}\right]^T$ is a quadrature point with an associated weight w_i.

Taylor series or Maclaurin series expansion of the analytic function $g(\boldsymbol{X})$ about the mean 0 results in Eq. (A.1) reducing to

$$E[g(\boldsymbol{X})] \approx \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_n=0}^{\infty}\frac{E\left[x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}\right]}{i_1!i_2!\cdots i_n!}\frac{\partial^{i_1+i_2+\cdots+i_n}g}{\partial x_1^{i_1}\partial x_2^{i_2}\cdots\partial x_n^{i_n}}(0) \tag{A.3}$$

and Eq. (A.2) resulting in

$$E[g(\boldsymbol{X})] \approx \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_n=0}^{\infty}\frac{\left(\sum_{i=1}^{N}w_i\left\{x_{(i,1)}^{i_1}x_{(i,2)}^{i_2}\cdots x_{(i,n)}^{i_n}\right\}\right)}{i_1!i_2!\cdots i_n!}\frac{\partial^{i_1+i_2+\cdots+i_n}g}{\partial x_1^{i_1}\partial x_2^{i_2}\cdots\partial x_n^{i_n}}(0). \tag{A.4}$$

Comparing Eqs. (A.3) and (A.4) results in the moment constraint equation:

$$\sum_{i=1}^{N}w_i\left\{x_{(i,1)}^{i_1}x_{(i,2)}^{i_2}\cdots x_{(i,n)}^{i_n}\right\} = E\left[x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}\right], \tag{A.5}$$

where $i_1 + i_2 + \ldots + i_n = d$ corresponds to the order of the moment.

Conjugate unscented transform (CUT) is an approach for identificatio of a sparse set of quadrature points for Gaussian and uniform distributions to match a specifie order of moments. For example, for a standard multivariate Gaussian distribution, all the odd-order moments are zero and the firs few even-order moments are shown in Table A1, where the moment constraint equations are determined for all possible permutations of the indices $\{i, j, k\} \in \{1, 2, 3, \ldots, n\}$, for the nth-dimensional Gaussian distributions. The solution to the moment constraint equations which capture the fourth-order moments is referred to as CUT-4. Assuming a symmetric distribution of the quadrature points about the origin along the principal axis σ_i at a distance r_1 and a second set of points lying symmetrically along the bisector axis c of all possible principal axis taken two at a time at a distance of r_2. For multivariate Gaussian distributions of dimension greater than 2, the closed form solution to the quadrature points and corresponding weights are shown in Table A2, where the total number of quadrature points $N = 2n + 2^n + 1$ and

$$r_1 = \sqrt{\frac{n+1}{2}}, \quad r_1 = \sqrt{\frac{n+2}{n-2}}, \tag{A.6}$$

$$w_1 = \frac{1}{r_1^4} = \frac{4}{(n+1)^2}, \quad w_2 = \frac{1}{2^n r_2^4} = \frac{(n-2)^2}{2^n(n+2)^2}. \tag{A.7}$$

Similarly quadrature points and associated weights for CUT-6 and CUT-8 points are determined by identifying points symmetric about the origin which satisfy all moments until the sixth and eighth orders, respectively. Details can be found in [20].

TABLE A1: First few even moments of standard Normal PDF

Moment	$\boldsymbol{E[X_i^2]}$	$\boldsymbol{E[X_i^2X_j^2]}$	$\boldsymbol{E[X_i^4]}$	$\boldsymbol{E[X_i^2X_j^4]}$	$\boldsymbol{E[X_i^6]}$	$\boldsymbol{E[X_i^2X_j^2X_k^2]}$
Value	1	1	3	3	15	1

TABLE A2: Quadrature points and weights for CUT-4

	Position	**Weights**
$1 \le i \le 2n$	$X_i = r_1\sigma_i$	$W_i = w_1$
$1 \le i \le 2^n$	$X_{i+2n} = r_2 c_i^n$	$W_{i+2n} = w_2$
Central weight	$X_0 = 0$	$W_0 = w_0$