Reliability Engineering and System Safety 142 (2015) 399–432

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Variable importance analysis: A comprehensive review

Pengfei Wei [a,*], Zhenzhou Lu [b,*], Jingwen Song [b]

[a] School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China
[b] School of Aeronautics, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form
30 March 2015
Accepted 25 May 2015
Available online 9 June 2015

Keywords:
Variable importance analysis
Difference-based
Regression technique
Random forest
Variance-based
Moment-independent
Graphic variable importance measures

ABSTRACT

Measuring variable importance for computational models or measured data is an important task in many applications. It has drawn our attention that the variable importance analysis (VIA) techniques were developed independently in many disciplines. We are strongly aware of the necessity to aggregate all the good practices in each discipline, and compare the relative merits of each method, so as to instruct the practitioners to choose the optimal methods to meet different analysis purposes, and to guide current research on VIA. To this end, all the good practices, including seven groups of methods, i.e., the difference-based variable importance measures (VIMs), parametric regression and related VIMs, nonparametric regression techniques, hypothesis test techniques, variance-based VIMs, moment-independent VIMs and graphic VIMs, are reviewed and compared with a numerical test example set in two situations (independent and dependent cases). For ease of use, the recommendations are provided for different types of applications, and packages as well as software for implementing these VIA techniques are collected. Prospects for future study of VIA techniques are also proposed.

Contents

* Corresponding authors.
E-mail addresses: wpf0414@163.com, pengfeiwei@nwpu.edu.cn (P. Wei), zhenzhoulu@nwpu.edu.cn (Z. Lu).

http://dx.doi.org/10.1016/j.ress.2015.05.018
0951-8320/

1. Introduction

Along with the rapid development of computer science and technique, a variety of computational models and numerical simulations have been developed for simulating and predicting the behavior of systems in nearly all fields of engineering and science such as aeronautical and astronautic engineering, chemistry and physics science, environmental science and technology, economics and education science. On the other hand, the last few decades have witnessed an explosive increase of the data volume in all kinds of large-scale scientific researches such as bioinformatics and related fields. To some degree, researchers from almost all the fields have reached an agreement on the necessity to perform variable importance analysis (VIA) based on these computational models and measured data. However, due to the wide dispersion of research fields and the lack of communication among different fields, the methodologies for VIA were independently developed in different research fields with different terminologies. These good practices in different disciplines, which will be reviewed in this article, are summarized in Fig. 1 with classification.

Researchers and practitioners working on computational models may face the problems of screening the relatively small group of important input variables from the tremendous candidate input variables (*variable prioritization setting*), fixing the large group of non-influential input variables at their nominal values without affecting the prediction accuracy or model output uncertainty (*variable fixing setting*), and determining how a reduction of the uncertainty of each input variable will influence the uncertainty in the output variable (*uncertainty reduction setting*) [1]. One can refer to Ref. [2] for an example of this type of analysis. VIA in these settings is mostly termed as "sensitivity analysis (SA)" in literature, where the word "sensitivity" used here is a general concept more related to "contribution" or "impact", not just the partial derivative which is commonly thought to be. This group of variable importance measures (VIMs) developed for computational models includes the difference-based VIMs, variance-based VIMs, moment-independent VIMs and the graphic VIMs, as shown in Fig. 1. This group of VIA techniques can also be termed as mathematical techniques.

In many disciplines such as bioinformatics, the objects operated by the analysts are measured data instead of computational models, and the analysts want to find the input variables that have obvious effect on the output variable based purely on data. This type of analysis is often dealt by statistical techniques such as measures of dependence, regression techniques and hypothesis tests. The correlation coefficient (CC), partial correlation coefficient (PCC), rank correlation coefficient (RCC), partial rank correlation coefficient (PRCC) and the moment-independent VIMs are all measures of dependence between the input and output variables. The parametric and nonparametric regression techniques aim at developing meta-model to approximate the true model response function. These techniques measure the variable importance either by the regression coefficients or by attributing the model output variance explained by the regression model to each of the input variables. The random forest, belonging to the group of nonparametric regression techniques, can provide the analysts with various types of VIMs, as indicated in Fig. 1. The hypothesis test techniques aim at testing the strength of relationship between the input and output variables, and use the probability-values (p-values) as measures of variable importance.

The reviews for "SA" methods developed for computational models are available in Refs. [3–12]. However, all these articles do not include the best practice for correlated input variables and the recently developed graphic VIMs. The reviews for statistical techniques (also called sampling-based techniques) are available

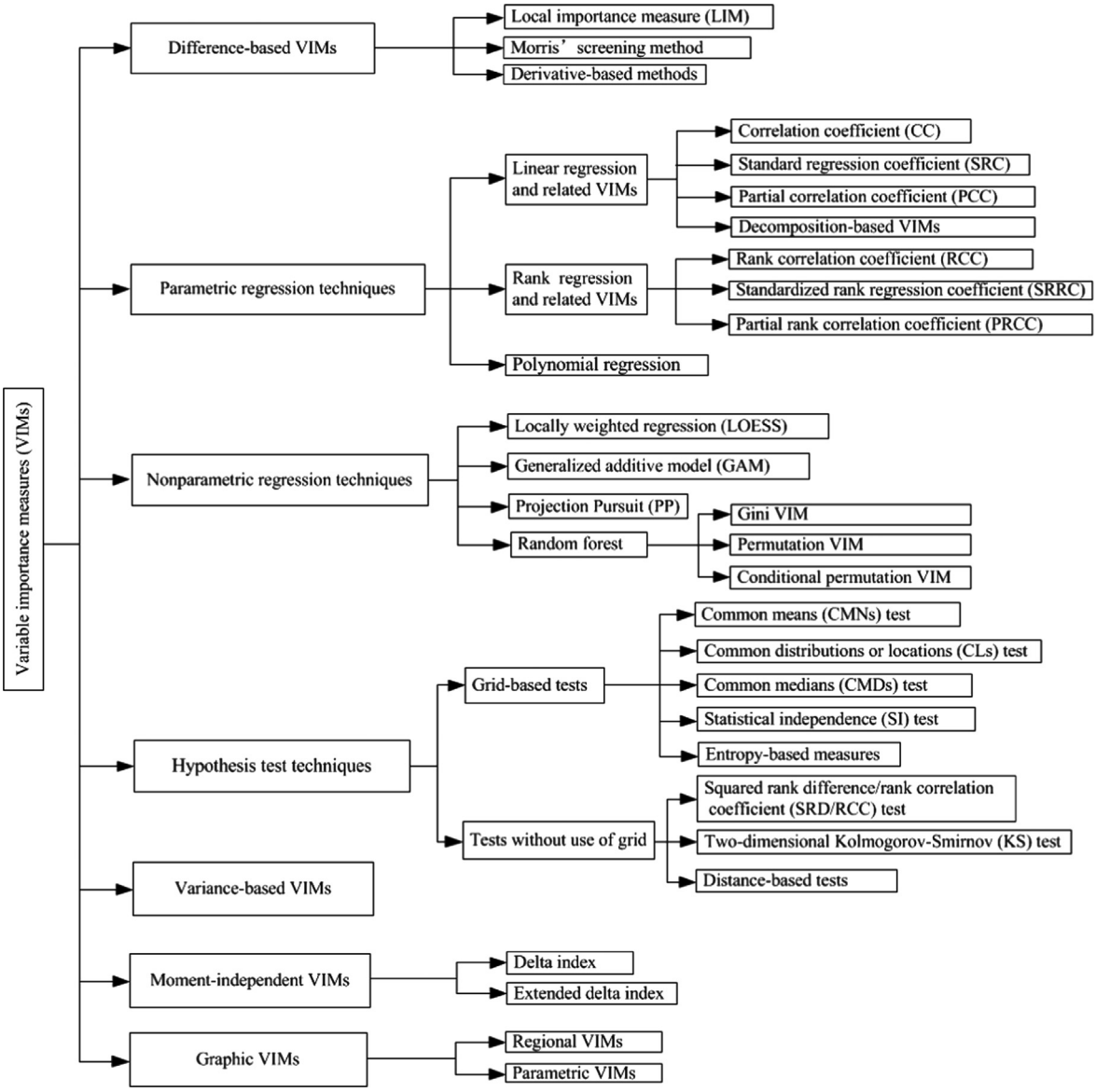

Fig. 1. Framework of this review.

in Refs. [13–16]. The nonparametric regression techniques, which belong to the group of statistical techniques, were reviewed and illustrated in Refs. [17–20]. In Refs. [21,22], the VIMs for correlated inputs based on multiple linear regression were reviewed. Refs. [23–25] provided reviews and comparisons of the random forest based VIMs. In Refs. [22,26], the VIMs based on multiple linear regression and random forest were compared. As far as we have learned, no review article has ever been presented for incorporating all these good practices and comparing the respective relative merits. This motivates us to carry out this work. In this review, all the activities of measuring variable importance are collectively termed as VIA.

Before the introduction of all these VIMs, there is a need to offer a definition for VIMs. Unfortunately, no unified definition can be carried out since that in different methods the importance of variables may be assessed in distinctly different ways. Summarily, VIM can be defined as

(1) a quantitative indicator that quantifies the change of model output value w.r.t. the change or permutation of one or a set of input variables, or
(2) an indicator that quantifies the contribution of the uncertainties of one or a set of input variables to the uncertainty of model output variable, or
(3) an indicator that quantifies the strength of dependence between the model output variable and one or a set of input variables.

Although defined in distinct forms and developed in different disciplines, all the VIMs summarized in Fig. 1 can be put into one of these three definitions. The difference-based VIMs and the three random forest based VIMs follows the first definition. The decomposition-based, variance-based, moment-independent and graphic VIMs are all belong to the second definition, and all the remaining VIMs in Fig. 1 as well as the moment-independent VIMs can be attributed to the third definition.

The purpose of this article is (a) to incorporate the good practices for VIA in nearly all the disciplines, (b) to compare the relative merits of each method, and (c) to explore the remaining challenges in VIA, so as (a) to guide the practitioners to choose the best methods to meet their special requirements, and (b) to instruct current research on VIA. To this end, we focus on these VIA techniques instead of the special applications of each method

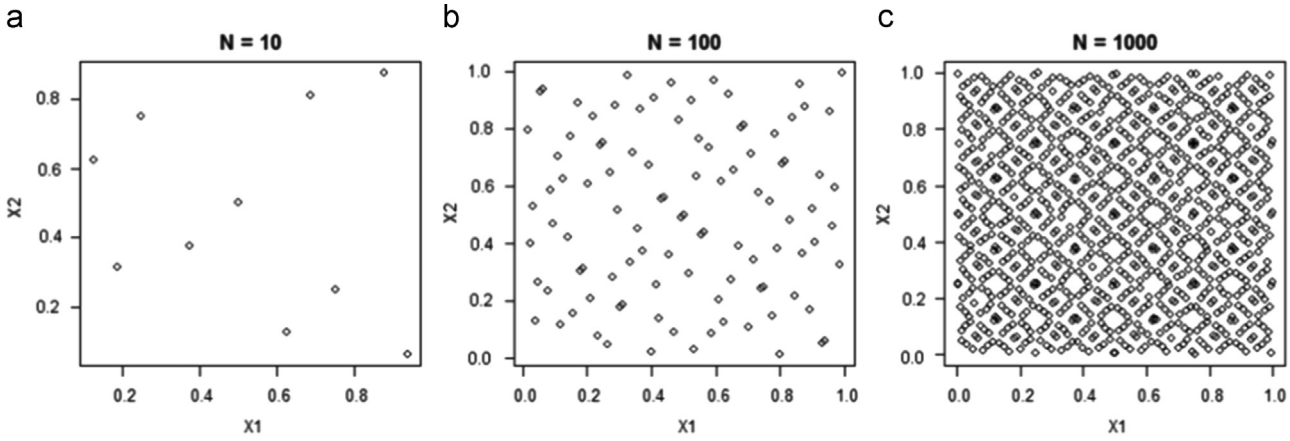

Fig. 2. Examples of LDS schedule for two-dimensional variables $\mathbf{X}=(X_1,X_2)$ independently and uniformly distributed between [0, 1], where the samples sizes N are set to be 10, 100 and 1000, respectively.

in specific discipline, and introduce a numerical example to test and compare these VIA techniques. For ease of application, we also collect available packages and software for implementing these VIA techniques.

The remaining of this article is organized according to the framework in Fig. 1. Section 2 gives some preparing works. Section 3 reviews the difference-based VIMs, followed by the parametric regression techniques and related VIMs in Section 4. Section 5 introduces the nonparametric regression techniques, but the random forest and related VIMs are organized in Section 6. The hypothesis test techniques for VIA are introduced in Section 7. In Section 8, the variance-based VIMs are introduced for independent and dependent inputs separately. The moment-independent VIMs are given in Section 9, followed by the newly developed graphic VIMs in Section 10. Section 11 concludes this article, gives discussions and comparisons of the reviewed VIA techniques, provides recommendations to meet different applications, summarizes the available packages and software and proposes some prospects.

2. Some preparing works

2.1. Uncertainty characterization and propagation

Uncertainties presented in real applications can be classified into epistemic uncertainties and aleatory ones depending on the sources of these uncertainties [27–33]. The epistemic uncertainties results from the insufficient knowledge of variables or events, thus can be reduced by collecting and learning more information on the events or variables; however, the aleatory uncertainties are due to the random nature of events or variables, thus cannot be reduced through further study. An example for epistemic uncertainties is the experts' opinion on the distribution of a random variable, and that for aleatory uncertainties is the instantaneous locations of molecule. Both types of uncertainties presented in model inputs will result in divergence of model output. The focus of uncertainty analysis is to propagate the uncertainties of the input variables through the computational model and measure the uncertainty of model output, while VIA aims at quantifying the effect of each input variable on the uncertainty of model output. Without loss of generality, we assume throughout this review that the uncertainties presented in the input variables are all due to epistemic uncertainties. For discussion of aleatory and epistemic uncertainties, see Refs. [27–31] for details. For VIA involving both aleatory and epistemic uncertainties, see Refs. [32,33] for details.

With the standardized probabilistic theory, the epistemic uncertainties of the n-dimensional model input variables $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ are characterized by the joint probability density function (PDF) $f_{\mathbf{X}}(\mathbf{x})$ or the joint cumulative distribution function (CDF) $F_{\mathbf{X}}(\mathbf{x})$ or the joint complementary cumulative distribution function (CCDF) $\overline{F}_{\mathbf{X}}(\mathbf{x})$. When the n input variables are independent, it holds that $f_{\mathbf{X}}(\mathbf{x})=\prod_{i=1}^{n}f_{X_i}(x_i)$, $F_{\mathbf{X}}(\mathbf{x})=\prod_{i=1}^{n}F_{X_i}(x_i)$ and $\overline{F}_{\mathbf{X}}(\mathbf{x})=\prod_{i=1}^{n}\overline{F}_{X_i}(x_i)$, where $f_{X_i}(x_i)$, $F_{X_i}(x_i)$ and $\overline{F}_{X_i}(x_i)$ are the marginal PDF, marginal CDF and marginal CCDF of X_i, respectively, and $\overline{F}_{X_i}(x_i)=1-F_{X_i}(x_i)$. The specifications of these CDFs (or PDFs or CCDFs) often involve an expensive analysis task, and are typically dealt by expert review. A relatively cheap strategy is to first specify raw characterizations for these CDFs, and perform VIA techniques to determine the important input variables that deserve to be further studied [16]. One should note that the supports of preliminarily specified crude CDFs should contain the supports of true CDFs so as not to neglect the sensitivity information of important variables (avoiding type II error). The expert review process commonly involves specifying the distribution types of variables by expert opinions and then estimating the distribution parameters with interval estimation, Bayesian estimation and/or other parameter estimation methods [16,34].

After the uncertainties of the input variables being characterized, these uncertainties need to be propagated through the computational model represented by $Y=g(\boldsymbol{X})$ so as to determine the uncertainty characterization of the model output Y and to study the effect of each input variable on the uncertainty of Y. In real application, the output variables may be multivariate or time-dependent. Developing VIMs for this kind of models is an active research fields nowadays (e.g., see Refs. [35–39]), but, by now, there is no widely accepted methods. In this review, only time-independent univariate model output is concerned. Given a computational model and the CDFs (or PDFs or CCDFs) of the input variables, the uncertainty analysis involves three steps, that is, (a) generating samples of input variables, (b) computing the model output values for each set of input samples by running the computational model, and (c) characterizing the uncertainty of model output with sample variance, PDF estimator, empirical CDF and/or other quantities based on the output samples obtained in step (b). This procedure often involves a large number of model runs, thus may be impractical for computationally expensive models. To deal with this type of problem, a simplified meta-model is usually first established to approximate the real model based on relatively small number of samples, and then the uncertainty is propagated by calling the meta-model instead of the real model. The first procedure of uncertainty analysis is known as "simulation", while the second procedure is named as "meta-model" or "response surface". Both procedures and

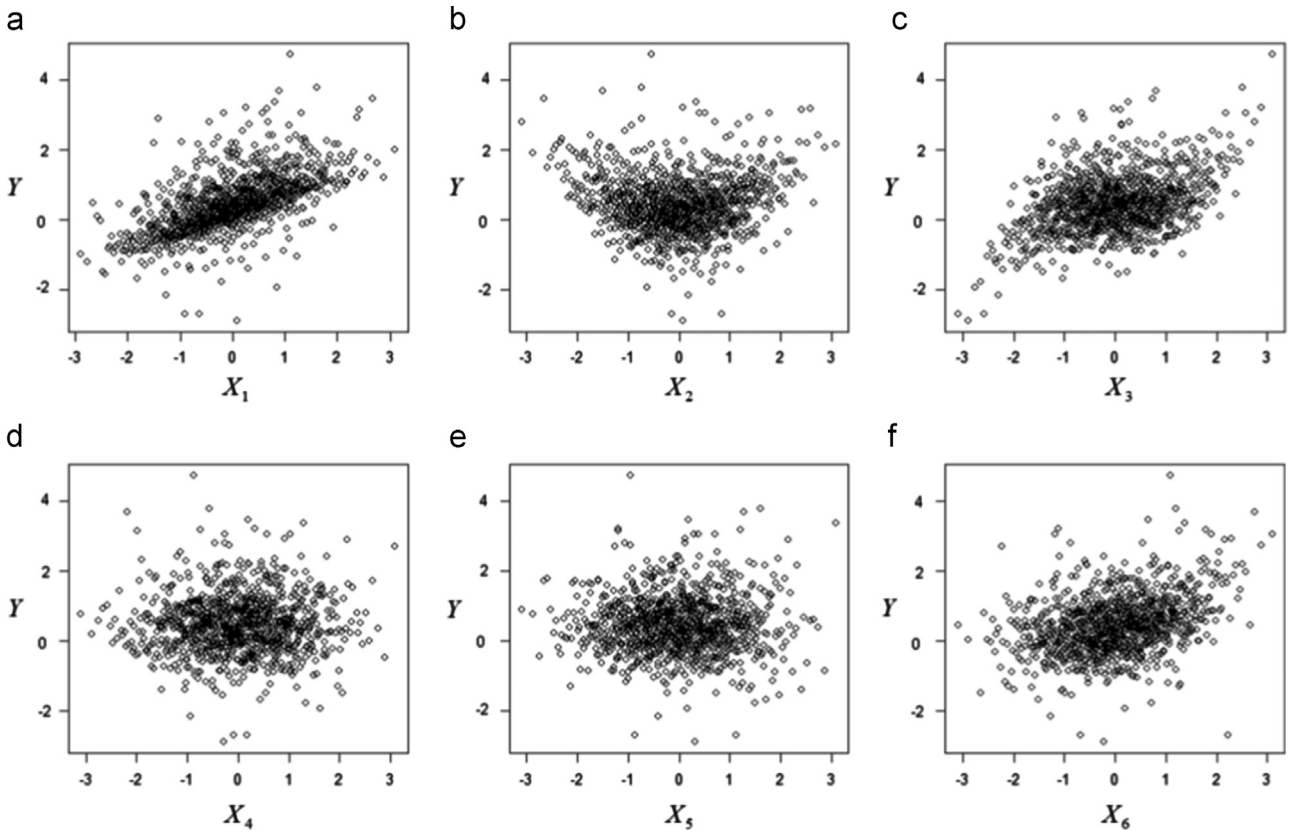

Fig. 3. Scatter plots of the response function in Eq. (1) for the case of independence with $N=1000$.

their products can be used for measuring variable importance, as will be indicated repeatedly in the reminder of this review.

2.2. Sampling schedules

Many procedures for uncertainty analysis and VIA involve generating random samples for the input variables. Many sampling schedules are available for this purpose. The commonly used schedules include simple random sampling (SRS), Latin hypercube sampling (LHS) [40], Latin supercube sampling (LSS) [41] and quasi-random sampling (e.g., Sobol's low-discrepancy sequence, LDS) [42,43]. Choosing an appropriate sampling schedule before performing the VIA techniques is certainly necessary as that different schedules result in definitely different convergence rates when used for estimating the various VIMs. Generally, the LHS, LSS and LDS schedules are nearly always more efficient than SRS. In the context of variance-based VIMs, the LDS schedule has been proved to be more efficient than both LHS and LSS [44,45]. Of course, the performances of different sampling schedules may depend on VIMs to be estimated, the LDS is superior to both LHS and LSS for variance-based VIMs does not necessarily indicate that it performs better for other VIMs. Comparing the performances of these sampling schedules is not the focus of this review, and only LDS schedule will be used in the subsequent contexts.

The LDS generates random sample matrix $\boldsymbol{M_X}$ of dimension $N \times n$ with each column of samples following uniform distribution in [0, 1] and each pair of columns being orthogonal. Each column of samples can then be transformed to samples following any other type of distribution with the corresponding inverse CDF. As $N \to \infty$, each column of samples tends to be best uniformly distributed, and for small N, the samples are also well distributed. The rationales of LDS will not be discussed in detailed here since it is not the focus of this review and will spend a lot of spaces. Throughout this review, the LDS schedule is implemented with the 'sobol' function in the package 'randtoolbox' [46] developed for R program [47], which is an efficient implementation of the algorithm in Ref. [48]. An example of the LDS schedule to generate samples of sizes $N=10$, 10^2 and 10^3 for 2-dimensional variables $\boldsymbol{X}=(X_1, X_2)$ independently and uniformly distributed between [0, 1] is shown in Fig. 2.

Correlations between input variables are ubiquitous in real applications, thus generating samples with desired correlation structure is necessary for VIA. However, generating samples for complex correlation structures is generally a rather challenging task. In Ref. [49], Iman and Conover developed an efficient procedure for injecting rank correlations to sample matrix generated with any sampling schedule. This procedure involves establishing an auxiliary sample matrix $\mathbf{R}^*$ of dimension $N \times n$ with desired rank correlations, and then rearranging the samples of $\mathbf{M_X}$ in each column so that they have the same ordering with the corresponding column in $\mathbf{R}^*$. In Refs. [16,40], this procedure has been used for injecting rank correlation structure to the sample matrix generated by LHS schedule. In this review, this procedure is extended to LDS schedule, and briefly summarized as follows.

Step 1 Generate a $N \times n$ sample matrix $\mathbf{M_X}$ with LDS schedule such that each column is uniformly distributed between 0 and 1, and each column is independent with others. Transform the ith column of $\mathbf{M_X}$ with the inverse CDF $F_{X_i}^{-1}(\bullet)$, then the ith column of $\mathbf{M_X}$ follows the marginal distribution $F_{X_i}(x_i)$.

Step 2 Write the target rank correlation matrix $\mathbf{C}$ as $\mathbf{C}=\mathbf{PP}'$, where $\mathbf{P}$ is a lower triangular matrix.

Step 3 Let $\mathbf{a}=\{\Phi(i/(N+1))\}_{i=1,2,\ldots,N}$, where $\Phi(\bullet)$ is the CDF of standard normal distribution. Let $\mathbf{R}$ be a $N \times n$ matrix with each column being a random permutation of $\mathbf{a}$.

Step 4 Estimate the Pearson correlation matrix $\mathbf{T}$ of $\mathbf{R}$, and write $\mathbf{T}=\mathbf{QQ}'$ so as to find a lower triangular matrix $\mathbf{Q}$.

Step 5 Let $\mathbf{S}=\mathbf{PQ}^{-1}$ and $\mathbf{R}^*=\mathbf{RS}'$.

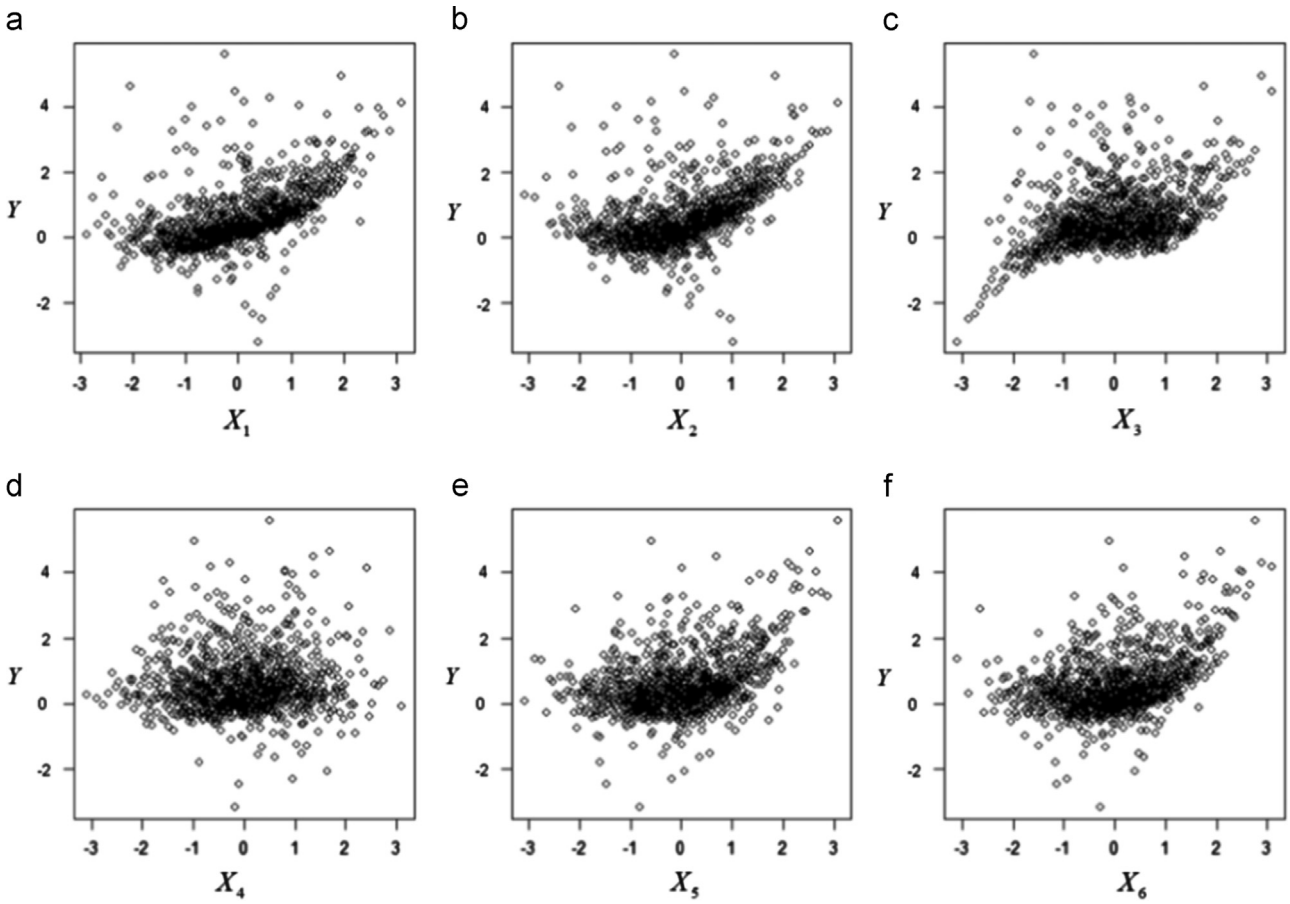

Fig. 4. Scatter plots of the response function in Eq. (1) for the case of dependence with $N=1000$.

Step 6 Rearrange the samples of $\mathbf{M_X}$ in each column so that they have the same ordering with the corresponding column in $\mathbf{R}^*$.

For detailed rationale of the above procedure, one can refer to Ref. [49].

2.3. Test example

Throughout this review, different VIA techniques are illustrated and compared with the computational model represented by the following response function:

$$Y = X_1/2 + X_2^2/4 + X_3^3/6 + X_5X_6\exp(X_4)/16 + X_6\exp(X_5)/12 + \exp(X_6)/8, \quad (1)$$

where $X_i (i=1,2,\ldots,6)$ are six random input variables following standard normal distribution. Two cases are considered: independence and dependent cases. For independent case, all these six input variables are independent with each other, and for dependent case, the RCCs ρ_{12} (between X_1 and X_2) and ρ_{56} (between X_5 and X_6) are both set to be 0.9, and the other RCCs are set to be 0. The scatter plots of Y against each input variable in the case of independence are shown in Fig. 3, where the sample size is set to be $N=1000$. As can be seen, the relationship between Y and X_1 is approximately linear, and that between Y and X_2 is nonlinear and non-monotonic. Y is nonlinearly but monotonically dependent on X_3. No specified pattern can be found for the relationship between Y and X_4 as well as X_5. The relationship between Y and X_6 is approximately monotonic but the nonlinearity cannot be specified. The scatter plots for the dependent case are shown in Fig. 4. As can be seen, the pattern of the relationship between Y and each correlated input variable is obviously different with that in the independent case.

3. Difference-based VIMs

In this section, the first group of VIA techniques, which are all based on differences or partial derivatives, is reviewed and compared. This group includes the local methods, Morris' screening method and the recently developed derivative-based method.

3.1. Local methods

The local methods are usually less informative than other VIMs, however, due to the simplicity and ease of calculation, they are still frequently used in many fields involving computational models. The most commonly used local measure is the local importance measure (LIM) [3,11].

Let $\boldsymbol{Y}=g(\boldsymbol{X})$ denote the model response function (also called g-function), where Y is the univariate model output of interest, and $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)^T$ is the vector of n-dimensional input variables. The LIM of an individual input X_i is defined as the partial derivative of model output w.r.t. X_i [11]:

$$Y_i' = \left.\frac{\partial Y}{\partial X_i}\right|_{\mathbf{x}^*}, \quad (2)$$

where $\mathbf{x}^*=(x_1^*,x_2^*,\ldots,x_n^*)$ is a fixed point in the input space, at which the LIM is computed.

The local index Y_i' measures the change of model output Y when the input X_i is perturbed at the point $\mathbf{x}_i^*$ and the other inputs are fixed at this point. The higher Y_i' is, the more sensitive Y is to X_i. The local index Y_i' is related to the first order Taylor expansion of the g-function:

$$Y = g(\mathbf{X}) \approx g(\mathbf{x}^*) + \sum_{i=1}^{n} Y_i'(X_i - x_i^*). \tag{3}$$

For a linear model, Y_i' keeps constant at any point. Given an increment (or decrement) of X_i by Δ_i, the increment (or decrement) of the model output Y can be computed by $\Delta_Y = Y_i'\Delta_i$. For a nonlinear model, the value of the LIM Y_i' varies with the nominal value $\mathbf{x}^*$, and the change Δ_Y of the model output Y w.r.t. the change Δ_i of X_i cannot be computed purely based on Y_i'.

With the first order approximation in Eq. (3), the variance $V(Y)$ of model output can be decomposed as [11]:

$$V(Y) = \sum_{i=1}^{n} Y_i'^2 V(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Y_i'Y_j' Cov(X_i, X_j). \tag{4}$$

Thus, when the input variables are independent with each other, the second term in Eq. (4) equals zero, and $Y_i'^2 V(X_i)/V(Y)$ equals the percentage of model output variance explained by X_i. Due to this reason, the LIM is commonly normalized as [11]:

$$LIM_i = \frac{SD(X_i)}{SD(Y)} \left.\frac{\partial Y}{\partial X_i}\right|_{\mathbf{x}^*}, \tag{5}$$

where $SD(X_i)$ and $SD(Y)$ indicates the standard deviations (SDs) of X_i and Y, respectively. One can also normalize the LIM by the nominal values of Y and X_i as:

$$LIM_i = \frac{x_i^*}{g(\mathbf{x}^*)} \left.\frac{\partial Y}{\partial X_i}\right|_{\mathbf{x}^*}. \tag{6}$$

In Ref. [50], the LIMs defined in Eqs. (5) and (6) have been generalized to constrained input variables.

When the input variables are correlated with each other, $Y_i'^2 V(X_i)/V(Y)$ can be interpreted as the uncorrelated contribution of X_i to $V(Y)$, and $2Y_i'Y_j' Cov(X_i, X_j)/V(Y)$ can be explained as the contribution of the correlation between X_i and X_j to $V(Y)$.

The importance index LIM_i only reflects the individual effect of the input X_i to the model output, and cannot tell the interaction effect of pair of input variables. For the later purpose, one can define the second order normalized LIM as:

$$LIM_{ij} = \frac{x_i^* x_j^*}{g(\mathbf{x}^*)} \left.\frac{\partial^2 Y}{\partial X_i \partial X_j}\right|_{\mathbf{x}^*} \quad \text{or} \quad LIM_{ij} = \frac{SD(X_i)SD(X_j)}{SD(Y)} \left.\frac{\partial^2 Y}{\partial X_i \partial X_j}\right|_{\mathbf{x}^*}. \tag{7}$$

Similarly, one can carry out higher order local importance index for a group of inputs for measuring their joint effect on the model output Y.

Other local VIMs include the differential importance measure (DIM) [50–52] and the finite change decomposition based VIMs [53], both of which attribute the finite change of model output to each of the input and/or the interactions of inputs with different strategies so as to measure the individual effect of the small change of each input and their interaction effects on the change of model output. One can refer to the respective reference for details.

The key to estimate LIM is the calculation of partial derivatives Y_i' for $i = 1, 2, \ldots, n$. The partial derivative Y_i' can be expressed in terms of the following limit:

$$Y_i' = \lim_{\Delta_i \to 0} \frac{g\left(x_1^*, \ldots, x_{i-1}^*, x_i^* + \Delta_i, x_{i+1}^*, \ldots, x_n^*\right) - g\left(x_1^*, \ldots, x_{i-1}^*, x_i^*, x_{i+1}^*, \ldots, x_n^*\right)}{\Delta_i}, \tag{8}$$

and can be simply approximated with the difference quotient:

$$Y_i' \cong \frac{g\left(x_1^*, \ldots, x_{i-1}^*, x_i^* + \Delta_i, x_{i+1}^*, \ldots, x_n^*\right) - g\left(x_1^*, \ldots, x_{i-1}^*, x_i^*, x_{i+1}^*, \ldots, x_n^*\right)}{\Delta_i}. \tag{9}$$

where Δ_i is a small perturbation of X_i, which should be carefully determined since unsuitable choice of the perturbation may produce misleading sensitivity information. This simple procedure is commonly called Brute-Force method. Eq. (9) is in fact the first order numerical differential formula derived from interpolation formula [54], and the approximation error is $-(\Delta_i/2)(\partial^2 Y/\partial X_i^2)|_{\mathbf{x}=\left(x_1^*, \ldots, x_{i-1}^*, \xi_i, x_{i+1}^*, \ldots, x_n^*\right)}$, where $\xi_i \in \left[x_i^*, x_i^* + \Delta_i\right]$ when $\Delta_i > 0$. One can also compute the partial derivatives with higher order numerical differential formula so as to improve the accuracy of estimates, but this indicates that more times of g-function evaluations are needed.

For a large system represented by a set of partial differential equations, there are also many other methods available for the computation of the partial derivatives such as the direct method [55,56], the Green's function method [57], the automated differentiation [54,58], the forward method and the adjoint method [59,60]. The readers can refer to Ref. [3] for more detail and comparison of these computational methods.

From the above discussion, several advantages of the local method can be summarized. First, it can be applied to both computational models with deterministic input variables and those with uncertain input variables. Second, for a linear model, the first order derivative Y_i' contains the information of global

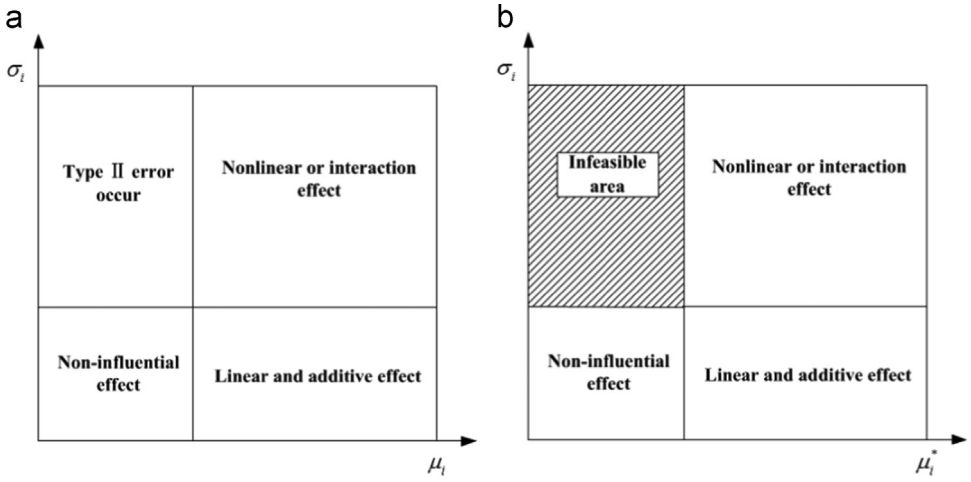

Fig. 5. Illustration of the three types of effect of the model inputs on output specified by the measures: (a) μ_i and σ_i, and (b) μ_i^* and σ_i.

importance measures. Third, compared with the global importance measures, the LIM is computationally cheaper. When the Brute-Force method is used, the total number of the g-function evaluations for computing all the first order LIMs is $(n+1)$. This can be especially appealing when the model under investigation is computationally expensive. However, one should bear in mind that the Brute-Force method is not always suitable for accurately estimating the partial derivatives as that small value of perturbation Δ_i may lead to round-off errors, and large value of Δ_i may result in truncation errors [54]. In many real applications, reasonable tradeoffs for the specification of perturbations are not always available, and in these cases, other advanced algorithms such as the adjoint method should be used.

The disadvantages of the local methods are also obvious. Due to their local nature, the LIM only reflects the local sensitivity information of model output to the input variables, and for nonlinear uncertain models with input variables characterized by intervals or probability distributions, the VIA results (variable importance ranking) obtained by LIM highly depend on the nominal point $\mathbf{x}^*$.

3.2. Morris' screening method

The screening method [60], proposed by Morris in 1991, is one of the most popular VIA techniques for screening the non-influential variables from a large number of model input variables with moderate computational cost.

For illustrating this method, it is usually assumed the input space to be the n-dimensional unit hypercube H^n. One notes that this can be easily extended to general cases since that the distribution function of any continuous random variable follows uniform distribution in the range [0, 1] and there is a one-to-one mapping between each input variable and its distribution function.

The basic idea of Morris' screening method is to first divide the range of each input variable into p levels so as to discretize the input space H^n into $(p-1)^n$ elements, and then compute the elementary effect of each variable using the information of the grid points, where the elementary effect $EE_i(\mathbf{x})$ of the input variable X_i at the point $\mathbf{x} \in H^n$ is defined as:

$$EE_i(\mathbf{x}) = \frac{g(\boldsymbol{x}+\Delta_i \boldsymbol{e_i}) - g(\boldsymbol{x})}{\Delta_i}, \tag{10}$$

where $\mathbf{e}_i$ is a n-dimensional vector with the ith element being unit and the other components being zero, Δ_i is a preselected step value in $\{1/(p-1), 2/(p-1), \ldots, 1-1/(p-1)\}$. The point $\mathbf{x}$ is selected to promise that $\mathbf{x}+\Delta_i \mathbf{e_i}$ is still in H^n.

Supposing the elementary effect $EE_i(\mathbf{x})$ follows distribution F_i, Morris proposed two statistics for measuring the importance of X_i, i.e., the mean μ_i and standard derivation σ_i of the distribution F_i, which are estimated by [60]

$$\mu_i = \frac{1}{J}\sum_{j=1}^{J} EE_i(\mathbf{x}_j), \quad \text{and} \quad \sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{J}\left(EE_i(\mathbf{x}_j) - \mu_i\right)^2} \tag{11}$$

respectively, where $EE_i(\mathbf{x}_j)$ is the jth sample of elementary effect $EE_i(\mathbf{x})$, and J is the total number of the estimated elementary effect of X_i. The sampling schedule of generating samples for the elementary effect will be discussed later in this subsection.

Based on the values of the two measures μ_i and σ_i, the model inputs can be classified into three groups. When μ_i is large and σ_i is small, then the model output Y is linear or at least additive w.r.t. X_i. If both μ_i and σ_i are small, then X_i is non-influential. If both μ_i and σ_i are large, then X_i has nonlinear or interaction effect on Y. These three situations are schematically illustrated in Fig. 5(a).

The main issue of the measure μ_i is that, when the model output Y is non-monotonic w.r.t. the input X_i, the elementary effects with opposite signs may cancel each other, leading to type ⓘ error, i.e. failing to identify the influential inputs. This type of error can be detected by the values of the statistics μ_i and σ_i. If μ_i is small and σ_i is large, then it is believed the type II error occurs, as shown in Fig. 5(a).

For avoiding the type II error, Campolongo et al. [61] suggested to use a new measure μ_i^* instead of μ_i:

$$\mu_i^* = \frac{1}{J}\sum_{j=1}^{J} \left|EE_i(\mathbf{x}_j)\right|. \tag{12}$$

The statistic μ_i^* quantifies the individual effect of X_i on the model output, and σ_i measures the nonlinear or interaction effects of X_i. An unfortunate reality is that σ_i cannot discriminate between nonlinear and interaction effects. Using the measures μ_i^* and σ_i, the three types of effect can be correctly identified, as shown in Fig. 5(b).

It is shown by Eqs. (11) and (12) that, for computing the statistics μ_i, μ_i^* and σ_i for one input, one needs to compute a total number of J elementary effects. The simplest way to do this is to first generate J samples of model inputs by LDS schedule, and then perturb the sample values of the input variables with Δ one by one at these sample points. This methods requires $2J$ simulations for each input, thus the total number of g-function evaluations is $2nJ$. In the past twenty years, many advanced strategies have been developed for relieving this computational cost.

In the original article, Morris [60] proposed an efficient strategy which consists of generating J trajectories, each of which provides one elementary effect for every input variable. The total number of model runs is $N = J(n+1)$. We denote this strategy as Morris' design. The J trajectories are generated by random design. Let $\boldsymbol{x}_j^{(i)}$ denote the ith point of the jth trajectory, where $i = 1, 2, \ldots, n+1$ and $j = 1, 2, \ldots, J$. Suppose now the range of each input has been divided into p levels, and the space H^n is discretized into $(p-1)^n$ elements. This discretization produces p^n grid points in the input space H^n. For building the jth trajectory, a base point $\mathbf{x}_j^*$ is first generated randomly from the p^n grid points. Then the first point $\mathbf{x}_j^{(1)}$ is obtained by increasing one or multiple components of $\mathbf{x}_j^*$ by Δ such that $\mathbf{x}_j^{(1)}$ is still in the space H^n. The second point $\mathbf{x}_j^{(2)}$ is generated by either increasing or decreasing the i_2th components of $\mathbf{x}_j^{(1)}$ with perturbation Δ, where i_2 is randomly selected in the set $\{1, 2, \ldots, n\}$. Whether decreasing or increasing the i_2th components of $\mathbf{x}_j^{(1)}$ is based on the principle that $\mathbf{x}_j^{(2)}$ is still in the input space H^n. The ith points $\mathbf{x}_j^{(i)}$ is obtained as $\mathbf{x}_j^{(i)} = \mathbf{x}_j^{(i-1)} + \mathbf{e}_{i_i}\Delta$ or $\mathbf{x}_j^{(i)} = \mathbf{x}_j^{(i-1)} - \mathbf{e}_{i_i}\Delta$, where i_i is randomly selected in the set $\{1, 2, \ldots, n\} - \{i_2, i_3, \ldots, i_{i-1}\}$.

Using the Morris' design, each point in the jth trajectory is obtained by linear and random transformation of the base point $\mathbf{x}_j^*$, thus a random matrix based scheme can be developed for generating the J trajectories [60]. For generating the jth trajectory, a matrix $\mathbf{B}_j^*$ of dimension $(n+1) \times n$ needs to be built. The ith row of $\mathbf{B}_j^*$ is the ith point $\mathbf{x}_j^{(i)}$ of the jth trajectory. For building $\mathbf{B}_j^*$, one should first choose a $(n+1) \times n$ matrix $\mathbf{B}$ with each element being either 0 or 1. The principle to build $\mathbf{B}$ is that, for each column $k = 1, 2, \ldots, n$ of $\mathbf{B}$, there are two rows differ only in their kth element. A commonly suggested matrix $\mathbf{B}$ is the strictly lower triangular matrix of 1's:

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{13}$$

Let $\mathbf{D}^*$ denotes a diagonal matrix, each element of which is randomly selected as -1 or 1 with equal probability, and $\mathbf{P}^*$ is a $n \times n$ matrix, each row of which is a random permutation of the n-dimensional vector $(1, 0, 0, \ldots, 0)$, and each row is not equal to any of the other rows. Let $\mathbf{O}_{n+1,n}$ denote a $(n+1) \times n$ matrix, each

element of which is 1. Then, given the randomly chosen base point $\mathbf{x}_j^*$, the matrix $\mathbf{B}_j^*$ is given as [1,60]:

$$\mathbf{B}_j^* = \left(\mathbf{O}_{n+1,1}\mathbf{x}_j^* + (\Delta/2)\left[(2\mathbf{B} - \mathbf{O}_{n+1,n})\mathbf{D}^* + \mathbf{O}_{n+1,n}\right]\right)\mathbf{P}^*. \tag{14}$$

The rationale behind Eq. (14) can be found in Ref. [60]. By repeating the above procedure for J times, one can generate the J trajectories.

As pointed out by Campolongo et al. [61], the Morris' design may lead to aggregation of the trajectories in the input space, i.e., insufficient exploration of the input space, especially when the input dimension is large. For avoiding this disadvantage, Campolongo et al. [61] developed an improved trajectory design method, which is denoted as optimization-based design in this article. This strategy first generates a large group of M(typically $M = 500 \sim 1000$) trajectories, and then chooses J trajectories from this group based on the principle that their dispersion is maximized in the input space. The dispersion between two trajectories k and m is measured by their distance:

$$d_{km} = \sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\sqrt{\sum_{z=1}^{n}\left(x_{k,z}^{(i)} - x_{m,z}^{(j)}\right)}, \tag{15}$$

where $x_{k,z}^{(i)}$ denotes the zth component of the ith point in the trajectory k. The dispersion D of J trajectories is measured by the square of the quadratic sum of the distance d_{km} between any pair of trajectories. This strategy, although produces the optimal exploration of the input space, needs to compute the dispersion D for a tremendous candidate combinations of J trajectories, which is often computationally demanding. A further improved version of the optimization-based design, which is more efficient but produces sub-optimal design, is developed in Ref. [62].

Other schemes include the Winding stairs design [63–66], radial design [67–69] and cell-based design [70], one can refer to the respective reference for details.

As aforementioned, the statistic σ_i cannot distinguish between the nonlinear and interaction effects, that is, one cannot know whether a high value of σ_i is caused by nonlinearities or interactions. Among the above five strategies, only the cell-based design provides us with the interaction effect, however, The current form of the cell-based design only favors number of inputs $n = r(r-1)/2$ with $r = 3, 4, 5, 6, \ldots$ [70]. It is necessary to develop effective and efficient screening strategies for separating the nonlinear effect with the interaction effect that favors any number of model inputs. Other attempts for measuring the interaction effect can be found in Refs. [71,72], in which the second order effect is considered. Many published works show [63,66–69] that, if the computational cost of model is tolerable, one can increase the number of trajectories of the above five strategies to estimate the Sobol's total effect indices, which are measures of the interaction effects. This will be discussed later in Section 8.1.2. Similar to the local method, when the increment Δ is selected to be too large, some nonlinear behavior may be missed, instead, if Δ is too small, the input space cannot be explored sufficiently and efficiently.

3.3. Derivative-based method

The derivative-based method was devised by Sobol and Kucherenko [73,74]. Two kinds of derivative-based indices, denoted as ν_i and $\tau_{\mathbf{X}_s}$, are proposed for an individual input variable X_i and a set of input variables $\mathbf{X}_s \subset \mathbf{X}$, respectively. The derivative-based indices share similar definitions with the Morris's indices, but have different explanations due to their links with the Sobol's total effect indices.

On the assumption that the input space being the n-dimensional unit hypercube H^n, the first derivative-based index ν_i for an individual input variable X_i is defined as [73]:

$$\nu_i = \int_{H^n}\left(\frac{\partial g(\mathbf{x})}{\partial x_i}\right)^2 \mathrm{d}\mathbf{x}. \tag{16}$$

The measure μ_i^* defined in Eq. (12) converges to $\int_{H^n}\left|\partial g(\mathbf{x})/\partial x_i\right|\mathrm{d}\mathbf{x}$, as shown in Ref. [73]. Thus, both μ_i^* and ν_i measure the average absolute change of model output w.r.t. the perturbation of X_i by exploring the full supports of all inputs although they are not equal.

The main contribution of Sobol and Kucherenko in Ref. [73] is the derivation of the link between ν_i and the Sobol's total effect index S_{Ti}, where S_{Ti} measures the average residual variance of model output when all the inputs except X_i are fixed over their full supports. Small value of S_{Ti} indicates X_i is non-influential. The Sobol's main and total effect indices are reviewed in detail in Section 8. The total effect index S_{Ti} is widely accepted by researchers for finding the non-influential input variables, but it generally needs a large number of model runs to compute especially when the simulation procedures are used, thus may be inapplicable to the computationally extensive models. An inequality between ν_i and S_{Ti} is derived by Sobol and Kucherenko as [73]:

$$S_{Ti} \le \frac{\nu_i}{\pi^2 V}, \tag{17}$$

where V is the total variance of model output. Eq. (17) implies that $\nu_i/\pi^2 V$ is an upper bound of S_{Ti}, and a small value of ν_i implies small S_{Ti}. Hence, the non-influential input variables can be detected by computing the values of ν_i $(i = 1, 2, \ldots, n)$.

The second derivative-based importance index $\tau_{\mathbf{X}_s}$ for a vector of inputs $\mathbf{X}_s$ is defined as [74]:

$$\tau_{\mathbf{X}_s} = \sum_{X_k \in \mathbf{X}_s}\int_{H^n}\left(\frac{\partial g(\mathbf{x})}{\partial x_k}\right)^2 \frac{1 - 3x_k + 3x_k^2}{6}\mathrm{d}\mathbf{x}. \tag{18}$$

Inequality between $\tau_{\mathbf{X}_s}$ and $S_{T\mathbf{X}_s}$ (the total effect indices for $\mathbf{X}_s$) is derived as [74]:

$$S_{T\mathbf{X}_s} \le \frac{24}{\pi^2}\frac{\tau_{\mathbf{X}_s}}{V}. \tag{19}$$

If the g-function is linear w.r.t. $\mathbf{X}_s$, it is proved that $S_{T\mathbf{X}_s} = \tau_{\mathbf{X}_s}/V$. Small value of $\tau_{\mathbf{X}_s}$ indicates small value of $S_{T\mathbf{X}_s}$. Thus $\tau_{\mathbf{X}_s}$ is helpful for identifying a group of non-influential input variables. When $\mathbf{X}_s$ contains only one input variable, say X_i, the importance index is reduced to:

$$\tau_i = \int_{H^n}\left(\frac{\partial g(\mathbf{x})}{\partial x_i}\right)^2 \frac{1 - 3x_i + 3x_i^2}{6}\mathrm{d}\mathbf{x}. \tag{20}$$

If the g-function is highly nonlinear w.r.t. X_i, the non-influential inputs can be correctly detected with small value of ν_i and τ_i, however, ranking the important input variables by the values of ν_i and τ_i may lead to false conclusions.

The Monte Carlo procedure for computing ν_i is presented in Ref. [75]. This procedure first generates a set of N samples of model input variables, and then computes the partial derivatives for each input at each sample point. The total number of g-function evaluations is $N(n+1)$. The procedure can be easily extended for computing τ_i and $\tau_{\mathbf{X}_s}$. The efficiency of the Monte Carlo procedure for computing ν_i and τ_i is investigated and compared with that for computing the total effect indices S_{Ti} in Ref. [74]. It is shown that, when the g-function is approximately linear w.r.t. X_i, to obtain the same standard deviations of estimates, the cost for computing S_{Ti} is much higher than those for computing ν_i and τ_i. However, when the g-function is highly nonlinear w.r.t. X_i, S_{Ti} is superior in the sense of computational cost.

When all the input variables follow independent normal distribution and the variance of X_i is σ_i^2, the links between the derivative-

based indices and the total effect indices are derived as [73,74]:

$$S_{Ti} \le \frac{\sigma_i^2 \nu_i}{V}, \quad \text{and} \quad S_{Ti} \le \frac{2\tau_i}{V}, \tag{21}$$

respectively. In Ref. [76], the link between ν_i and S_{Ti} is extended to cases that the model inputs follow non-uniform and non-normal distributions.

Summarily, two important conclusions are given below:

- The indices ν_i and τ_i are able to detect the non-influential input variables with relative low computational cost whether the g-function is linear or not, but they are not applicable for ranking important input variables when the g-function is highly nonlinear.
- The indices ν_i and τ_i are computationally more efficient than S_{Ti} only when the g-function is approximately linear.

3.4. Implementations and comparisons of difference-based VIMs

We use the numerical model expressed in Eq. (1) to illustrate and compare the difference-based VIMs. We assume that the LIM is normalized with the SDs of Y and X_i, as shown in Eq. (5). We then compute LIM_i by varying the nominal value of X_i from -3 to 3 and fixing the other inputs at zero. The results are shown in Fig. 6. As can be seen, the LIMs of X_4 and X_5 all keep zero w.r.t. the change of the nominal points, thus these two variables can be thought to be non-influential. It is also shown that, the LIMs of the other four variables as well as the importance ranking induced by the LIM change w.r.t. the shifts of base points, thus one can not judge the relative importance of one input with the LIM computed at one base point. For example, the LIMs of X_2 and X_3 computed at the mean point all equal to zero, does not indicating that X_2 and X_3 have no effect on Y. In fact, as pointed out by an anonymous reviewer, zero values of LIM_i could result because either (i) the changes in the input variable have no effect on the model output or (ii) the partial derivatives are computed at a local minimum or maximum. In the latter case, it is possible that the change in the input variables have large effects on the output variables. To measure the overall effect of each input variable on Y, one should use Morris' screening method or the derivative-based method.

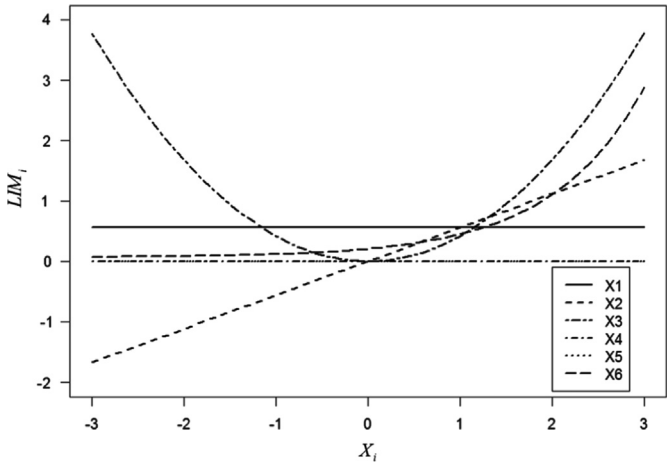

Fig. 6. Results of LIMs as a function of the base points at which the partial derivatives are computed. The results are computed for the independent case, and the base point for each input varies from -3 to 3.

The Morris' screening method is then performed with the number of trajectories J changes from 10 to 200, and the results are plotted in Fig. 7. The corresponding number of levels (p) and the total number of function evaluations (N) are labeled on each graph. As can be seen, as J is set to be larger than 50, the four influential variables (X_1,X_2,X_3 and X_6) are distinguished from the other two relatively unimportant variables. The results in Fig. 7 also reveal the type of relationship between the input and output variables. For example, small σ_1 and large μ_1^* imply that the relationship between Y and X_1 is linear, large σ_6 and μ_6^* indicate that the effect of X_6 on Y is greatly due to nonlinearity or interaction. In practical applications, the Morris' screening design is most commonly used for screening non-influential inputs. As can be seen from Fig. 7, among the six input variables, only X_4 is identified as non-influential.

The derivative-based measures ν_i and τ_i are then computed by Monte Carlo procedure with the sample size N varying from 10 to 10^4, and the results are plotted in Fig. 8, where the samples are drawn by LDS schedule. As can been seen, when the sample size N exceeds 50, the results of both ν_i and τ_i produce convincing importance ranking.

All the three difference-based methods are based on the computation of the partial derivatives of model output to the

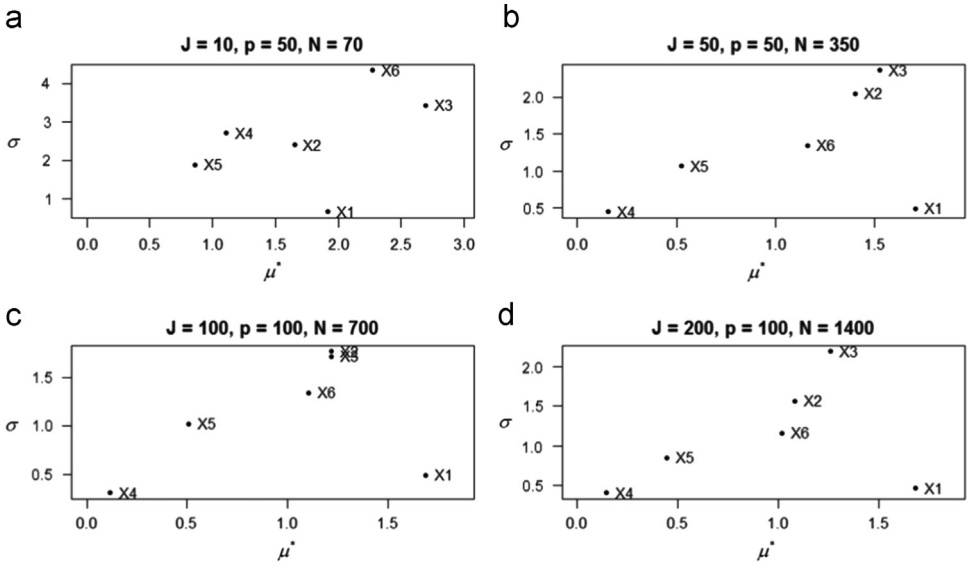

Fig. 7. Results of Morris' screening method for the independent case implemented with optimization-based design strategy, where the number of trajectories J is set to be: (a) 10, (b) 50, (c) 100 and (d) 200.

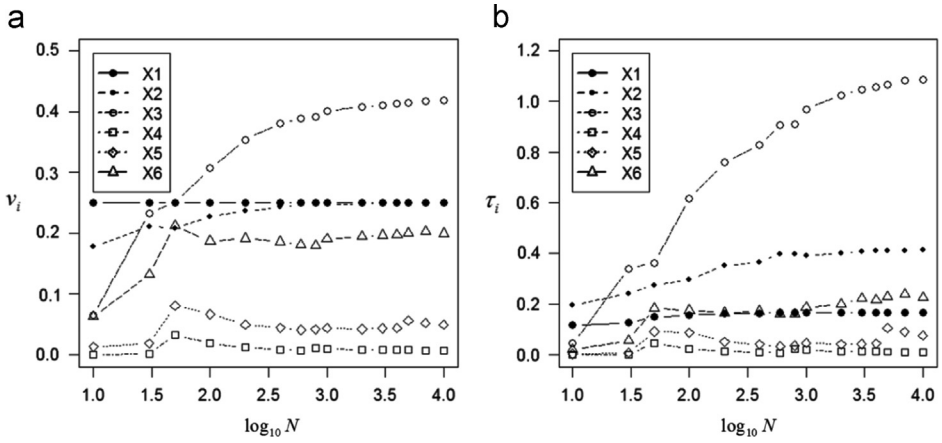

Fig. 8. Results of the derivative-based VIMs for the independent case computed with the Monte Carlo procedure, where x-labels indicates the log of the sample size N, and y-labels refer to the estimates of: (a) v_i and (b) τ_i.

input variables, thus they suffer by a major disadvantage, that is, for computational models with non-smooth response functions, all these methods are not applicable. There are also other difference-based methods for VIA such as the one based on finite difference decomposition [77]. Here we do not go into further discussion.

4. Parametric regression techniques

Different with the difference-based VIMs discussed in Section 2, the VIMs concerned in this section do not need any derivative information, but only the sample matrix $\mathbf{M}_X = (x_{ij})_{i=1,2,\ldots,N,j=1,2,\ldots,n}$ $(N > n+1)$ as well as the corresponding model output values $\mathbf{M}_Y = (y_1, y_2, \ldots, y_N)^T$ are needed. Suppose now the sample matrix $\mathbf{M}_X$ has been obtained with the LDS schedule and the vector $\mathbf{M}_Y$ have been computed by calling the model response function. We start with the parametric regression and related VIMs.

4.1. VIMs for linear dependence

The VIMs discussed in this subsection include Pearson's CC, SRC, PCC and decomposition-based measures. All of them are based on modeling the linear relationship between the output variable and one or a set of input variables.

4.1.1. Correlation coefficients (CCs)

The CC between Y and X_i is defined as:

$$r_i = r(X_i, Y) = \frac{E\{[X_i - E(X_i)][Y - E(Y)]\}}{\sqrt{V(X_i)V(Y)}}, \tag{22}$$

and can be estimated by

$$\hat{r}_i = \frac{\sum_{j=1}^{N} (x_{ji} - \bar{x}_i)\left(y_j - \bar{y}\right)}{\sqrt{\left(\sum_{j=1}^{N} (x_{ji} - x_i)^2\right)\left(\sum_{j=1}^{N} \left(y_j - \bar{y}\right)^2\right)}}, \tag{23}$$

where $\bar{x}_i = \sum_{j=1}^{N} x_{ji}/N$ and $\bar{y} = \sum_{j=1}^{n} y_j/N$ are the estimates of the expectations $E(X_i)$ and $E(Y)$, respectively.

The CC r_i takes value between -1 and 1, where $r_i = 1$ indicates that the relationship between Y and X_i are exactly and positively linear, i.e., $Y = a + bX_i$ with b being a positive value, $r_i = -1$ indicates that exactly and negatively linear relationship exists between Y and X_i, i.e., $Y = c + dX_i$ with d being a negative value, and 0 implies there is no linear relationship between Y and X_i. An absolute value of r_i between 0 and 1 indicates that part of Y linearly depends on X_i, and the larger the absolute value is, the stronger the linear dependence is. One should note that r_i only reflects the linear dependence between Y and X_i, but cannot reflect other types of dependence. For example, for $Y = X_i^2$ with X_i following standard normal distribution, the CC r_i is computed to be 0.

4.1.2. Linear regression and standardized regression coefficients (SRCs)

The multiple linear regression aims at constructing a linear model with the principle of least square error so as to fit the relationship between the model output and input variables [78,79]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + e, \tag{24}$$

where β_i is the regression coefficient of the input variable X_i, e is the prediction error and it is often assumed that $e \sim N(0, \sigma^2)$. The regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_n)^T$ as well as the variance σ^2 of e are estimated by minimizing the mean square error (MSE):

$$\sum_{j=1}^{N} \left(y_j - \hat{y}_j\right)^2 = \sum_{j=1}^{N} \left[y_j - \left(\beta_0 + \sum_{i=1}^{N} \beta_i x_{ji}\right)\right]^2, \tag{25}$$

and the estimators are derived as [78,79]:

$$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n\right)^T = \left(\overline{\mathbf{M}}_{\mathbf{X}}^T \overline{\mathbf{M}}_{\mathbf{X}}\right)^{-1} \overline{\mathbf{M}}_{\mathbf{X}}^T \mathbf{M}_Y \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{N-n-1} \sum_{i=1}^{N} \left[y_j - \left(\hat{\beta}_0 + \sum_{i=1}^{N} \hat{\beta}_i x_{ji}\right)\right]^2, \tag{26}$$

respectively, where $\overline{\mathbf{M}}_{\mathbf{X}}$ indicates a sample matrix of dimension $N \times (n+1)$ with all the elements of first column being 1 and the other n columns being $\mathbf{M}_{\mathbf{X}}$, the superscripts T and -1 refer to the matrix transposition and inversion, respectively.

The regression coefficients $\hat{\beta}_i$ $(i = 1, 2, \ldots, n)$ measure the sensitivity of Y to X_i's, but can not be used for quantifying the relative importance of each input variables as their values are influenced by the units of input variables. With the standard deviation $\hat{s} = \sqrt{\sum_{j=1}^{N} \left(y_j - \bar{y}\right)^2/(N-1)}$ of the output samples and the SDs $\hat{s}_i = \sqrt{\sum_{j=1}^{N} (x_{ji} - \bar{x}_i)^2/(N-1)}$ of input samples, the fitted regression model can be rewritten as:

$$\left(\hat{Y} - \bar{y}\right)/\hat{s} = \sum_{i=1}^{N} \left(\hat{\beta}_i \hat{s}_i/\hat{s}\right)(X_i - \bar{x}_i)/\hat{s}_i, \tag{27}$$

where $b_i = \hat{\beta}_i \hat{s}_i/s$ is called standard regression coefficient (SRC) of X_i, which, when the input variables are independent with each other, reflects the sensitivity of the standardized output $(Y - E(Y))/\sqrt{V(Y)}$ to

the standardized input variable $(X_i - E(X_i))/\sqrt{V(X_i)}$. Thus the influence of units has been moved, and $|b_i|$ $(i = 1, 2, \ldots, n)$ can now be used for measuring the relative importance of model inputs. The larger $|b_i|$ is, the more important X_i is. When some types of dependences exist among the input variables, SRC is no longer suitable for measuring the relative importance of input variables [16].

The fitted regression model is not only suitable for indicating the variable importance, but can also be used for prediction. Let $\hat{s}_{tot}^2 = (N-1)\hat{s}^2 = \sum_{j=1}^{N} \left(y_j - \bar{y}\right)^2$ indicate the total sum of squares. With the model parameters in Eq. (24) estimated with the principle of least squares, the variance $\hat{s}^2$ of model output samples can be decomposed as [16,17]:

$$\hat{s}_{\text{tot}}^2 = \hat{s}_{\text{reg}}^2 + \hat{s}_{\text{res}}^2, \tag{28}$$

where $\hat{s}_{\text{reg}}^2 = \sum_{j=1}^{N} \left(\hat{y}_j - \bar{y}\right)^2$ is the regression sum of squares, measuring the quantity of variance explained by the fitted regression model, $\hat{s}_{\text{res}}^2 = \sum_{j=1}^{N} \left(\hat{y}_j - y_j\right)^2$ is the residual sum of squares, indicating the part of sample variance not explained by the fitted model, and $\hat{y}_j$ is the prediction for the input sample $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$. Thus, from this point of view, the ratio

$$R^2 = \hat{s}_{\text{reg}}/\hat{s}_{\text{tot}}^2 = \sum_{j=1}^{N} \left(\hat{y}_j - \bar{y}\right)^2 / \sum_{j=1}^{N} \left(y_j - \bar{y}\right)^2 \tag{29}$$

can be explained as the measure of precision of the fitted regression model. Due to Eq. (28), R^2 has a value between 0 and 1. The larger R^2 is, the better the regression model match the samples. In real application, only when R^2 is sufficiently large (typically higher than 0.7), can the SRCs correctly reflect the relative importance of the input variables. When the input variables are independent, R^2 can be decomposed as [17]:

$$R^2 = R_1^2 + R_2^2 + \cdots + R_n^2, \tag{30}$$

where R_i^2 indicates the fraction of the sample variance explained by the univarite linear regression model of Y on X_i. With Eq. (30), R_i^2 is explained as the contribution of X_i to R_i^2, thus measures the relative importance of X_i.

When the input variables are independent with each other, the CC and SRC are equal, and the square of CC r_i is equal to R_i^2. Thus, both CC and SRC are applicable to linear or approximately linear computational model with independent input variables. When the model behavior is highly nonlinear or the input variables are highly dependent, both measures are not applicable [80]. For linear model with linearly correlated input variables, the decomposition-based measures are more suitable for measuring different types of effects.

When used for VIA of high-dimensional inputs, implementing the linear regression in a stepwise manner is more practical [16]. This will be further discussed in Section 5.4.

The univariate linear regression models of Y w.r.t. X_1, X_2 and X_3, respectively, in the case of independence, are shown in Fig. 9. As can be seen, the relationship between Y and X_1 is most satisfactorily represented. For (Y, X_3), the samples are only fitted well around the origin. For (Y, X_2), the relationship is not well fitted anywhere.

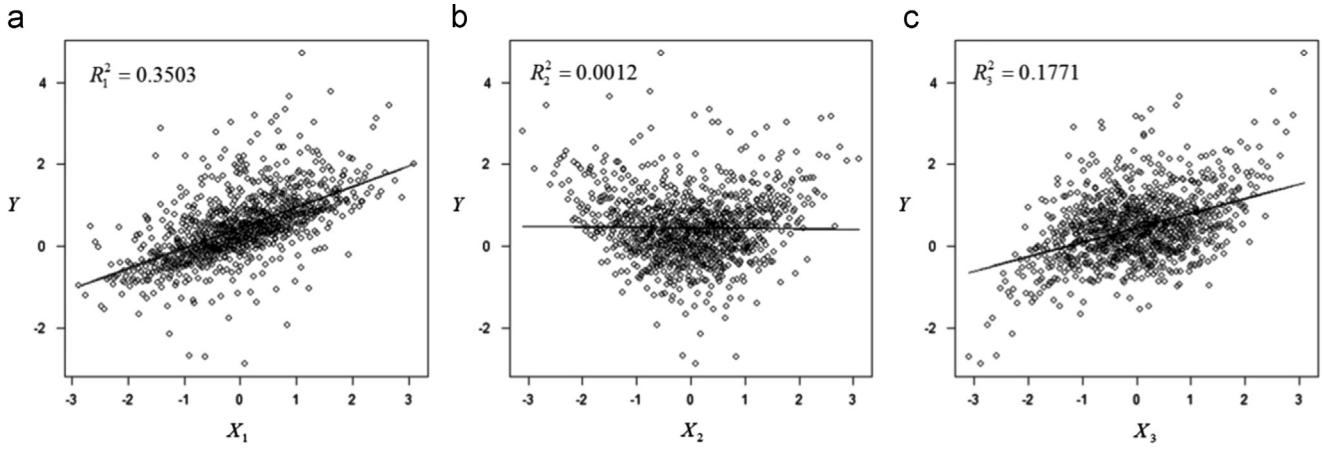

Fig. 9. The univariate linear regression models of Y on X_1, X_2 and X_3, respectively, for the case of dependence, where 1000 samples are used for performing this analysis.

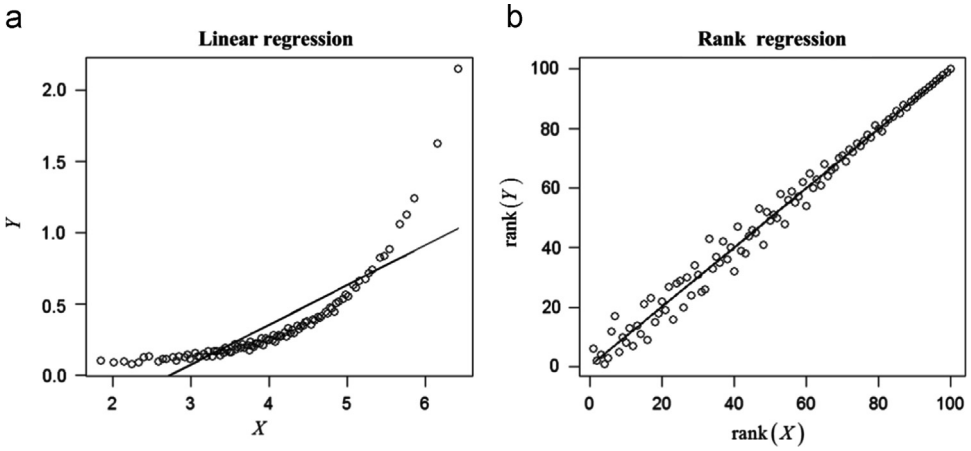

Fig. 10. An example of rank regression on monotonic relationship.

4.1.3. Partial correlation coefficient (PCC)

When the input variables are correlated, the CC r_i contains not only the effect of X_i on Y, but also the correlated effects induced by the other inputs correlated with X_i. The PCC makes a nice correction to wipe out these correlated effects. For removing the correlated effects, the following two regression models are first constructed:

$$\begin{cases} \hat{X}_{ic} = \alpha_0 + \sum_{j=1, j\neq i}^{n} \alpha_j X_j \\ \hat{Y}^{(-i)} \end{cases} = \beta_0 + \sum_{j=1, j\neq i}^{n} \beta_j X_j, \tag{31}$$

and then the PCC p_i is defined as the Pearson CC between $X_{iu} = X_i - \hat{X}_{ic}$ and $Y - \hat{Y}^{(-i)}$, where $\hat{X}_{ic}$ and X_{iu} can be explained as the correlated and uncorrelated parts of X_i, respectively. While all the inputs are linearly correlated, $\hat{X}_{ic}$ is in fact the unbiased estimation of $E(X_i|\mathbf{X}_{\sim i})$, where $\mathbf{X}_{\sim i}$ is a vector containing all the input variables except X_i. By subtracting $\hat{X}_{ic}$ from X_i, the part of X_i explained by the other input variables are all removed. Similarly, when Y is linearly dependent on all the input variables, by subtracting $\hat{Y}^{(-i)}$ from Y, the part of Y explained by $\mathbf{X}_{\sim i}$ is fully removed. Thus p_i measures the linear dependence between Y and X_i by removing the correlated effects.

4.1.4. Decomposition-based measures

When the input variables are independent, R^2 can be decomposed as (30), and R_i^2 can be explained as the contribution of X_i to R_i^2. When the input variables are correlated, R_i^2 contains not only the uncorrelated effects of X_i but also the correlated effects induced by its correlations with other input variables. The decomposed-based method aims at decomposing R_i^2 into two parts: the uncorrelated parts R_{iu}^2 representing the uncorrelated contribution of X_i to R^2 and the correlated part R_{ic}^2 indicating the correlated contribution of X_i to R^2 [81]. With Eq. (31), X_i can be decomposed as $X_i = X_{ic} + X_{iu}$, where X_{iu} refers to the uncorrelated part of X_i. The uncorrelated contribution R_{iu}^2 is then defined as the fraction of sample variance explained by the univariate linear regression model of Y on X_{iu}, and the correlated contribution R_{ic}^2 is computed by $R_{ic}^2 = R_i^2 - R_{iu}^2$. In real application, R_{iu}^2 is used as VIM. When the input variables are independent, R_{iu}^2 degrades into R^2, and R_{ic}^2 equals zero.

In some applications, the analysts may also be interested in the contribution to R^2 made by the correlation between pair of input variables. Denote the contribution made by the correlation between X_i and X_j as R_{ijc}^2. Then it can be computed with the following procedures [82,83]. Construct a regression model between Y and $\mathbf{X}_{-ij}$, and compute the fraction of sample variance R_{-ij}^2 explained by this regression model, where $\mathbf{X}_{-ij}$ denotes the input vector containing all the input variables but X_i and X_j. Similar to R_i^2, R_{-ij}^2 includes two parts: the uncorrelated contribution made by the uncorrelated part of $\mathbf{X}_{-ij}$ and the correlated contribution made by the correlations between $\mathbf{X}_{-ij}$ and (X_i, X_j). By subtracting R_{-ij}^2 from R^2, the residual $R_{ijt}^2 = R^2 - R_{-ij}^2$ contains three parts: the uncorrelated contributions R_{iu}^2 and R_{ju}^2 as well as the correlated contribution R_{ijc}^2 made by the correlation between X_i and X_j. Thus R_{ijc}^2 can be computed by $R_{ijc}^2 = R_{ijt}^2 - R_{iu}^2 - R_{ju}^2$.

Then an importance matrix can be obtained as [83]:

$$\mathbf{R} = \begin{pmatrix} R_{1u}^2 & R_{12c}^2 & \cdots & R_{1nc}^2 \\ R_{21c}^2 & R_{2u}^2 & \cdots & R_{2nc}^2 \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1c}^2 & R_{n2c}^2 & \cdots & R_{nu}^2 \end{pmatrix}, \tag{32}$$

where the diagonal contains the uncorrelated contribution of each variable, and the counter-diagonal elements R_{ijc}^2 $(i \neq j)$ are the correlated contributions between pair of input variables. It is obvious that $\boldsymbol{R}$ is symmetric. It is shown by Hao et al. [83] that $R_{ic}^2 = \sum_{k=1}^{n} R_{ikc}^2$ and

$$\sum_{k=1}^{n} \left(R_{ku}^2 + \sum_{j=k+1}^{n} R_{kjc}^2 \right) = R^2 \tag{33}$$

hold. When the input variables are independent, all the R_{kjc}^2's equal zero, and the decomposition in Eq. (33) degrades into Eq. (30).

One should note that, different with the CC, SRC and PCC which aim at measuring the dependence between the input and output variables, the importance matrix $\boldsymbol{R}$ attributes the variance explained by the linear regression model to each of the input variable and their correlations, thus tell the sources of the model output uncertainty (measured by variance). This makes the importance matrix $\boldsymbol{R}$ especially useful for reducing the model output uncertainty when the g-function is approximately linear and the input variables are linearly correlated.

4.2. Rank regression and related VIMs

The linear regression model often fails to produce satisfactory representation for highly nonlinear response functions. However, as the relationship between the output and input variables are monotonic, the rank transformation can be used for improving the performance of the linear regression model. With the rank transformation, the samples of each variable is ranked according to the magnitudes of their values, that is, the smallest sample is given rank 1; the second-smallest sample is given rank 2; and this process is repeated until the largest sample is given rank N. Then the linear regression procedure is performed based on the rank-transformed samples instead of the raw sample values. The SRC

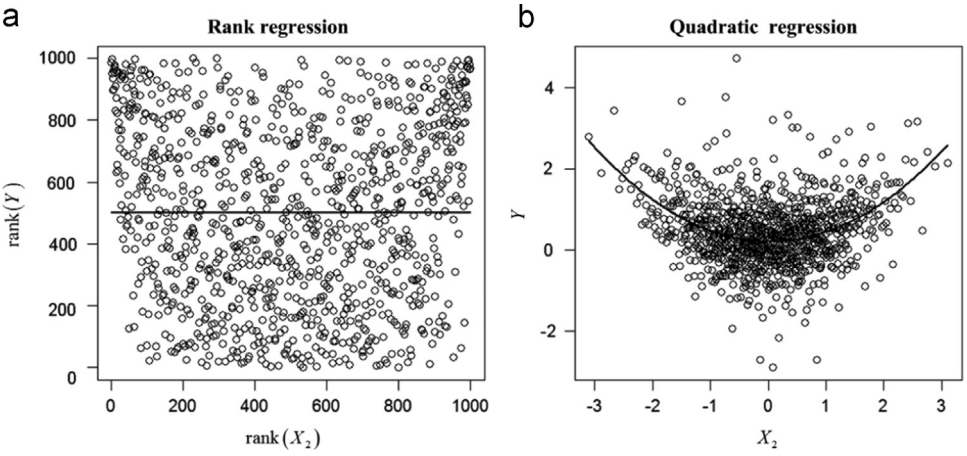

Fig. 11. Rank regression (a) and quadratic regression (b) of Y on X_2 for the case of independence.

Table 1
Results of CCs, SRCs, PCCs, RCCs, SRRCs and PRCCs for the independent case computed with 1000 sample points.

Input variables	X_1	X_2	X_3	X_4	X_5	X_6	R^2
CCs	0.5913	−0.0127	0.4198	0.0132	−0.0000	0.3494	–
SRCs	0.5954	−0.0025	0.4259	0.0194	0.0048	0.3553	0.6557
PCCs	0.7120	−0.0043	0.5872	0.0330	0.0081	0.5176	–
RCCs	0.6480	−0.0015	0.3315	0.0142	0.0069	0.3397	–
SRRCs	0.6496	0.0031	0.3341	0.0170	−0.0060	0.3421	0.6485
PRCCs	0.7384	0.0053	0.4908	0.0287	−0.0102	0.4996	–

Table 2
VIA results based on CCs, SRCs, PCCs, RCCs, SRRCs and PRCCs and sample size $N=1000$ for the dependent case of the test model.

Input variables	X_1	X_2	X_3	X_4	X_5	X_6	R^2
CCs	0.5269	0.4753	0.3834	0.0431	0.3658	0.3969	–
SRCs	0.5027	0.0021	0.3739	0.0499	0.0911	0.3061	0.5489
PCCs	0.3334	0.0042	0.5098	0.0661	0.0272	0.2370	–
RCCs	0.6154	0.5375	0.3574	0.0487	0.3521	0.3857	–
SRRCs	0.6739	−0.0658	0.3540	0.0456	0.0501	0.3376	0.6563
PRCCs	0.4618	−0.0508	0.5165	0.0774	0.0389	0.2535	–

computed with the rank-transformed data is often termed as SRRC. The Pearson's CC and PCC can then be computed for the rank-transformed samples, which are called RCC and PRCC, respectively. As the rank transformation converts the monotonic relationship into linear relationship, the RCC, SRRC and PRCC can successfully capture the monotonic dependence between the output and input variables [84].

An example of rank regression on monotonic relationship is shown in Fig. 10. As can be seen, through the rank transformation, the nonlinear and monotonic relationship has been successfully reverted into linear relationship, and with the rank regression, the samples are well fitted. However, the rank regression is not suitable for representing nonlinear and non-monotonic relationship, as shown by Fig. 11(a).

4.3. Polynomial regression

Linear regression fails to capture nonlinear relationship in analysis, as shown in Fig. 9(b), and the rank regression is only suitable for modeling monotonic nonlinear relationship, but cannot be used for capturing non-monotonic relationship. This situation can be improved by introducing higher order polynomial terms (e.g., variable squares x_j^2 and products $x_j x_k$) to a linear regression model. For example, a quadratic regression model is formulated as:

$$Y=\beta_0+\sum_{j=1}^{n}\beta_j X_j+\sum_{j=1}^{n}\sum_{k=j}^{n}\beta_{jk}X_jX_k+e, \tag{34}$$

where the regression coefficients as well as the variance of the residual e can be similarly estimated with the principle of minimizing the MSE. One can also introduce higher order polynomial terms to further improve the performance of regression model. However, with higher order polynomial terms, more samples are usually needed for estimating the regression coefficients.

As can be seen from Fig. 11(b), compared with the linear regression (Fig. 9(b)) and rank regression (Fig. 11(a)), the quadratic regression produces much more satisfactory representation to the nonlinear and non-monotonic relationship between Y on X_2.

4.4. Results and discussions of parametric techniques

We now apply the various VIMs introduced in this section to the computational model in Eq. (1), and compare their relative merits. With 1000 samples generated by the LDS schedule, the CCs, SRCs, PCCs, RCCs, SRRCs and PRCCs are computed for both the independent and dependent cases, and the results are shown in Tables 1 and 2, respectively, with the fractions of sample variance explained by the linear and rank linear regression models displayed in the last columns. The importance matrix **R** for the dependent case is computed as:

$$\mathbf{R}=\begin{bmatrix} 0.0480 & 0.2065 & -0.0001 & -0.0005 & -0.0005 & 0.0007 \\ 0.2065 & 0.0000 & -0.0006 & -0.0005 & -0.0005 & -0.0005 \\ -0.0010 & -0.0006 & 0.1398 & -0.0006 & -0.0006 & -0.0005 \\ -0.0005 & -0.0005 & -0.0006 & 0.0025 & -0.0005 & -0.0005 \\ -0.0005 & -0.0005 & -0.0005 & -0.0005 & 0.0016 & 0.1327 \\ -0.0007 & -0.0005 & -0.0008 & -0.0005 & 0.1327 & 0.0178 \end{bmatrix}. \tag{35}$$

As can be seen from Table 1, for the independent case, all these six VIMs identify X_1, X_3 and X_6 as important variables, and the other three variables are identified as non-influential variables. With the VIMs (CCs, SRCs and PCCs) based on raw data, X_3 is supposed to be obviously more important than X_6; however, the other three VIMs (RCCs, SRRCs and PRCCs) based on rank-transformed data show that X_3 and X_6 are nearly equally important. As the relationship between Y and X_3 as well as that between Y and X_6 are nonlinear and monotonic, the results based on rank-transformed data may be more convinced. From the last column of Table 1, one can see that the fractions of sample variance explained by the linear and rank linear regression models are nearly the same, indicating that these two linear models have nearly the same approximation accuracy. It is also shown that all these six VIMs fail to capture the important effect of X_2 due to the nonlinear and non-monotonic relationship between Y and X_2.

From Table 2, for the dependent case, the CCs and RCCs identify all the input variables except X_4 as influential variables, the other four VIMs, however, think only X_1, X_3 and X_6 are influential and the other three variables are non-influential. It seems that the CCs and RCCs successfully identify the important effects of X_2, however, this is not really the case. The large values of the CC and RCC for X_2 actually result from its high correlation with the influential variable X_1, and do not imply that the nonlinear and non-monotonic relationship between Y and X_2 is captured by these two VIMs. From this point of view, the CC and RCC are not suitable for the situation of variable dependence. As can be seen, PCCs and PRCCs produce the same importance ranking for the three influential variables: $X_3>X_1>X_6$, while the importance ranking obtained with SRCs and SSRCs are $X_1>X_3>X_6$ and $X_1>X_3\approx X_6$. As the SRC and SSRC do not provides reliable information of variable importance in the case of dependence [80], the importance rankings obtained by PCCs and PRCCs are more convincing. However, this does not indicate that the PCCs and PRCCs are good practices for correlated input variables as that, when the input variables are highly correlated, they do not provide meaningful indication of the uncorrelated effects of the correlated inputs.

The importance matrix conceives distinct different importance information. It aims at specifying the source of model output variation, and determining whether the model output variance results from the variations of input variables or from their correlations. The fraction of sample variance explained by the linear regression model is about 0.5489. Then the importance matrix shown in Eq. (35) explains the sources of the explained variance. As can be seen, the most part comes from the correlation between X_1 and X_2 (about 0.2065), about 0.1398 comes from the variation of X_3, and about 0.1327 comes from the correlation between X_5 and X_6. The remaining 0.0699 comes from the other sources. This indicates that, for reducing the model output

variance, the analysts should focus on the variation of X_3, the correlation between X_1 and X_2 as well as the correlation between X_5 and X_6. Compared with the PCCs and PRCCs, the diagonal elements of the importance matrix provide more meaningful indication of the uncorrelated effects of each correlated input variable.

Summarily, we draw the following two conclusions.

- When the input variables are independent, all the CCs, SRCs, and PCCs based on raw data can successfully capture the linear dependence between the output and input variables; however, for nonlinear relationship, all these three methods are not suitable. When the relationship between the output and input variables are nonlinear and monotonic, the three VIMs (RCC, SRRC and PRCC) based on rank-transformed data are applicable.
- When the input variables are correlated, only the decomposition-based measures (or importance matrix) are applicable, which not only identify the sources of model output variance explained by the linear regression model, but also provide importance ranking for each inputs by quantifying the uncorrelated effect of each variable.

Another group of linear regression based techniques for measuring the importance of correlated input variables are the relative importance analysis (RIA) techniques. These techniques can be further divided into two groups: ordering-based methods (e.g., the average squared semi-partial correlation [85–89], proportional marginal variance decomposition [90] and dominance analysis [91–93]) and transformation-based methods (e.g., Johnson's epsilon [21,94] and the omega measure recently developed by Zuber and Strimmer [95]). For reviews of these techniques see Refs. [21,22]. Here we do not discuss them in detail.

5. Nonparametric regression techniques

The parametric regression model and related VIMs have several disadvantages. First, the form of the regression model (linear or quadratic) needs to be specified, which is unavailable in most practical applications. Second, the parametric regression techniques are usually not good at approximating the local behavior of computational model [13]. Comparably, the nonparametric regression techniques directly estimate the regression function other than any regression parameters (e.g., regression coefficients in linear regression), thus provide a more flexible strategy that does not need any prior knowledge on the model behavior, and often produce better approximation to local behaviors of response function. The nonparametric regression techniques introduced in this section include (i) locally weighted regression (LOESS), (ii) generalized additive model (GAM), and (iii) projection pursuit regression (PPR). Another popular nonparametric regression technique is random forest. However, as that random forest has several popular byproducts for VIA, we arrange this technique in the next section.

5.1. Locally weighted regression (LOESS)

The LOESS aims at representing the relationship between Y and $\boldsymbol{X}$ with the information of sample points near $\boldsymbol{X}$ by polynomial regression technique. The first order approximation function is assumed to be of the form[1]

$$Y=\beta_0(\mathbf{X})+\sum_{i=1}^{n}\beta_i(\mathbf{X})X_i+\varepsilon, \tag{36}$$

where the regression coefficient $\beta_j(\mathbf{X})$ $(j=0,1,\ldots,n)$ for a given value of $\mathbf{X}$ can be estimated by minimizing the sum [96,97]

$$\sum_{j=1}^{N}\left(y_j-\beta_0-\sum_{i=1}^{n}\beta_i x_{ji}\right)^2 W\left[\frac{D(\mathbf{x},\mathbf{x}_j)}{h}\right], \tag{37}$$

where $D(\mathbf{x},\mathbf{x}_j)=\sqrt{\sum_{i=1}^{n}(z_j-z_{ji})^2}$ is the normalized Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_j=(x_{j1},x_{j2},\ldots,x_{ji})$ with $z_{ji}=(x_{ji}-\overline{x}_i)/\hat{s}_i$, h is the half-width of the neighborhood, which is usually adjusted to be the normalized distance $d_r(\mathbf{x})$ to the rth nearest neighbor (NN) of $\mathbf{x}$, and the weight function $W(\bullet)$ is expressed as:

$$W(z)=\begin{cases}[1-|z|^3]^3 & \text{if } |z|<1\\ 0 & \text{else}\end{cases}. \tag{38}$$

5.2. Generalized additive model (GAM)

The GAM technique is based on the assumption that the model response function can be decomposed as the summation of univariate function terms:

$$Y=\sum_{i=1}^{n}g_i(X_i)+\varepsilon, \tag{39}$$

where the univariate function $g_i(X_i)$ can be estimated by any univariate nonparametric regression techniques such as LOESS and smoothing spline.

The smoothing spline for univariate relationship between Y and X is formulated as follows: find the function $g(X)$ with two continuous derivatives that minimize the following penalized residual sum of squares:

$$RSS(h)=\sum_{j=1}^{N}\left[y_j-g(x_j)\right]^2+h\int_{x_{\min}}^{x_{\max}}|g''(x)|^2\mathrm{d}x, \tag{40}$$

where h is a smoothing parameter determining the smoothness of the spline, which can be specified by cross-validation [98]. The first term in Eq. (40) is the sum of square errors, promising the fitness of the estimation to the training data, and the second term is a roughness penalty, promising the smoothness of the estimation $\hat{g}(X)$.

The unique solution $\hat{g}(x)$ that minimizes $RSS(h)$ is a natural cubic polynomial spline with knots at each sample point of X, where the cubic spline is a piece-wise continuous polynomial function with both the first and second derivatives being continuous at the knots.

The univariate estimates of the relationships between Y and X_i $(i=1,2,3)$ by LOESS and smoothing spline in the case of independence are compared in Fig. 12. As can be seen, these two regression techniques produce similar results.

5.3. Projection pursuit (PP)

PP regression involves first transforming the input spaces linearly to a lower dimensional space, and then performing additive regression on this lower dimensional space. The approximation is assumed to be of the form [99]

$$Y=\sum_{s=1}^{nL}g_s(\boldsymbol{\alpha}_s\mathbf{X})+\varepsilon, \tag{41}$$

where $\boldsymbol{\alpha}_s=(\alpha_{s1},\alpha_{s2},\ldots,\alpha_{s,n})$, and for $s\neq t$, $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_t$ are orthogonal,$\mathbf{X}=(X_1,X_2,\ldots,X_n)^T$, and $g_s(\bullet)$ is an arbitrary univariate function. The nL (usually smaller than n) linear combinations $\boldsymbol{\alpha}_s\mathbf{X}$ $(s=1,2,\ldots,nL)$ form a set of new bases in the input space.

The estimations for g_s, $\boldsymbol{\alpha}_s$ and nL are determined in a recursive procedure as follows. First estimate $\boldsymbol{\alpha}_1$ and g_1 by minimizing the

[1] The pth order approximation function is expressed as:$Y=\beta_0(\mathbf{X})+\sum_{j_1=1}^{n}\beta_{j_1}(\mathbf{X})X_{j_1}+\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\beta_{j_1j_2}(\mathbf{X})X_{j_1}X_{j_2}+\cdots+\sum_{j_1=1},\ldots,\sum_{j_p=1}^{n}\beta_{j_1,\ldots j_p}(\mathbf{X})X_{j_1}X_{j_2},\ldots,X_{j_p}+\varepsilon$.

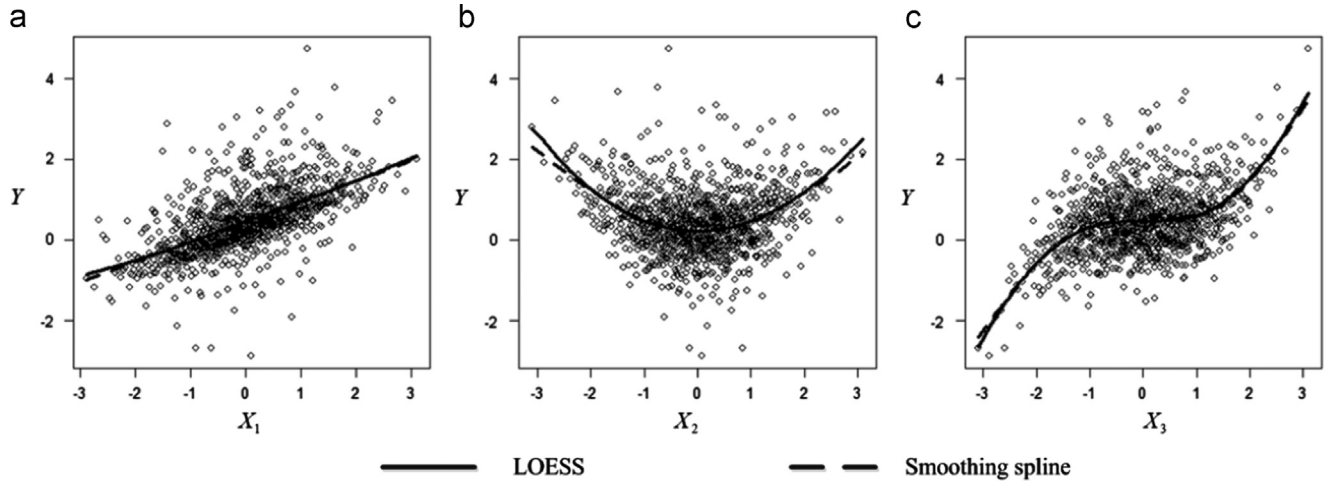

Fig. 12. Univariate LOESS and smoothing spline estimates in the case of variable independence.

Table 3
Comparison of importance rankings in the case of independence obtained with parametric regression (i.e., LIN_REG, RANK_REG and QUAD_REG) and nonparametric regression (i.e., LOESS, GAM and PP_REG).

LIN_REG		RANK_REG		QUAD_REG		LOESS		GAM		PP_REG	
Order	R^2	Order	R^2	Order	R^2	Order	R^2	Order	R^2	Order	R^2
X_1	0.3503	X_1	0.4205	X_1	0.3504	X_1	0.3507	X_1	0.3507	X_1	0.3510
X_3	0.5292	X_6	0.5367	X_3	0.5293	X_3	0.6059	X_3	0.6175	X_3	0.6211
X_6	0.6553	X_3	0.6567	X_2	0.6920	X_2	0.7537	X_2	0.7793	X_2	0.7793
				X_6	0.8396	X_6	0.8770	X_6	0.9249	X_6	0.9349
										X_5	0.9695

sum:

$$\sum_{j=1}^{N}\left[y_j - g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}\mathbf{x}_j)\right]^2, \tag{42}$$

where $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\ldots,\alpha_n)$ with $\|\boldsymbol{\alpha}\|=1$, and for each realization of $\boldsymbol{\alpha}$, $g_{\boldsymbol{\alpha}}$ is estimated by nonparametric univariate regression techniques (e.g., LOESS and smoothing spline). With the estimations $\hat{\boldsymbol{\alpha}}_1$ and $\hat{g}_1$, $\boldsymbol{\alpha}_2$ and g_2 can then be estimated by minimizing the sum:

$$\sum_{j=1}^{N}\left[y_j - \hat{g}_1(\hat{\boldsymbol{\alpha}}_1\mathbf{x}_j) - g_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}\mathbf{x}_j)\right]^2, \tag{43}$$

where $\|\boldsymbol{\alpha}\|=1$, $\hat{\boldsymbol{\alpha}}_1\boldsymbol{\alpha}=0$, and $g_{\boldsymbol{\alpha}}$ is also estimated by univariate regression techniques. The above process is repeated until no obvious reduction on the MSE can be achieved.

5.4. Implementations of the nonparametric regression techniques

In real applications, the input vector can be very high-dimensional, resulting in poor performance or acquirement of increasing number of training sample points when the nonparametric regression techniques are directly applied. For avoiding this shortcoming, the nonparametric regression procedures can be performed in a stepwise manner [17]. In the first step, perform the regression technique on each of the input variables, and determine the one, denoted as X_{i_1}, that results in the univariate regression model with the most explained sample variance (denoted as $R^2_{i_1}$). Then X_{i_1} can be thought as the most influential variable. In the second step, perform a regression procedure on X_{i_1} and each of the $n-1$ remaining variables, and choose the one, denoted as X_{i_2}, that results in a bivariate regression model with the most explained sample variance (denoted as $R^2_{i_1 i_2}$). Then X_{i_2} is thought to be the second most important variable. This procedure is repeated until some stopping criterion is reached. The order that each input variable enters the regression model and the increments in the explained variance when each variable is added to the regression model can be used to measure the relative importance of input variables. The increment of the explained sample variance or the F-statistic with appropriate degrees of freedom (see Ref. [17] for details) can be used for determining the stopping criterion. If the explained sample variance does not increase obviously by adding any one of remaining variables, the stepwise variable selection procedure stops.

The orders of selected variables as well as the cumulative explained variances R^2 obtained with LOESS, GAM and PP regression (PP_REG) are reported in Table 3, together with the results of linear regression (LIN_REG), linear rank regression (RANK_REG) and quadratic regression (QUAD_REG). As can be seen, both LIN_REG and RANK_REG correctly identify the three variables X_1, X_3 and X_6 as the influential variables, but fail to capture the effect of X_2. Both models explain about 65 percent of sample variance, thus do not capture the full model behavior. Comparably, QUAD_REG explains about 84 percent of the sample variance, thus approximates the model function more accurately. This indicates that both LIN_REG and RANK_REG do not capture the non-monotonic relationship between Y and X_2, but QUAD_REG does. It is shown that the LOESS explains about 87.7 percent of the sample variance, thus provides a more accurate approximation to the computational model than QUAD_REG. The univariate terms of GAM are estimated by smoothing spline. As can be seen, it has higher approximation accuracy than LOESS. PP_REG explains about 97 percent of the sample variance, thus provides the most accurate approximation to the computational model. It is shown that the stepwise procedures with QUAD_REG, GAM, LOESS and

PP_REG produce the same importance ranking for the four most important variables, i.e., $X_1 > X_3 > X_2 > X_6$.

Other nonparametric regression techniques include neural network [100], support vector regression [101,102], polynomial chaos expansion regression [103], state dependent regression (SDR) [104,105] and so on. Detailed descriptions of these techniques are beyond the scope of this article. One can refer to the respective reference for detail. For additional discussion and illustration of the use of nonparametric regression procedures in VIA, one can refer to Section 6.8 of Ref. [16] and Refs. [17,18].

6. Random forest

6.1. Brief introduction to random forest

The random forest is a machine learning algorithm for regression and classification based on a set of training data $\mathbf{M} = (\mathbf{M}_{\mathbf{X}}, \mathbf{M}_Y)$. When used for regression, it belongs to the nonparametric regression technique. A random forest regression model consists of a collection of *ntree* regression trees $\{h(\mathbf{X}, \boldsymbol{\Theta}_k), k = 1, 2, \ldots, ntree\}$, where $\{\boldsymbol{\Theta}_k\}$ are *ntree* sets of independent identically distributed random samples, each tree provides a prediction at input variables $\boldsymbol{X}$ and the prediction of the random forest at $\boldsymbol{X}$ takes the mean value of the *ntree* predictions. In the milestone article of random forest [106], Breiman suggested using the Classification And Regression Trees (CART) algorithm developed in Ref. [107] for growing each individual tree. Other algorithms for growing individual tree includes THAID [108], C4.5 [109,110] and Conditional Inference Trees (CIT) [111]. Here, the random forest based on CART and CIT are described since the VIMs we later review are based on them. Readers with interest on this and related topics can refer to Refs. [23,24] for details. We denote the random forest constructed with CART and CIT algorithms as CART-RF and CIT-RF, respectively.

A random forest is constructed by first drawing *ntree* subsamples $\{\boldsymbol{\Theta}_k\}$ $(k = 1, 2, \ldots, ntree)$ from the training data set $\mathbf{M} = (\mathbf{M}_{\mathbf{X}}, \mathbf{M}_Y)$ by bootstrap with or without replacement, and then establishing the individual tree $h(\mathbf{X}, \boldsymbol{\Theta}_k)$ from the subsamples $\{\boldsymbol{\Theta}_k\}$ by CART or CIT algorithm. When the subsamples are generated with replacement, about 63.2 percent of data will be included in each subsample $\boldsymbol{\Theta}_k$. Thus it is usually suggested, when bootstrap without replacement is conducted, sample size of each subsample is set to be 0.632 times the sample size N of the training data set $\mathbf{M}$. Then for each subsample $\boldsymbol{\Theta}_k$, a set of out-of-bag (OOB) data is obtained by $\mathbf{B}_k = \mathbf{M} - \boldsymbol{\Theta}_k$. The OOB data $\mathbf{B}_k$ is then used for measuring the prediction error of the kth decision tree and for measuring the importance of each input, as will be shown in the next subsection.

Both CART and CIT algorithms grow the individual decision tree by recursively partitioning the subsamples from the root node down to the terminal nodes so that the subsamples are divided into more and more homogeneous parts. In CART-RF, the bootstrap subsamples are generated with replacement. At each node, *ntry* ($ntry \ll n$) inputs are selected randomly from the n inputs to form the candidate splitting variables, and then the splitting variable and the cutpoint are specified based on the principle of maximizing the reduction of node impurity, where the impurity of a node is measured by the Gini impurity index [106,107]. Each node splits downward recursively until some stop criteria are reached. A simple stop criterion is when the impurity of the current node is lower than a predetermined threshold value. Based on the above procedure, all the *ntree* trees are grown sufficiently without pruning. From the complete forest formed by these *ntree* trees, the model output value corresponding to any input value can be predicted as an average vote or value of the predictions of all trees. The random feature of CART-RF is shown in two aspects. First, the bootstrap subsamples are randomly generated. Second, the *ntry* candidate splitting variables are randomly chosen. These two random features, as shown by Breiman [106], sufficiently improve the accuracy of prediction.

As shown by many studies [112–114], the CART algorithm faces the problem of biased variable selection when the model includes different types of inputs, or categorical inputs with different numbers of status, or inputs with many missing values and/or correlated inputs. Strobl et al. [114] also showed that, another source of bias, which is induced by the bootstrap sampling with replacement, exists in CART-RF. For avoiding these two sources of bias, Strobl et al. [114] suggested using CIT algorithm developed by Hothorn et al. [111] to grow each individual tree and sampling the bootstrap subsamples without replacement. CIT is different from CART in three aspects: the selection of splitting variables, the splitting criterion and the stop criterion. At each node, the CIT selects the splitting variable by testing the dependence between each input and output variables. The partial null hypothesis of X_i is $H_0^i : F(Y|X_i) = F(Y)$, and the global null hypothesis is $H_0 = \cap_{i=1}^{n} H_0^i$, where $F(\bullet)$ is the distribution function. At current node, if the null hypothesis is not rejected, it indicates that this node covers no information of model output, this node is then set to be a leaf node; however, if the global null hypothesis is rejected, the current node should split, and the input variable with the strongest dependence with the output variable is selected as the splitting variable, where the dependence is measured by the p-value of the partial hypothesis test.

The decision trees grown by CART are usually quite large, and need to be pruned when individual tree is used for prediction so as to avoid overfitting. The decision tree grown by CIF is usually much smaller than that grown by CART, and need not to be pruned. When used for growing random forest, both algorithms do not need to prune the branches. Practical applications of CART-RF and CIT-RF involve proper identification of the parameters such as the number of candidate splitting inputs *ntry* and the number of trees. Since this review only concerns the VIMs derived from random forest instead of the process for developing random forest, we do not go further into these contents. The readers interested in these details can refer to Refs. [23–25,106]. The R packages "randomForest" [115] developed by Liaw A and Wiener, and "party" developed [116] by Hothor et al. are available for implementing CART-RF and CIT-RF, respectively. The Matlab package randomforest-matlab developed by Jaiantilal for implementing CART-RF is available in Ref. [117].

6.2. Random forest based VIMs

The popular VIMs based on random forest include the Gini VIM (GVIM) [106–118], permutation VIM (PVIM) [106,107] and conditional permutation VIM (CPVIM) [119]. Both GVIM and PVIM are proposed along with CART-RF, and can be implemented using the "randomForest" package. PVIM can also be derived from CIT-RF, and computed with the "party" package. CPVIM is based on CIT-RF, and can be implemented using the "party" package.

6.2.1. Gini VIM

At each father node, the choice of the splitting variable from a set of candidate splitting variables as well as the splitting criteria are based on maximizing the decrease of the impurity index of this father node. Suppose now the random forest has been grown. If the model output is categorical variable, let $P_F(\omega_i)$ denote the frequency of data dropping into the category ω_i in the father node. Then the Gini impurity index of this father node is defined as [107,118]:

$$GI_F = \sum_{i \neq j} P_F(\omega_i) P_F(\omega_j) = 1 - \sum_i P_F^2(\omega_i). \tag{44}$$

For continuous output, the impurity index is defined as the MSE of the output values in the father node. Let GI_L and GI_R denote

the impurity indices of the left and right child nodes, and p_l and p_r stand for the fractions of data sent to the left and right child nodes, respectively. Then the decrease of impurity in splitting this father node is computed by [107,118]:

$$\Delta GI = GI_F - p_r GI_R - p_l GI_L. \tag{45}$$

The GVIM index $GVIM_i^{(k)}$ of X_i in the kth tree is defined as the sum of decreases of impurity indices of the nodes whose splitting variable is X_i, and the overall GVIM of X_i, denoted as $GVIM_i$, is then defined by summing or averaging $GVIM_i^{(k)}$ across all the $ntree$ trees.

The interpretation of GVIM is straightforward. At any father node, the input, which is chosen as the splitting variable, leads to the most decrease of node impurity, thus can be thought as the most influential input variable among these candidate splitting variables. Summing all the impurity decreases resulting from X_i across each tree in the forest provides an overall measure of the contribution of X_i to the accuracy of model prediction.

6.2.2. Permutation VIM

The PVIM is defined with the OOB data. Let $\mathbf{B}_k = \left\{\left(y_j^{(k)}, \mathbf{x}_j^{(k)}\right)\right\}$ ($k = 1, 2, \ldots, ntree$ and $j = 1, 2, \ldots, noob$) stand for the OOB data of the kth tree, where $noob$ is the number of sample points in $\mathbf{B}_k$. The rationale behind PVIM for continuous output is given as follows. For the kth tree, the MSE of the OOB data $\mathbf{B}_k$ before and after randomly permuting the values of X_i is computed as:

$$MSE_k = \frac{1}{noob}\sum_{j=1}^{noob}\left(y_j^{(k)} - \hat{y}_j^{(k)}\right)^2 \quad \text{and} \quad MSE_{k,i} = \frac{1}{noob}\sum_{j=1}^{noob}\left(y_j^{(k)} - \hat{y}_{j,i}^{(k)}\right)^2, \tag{46}$$

respectively, where $\hat{y}_j^{(k)}$ and $\hat{y}_{j,i}^{(k)}$ are the model output values of the OOB data predicted by the kth tree before and after randomly permuting the values of X_i, respectively. Then the PVIM for X_i in the kth tree is defined as $PVIM_i^{(k)} = MSE_{k,i} - MSE_k$ [106,107], and the overall PVIM of X_i, denoted as $PVIM_i$, is defined by averaging $PIVM_i^{(k)}$ across all trees, i.e., $PVIM_i = \sum_{k=1}^{ntree} PVIM_i^{(k)}/ntree$. For categorical output, $PVIM_i$ is defined as the average difference between the error rates of the OOB data after and before permuting values of X_i.

In the kth tree, if X_i is not chosen as the splitting variable of any node, then in Eq. (46), $\hat{y}_j^{(k)} = \hat{y}_{j,i}^{(k)}$ holds for all j and further $PVIM_i^{(k)} = 0$ hold. This property of PVIM is consistent with GVIM. $PVIM_i$ measures the average difference between the MSEs of the OOB data computed after and before permuting the value of X_i. It can also be interpreted as the contribution of X_i to the model prediction accuracy with the consideration of its interaction effects with other inputs. Permuting an importance input variable usually intends to increase the prediction error of OOB data, leading to higher value of PVIM.

6.2.3. Conditional permutation VIM

An alternative revised version of PVIM is CPVIM. Strobl et al. [119] noticed that, the PVIMs contain the correlated contributions when the input variables are correlated, thus tend to overestimate the importance of correlated inputs. This phenomenon, as explained by Strobl et al. [119], on the one hand, is caused by the preference of correlated inputs as the splitting variables at the early stage (in fact, this is only true for correlated inputs that are associated with output, as shown by Nicodemus et al. [120]), on the other hand, results from the fact the global permutation of X_i breaks not only the association of X_i with the output, but also the correlations of X_i with the other inputs. In many applications such as disease studies [119], the practitioners may be also interested in measuring the marginal importance of each input, regardless of the portion of contributions due to the correlations between inputs. For this purpose, Strobl et al. [119] developed the CPVIM.

CPVIM is different with PVIM in two aspects: the algorithm for establishing the random forest and the permutation scheme. Strobl et al. [119] recommended using the CIT algorithm to grow each individual tree so as to avoid the first source of correlated effects, and permuting X_i with a conditional permutation scheme so that the association of X_i with Y is broken, but the correlations of X_i with other inputs can be kept. The conditional permutation scheme was suggested by Strobl et al. as follows: for an individual tree, use all the cutpoints of those inputs $\boldsymbol{Z}$ highly correlated with X_i to form a set of grids, and then permute X_i in each grid. The CPVIM of X_i for this tree can be computed as the prediction accuracy loss before and after permutation, as in Eq. (46), and the overall CPVIM of X_i is computed by averaging the CPVIMs over all trees in the forest. An important issue in CPVIM is the identification of inputs $\boldsymbol{Z}$ correlated with X_i. If the input variables are all continuous, Strobl et al. suggested using the Pearson correlation. If the model involves multiple types of inputs, the p-value of conditional inference test, which is used in CIT algorithm for choosing the splitting variables, is suggested.

6.3. Comparisons and implementations of random forest based VIMs

The GVIM and PVIM have been compared by several researchers using both simulated and real data [121–124]. It was concluded that, the GVIMs, derived from CART-RF, tend to be biased in many scenarios. This disadvantage is mainly due to the variable selection bias, that is, the Gini splitting criterion prefers those inputs with more categories in selecting the splitting variables, leading to higher GVIMs of those inputs. The variable selection bias in CART-RF, as shown by Boulesteix et al. [24], would not transfer into PVIM. It is usually believed that PVIM is superior to GVIM. However, there are also cases where GVIM is preferred. If the input variables are continuous and mutually uncorrelated, the variable selection bias will not emerge, and the GVIM does not tend to be biased. Under this premise, while the output is categorical variable with strongly unbalanced categories, GVIM is expected to produce better results than PVIM [24,123]. It is also found that, computing PVIM based on CIT-RF without replacement usually leads to more robust results than using CART-RF [24].

Comparison between PVIM and CPVIM has also been conducted by several articles [119,120]. Both PVIM and CPVIM can be produced by CIT-RF. PVIM includes both the uncorrelated effect (reflects the structural importance of each input) and correlated effect (results from the correlations with other inputs). In CPVIM, the second part, i.e., the correlated effect, is greatly weakened, but not completely removed [120]. In practical applications, whether to choose PVIM or CPVIM depends on what information the analysts want to extract. As concluded by Nicodemus et al. [120], CPVIM would be preferable while the analysts' purpose is to identify a set of truly influential inputs without considering the correlated effects, in other cases, the correlations are inherent mutual property of inputs, and the analysts want to screen all the influential inputs without eliminating the correlated effects, PVIM is more suitable.

The advantages of random forest based VIMs are as follows. First, they can incorporate all types of model inputs such as categorical and continuous inputs. Second, these methods can be applied to "small N large n" problems, that is, they can screen important inputs from a large amount of inputs (e.g., several thousands) with relatively small set of samples (usually, $N < n$). The excellent performance of these VIMs based on random forest for "small N large n" results from the fact that, usually a small group of model inputs cover the most information of model output no matter how large the number of model inputs is, and the

Table 4
The results of GVIM (computed by CART-RF), PVIM (computed by CART-RF and CIT-RF) and CPVIM (computed by CIT-RF) for both independent and dependent cases, where the superscripts indicate the importance rankings.

Variables	X_1	X_2	X_3	X_4	X_5	X_6	R^2
Independent case, $N=1000$							
GVIM$^{(CART\text{-}RF)}$	$267.19^{(2)}$	$103.61^{(4)}$	$322.06^{(1)}$	$16.786^{(6)}$	$39.005^{(5)}$	$144.82^{(3)}$	0.7965
PVIM$^{(CART\text{-}RF)}$	$0.4432^{(1)}$	$0.1443^{(4)}$	$0.3776^{(2)}$	$0.0005^{(6)}$	$0.0161^{(5)}$	$0.2166^{(3)}$	
PVIM$^{(CIT\text{-}RF)}$	$0.4544^{(1)}$	$0.0290^{(4)}$	$0.3808^{(2)}$	$-0.0004^{(6)}$	$0.0181^{(5)}$	$0.2259^{(3)}$	0.7304
Dependent case, $N=1000$							
GVIM$^{(CART\text{-}RF)}$	$233.72^{(1)}$	$159.07^{(3)}$	$171.57^{(2)}$	$11.88^{(6)}$	$179.29^{(5)}$	$157.50^{(4)}$	0.9115
PVIM$^{(CART\text{-}RF)}$	$0.3705^{(1)}$	$0.1940^{(5)}$	$0.2716^{(2)}$	$0.0033^{(6)}$	$0.2268^{(4)}$	$0.2391^{(3)}$	
PVIM$^{(CIT\text{-}RF)}$	$0.3802^{(1)}$	$0.1272^{(4)}$	$0.2598^{(3)}$	$0.0011^{(6)}$	$0.0790^{(5)}$	$0.3379^{(2)}$	0.8350
CPVIM$^{(CIT\text{-}RF)}$	$0.0180^{(1)}$	$0.0069^{(4)}$	$0.0884^{(2)}$	$0.0001^{(6)}$	$0.0048^{(5)}$	$0.0078^{(3)}$	

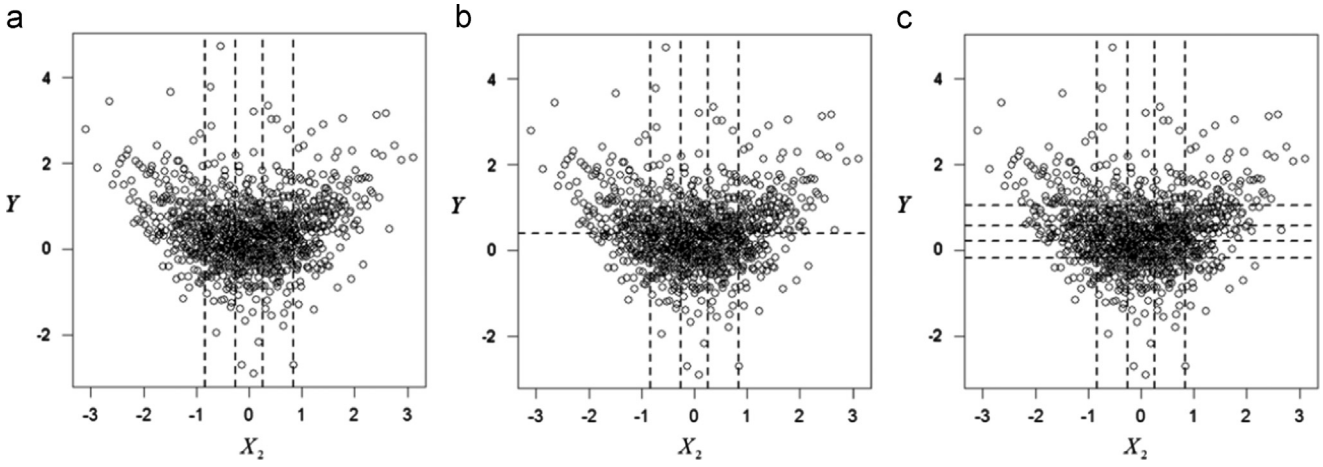

Fig. 13. Grids for hypothesis tests: (a) division of the range of X_2 for CMNs and CLs tests, (b) division of the ranges of X_2 and Y for CMDs, and (c) division of the ranges of X_i and Y for SI test and entropy-based VIMs.

variable selection procedure always prefers this small group of influential inputs as the splitting variables. Due to these advantages, the random forest based VIMs have been regarded as standard techniques in bioinformatics for extracting information from microarray data [125,126].

The disadvantages of random forest based VIMs are twofold. First, these measures may fail to identify the important variables when the input dimension is high and most of the input variables are influential. Second, for high-dimensional data, the ability of random forest to capture the interaction effects may decline [127].

The results of the three random forest based VIMs for both the independent and dependent cases are computed with 1000 sample points, and the results are listed in Table 4, where the PVIM are computed by both CART-RF and CIT-RF. As can be seen, for the independent case, the importance ranking induced by GVIM and PVIM computed with CART-RF and CIT-RF are nearly the same except that of the two most important variables. For the dependence case, the importance of the correlated variables (e.g., X_5) are enhanced due to the correlated effects. With the CPVIM computed by CIT-RF, the correlated effects are weakened, and the induced importance ranking is the same as that obtained by PVIM in the independent case.

7. Hypothesis tests and related VIMs

The hypothesis test based VIMs aim at testing the dependence between each input and output variables. These techniques can be divided into two groups based on whether the sample space needs to be divided into grids.

7.1. Grid-based hypothesis tests

This group of hypothesis test techniques is based on splitting the sample space of (Y,X_i) into grids and then testing whether the patterns of sample distributions across different grid cells are random. If the null hypothesis (patterns are random) is rejected, then X_i is believed to be influential. The commonly used test techniques in this group includes (i) common means (CMNs) test, (ii) common distributions or locations (CLs) test, (iii) common medians (CMDs) test, (iv) statistical independence (SI) test, and (v) entropy-based VIMs.

7.1.1. Common means (CMNs) test

The CMNs test involves first dividing the samples of X_i into nX classes, and then testing whether the samples mean in the nX classes are the same. With this division, the region of X_i is divided into nX mutually exclusive and exhaustive subintervals, and each subinterval contains equal number of sample points. Let $\mathbf{X}_d$ $(d=1,2,\ldots,nX)$ denote the sample set of X_i contained in the dth subinterval, and nX_d indicate the number of sample points contained in $\mathbf{X}_d$. This type of space division is schematically illustrated in Fig. 13(a), in which the 1000 sample points of X_2 is divided into five groups.

The statistics used for CMNs test is [15,16]

$$F=\frac{\left(\sum_{d=1}^{nX} nX_d\bar{y}_d^2-N\bar{y}^2\right)/(nX-1)}{\left(\sum_{j=1}^{N} y_j^2-\sum_{d=1}^{nX} nX_d\bar{y}_d^2\right)/(N-nX)}, \tag{47}$$

are normally distributed with equal mean values, then the statistic F in Eq. (47) follows F-distribution with freedom of $(nX-1, N-nX)$. With the assumption of normal distribution holding, the probability $prob\left(F>\hat{F}\,|\,F\sim F(nX-1,N-nX)\right)$ that the random variable F with distribution $F(nX-1,N-nX)$ exceeds the estimate $\hat{F}$ can be computed and served as VIM. A low p-value implies that X_i has obvious effect on Y.

7.1.2. Common distributions or locations (CLs) test

The division of scatterplot for CLs test is the same as that in CMNs test, as shown in Fig. 13(a). The statistic used is the Kruskal–Wallis test statistic T expressed as follows [15,16]:

$$T=\left[\sum_{d=1}^{nX}\left(R_d^2/nX_d\right)-N(N+1)^2/4\right]/s^2, \tag{48}$$

where

$$R_d=\sum_{x_{ji}\in\mathbf{X}_d} r\left(y_j\right),\quad s^2=\left[\sum_{j=1}^{N} r\left(y_j\right)^2-N(N+1)^2/4\right]/(N-1), \tag{49}$$

and $r\left(y_j\right)$ refers to the rank of y_j. If the assumption that the samples of Y in each class follows the same distribution holds, the statistic T in Eq. (48) approximately follows χ^2 distribution with $(nX-1)$ degree of freedom. Then the probability $prob(T>\hat{T}\,|\,T\sim\chi^2(nX-1))$ that the random statistic T exceeds the estimate $\hat{T}$ can be computed and used as VIM. The lower the p-value is, the more influential X_i is.

7.1.3. Common medians (CMDs) test

For CMDs test, based on the division for CMNs, the region of Y needs to be divided into two parts by the line $y=y_{0.5}$, as shown in Fig. 13(b), where $y_{0.5}$ is the median of Y computed from all the N sample points, i.e.,

$$y_{0.5}=\begin{cases} y_{(\lfloor 0.5N\rfloor+1)} & \text{if } N \text{ is an odd number}\\ (y_{(0.5N)}+y_{(0.5N+1)})/2 & \text{else}\end{cases}, \tag{50}$$

where $y_{(j)}$ indicates the ordering of the sample values of y such that $y_{(i)}\le y_{(i+1)}$ and $\lfloor 0.5N\rfloor$ denotes the greatest integer no larger than $0.5N$. With this division, the sample space of (Y,X_i) is divided into $2nX$ cells. Let nX_{rd} denote the number of sample points contained in cell (r,d) with $r=1$ referring to the cell above $y=y_{0.5}$ in $\mathbf{X}_d$, and $r=2$ indicating the cells below $y=y_{0.5}$.

The statistic for CMDs test is defined as follows [15,16]

$$T=\sum_{d=1}^{nX}\sum_{r=1}^{2}(nX_{rd}-nE_{rd})^2/nE_{rd}, \tag{51}$$

where

$$nE_{rd}=\left(\sum_{p=1}^{2} nX_{pd}/N\right)\left(\sum_{q=1}^{nX} nX_{rq}/N\right)/N=\left(\sum_{p=1}^{2} nX_{pd}\right)\left(\sum_{q=1}^{nX} nX_{rq}\right)/N \tag{52}$$

refers to the expected number of sample points in cell (r,d). If the assumption that each individual classes have the same median holds, the statistic T in Eq. (51) approximately follows distribution of $\chi^2(nX-1)$. Then the p-value $prob\left(T>\hat{T}\,|\,T\sim\chi^2(nX-1)\right)$ that the random statistic T exceeds the estimate $\hat{T}$ quantifies the effect of X_i on the behavior of Y. The lower the p-value is, the more important X_i is.

7.1.4. Statistical independence (SI) test

For SI test, the range of X_i is divided in the same manner as that for CMNs and CLs tests, and the range of Y is divided into nY subintervals in an analogous way as that used for X_i. Let $\mathbf{Y}_r$ indicate the sample set of Y in the rth subinterval, and nY_r $(r=1,2,\ldots,nY)$ denote the number of samples contained in $\mathbf{Y}_r$. The divisions of the ranges of Y and X_i are illustrated in Fig. 13(c), where both ranges are divided into five subintervals with the same number of sample points. This type of partition results in $nX\times nY$ cells. Let $\mathbf{O}_{rd}$ indicate the set of sample points contained in cell (r,d), and $\left(x_{ji},y_j\right)\in\mathbf{O}_{rd}$ if and only if $x_{ji}\in\mathbf{X}_d$ and $y_j\in\mathbf{Y}_r$. Denote the total number of elements contained in $\mathbf{O}_{rd}$ as nO_{rd}.

The statistic for SI test is then defined as follows [15,16]:

$$T=\sum_{d=1}^{nX}\sum_{r=1}^{nY}(nO_{rd}-nE_{rd})^2/nE_{rd}, \tag{53}$$

where $nE_{rd}=(nY_r/N)(nX_c/N)N=nY_r nX_c/N$ refers to the estimate of the expected number of sample points that should fall in cell (r,d). If the assumption that X_i and Y are independent holds, the statistic T in Eq. (53) approximately follows distribution of $\chi^2[(nX-1)(nY-1)]$. Then the p-value $prob\left(T>\hat{T}\,|\,T\sim\chi^2[(nX-1)(nY-1)]\right)$ measures the effect of X_i on Y. A small p-value indicates that X_i is influential.

7.1.5. Entropy-based VIMs

The entropy-based VIMs provide a set of measures for quantifying the nonlinear dependence between Y and X_i. The partitions of the sample space of Y and X_i are the same as that for SI test. The entropy of Y and X_i are defined as

$$H(Y)=-\sum_{r=1}^{nY}(nY_r/N)\ln(nY_r/N) \tag{54}$$

and

$$H(X_i)=-\sum_{d=1}^{nX}(nX_d/N)\ln(nX_d/N), \tag{55}$$

respectively. These two quantities measure the uncertainties of the samples of Y and X_i respectively. The joint entropy $H(Y,X_i)$ for quantifying the uncertainty associated with the joint samples of Y and X_i is defined as follows:

$$H(Y,X_i)=-\sum_{r=1}^{nY}\sum_{d=1}^{nX}(nO_{rd}/N)\ln(nO_{rd}/N). \tag{56}$$

The expected entropy $H(X_i\,|\,Y)$ of X_i conditional on Y and the expected entropy of Y conditional on X_i are computed by

$$\begin{aligned}H(X_i\,|\,Y)&=\sum_{r=1}^{nY}\left\{\frac{nY_r}{N}\right\}\left\{-\sum_{d=1}^{nX}\left(\frac{nO_{rd}}{N}\right)/\left(\frac{nY_r}{N}\right)\ln\left[\left(\frac{nO_{rd}}{N}\right)/\left(\frac{nY_r}{N}\right)\right]\right\}\\&=-\sum_{r=1}^{nY}\sum_{d=1}^{nX}\left[\frac{nO_{rd}}{N}\ln\left(\frac{nO_{rd}}{nY_r}\right)\right]=H(Y,X_i)-H(Y)\end{aligned} \tag{57}$$

and

$$\begin{aligned}H(Y\,|\,X_i)&=\sum_{d=1}^{nX}\left\{\frac{nX_d}{N}\right\}\left\{-\sum_{r=1}^{nY}\left(\frac{nO_{rd}}{N}\right)/\left(\frac{nX_d}{N}\right)\ln\left[\left(\frac{nO_{rd}}{N}\right)/\left(\frac{nX_d}{N}\right)\right]\right\}\\&=-\sum_{c=1}^{nX}\sum_{r=1}^{nY}\left[\frac{nO_{rd}}{N}\ln\left(\frac{nO_{rd}}{nX_d}\right)\right]=H(Y,X_i)-H(X_i)\end{aligned} \tag{58}$$

respectively. Then the contribution of the uncertainty in X_i to the entropy (uncertainty) of Y can be measured by

$$U(Y\,|\,X_i)=[H(X_i)-H(X_i\,|\,Y)]/H(X_i)=[H(Y)+H(X_i)-H(Y,X_i)]/H(X_i), \tag{59}$$

and the strength of the association between Y and X_i can be estimated by

$$U(Y,X_i)=2[H(Y)+H(X_i)-H(Y,X_i)]/[H(Y)+H(X_i)]. \tag{60}$$

Both $U(Y\,|\,X_i)$ and $U(Y,X_i)$ can be served as measures of variable importance [16]. If Y is independent of X_i, both measures equal zero; if Y is uniquely and fully determined by X_i, both measures equal unit. Values between zero and unit indicate the strength of the association between Y and X_i, and the higher the values are, the stronger the association is.

Table 5
Results of CMNs, CLs, CMD and IS tests for the independent case computed with formal statistical procedures and MCS procedure ($Np=10,000$), where the superscripts (MCS) in the first column indicates that the p-values in the corresponding rows are estimated with MCS procedure, and the superscripts in the other columns indicate the importance ranks induced by the p-values.

Variables	X_1	X_2	X_3	X_4	X_5	X_6
CMNs: $nX=10$	$0.0000^{(1)}$	$0.0000^{(4)}$	$0.0000^{(2)}$	$0.9974^{(6)}$	$0.9792^{(5)}$	$0.0000^{(3)}$
CMNs$^{(MCS)}$: $nX=10$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.9970^{(6)}$	$0.9811^{(5)}$	$0.0000^{(2.5)}$
CLs: $nX=10$	$0.0000^{(1)}$	$0.0000^{(3)}$	$0.0000^{(2)}$	$0.9964^{(6)}$	$0.9866^{(5)}$	$0.0000^{(4)}$
CLs$^{(MCS)}$: $nX=10$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.9946^{(6.0)}$	$0.9858^{(5.0)}$	$0.0000^{(2.5)}$
CMDs: $nX=10,\ nY=2$	$0.0000^{(1)}$	$0.0000^{(2)}$	$0.0000^{(3)}$	$0.0911^{(5)}$	$0.9659^{(6)}$	$0.0000^{(4)}$
CMDs$^{(MCS)}$: $nX=10,\ nY=2$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.9054^{(5.0)}$	$0.9675^{(6.0)}$	$0.0000^{(2.5)}$
SI: $nX=10, nY=5$	$0.0000^{(1)}$	$0.0000^{(4)}$	$0.0000^{(2)}$	$0.8791^{(5)}$	$0.9911^{(6)}$	$0.0000^{(3)}$
SI$^{(MCS)}$: $nX=10,\ nY=5$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.8767^{(5.0)}$	$0.9899^{(6.0)}$	$0.0000^{(2.5)}$

Table 6
Results of entropy-based measures together with the results of SI test for the independent case, where the subscripts indicate the importance rankings.

Variables		X_1	X_2	X_3	X_4	X_5	X_6
$nX=5,\ nY=5$							
SI test	χ^2	577.3	90.8	129.0	13.6	7.6	133.0
	p-Value	$0.0000^{(1)}$	$0.0000^{(4)}$	$0.0000^{(3)}$	$0.6322^{(5)}$	$0.9611^{(6)}$	$0.0000^{(2)}$
Joint entropy	$U(Y,X_i)$	$0.1719^{(1)}$	$0.0284^{(4)}$	$0.0374^{(3)}$	$0.0043^{(5)}$	$0.0023^{(6)}$	$0.0414^{(2)}$
Cond. entropy	$U(Y\|X_i)$	$0.1719^{(1)}$	$0.0284^{(4)}$	$0.0374^{(3)}$	$0.0043^{(5)}$	$0.0023^{(6)}$	$0.0414^{(2)}$
R measure	$R(Y,X_i)$	$0.6519^{(1)}$	$0.2958^{(4)}$	$0.3367^{(3)}$	$0.1167^{(5)}$	$0.0866^{(6)}$	$0.3531^{(2)}$
$nX=10,\ nY=5$							
SI test	χ^2	665.2	149.9	197.9	26.4	19.0	164.8
	p-Value	$0.0000^{(1)}$	$0.0000^{(4)}$	$0.0000^{(2)}$	$0.8791^{(5)}$	$0.9911^{(6)}$	$0.0000^{(3)}$
Joint entropy	$U(Y,X_i)$	$0.1558^{(1)}$	$0.0379^{(4)}$	$0.0450^{(2)}$	$0.0068^{(5)}$	$0.0050^{(6)}$	$0.0411^{(3)}$
Cond. entropy	$U(Y\|X_i)$	$0.1893^{(1)}$	$0.0460^{(4)}$	$0.0547^{(2)}$	$0.0082^{(5)}$	$0.0061^{(6)}$	$0.0500^{(3)}$
R measure	$R(Y,X_i)$	$0.6755^{(1)}$	$0.3711^{(4)}$	$0.4017^{(2)}$	$0.1614^{(5)}$	$0.1397^{(6)}$	$0.3856^{(3)}$

Another useful measure of association based on entropy is the defined as follows [128]:

$$R(Y,X_i)=\sqrt{1-\exp\{-2[H(X_i)+H(Y)-H(Y,X_i)]\}}. \tag{61}$$

which also takes values between zero and unit, and the value indicates the strength of the association between Y and X_i. If Y and X_i follow bivariate normal distribution, then $R(Y,X_i)$ tends to the absolute value of the CC as N, nX and nY increase [128].

7.1.6. Implementations of the grid-based test techniques

The CMNs, CLs, CMDs and SI tests are all based on estimating the p-values under proper assumptions, which may certainly not hold in many practical applications. It is necessary to test the effectiveness of these p-values. To deal with this problem, a Monte Carlo simulation (MCS) procedure has been proposed for numerically estimating the p-values [15,16]. This procedure is briefly described as follows. First, randomly permute the samples of X_i (or Y) so as to obtain a new set of sample pairs $\left(\tilde{x}_{ji},\tilde{y}_j\right)$ $(j=1,2,\ldots,N)$. Second, compute the value of statistics corresponding to any of the four tests with the new set of sample pairs $\left(\tilde{x}_{ji},\tilde{y}_j\right)$. These two steps are repeated for Np times, and Np sample values for the statistic of interest can be obtained. As the random permutation has broken the relationship between Y and X_i, these Np sample values of the statistic must follow the true distribution of the statistic other than the derived distribution based on any assumption, thus can be used for numerically estimating the p-values. These numerically estimated p-values can then be applied to test the effectiveness of the p-values computed based on assumptions.

With 1000 sample points generated with LDS schedule, the p-values for CMNs, CLs, CMDs and IS tests are computed for the independent case and the results are shown in Table 5. For demonstrating the effectiveness of these p-values, the MCS procedure is also performed for these four statistical tests and the results are listed in Table 5, where the number Np of replication is set to be 10,000. As can be seen, the results of all these five test techniques computed with formal statistical procedures are in good agreement with the respective results estimated by MCS procedure, indicating that these results from formal statistical procedures are accurate. It is shown that, although small differences exist among the importance ranking produced by these four test techniques, they all identify X_1, X_2, X_3 and X_6 as influential variables, and recognize X_4 and X_5 as non-influential variables.

The results for entropy-based measures with 5×5 and 10×5 grids are displayed in Table 6. For comparison, the results of SI test are also listed. As can be seen, although there are differences among the values of the three entropy-based measures of each variable, they produce the same importance ranking with the SI test, and the importance ranking is $X_1>X_6>X_3>X_2>X_4>X_5$. These results are also in accordance with those obtained by CMNs, CLs and CMD tests. For more examples on the grid-based VIMs, one can refer to Section 6.6 of Ref. [2].

7.2. Hypothesis tests without use of grid

There are also hypothesis test techniques that do not need to divide the sample space such as the squared rank difference/rank correlation coefficient (SRD/SRC) test, two dimensional Kolmogorov–Smirnov (KS) test and distance-based tests.

7.2.1. Squared rank difference/rank correlation coefficient (SRD/RCC) test

The SRD/RCC test use a statistic derived from the p-values of two hypothesis tests, i.e., SRD test and RCC test, both of which are based on rank-transformed data. The rationale is briefly described as follows.

The SRD test is based on the statistic

$$Q_i = \sum_{j=1}^{N-1} (r_{j+1,i} - r_{ji})^2, \tag{62}$$

where r_{ji} refers to the rank of Y obtained with the sample element in which X_i has rank j. If the assumption that there is no relationship between X_i and Y, Q_i approximately follows a normal distribution with mean $N(N^2-1)/6$ and SD $\sqrt{N^5}/N$ when $N>40$. Thus, the p-value $p_{srdi} = prob\left(Q_i > \hat{Q}_i | Q_i \sim N\left(N(N^2-1)/6, \sqrt{N^5}/N\right)\right)$ that the statistic Q_i exceeds the observed value $\hat{Q}_i$ can be used as a measure of the strength of the association between Y and X_i.

The statistic for RCC test is in fact the RCC between Y and X_i:

$$RC_i = \frac{\sum_{j=1}^{N} \left[r(x_{ji}) - (N+1)/2\right]\left[r\left(y_j\right) - (N+1)/2\right]}{\sqrt{\left\{\sum_{j=1}^{N} \left[r(x_{ji}) - (N+1)/2\right]^2\right\}\left\{\sum_{j=1}^{N} \left[r\left(y_j\right) - (N+1)/2\right]^2\right\}}}, \tag{63}$$

where $r(x_{ji})$ and $r\left(y_j\right)$ are the ranks of the samples x_{ji} and y_j, respectively. If the assumption that there is no rank correlation between Y and X_i holds the statistic RC_i follows a known distribution [129]. For $N \le 30$, the pth quantile of RC_i can be directly read from Table A10 of Ref. [129]; while for large N, the pth quantile of RC_i can be computed by $\omega_p = z_p/\sqrt{N-1}$, where z_p refers to the pth quantile of standard normal distribution. Then the p-value p_{rcci} that RC_i exceeds the estimate $\widehat{RC}_i$ indicates the strength of the monotonic dependence between Y and X_i.

By combining the p-values p_{srdi} and p_{rcci}, the statistic for SRD/RCC test is defined by [130]

$$\chi_4^2 = -2[\ln(p_{srdi}) + \ln(p_{rcci})], \tag{64}$$

which follows chi-square distribution with freedom of four degrees (see Section 2.8 of Ref. [131]). Then the p-value that the statistic χ_4^2 exceeds the estimate $\hat{\chi}_4^2$ can be used for measuring the strength of the dependence between Y and X_i.

7.2.2. Two-dimensional Kolmogorov–Smirnov (KS) test

With each sample x_{ji} of X_i, the sample space of (X_i, Y) can be divided into two distinct parts:

$$\mathbf{X}_1 = \{x_i | x_i > x_{ji}\}, \quad \mathbf{X}_2 = \{x_i | x_i < x_{ji}\}, \tag{65}$$

and with each sample y_j of Y, the sample space of (X_i, Y) is also divided into two parts

$$\mathbf{Y}_1 = \left\{y | y > y_j\right\}, \quad \mathbf{Y}_2 = \left\{y | y < y_j\right\}. \tag{66}$$

Let nX_1 and nX_2 denote the numbers of sample points contained in $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, nY_1 and nY_2 refer to the numbers of sample points contained in $\mathbf{Y}_1$ and $\mathbf{Y}_2$, respectively. Then, with each sample point $\left(x_{ji}, y_j\right)$ of (X_i, Y), the sample space can be divided into four quadrants

$$\begin{aligned} &Q_{j1} = \{(x_i, y) | x_i \in \mathbf{X}_1, y \in \mathbf{Y}_1\}, \quad Q_{j2} = \{(x_i, y) | x_i \in \mathbf{X}_2, y \in \mathbf{Y}_1\}, \\ &Q_{j3} = \{(x_i, y) | x_i \in \mathbf{X}_2, y \in \mathbf{Y}_2\}, \quad Q_{j4} = \{(x_i, y) | x_i \in \mathbf{X}_1, y \in \mathbf{Y}_2\}, \end{aligned} \tag{67}$$

as illustrated by Fig. 14.

Let fE_{jk} denote the expected fraction of sample points contained in Q_{jk}, which can be estimated by

$$\begin{aligned} &fE_{j1} = nX_1 nY_1/N^2,\ fE_{j2} = nX_2 nY_1/N^2,\ fE_{j3} = nX_2 nY_2/N^2, \\ &fE_{j4} = nX_1 nY_2/N^2, \end{aligned} \tag{68}$$

where fO_{jk} indicates the actual fraction of sample points contained in Q_{jk}. Then the statistic of KS test is given by [16,132]

$$D = \max\{|fE_{jk} - fO_{jk}|, j = 1, 2, \ldots, N, k = 1, 2, 3, 4\}. \tag{69}$$

If there is no relationship between X_i and Y, the statistic D is a random variable, and the p-value that D exceeds the estimate $\hat{D}$ computed with sample points $\left(x_{ji}, y_j\right)$ $(j = 1, 2, \ldots, N)$ can be approximated by

$$\text{prob}\left(D > \hat{D}\right) \cong Q_{KS}\left(\frac{D\sqrt{N}}{1 + \left[0.25 - 0.75/\sqrt{N}\right]\sqrt{1 - r_i}}\right), \tag{70}$$

where r_i is the CC between X_i and Y (see Eq. (23)), and $Q_{KS}(\bullet)$ is expressed as follows:

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty} (-1)^{j-1} \exp\left(-2j^2\lambda^2\right). \tag{71}$$

The p-value $\text{prob}\left(D > \hat{D}\right)$ can also be computed with the MCS procedure which has been illustrated in Section 7.1.6 for grid-based test techniques. The p-value $\text{prob}\left(D > \hat{D}\right)$ indicates the strength of the association between Y and X_i. The lower the p-value is, the stronger the relationship is.

7.2.3. Distance-based tests

This group of techniques includes the nearest neighbor (NN) test, total distance (TD) test and coefficient of aggregation (CA) test, all of which are based on the distances between sample points.

The statistic d_{NN} of the NN test is defined as [16,133]

$$d_{NN} = \sum_{j=1}^{N} d_{ji}/N, \tag{72}$$

where d_{ji} refers to the distance between $\left(x_{ji}, y_j\right)$ and its NN among the other $N-1$ sample points. With the assumption that there is no relationship between X_i and Y, the statistic d_{NN} is a random variable, and the p-value that d_{NN} is smaller than the estimate $\hat{d}_{NN}$ can be served as a measure of the effect of X_i on Y. The smaller the p-value is, the more effect X_i has on Y. The probability distribution of the statistic d_{NN} can be estimated with the MCS procedure as described in Section 7.1.6.

The statistic d_{TD} for TD test is given by [16]

$$d_{\text{TD}} = \sum_{j=1}^{N} \sum_{k=j+1}^{N} d_{jk}/[N(N-1)/2], \tag{73}$$

where d_{jk} is the distance between $\left(x_{ji}, y_j\right)$ and (x_{ki}, y_k). The p-value of obtaining a value for d_{TD} smaller than the estimate $\hat{d}_{\text{TD}}$ can be estimated by the MCS procedure and served as VIM.

The statistic d_{CA} for CA test is expressed as [16,134]

$$d_{\text{CA}} = \sum_{j=1}^{N} \tilde{d}_{ji}/\left[\sum_{j=1}^{N} \tilde{d}_{ji} + \sum_{j=1}^{N} d_{ji}\right], \tag{74}$$

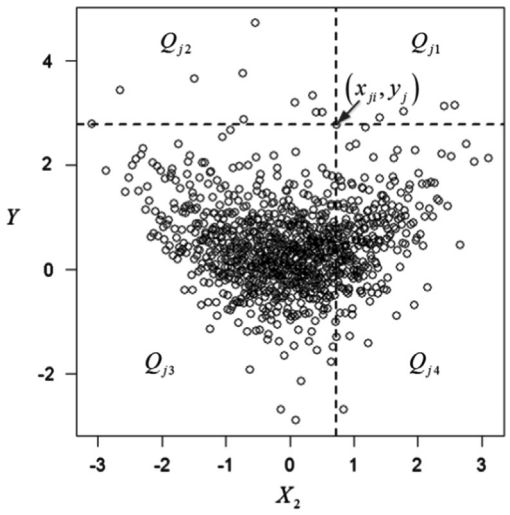

Fig. 14. Illustration of the four quadrants used for KS test.

where d_{ji} denotes the same notation as it in Eq. (72), and $\tilde{d}_{ji}$ is the distance between $(\tilde{x}_{ji}, \tilde{y}_j)$ and its NN among the sample points $(\tilde{x}_{ki}, \tilde{y}_k)$ $(k=1,2,\ldots,j-1,j+1,\ldots,N)$ generated by random permuting the respective samples of X_i and Y contained in (x_{ki}, y_k) $(k=1,2,\ldots,N)$. Similar to d_{NN} and d_{TD}, the distribution of d_{CA} can be estimated with the MCS procedure, and the p-value of obtaining the value for d_{CA} larger than the estimate $\hat{d}_{CA}$ can be computed based on this distribution. A low p-value indicates that X_i is influential.

7.2.4. Implementations of the statistical test techniques without using grid

The advantage of the non-grid based methods over the grid-based ones is that the results do not depend on the definition of grids, thus given a set of sample points, the results of tests are certain. The results of the SRD/RCC, KS, NN, TD and CA tests are computed for the independent case based on 1000 sample points generated by the LDS schedule, and the results are listed in Table 7, together with the results of SI test computed with 5×5 grid.

As can be seen, the results of SRD/RCC produce misleading information on variable importance ranking. Compared with the results of SI test, the SRD/RCC test correctly identifies the three important variables (X_1, X_3 and X_6), but mistakenly recognizes the influential variable X_2 as non-influential variable and recognizes the non-influential variable X_4 as influential variable.

The p-values of the KS test computed with Eq. (70) are much larger than those obtained by MCS procedure, thus are unconvincing. Comparable, the p-values of KS test computed by MCS procedure correctly identify the four influential input variables (X_1, X_2, X_3 and X_6) and two non-influential variables (X_4 and X_5), but does not show the relative importance of the four influential variables.

For computing the p-values of the NN and CA tests with MCS procedure, the R package "RANN" [135] is used for efficiently searching the NN for each sample point. With the results of the NN test, the four most influential variables are correctly identified, but the importance ranking is different with that obtained by the SI test. With the TD test, the four influential variables are also distinguished from the two non-influential variables, but the relative importance of the four influential variables is not distinguished. The CA test produces the same importance ranking with the NN test. For additional examples on the SRD/SRC test, KS test and distance-based tests, one can refer to Section 6.11 of Ref. [2].

8. Variance-based VIMs

The variance-based VIMs, also called Sobol's indices [136,137], are one of the most popular practice in many disciplines involving computational models. They measure the relative importance of one input variable by the partial variance of model output explained by this variable. The classical Sobol's indices are only defined for independent input variables based on high-dimensional model representation (HDMR) decomposition. In recent years, several works have been done to extend the Sobol's indices to correlated variables. Thus, we introduce the Sobol's indices for independent and dependent cases separately.

8.1. Independent case

8.1.1. Definitions and interpretations

The Sobol's indices aim at attributing the total variance of model output, instead of variance explained by any meta-model (e.g., multiple linear regression model), to each input variable with the consideration of the interaction effects among variables. By HDMR decomposition, the g-function can be uniquely decomposed into 2^n functional terms of increasing dimensions [136]:

$$Y=g(\mathbf{X})=g_0+\sum_i g_i(X_i)+\sum_i\sum_{j>i} g_{ij}(X_i,X_j)+\cdots+g_{12,\ldots,n}, \tag{75}$$

where $g_0=E(Y)$, $g_i=E(Y|X_i)-g_0$ and $g_{ij}=E(Y|X_i,X_j)-g_i-g_j-g_0$. As shown by Sobol' [136], if the g-function is square integrable and the input variables are independent with each other, then all the 2^n terms are orthogonal with each other. Taking variances to both sides of Eq. (75) yields:

$$V(Y)=\sum_i V_i+\sum_i\sum_{j>i} V_{ij}+\cdots+V_{12,\ldots,n}, \tag{76}$$

where $V_i=V(g_i)=V(E(Y|X_i))$ and $V_{ij}=V\left(g_{ij}\right)=V(E(Y|X_i,X_j))-V_i-V_j$.

Due to the total variance law, $V_i=V(Y)-E(V(Y|X_i))$, where $E(V(Y|X_i))$ is interpreted as the average residual variance of model output when X_i is fixed over its full range. Thus, the first order partial variance V_i can be explained as the average reduction of model output variance resulting from fixing X_i, that is, V_i measures the individual contribution of X_i to the total variance $V(Y)$. The larger V_i is, the more reduction of output variance can be obtained by reducing the uncertainty of X_i. The second order partial variance V_{ij} quantifies the interaction effect between X_i and X_j. Similar interpretations can be given to the higher order partial variances.

Another commonly used measure is the total partial variance V_{Ti} [137], which is defined as the summation of all terms in Eq. (76) with subscripts including i, that is, V_{Ti} incorporates both the individual effect of X_i and its interaction effects with all the other $n-1$ input variables $\mathbf{X}_{\sim i}=(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$. The smaller V_{Ti} is, the less X_i contributes to the model output variance. V_{Ti} can also be computed by subtracting the main effect of $\mathbf{X}_{\sim i}$ from the total variance, that is, $V_{Ti}=V(Y)-V(E(Y|\mathbf{X}_{\sim i}))=E(V(Y|\mathbf{X}_{\sim i}))$. Thus, V_{Ti} also measures the average residual variance of model output when all the inputs but X_i are fixed over their full ranges.

Standardizing V_i and V_{Ti} by the total variance $V(Y)$, the main effect index S_i and total effect index S_{Ti} are defined as:

$$S_i=\frac{V_i}{V(Y)}=\frac{V(E(Y|X_i))}{V(Y)} \quad \text{and}$$

$$S_{Ti}=\frac{V_{Ti}}{V(Y)}=\frac{V(Y)-V(E(Y|\mathbf{X}_{\sim i}))}{V(Y)}$$

Table 7
Comparison of the results of the non-grid based tests for the case of independence, where the superscripts indicate the ranks.

Variables	X_1	X_2	X_3	X_4	X_5	X_6
SI: $nX=nY=5$	$0.0000^{(1)}$	$0.0000^{(4)}$	$0.0000^{(3)}$	$0.6322^{(5)}$	$0.9611^{(6)}$	$0.0000^{(2)}$
SRD/RCC	$0.0000^{(2)}$	$0.0323^{(5)}$	$0.0000^{(2)}$	$0.0000^{(4)}$	$0.8880^{(6)}$	$0.0000^{(2)}$
KS	$0.0000^{(1)}$	$0.2527^{(4)}$	$0.0467^{(3)}$	$0.9996^{(5)}$	$0.9999^{(6)}$	$0.0079^{(2)}$
$\mathrm{KS}^{(\mathrm{MCS})}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.7970^{(5)}$	$0.9852^{(6)}$	$0.0000^{(2.5)}$
$\mathrm{NN}^{(\mathrm{MCS})}$	$0.0039^{(2)}$	$0.1291^{(4)}$	$0.0000^{(1)}$	$0.3341^{(5)}$	$0.7402^{(6)}$	$0.0655^{(3)}$
$\mathrm{TD}^{(\mathrm{MCS})}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.0000^{(2.5)}$	$0.3996^{(6)}$	$0.0203^{(5)}$	$0.0000^{(2.5)}$
$\mathrm{CA}^{(\mathrm{MCS})}$	$0.0080^{(2)}$	$0.2791^{(4)}$	$0.0006^{(1)}$	$0.7423^{(5)}$	$0.7746^{(6)}$	$0.1755^{(3)}$

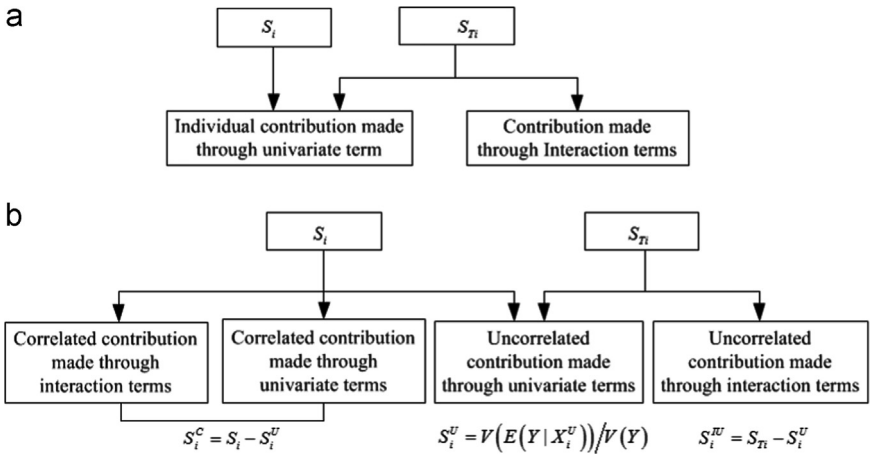

Fig. 15. Components contained in S_i and S_{Ti}: (a) X_i is uncorrelated with other inputs; (b) X_i is correlated with other inputs.

$$= \frac{E(V(Y \mid \mathbf{X}_{\sim i}))}{V(Y)}, \tag{77}$$

respectively. With definitions in Eq. (77), both S_i and S_{Ti} are bounded in [0, 1], and $S_i \le S_{Ti}$. Usually S_i is used for selecting important variables, while S_{Ti} is more suitable for screening non-influential variables. The difference $S_{Ti} - S_i$ measures the interaction effects of X_i with other inputs. If $S_i = S_{Ti}$ holds for all i, then there is no interaction effects, and the g-function is said to be additive.

8.1.2. Computational issues

The methods for computing Sobol's indices available in the literature can be divided into three groups: Fourier Amplitude Sensitivity Test (FAST), meta-model and Monte Carlo simulations.

The classical FAST method was developed by Cukier et al. [138,139] in 70s, much earlier than the proposition of Sobol's indices. It was later shown by Saltelli et al. [140] in 1999, that the statistical quantity estimated by the FAST method is in fact the main effect index S_i. In Ref. [140], Saltelli et al. also extended the FAST method for computing the total effect index. The rationale behind the classical FAST is to first introduce each variable a periodic signal by a well-designed space-filling curve and then compute the partial variances using the Fourier transformation. The classical FAST method faces two challenges. First, the minimum number of points increases with the input dimension. Second, a quite complex algorithm is necessary to choose a frequency for each input variable so that the higher harmonics of each frequency do not interfere with those of the others. Improper choice of frequencies will lead to low-accuracy of estimates. For conquering these shortcomings, several improvements have been made in the past years [140–142]. Until now, the most widely accepted version of FAST is the Random Balance Design (RBD) [141]. In this method, the choice of frequencies for each input is avoided, and the computational cost does not increase with the input dimension. Unfortunately, although several works have shown that the FAST method can be used to compute the total effect index [140,142,143], the FAST method is usually only used for computing the main effect index.

The basic idea behind the meta-model methods is to first approximate the g-function with an explicit or semi-explicit function (called meta-model) and then compute the Sobol's indices based on this meta-model. Commonly used meta-model methods are polynomial chaos expansion [103], Bayesian approach [144], sparse grid interpolation [145], polynomial dimensional decomposition [146], cut-HDMR [147], random sampling (RS)-HDMR [148], Neural network [149], state dependent regression (SDR) [104,105], random forest, LOESS, GAM, projection pursuit [19,20] and Kriging interpolation [150]. These meta-model methods are able to compute the Sobol's indices with a relatively small computational cost, thus can be particularly useful when the g-function is computationally expensive. The main drawback of the meta-model methods is that they are only capable of computing the lower order effects. When the g-function is mainly governed by interaction effects, these methods are weak in computing the total effect index.

The Monte Carlo estimators for Sobol's indices have been studied substantially [63,67,68,136,137,151]. Given two $(N \times n)$ sample matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of input variables, one can obtain another sample matrix $\mathbf{A}_B^{(i)}$ by adapting the ith column from $\boldsymbol{A}$ and the other column from $\boldsymbol{B}$. Let $\mathbf{Y}_A = \left(y_{A(j)}\right)_{j=1,\ldots,N}$, $\mathbf{Y}_B = \left(y_{B(j)}\right)_{j=1,\ldots,N}$ and $\mathbf{Y}_{\mathbf{A}_B^{(i)}} = \left(y_{\mathbf{A}_B^{(i)}(j)}\right)_{j=1,\ldots,N}$ denote the output value vectors corresponding to $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{A}_B^{(i)}$. All the available estimators are collected and compared in Ref. [68]. Among the many estimators, Saltelli et al. [68] have shown that the best practice for V_i and V_{Ti} are:

$$V_i = \frac{1}{N}\sum_{j=1}^{N} y_{B(j)}\left(y_{\mathbf{A}_B^{(i)}(j)} - y_{A(j)}\right) \quad \text{and} \quad V_{Ti} = \frac{1}{2N}\sum_{j=1}^{N}\left(y_{A(j)} - y_{\mathbf{A}_B^{(i)}(j)}\right)^2, \tag{78}$$

respectively. The key to compute the Sobol's indices by the Monte Carlo estimators is the generation of three sample matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{A}_B^{(i)}$. Generally, two groups of methods are available in literature. The first group of methods generates the sample matrices by sampling techniques. The commonly used schedules are SRS, LHS [40] and LDS [42,43]. The latter two techniques are mostly used since they lead to higher convergence rate than the simple random sampling. These two sampling techniques are compared in Ref. [152]. The authors showed that, in almost all the cases they investigated, the LDS schedule performs better than the LHS schedule. The second group of methods includes the five strategies introduced in Section 3.2 for implementing Morris' methods, that is, the Morris' design [60], optimization-based design [61], Winding Stairs design [63–66], radial design [67–69] and cell-based design [70]. These five strategies allow for the design of matrices $\mathbf{A}$ and $\mathbf{A}_B^{(i)}$, thus can be used for estimating the total effect indices. If the g-function is computationally expensive, one can use small J (number of trajectories) to perform Morris' method, while the computational cost allows, one can increase J to compute the total effect indices [69]. Except the above two groups of methods, one can also generate the sample matrix $\mathbf{B}$ by randomly permuting each

column of $\mathbf{A}$ individually, and then establish $\mathbf{A}_B^{(i)}$ by assigning all but the ith column of $\mathbf{B}$ to $\mathbf{A}_B^{(i)}$ and the ith column of $\mathbf{A}$ to $\mathbf{A}_B^{(i)}$.

8.2. Dependent case

Kucherenko et al. [153] showed that, if X_i is correlated with other inputs, S_i can be higher than S_{Ti} depending on the level of correlations. This phenomenon is interpreted in Refs. [81,83,154] for additive model (without interactions) and in Ref. [155] for non-additive model (with interactions). When all the inputs are independent with each other, S_i only reflects the individual contribution of X_i and S_{Ti} includes both the individual contribution of X_i and the interaction contributions of X_i with the other input variables, as shown in Fig. 15(a). However, when X_i is correlated with others, Hao et al. [155] showed that, for a non-additive model, S_i consists of three components and S_{Ti} consists of two components, as shown in Fig. 15(b). Comparing Fig. 15(a) and (b), one can find that, as X_i is correlated with the other input variables, two components arise in S_i: correlated contribution made through interaction terms and correlated contribution made through univariate terms. The higher the correlations are, the more correlated contributions may be introduced to S_i. With some levels of correlation, S_i may be larger than S_{Ti}. Due to the above reason, researchers became aware of the necessity of separating the different types of contributions from S_i and S_{Ti} [81,83,154–158]. Among all these works, Xu and Gertner's decomposition and its improved versions have received the most attention. In Section 4.1.4, Xu and Gertner's decomposition based on linear regression has been introduced. In recent years, Xu and Gertner's decomposition has been extended to general model with linearly or nonlinearly correlated input variables, which is reviewed below.

As X_i is correlated (linearly or nonlinearly) with the remaining inputs $\mathbf{X}_{\sim i}$, Xu and Gertner [81,158] decomposed X_i into two components X_i^U and X_i^C, where $X_i^C = E(X_i|\mathbf{X}_{\sim i})$ and $X_i^U = X_i - X_i^C$. X_i^U is linearly independent of X_i^C and $\mathbf{X}_{\sim i}$ (see Refs. [157,158] for detail). Thus, X_i^U represents the uncorrelated variation of X_i, and the main effect of X_i^U, i.e., $S_i^U = V\left(E\left(Y|X_i^U\right)\right)/V(Y)$, measures the uncorrelated contribution made through univariate terms, as shown in Fig. 15(b). Further, $S_i^C = S_i - S_i^U$ quantifies the correlated contributions made through both the univariate and interaction terms. Based on the above work, Hao et al. [155] showed that the uncorrelated contribution made through interaction terms can be computed as $S_i^{IU} = S_{Ti} - S_i^U$ (see the last component in Fig. 15(b)).

Several numerical methods have been introduced to implement the above decomposition. Monte Carlo estimators based on copula were derived by Kucherenko et al. in Ref. [153] for S_i and S_{Ti}. The FAST method for computing S_i was introduced by Xu and Gertner in Refs. [159,160]. In Ref. [158], Xu extended the FAST method for computing S_i^U and S_i^C. Monte Carlo estimators for S_i^U and S_i^C were derived in Ref. [161].

Additive models are widely used in practical applications. In this case, all the contributions made through the interaction terms disappear, and the uncorrelated contribution S_i^U is equal to the total effect index S_{Ti}, i.e., $S_i^U = S_{Ti}$. Meanwhile, the correlated contribution S_i^C can be computed as $S_i^C = S_i - S_i^U$. Based on this idea, artificial neural network (ANN) [83], point estimation procedure [162] and SDR meta-model [163] have been introduced for computing S_i^U and S_i^C in the case of additive model.

8.3. Implementations and discussions of variance-based VIMs

The Sobol's indices attribute the model output variance to each individual input variable and their interactions. Compared with the linear regression based methods, the attributed variance is the total model output variance other than variance explained by regression model, thus they are model free. The Sobol's indices not only produce robust importance ranking of input variables, make clear the sources of model output variance (from variation of input variables or from correlation between input variables), but can also reflect the model function behavior (additive or non-additive). The Sobol's indices are frequently used for reducing the variance (uncertainty) of model output, however, it can also be used for other purposes. For example, in Ref. [164], Wei et al. extended the Sobol's indices to structural reliability analysis, and proposed the global reliability sensitivity analysis technique. In Ref. [165], Sobol' employed the Sobol's indices to estimate the approximation error when fixing non-influential variables.

Researchers' doubt on Sobol's indices is mainly in three aspects. First, compared with other methods, such as Morris' screening method and the random forest based methods, the Sobol's indices are computationally more expensive and not appropriate for very high-dimensional problems. Second, the premise of Sobol's indices – the model input variables can be fixed at some points through further research – is not reasonable especially when the inputs contain aleatory uncertainty (inherent uncertainty that cannot be reduced through further study) [166]. For avoiding this disadvantage, Wei et al. [167] developed the W-indices with the premise of reducing the ranges of input variables instead of fixing the input variables. Third, the variance may not be sufficient to measure the uncertainty of model output, thus the moment-independent VIMs have been proposed, which will be introduced in the next section. Despite this doubts, the Sobol's indices are unquestionable one of the most popular methods in many disciplines involving computational models.

For the independent case, the main and total effect indices are estimated by the estimators in Eq. (78) (with sample size $N = 2000$) and SDR regression (with training sample size $N = 1024$), and the results are reported in Table 8, where the sample points are all generated by LDS schedule. As can be seen, the results obtained with the two computational procedures are in good agreement. The main and total effect indices produce the same importance ranking of $X_1 > X_3 > X_6 > X_2 > X_5 > X_4$. It is also shown that the total effect indices of X_5 and X_6 are obvious higher than their respective main effect indices, indicating that there are interaction effects.

The main effect indices for the dependent case (2nd row of Table 9) show that, when the correlation structure is injected, the relative importance of the correlated inputs (e.g., X_5 and X_6) are enhanced compared with that in the independent case. This is due to the fact that the correlated effects are included in the main effect indices. It is shown that, the summation of the uncorrelated effect indices (3th row of Table 9) is much smaller than one, indicating that the most of the model output variance comes from the correlations, as indicated by the correlated effect indices listed in the 4th row. As that the total effect indices does not include the correlated effects, their values are much smaller than those in the independent case.

9. Moment-independent VIMs

With the motivation that variance may not be sufficient to describe the uncertainty of model output, a group of VIMs by looking at the full distribution range of the output variable have been developed [168–173]. Among all these measures, the delta index proposed by Borgonovo [172] has received the most attentions. The delta index δ_i of X_i is defined by the average distance between the unconditional output density $f_Y(y)$ and the conditional density $f_{Y|X_i}(y)$ when X_i is fixed over its full distribution

Table 8
Main and total effect indices for the independent case computed with the estimators in Eq. (78) and the SDR technique.

Measures	X_1	X_2	X_3	X_4	X_5	X_6	Costs
Monte Carlo estimators in Eq. (78) with sample size $N=2000$							
S_i	$0.3097^{(1)}$	$0.1544^{(4)}$	$0.2832^{(2)}$	$0.0010^{(6)}$	$0.0062^{(5)}$	$0.1606^{(3)}$	16,000
S_{Ti}	$0.3116^{(1)}$	$0.1526^{(4)}$	$0.2860^{(2)}$	$0.0147^{(6)}$	$0.0948^{(5)}$	$0.2804^{(3)}$	
SDR regression with training sample size $N=1024$							
S_i	$0.3049^{(1)}$	$0.1433^{(4)}$	$0.2741^{(2)}$	$0.0000^{(5.5)}$	$0.0000^{(5.5)}$	$0.1574^{(3)}$	1024
S_{Ti}	$0.3049^{(1)}$	$0.1433^{(4)}$	$0.2741^{(2)}$	$0.0000^{(6)}$	$0.0718^{(5)}$	$0.2392^{(3)}$	

Table 9
The results for the dependent case computed by SDR technique with 1024 sample points.

Measures	X_1	X_2	X_3	X_4	X_5	X_6
S_i	$0.2802^{(4)}$	$0.2807^{(3)}$	$0.1613^{(5)}$	$0.0092^{(5)}$	$0.3780^{(2)}$	$0.3940^{(1)}$
S_i^U	0.0256	0.0037	0.1612	0.0090	0.0027	0.0209
S_i^C	0.2516	0.2770	0.0000	0.0002	0.3507	0.3778
S_{Ti}	0.0377	0.0540	0.2011	0.0493	0.0780	0.1068
S_i^{IU}	0.0121	0.0503	0.0399	0.0403	0.0753	0.0879

range, that is,

$$\delta_i = \frac{1}{2}E\left\{\int |f_Y(y) - f_{Y|X_i}(y)|\mathrm{d}y\right\}. \tag{79}$$

δ_i can be interpreted as the average change of model output uncertainty resulting from fixing X_i, where the uncertainty of model output is measured by the change of density instead of any moment. That is why the delta index is said to be moment-independent. Wei et al. [174] and Plischke et al. [175] independently showed that

$$\delta_i = \frac{1}{2}\int\int |f_{YX_i}(y,x_i) - f_Y(y)f_{X_i}(x_i)|\mathrm{d}y\mathrm{d}x_i, \tag{80}$$

where $f_{X_i}(x_i)$ and $f_{YX_i}(y,x_i)$ are the marginal density of X_i and the joint density of (Y,X_i), respectively. Eq. (80) indicates that δ_i can also be interpreted as the measure of dependence between Y and X_i.

As pointed out by Wei et al. [173], the dependence between two random variables is fully governed by their copula, where a copula is function that couples the joint distribution of a set of random variables with their marginal distribution functions (for detail of copula see Refs. [176,177]). Let $c(u,v_i)$ and $C(u,v_i)$ denote the copula density and copula function of (Y,X_i), respectively, where u and v are the marginal distribution functions of Y and X_i, respectively. It is shown by Wei et al. that [173]

$$\delta_i = \frac{1}{2}\int_0^1\int_0^1 |c(u,v_i) - 1|\mathrm{d}u\mathrm{d}v_i. \tag{81}$$

With the motivation that dependence measures can be used for measuring variable importance, Wei et al. [173] introduced the following extended delta index, first proposed by Schweizer and Wolff [178,179], as a new moment-independent VIM.

$$\delta_i^E = 12\int_0^1\int_0^1 |C(u,v_i) - uv_i|\mathrm{d}u\mathrm{d}v_i. \tag{82}$$

With Eqs. (80) and (81), the computation of the delta index is straightforward as long as the densities $f_{YX_i}(y,x_i)$ and $f_Y(y)$ or the copula density $c(u,v_i)$ can be estimated. Given a set of sample points, Wei et al. suggested using the kernel density estimators developed by Botev et al. [180] (with Matlab package available in Ref. [181]) for computing these densities. For estimating the extended delta index from given data, Wei et al. suggested using the empirical copula function [173].

Other types of moment-independent VIMs are also available such as the one based on KS metric [182]:

$$\delta_i^{KS} = E\left\{\sup_{y\in[-\infty,+\infty]}\left|F_Y(y) - F_{Y|X_i}(y)\right|\right\}, \tag{83}$$

and the one by averaging the Kullback–Leibler divergence on densities [182]:

$$\delta_i^{KL} = E\left\{\int_{y\in[-\infty,+\infty]} f_{Y|X_i}(y)\ln\left(\frac{f_{Y|X_i}(y)}{f_Y(y)}\right)\right\}. \tag{84}$$

For more details of on δ_i^{KS} and δ_i^{KL} one can refer to Ref. [182].

The constants $1/2$ in Eqs. (79) and (12) in Eq. (82) promise the normalization of δ_i and δ_i^E between 0 and 1. Both δ_i and δ_i^E can be served as measures of dependence between Y and X_i. $\delta_i = 0$ (or $\delta_i^E = 0$) indicates that Y is absolutely independent of X_i, and X_i is not in the model response function. $\delta_i = 1$ (or $\delta_i^E = 1$) implies that Y is fully and uniquely dependent on X_i, and there is no other variables but only X_i in the model response function. If δ_i (or δ_i^E) takes value between 0 and 1, then Y is partially dependent on X_i, and the larger the index is, the stronger the dependence is. This property of moment-independent VIMs makes them more useful than the variance-based VIMs when used for variable fixing and uncertainty reduction [182]. The second advantage of moment-independent VIMs over variance-based VIMs is that they are monotonic transformation invariant [182]. This property enables us to compute the moment-independent VIMs efficiently and accurately by performing monotonic transformation on the model output, when the output is severely skewed and has ranges over several orders of magnitude [182]. The third advantage of moment-independent VIMs is their suitability for model with huge number of inputs [175]. The fourth advantage is that all the moment-independent VIMs are well posed when the input variables are correlated. Compared to the variance-based VIMs, the main disadvantage of moment-independent VIMs is that they cannot reflect the behavior of the model response function.

The delta and extended delta indices for both the independent and dependent cases are computed by the copula-based methods proposed in Ref. [173], and the results are shown in Fig. 16. Both methods compute the respective indices with the same set of sample points generated by LDS schedule. As can be seen, 500–1000 sample points are generally sufficient for generating converged results in both cases. In the independent case, the

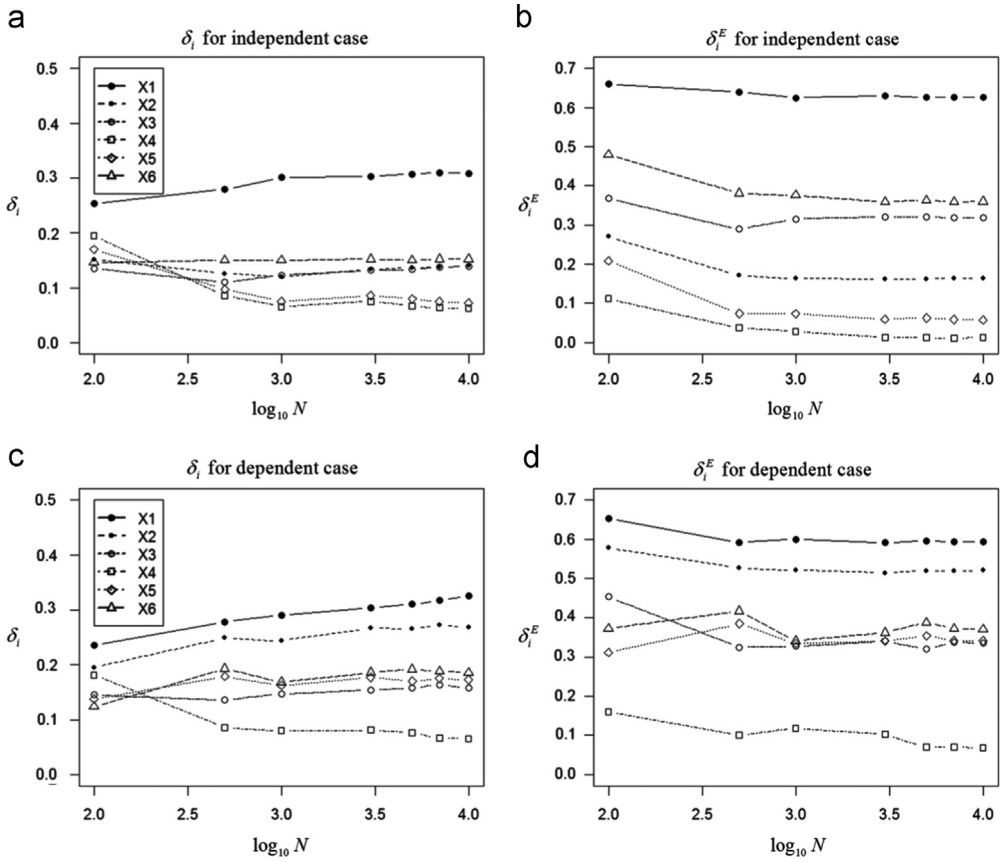

Fig. 16. Estimates of the delta and extended delta indices as a function of the log of sample size N.

importance rankings induced by δ_i and δ_i^E are $X_1 > X_6 > X_3 \approx X_2 > X_5 > X_4$ and $X_1 > X_6 > X_3 > X_2 > X_5 > X_4$, respectively. Thus, δ_i and δ_i^E result in nearly the same importance ranking except that δ_i^E identify X_3 as more influential than X_2 but δ_i think that X_2 and X_3 are equally important. The importance rankings induced by both indices are different with the results derived from the variance-based VIMs, as illustrated in Table 8.

It is shown in Fig. 16(c) and (d) that, as the correlations are introduced, the relative importance of X_2 and X_5 evaluated by both δ_i and δ_i^E have been largely enhanced, indicating that the correlated contributions are also involved in δ_2 and δ_5 as well as δ_2^E and δ_5^E.

10. Graphic VIMs

All the aforementioned methods aim at determining the relative importance of the input variables by defining one or multiple importance indices to each input. Then, the next problem is what we can do with the importance rankings after we have them. The graphic importance measures deal with this type of problem. The settings of the graphic importance measures are given as follows:

- Understanding the behavior of g-function.
- Investigating the relative importance of input variables.
- Measuring the effect of different subregions of input variables on the output variable.
- Quantifying the uncertainty reductions of model output when the uncertainties of input variables are reduced.

The first setting is also one of the objectives of Sobol's indices. As we will see, the information on model behavior provided by the Sobol's indices can be enriched by the graphic VIMs especially the regional VIMs. The second setting is also that of all the aforementioned importance measures. The latter two settings are the "unique skills" of the graphic VIMs.

Commonly used graphic VIMs are scatterplot (Section 1.2.3 in Ref. [1]), meta-model plot [104,105], regional VIMs [183–187] and parametric VIMs [188]. Given a set of sample points, one can plot the samples of each pair (Y, X_i) on a two-dimensional plane. By the shape of the cloud of the points, one cannot only empirically judge the relative importance of each input variable, but can also investigate the behavior of the g-function. Two dimensional scatter plot can only reflect the behaviors of the univariate components $E(Y|X_i)$. Compared with the scatterplot, the meta-model methods can explicitly (other than empirically) estimate the conditional moments. As we will see later, these information on the conditional moments are also included in the regional VIMs.

The regional VIA technique was first developed by Sinclair [183], and further developed by Bolado-Lavin et al. [184], Tarantola et al. [185] and Wei et al. [187], although the concept "regional VIA" was first introduced by Wei et al. [186,187]. The first regional VIM, developed by Sinclair [183] and revived by Bolado-Lavin et al. [184], is the contribution to sample mean (CSM) plot. The CSM function for X_i is defined as follows:

$$CSM_i(q) = \frac{1}{E(Y)} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{-\infty}^{F_i^{-1}(q)} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}_{\sim i} \mathrm{d}x_i, \tag{85}$$

where $F_i^{-1}(q)$ is the inverse distribution function of X_i at quantile q and $f_{\mathbf{X}}(\mathbf{x})$ is the joint density of the input variables. The n-dimensional integral in Eq. (85) is computed on the full ranges for all input variables except X_i, for which it is computed from the lower bound to the quantile q. One should note that the value of $CSM_i(q)$ may not be bounded in [0, 1], as shown by Wei et al. [187].

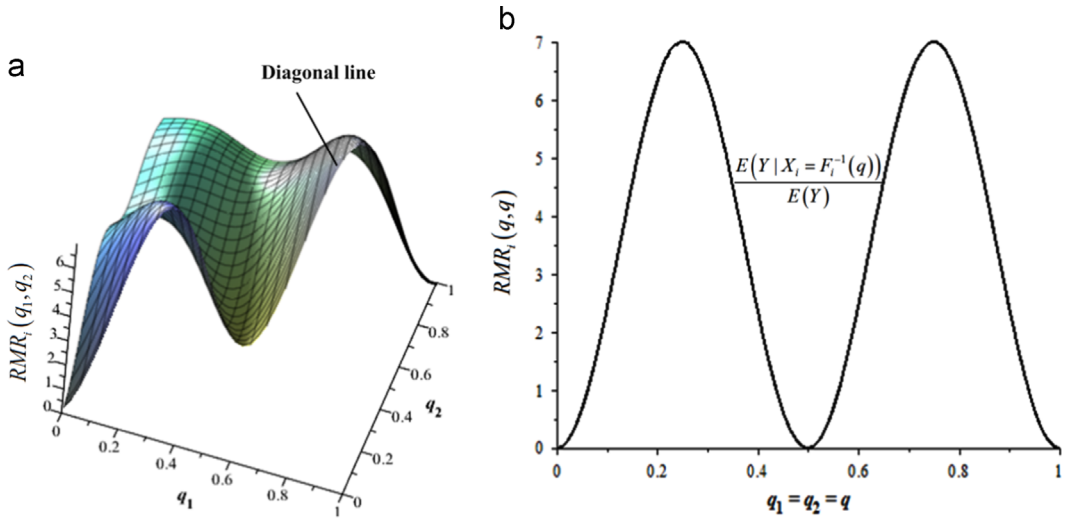

Fig. 17. Graphic illustration of the mean ratio function: (a) 3D plot of $RMR_i(q_1,q_2)$, and (b) diagonal line of $RMR_i(q_1,q_2)$.

$CSM_i(q)/q$ quantifies the amount of relative change of model output when the upper bound of X_i is reduced to $F_i^{-1}(q)$. Tarantola et al. [185] and Wei et al. [187] suggested a regional mean ratio (RMR) function defined as follows:

$$RMR_i(q_1,q_2)=\frac{CSM_i(q_2)-CSM_i(q_1)}{q_2-q_1}=\frac{1}{E(Y)}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{F_i^{-1}(q_1)}^{F_i^{-1}(q_2)}g(\mathbf{x})\frac{f_{\mathbf{X}}(\mathbf{x})}{q_2-q_1}\mathrm{d}x_i\mathrm{d}\mathbf{x}_{\sim i}, \tag{86}$$

where $0\le q_1\le q_2\le 1$. A 3D plot of $RMR_i(q_1,q_2)$ is shown in Fig. 17 (a), from which one can directly read the amount of relative change of model output mean when the range of X_i is reduced to any subregion $\left[F_i^{-1}(q_1),F_i^{-1}(q_2)\right]$. Let $q_1=q_2=q$, then the diagonal line $RMR_i(q,q)$ is plotted in Fig. 17(b). It is shown by Wei et al. [167,187] that this diagonal line is in fact the conditional expectation $E\left(Y|X_i=F_i^{-1}(q)\right)/E(Y)$, which reflects the behavior of the univariate functional component in the HDMR decomposition (see Eq. (75)). This observation has two potential applications. First, this can be applied for constructing the additive meta-model so as to approximate to g-function by omitting the higher order (≥ 2) functional components in Eq. (75). Second, the conditional expectation $E\left(Y|X_i=F_i^{-1}(q)\right)$ derived from the diagonal line of $RMR_i(q_1,q_2)$ can be used for estimating the main effect index, i.e., $V_i=V\left(E\left(Y|X_i=F_i^{-1}(q_i)\right)\right)$, where q_i follows uniform distribution between 0 and 1.

Inspired by the CSM function, Tarantola et al. proposed [185] the contribution to sample variance (CSV) function and a variance ratio function, both of which reflect the amount of deviation of model output mean from the original mean when the range of one input is reduced, but cannot tell the actual reduction of model output variance due to the reduced range. For the latter purpose, Wei et al. [187] developed a regional variance ratio (RVR) function defined as

$$RVR_i(q_1,q_2)=\frac{1}{V(Y)}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{F_i^{-1}(q_1)}^{F_i^{-1}(q_2)}\left(g(\mathbf{x})-E\left(Y^{[q_1,q_2]}\right)\right)^2\frac{f_{\mathbf{X}}(\mathbf{x})}{q_2-q_1}\mathrm{d}x_i\mathrm{d}\mathbf{x}_{\sim i}, \tag{87}$$

where $E\left(Y^{[q_1,q_2]}\right)$ is the residual model output mean due to the reduced range $\left[F_i^{-1}(q_1),F_i^{-1}(q_2)\right]$, and $E\left(Y^{[q_1,q_2]}\right)=RMR_i(q_1,q_2)E(Y)$. From the 3D plot of $RVR_i(q_1,q_2)$, the actual reduction of model output variance due to any reduced ranges of input variables can be directly obtained. The diagonal line $RVR_i(q,q)$ is the conditional variance $V\left(Y|X_i=F_i^{-1}(q)\right)/V(Y)$, and the area covered by this diagonal line is equal to $1-S_i$. This reveals the connection between the regional VIMs and the Sobol's indices [167]. In Eq. (87), if we let $q_1=q$ and $q_2=1-q$ with $q\in[0,0.5]$, then the counter-diagonal line $RVR(q,1-q)$ measures the residual sample variance when the range of X_i is symmetrically reduced to $\left[F_i^{-1}(q_1),F_i^{-1}(q_2)\right]$.

Another group of graphic VIMs are the parametric VIMs, which reflect the changes of model probabilistic responses (e.g., model output variance) w.r.t to the changes of the distribution parameters of model inputs. in Ref. [188], Wei et al. developed the univariate parametric mean ratio (PMR) and parametric variance ratio (PVR) functions w.r.t. the reduced variance of X_i as:

$$PMR_i(q)=\frac{E\left(Y^{(q\sigma_i^2)}\right)}{E(Y)}=\frac{1}{E(Y)}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}g(\mathbf{x})f_{\mathbf{X}}^*(\mathbf{x};q)\mathrm{d}\mathbf{x} \tag{88}$$

and

$$PVR_i(q)=\frac{1}{V(Y)}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left(g(\mathbf{x})-E\left(Y^{(q\sigma_i^2)}\right)\right)^2f_{\mathbf{X}}^*(\mathbf{x};q)\mathrm{d}\mathbf{x}, \tag{89}$$

respectively, where $f_{\mathbf{X}}^*(\mathbf{x};q)$ is the joint density of the input variables when the variance of X_i is reduced from σ_i^2 to $q\sigma_i^2$ with $0\le q\le 1$. One can also define parametric VIMs w.r.t. any other distribution parameters of input variables. The interpretations of $PMR_i(q)$ and $PVR_i(q)$ are straightforward. $PVR_i(q)$ measures the amount of residual variance of model output when the variance of X_i is reduced to $q\sigma_i^2$.

All the regional and parametric VIMs can be computed with a set of sample points. In Refs. [184,185,187] Monte Carlo estimators based on ordering the sample points of model inputs are derived for estimating the CSM, CSV, RMR and RVR functions. In Ref. [188], Wei derived Monte Carlo estimators for the PMR and PVR functions. For example, given the sample matrix $\mathbf{M}_x=(x_{ij})$ with $i=1,2,\ldots,N$and $j=1,2,\ldots,n$ generated from the original joint PDF $f_{\mathbf{X}}(\mathbf{x})$, and the corresponding output values $\mathbf{M}_y^T=(y_1,y_2,\ldots,y_n)$, the Monte Carlo estimator for $PMR_i(q)$ is given as follows:

$$PMR_i(q)\cong\frac{\sum_{j=1}^N y_jf_X^*(\mathbf{x}_j;q)/f_{\mathbf{X}}(\mathbf{x}_j)}{\sum_{j=1}^N y_j}. \tag{90}$$

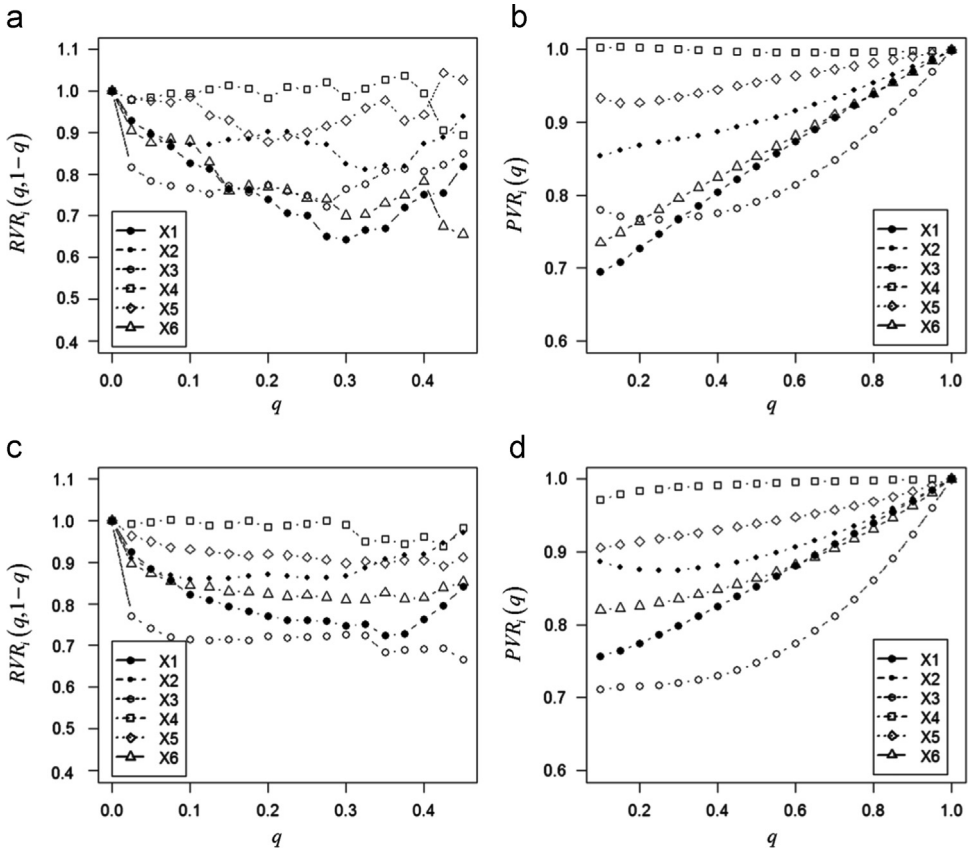

Fig. 18. Results of regional and parametric VIMs for the independent case, where (a) and (b) are the counter-diagonal lines $RVR(q,1-q)$ of the regional variance ratio functions and the parametric variance ratio functions computed with 1000 sample points, and (c) and (d) refer to the results computed with 10,000 sample points.

This estimator can also be extended to estimate the RMR and RVR functions so that ordering the sample points can be avoided.

Compared with Sobol's indices, the graphic VIMs not only provide more information on both the relative importance of inputs and the model behavior, but also show how a reduction in the uncertainty of each input variable will influence the uncertainty of the output. Both the regional and parametric VIMs can be extended to multivariate cases for measuring the interaction effects [167,188]. The disadvantage of the graphic VIMs, when compared with Sobol's indices, is that the higher order (>2) interaction effects and the total effects cannot be shown visually. For ease of reading, it is suggested to show the bivariate regional and parametric VIMs by contour plots instead of 3D plots [188].

The counter-diagonal lines $RVR(q,1-q)$ of the regional variance ratio function as well as the parametric variance ratio function are computed by the estimators in Ref. [188] with 1000 and 10,000 sample points generated by LDS schedule, and the results are plotted in Fig. 18. It is shown that, for this test model, the estimates of the parametric variance ratio functions with 1000 sample points are much more robust and accurate than the estimates of regional variance ratio functions. From Fig. 18(c), the reduction of model output variance due to the symmetric reduction of the region of each input can be directly read. For example, as the region of the X_1 is reduced to $\left[F_i^{-1}(0.1),F_i^{-1}(0.9)\right]$, the model output variance is reduced by 30 percent approximately. From Fig. 18(d), when the variances of the six inputs are reduced to 0.8, X_3 results in the most reduction of model output variance, followed by X_1 and X_6, and then X_2. With the reduction of the variance of X_4, the model output variance almost keeps constant.

11. Conclusions, discussions, recommendations and prospects

This review article concerns the collection of all the good practices for VIA developed in different disciplines. These VIMs can be divided into two groups: mathematical techniques and statistical techniques. The mathematical techniques include the difference-based VIMs (Section 2), variance-based VIMs (Section 8), moment-independent VIMs (Section 9) and graphic VIMs (Section 10), and the other VIMs belong to the statistical techniques.

The mathematical techniques are commonly developed for measuring the importance of input variables of computational models, and most of them need to compute the model response function at prescribed or well-designed points. For example, for computing the VIMs of Morris' screening method, we need to compute the model output values at the points on the preselected trajectories. This feature makes these methods not suitable for analysis with only data. However, some VIMs in this groups can be applied to data such as the moment-independent VIMs and the variance-based VIMs (computed with meta-model or RBD).

The statistical techniques are all especially designed for extracting the variable importance information based on data, where the data can be generated by calling the response function at the sample points of input variables obtained with sampling schedule (e.g., LDS) or generated from experimental measurement. We call this type of property of the statistical techniques as "data-driven". This property makes these statistical techniques applicable to both the computational model and data.

In summary, the recommendations are as follows:

– In the case of correlated input variables, the best practice till now for separating the correlated and uncorrelated effects are

Xu and Gertner's decomposition and its extended versions (see Sections 4.1.4 and 8.2). If the practitioners' interest is on investigating the different types of contributions, then the CPVIM based on CIT-RF as well as the decomposition-based VIMs are suggested (if $R^2 > 0.7$, the decomposition-based VIMs based on linear regression model are suggested, otherwise, the version for nonlinear model is suggested, see Section 8.2). However, when the correlations are inherent mutual property of the input variables and practitioners want to rank the input variables without eliminating the correlated effects, then the available methods are parametric regression based VIMs (CC, RCC, SRC, SRRC etc.), nonparametric regression techniques, random forest based VIMs (GVIM and PVIM), hypothesis test techniques, moment-independent VIMs (delta and extended delta indices) and variance-based VIMs (main and total effect indices).

- If the model response function is linear or approximately linear, then the best practice is the multiple linear regression based VIMs (CC, SRC and PCC). For nonlinear but monotonic response function, the rank regression technique and the related VIMs (RCC, SRRC and PRCC) can be used. For nonlinear and non-monotonic response function, the polynomial regression or nonparametric regression (LOESS, GAM and PP_REG) techniques can be applied. When the input dimension is high, the stepwise implementation of the regression techniques can be performed, and the incremental changes in R^2 values due to addition of new variables can be computed as used as VIMs.
- For problem with high-dimensional input variables, the available methods are Morris' screening method, parametric and nonparametric regression implemented with stepwise procedure, random forest based VIMs, hypothesis test techniques, moment-independent VIMs and graphic VIMs. The computational cost of Morris' screening method increases with the increase of input dimension or the degree of nonlinearity. Both the regression techniques implemented with stepwise procedure and random forest VIMs can incorporate "large n small N" problem with the consideration of interaction effects when only a relatively small number of input variables are influential. The hypothesis test techniques and the moment-independent VIMs can also be applied to deal with "large n small N" problem even most of the input variables are influential, but both are not good at identifying interactions effects, thus both can be used for identifying important variables but are not suitable for eliminating non-influential variables. The graphic VIMs can incorporate the low order interaction effects, as shown in Ref. [188].
- If the object of analysis is data other than computational models, the available methods are all the statistical techniques, the moment-independent VIMs, the variance-based VIMs computed with RBD or meta-models and the graphic VIMs. The superiority of the delta index over other VIMs is the property of monotonic transformation invariance, which enables us to deal with the problems with large-scale output. The regional VIMs can be especially applicable in the uncertainty reduction setting.
- For learning the model behavior, the available methods are variance-based VIMs and regional VIMs. Both methods have respective advantages and disadvantages. With the variance-based method, every order interaction effect and total effect can be measured, revealing the amount of interaction effects and whether the model is additive. The diagonal lines of the regional VIMs (RMR and RVR functions) show the univariate conditional moments $E(Y|X_i)$ and $V(Y|X_i)$ graphically.
- If the practitioners' intention is to reduce model output uncertainty, then the variance-based VIMs, the moment-independent VIMs (which look at uncertainty reduction the most) and the graphic VIMs (regional and parametric VIMs) can be used. Both the variance-based and moment-independent VIMs measure the amount of the reduction of the model output uncertainty when the true value of each input is learned, thus both can be used for specifying the key uncertainty drivers. After the influential input variables being specified by the moment-independent or variance-based VIMs, the graphic VIMs can then be performed on these influential input variables so as to provide quantitative information on how the model output uncertainty will change w.r.t to the change of model input uncertainties.
- While the factor prioritization is of concern, the regression techniques implemented in stepwise manner, the variance-based VIMs and the moment-independent VIMs can be applied, where the moment-independent VIMs are the only ones to possess the property that their value is null if and only if the model output Y is independent of the model input X_i.
- While the computational model or available data involving multiple types of input variables (such as logic and categorical variables) or missing data (see Ref. [189]), the random forest based VIMs are suggested.

Despite many decades of intensive research and many available methods, there are still many problems left to be solved in VIA. First, large-scale numerical experiments should be carried out to test the effectiveness of each method in different types of problems and compare the relative merits of all these methods. Second, the existing methods should be extended or new VIA

Table 10
Packages and Software for implementing VIMs.

Name	Source	Description
SimLab	Ref. [193]	A free framework for uncertainty analysis and VIM, where the VIM methods include Morris' method, Variance-based method (FAST, extended FAST and Monte Carlo simulation) and multiple linear regression (only includes the regression coefficients and correlation coefficients)
GUI-HDMR	Refs. [194,195]	A software tool for variance-based VIM, where the Sobol's indices are computed with RS-HDMR meta-model
kde, kde2d	Ref. [181]	Matlab package for estimating univariate and bivariate densities from data with nonparametric kernel density estimators. This package can be used for estimating the delta index
sensitivity	Ref. [196]	R package for implementing various VIMs such as parametric regression based VIMs (SRC, SRRC, PCC and PRCC), Morris' screening methods and algorithms for computing the variance-based VIMs (including Monte Carlo procedures, FAST and Kriging-based procedure)
randomForest	Ref. [115]	R package for growing CART-RF, in which GVIM and PVIM are include.
party	Ref.[116]	R package for growing CIT-RF, in which PVIM and CPVIM are included
Random Jungle (RJ)	Ref.[197]	Package for fast implementation of random forest. The package is written with C++ language but can also be used with R program.
randomforest-matlab	Ref. [117]	Matlab version of randomForest

techniques should be developed to solve many specific types of problems. Some of these problems are described as follows.

- For categorical output, the random forest (with classification trees) based VIMs provide a sound strategy. For categorical output with only two groups, the regionalized sensitivity analysis [190] (see also Section 5.2 of Ref. [1]) is a reasonable method. However, for unbalanced data or rare event, both methods are impracticable. The unbalanced data and prediction of the rare event are widespread problems in many disciplines. For example, in structural reliability analysis, the failure of a component or a structural system is certainly a rare event (the probability of event is usually less than 10^{-3}). Then the problem is how to handle VIA in these applications. In Ref. [164], Wei et al. proposed the global reliability sensitivity analysis based on the Sobol's indices to deal with VIA in rare event problem. However, this method is not applicable to high-dimensional problem. Janitza et al. [191] proposed an improved PVIM for dealing with the unbalanced data. However, there is still a long way to go to deal with this kind of problem soundly.
- The random forest based VIMs are currently the most reasonable strategies for high-dimensional problem especially when the variable dimension is higher than the sample size. However, for correlated input variables, none of the three methods in this group can ideally separate the correlated and uncorrelated effects. There is a need to develop VIMs based on random forest to measure the different types of effects.
- The moment-independent VIMs are well posed in the presence of variable correlations. However, as indicates in Fig. 16, when the input variables are correlated, the delta and extended delta indices include both the correlated and uncorrelated contributions, and there is need to discriminate these two types of contributions (see Ref. [192] for an attempt).
- For several skewed and large-scale output, the direct computation of the VIMs often leads to poor accuracy [182]. Although some works have been done (e.g., see [171,182]), there is still a requirement for VIA techniques that can handle this type of problem with relatively low cost especially in high dimension.
- For problem with multiple outputs or time-dependent output, although some feasible strategies are available (e.g., see Refs. [36–39]), we rightly expect more applicable methods.

There are also other types of problems left to be solved in specific applications such as VIA of dynamic system. Here we do not dig deeper into them.

Many packages and software are available for implementing VIA techniques. For ease of application, we summarize these packages or software in Table 10 with details. In the download page of SimLab, many other routines for VIM techniques (e.g., SDR meta-model) are also provided. Bi [22] showed that the program "cforest" in the package "party" for growing CIT-RF does not work for data with small size. We also find that, when using the function "varimp" in the "party" package for VIM, the computation of CPVIM is much more expensive than that of PVIM, and in the case of high dimension (e.g., n is several hundred), it is impractical.

Acknowledgments

This work is supported by National Natural Science Foundation of China (NSFC 51475370) and Excellent Doctorate Foundation of Northwestern Polytechnical University. The authors are thankful to the anonymous reviewers for their valuable comments.

References

[1] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global sensitivity analysis. The primer. Chichester: John Wiley & Sons; 2008.
[2] Helton JC, Hansen CW, Sallaberry CJ. Conceptual structure and computational organization of the 2008 performance assessment for the proposed high-level radioactive waste repository at Yucca mountain, Nevada. Reliab Eng Syst Saf 2014;122:223–48.
[3] Ionescu-Bujor M, Cacuci DG. A comparative review of sensitivity and uncertainty analysis of large-scale systems—I: Deterministic methods. Nucl Sci Eng 2004;147(3):139–203.
[4] Cacuci DG, Ionescu-Bujor A. comparative review of sensitivity and uncertainty analysis of large-scale systems—II: Statistical methods. Nucl Sci Eng 2004;147(3):204–17.
[5] Cacuci DG, Ionescu-Bujor M. A comparative review of sensitivity and uncertainty analysis of large-scale systems—II: Statistical methods. Nucl Sci Eng 2004;147(3):204–17.
[6] Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis for chemical models. Chem Rev 2005;105(7):2811–28.
[7] Borgonovo E. Measuring uncertainty importance: investigation and comparison of alternative approaches. Risk Anal 2006;26(5):1349–62.
[8] Hall JW, Boyce SA, Wang Y, Dawson RJ, Tarantola S, Saltelli A. Sensitivity analysis for hydraulic models. J Hydraul Eng—ASCE 2009;135(11):959–69.
[9] Tian W. A review of sensitivity analysis methods in building energy analysis. Renewable Sustainable Energy Rev 2013;20:411–9.
[10] Borgonovo E. Sensitivity analysis in decision making. Wiley Encyclopedia of Operational Research and Management Science; 2013. p. 1–11.
[11] Helton JC. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Syst Saf 1993;42(2-3):327–67.
[12] Frey HC, Patil SR. Identification and review of sensitivity analysis methods. Risk Anal 2002;22(3):553–78.
[13] Helton JC, Davis FJ. Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Anal 2002;22(2):591–622.
[14] Saltelli A, Marivoet J. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. Reliab Eng Syst Saf 1990;28(2):229–53.
[15] Kleijnen JPC, Helton JC. Statistical analyses of scatterplots to identify important factors in large-scale simulation, I: Review and comparison of techniques. Reliab Eng Syst Saf 1999;65(2):147–85.
[16] Helton JC, Johnson JD, Sallaberry CJ, Storlie CB. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab Eng Syst Saf 2006;91(10-11):1175–209.
[17] Storlie CB, Helton JC. Multiple predictor smoothing methods for sensitivity analysis: description of techniques. Reliab Eng Syst Saf 2008;93(1):28–54.
[18] Storlie CB, Helton JC. Multiple predictor soothing methods for sensitivity analysis: example results. Reliab Eng Syst Saf 2008;93(1):55–77.
[19] Storlie CB, Swiler LP, Helton JC, Sallaberry CJ. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliab Eng Syst Saf 2009;94(11):1735–63.
[20] Storlie CB, Reich BJ, Helton JC, Swiler LP, Sallaberry CJ. Analysis of computationally demanding models with continuous and categorical inputs. Reliab Eng Syst Saf 2013;113(1):30–41.
[21] Johnson JW, Lebreton JM. History and use of relative importance indices on organizational research. Organ Res Methods 2004;7(3):238–57.
[22] Bi J. A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. J Sens Stud 2012;27(2):87–101.
[23] Strobl C, Malley J, Tutz G. An introduction to recursive portioning: rationale, application and characteristics of classification and regression trees, bagging and random forest. Psychol Methods 2009;14(4):323–48.
[24] Boulesteix A-L, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIREs Data Min Knowledge Discovery 2012;2(6):493–507.
[25] Siroky DS. Navigating random forest and related advances in algorithmic modeling. Stat Surv 2009;3:147–63.
[26] Grömping U. Variable importance assessment in regression: linear regression versus random forest. Am Stat 2009;63(4):308–19.
[27] Helton JC, Burmaster DE. Guest Editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems. Reliab Eng Syst Saf 1996;54(2-3):91–4.
[28] Paté-Cornell ME. Uncertainties in risk analysis: six levels of treatment. Reliab Eng Syst Saf 1996;54(2-3):95–111.
[29] Parry GW. The characterization of uncertainty in probabilistic risk assessments of complex systems. Reliab Eng Syst Saf 1996;54(2-3):119–26.
[30] Hora SC. Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. Reliab Eng Syst Saf 1996;54(2-3):217–23.
[31] Kiureghian AD, Ditlevsen O. Aleatory or epistemic? Does it matter? Struct Saf 2009;31(2):105–12.
[32] Helton JC. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. J Stat Comput Simul 1997;57(1-4):3–76.
[33] Helton JC, Sallaberry CJ. Uncertainty and sensitivity analysis: from regulatory requirements to conceptual structure and computational implementation.

IFIP Advances in Information and Communication Technology. AICT 2012:60–76.
[34] Rohatg VK. An introduction to probability theory and mathematical statistics. New York, NY: Wiley; 1976.
[35] Campbell K, KcKay MD, Williams BJ. Sensitivity analysis when model outputs are functions. Reliab Eng Syst Saf 2006;91:1468–72.
[36] Lamboni M, Makowski D, Lehuger, Gabrielle B, Monod H. Multivariate global sensitivity analysis for dynamic crop models. Field Crops Res 2009;113:312–20.
[37] Lamboni M, Monod H, Makowski D. multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliab Eng Syst Saf 2011;96:450–9.
[38] Garcia-Cabrejo O, Valocchi A. Global sensitivity analysis for multivariate output using polynomial chaos expansion. Reliab Eng Syst Saf 2014;126:25–36.
[39] Cao J, Du FR, Ding ST. Global sensitivity analysis for dynamic systems with stochastic input processes. Reliab Eng Syst Saf 2013;118:106–17.
[40] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81 (1):23–69.
[41] Owen AB. Latin supercube sampling for very high-dimensional simulations. ACM Trans Modell Comput Simul 1998;8(1):71–102.
[42] Sobol' IM. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput Math Math Phys 1967;7(4):86–112.
[43] Sobol' IM. Uniformly distributed sequences with an additional uniform property. USSR Comput Math Math Phys 1976;16(5):236–42.
[44] Kucherenko S, Feil B, Shah N, Mauntz W. The identification of model effective dimensions using global sensitivity analysis. Reliab Eng Syst Saf 2011;96(4):440–9.
[45] Tarantola S, Becker W, Zeitz D. A comparison of two sampling methods for global sensitivity analysis. Comput Phys Commun 2012;183:1061–72.
[46] Chalabi Y, Dutang C, Savicky P, Wuertz D. randtoolbox: toolbox for pseudo and quasi random number generation and RNG tests. Available at: ⟨http://cran.r-project.org/web/packages/randtoolbox/index.html⟩; 2013 (accessed 14 January 2014).
[47] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL ⟨http://www.R-project.org⟩; 2010.
[48] Bratley P, Fox BL. Algorithm 659: implementing Sobol's quasi-random sequence generator. ACM Trans Math Software 1988;14:88–100.
[49] Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. Commun Stat-Simul Comput 1982;11 (3):311–34.
[50] Borgonovo E. Sensitivity analysis of model output with input constraints: a generalized rationale for local methods. Risk Anal 2008;28(3):667–80.
[51] Borgonovo E, Apostolakis GE. A new importance measure for risk-informed decision making. Reliab Eng Syst Saf 2001;72(2):193–212.
[52] Borgonovo E. Differential, criticality and Birnbaum importance measures: an application to basic event, groups and SSCs in event trees and binary decision diagrams. Reliab Eng Syst Saf 2007;92(10):1458–67.
[53] Borgonovo E. Sensitivity analysis with finite changes: an application to modified EOQ models. Eur J Oper Res 2010;200(1):127–38.
[54] Griewank A, Walther A. Evaluating derivatives: principles and techniques of algorithmic differentiation. 2nd ed.. Philadelphia: SIAM; 2008.
[55] Dunker AM. Efficient calculation of sensitivity coefficients for complex atmospheric models. Atmos Environ 1981;15(7):1155–61.
[56] Dunker AM. The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. J Chem Phys 1984;81(5):2385–93.
[57] Kramer MA, Calo JM. An improved computational method for sensitivity analysis: green's function method with 'AIM'. Appl Math Modell 1981;5 (6):432–41.
[58] Bartholomew-Biggs M, Brown S, Christianson B, Dixon L. Automatic differentiation of algorithms. J Comput Appl Math 2000;124:171–90.
[59] Cacuci DG. Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. J Math Phys 1981;22(12):2794–802.
[60] Cacuci DG. Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. J Math Phys 1981;22(12):2803–12.
[61] Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33(2):161–74.
[62] Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. Environ Modell Software 2007;22 (10):1509–18.
[63] Ruano MV, Ribes J, Seco A, Ferrer J. An improved sampling strategy design for application of Morris method to systems with many input factors. Environ Modell Software 2012;37(10):103–9.
[64] Jansen MJW. Analysis of variance designs for model output. Comput Phys Commun 1999;117(1):35–43.
[65] Jansen MJW, Rossing WAH, Daamen RA. Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In: Grasman J, Van Straten G, editors. Predictability and nonlinear modeling in natural sciences and economics. Dordrecht: Kluwer; 1994. p. 334–43.
[66] Jansen MJW. Wingding stairs sample analysis program WINDINGS 2.0. Technical report, Private communication.
[67] Chan K, Saltelli A, Tarantola S. Winding stairs: a sampling tool to compute sensitivity indices. Stat Comput 2000;10(3):187–96.
[68] Saltelli A. Making best use of model evaluations to compute sensitivity indices. Comput Phys Commun 2002;145(2):280–97.
[69] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 2010;181(2):259–70.
[70] Campolongo F, Saltelli A, Cariboni J. From screening to quantitative sensitivity analysis. A unified approach. Comput Phys Commun 2011;182 (4):978–88.
[71] Saltelli A, Campolongo F, Cariboni J. Screening important inputs in models with strong interaction properties. Reliab Eng Syst Saf 2009;94(7):1149–55.
[72] Campolongo F, Braddock R. The use of graph theory in the sensitivity analysis of the model output: a second order screening method. Reliab Eng Syst Saf 1999;64(1):1–12.
[73] Cropp RA, Braddock RD. The new Morris method: an efficient second-order screening method. Reliab Eng Syst Saf 2002;78(1):77–83.
[74] Sobol' IM, Kucherenko S. Derivative based global sensitivity measures and their link with global sensitivity indices. Math Comput Simul 2009;79 (10):3009–17.
[75] Sobol' IM, Kucherenko S. A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. Comput Phys Commun 2010;181(7):1212–7.
[76] Kucherenko S, Rodriguez-Fernandez M, Pantelides C, Shah N. Monte Carlo evaluation of derivative-based global sensitivity measures. Reliab Eng Syst Saf 2009;94(7):1135–48.
[77] Lambonia M, Iooss B, Popelin A-L, Gamboa F. Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests. Math Comput Simul 2013;87:45–54.
[78] Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6th ed.. Upper Saddle River, NJ: Pearson Prentice Hall; 2007.
[79] Chatterjee S, Hadi AS. Regression analysis by example. 4th ed.. Hoboken, NJ: John Wiley & Son; 2006.
[80] Helton JC, Davis FJ. Sampling-based methods. In: Saltelli A, Shan K, Scott EM, editors. Sensitivity analysis. New York, NY: Wiley; 2000. p. 101–53.
[81] Xu C, Gertner GZ. Uncertainty and sensitivity analysis for models with correlated parameters. Reliab Eng Syst Saf 2008;93(10):1563–73.
[82] Hao W, Lu Z, Tian L. A novel method for analyzing variance based importance measures of correlated input variables. Acta Aeronaut Astronaut Sin 2011;32(9):1637–43.
[83] Hao W, Lu Z, Wei P, Feng J, Wang B. A new method on ANN for variance importance measure analysis of correlated input variables. Struct Saf 2012;38:56–63.
[84] Iman RL, Conover WJ. The sue of the rank transform in regression. Technometrics 1978;21(4):499–509.
[85] Grömping U. Estimators of relative importance for linear regression based on variance decomposition. Am Stat 2007;61(2):139–47.
[86] Lindeman RH, Merenda PF, Gold RZ. Introduction to bivariate and multivariate analysis. Glenview, IL: Scott, Foresman; 1980.
[87] Kruskal W. Relative importance by averaging over orderings. Am Stat 1987;41(1):6–10.
[88] Kruskal W. Correction to "Relative importance by averaging over orderings". Am Stat 1987;41:341.
[89] Grömping U. Relative importance for linear regression in R: the package relaimpo. J Stat Software 2006;17(1):1–27.
[90] Feldman B. Relative importance and value. Unpublished manuscript, downloadable at ⟨http://www.prismanalytics.com/docs/RelativeImportance.pdf⟩; 2013 (accessed November 25, 2013).
[91] Budescu DV. Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. Psychol Bull 1993;114 (3):542–51.
[92] Azen R, Budescu DV. The dominance analysis approach for comparing predictors in multiple regression. Psychol Methods 2003;8(2):129–48.
[93] Budescu DV, Azen R. Beyond global measures of relative importance: some insights from dominance analysis. Organ Res Methods 2004;7(3):341–50.
[94] Johnson JW. A heuristic method for estimating the relative weight of predictor variables in multiple regression. Multivariate Behav Res 2000;35 (1):1–19.
[95] Zuber V, Strimmer K. High-dimensioanl regression and variable selection using CAR Scores. Stat Appl Genet Mol Biol 2011;10(1) Article 34.
[96] Wood SN. Modelling and smoothing parameter estimation with multiple quadratic penalties. J R Stat Soc B 2000;62:413–28.
[97] Loader C. Local regression and likelihood. New York, NY: Springer; 1999.
[98] Fox J. Nonparametric regression: appendix to an R and S-PLUS companion to applied regression. Encycl Stat Behav Sci 2002.
[99] Fredman JH, Stuetzle W. Projection pursuit regression. J Am Stat Assoc 1981;76(376):817–23.
[100] Specht DF. A general regression neural network. IEEE Trans Neural Networks 1991;2(2):568–76.
[101] Basak D, Pal S, Patranabis DC. Support vector regression. Neural Inf Process—Lett Rev 2007;11(10):203–24.
[102] Clarke SM, Griebsch JH, Simpson TW. Analysis of support vector regression for approximation of complex engineering analyses. J Mech Des 2005;127 (6):1077–87.
[103] Sudret B. Global sensitivity analysis using polynomial chaos expansion. Reliab Eng Syst Saf 2008;93(7):964–79.
[104] Ratto M, Pagano A, Young PC. State dependent parameter metamodelling and sensitivity analysis. Comput Phys Commun 2007;177(11):863–76.

[105] Ratto M, Pagano A, Young PC. Non-parametric estimation of conditional moments for sensitivity analysis. Reliab Eng Syst Saf 2009;94(2):237–43.
[106] Breiman L. Random forest. Mach Learn 2001;45(1):5–32.
[107] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Belmont, CA: Wadsworth; 1984.
[108] Fielding A, O'Muircheartaigh CA. Binary segmentation in survey analysis with particular reference to AID. Statistician 1977;25:17–28.
[109] Quinlan JR. Introduction of decision trees. Mach Learn 1986;1(1):81–106.
[110] Quilan JR. C4.5: programs for machine learning. San Francisco, CA: Morgan Kaufman Publishers Inc; 1993.
[111] Hothorn T, Hornik K, Zeileis A. Unbiased recursive portioning: a conditional inference framework. J Comput Graph Stat 2006;15(3):651–74.
[112] White AP, Liu WZ. Bias in information-based measures in decision tree induction. Mach Learn 1994;15(3):321–9.
[113] Shih Y-S, Tsai H-W. Variable selection bias in regression trees with constant fits. Comput Stat Data Anal 2004;45(3):595–607.
[114] Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinf 2007;8(1):25.
[115] Liaw A, Wiener M. randomForest: Breiman and Cutler's random forests for classification and regression. Available at: ⟨http://cran.r-project.org/web/packages/randomForest/index.html⟩; 2012 (accessed 13 December 2013).
[116] Hothorn T, Hornik K, Strobl C, Zeileis A. party: a laboratory for recursive partytioning. Available at: ⟨http://mirrors.ustc.edu.cn/CRAN/web/packages/party/index.html⟩; 2013 (accessed 13 December 2013).
[117] Jaiantilal A. randomforest-matlab: random forest (regression, classification and clustering) implementation for MATLAB (and Standalone). , Available at: ⟨http://code.google.com/p/randomforest-matlab/⟩; 2010 (accessed 14 December 2013).
[118] Raileanu LE, Stoffel K. Theoretical comparison between the Gini Index and Information Gain criteria. Ann Math Artif Intell 2004;41(1):77–93.
[119] Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forest. BMC Bioinf 2008;9(1):307.
[120] Nicodemus KK, Malley JD, Strobl C, Ziegler A. The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinf 2010;11(1):110.
[121] Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Comput Stat Data Anal 2008;52(4):2249–60.
[122] Boulesteix AL, Slawski M. Stability and aggregation of ranked gene lists. Brief Bioinf 2009;10(5):556–68.
[123] Calle ML, Urrea V. Letter to the editor: stability of random forest importance measures. Brief Bioinf 2010;12(1):86–9.
[124] Nicoswmus KK. Letter to the editor: on the stability and ranking of predictors from random forest variable importance measures. Brief Bioinf 2011;12(4):369–73.
[125] Díaz-Uriarte R, de Andrés SA. Gene selection and classification of microarray data using random forest. BMC Bioinf 2006(1): 7:3.
[126] Goldstein BA, Polley EC, Briggs FBS. Random forests for genetic association studies. Stat Appl Genet Mol Biol 2011;10(1):1–34.
[127] Winham SJ, Colby CL, Freimuth RR, Wang X, de Andrade M, Huebner M, et al. SNP interaction detection with random forests in high-dimensional genetic data. BMC Bioinf 2012;13:164.
[128] Gtanger C, Lin J-L. Using the mutual information coefficient to identify lags in nonlinear models. J Time Ser Anal 1994;15(4):371–84.
[129] Conover WH. Practical nonparametric statistics. 3rd ed.. New York, NY: Wiley; 1999.
[130] Hora SC, Helton JC. A distribution-free test for the relationship between model input and output when using Latin hypercube sampling. Reliab Eng Syst Saf 2003;79(3):333–9.
[131] Winer BI. Statistical principles in experimental design. 2nd ed.. New York, NY: McGraw; 1971.
[132] Peacock JA. Two-dimensional goodness-of-fit testing in astronomy. Mon Not R Astron Soc 1983;202(2):615–27.
[133] Clark PJ, Evans FC. Distance to nearest neighbor as a measure of sparial relationships in populations. Ecology 1954;35:23–30.
[134] Diggle PJ, Cox TF. Some distance-based tests of independence for sparsely sampled multivariate spatial point patterns. Int Stat Rev 1983;51(1):11–23.
[135] Arya S, Mount D, Kemp SE, Jefferis G. RANN: fast neighbor search (wraps Arya and Mount's ANN library). ⟨http://cran.r-project.org/web/packages/RANN0/index.html⟩; 2014 (accessed 22 September 2014).
[136] Sobol' IM. Sensitivity analysis for non-linear mathematical models. Math Modell Comput Exp 1993;1:407–14 Translated from Russian;
Sobol' IM. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie 1990;2:112–8.
[137] Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. Reliab Eng Syst Saf 1996;52(1):1–17.
[138] Cukier RI, Fortuin CM, Shuler KE, Petschek AG, Schaibly JH. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. J Chem Phys 1973;59(8):3873–8.
[139] Cukier RI, Levine HB, Shuler KE. Nonlinear sensitivity analysis of multiparameter model systems. J Comput Phys 1978;26(1):1–42.
[140] Saltelli A, Tarantola S, Chan KPS. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 1999;41(1):39–56.
[141] Tarantola S, Gatelli D, Mara T. Random balance designs for the estimation of first order global sensitivity indices. Reliab Eng Syst Saf 2006;91(6):717–27.
[142] Xu C, Gertner G. Understanding and comparison of different sampling approaches for the Fourier amplitudes sensitivity test (FAST). Comput Stat Data Anal 2011;55(1):184–98.
[143] Mara TA. Extension of the RBD-FAST method to the computation of global sensitivity indices. Reliab Eng Syst Saf 2009;94(8):1274–81.
[144] Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc: Ser B 2004;66(3):751–69.
[145] Buzzard GT, Xiu DB. Variance-based global sensitivity analysis via sparse-grid interpolation and cubature. Commun Comput Phys 2011;9(3):542–67.
[146] Rahman S. Global sensitivity analysis by polynomial dimensional decomposition. Reliab Eng Syst Saf 2011;96(7):825–37.
[147] Li G, Rosenthal C, Rabitz H. High dimensional model representations. J Phys Chem A 2001;105(33):7765–77.
[148] Li G, Wang S-W, Rabitz H. Practical approaches to construct RS-HDMR component functions. J Phys Chem A 2002;106(37):8721–33.
[149] Marseguerra M, Masini R, Zio E, Glacomo C. Variance decomposition-based sensitivity analysis via neural networks. Reliab Eng Syst Saf 2003;79(2):229–38.
[150] Kleijnen JPC. Kriging metamodeling in simulation: a review. Eur J Oper Res 2009;192(1):707–16.
[151] Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 2001;55(1-3):271–80.
[152] Tarantola S, Becker W, Zeitz D. A comparison of two sampling methods for global sensitivity analysis. Comput Phys Commun 2012;183(5):1061–72.
[153] Kucherenko S, Tarantola S, Annoni P. Estimation of global sensitivity indices for models with dependent variables. Comput Phys Commun 2012;183(4):937–46.
[154] Hao W, Lu Z, Tian L. Importance measure of correlated normal variables and its sensitivity analysis. Reliab Eng Syst Saf 2012;99:151–60.
[155] Hao W, Lu Z, Li L. A new interpretation and validation of variance based importance measures for models with correlated inputs. Comput Phys Commun 2013;184(5):1401–13.
[156] Li G, Rabitz H, Yelvington PE, Oluwole O, Bacon F, Kolb CE, Schoendorf J. Global sensitivity analysis for systems with independent and/or correlated inputs. J Phys Chem A 2010;114(19):6022–32.
[157] Mara TA, Tarantola S. Variance-based sensitivity indices for models with dependent inputs. Reliab Eng Syst Saf 2012;107:115–21.
[158] Xu C. Decoupling correlated and uncorrelated parameter uncertainty contributions for nonlinear models. Appl Math Modell 2013;37(24):9950–69.
[159] Xu C, Gertner GZ. Extending a global sensitivity analysis technique to models with correlated parameter. Comput Stat Data Anal 2007;51(12):5579–90.
[160] Xu C, Gertner GZ. A general first-order global sensitivity analysis method. Reliab Eng Syst Saf 2008;93(7):1060–71.
[161] Most T. Variance-based sensitivity analysis in the presence of correlated input variables. In: Fifth international conference on reliable engineering computing (REC), Brno, Czech Republic; 2012.
[162] Zhou C, Lu Z, Li L, Feng J, Wang B. A new algorithm for variance based importance analysis of models with correlated inputs. Appl Math Modell 2013;37(3):864–75.
[163] Li L, Lu Z, Zhou C. Importance analysis for models with correlated input variables by the state dependent parameter method. Comput Math Appl 2011;62(12):4547–56.
[164] Wei P, Lu Z, Hao W, Feng J, Wang B. Efficient sampling methods for global reliability sensitivity analysis. Comput Phys Commun 2012;183(8):1728–43.
[165] Sobol' IM, Tarantola S, Gatelli D, Kucherenko S, Mauntz W. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliab Eng Syst Saf 2007;92(7):957–60.
[166] Allairea DL, Willcox KE. Distributional sensitivity analysis. Procedia-Soc Behav Sci 2010;2(6):7595–6.
[167] Wei P, Lu Z, Song J. A new variance-based global sensitivity analysis technique. Comput Phys Commun 2013;184(11):2540–51.
[168] Park CK, Ahn KI. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. Reliab Eng Syst Saf 1994;46(3):253–61.
[169] Chun M-H, Han S-J, Tak N-I. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. Reliab Eng Syst Saf 2000;70(3):313–21.
[170] Tang Z, Lu Z, Jiang B, Wang P, Zhang F. Entropy-based importance measure for uncertainty model inputs. AIAA J 2013;51(10):2319–34.
[171] Baucells M, Borgonovo E. Invariant probabilistic sensitivity analysis. Manage Sci 2013;59(11):2536–49.
[172] Borgonovo E. A new uncertainty importance measure. Reliab Eng Syst Saf 2007;92(6):771–84.
[173] Wei P, Lu Z, Song J. Moment-independent sensitivity analysis using copula. Risk Anal 2014;34(2):210–22.
[174] Wei P, Lu Z, Yuan X. Monte Carlo simulation for moment-independent sensitivity analysis. Reliab Eng Syst Saf 2013;110:60–7.
[175] Plischke E, Borgonovo E, Smith CL. Global sensitivity measures from given data. Eur J Oper Res 2012;226(3):536–50.
[176] Nelsen RB. An Introduction to Copulas. 2nd ed.. New York, NY: Springer; 2006.
[177] Genest C, Favre A-C. Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng 2007;12(4):347–68.
[178] Wolff EF. N-dimensional measures of dependence. Stochastica 1980;4(3):175–88.

[179] Schweizer B, Wolff EF. On nonparametric measures of dependence for random variables. Ann Stat 1981;9:879–85.
[180] Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Stat 2010;38(5):2916–57.
[181] Botev ZI. Kernel density estimation using Matlab. Available at ⟨http://www.mathworks.us/matlabcentral/fileexchange/authors/27236⟩; 2012 (accessed April 5, 2012).
[182] Borgonovo E, Tarantola S, Plischke E, Morris MD. Transformations and invariance in the sensitivity analysis of computer experiments. J R Stat Soc B 2014;76(5):925–47.
[183] Sinclair J. Response to the PSACOIN Level S exercise. PSACOIN Level S intercomparison. Nucl Energy Agency. Organ Econ Cooperation Dev 1993.
[184] Bolado-Lavin R, Castings W, Tarantola S. Contribution to the sample mean plot for graphical and numerical sensitivity analysis. Reliab Eng Syst Saf 2009;94(6):1041–9.
[185] Tarantola S, Kopustinskas V, Bolado-Lavin R, Kaliatka A, Uščuras E, Vaišnoras M. Sensitivity analysis using contribution to sample variance plot: application to a water hammer model. Reliab Eng Syst Saf 2012;99:62–73.
[186] Wei P, Lu Z, Wu D, Zhou C. Moment-independent regional sensitivity analysis: application to an environmental model. Environ Modell Software 2013;47:55–63.
[187] Wei P, Lu Z, Ruan W, Song J. Regional sensitivity analysis using revised mean and variance ratio functions. Reliab Eng Syst Saf 2014;121:121–35.
[188] Wei P, Lu Z, Song J. Uncertainty importance analysis using parametric moment ratio function. Risk Anal 2014;34(2):223–34.
[189] Hapfelmeier A, Hothorn T, Ulm K, Strobl C. A new variable importance measure for random forests with missing data. Stat Comput 2014;24(6):21–34.
[190] Young PC. Data-based mechanistic modeling, generalized sensitivity and dominant model analysis. Comput Phys Commun 1999;117(1):113–29.
[191] Janitza S, Strobl C, Boulesteix A-L. An AUC-based permutation variable importance measure for random forests. BMC Bioinf 2013;14(1):19.
[192] Zhou CC, Lu ZZ, Zhang LG, Hu JX. Moment-independent sensitivity analysis with corelations. Appl Math Modell 2014;38(19-20):4885–96.
[193] Joint Research Centre of European Commission. Simlab: a free development framework for sensitivity and uncertainty analysis. Available at: ⟨http://ipsc.jrc.ec.europa.eu/?id=756⟩; 2013 (accessed 15 December 2013).
[194] Ziehn T, Tomlin AS. GUI-HDMR—a software tool for global sensitivity analysis of complex models. Environ Modell Software 2009;24(7):775–85.
[195] Ziehn T, Tomlin A. GUI-HDMR: a software tool for global sensitivity analysis. Available at: ⟨http://gui-hdmr.de/⟩; 2011 by contacting Tilo Ziehn or Alison Tomlin (accessed 21 November 2011).
[196] Pujol G, Iooss B, Janon A. sensitivity: a collection of functions for factor screening, global sensitivity analysis and reliability sensitivity analysis of model output. Available at: ⟨http://cran.r-project.org/web/packages/sensitivity/index.html⟩; 2013 (accessed 22 December 2013).
[197] Schwarz DF, König IR, Ziegler A. On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. accessed 15 March 2014. Bioinformatics 2010;26:1752–8 Package available at: ⟨http://imbs-luebeck.de/imbs/de/node/227⟩.