More powerful HSIC-based independence tests, extension to space-filling designs and functional data

Mohamed Reda El Amri, Amandine Marrel

More powerful HSIC-based independence tests, extension to space-filling designs and functional data

Mohamed Reda El Amri[a,*], Amandine Marrel[a,b]

[a] *CEA, DES, DER, F-13108 Saint-Paul-Lez-Durance, France*
[b] *Institut de Mathématiques de Toulouse, 31062, Toulouse, France*

Abstract

The Hilbert-Schmidt Independence Criterion (HSIC) is a dependence measure based on reproducing kernel Hilbert spaces. This measure can be used for the global sensitivity analysis of numerical simulators whose objective is to identify the most influential inputs on the output(s) of the code. For this purpose, HSIC-based sensitivity measures and independence tests can be used for the ranking and the screening of inputs, respectively. In this framework, our work proposes several improvements in the use of HSIC to increase their application spectrum and make the associated independence tests more powerful. First, we introduce a new method to perform the tests in a non-asymptotic framework. This method is much less CPU-time expensive than the one based on permutation, while remaining as efficient. Then, the use of HSIC-based independence tests is extended to the case of some space-filling designs, where the *independent and identically distributed* condition of the observations is lifted. For this, a new procedure based on conditional randomization test is used. In addition, we also propose a more powerful test that relies on a well-chosen parameterization of the HSIC statistics: the kernel bandwidth parameter is optimized instead of the standard choices. Numerical studies are performed to assess the efficiency of these procedures and compare it to existing tests in the literature. Finally, HSIC-based indices for functional outputs are defined: they rely on appropriate and relevant kernels for this type of data. Illustrations are provided on temporal outputs of an analytical function and a compartmental epidemiological model.

Keywords: Hilbert Schmidt Independence Criterion, Space-Filling Designs, Conditional Randomization Test, Pearson approximation, Bandwidth selection, Functional random variables.

1. Introduction

As part of safety studies for nuclear reactors, numerical simulators are fundamental tools for understanding, modelling and predicting physical phenomena. These tools can take a large number of input parameters, characterizing the studied phenomenon or related to its physical and numerical modelling. The information related to some of these parameters is often limited or uncertain, this can be due to the lack or absence of data, measurement or modelling errors, or a natural variability of the parameters. These input parameters, and consequently the simulator output, are thus uncertain. This is referred to as uncertainty propagation. In this framework, Sensitivity Analysis (SA) methods aim at determining how the variability of model's inputs affects the variance of its output. Usually, two main fields are distinguished: the Local SA (LSA) and the Global SA (GSA). The first one studies the output variation for small input shifts near their reference values (also called nominal values) while the global approach assesses and quantifies the global impact of each input uncertainty on the output. GSA therefore requires to characterize the input uncertainties over their variation range, for example by assigning a probability distribution to the input vector. Within this framework, the statistical methods for GSA are mostly based on Monte Carlo simulations of the model, i.e. on a random sampling of inputs according to their probability distributions. An approach conventionally used for the GSA is based on the decomposition of the output variance, where each term of the decomposition represents the part of the contribution of an input (or a group of inputs) to the output variance. Originally introduced in [Hoe48], this decomposition is commonly called the ANOVA decomposition (for *ANalysis Of VAriance*). Sensitivity indices are directly

*Corresponding author
Email address: mohamed-reda.elamri@cea.fr (Mohamed Reda El Amri)

derived from this decomposition: these are the Sobol' indices [Sob93], mentioned above. Sobol's indices are easily interpretable, but their expressions involve multidimensional integrals whose estimation by Monte-Carlo methods requires in practice a very large number of model simulations (several tens of thousands). Their direct estimation is therefore very often impossible for time-consuming simulators. Even if very recent work makes it possible to significantly reduce this cost [GGKL20], the fact remains that Sobol' indices only consider the variance of the output. They do not evaluate the impact of each input on the whole probability distribution of the output. The nullity of a Sobol' index is not equivalent to the independence between the output and an input.

The dependence measures introduced for GSA by [DV15], make it possible to overcome several of the limitations listed above. First, these measures quantify from a probabilistic point of view the dependence between each input and output. Thus, under some parametric choices, the nullity of a dependence measure between an input and the output is equivalent to the independence of these two random variables. These measures can be used quantitatively to rank the inputs by order of influence on the output, as well as qualitatively to perform a screening. For this, statistical independence tests built upon dependence measures can be used [DLM16]. These tests provides a more rigorous and accurate statistical and mathematical framework than a simple assessment and comparison of sensitivity measure values. In particular, this avoids the arbitrary choice of a threshold value for sensitivity measures, beyond which an input variable is considered as influential. Among the existing dependence measures, the Hilbert-Schmidt Independence Criterion denoted HSIC [GBSS05] has many advantages. HSIC generalizes the notion of covariance between two random variables and and thus makes it possible to capture a very wide spectrum of forms of dependence between the variables. Moreover, HSIC suffers less from the *Curse of dimensionality* in its estimation and can be applied to random variables of various types: scalar, vector, categorical, etc. For this reason, [DV15] and then [DLM16] investigated the use of HSIC for GSA, compared them to Sobol' indices and showed their low estimation cost in practice to obtain reliable results (in practice a few hundred simulations compared to several tens of thousands for Sobol' indices). In addition, many recent works deal with asymptotic [GFT$^+$08] but also non-asymptotic independence HSIC-based tests [DLM16, Mey19, EAM]. For all these reasons, we focus here on HSIC-based independence tests for GSA of numerical simulators. Our work aims to propose several improvements for the use of HSIC-based sensitivity measures and independence test.

The main contributions of this work are as follows:

- First, new method to perform HSIC-based independence test in a non-asymptotic framework is proposed. More precisely, the HSIC-based SA is built upon a n-size Monte Carlo sample of inputs and output of the studied model. The HSIC independence test then relies on the estimation of the p-value under independence hypothesis. There are several existing methods for approximating the p-values (namely, permutation-based method or asymptotic Gamma approximation, e.g.). Beside these two main approaches, we introduce a new one based on a Pearson type III approximation. This yields results as efficient as the permutation-based method (regardless the sample size) while requiring a computation time almost as negligible as the Gamma approximation.

- Second, we introduce a new algorithm based on the Conditional Randomization Test (CRT). The CRT is dedicated to correct the HSIC-based independence tests built upon Space-Filling Designs (SFD). Indeed, the counterpart to the use of SFD rather than a simple Monte Carlo sampling is that the simulations are no longer *independent and identically distributed* (i.i.d.). Yet, this i.i.d. assumption of the realizations is required for the application of fundamental statistics theorems such as central limit theorem, which is required to demonstrate the convergence of statistical estimators, build confidence intervals or even implement statistical tests. Through different examples, we show the efficiency and the robustness of the CRT for most SFD.

- Then, the HSIC-based SA mainly depends on the kernel bandwidth parameters. To avoid a heuristic choice of these parameters, we propose a fast and efficient bandwidth selection. To achieve this, we derive a closed-form expression for the gradient of the HSIC with respect to the bandwidth parameters, which aims at facilitating its maximization using gradient-based ascent algorithms. Subsequently, a new HSIC based independence test is proposed accordingly. The proposed method is evaluated on various analytical examples, and we have shown that our method increases the statistical test accuracy.

- Finally, we extend the use of the HSIC-based SA to functional outputs. As previously mentioned, the

HSIC can be applied to scalar, vector or even categorical random variables. This holds true as the notion of distance are well established in such cases. In the case of functional random variables, the kernel literature has already proposed several ways to handle curves or images for regression or classification purposes. For instance, the PCA-kernel [CFS99] can be used, as illustrated in [DV15]. [CVBM07] proposes the Global Alignment Kernel (GAK) which operates on time series, and it is based on the well-known Dynamic Time Warping [SC78] family of distances. The GAK has a quadratic time and space complexity that limits its use to only small time series data sets, and thus this kernel will not be considered. In this work, we compare various definitions of kernels, all based on dimension reduction technique and different weighting functions.

The remainder of the article is structured as follows. In the next section, we define the HSIC measures and their use to test independence. A novel p-value's approximation method for HSIC-tests and an extension of the use of the HSIC to the case of SFD samples are presented in Section 3. Section 4 describes a new and more powerful HSIC-based independence test based on the optimal bandwidth parameters. In section 5, we deal with the extension of HSIC to functional outputs. Finally, general conclusions on this work and future research topics are drawn in Section 6.

2. HSIC: general principle and use for testing independence

In this section, we describe the Hilbert Schmidt Independence Criterion (HSIC) and its associated SA measures which are at the center of all the developments proposed in this work.

2.1. Definition and estimation

HSIC are built upon kernel-based approaches for detecting dependence, and more particularly on cross-covariance operators in reproducing kernel Hilbert spaces (RKHS) composed of mapping functions (features) and characterized by positive definite kernel function. This amounts to considering covariance between feature functions applied to the two variables (here X_i and Y). This set of functions (possibly non-linear), which is defined by the space and the kernel, can be of infinite dimension and allow to capture a very broad spectrum of forms of dependency. The HSIC is then defined as the squared Hilbert-Schmidt norm of the cross-covariance operator and somehow "summarizes" the set of covariances between features:

$$||C_{X_iY}||^2_{HS} = \sum_{l,m} |\mathbb{COV}(u_l(X_i), v_m(Y))|^2,$$

where $(u_l)_{l\geq 0}$ and $(v_m)_{m\geq 0}$ are orthonormal bases of the RKHS associated X_i and Y respectively. The kernel trick allows to get rid of an explicit expression of the feature map and consequently of $(u_l)_{l\geq 0}$ and $(v_m)_{m\geq 0}$: HSIC can be directly expressed with kernels and estimated in a very simple and low cost way. If we note the kernels k_i and k of the RKHS associated to X_i and Y respectively, one can show that the HSIC measure can be written using expected values of kernels:

$$\begin{aligned}\mathrm{HSIC}(X_i, Y) =& \mathbb{E}[k_i(X_i, X_i')k(Y, Y')] + \mathbb{E}[k_i(X_i, X_i')]\mathbb{E}[k(Y, Y')] \\ &- 2\mathbb{E}[\mathbb{E}[k_i(X_i, X_i')|X_i]\mathbb{E}[k(Y, Y')|Y]],\end{aligned} \tag{1}$$

where (X_i', Y') is an independent and identically distributed (i.i.d.) copy of (X_i, Y).

Concerning the choice of kernel, characteristic kernels (meaning that the feature map is injective) allow to fully characterize the independence with HSIC:

$$\mathrm{HSIC}(X_i, Y) = 0 \Longleftrightarrow X_i \perp\!\!\!\perp Y, \tag{2}$$

$\perp\!\!\!\perp$ meaning that X_i and Y are independent. For GSA purposes, the nullity of the measure indicates that X_i does not influence Y. The characteristic property of the kernel is crucial for testing dependence. More precisely, we will use here the characteristic Gaussian kernel which is defined, for real variables $(\mathbf{z}, \mathbf{z}') \in \mathbb{R}^q \times \mathbb{R}^q$ as:

$$\ell_\lambda(\mathbf{z}, \mathbf{z}') = \exp(-\frac{||\mathbf{z} - \mathbf{z}'||_2^2}{2\lambda^2}),$$

where λ is the *bandwidth parameter* and $||.||_2$ is the Euclidean norm in $\mathbb{R}^q$. Usually, for real scalar variable Z, one uses in practice $\lambda = \sigma_z$ with σ_z being the empirical standard deviation of the sample of Z. The optimal choice of λ will be discussed in Section 4.

To conclude on HSIC, it remains to deal with its estimation in practice. From a n-sample $(X_i^{(m)}, Y^{(m)})_{1\leq m\leq n}$ of (X_i, Y), an estimator of the measure HSIC is proposed in [GBSS05]:

$$\widehat{\mathrm{HSIC}}(X_i, Y) = \frac{1}{n^2}\mathrm{Tr}(L_i H L H) = \frac{1}{n^2}\mathrm{Tr}(H L_i H L) = \frac{1}{n^2}\mathrm{Tr}(L_i^{db} L^{db}), \tag{3}$$

where the Gram matrices L_i and L are defined by $L_i = (k_i(X_i^{(l)}, X_i^{(m)}))_{1\leq l,m\leq n}$ and $L = (k(Y^{(l)}, Y^{(m)}))_{1\leq l,m\leq n}$, and $H = (\delta_{lm} - 1/n)_{1\leq l,m\leq n}$, with δ_{lm} is the Kronecker operator. L_i^{db} and L^{db} are the corresponding double centered matrices.

For GSA and ranking purpose, a normalization of the HSIC has been proposed by [DV15] to provide a sensitivity index included in the range $[0, 1]$ which makes its interpretation easier:

$$\mathrm{R}^2_{\mathrm{HSIC},i} = \frac{\mathrm{HSIC}(X_i, Y)}{\sqrt{\mathrm{HSIC}(X_i, X_i)\mathrm{HSIC}(Y, Y)}}. \tag{4}$$

In practice, $\mathrm{R}^2_{\mathrm{HSIC},i}$ can be estimated using a plug-in approach by:

$$\widehat{\mathrm{R}}^2_{\mathrm{HSIC},i} = \frac{\widehat{\mathrm{HSIC}}(X_i, Y)}{\sqrt{\widehat{\mathrm{HSIC}}(X_i, X_i)\widehat{\mathrm{HSIC}}(Y, Y)}}. \tag{5}$$

2.2. Statistical independence tests built upon HSIC

In a screening context, the objective is to separate the input parameters into two sub-groups, the significant ones and the non-significant ones. For this, statistical hypothesis tests can be built upon HSIC. For a given input X_i, it aims at testing the null hypothesis "$(\mathcal{H}_0^i)$: X_i and Y are independent", against, its alternative "$(\mathcal{H}_1^i)$: X_i and Y are dependent". Since the nullity of HSIC (with characteristic kernels) is equivalent to the independence of variables, testing independence is equivalent to test:

$$(\mathcal{H}_0^i) : \mathrm{HSIC}(X_i, Y) = 0 \qquad \text{versus} \qquad (\mathcal{H}_1^i) : \mathrm{HSIC}(X_i, Y) > 0.$$

The estimator $\widehat{\mathcal{T}}_i = n\widehat{\mathrm{HSIC}}(X_i, Y)$ is then a relevant statistic for this test, with a rejection region defined by the values of $\widehat{\mathcal{T}}_i$ significantly greater than 0. Remember that a statistical independence test is a decision-making procedure between two hypotheses and, according to the available sample, this statistical decision has a non-zero probability of being false. The quality of a statistical test is naturally determined by the risks of misjudging dependence and independence between X_i and Y. The probability of being wrong under the null hypothesis $(\mathcal{H}_0^i)$ is generally called *first-kind error* or *level of test* and denoted α. The probability that the test is wrong under the alternative hypothesis is called *second-kind error* and typically known as β. *The power of hypothesis test* is the probability of rejecting the null hypothesis when, in fact, it is false. Simply put, power is the probability of not making a second-kind error and mathematically it is equal to $1 - \beta$. The closer the power is to 1, the greater the capacity of detecting a false null hypothesis is. Theoretical and practical control of the test level α is possible and generally set at a threshold of $\alpha = 5\%$ or 10%. By contrast, there is currently no theoretical or practical control of β.

The test can be defined in an equivalent way using the *p-value*[1] which is defined as the probability that, under $(\mathcal{H}_0^i)$, the test statistic (in this case, $\widehat{\mathcal{T}}_i$) is greater than or equal to the observed value on the data $\widehat{\mathcal{T}}_{i,obs} = n \times \widehat{\mathrm{HSIC}}(X_i, Y)_{obs}$. Therefore, the p-value is defined by:

$$p_{val,i} = \mathbb{P}_{(\mathcal{H}_0^i)}(\widehat{\mathcal{T}}_i \geq \widehat{\mathcal{T}}_{i,obs}).$$

[1]More generally, the p-value is defined as the largest probability, under the null hypothesis, to have observed a value of the statistics at least as extreme as the value actually observed. The p-value can therefore be viewed as an indicator of the likelihood of the observed statistics under the null hypothesis.

The independence test rejects the null hypothesis ($\mathcal{H}_0^i$) if $p_{val,i} < \alpha$, which means here that the input X_i is significantly influential.

Thus, to compute this p-value, it is necessary to know the cumulative distribution function of the statistic $\widehat{\mathcal{T}}_i$ under ($\mathcal{H}_0^i$). Unfortunately, this function is analytically unknown and approximation methods must be used, as described in the following.

2.3. Methods for approximating the p-values in HSIC-based independence tests

In this section, we discuss different ways to approximate the distribution of HSIC-based statistics $\widehat{\mathcal{T}}_i$ under independence hypothesis ($\mathcal{H}_0$).

2.3.1. Asymptotic approximation for large n

Under asymptotic convergence (i.e., when n is sufficiently large), [GFT$^+$08] have proved that the asymptotic law of $\widehat{\mathcal{T}}_i$ converges in distribution to $\sum_{l\geq 0} \tau_l Z_l^2$ where the standard normal variables Z_l are independent and where the coefficients τ_l are the solutions of a eigenvalue problem (for details, see [GFT$^+$08]). Since the distribution of an infinite weighted sum of independent chi-squared variables can be approached by a Gamma distribution, it follows that the distribution of $\widehat{\mathcal{T}}_i$ can be asymptotically approached by a Gamma distribution whose parameters are:

$$\hat{\gamma} = \frac{E^2}{V} \text{ and } \hat{\beta} = \frac{nV}{E}, \tag{6}$$

where

- $E = n^{-1}(1 + E_x E_y - E_x - E_y)$,
- $E_x = \frac{1}{n-1}\sum_{l\neq h}(L_i)_{lh}$,
- $E_y = \frac{1}{n-1}\sum_{l\neq h}(L)_{lh}$,
- $V = \frac{(n-4)(n-5)}{n(n-1)(n-2)(n-3)}\mathbf{1}^\top(B - \text{diag}(B))\mathbf{1}$, where $B = ((HL_iH)\odot(HLH))^{\cdot 2}$, $\odot$ is the element-wise multiplication and $(M)^{\cdot 2}$ is the element-wise matrix power for all $M \in \mathcal{M}_n(\mathbb{R})$.

2.3.2. Permutation-based method for smaller n

Outside the asymptotic framework, i.e., when n is rather small (e.g., lower than thousand), the Gamma approximation cannot be used anymore. In this case, non-asymptotic versions of the test ([DLM16, Mey19]) based on permutation method offer a suitable and relevant alternative. Successfully used in a wide range of applications, permutation-based tests require few assumptions concerning statistical distributions (just exchangeability) and are very flexible. The permutation principle is very intuitive and consists in simulating the null hypothesis by breaking the structure inputs/output (i.e., the potential relationship) by permuting the rows of $(X_i^{(m)})_{m=1}^n$ or $(Y^{(m)})_{m=1}^n$. More precisely, the rows of one vector (for example Y) are permuted (the observations of Y are labeled at random) and the HSIC value is computed. This operation is repeated for each possible permutation ($B = n!$ permutations). A sample of the HSIC statistic under ($\mathcal{H}_0$) is thereby simulated. A permutation-based test can be derived: the p-value is computed as the proportion of the values (including the observed one) more extreme than or equal to the observed one

2.4. Limitations

When the sample size n is sufficiently small (smaller than 9 for example), the exact distribution of permuted HSIC can be obtained (enumerating all possible arrangements; $9! = 362880$ permutations are computationally achievable) and an exact test can be performed. However when n becomes larger, it is not possible to perform all the permutations in terms of computational cost. Thus, one way to approximate the distribution of permuted HSIC is to randomly sample a reduced number of permutations B_{rand} among all possible permutations. However, the estimation of the p-value mainly depends on the size of this random sample B_{rand}. One solution can consist in considering systematically a very large value of B_{rand}, e.g. $B_{rand} = 5000$, to ensure convergence of the estimated p-value. To avoid an a priori choice of B_{rand} and optimize its value to reduce the computational cost of the procedure, three algorithms are proposed in [EAM] for sequentially estimating the p-values, according to the GSA purpose. As an alternative, we show in Section 3.1 that the distribution of the permuted HSIC (i.e. if all the $n!$ permutations were used) is a generalized Gamma distribution. We also provide analytical formulas

to compute its first 3 moments. Through numerical experiments, we show that the proposed method is more suitable for low-size sample.

Another limitation of current HSIC tests concerns their applications to samples where the observations are no longer independent and identically distributed (i.i.d.). Space-Filling Designs (SFD) are part of them. Indeed, the i.i.d. condition is required to demonstrate the convergence of HSIC estimators as well as yo compute p-value of HSIC test. In the next section, we highlight the errors induced by a direct application of the HSIC-based tests on a SFD sample. To circumvent this, we propose a corrected HSIC-based tests

3. Extension to low-size sample and current Space-Filling Designs

3.1. Generalized Gamma approximation

3.1.1. Definition

In this section, we propose a new method for approximating p-value in HSIC test. This approach consists of approximating the permutation distribution by a continuous distribution without doing any permutation and using the first moment of the exact permutation distribution under the null hypothesis. In [KAHSL95], the authors have demonstrated analytical formulas of the first three moments for any trace statistic of general form $\mathcal{S} = \mathrm{Tr}(\widetilde{L_i}\widetilde{L})$, where $\widetilde{L_i}$ and $\widetilde{L}$ are any square, symmetric and centered kernel matrices. With these first three moments, [JPH08] compared different moment based methods such as the normal, the log-transformation and Pearson family approximation and pointed out the quality of the Pearson type III approximation for permutation distributions. In our case, the statistic under consideration respects the general form of $\mathcal{S}$ (see Eq. (3)). Thus a direct application of the existing formulas gives us the first three moments of the HSIC-statistic under the permutation distribution. In particular, the first two moments are given by:

$$\begin{aligned}
\mathbb{E}_{\mathtt{perm}}(\mathcal{S}) &= \frac{T}{n-1}, \\
\mathbb{VAR}_{\mathtt{perm}}(\mathcal{S}) &= \mathbb{E}_{\mathtt{perm}}(\mathcal{S}^2) - \mathbb{E}_{\mathtt{perm}}(\mathcal{S})^2 \\
&= \frac{A}{n} + \frac{(T^2 - A) + 2(T_2 - A) + 4A}{n(n-1)} + \frac{2T_2 - 6A + T^2}{n(n-1)(n-2)(n-3)} \\
&\quad + \frac{4(2A - T_2) + 2(2A - T^2)}{n(n-1)(n-2)},
\end{aligned} \tag{7}$$

where

- $T = \mathrm{Tr}(\widetilde{L_i})\mathrm{Tr}(\widetilde{L})$,
- $T_2 = \sum_{lh}(\widetilde{L_i})^2_{lh} \sum_{lh}(\widetilde{L})^2_{lh}$,
- $A = \sum_{l}(\widetilde{L_i})^2_{ll} \sum_{l}(\widetilde{L})^2_{ll}$.

The expression of the skewness (the third moment) $\gamma = \gamma_{\mathtt{perm}}(\mathcal{S})$ is detailed in [KAHSL95]. Then, the standardized variable $\mathcal{S}_s = (\mathcal{S} - \mathbb{E}_{\mathtt{perm}}[\mathcal{S}])/\sqrt{\mathbb{VAR}_{\mathtt{perm}}(\mathcal{S})}$ is assumed to follow a Pearson Type III under the null hypothesis.

3.1.2. Numerical comparison of the different approximations of p-values

Here we compare the above presented methods for estimating the p-values and performing the HSIC-based statistical test. For this, we consider two independent standardized Gaussian variables $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ and apply them the 3 following transformations:

- The identity transformation, $z \mapsto z$.
- The square root transformation, $z \mapsto sgn(z)|z|^{1/2}$.
- The cubic square root transformation, $z \mapsto sgn(z)|z|^{3/2}$,

where the sign function, sgn, is defined as $sgn(z) := \{-1 \text{ if } z < 0,\ 0 \text{ if } z = 0 \text{ and } 1 \text{ if } z > 0\}$. The performances of the approximations are assessed for different sample sizes n. For all the tests, we consider the Gaussian kernel with bandwidth set at the inverse of the empirical variance. Note that for the permutation-based method, we use $B = 3000$ permutations. For each sample size, 1000 repetitions were carried out by generating independent samples of (Z_1, Z_2) under $(\mathcal{H}_0)$. Figure 1 gives the percentage (over the 1000 repetitions) of null hypotheses rejected for each method (the expected value is equal to $\alpha = 5\%$). Two observations can be made:

- the three methods give very close results for large n (> 100);
- permutation and Pearson Type III based methods give very similar results for smaller n;
- the error rates for the permutation-based method and the Pearson type III approximation are all close to 5% and provide a far more accurate approximation than the Gamma approximation.

For all the considered transformations, the Pearson approximation provides a very reliable of the nominal level (0.05) and just as accurate as the permutation-based method even for low sample sizes n. Same observations have been done on other analytical examples. On the other hand, the Gamma approximation becomes suitable only since larger n (here $n \geq 50$).

In the following, we study the impact of the approximation method on the distribution of the p-value. More precisely, suppose that the test statistic $\widehat{\mathcal{S}}_i$ has the cumulative distribution function F under the null hypothesis. Assuming that F is invertible, we can derive the distribution of the random p-value $P_{val,i} = F(\widehat{\mathcal{S}}_i)$ as follows:

$$\begin{aligned}\mathbb{P}(P_{val,i} < p) &= \mathbb{P}(F(\widehat{\mathcal{S}}_i) < p),\\ &= \mathbb{P}(\widehat{\mathcal{S}}_i < F^{-1}(p)) = F(F^{-1}(p)) = p,\end{aligned}$$

from which we can conclude that the distribution of $P_{val,i}$, under $(\mathcal{H}_0)$, is uniform on [0,1]. A simple simulation of the distribution of $P_{val,i}$ can be obtained by Algorithm 1. In the following, the parameter L is set to 1000 to obtain the empirical CDF of the p-value.

Algorithm 1 – empirical Cumulative Distribution Function (eCDF) of the p-value.

Require: Sample size n and L
1: **for** $l = 1$ to L **do**
2: Generate D_n a sample of n realizations of the variables Z_1 and Z_2
3: apply one of the transformations to the sample D_n
4: Estimate the p-value $p_{val}^{(l)}$ by one of the three strategies previously presented
5: **end for**
6: Discrete distribution of the p-value P_{val} over the set $\{p_{val}^{(1)}, \ldots, p_{val}^{(L)}\}$

Figure 2 shows the empirical Cumulative Distribution Function (eCDF) of the p-values under independence hypothesis. These eCDF are obtained either by permutation-based method, Pearson type III or Gamma approximation. The analysis of the convergence of the eCDF according to n yealds:

- regardless of the sample size n and the transformation function, the permutation-based method provides an uniformly distributed p-value. In other words, the error rate is almost equal to the pre-chosen nominal level α as demonstrated in [Mey19];
- regardless the transformation function, the Pearson type III approximation gives an error rate close to the nominal level if $\alpha \leq 0.2$ for $n = 10$ and $n = 30$ for larger α. One note that this is a satisfying result as the level is set the most of the time at 0.05 or 0.01 ($\alpha \leq 0.1$);
- regarding the Gamma approximation, the more we increase the sample size n (moving towards an asymptotic framework), the closer the error rate is to the nominal level.

To conclude this comparison, Figure 3 shows the logarithmic CPU time required to compute the p-value as a function of sample size n for the three methods. These approaches are implemented in the software R and all computations are performed on a Intel 2.30 Ghz processor. From this figure, the running time of the

Gamma and Pearson Type III methods grows much more slowly than the permutation-based method. Indeed, the discrete distribution of the HSIC under independence hypothesis in the permutation-based method involves an empirical estimation of the HSIC (which require at least $\mathcal{O}(n^2)$) for each permutation, thus the algorithm has complexity $\mathcal{O}(B \times n^2)$. The latter becomes time consuming as n and B increase comparing to the Gamma or Pearson type III approximations which has complexity $\mathcal{O}(2 \times n^2)$ and $\mathcal{O}(4 \times n^2)$, respectively.

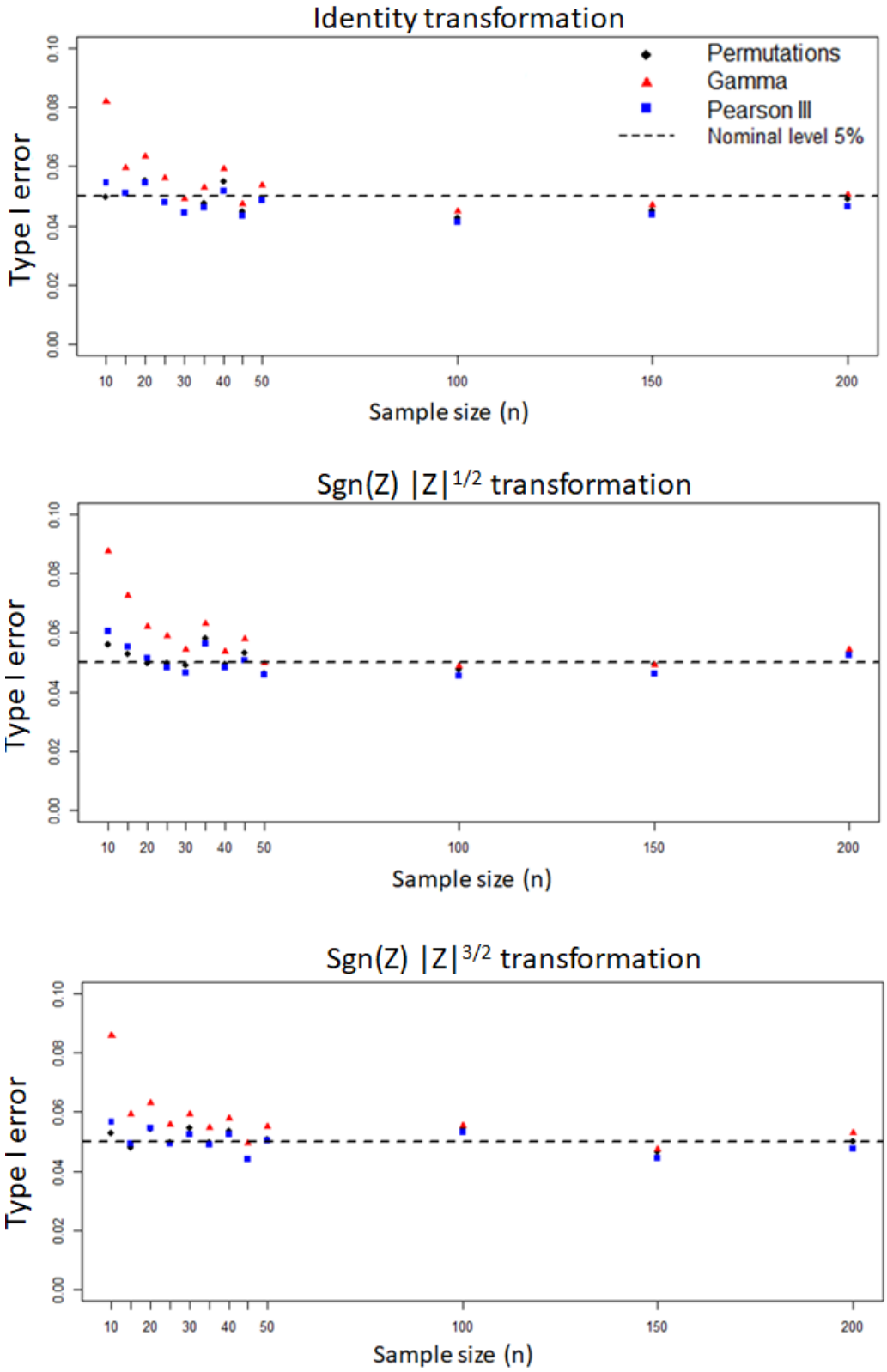

Figure 1: Type I error rates of HSIC for Gaussian distributions (here, the nominal level α is equal to 5% and represented in dotted black). Error rates are calculated using permutations ($B = 3000$), Gamma distribution and Pearson type III distribution for different data transformations.

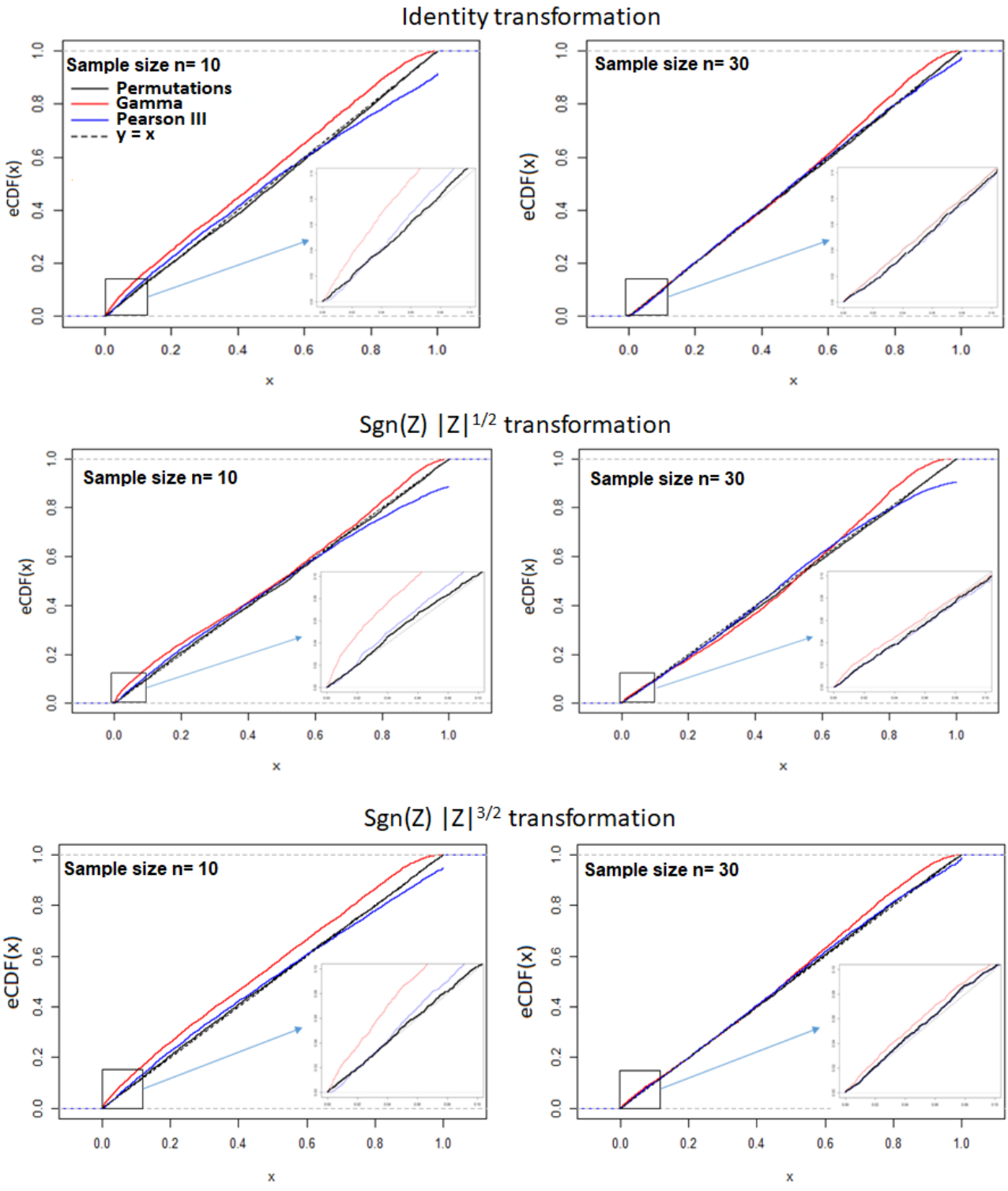

Figure 2: Empirical Cumulative Density Functions (eCDF) of the p-value under independence hypothesis for different data transformations and sample sizes. The p-value is approximated using permutations ($B = 3000$), Gamma distribution and Pearson type III distribution. The reference line $y = x$ is also plotted in dotted black.

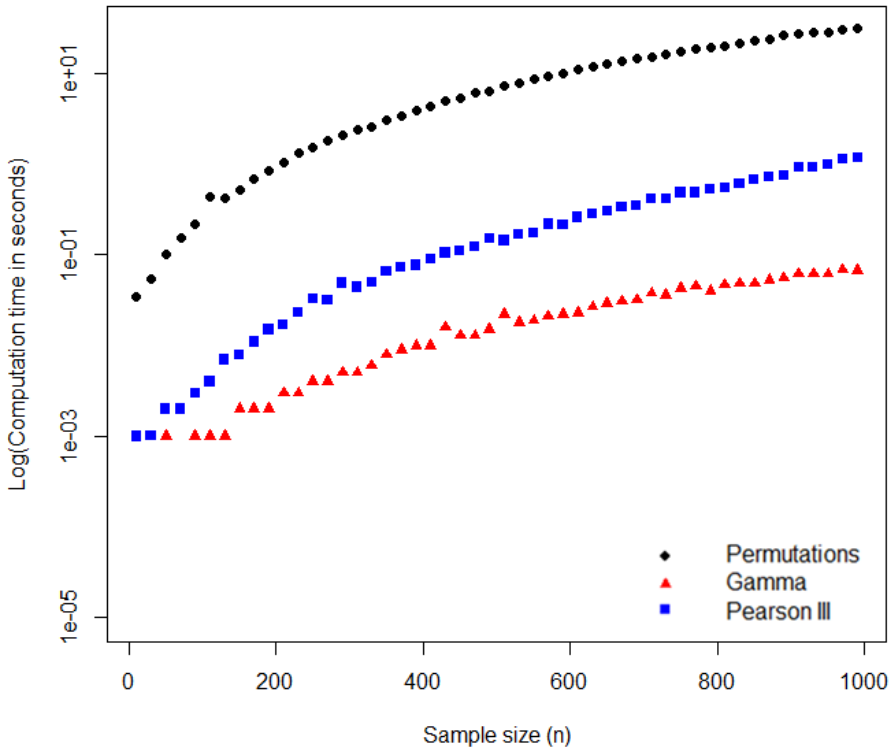

Figure 3: Logarithmic of computation time (in seconds) to estimate the p-value for the different approximation methods, according to the sample size n.

3.2. HSIC-based independence test for Space-Filling Designs

Either for screening or ranking purposes, the HSIC measures mainly depend on the learning n-sample of inputs and associated outputs $(\mathbf{X}^{(m)}, Y^{(m)})_{m=1}^{n}$. Moreover, it usually assumed that the input samples are independent and identically distributed (i.i.d.). However, there are other techniques to generate experimental designs, still within the framework of Monte-Carlo (or Quasi-Monte Carlo) approaches which are often used in uncertainty treatment in numerical simulations. A survey and recent comparison studies of these approaches can be found in [DCI13]. This section aims at providing an overview of different existing methods to generate a sample of n observations $(X_1^{(1:n)}, \ldots, X_d^{(1:n)})^\top$ and also introduce the notion of Space-Filling Designs. We highlight the influence of such sampling methods on the statistical independence tests. In addition, we propose a new method to adapt the HSIC-based independence test conditionally to the sampling strategy.

3.2.1. Review of sampling methods

In what follows, we assume that the inputs variables $X_1, \ldots, X_d$ follow a uniform distribution over the support $[0,1]^d$. Otherwise, one can always work under uniformity and then apply the inverse transformation to place the support back on the original scale and retrieve the original distribution.

Simple Random Sampling (SRS). In order to build the numerical Design of Experiments (DoE), i.e. to select the points at which the model will be evaluated, the simplest sampling method is to generate independently for each variable X_i a set of n independent realizations of a uniform random variable on $[0,1]$. More formally, the set of points resulting from such sampling is denoted $\mathcal{P} = \{\mathbf{X}^{(m)}\}_{m=1:n}$ with

$$\mathbf{X}^{(m)} = (U_1^{(m)}, \ldots, U_d^{(m)}) \ , \ m \in \{1, \ldots, n\},$$

where the $U_i^{(m)}$, $i \in \{1, \ldots, d\}$, are independent random variables uniformly distributed. The resulting DoE $\mathcal{P}$ is called a (simple) random design. Working with fixed n, such a design does not guarantee a good distribution of points in space. A random design very often reveals both gaps and areas where the sampled points are concentrated. This phenomenon is illustrated by Figure 4. This kind of problems can be improved by considering DoE resulting from sampling methods consisting in partitioning the input space into distinct sub-domains. Once

this partitioning is done, a point is randomly sampled in each sub-domain. The DoE described in the following are based on this idea of cutting and are called "stratified" designs. We focus in particular on the case of Latin Hypercubes.

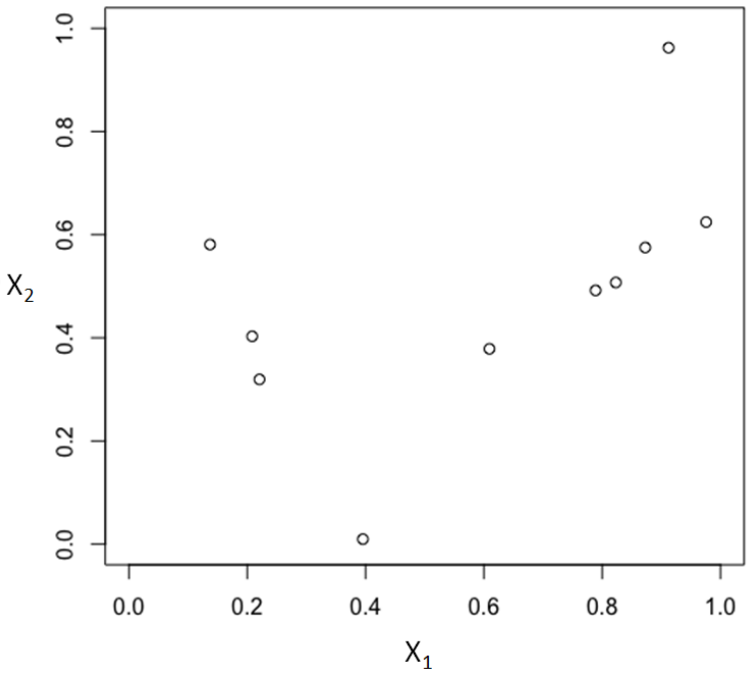

Figure 4: A random design of size $n = 10$ and $d = 2$.

Random Latin Hypercube Sampling (RLHS). Widely applied in computational engineering, RLHS was developed by [MBC79]. It is one form of sampling that can reduce the variance in the Monte Carlo estimate of an integrand. RLHS, which is an extension of stratified sampling, aims at ensuring that each of the scalar input has the whole of its range well scanned, according to a probability distribution. Let the range of each variable X_i, $i = 1, \ldots, d$, be simultaneously partitioned into n equally probable intervals. We note $(X_i^{(m)})_{m=1:n}$ the n-sample of X_i. A RLHS of size n is obtained from a random selection of n values - one per strata - for each X_i. Thus we obtain d n-tuples that form the d columns of the $n \times d$ matrix experiments $\mathbf{X^{(1:n)}}$ generated by RLHS. Once a point is selected in an interval, no other point could be selected in this interval (see Figure 5). A RLHS can also be defined in its centered version for which the points are then placed in the center of the strata. RLHS combines stratification with uniform distribution. Theorem 1 of [Owe13] shows that each of the RLHS points $\mathbf{X^{(m)}} = (X_1^{(m)}, \ldots, X_d^{(m)})$ has the $\mathcal{U}[0,1]^d$ distribution.

Theorem 1. *Let $d \geq 1$ and $n \geq 1$ be integers. Let $\mathbf{X^{(1:n)}}$ be a Latin Hypercube Sample, then $\mathbf{X^{(m)}} \sim \mathcal{U}[0,1]^d$ holds for each $m = 1, \ldots, n$.*

Proof. We refer the reader to the proof of Theorem 10.1 in [Owe13].

Let a numerical simulator $f \in \mathbb{L}^2([0,1]^d)^*$ of which the mean and the variance are denoted by $\mu = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x}$ and $\sigma^2 = \int_{[0,1]^d} (f(\mathbf{x}) - \mu)^2 d\mathbf{x}$, respectively. The empirical estimator of μ by a RLHS, $\mathbf{X^{(m)}}$, is written as follows:

$$\hat{\mu}_{lhs} = \frac{1}{n} \sum_{m=1}^{n} f(\mathbf{X^{(m)}}).$$

An immediate consequence of Theorem 1 is that $\mathbb{E}[\hat{\mu}_{lhs}] = \mu$. [Ste87] asymptotically compares the variances $\mathbb{VAR}(\hat{\mu}_{lhs})$ and $\mathbb{VAR}(\hat{\mu}_{iid})$, where $\hat{\mu}_{iid}$ is the empirical estimator build from a simple random sample. The

*We denote by $\mathbb{L}^2([0,1]^d)$ the space of all functions $f : [0,1]^d \to \mathbb{R}$ whose square is Lebesgue-integrable.

result is that $\mathbb{VAR}(\hat{\mu}_{lhs})$ is asymptotically always smaller than $\mathbb{VAR}(\hat{\mu}_{iid})$. This result is generalized in [Loh96] in the case of multidimensional outputs. Also Loh proves the existence of a central limit theorem for a Latin hypercube under the condition $f \in \mathbb{L}^2([0,1]^d)$. In [Owe97], a non-asymptotic result for $\mathbb{VAR}(\hat{\mu}_{lhs})$ at fixed size n is given by the following proposition.

Proposition 1. *Let $n \geq 2$ be integer. Let $\mathbf{X^{(1:n)}}$ be a Latin Hypercube Sample and $f \in \mathbb{L}^2([0,1]^d)$. Then $\mathbb{VAR}(\hat{\mu}_{lhs}) \leq \sigma^2/(n-1)$.*

Knowing that $\mathbb{VAR}(\hat{\mu}_{iid}) = \sigma^2/n$, Proposition 1 tells us that in the worst case $\hat{\mu}_{lhs}$ has a variance only $n/(n-1)$ times greater than that of $\hat{\mu}_{iid}$.

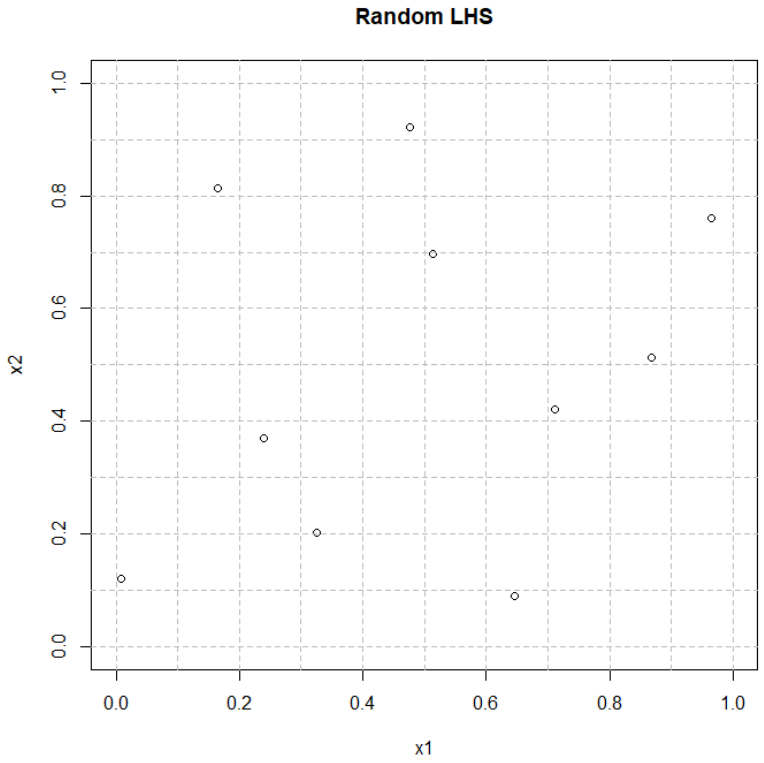

Figure 5: Example of RLHS of size $n = 10$ over $[0,1]^2$ (with regular intervals): each of the n rows (each of the n columns, respectively), which corresponds to an interval of X_1 (of X_2, resp.), contains one and only one draw.

However, this stratification scheme is built by superimposing well stratified one-dimensional samples, and cannot be expected to provide in principle good uniformity properties in a d-dimensional unit hypercube $[0,1]^d$. For example, a bad configuration corresponds to getting all the points aligned in the diagonal of $[0,1]^d$. Alternatives have been proposed to build *better* LHS. One of these alternatives consists in constructing Latin hypercubes for which the sample correlations among the variables are small (see [Owe94, HHY95] for more details). A second alternative consists in constructing LHS for which a Space-Filling criterion is optimized (see [DCI13], for additional information). But generating some few simulations of SFD-LHS, we observe that the usual optimization algorithm for optimal LHS yields a non uniform recovery of the input space. An illustration is given by Figure 6 where 100 optimized-LHS of size $n = 5$ (left) and $n = 10$ (right) are overlaid on the same plot. We can observe that the points does not uniformly cover the input space and some areas will never be explored: these undesirable patterns show that the realizations resulting from this way of sampling cannot be considered as a set of i.i.d. realizations. For this reason, only Random LHS will considered in this work.

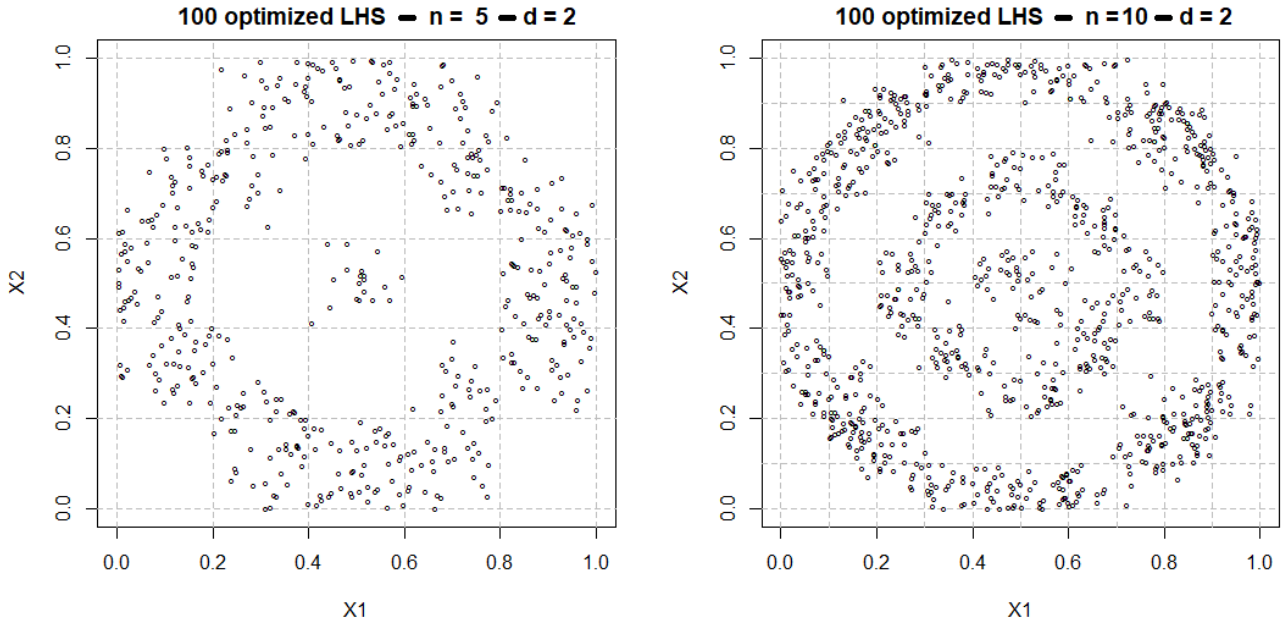

Figure 6: Distribution of the points for 100 optimized-LHS of size $n = 50$ (left) and 10 (right) given by the R *DiceDesign* package and the simulated annealing optimization.

Scrambled Low-discrepancy sequences. Low-discrepancy sequences were initially used in numerical integration to replace the random sequences used in the Monte Carlo method. The method has then been called Quasi-Monte Carlo (QMC). By construction of the sequence, the points are distributed almost uniformly throughout the domain $[0, 1]^d$. Most of these sequences are generated by deterministic algorithms using the inverse radical function in base b. These deterministic sequences have the property that they are well spread throughout the unit cube and are known as low discrepancy sequences. The monograph by [Nie92] provides an excellent discussion of these sequences. A major drawback of the classical LD approach has been the absence of a reliable practical error bound. Even though there exists a deterministic upper bound, this theoretical bound significantly overestimates the actual error in practice. This is in contrast to the crude Monte Carlo method for which the standard error of the estimate is readily available. Examples of these sequences are given by Halton, Sobol, Faure and Niederreiter [Nie92]. In this study, we will focus only on Sobol sequence as it is widely considered in the literature.

Randomization techniques for QMC. LD methods may be thought of as derandomized Monte Carlo. Randomized QMC (RQMC) methods re-randomize them. In RQMC, one takes a LD sequence $(\mathbf{x}^{(m)})_{m=1}^n$ and transforms it into random points $(\mathbf{z}^{(m)})_{m=1}^n$ such that $\mathbf{z}^{(m)}$ contains a LD property. The simplest way to achieve the latter property is to have each $\mathbf{z}^{(m)} \sim \mathcal{U}([0, 1]^d)$. For this, Cranley and Patterson ([CP76]) proposed a rotation modulo one

$$\mathbf{z}^{(m)} = (\mathbf{x}^{(m)} + U) \bmod 1, \tag{8}$$

where $U \sim \mathcal{U}([0, 1]^d)$ and both addition and remainder modulo one are interpreted componentwise. More precisely, we start by adding the same instance of U to all the realizations $\mathbf{x}^{(m)}$ and then we apply the modulo operation to $(\mathbf{x}^{(m)} + U)$. It is easy to see that each $\mathbf{z}^{(m)} \sim \mathcal{U}([0, 1]^d)$. This method is known as *random shift* technique. An alternative is to use *scrambling* methods. These methods work in depth on the digital structure of the sequence itself. For more details on this method of randomization, we refer the reader to [Owe95]. Theoretically, Owen's nested scrambling is powerful, however it is too much bookkeeping in the various implementations. A modified and simplified Owen's scrambling methods were explored and linear scrambling is considered to be suitable choices for scrambling methods. From the implementational point of view, linear scrambling is the simplest and most effective scrambling method to improve the quality of RQMC sequences.

In our study case, the above Space-Filling properties are relevant if we are interested in estimating the HSIC values, as it is based on expectations. However, the use of HSIC-based independence tests assumes the

independence of observations. This hypothesis is verified in the case of Monte Carlo strategy but is no longer satisfied with SFD. In Section 3.2.2, we highlight the influence of the SFD on the HSIC-based independence tests. In Section 3.2.3, we introduce a new method to adapt the HSIC-based independence test conditionally to the sampling strategy.

3.2.2. The impact of the Space-Filling Designs on the HSIC-based independence test

We consider the modified Sobol G-function:

$$f_G(\mathbf{X}) = f(X_1, X_2, X_3, X_4) = \prod_{i=1}^{3} \frac{|4X_i - 2| + a_i}{1 + a_i}, \tag{9}$$

where $(a_1, a_2, a_3) = (0, 1, 99)$ and $\forall i \in \{1, \dots, 4\}, X_i \sim \mathcal{U}[0, 1]$, meaning that we add a dummy input variable X_4 for analysis purposes. Thus under $(\mathcal{H}_0)$, we simulate the distributions of $P_{val,4}$ by Algorithm 1 where at Step 1 the learning sample is generated by one of the three strategies: SRS, RLHS or scrambled-Sobol. We consider the permutation-based strategy (here $B = 3000$) to estimate the p-values at Step 5. Results obtained with $L = 1000$ repetitions are given by Figure 7. For the random designs SRS and RLHS where the observations are i.i.d., the distribution is uniform, regardless of n, as expected. On the contrary, for scrambled-Sobol sequence, the distribution moves away from uniformity as the sample size increases. Besides, this impacts the level of the test. More precisely, from $n \geq 50$, the first-kind error ("observed" level) is much lower than the prescribed level and a correction is thus required.

3.2.3. Proposed correction with Conditional Randomization Test (CRT)

In what follows, we propose a general new method for testing the independence of the variables X_i and Y knowing that the input variable X_i is sampled according to a certain algebraic structure (Simple random sample, Random LHS, scrambled-Sobol). To overcome this problem and recover the nominal level, we propose a new method to estimate the p-value at Step 5 in Algo. 1. The proposed approach consists in replacing the permutation step by a sampling calculation (see Step 2 in Algo. 2). We recall that the permutations are performed to break the dependency. Here we propose to generate a new sample $X_i^{[b]}$ according to the used sampling method instead of making one permutation. Our work builds on the Conditional Randomization Test (CRT) introduced in [CFJL16]. The CRT is a very general framework for testing the conditional independence of variables X_i and Y given a potentially high dimensional random vector Z. Assuming then that the distribution of $X_i|Z$ is known, the CRT operates by sampling a new copy of the X_i-values in the data set. To construct a test of $(\mathcal{H}_0^i)$, then, the CRT repeats this process B times, sampling

$$\mathbf{X}_i^{[1]}, \dots, \mathbf{X}_i^{[B]} \text{ independently.}$$

Under the null hypothesis, the couples $(\mathbf{X}_i^{[1]}, \mathbf{Y}), \dots, (\mathbf{X}_i^{[B]}, \mathbf{Y})$ are all identically distributed; in fact, they are exchangeable. Thus the random variables

$$\widehat{\text{HSIC}}_1(X_i, Y), \dots, \widehat{\text{HSIC}}_B(X_i, Y) \tag{10}$$

are therefore exchangeable as well. We can compute a p-value by ranking the value that is obtained from the learning sample against the values that are obtained from the CRT's copies:

$$\hat{p}_{val,i}^B = \frac{1 + \sum_{b=1}^{B} \mathbb{1}_{\widehat{\text{HSIC}}_b(X_i,Y) > \widehat{\text{HSIC}}_{obs}(X_i,Y)}}{B + 1}$$

The exchangeability of the random variables in expression (10) ensures that this is a valid p-value under the null, i.e. it satisfies that $\mathbb{P}(\hat{p}_{val,i}^B < \alpha) = \alpha$ for all $\alpha \in [0, 1]$ if the null hypothesis $(\mathcal{H}_0^i)$ is true. In our context, the input variable X_i is sampled according to the chosen sampling strategy (SRS, RLHS, scrambled-Sobol). Thus we can consider that the sampling strategy plays the role of the variable Z. For instance, in the case of MC sampling (SRS), we have no conditioning because the samples are obtained randomly. On the other hand, Random LHS (which is an extension of stratified sampling) is obtained in order to guarantee certain constraints (conditioning to a hidden variable Z), just like the scrambled-Sobol sampling.

Figure 8 illustrates the performance of the proposed independence test approach under $(\mathcal{H}_0)$ for the G-Sobol function f_G. The uniformity of the p-value distribution is now recovered for all the sampling methods, and regardless of n. The resulting tests will be of prescribed level.

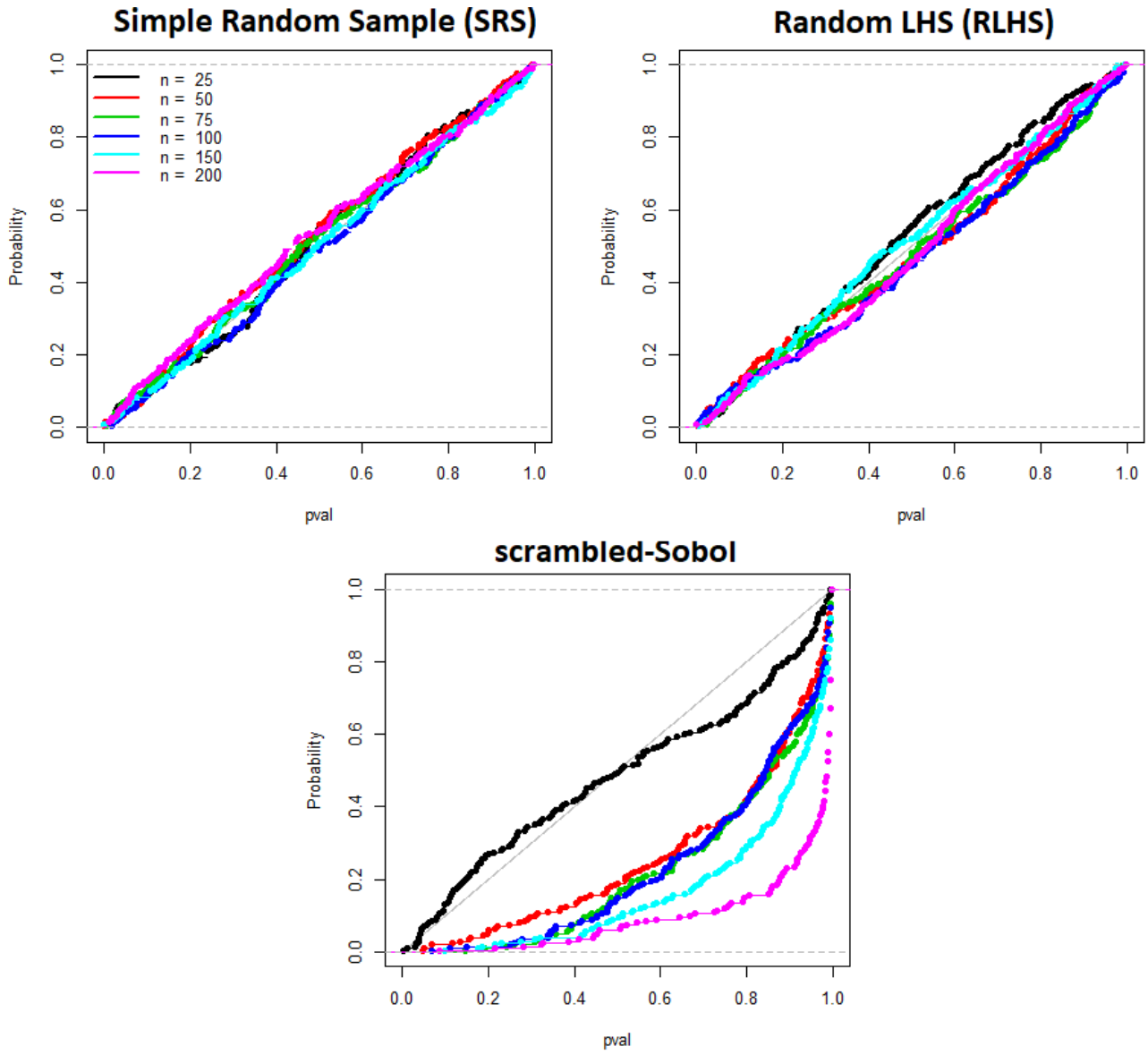

Figure 7: G-Sobol function f_G – Empirical Cumulative Density Functions (eCDF) of p-values under ($\mathcal{H}_0$) for different sample sizes n and sampling methods.

Algorithm 2 – Conditional Randomization Test (for each X_i)

Require: The learning sample $(\mathbf{X}_i, \mathbf{Y})$ of n inputs/outputs and B

1: Compute $\widehat{\text{HSIC}}_{obs}(X_i, Y)$ from Eq. (3)

2: Generate B new samples $\left(\mathbf{X}_i^{[b]}, \mathbf{Y}\right)_{1\leq b\leq B}$ according to the used sampling method

3: Compute the B sampling-based estimators $\left(\widehat{\text{HSIC}}_b(X_i, Y)\right)_{1\leq b\leq B}$ by replacing X_i by $\mathbf{X}_i^{[b]}$ in Eq. (3)

4: Estimate the p-value by $\hat{p}^B_{val,i} = \dfrac{1}{B+1}\left(1 + \displaystyle\sum_{b=1}^{B} \mathbb{1}_{\widehat{\text{HSIC}}_b(X_i,Y) > \widehat{\text{HSIC}}_{obs}(X_i,Y)}\right)$

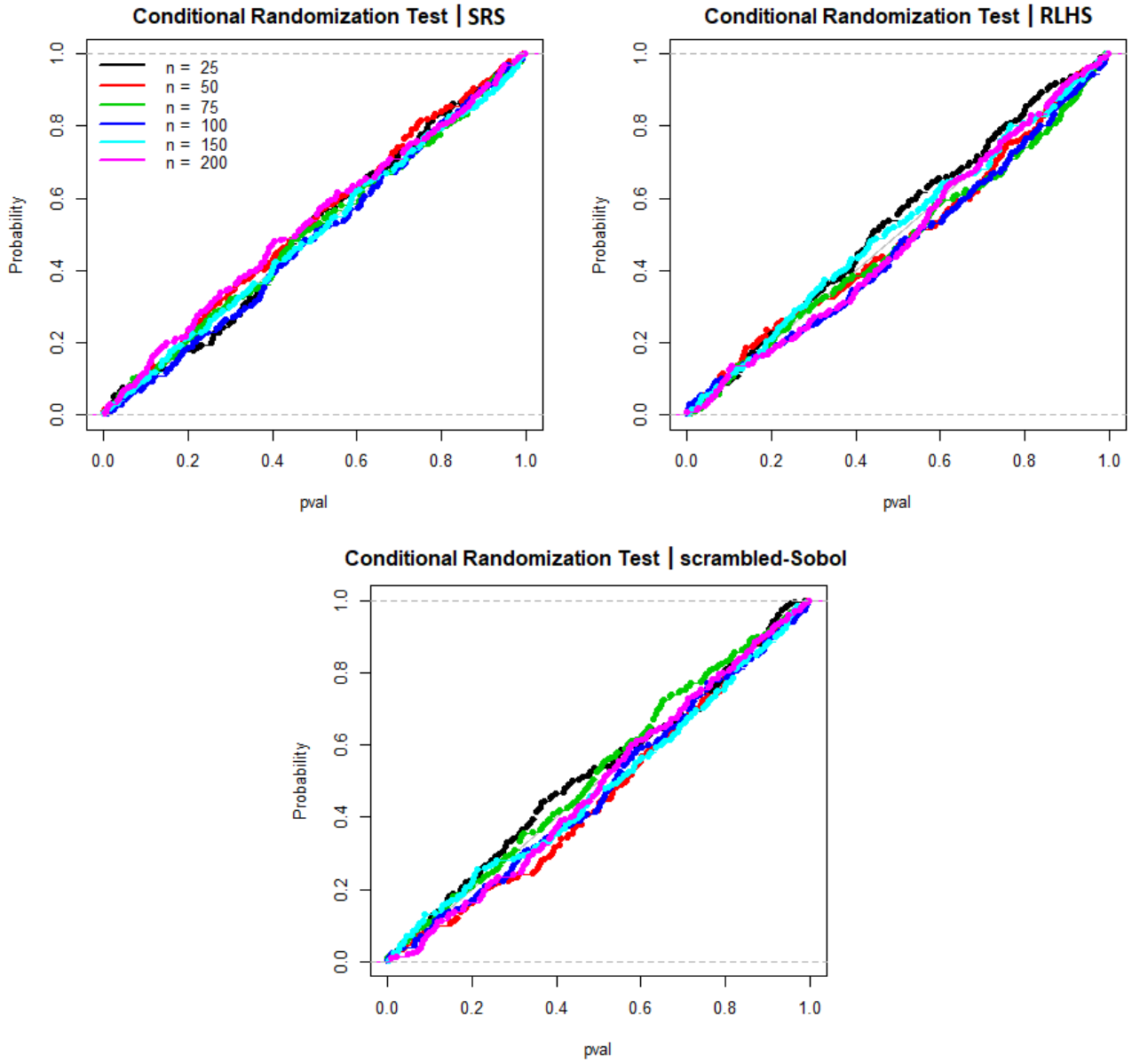

Figure 8: G-Sobol function f_G – Empirical Cumulative Density Functions (eCDF) of p-values under ($\mathcal{H}_0$) obtained by CRT approach for different sample sizes n and sampling methods.

3.2.4. Impact of CRT on test power

We now consider all the inputs of f_G function where each X_i is more influential the lower a_i is. We evaluate the selection rate (the percentage of perfect screening) obtained with the CRT tests and the different sampling methods. Figure 9 shows the results obtained for different sample sizes and for a level $\alpha = 0.05$. For the most influential inputs X_1 and X_2, we note that the larger the sample size n, the better the selection rate. Regardless of the considered sampling method, we note that the CRT method allows to select the variable as influential with a very small sample size. On the other hand, since the variable X_2 is relatively important, we notice that the scrambled-Sobol sampling allows to have a better selection rate for small sample sizes. Regarding the non influential inputs X_3 and X_4, we note that whatever the sample size n, these inputs are correctly screened (as non influential) in 95% of cases. The error of screening of 5% corresponds to the level of the test (here $\alpha = 0.05$). Regardless of the considered sampling method we note that the permutation-based method allows to select the influential inputs from a very small sample size.

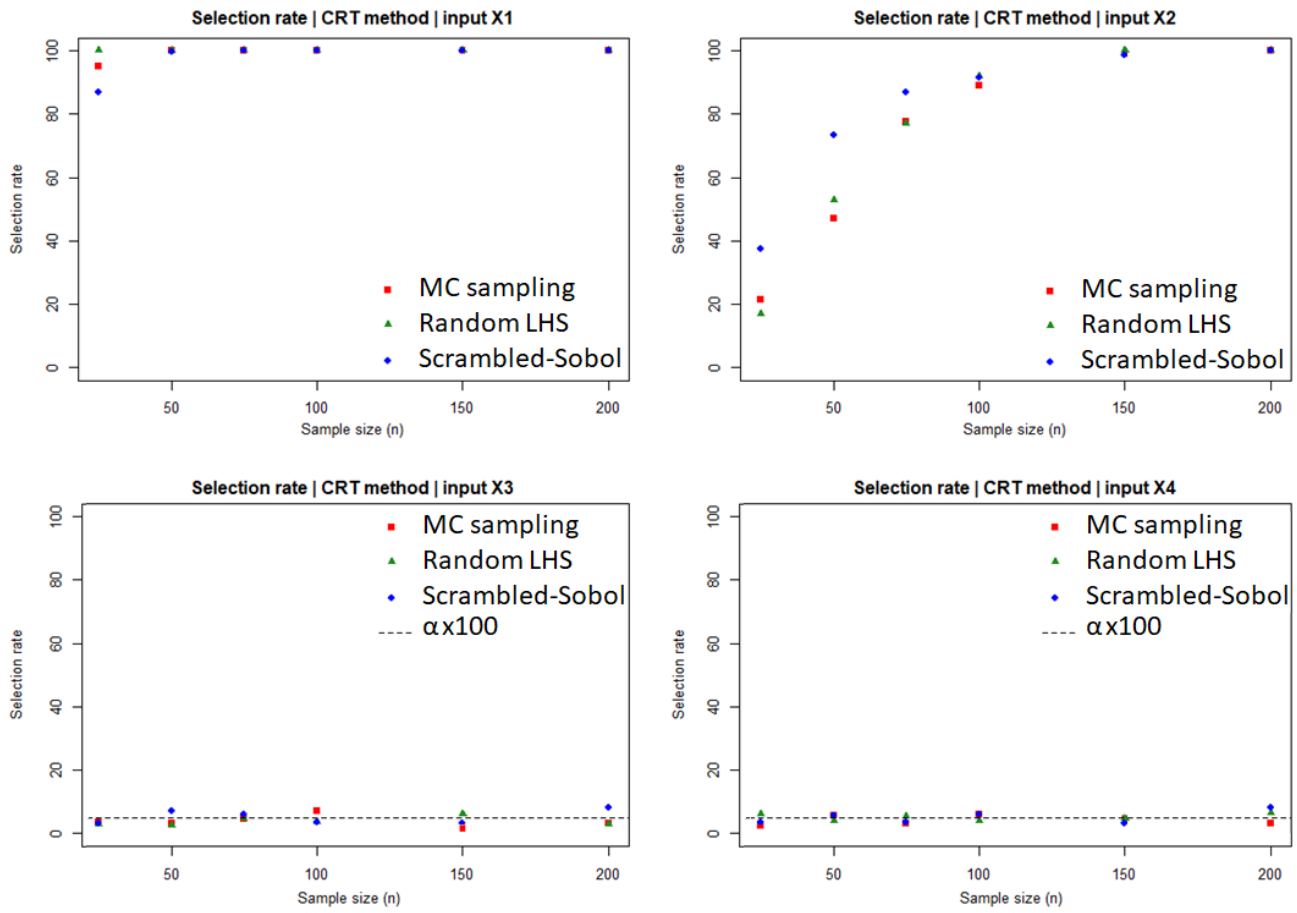

Figure 9: G-Sobol function – Selection rate of the inputs provided by screening based on CRT independence tests, for a level $\alpha = 0.05\%$ and for different samplings.

4. Optimal bandwidth selection

A key element in the efficiency of HSIC in detecting influence is the choice of the kernel and its parameters. Especially, Gaussian radial basis (RBF) kernel has shown to be an optimal choice of kernel for variety of applications and objectives. The significance of the RBF kernel depends on its bandwidth parameter. Thus, an improper choice could significantly worsen the results of screening and ranking. In literature, the bandwidth parameter of RBF kernel is usually selected based on traditional empirical standard deviation or empirical median. Recently in [ALMM21] a method is proposed to avoid such a bandwidth parameter choice. The authors aggregate several single HSIC-based independence tests with different bandwidths, and prove sharp upper bounds for the uniform separation rate of the aggregated procedure over Sobolev balls. Here we propose a fast and efficient bandwidth selection method to improve the power of the test, while optimizing the computation time of the procedure.

4.1. Proposed bandwidth parameter selection

Let λ_i and λ be the kernel bandwidth parameters associated with the variables X_i and Y. We denote by $\widehat{\mathrm{HSIC}}(X_i, Y|(\lambda_i, \lambda))$, the estimated HSIC by these parameters (see Eq. (3)). For GSA purpose, the selection of optimal bandwidth can be formulated as follows

$$\begin{aligned}(\lambda_i, \lambda)_n^* &= \texttt{arg}\max_{(\lambda_i,\lambda)} \widehat{\mathrm{HSIC}}\big(X_i, Y \mid (\lambda_i, \lambda)\big),\\ \widehat{\mathrm{HSIC}^*}\big(X_i, Y\big) &= \widehat{\mathrm{HSIC}}\big(X_i, Y \mid (\lambda_i, \lambda)_n^*\big) = \max_{(\lambda_i,\lambda)} \widehat{\mathrm{HSIC}}\big(X_i, Y \mid (\lambda_i, \lambda)\big),\end{aligned} \tag{11}$$

where $(\lambda_i, \lambda)^*$ is the optimal choice of bandwidth parameters. In [Bhu18], the authors propose a grid search method to solve the optimization problem (11). However, the accuracy of the solution mainly depend on the discretization of the search space. From our experimental evaluation, it seems that the computational complexity is higher, because it might take several iterations to converge to the solution of the optimization problem. Here, we propose to use the gradient of the HSIC to provide the search direction that allows us to reach the optimal solution more faster. By applying the chain rule, and first-order derivatives of matrices, the derivatives of HSIC w.r.t. (λ_i, λ) are obtained:

$$\begin{cases} \dfrac{\partial \widehat{\text{HSIC}}}{\partial \lambda_i} &= \frac{1}{n^2}\text{Tr}\big(HLHM\big) \\ \dfrac{\partial \widehat{\text{HSIC}}}{\partial \lambda} &= \frac{1}{n^2}\text{Tr}\big(HL_iHN\big) \end{cases} \tag{12}$$

where $M = \big(\dfrac{\partial k_i}{\partial \lambda_i}(X_i^{(l)}, X_i^{(m)})\big)_{1\leq l,m\leq n}$ and $N = \big(\dfrac{\partial k}{\partial \lambda}(Y^{(l)}, Y^{(m)})\big)_{1\leq l,m\leq n}$. Table 4.1 recalls the usual characteristic kernels and their derivatives. We may note that both $\widehat{\text{HSIC}}$ and its derivatives lead to closed-form solutions, just involving simple matrix multiplication and a trace operation.

	Kernel function	**Derivative kernel function**
Gaussian	$\exp(-\frac{r^2}{2\ell^2})$	$\frac{r^2}{\ell^3}\exp(-\frac{r^2}{2\ell^2})$
Laplacian	$\exp(-\frac{r}{\ell})$	$\frac{r}{\ell^2}\exp(-\frac{r}{\ell})$
Matérn 3/2	$\left[1+\frac{r\sqrt{3}}{\ell}\right]\exp\left(-\frac{r\sqrt{3}}{\ell}\right)$	$\frac{3r^2}{\ell^3}\exp\left(-\frac{r\sqrt{3}}{\ell}\right)$
Matérn 5/2	$\left[1+\frac{r\sqrt{5}}{\ell}+\frac{5r^2}{3\ell^2}\right]\exp\left(-\frac{r\sqrt{5}}{\ell}\right)$	$\frac{5r^2}{3\ell^3}\exp\left(-\frac{r\sqrt{5}}{\ell}\right)\left[1+\frac{r\sqrt{5}}{\ell}\right]$
Rational quadratic	$1-\frac{(r/\ell)^2}{1+(r/\ell)^2}$	$\frac{2r^2/\ell^3}{\left(1+(r/\ell)^2\right)^2}$
Inverse multiquadratic	$\frac{1}{\sqrt{1+(r/\ell)^2}}$	$\frac{r^2}{\ell^3\left(1+(r/\ell)^2\right)^{3/2}}$

Table 1: Usual characteristic kernel functions and their derivatives.

HSIC-based independence test with optimal bandwidth. The previously methods cited in Sections 2.3 and 3.1 for approximating p-values are no longer directly applicable to the HSIC* statistic, since $(\lambda_i, \lambda)^*$ is no longer a constant and dependent on the learning sample $(X_i^{(m)}, Y^{(m)})_{1\leq m\leq n}$. For instance, for the Gamma-based approximation, we have no theoretical guarantee that the asymptotic distribution of the HSIC with optimal bandwidth can be approximated by a Gamma law. For the Pearson type III, as it is based on the exchangeability of the permutations, this condition is no more verified when we use the optimal bandwidths. This leaves the permutation-based method which can be adapted by re-estimating the optimal bandwidths for each permuted sample. This yields to the modified permutation test given by Algorithm 3.

Algorithm 3 – Optimal permutation-based independence test (optimal bandwidth parameter)

Require: The learning sample $(\mathbf{X}_i, \mathbf{Y})$ of n inputs/outputs and the number of permutations B

1: $\widehat{\text{HSIC}}^*_{obs} = \widehat{\text{HSIC}}^*\big(X_i, Y\big)$

2: Generate B permutation-based samples $\big(X_i, Y_{[b]}\big)_{1\leq b\leq B}$

3: Compute the B permutation-based estimators: $\widehat{\text{HSIC}}^*_b = \widehat{\text{HSIC}^*}\big(X_i, Y_{[b]}\big)$, $\forall b \in 1, \ldots, B$

4: Estimate the p-value: $\hat{p}^B_{val,i} = \dfrac{1}{B+1}\big(1 + \sum_{b=1}^{B} \mathbb{1}_{\widehat{\text{HSIC}}^*_b > \widehat{\text{HSIC}}^*_{obs}}\big)$

5: IF $\hat{p}^B_{val,i} < \alpha$, then reject $(\mathcal{H}^i_0)$. ELSE accept $(\mathcal{H}^i_0)$

At Steps 1 and 3, the optimization is performed using a gradient-based ascent algorithm. Regarding the number of permutations B, we will consider in the following a large value ($B = 3000$). However, to avoid a priori choice, the algorithms proposed in [EAM] could be used to optimize B (according to the GSA purpose).

4.2. Numerical tests

The proposed optimal permutation test (Algo 3) is compared with the usual test with standard deviation for (λ_i, λ), the Mutual Information-based test (MINT) proposed by [BS19] and with the aggregated procedure of [ALMM21]. For this, we consider the same analytical test cases as in [ALMM21]. These examples denoted (i) to (iii) consider the three following different mechanism of dependence, each one varying by a shape parameter (denoted l or ρ):

(i) For $l \in \{1, \dots, 10\}$ and $(X, Y) \in \mathcal{U}([-\pi, \pi])^2$, with joint density function $f_l(x, y) = \frac{1}{4\pi^2}\big(1 + \sin(lx)\sin(ly)\big)$.

(ii) Let L, Θ, ϵ_1 and ϵ_2 be independent with $L \sim \mathcal{U}(\{1, \dots, l\})$ for some $l \in \{1, \dots, 10\}$, $\Theta \sim \mathcal{U}([0, 2\pi])$, and $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, 1)$. Set

- $X = L\cos(\Theta) + \epsilon_1/4$,
- $Y = L\sin(\Theta) + \epsilon_2/4$.

(iii) Let X and ϵ be independent with $X \sim \mathcal{U}([-1, 1])$ and $\epsilon \sim \mathcal{N}(0, 1)$, and for a parameter $\rho \in \{0.1, \dots, 1\}$, let $Y = |X|^\rho\ \epsilon$.

For each example, the power of the different testing procedures is estimated using 1000 different samples of (X, Y) of sizes $n = 50$, 100 and 200. The obtained power curves are given by Figure 10, w.r.t parameters l and ρ. Regarding the sample size, our procedure of testing constantly yields the best performances. For the case (i), the HSIC test with optimal bandwidth provides results much better than all the other tests. For the examples (ii) and (iii), the proposed method is the most powerful procedure. Regardless of the mechanism of dependence between the variables, the optimal HSIC procedure performs better than the state-of-the-art methods.

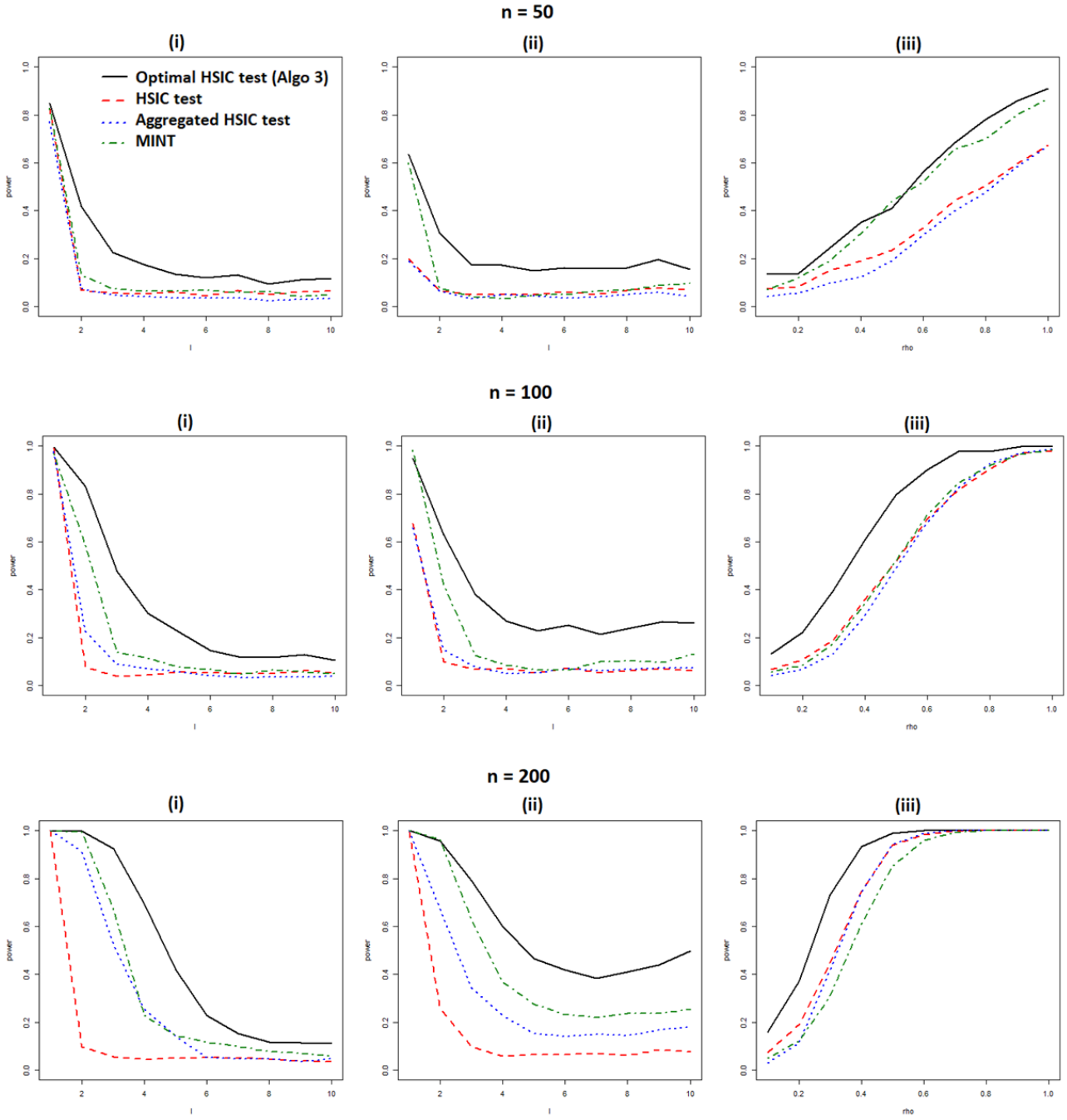

Figure 10: Power curves of independence tests, according to the shape parameters (l or ρ) for the three types of dependence (i) to (iii) and for different sample sizes n.

5. HSIC for functional output

Time series has received a large interest over the last decades within the data mining and machine learning communities. The most used dissimilarity measures for such type of data are the euclidean distance (ED) and the dynamic time warping (DTW). The computational cost of ED is lower than the one of DTW, but ED is not able to deal with temporal distortions. [CVBM07] introduces the Global Alignment Kernel (GAK) that takes into account all possible alignments in order to produce a reliable similarity measure to be used at the core of standard kernel methods such as Support Vector Machines (SVM). On the other hand, Da Veiga [DV15] advises the use of kernels depending on semi-metrics dedicated to functional data for GSA purpose. In the following,

we introduce some basic notions of functional dimension reduction on which the proposed kernels are based. Then, several modified kernels based on weighting functions are proposed to obtain more efficient results.

5.1. Functional dimension reduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We assume that the random process $\mathbf{Y}$ belongs to the functional space $\mathcal{H} = \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{Y})$ with

$$\mathcal{Y} = \left\{ \mathbf{y} : [0,T] \to \mathbb{R}, ||\mathbf{y}|| = (< y, y >)^{1/2} = \left(\int_0^T \mathbf{y}(t)^2 \mathbf{dt} \right)^{1/2} < +\infty \right\}.$$

Moreover, we assume that $\mathbf{Y} \in \mathcal{H}$ has zero mean and continuous covariance function $C(t,s)$. Then

$$\mathbf{Y}(t) = \sum_{h=1}^{\infty} U_h \psi_h(t) \text{ for } t \in [0,T] \tag{13}$$

where $\{\psi_h\}_{h=1}^{\infty}$ is an orthonormal basis of eigenfunctions of the integral operator corresponding to C:

$$\theta_h \psi_h(t) = \int_0^T C(t,s)\psi_h(s)\mathrm{ds}, \tag{14}$$

and with $\{U_h\}_{h=1}^{\infty}$ denoting a set of uncorrelated random variables with zero mean and variance θ_h. The eigenvalues θ_h measure the variation of the projected process $\mathbf{Y}$ onto the U_h direction. Each θ_h represents the part of variance of $\mathbf{Y}$ which is explained by the component U_h (alone) and their sum is equal to the total variance. Decomposition Eq. (13) is known as the Karhunen-Loève (KL) expansion of $\mathbf{Y}$ ([LK10]). In the following, we denote the truncated version of $\mathbf{Y}$ as

$$\mathbf{Y}_q(t) = \sum_{h=1}^{q} U_h \psi_h(t), \tag{15}$$

which represents, in the mean square error sense, the optimal q-term approximation of $\mathbf{Y}$ ([LK10]).

Computational details for functional PCA. In practice, the covariance structure of the process $\mathbf{Y}$ might be unknown and has to be estimated from the data. More precisely, $C(s,t)$ can be estimated from the sample $\{Y^{(1)}, \ldots, Y^{(n)}\}$ by:

$$\hat{C}(s,t) = \frac{1}{n} \sum_{m=1}^{n} Y^{(m)}(s) Y^{(m)}(t). \tag{16}$$

The eigenvalue problem defined by Eq. (14) is then solved by replacing C by $\hat{C}$ (see, [CFS99] for convergence results). That approximated eigenvalue problem can be solved by discretizing the trajectories at several discrete time points $t_1, \ldots, t_{N_T}$. This yields the matrix of n discretized trajectories $\mathrm{Y} \in \mathcal{M}_{n,N_T}(\mathbb{R})$ wih $\mathrm{y}_{i,j} = [Y^{(i)}(t_j)]$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, N_T$. It leads to the empirical covariance matrix defined as $\hat{C}_n = \frac{1}{n}{}^{\top}\mathrm{YY} \in \mathcal{M}_{N_T,N_T}(\mathbb{R})$. We then have to solve a classical multivariate PCA with N_T variables given by a sample of size n. Other approaches to implement functional PCA can be found in the literature. In [Ram06], e.g., the authors propose to expand the curves as linear combinations of spline basis functions, and to apply PCA to the coefficients of the curves on the spline basis. There also exist different criteria for the choice of the truncation argument q [Jac93]. One can cite the Kaiser-Guttman criterion which consists of choosing the first components with eigenvalues higher than 1. Instead of using the variance provided by θ, the choice of q could be based on the "percentage of variance" given by the proportion $\frac{\theta_h}{\sum_{i=1}^{n} \theta_i}$. By this way, we choose q so that the cumulated ratio of variance explained by the first components exceeds a given threshold. Alternatively, to avoid the arbitrary choice of the threshold, the eigenvalues are often displayed in a downward curve and q is chosen according to the stagnation of the slope.

5.2. Kernel for functional random variable

Since we place ourselves in a finite space thanks to aforementioned reduction, one can use standard kernels, like Gaussian or Laplace kernels. For example, let $d_q(.,.)$ be such a semi-metric defined on $\mathcal{Y} \times \mathcal{Y}$ ([CFS99]). The kernel associated to $\mathcal{Y}$ is then given by:

$$k_q(Y^{(m)}, Y^{(l)}) = k(d_q(Y^{(m)}, Y^{(l)})) = k(||\mathbf{U}^{(m)} - \mathbf{U}^{(l)}||_{\mathbb{L}^2}), \text{ for any } l, m \in \{1, \dots, n\}, \tag{17}$$

where k is a kernel acting on $\mathbb{R}$, $\mathbf{U}^{(\bullet)} = (< Y^{(\bullet)}, \psi_1 >, ..., < Y^{(\bullet)}, \psi_q >)^{\mathrm{T}}$ is the projection of the observation $Y^{(\bullet)}$ on the q-dimensional truncated space given by Karhunen-Loève expansion (Eq. (13)). For instance, in the case of Gaussian kernel, we obtain:

$$\begin{aligned} k_q^{rbf}(Y^{(m)}, Y^{(l)}) &= \exp\left(- \frac{||\mathbf{U}^{(m)} - \mathbf{U}^{(l)}||_{\mathbb{L}^2}^2}{2\sigma^2} \right), \\ &= \prod_{h=1}^{q} \exp\left(- \frac{|U_h^{(m)} - U_h^{(l)}|^2}{2\sigma^2} \right), \end{aligned} \tag{18}$$

where σ is estimated by the empirical standard deviation associated to the sample $\{\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(n)}\}$ or by the empirical median associated to $\left(||\mathbf{U}^{(m)} - \mathbf{U}^{(l)}||_{\mathbb{L}^2}^2\right)_{1 \leq m,l \leq n}$. In the following, we use the closure properties of the kernels reminded by Proposition 2, which allow us to create more complicated kernels from simple building blocks.

Proposition 2 (Closure properties)**.** *Let $k_1, \dots, k_q$ be kernels over $\mathbb{R} \times \mathbb{R}$, $\mathbf{u} = (u_1, \dots, u_q)^\top \in \mathbb{R}^q$, and $\forall i \in \{1, \dots, q\}, w_i \in \mathbb{R}^+$. The following functions are still kernels:*

(i) Sum of weighted 1-dimensional kernels: $k_{\sum_w}(\mathbf{u}, \mathbf{u}') = \sum_{h=1}^{q} w_h k_h(u_h, u_h')$,

(ii) Product of weighted 1-dimensional kernels: $k_{\prod_w}(\mathbf{u}, \mathbf{u}') = \prod_{h=1}^{q} w_h k_h(u_h, u_h')$.

The weights $w_1, \dots, w_q$ control the relative contribution of the different factors $U_1, \dots, U_q$ in the global kernel. In practice, we recommend to consider the percentage of variance explained by each factor given by $w_h = \frac{\theta_h}{\sum_{i=1}^{q} \theta_i}$. Regarding the bandwidth parameters of the kernels $k_1, \dots, k_q$, we consider the empirical standard deviation associated to the sample $\{U_h^{(1)}, \dots, U_h^{(n)}\}, h = 1, \dots, q$. We note that the kernel given by Eq. (18) is a special case of the product of kernels where all the weights are set to 1 and σ is taken to be the same for all the kernels $k_1, \dots, k_q$.

Note that, to avoid any confusion with the usual $\mathrm{R}^2_{\mathrm{HSIC}}$ for scalar output (cf. Eq. (5)), we denote by $\mathrm{R}^2_{\mathrm{HSIC,PCA}}$ the counterpart obtained with the PCA-based reduction and the kernels given by Proposition 2.

5.3. Experimental results and discussion

In this section, we illustrate the behavior of the functional HSIC-based sensitivity indices on several examples representative of typical GSA industrial applications. To do so, we consider the following arctangent-based temporal function:

$$\mathbf{Y}_{atan}(t, X_1, X_2, X_3) = atan(X_1)\cos(t) + 50 atan(X_2)\sin(t), \tag{19}$$

where $t \in [0, 2\pi]$, $\forall i, 1, \dots, 3,\ X_i \sim \mathcal{U}([-7, 7])$. Note that X_3 is a dummy input variable. As the function is additive, we can directly see that the $\mathbf{Y}_{atan}$ mainly depends on the variable X_2. To support our statement, we estimate the $\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}$ for the two first PCA components U_1 and U_2, by an intensive Monte Carlo sampling (sample of size $n = 10000$ to ensure convergence of estimates). Note that a temporal discretization with $N_T = 100$ equally spaced points is used, this number being sufficient to ensure the stability of the results considering the regularity of the curves. The results are summarized in Table 5.3. These two first components explain 96.96% and 3.03% of the output variance, respectively. The effect conveyed by U_1 which strongly dominates all the others is only driven by X_2 which is therefore by far the most influential input. The second principal component U_2 defines a less important mode of variation of $\mathbf{Y}_{atan}$ and carries the much smaller influence of X_1.

	X_1	X_2
U_1 (96.96%)	5.69×10^{-3}	0.904
U_2 (3.03%)	0.903	6.32×10^{-3}

Table 2: Temporal function $\mathbf{Y}_{atan}$. Estimated $\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}$ for the two first PCA components (written in brackets is the percentage of the explained variance by the each component alone).

The objective is now to examine how the $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ indices built upon the different kernels of Proposition 2 reflect the influence of each input, when applied directly on the matrix of n discretized trajectories. More particularly, we want to compare the weighted kernels with $w_h = \frac{\theta_h}{\sum_{i=1}^{q} \theta_i}$ with the no-weighted ones ($w_h = 1$). To do so, we generate $n = 100$ trajectories of $Y(t)$ and we calculate the $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ with k_{Σ_w} and k_{Π_w}, for different basis sizes q. Figure 11 shows the results of all the procedures. The simplest approach (subplot (a)) with product of all kernels and equal weights yields the most misleading analysis. Indeed, as q increases, all the inputs end up having the same $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$, including the non-influential input X_3. From subplot (b), we remark that the sum of kernels with equal weights avoid this mistake: X_2 is found as the last input in the ranking, as expected. However, the influence of X_1 and X_2 strongly varies according to the size of the base q. The influence of X_1 evolves in a non-monotonic way and progressively decreases from $q \geq 3$. Same remark could be made for X_2. With this kernel, the $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ strongly depends on the basis size.

On the other hand, if a weighting by the percentage of variance explained by each component is applied, the $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ indices lead to more consistent results, with a monotonic convergence according to q. More precisely, as q increases, the variable X_2 is considered as the most influential and X_3 is ranked last. Regarding the weighted product of the kernels (subplot (c)), we notice that only the effect of X_2 dominates; it does not allow to highlight the secondary influence of X_1. For these reasons and based on our feedback on other analytical examples, we therefore recommend the use of kernel k_{Σ_w} with $w_h = \frac{\theta_h}{\sum_{i=1}^{q} \theta_i}$. Besides, this kernel will be used for the second illustrative test case.

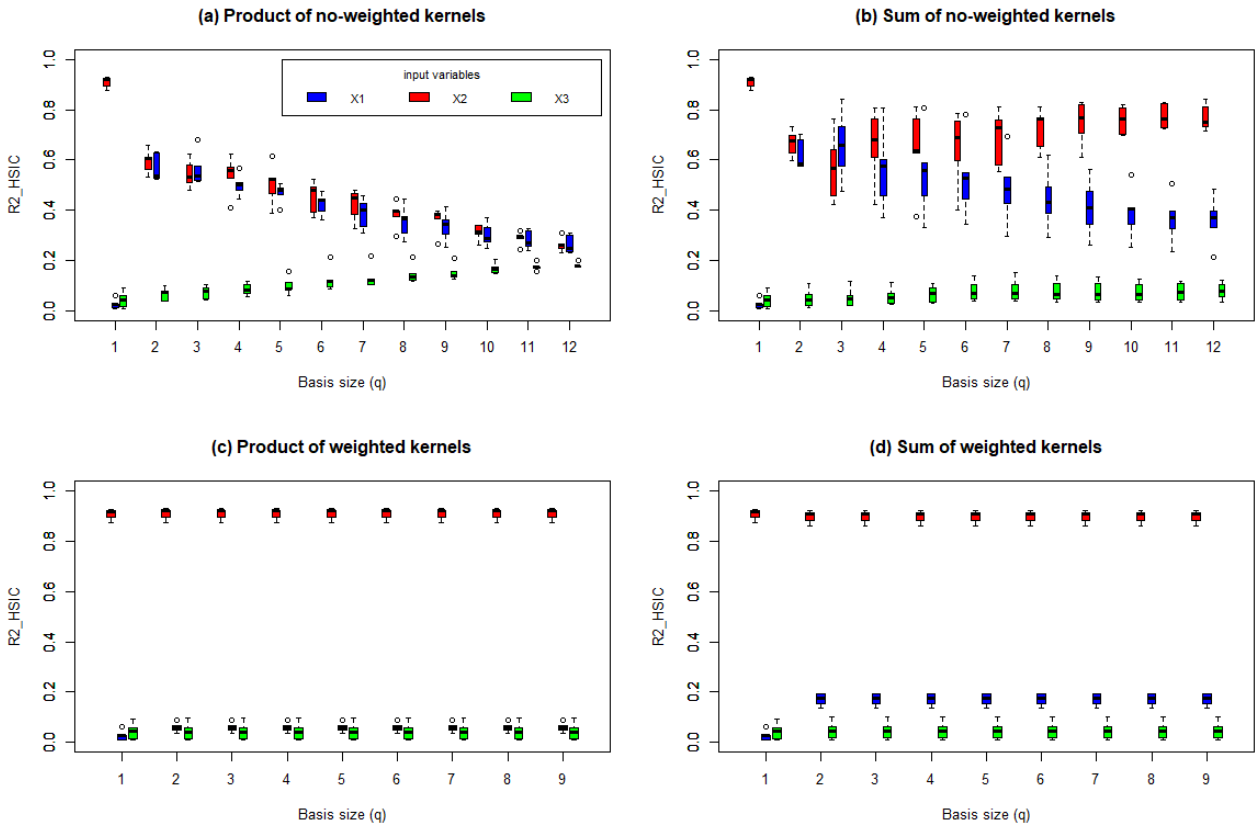

Figure 11: Temporal function $\mathbf{Y}_{atan}$. Estimated $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ sensitivity indices defined by the different weighted and no-weighted kernels according to the basis size q, with $n = 100$ and 50 replicates.

For our second illustration, we consider a compartmental epidemiological model inspired by previous works on COVID-19 (see, [MW20] for more details). The model is a modified Susceptible - Infected - Recovered (SIR) model that accounts for two different types of infectious people: the reported cases and the unreported cases. This model consists of the following system of ordinary differential equations:

$$\begin{cases} S'(t) = -\tau(t)S(t)\big(I(t) + U(t)\big) \\ I'(t) = \tau(t)S(t)\big(I(t) + U(t)\big) - \nu I(t) \\ R'(t) = f\nu I(t) - \eta R \\ U'(t) = (1-f)\nu I(t) - \eta U. \end{cases} \tag{20}$$

Here $t \geq t_0$ is the time in days, t_0 is the beginning date of the model of the epidemic (time at which the epidemic started), $S(t)$ is the number of individuals likely to be infected at time t, $I(t)$ is the number of asymptomatic infectious individuals for an average period of $1/\nu$ days. $R(t)$ and $U(t)$ is the number of reported and unreported symptomatic infectious individuals for an average period of $1/\eta$ days. The transmission rate is given by $\tau(t) = \tau_0 \exp\{-\mu \max(t - N, 0)\}$, and it decreases with an exponential decay with rate μ as soon as social distancing and lockdown start to have an effect after N days. The fraction of asymptomatic infectious who become reported as symptomatic is denoted by f, and $1 - f$ refers to the unreported symptomatic ones. The values of the initial conditions are:

$$S_0 = 66.99 \times 10^6(\text{ initial population }) \ , \ I_0 = \frac{\chi 2}{f\nu} \ , \ R_0 = 1 \ , \ U_0 = \frac{(1-f)\nu}{\eta + \chi 2} I_0, \tag{21}$$

where $\chi 2$ is estimated from the available data. We note that these equations involve many input parameters and model the dynamic between the different compartments. From a GSA perspective, we assume that we have uncertainty on the following 6 input variables: τ_0 (transmission rate before the epidemic outbreak), μ

(rate of the exponential decay), N, η, ν (days until symptoms and recovery) and $\chi 2$ (which impacts the initial conditions of the dynamic model (see, Eq. (21))). These inputs are assumed to be independent and follow uniform distributions with ranges consistent with the values from [MW20], i.e. $\tau_0 \sim \mathcal{U}([5.9 \times 10^{-9}, 6.1 \times 10^{-9}])$, $\mu \sim \mathcal{U}([0.028, 0.036])$, $N \sim \mathcal{U}([8, 15])$, $1/\eta \sim \mathcal{U}([5, 9])$, $1/\nu \sim \mathcal{U}([5, 9])$ and $\chi 2 \sim \mathcal{U}([0.32, 0.4])$. The weighted sum of kernels k_{Σ_w} is used to estimate the global $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ of each input. The results obtained with 50 repetitions of a sample of size $n = 200$, with $N_T = 100$ and for different basis sizes q are reported in Figure 12. For both compartments (I and R), the most influential input is $\chi 2$, as expected since it defines the initial values. However, the influence and the ranking of the other inputs differs according to the output. Concerning the inputs related to the transmission rate, μ has a strong influence on I and less on R, while N is influential on both. Finally, ν has a strong impact on compartment I and negligible on R, conversely to η which impacts only R. This remark is consistent with the model equations (Eq. (20)).

In addition, Figure 13 shows the evolution of the indices according to the sample sizes n and for a fixed basis size $q = 8$. We note that the interquartile range decreases as n increases, becoming minimal from $n = 200$. We also note that a good ranking of the inputs is ensured from $n \geq 150$.

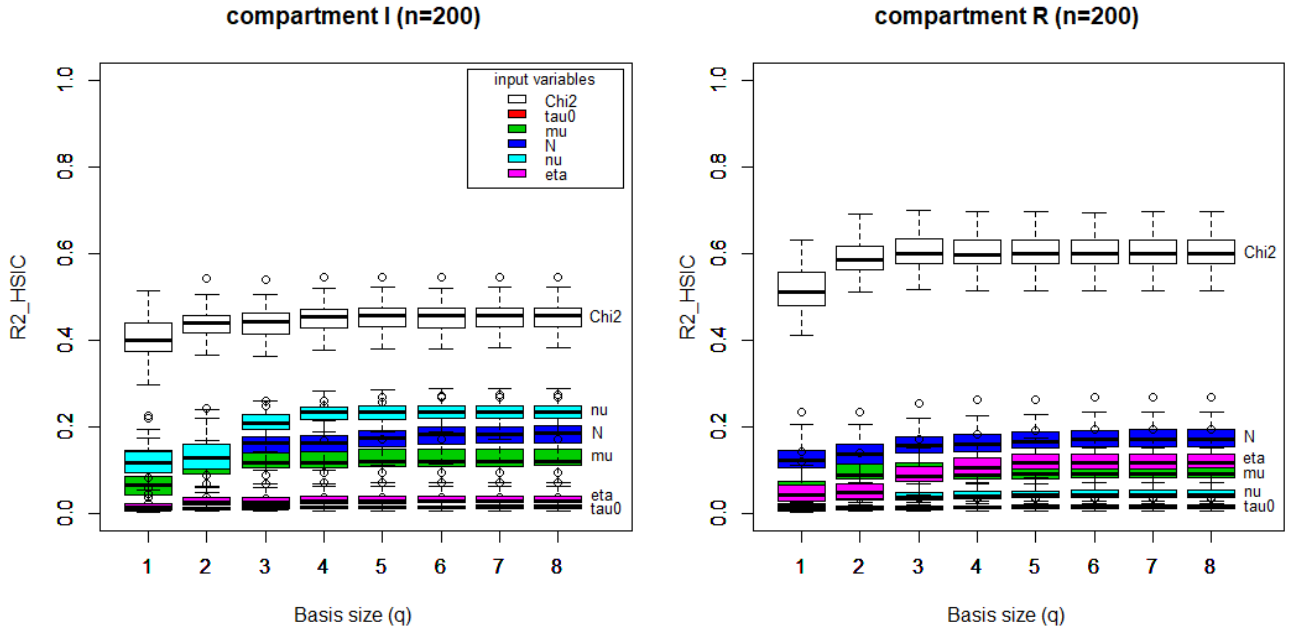

Figure 12: Temporal SIR model. Estimated $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ sensitivity indices for compartments I (left) and R (right), $n = 200$, 50 replicates.

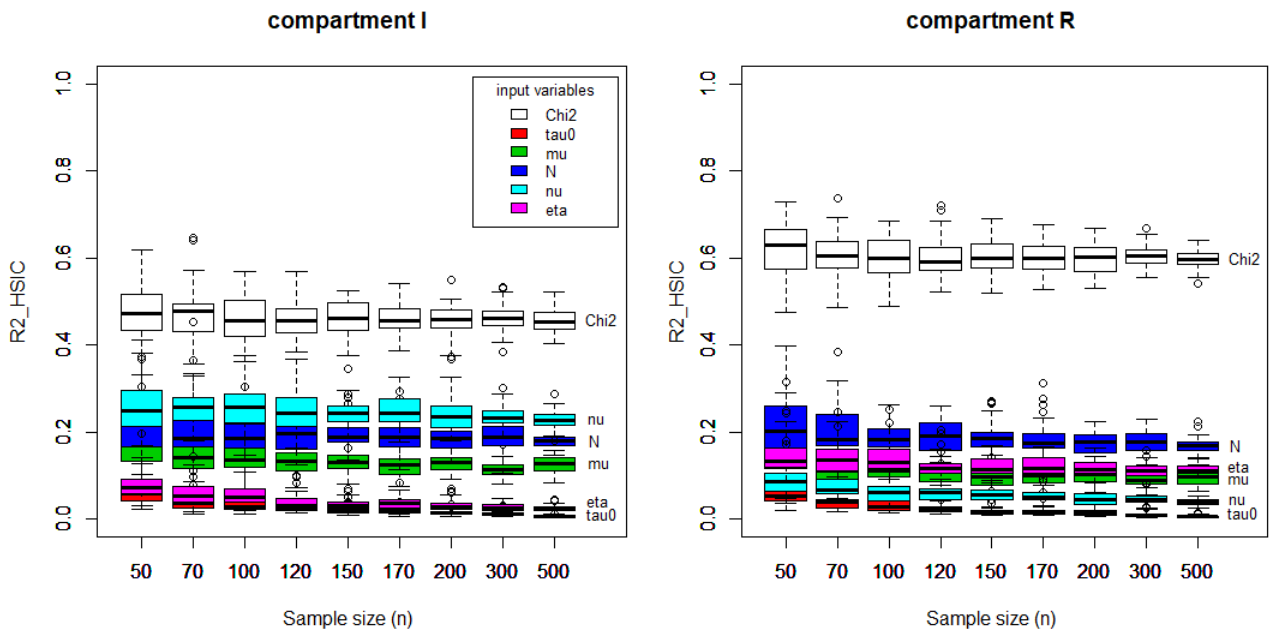

Figure 13: Temporal SIR model. Estimated $\widehat{\mathrm{R}}^2_{\mathrm{HSIC,PCA}}$ sensitivity indices for compartments I (left) and R (right) according to different sample sizes n, 50 replicates.

6. Conclusions & perspectives

In this paper, new improvements have been proposed to extend the applicative spectrum of HSIC and improve their use for screening and ranking purposes, in the context of global sensitivity analysis of numerical simulators.

First, for screening purpose, three methods to perform the HSIC-based independence tests have been presented and compared. On the one hand, we have recalled two main methods existing in the literature, namely the permutation-based method and the Gamma approximation, for non-asymptotic and asymptotic framework, respectively. On the other hand, we have proposed a new test based on Pearson type III approximation. Through numerical examples, we have highlighted the advantages and disadvantages of the different methods. The permutation-based method gives very satisfactory results regardless the sample size, but might be very expensive in terms of CPU time. In contrast, the proposed approximation based on Pearson Type III distribution is much less expensive, while providing as efficient results as the permutation-based method, even for small sample sizes. Finally, the use of Gamma approximation is to be reserved for large samples, for which the asymptotic framework is reached.

Second, we have addressed the problem of extending the HSIC-based tests to samples that are not purely random, and more specifically to samples from Space-Filling Designs (SFD). We have highlighted the impact of some of these designs on the level of the aforementioned usual tests. Indeed, the prescribed level is correct for the simple Latin hypercube, but this is no longer the case with a low-discrepancy Quasi-Monte Carlo sequence or an optimized Latin Hypercube, since the hypothesis of independence of the observations is no longer verified. To overcome this problem, a new algorithm based on Conditional Randomization Test (CRT) has been proposed to correct the HSIC-based independence tests built-upon scrambled-Sobol sequences. Note that this type of design which is part of low-discrepancy Quasi-Monte Carlo sequences is very popular in computer experiments and sensitivity analysis communities. Thanks to the proposed CRT test, the right level is recovered for this design, while keeping the correct level also if we apply it to a pure Monte Carlo sample or a simple Latin Hypercube.

In addition for the independence tests, we have proposed a fast and an efficient selection of the kernel bandwidth parameters of the HSIC, in order to avoid a heuristic choice of these parameters (which is the current practice). For this, a closed-form of the gradient of the HSIC has been derived with respect to these parameters and a new test has been proposed, according to the optimal value of these parameters. Through

numerical tests, the procedure with optimal bandwidths has been shown to outperform the usual HSIC test (with standard choices for kernel parameters) and other existing tests, for several mechanism of dependence.

Finally, we have introduced a new kernel more suitable for the application of HSIC on functional output. This kernel relies on both functional principal component analysis and a relevant weighted combination of (one-dimensional) kernels applied to each component. We have highlighted the performance of the resulting kernel through two analytical examples, by showing that it led to a ranking of the inputs more consistent with the real influence of the variables. Of course, the relevance of this kernel defined for functional data needs to be further evaluated on other datasets (real data among others), as well as on functional inputs. Moreover, its adaptation with other dimension reduction techniques, such as those based on B-splines or wavelets basis, as well as its comparison with DTW-based kernels, also seems to us an interesting line of research.

For future work, the recent developments of [DV21] around the HSIC will also be studied with the greatest interest. The author proposes new multivariate kernels for the formulation of HSIC-based indices, which pave the way for a more relevant interpretation of the indices (definition of indices of the first and total order made possible). Further investigations and experiments should be performed to better understand the behavior of these new indices, for ranking but also screening purposes, and the key role played by the specific kernels used.

References

[ALMM21] Mélisande Albert, Béatrice Laurent, Amandine Marrel, and Anouar Meynaoui. Adaptive test of independence based on hsic measures. To appear in Annals of Statistics, 2021.

[Bhu18] Bharath Bhushan Damodaran. Fast Optimal Bandwidth Selection for RBF Kernel using Reproducing Kernel Hilbert Space Operators for Kernel Based Classifiers. *arXiv e-prints*, page arXiv:1804.05214, April 2018.

[BS19] Thomas B Berrett and Richard J Samworth. Nonparametric independence testing via mutual information. *Biometrika*, 106(3):547–566, 2019.

[CFJL16] Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection. *arXiv preprint arXiv:1610.02351*, 2016.

[CFS99] Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Functional linear model. *Statistics & Probability Letters*, 45(1):11–22, 1999.

[CP76] Roy Cranley and Thomas NL Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976.

[CVBM07] Marco Cuturi, Jean-Philippe Vert, Oystein Birkenes, and Tomoko Matsui. A kernel for time series based on global alignments. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 2, pages II–413. IEEE, 2007.

[DCI13] Guillaume Damblin, Mathieu Couplet, and Bertrand Iooss. Numerical studies of space-filling designs: optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*, 7(4):276–289, 2013.

[DLM16] Matthias De Lozzo and Amandine Marrel. New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, 86(15):3038–3058, 2016.

[DV15] Sébastien Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305, 2015.

[DV21] Sébastien Da Veiga. Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. *arXiv preprint arXiv:2101.05487*, 2021.

[EAM] Mohamed Reda El Amri and Amandine Marrel. Optimized hsic-based tests for sensitivity analysis: Application to thermalhydraulic simulation of accidental scenario on nuclear reactor. *Quality and Reliability Engineering International*.

[GBSS05] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.

[GFT+08] Arthur Gretton, Kenji Fukumizu, Choon H Teo, Le Song, Bernhard Schölkopf, and Alex J Smola. A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592, 2008.

[GGKL20] Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global Sensitivity Analysis: a new generation of mighty estimators based on rank statistics. *Preprint hal-02474902*, 2020.

[HHY95] Carl M Harris, Karla L Hoffman, and Leslie-Ann Yarrow. Obtaining minimum-correlation latin hypercube sampling plans using an ip-based heuristic. *Operations-Research-Spektrum*, 17(2-3):139–148, 1995.

[Hoe48] Wassily Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325, 1948.

[Jac93] Donald A Jackson. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214, 1993.

[JPH08] Julie Josse, Jérome Pagès, and François Husson. Testing the significance of the rv coefficient. *Computational Statistics & Data Analysis*, 53(1):82–91, 2008.

[KAHSL95] Frédérique Kazi-Aoual, Simon Hitier, Robert Sabatier, and Jean-Dominique Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational statistics & data analysis*, 20(6):643–656, 1995.

[LK10] Olivier Le Maître and Omar M. Knio. *Spectral Methods for Uncertainty Quantification*. Scientific Computation. Springer, Dordrecht, 2010.

[Loh96] Wei-Liem Loh. On latin hypercube sampling. *Annals of Statistics*, 24(5):2058–2080, 1996.

[MBC79] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[Mey19] Anouar Meynaoui. *New developments around dependence measures for sensitivity analysis: application to severe accident studies for generation IV reactors*. PhD thesis, University of Toulouse, 2019.

[MW20] Pierre Magal and Glenn Webb. Predicting the number of reported and unreported cases for the covid-19 epidemic in south korea, italy, france and germany. *Available at SSRN: https://ssrn.com/abstract=3557360 or http://dx.doi.org/10.2139/ssrn.3557360*, 2020.

[Nie92] Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[Owe94] Art B Owen. Controlling correlations in latin hypercube samples. *Journal of the American Statistical Association*, 89(428):1517–1522, 1994.

[Owe95] Art B Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pages 299–317. Springer, 1995.

[Owe97] Art B Owen. Monte carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997.

[Owe13] Art B Owen. Monte carlo theory, methods and examples. *Monte Carlo Theory, Methods and Examples. Art Owen*, 2013.

[Ram06] James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

[SC78] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

[Sob93] Ilya M Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments*, 1(4):407–414, 1993.

[Ste87] Michael Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.