Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2023) 237-250

ECCOMAS Proceedia

UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12-14 June 2023

NEW DEPENDENT MEASURES OF ASSOCIATION BETWEEN DYNAMIC MODEL OUTPUTS AND INPUTS USING KERNELS

M. Lamboni[1,2]

[1] University of Guyane, DFRST, 97346 Cayenne, French Guiana, France
[2] 228-UMR Espace-Dev, Univ. of Guyane, Univ. of Réunion, IRD, Univ. of Montpellier, France
e-mail: matieyendou.lamboni[at] gmail.com / matieyendou.lamboni[at] univ-guyane.fr

Abstract. *We propose new (in)-dependent measures for assessing the impacts of the model inputs on dynamic outputs that make use of the asymmetric role between the model outputs and inputs and satisfy the Renyi axioms. Our dependent measures, which result from the test statistics of independence between dynamic outputs and the model inputs, rely on kernel methods. Such measures give a theoretical and statistical foundation of multivariate sensitivity analysis and allow its extension thanks to sensitivity functionals (SFs) and kernel methods. We provide kernel-based sensitivity indices for dynamic models or time-depending processes that i) cope with every distribution of SFs, ii) account for the necessary or desirable moments of SFs and the interactions among the model inputs.*

Keywords: Independence tests, Kernel methods, Sensitivity functionals and indices, Uncertainty quantification, RKHS.

doi: 10.7712/120223.10335.19568

1 Introduction

Human-induced dynamic systems are widely encountered in engineering. Such systems are often represented by complex dynamic models or time-depending processes, which are used for better understanding interactions and drivers that affect predictions of a wide range of problems and issues. Such mathematical representations of physical realities are often non-linear with important interactions among their components, and uncertainty arises in many guises such as the systemic uncertainty, the model discrepancy, the distribution of the model inputs (parameters and input variables).

In uncertainty quantification, multivariate sensitivity analysis (MSA) ([1, 2, 3, 4, 5, 6, 7, 8]) and dependent MSA (dMSA) ([9, 10]) are valuable tools for identifying unnecessary inputs regarding their effects on the model outputs, and for assessing interactions among the main drivers of complex dynamic models. Since MSA and dMSA make use of the variance-covariance as the importance measure, such approaches account only for the second-order moments of sensitivity functionals (SFs) and do not exist when a SF follows the Cauchy distribution for instance. Sensitivity functionals for dynamic models are time random processes and contain different information brought by each input or group of inputs on the whole model outputs.

In this paper, we develop new (in)-dependent measures of association between dynamic model outputs and independent inputs that i) cope with every distribution of SFs, ii) account for the necessary or desirable moments of SFs, and iii) account for interactions among the model inputs. Using the asymmetric role of dynamic outputs and the model inputs, Section 3 deals with the statistical test of independence between the dynamic model outputs and independent inputs using SFs and kernel methods. Formally, our approach consists in embedding SFs into the appropriate Reproducing Kernel Hilbert Space (RKHS) thanks to symmetric and positive definite kernels ([11, 12]). When the model outputs depend on some inputs, the associated test statistics can be used as the crude dependent measures. In Section 4, we propose new dependent measures that satisfy the Renyi axioms ([13]) keeping in mind the asymmetric role between the model outputs and inputs. Our new dependent measures give a theoretical and statistical foundation of MSA and generalize MSA. Moreover, such dependent measures lead to the definitions of the kernel-based sensitivity indices (Kb-SIs) for dynamic models or time-depending processes. We distinguish the first-order and total Kb-SIs with the former index less than the latter. Empirical Kb-SIs are provided for computational issues in Section 5, and a case study is provided in Section 6. We conclude this work in Section 7.

General notation

For an integer $d > 0$, we use $\mathbf{X} := (X_1, \ldots, X_d)$ for a random vector of the model inputs. For $u \subseteq \{1, \ldots, d\}$, we use $\mathbf{X}_u := (X_j, \forall\, j \in u)$; $\mathbf{X}_{\sim u} := (X_j, \forall\, j \in \{1, \ldots, d\} \setminus u)$ and we have the following partition $\mathbf{X} = (\mathbf{X}_u, \mathbf{X}_{\sim u})$. We use $||\cdot||_p$ for the p-norm, $\mathbb{E}[\cdot]$ for the expectation and $\mathbb{V}[\cdot]$ for the variance-covariance.

2 Preliminary: kernel methods

The theory of RKHS and kernel methods have been successfully used for testing and distinguishing two different distributions of random vectors by embedding such random vectors into the appropriate RKHS or feature spaces ([14, 11, 15, 16]).

Consider an arbitrary space $\mathcal{X}$, an Hilbert space $\mathcal{H}$ endowed with the inner product $\langle \cdot, \cdot \rangle$. The function $\phi : \mathcal{X} \to \mathcal{H}$ given by $\phi(x) =: k(\cdot, x)$ is called a feature map; and the function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ given by

$$k(x, x') = \langle k(\cdot, x),\, k(\cdot, x') \rangle \ ,$$

is called a valid kernel, that is, a symmetric and positive-definite kernel ([14]). Thus, for a random variable Y having F as the cumulative distribution function (i.e., $Y \sim F$), the transformation $Y \mapsto k(\cdot, Y)$ aims at embedding Y into the RKHS induced by the kernel k. The mean element defined by ([15, 16, 17])

$$\mu_F(Y) := \mathbb{E}\left[k(\cdot, Y)\right] \ ,$$

is a linear statistic in the new RKHS, and it accounts for all the moments or desirable moments of Y depending on the kernel k. The mean element can be used for entirely characterizing the distributions of random variables, and such characterization is often given as follows: for all $Y \sim F$ and $Z \sim G$,

$$\mu_F(Y) = \mu_G(Z) \Longleftrightarrow F = G\,,$$

for only the set of characteristic kernels. It means that the mean element associated with a characteristic kernel is one to one. By taking the distance between $\mu_G(Z)$ and $\mu_F(Y)$ in the RKHS, we obtain the maximum mean discrepancy (MMD), that is,

$$MMD^2(Y, Z) := ||\mu_F(Y) - \mu_G(Z)||_{\mathcal{H}}^2 \ .$$

Thus, if k is a characteristic kernel, then we can discriminate the distribution of Y and Z using $MMD^2(Y, Z)$.

3 Testing independence between inputs and dynamic outputs

3.1 Dynamic outputs and time-depending sensitivity functionals

For a given time period T, consider $t \in [0,\, T]$ and a dynamic model $f(\cdot, t) : \mathbb{R}^d \to R$ that includes d independent random inputs $\mathbf{X} := (X_1, \ldots, X_d)$ (i.e., assumption (A1)) and provides $(f(\mathbf{X}, t),\, \forall\, t \in [0,\, T])$ as dynamic outputs. Since $f(\mathbf{X}, t) \in \mathbb{R}$, the SFs associated with the inputs $\mathbf{X}_u$ with $u \subseteq \{1, \ldots, d\}$ are given below. Under assumption (A1), the first-order and total SFs of $\mathbf{X}_u$ are respectively defined by ([18, 7, 19])

$$f_u^{fo}(\mathbf{X}_u, t) := \mathbb{E}\left[f(\mathbf{X}, t)|\mathbf{X}_u\right] - \mathbb{E}\left[f(\mathbf{X}, t)\right], \qquad \forall\, t \in [0,\, T]\,;$$

$$f_u^{tot}(\mathbf{X}, t) := f(\mathbf{X}, t) - \mathbb{E}\left[f(\mathbf{X}, t)|\mathbf{X}_{\sim u}\right], \qquad \forall\, t \in [0,\, T]\,.$$

When $u = \{1, \ldots, d\}$, the total SF comes down to the centered model outputs, that is,

$$f^c(\mathbf{X}, t) := f(\mathbf{X}, t) - \mathbb{E}\left[f(\mathbf{X}, t)\right], \qquad \forall\, t \in [0,\, T]\,,$$

It is worth noting that $\left(f_u^{fo}(\mathbf{X}_u, t),\, t \in [0,\, T]\right)$; $(f_u^{tot}(\mathbf{X}, t),\, t \in [0,\, T])$ and $(f^c(\mathbf{X}, t),\, t \in [0,\, T])$ are zero-mean random processes, whose components may be correlated or dependent. For Gaussian SFs (random processes), the Karhunen-Loeve expansion or functional principal component analysis allows for concentrating some information contained in such random processes in few uncorrelated components ([20, 21]). Recall that the Karhunen-Loeve expansion relies on the theory of RKHS.

3.2 Independence tests between inputs and outputs

For a given output identified by $t_0 \in [0,\, T]$ (i.e., $f(\mathbf{X}, t_0)$), it is known in [22] (Proposition 1) that under (A1), $f(\mathbf{X}, t_0)$ does not depend on $\mathbf{X}_u$ if and only if $f_u^{tot}\left((\mathbf{X}, t_0) = 0\right.$. Therefore, the initial null hypothesis of our statistical test of independent between $\mathbf{X}_u$ and the dynamic outputs is given by

$$f_u^{tot}\left((\mathbf{X}, t) = 0,\; \forall\, t \in [0,\, T] \quad a.s.\,,\right.$$

because some inputs can contribute to the model outputs at the beginning or at the end of the phenomenon of interest ([1]). Such null hypothesis is equivalent to

$$\mathbb{V}\left[f_u^{tot}\left((\mathbf{X}, t)\right] = 0,\; \forall\, t \in [0,\, T]\,,\right.$$

when the total SF is a Gaussian random process. In what follows, we are going to embed the total SF into a RKHS so as to account for desirable or necessary moments. To that end, we assume that

(A2) $\int_0^T \mathbb{E}\sqrt{k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X},t\right)\right)}\, dt < +\infty$ for all $u \subseteq \{1,\ldots,d\}$;
and we need a new concept (see Definition 1).

Definition 1 *Let $\mathbf{X}'$ be an i.i.d. copy of $\mathbf{X}$ and assume that (A1)-(A2) hold. A kernel k is said to be equivalent for the independence test between $\mathbf{X}_u$ and $(f(\mathbf{X},t),\, t \in [0,\, T])$ whenever $\forall\, t \in [0,\, T]$*

$$\mathbb{E}\left[k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X}',t\right)\right)\right] - k(0,0) = 0 \Longrightarrow f_u^{tot}\left(\mathbf{X},t\right) = 0 \; a.s.\,.$$

We can see that the quadratic kernel given by $k_q(y,y') := y^2(y')^2$ and the absolute kernel given by $k_a(y,y') := |y||y'|$ are both equivalent kernels for the independence test.

For the set of equivalent kernels for the independence test (i.e., $\mathcal{K}_E$), the kernel-based test hypotheses are given as follows: $\forall\, k \in \mathcal{K}_E$ and $\forall\; q > 0$,

$$\begin{cases} H_0: & \int_0^T \left|\mathbb{E}\left[k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X}',t\right)\right)\right] - k(0,0)\right|^q\, dt = 0 \\ H_1: & \int_0^T \left|\mathbb{E}\left[k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X}',t\right)\right)\right] - k(0,0)\right|^q\, dt \neq 0 \end{cases}. \tag{1}$$

To derive the test statistic, we are given two independent samples $\{f_u^{tot}\left(\mathbf{X}_i,t\right)\}_{i=1}^m$ and $\{f_u^{tot}\left(\mathbf{X}_i',t\right)\}_{i=1}^m$ from $f_u^{tot}\left(\mathbf{X},t\right)$. The usual, unbiased and consistent estimator of $\mu_k^{tot}(t) := \mathbb{E}\left[k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X}',t\right)\right)\right]$ is given as follows: $\forall\, t \in [0,\, T]$,

$$\widehat{\mu_k^{tot}}(t) := \frac{1}{m}\sum_{i=1}^m k\left(f_u^{tot}\left(\mathbf{X}_i,t\right), f_u^{tot}\left(\mathbf{X}_i',t\right)\right)\,.$$

The law of large numbers (LLN) and the central limit theorem (CLT) ensure that

$$\widehat{\mu_k^{tot}}(t) \xrightarrow{P} \mu_k^{tot}(t); \qquad \sqrt{m}\left(\widehat{\mu_k^{tot}}(t) - \mu_k^{tot}(t)\right) \xrightarrow{D} \mathcal{N}\left(0,\, \sigma_k^{tot}(t)\right)\,, \tag{2}$$

where $\sigma_k^{tot}(t) := \mathbb{V}\left[k\left(f_u^{tot}\left(\mathbf{X},t\right),\, f_u^{tot}\left(\mathbf{X}',t\right)\right)\right]$. In general, the variance is unknown, and it can be estimated using the formula

$$\widehat{\sigma_k^{tot}}(t) := \frac{1}{m-1}\sum_{i=1}^m \left(k\left(f_u^{tot}\left(\mathbf{X}_i,t\right), f_u^{tot}\left(\mathbf{X}_i',t\right)\right) - \widehat{\mu_k^{tot}}(t)\right)^2 \xrightarrow{P} \sigma_k^{tot}(t)\,.$$

But, under the null hypothesis H_0, the unbiased and consistent estimator of $\sigma_k^{tot}(t)$ is given by

$$\widehat{\sigma_{k,H_0}^{tot}}(t) := \frac{1}{m}\sum_{i=1}^m \left(k\left(f_u^{tot}\left(\mathbf{X}_i,t\right), f_u^{tot}\left(\mathbf{X}_i',t\right)\right) - k(0,0)\right)^2\,.$$

Based on these elements, the test statistic under the null hypothesis and its asymptotic distribution are provided in Corollary 1.

Corollary 1 *Let $G(t) \sim \mathcal{N}\left(0,\, \widehat{\sigma_{k,H_0}^{tot}}(t)\right)$, $t \in [0,\, T]$ be the Gaussian random process and assume (A1)-(A2) hold. If $k \in \mathcal{K}_E$ (i.e., (A3)), then the test statistic is given by*

$$T_{k,H_0}^{tot} := m \int_0^T \left| \widehat{\mu_k^{tot}(t)} - k(0,0) \right|^2 \, dt \xrightarrow{D} \sum_{k=1}^{+\infty} \lambda_k Z_k^2 \,, \tag{3}$$

where $\{Z_k \sim \mathcal{N}(0,1)\}_{k=1}^{\infty}$ are i.i.d. and $\{\lambda_k\}_{k=1}^{\infty}$ are the eigenvalues related to the Karhunen-Loeve expansion of $(G(t),\, t \in [0,\, T])$.

Proof. *See Appendix A.*

□

In case of independence between inputs and outputs, the test statistic T_{k,H_0}^{tot} vanishes using either the total SF or the first-order SF. Therefore, we are going to use a part of T_{k,H_0}^{tot} as the measure of deviation from independence.

Remark 1 *When the analytical expression of the total SF of $\mathbf{X}_u$ is not available, the following plug-in estimators are going to be used. The usual consistent estimator of $\mathbb{E}_{\mathbf{X}_u}\left[f(\mathbf{X}_u, \mathbf{X}_{\sim u}, t)\right]$ is given by*

$$\widehat{\mu}(\mathbf{X}_{\sim u}, t) := \frac{1}{m} \sum_{i=1}^{m} f(\mathbf{X}_{i,u}, \mathbf{X}_{\sim u}, t) \xrightarrow{P} \mathbb{E}_{\mathbf{X}_u}\left[f(\mathbf{X}_u, \mathbf{X}_{\sim u}, t)\right] \,,$$

and the estimator of $\mu_k^{tot}(t)$ is given by

$$\widehat{\mu_k^{tot'}(t)} := \frac{1}{m} \sum_{i=1}^{m} k\left(f(\mathbf{X}_i, t) - \widehat{\mu}(\mathbf{X}_{i,\sim u}, t),\; f(\mathbf{X}_i', t) - \widehat{\mu}(\mathbf{X}_{i,\sim u}', t)\right) \,.$$

4 Dependent measures using sensitivity functionals

This section aims at providing the dependent measures between dynamic outputs and the model inputs that satisfy the Renyi axioms ([13]) using the results provided in Corollary 1.

Formally, when the inputs $\mathbf{X}_u$ contribute to dynamic outputs, the quantity

$$\mathcal{D}_k^{1,q}(f_u^{tot}) := \int_0^T \left| \mathbb{E}\left[k\left(f_u^{tot}(\mathbf{X}, t),\, f_u^{tot}(\mathbf{X}', t)\right)\right] - k(0,0) \right|^q \, dt, \quad \forall\, q > 0 \,, \tag{4}$$

is used to measure the deviation from independence. To account for the cross-components (including auto-correlations), the above deviation is extended as follows:

$$\mathcal{D}_k^{2,q}(f_u^{tot}) := \int_0^T \int_0^T \left| \mathbb{E}\left[k\left(f_u^{tot}(\mathbf{X}, t_1),\, f_u^{tot}(\mathbf{X}, t_2)\right)\right] - k(0,0) \right|^q \, dt_1 dt_2 \,. \tag{5}$$

The generalized measure of deviation from independence is introduced below. Consider two random vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ identically distributed as $\mathbf{X}$ and $t_1, t_2 \in [0, T]$, and define

$$\mathcal{D}_k^{3,q}(f_u^{tot}) := \int_0^T \int_0^T \left|\mathbb{E}\left[k\left(f_u^{tot}(\mathbf{X}^{(1)}, t_1),\, f_u^{tot}(\mathbf{X}^{(2)}, t_2)\right)\right] - k(0,0)\right|^q dt_1 dt_2 \,. \tag{6}$$

We can see that $\mathcal{D}_k^{1,q}(f_u^{tot})$ and $\mathcal{D}_k^{2,q}(f_u^{tot})$ are particular cases of $\mathcal{D}_k^{3,q}(f_u^{tot})$. In what follows, we are going to focus on $\mathcal{D}_k^{1,q}(f_u^{tot})$ and $\mathcal{D}_k^{2,q}(f_u^{tot})$ in order to define new dependent measures.

A reasonable measure of deviation from independence (or kernel) should take into account the fact that the first-order SF of $\mathbf{X}_u$ brings partial information compared to the total SF. This leads to the following definition.

Definition 2 *Assume that (A1)-(A3) hold.*
A kernel k is said to be sensitivity-compatible whenever $\forall\, t_1, t_2 \in [0, T]$

$$\begin{aligned} &\left|\mathbb{E}\left[k\left(f_u^{fo}\left(\mathbf{X}_u^{(1)}, t_1\right),\, f_u^{fo}\left(\mathbf{X}_u^{(2)}, t_2\right)\right)\right] - k(0,0)\right| \leq \\ &\left|\mathbb{E}\left[k\left(f_u^{tot}\left(\mathbf{X}^{(1)}, t_1\right),\, f_u^{tot}\left(\mathbf{X}^{(2)}, t_2\right)\right)\right] - k(0,0)\right| \,. \end{aligned}$$

We can check that the quadratic kernel $k_q(y, y') := y^2(y')^2$ and the absolute kernel $k_a(y, y') := |y||y'|$ are both equivalent and sensitivity-compatible kernels thanks to Lemma 1.

Lemma 1 *Consider $\mathcal{D}_k^{1,q}(\bullet)$ and $\mathcal{D}_k^{2,q}(\bullet)$, and assume that (A1)-(A3) hold. If $k(\cdot, x)$ is convex and $k(\cdot, 0) = 0$, then k is a sensitivity-compatible kernel.*

Proof. *Since $f_u^{fo}(\mathbf{X}_u, t) = \mathbb{E}_{\mathbf{X}_{\sim u}}\left[f_u^{tot}(\mathbf{X}, t)\right]$, the result holds by applying the Jensen inequality.*

□

Now, we have all the elements in hand to introduce the new dependent measures between dynamic outputs and the model inputs. For the inputs $\mathbf{X}_u$, the first-type dependent measures are defined as follows:

$$S_{T_u}^{1,k,q} := \frac{\mathcal{D}_k^{1,q}(f_u^{tot})}{\mathcal{D}_k^{1,q}(f^c)} \,, \tag{7}$$

can be seen as the total kernel-based SIs (Kb-SIs) of $\mathbf{X}_u$. We then define the first-order Kb-SIs of $\mathbf{X}_u$ by

$$S_u^{1,k,q} := \frac{\mathcal{D}_k^{1,q}(f_u^{fo})}{\mathcal{D}_k^{1,q}(f^c)} \,. \tag{8}$$

Likewise, the second-type dependent measures and kernel-based SIs of $\mathbf{X}_u$ are defined as follows:

$$S_{T_u}^{2,k,q} := \frac{\mathcal{D}_k^{2,q}(f_u^{tot})}{\mathcal{D}_k^{2,q}(f^c)}\,, \tag{9}$$

$$S_u^{2,k,q} := \frac{\mathcal{D}_k^{2,q}(f_u^{fo})}{\mathcal{D}_k^{2,q}(f^c)}\,. \tag{10}$$

Again, the second-type kernel-based SIs aim at accounting for the cross-components of the time depending SFs. While the linear kernel $k_l(y, y') = yy'$ allows for assessing the auto-correlations of SFs, it is to be noted that such kernel is not a characteristic kernel in general. Indeed, it is a characteristic kernel for the class of Gaussian time processes when using $S_{T_u}^{2,k,q}$ only. For instance, the centered and characteristic kernels that are convex on the support of SFs may be used so as to assess all the moments of SFs. Interesting properties of the kernel-based SIs are derived in the following theorem.

Theorem 1 *Let $k \in \mathcal{K}_E$ be a sensitivity-compatible kernel, and assume (A1)-(A3) hold. Then,*

$$0 \leq S_u^{1,k,q} \leq S_{T_u}^{1,k,q} \leq 1\,; \tag{11}$$

and

$$0 \leq S_u^{2,k,q} \leq S_{T_u}^{2,k,q} \leq 1\,. \tag{12}$$

Proof. *The proof is straightforward using Definition 2 and the conditional Jensen inequality.*

□

It is to be noted that for the quadratic kernel $k_q(y, y') := y^2(y')^2$, the first-type Kb-SIs ($S_\bullet^{1,k_q,1/2}$) are equal to the first-type generalized sensitivity indices (GSIs) ([2, 3]). In the case of the linear kernel $k_l(y, y') = yy'$, the second-type Kb-SIs ($S_\bullet^{2,k_l,2}$) are equivalent to the second-type GSIs ([7, 19]). The absolute kernel $k_a(y, y') := |y||y'|$ leads to new sensitivity indices for dynamic models, and such indices aim at accounting for small variations compared to GSIs. Of course, we can generalize such indices using the l_p-norm.

5 Empirical kernel-based sensitivity indices

Since SFs are analytically unknown in general, we start this section with the estimators of SFs, followed by those of the Kb-SIs. To that end, we are given two independent samples $\{\mathbf{X}_i\}_{i=1}^m$ and $\{\mathbf{X}'_i\}_{i=1}^m$ from $\mathbf{X}$. Keeping in mind the method of moments, let us recall the following consistent estimators:

$$\widehat{\mu}(\mathbf{X}_u, t) := \frac{1}{m}\sum_{i=1}^{m} f(\mathbf{X}_u, \mathbf{X}_{i,\sim u}, t) \xrightarrow{P} \mathbb{E}_{\mathbf{X}_{\sim u}}\left[f(\mathbf{X}, t)\right]\,,$$

$$\widehat{\mu}(\mathbf{X}_{\sim u}, t) := \frac{1}{m}\sum_{i=1}^{m} f(\mathbf{X}_{i,u}, \mathbf{X}_{\sim u}, t) \xrightarrow{P} \mathbb{E}_{\mathbf{X}_u}\left[f(\mathbf{X}, t)\right] ,$$

$$\widehat{\mu}(t) := \frac{1}{m}\sum_{i=1}^{m} f(\mathbf{X}_{i}, t) \xrightarrow{P} \mathbb{E}\left[f(\mathbf{X}, t)\right] .$$

Using the Slutsky theorem, the following consistent estimators of SFs are derived:

$$\widehat{\mu}(\mathbf{X}_u, t) - \widehat{\mu}(t) \xrightarrow{P} f_u^{fo}(\mathbf{X}_u, t) ,$$

$$f(\mathbf{X}, t) - \widehat{\mu}(\mathbf{X}_{\sim u}, t) \xrightarrow{P} f_u^{tot}(\mathbf{X}, t) ,$$

$$f(\mathbf{X}, t) - \widehat{\mu}(t) \xrightarrow{P} f^{c}(\mathbf{X}_u, t) .$$

Using the above estimators of SFs and bearing in mind the plug-in approach, the consistent estimators of Kb-SIs are provided in corollaries 2-3.

Corollary 2 *Let $k \in \mathcal{K}_E$ be a sensitivity-compatible kernel, and assume (A1)-(A3) hold. If the output is observed at $t_\ell \in [0,\, T]$ with $\ell = 1, \ldots, L$, then the estimators of the first-type Kb-SIs are given as follows:*

$$\widehat{S_u^{1,k,q}} := \frac{\sum_{\ell=1}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(\widehat{\mu}(\mathbf{X}_{i,u}, t_\ell) - \widehat{\mu}(t_\ell),\, \widehat{\mu}(\mathbf{X}'_{i,u}, t_\ell) - \widehat{\mu}(t_\ell)\right) - k(0,0)\right|^q}{\sum_{\ell=1}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_\ell) - \widehat{\mu}(t_\ell),\, f(\mathbf{X}'_{i}, t_\ell) - \widehat{\mu}(t_\ell)\right) - k(0,0)\right|^q} ;$$

$$\widehat{S_{T_u}^{1,k,q}} := \frac{\sum_{\ell=1}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_\ell) - \widehat{\mu}(\mathbf{X}_{i,\sim u}, t_\ell),\, f(\mathbf{X}'_{i,u}, t_\ell) - \widehat{\mu}(\mathbf{X}'_{i,\sim u}, t_\ell)\right) - k(0,0)\right|^q}{\sum_{\ell=1}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_\ell) - \widehat{\mu}(t_\ell),\, f(\mathbf{X}'_{i}, t_\ell) - \widehat{\mu}(t_\ell)\right) - k(0,0)\right|^q} .$$

Likewise, Corollary 3 provides the estimators of the second-type Kb-SIs.

Corollary 3 *Let $k \in \mathcal{K}_E$ be a sensitivity-compatible kernel, and assume (A1)-(A3) hold. If the output is observed at $t_\ell \in [0,\, T]$ with $\ell = 1, \ldots, L$, then the estimators of the second-type Kb-SIs are given as follows:*

$$\widehat{S_u^{2,k,q}} := \frac{\sum_{\substack{\ell_1=1\\ \ell_2=1}}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(\widehat{\mu}(\mathbf{X}_{i,u}, t_{\ell_1}) - \widehat{\mu}(t_{\ell_1}),\, \widehat{\mu}(\mathbf{X}_{i,u}, t_{\ell_2}) - \widehat{\mu}(t_{\ell_2})\right) - k(0,0)\right|^q}{\sum_{\substack{\ell_1=1\\ \ell_2=1}}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_{\ell_1}) - \widehat{\mu}(t_{\ell_1}),\, f(\mathbf{X}_{i}, t_{\ell_2}) - \widehat{\mu}(t_{\ell_2})\right) - k(0,0)\right|^q} ;$$

$$\widehat{S_{T_u}^{2,k,q}} := \frac{\sum_{\substack{\ell_1=1\\ \ell_2=1}}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_{\ell_1}) - \widehat{\mu}(\mathbf{X}_{i,\sim u}, t_{\ell_1}),\, f(\mathbf{X}_{i,u}, t_{\ell_2}) - \widehat{\mu}(\mathbf{X}_{i,\sim u}, t_{\ell_2})\right) - k(0,0)\right|^q}{\sum_{\substack{\ell_1=1\\ \ell_2=1}}^{L} \left|\frac{1}{m}\sum_{i=1}^{m} k\left(f(\mathbf{X}_{i}, t_{\ell_1}) - \widehat{\mu}(t_{\ell_1}),\, f(\mathbf{X}_{i}, t_{\ell_2}) - \widehat{\mu}(t_{\ell_2})\right) - k(0,0)\right|^q} .$$

It is to be noted that the generic expressions of the estimators provided in corollaries 2-3 can have simple expressions for specific kernels. Moreover, efficient estimators can be deduced for specific kernels such as the quadratic kernel.

6 Test case study

We consider the mean component of the dynamic and stochastic risk model used in [23] and provided in Appendix B. Such model simulates the concentration of $L.\ monocytogenes$ in a cheese at an hourly time-step under some technical assumptions or crude simplifications (see [23] for more details). The model includes four input factors such as the temperatures during the three main periods of cheeses processing and the initial concentration of $L.\ monocytogenes$. The new uncertainties on these input factors are listed in Table 1. Moreover, only the outputs at days $1, 2, \ldots, 28$ are considered in this paper. The Kb-SIs associated with the absolute kernel, the linear and quadratic kernels are listed in Table 2.

Inputs	Description	Unit	Probability distribution
T_1	Temperature of first period	^oC	Truncated normal $\mathcal{N}(5,5)$ on $[-1.72, 45.5]$
T_2	Temperature of second period	^oC	Truncated normal $\mathcal{N}(7,7)$ on $[-1.72, 45.5]$
T_3	Temperature of third period	^oC	Truncated normal $\mathcal{N}((8,8)$ on $[-1.72, 45.5]$
C_0	Initial concentration	log CFU	Uniform $\mathcal{U}(-4,1)$

Table 1: Uncertainties on the input factors. In $\mathcal{N}(5,10)$, we have mean=5, sd=10.

Kernels	T_1	T_2	T_3	C_0	Type of Kb-SIs
First-order Kb-SIs					
Absolute	0.001	0.001	0.341	0.592	$S_j^{1,k_a,1/2}$
Quadratic	0.001	0.001	0.494	0.500	$S_j^{1,k_q,1/2}$
Linear	0.000	0.000	0.156	0.143	$S_j^{2,k_l,2}$
Total Kb-SIs					
Absolute	0.000	0.000	0.340	0.593	$S_{T_j}^{1,k_a,1/2}$
Quadratic	0.000	0.000	0.495	0.502	$S_{T_j}^{1,k_q,1/2}$
Linear	0.000	0.000	0.156	0.144	$S_{T_j}^{2,k_l,2}$

Table 2: Kernel-based sensitivity indices for the dynamic risk model using $m = 1000$.

The three indices show that T_3 and C_0 are the most influential input variables. According to the absolute kernel and the first-type Kb-SIs, the initial concentration is the most important driver of the concentration of $L.\ monocytogenes$ in a cheese, while the linear kernel (second-type indices) and the quadratic kernel (first-type) show that both inputs have the same importance regarding our ability to control $L.\ monocytogenes$ in a cheese.

7 Conclusion

In this paper, we have proposed the kernel-based (in)-dependent measures between the dynamic model outputs and the model inputs based on the statistical tests of independence. Using such dependent measures, two types of kernel-based sensitivity indices are provided, including their estimators. These new indices extend generalized sensitivity indices introduced in [1, 2] and are flexible enough to i) account explicitly for desirable and necessary moments, including the auto-correlations of sensitivity functionals; ii) cope with every distribution of sensitivity functionals.

Numerical results show that different ranking of inputs can be obtained using different kernels. Among the set of kernels that are equivalent for the independence tests and sensitivity-compatible, any choice of a kernel should be based on what the modelers want to account for and the statistical properties of sensitivity functionals. In next future, it is worth investigating the extension of the proposed methods to cope with dependent or correlated inputs thanks to the dependency models provided in [9, 10, 24].

Appendix A Proof of Corollary 1

Since $\left(\sqrt{m}(\widehat{\mu_k^{tot}(t)} - k(0,0)),\, t \in [0,\, T]\right)$ is a zero-mean and asymptotically Gaussian process (see Equation 2), the result holds using the Karhunen-Loeve expansion or knowing that the squared norm of a zero-mean Gaussian process can be represented as $\sum_{k=1}^{+\infty} \lambda_k Z_k^2$ (see [20], Chapter 1).

Appendix B Dynamic risk model

The dynamic risk model is based on the following equations that involve the characteristics of $L.\, monocytogenes$ and the cheeses processing. The model output at a given time t (i.e., $Y(t)$) is the log-concentration of $L.\, monocytogenes$ (CFU) in a cheese, and it is given by

$$Y(t) = 9 - \log_{10}\left[1 + \left(\frac{10^9}{10^{C_0}} - 1\right)\exp(-r(t) * t)\right] ,$$

where the growth rate (i.e., $r(t)$) is the product of the optimal growth rate (i.e., $\mu_{opt} = 1/4$), the factors of temperature (i.e., $\gamma(T)$), pH (i.e., $\gamma(pH)$) and water activity (i.e., $\gamma(aw)$). The growth rate $r(t)$ is defined by

$$r(t) = \frac{1}{4} \times \gamma(T(t)) \times \gamma(pH(t)) \times \gamma(aw),$$

where $T(t)$ and $pH(t)$ are the temperature and pH at time t, respectively. We use $T_1 = T(t)$ from days 0 to 7; $T_2 = T(t)$ from days 8 to 14; $T_3 = T(t)$ from days 15 to 28; and

$$\gamma(T(t)) = \frac{(T(t) - 45.5)(T(t) + 1.72)^2}{(38.72\,[38.72(T(t) - 37) + 8.5(35.28 - 2T(t))])},$$

$$\gamma(pH(t)) = \begin{cases} \frac{(pH(t)-9.61)(pH(t)-4.71)}{2.59(pH(t)-7.1)+2.51(4.71-pH(t))} & if\, 4.71 < pH(t) < 9.61 \\ 0 & \text{otherwise} \end{cases},$$

with

$$pH(t) = -1.14 \times 10^{-9}t^3 + 3.04 \times 10^{-6}t^2 - 5.2 \times 10^{-4}t + 4.58\,, \qquad \gamma(aw) = 0.91666\,.$$

REFERENCES

[1] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod, Multivariate global sensitivity analysis for dynamic crop models, Fields Crop Reasearch 113 (2009) 312–320.

[2] M. Lamboni, H. Monod, D. Makowski, Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models, Reliability Engineering and System Safety 96 (2011) 450–459.

[3] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, Sensitivity indices for multivariate outputs, Comptes Rendus Mathematique 351 (7) (2013) 307–310.

[4] S. Xiao, Z. Lu, P. Wang, Multivariate global sensitivity analysis for dynamic models based on wavelet analysis, Reliability Engineering & System Safety 170 (2018) 20–30.

[5] A. Alexanderian, P. A. Gremaud, R. C. Smith, Variance-based sensitivity analysis for time-dependent processes, Reliability Engineering & System Safety 196 (2020) 106722.

[6] M. B. Heredia, C. Prieur, N. Eckert, Nonparametric estimation of aggregated sobol' indices: Application to a depth averaged snow avalanche model, Reliability Engineering & System Safety 212 (2021) 107422.

[7] M. Lamboni, Multivariate sensitivity analysis: Minimum variance unbiased estimators of the first-order and total-effect covariance matrices, Reliability Engineering & System Safety 187 (2019) 67 – 92.

[8] T. Perrin, O. Roustant, J. Rohmer, O. Alata, J. Naulin, D. Idier, R. Pedreros, D. Monculon, P. Tinard, Functional principal component analysis for global sensitivity analysis of model with spatial output, Reliability Engineering & System Safety 211 (2021) 107522.

[9] M. Lamboni, S. Kucherenko, Multivariate sensitivity analysis and derivative-based global sensitivity measures with dependent variables, Reliability Engineering & System Safety 212 (2021) 107519.

[10] M. Lamboni, On dependent generalized sensitivity indices and asymptotic distributions, arXiv preprint arXiv2104.12938 (2022).

[11] A. Berlinet, C. Thomas, T. A. Gnan, Reproducing Kernel Hilbert Space in probability and statistics, Kluwer Academic, 2004.

[12] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, B. Schölkopf, Kernel methods for measuring independence, Journal of Machine Learning Research 6 (2005) 2075–2129.

[13] A. Renyi, On measures of dependence, Acta Mathematica Academiae Scientiarum Hungarica 10 (3-4) (1959) 441–451.

[14] N. Aronszajn, Theory of reproducing kernels, Transactions of the American Mathematical Society 68 (1950) 337–404.

[15] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola, A kernel method for the two-sample-problem, in: B. Schölkopf, J. Platt, T. Hoffman (Eds.), Advances in Neural Information Processing Systems, Vol. 19, MIT Press, 2007.

[16] A. Smola, A. Gretton, L. Song, B. Schölkopf, A hilbert space embedding for distributions, in: International Conference on Algorithmic Learning Theory, Springer, 2007, pp. 13–31.

[17] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.

[18] M. Lamboni, Uncertainty quantification: a minimum variance unbiased (joint) estimator of the non-normalized Sobol' indices, Statistical Papers (2018) –doi: https://doi.org/10.1007/s00362-018-1010-4.

[19] M. Lamboni, Weak derivative-based expansion of functions: Anova and some inequalities, Mathematics and Computers in Simulation 194 (2022) 691–718.

[20] H. H. Kuo, Gaussian Measures in Banach Spaces. Lecture Notes in Mathematics 463, Springer, Berlin, 1975.

[21] J. O. Ramsay, B. W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer-Verlag, Berlin, 2002.

[22] M. Lamboni, Derivative-based integral equalities and inequality: A proxy-measure for sensitivity analysis, Mathematics and Computers in Simulation 179 (2021) 137 – 161.

[23] M. Lamboni, M. Sanaa, F. Tenenhaus-Aziza, Sensitivity analysis for critical control points determination and uncertainty analysis to link fso and process criteria: Application to listeria monocytogenes in soft cheese made from pasteurized milk, Risk Analysis 34 (4) (2014) 751–764.

[24] M. Lamboni, Efficient dependency models: simulating dependent random variables, Mathematics and Computers in Simulation , submitted on 03/01/2021 (2021).