USE OF HSIC-BASED SENSITIVITY ANALYSIS FOR FUNCTIONAL DATA

**Hilbert Schmidt Independence Criterion*

FROM RESEARCH TO INDUSTRY

A. Marrel[1,2] & R. El Amri[1,3]

[1]CEA DES/IRESNE/DER, [2]IMT Toulouse

[3]Now at IRT Saint Exupéry, Toulouse

10th SAMO Conference, Tallahassee, USA

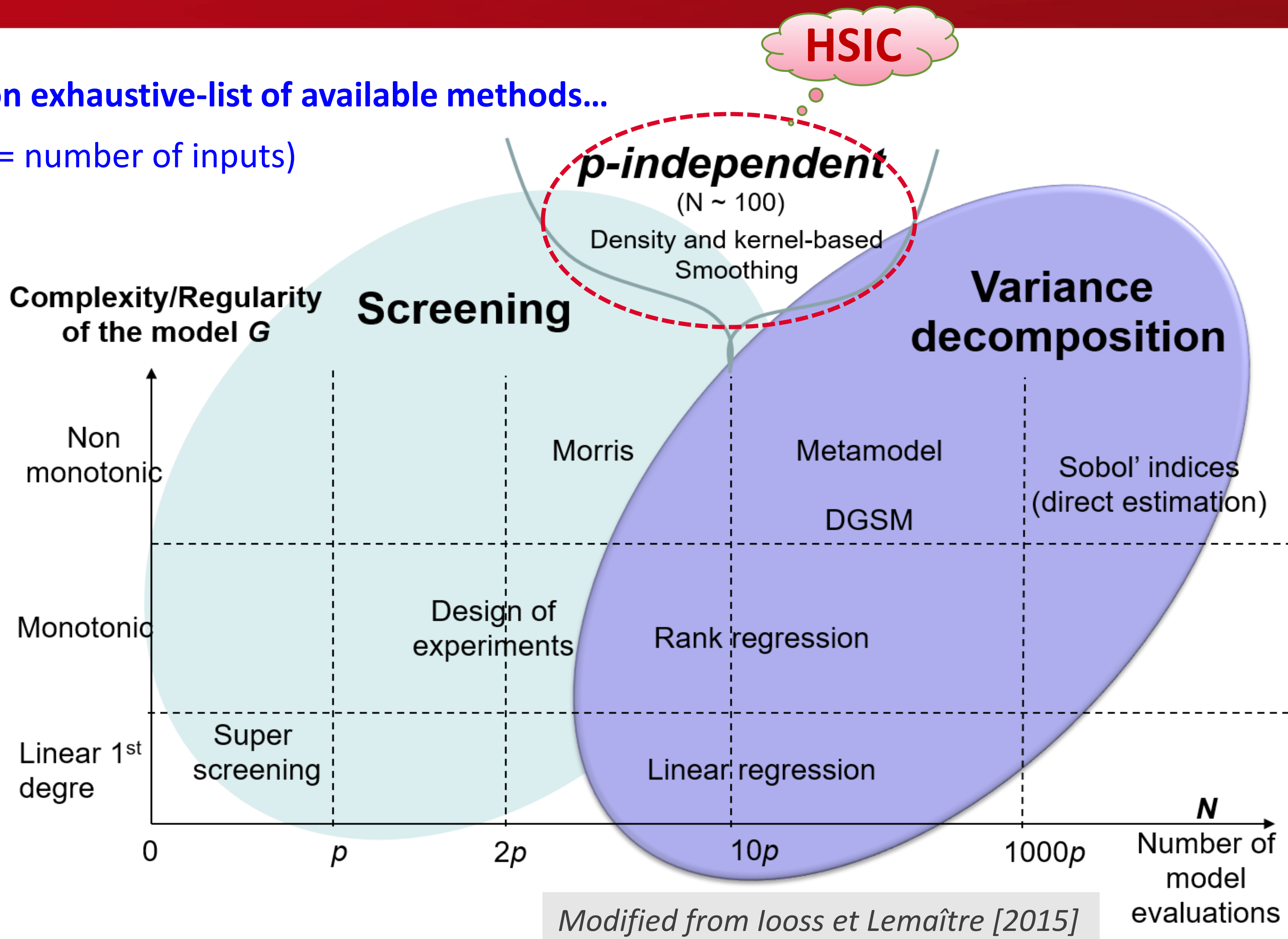
***IRESNE** / DER / SESI / LEMS*

Research Institute on Nuclear Systems for Low Carbone Energy Production

Global Sensitivity Analysis of numerical simulators

Non exhaustive-list of available methods…

(p = number of inputs)



Modified from Iooss et Lemaître [2015]

HSIC* Review

**Hilbert Schmidt Independence Criterion*

IRESNE | **DER** | **SESI** | **LEMS**

Research Institute on Nuclear Systems for Low Carbone Energy Production

HSIC review: a few notations

Black-box model

$$\boldsymbol{Y} = \mathcal{M}(X_1, \dots, X_d)$$

- $X_1, \dots, X_d$ are *d* independent inputs, evolving in domain $\mathcal{X}_1, \dots, \mathcal{X}_d$
- $\boldsymbol{Y}$ evolves in domain $\mathcal{Y}$
- $\boldsymbol{P}_X$ denotes the probability distribution of X
- $\boldsymbol{P}_{X,Y}$: the joint probability measure and $\boldsymbol{P}_Y \otimes \boldsymbol{P}_X$ the product of marginal distributions

Only a *n*-sample of simulations is available

$\mathcal{M}$ unknown, only Monte-Carlo sample $\left(\boldsymbol{X}^{(j)}, \boldsymbol{Y}^{(j)}\right)_{1 \le j \le n}$ where $Y^{(j)} = \mathcal{M}\left(X^{(j)}\right)$

HSIC review: GSA built upon RKHS framework

- **How to evaluate the sensitivity in a probabilistic way? ⇔ independence**

  → **By comparing $P_{X,Y}$ with $P_X \otimes P_Y$**

$$S_i = d\left(P_{X_i,Y}, P_{X_i} \otimes P_Y\right)$$

where *d* a dissimilarity measure between two probablity distributions

d can be based on **Maximum Mean Discrepancy**:

$$\mathrm{MMD}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \left[\mathbb{E}_{\mathbb{P}} f(Y) - \mathbb{E}_{\mathbb{Q}} f(Y)\right]$$

with $\mathcal{F}$ = **unit ball in a (characteristic) RKHS (Reproducing Kernel Hilbert Space)**

Sriperumbudur et al. [2008]

$$\Rightarrow S_i = MMD^2\left(P_{X,Y}, P_Y \otimes P_X\right) = HSIC(X, Y)$$

Hilbert-Schmidt Independence Criterion

HSIC review: definition from RKHS

▶ **MMD² applied between $P_{X,Y}$ and $P_Y \otimes P_X \Rightarrow HSIC(X,Y)_{\mathcal{F}_{X_i},\mathcal{F}_Y}$**

$\mathcal{F}_X$ and $\mathcal{F}_Y$ **RKHS** associated to *X* and *Y*, resp :

Kernel $k_X: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with feature space $\mathcal{F}_X$ and feature map φ_X

Kernel $k_Y: \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ with feature space $\mathcal{F}_Y$ and feature map φ_Y

kernel defines the inner product in the RKHS

$$\Rightarrow HSIC(X,Y)_{\mathcal{F}_X,\mathcal{F}_Y} = MMD^2_{\mathcal{F}_X,\mathcal{F}_Y}\left(P_{X,Y}, P_Y \otimes P_X\right) = \left\|C_{X,Y}\right\|^2_{HS}$$

With $C_{X,Y}$ **the covariance operator** between features maps:

$$C_{X,Y} = \mathbb{E}_{X,Y}[\varphi_X(X) \otimes \varphi_Y(Y)] - \mathbb{E}_X[\varphi_X(X)] \otimes \mathbb{E}_Y[\varphi_Y(Y)]$$

Gretton et al. [2005]

*HSIC **"summarizes" the cross-cov** between **feature maps***

*$\Rightarrow$ **Large panel of input-output dependency** can be captured.*

HSIC review: estimation

▶ **Kernel trick** in action (reproducing property):

$$\mathrm{HSIC}(X,Y) = \mathbb{E}[k_X(X,X')k_Y(Y,Y')] + \mathbb{E}[k_X(X,X')]\mathbb{E}\left[k_y(Y,Y')\right] - 2\mathbb{E}\left[\mathbb{E}[k_X(X,X')|X]\mathbb{E}[k_Y(Y,Y')|Y]\right]$$

where (X',Y') is an independent and identically distributed copy of $(X,Y) \sim P_{X,Y}$.

▶ **Characteristic kernels and RKHS** ⇨ Embedding of probability measures is **injective**

⇨ ***Equivalence to independence:*** $\boldsymbol{HSIC(X,Y) = 0 \Leftrightarrow X \perp Y}$ *Ex: Gaussian Kernel*

$$k_\lambda(x_i,x_i') = exp\left(-\frac{(x_i - x_i')^2}{2\lambda^2}\right)$$

▶ **Monte-Carlo estimation:** from a ***n*-sample** $\left(X_i^{(j)},Y^{(j)}\right)_{1\le j\le n}$

$$\widehat{\mathrm{HSIC}}(X,Y) = \frac{1}{n^2}Tr(K_X H L_Y H)$$

where $H = I_n - \frac{1}{n}$, $K_X = \left(k_X\left(X^{(j)},X^{(j')}\right)\right)_{1\le j,j'\le n}$ and $L_Y = \left(k\left(Y^{(j)},Y^{(j')}\right)\right)_{1\le j,j'\le n}$

Statistical properties of $\widehat{\mathrm{HSIC}}$: Asymptotically unbiased, variance of order $O(1/n)$, asymptotically convergence of $n\,\widehat{\mathrm{HSIC}}$ if $X \perp Y$ *Gretton et al. [2008]*

HSIC review: sentivity indices and independence test

▶ Normalization for sensitivity analysis:

Da Veiga [2015]

$$R^2_{HSIC} = \frac{HSIC(X,Y)}{\sqrt{HSIC(X,X)HSIC(Y,Y)}}$$

⇨ $R^2_{HSIC} \in [0,1]$ for easier interpretation

$$\text{Influence}(X_{[1]}) > \text{Influence}(X_{[2]}) > \cdots > \text{Influence}(X_{[d]})$$

Where order $[\cdot]$ is such that $\hat{R}^2_{H,X_{[1]}} > \hat{R}^2_{H,X_{[2]}} > \cdots > \hat{R}^2_{H,X_{[d]}}$

⇒ Use for ranking of inputs

▶ Independence tests: $HSIC(X,Y) = 0 \Leftrightarrow X \perp Y$ (*with <u>characteristic</u> kernels!*)

- Null hypothesis: $\mathcal{H}_0 : X \perp Y$ against $\mathcal{H}_1 : X \not\perp Y$
- Test statistics: $n\widehat{\text{HSIC}}(X,Y)$
- Decision rule: $\mathcal{H}_0$ rejected iff $n\widehat{\text{HSIC}}(X,Y) > q_{1-\alpha}$

  where $q_{1-\alpha}$ is the $(1-\alpha)$quantile of $n\widehat{\text{HSIC}}(X,Y)$ under $\mathcal{H}_0$

⇒ Use for screening of inputs

Gretton et al. [2008]
De Lozzo & Marrel [2016]

Extension to Functional data

IRESNE | DER | SESI | LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

HSIC for functional output

Output = Functional data

Output is a random function of time or space $\Rightarrow$ which kernel to be used?

- $\boldsymbol{Y}$ belongs to the functional space $\mathcal{H} = \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{Y})$ with

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

- Assume that $\boldsymbol{Y} \in \mathcal{H}$ has mean $\boldsymbol{\mu(t)}$ and continuous covariance function $\boldsymbol{C(t,s)}$.

Step 1: Functional dimension reduction

El Amri & Marrel [2021]

Functional Principal Component Analysis (FPCA) and truncation with the ***q* first terms**:

$$Y(t) - \mu(t) \approx \sum_{k=1}^{\boldsymbol{q}} \boldsymbol{U}_k \, \varphi_k(t)$$

where $\{\varphi_k\}_{k=1}^{\infty}$ is an orthonormal basis of eigenfunctions of the integral operator corresponding to $\boldsymbol{C}$, $\{\boldsymbol{U}_k\}_{k=1}^{\infty}$ denoting a set of uncorrelated random variables with zero mean and variance $\{\theta_k\}_{k=1}^{\infty}$ (eigenvalues of operator $\boldsymbol{C}$).

HSIC for functional output

▶ Step 1: Functional dimension reduction (in practice)

El Amri & Marrel [2021]

C estimated by

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

The eigenvalue problem solved by **replacing C by $\widehat{C}$** and by discretizing the trajectories at several discrete time points $t_1, \dots, t_{N_T}$.

- **Choice of *q*** (number of terms kept in FPCA approximation): minimal number such that the cumulated ratio of variance explained exceeds a given threshold
  - $\Rightarrow$ Control the percentage of output variance explained by the truncation

Alternative with **spline-based FPCA:** trajectories are expanded as linear combinations of spline basis functions, before applying PCA to the coefficients on the spline basis

HSIC for functional output

Step 2: Build FPCA-based kernel

Weigthed sum of 1-D kernels with the *q* first FPCA (random) coefficients $(\boldsymbol{U}_k)_{\boldsymbol{k=1,\ldots,q}}$

$$k_{\Sigma_w}\left(Y^{(j)}, Y^{(j')}\right) = \sum_{k=1}^{q} w_k k_U\left(U_k^{(j)}, U_k^{(j')}\right)$$

weigths $(w_k)_{\boldsymbol{k=1,\ldots,q}}$ control the relative contribution of the main *q* components

Practical choices to build $\boldsymbol{k_{\Sigma_w}}$

- **For weights $\boldsymbol{w_k}$**: percentage of variance explained by each component $\boldsymbol{U_k}$

$$w_k = \frac{\theta_k}{\sum_{k=1}^{q} \theta_k} \qquad (\ \theta_k \text{ eigenvalues from FPCA})$$

- **For 1-D kernels $\boldsymbol{k_U}$**: usual kernel for real variables (Gaussian e.g.)

$\Rightarrow$ Estimators $\widehat{HSIC}_{PCA}$ and $\hat{R}^2_{HSIC,PCA}$ from FPCA reduction and kernel $\boldsymbol{k_{\Sigma_w}}$

Other « aggregated » kernels proposed in *El Amri & Marrel [2021].*

Note that $\boldsymbol{k_{\Sigma_w}}$ is not characteristic due to FPCA truncation.

HSIC for functional output

▶ **Illustration on analytical example:** arctangent-based temporal function

$$Y_{atan}(t, X_1, X_2, X_3) = \text{atan}(X_1)\cos(t) + 50\,\text{atan}(X_2)\sin(t)$$

Where $t \in [0,2\pi]$ and $X_i \sim \mathcal{U}([-7;7])\,\forall i = 1, \dots, 3$

➢ « Theoretical analysis » from intensive Monte-Carlo estimation (*n = 100000)*

⇒ Two first components $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ explain **around 97% and 3% of variance**, resp.

⇒ **Strong influence** of $\boldsymbol{X}_2$ on $\boldsymbol{U}_1$, far the most influential input

⇒ Much smaller influence of $\boldsymbol{X}_1$, mainly conveyed by $\boldsymbol{U}_2$

⇒ $\boldsymbol{X}_3$ = dummy variable

	$\widehat{R}^2_{HSIC,X_1}$	$\widehat{R}^2_{HSIC,X_2}$
U_1 (96.96%)	5.69×10^{-3}	0.904
U_2 (3.03%)	0.903	6.32×10^{-3}

Individual $\widehat{R}^2_{HSIC}$ on each FPCA component

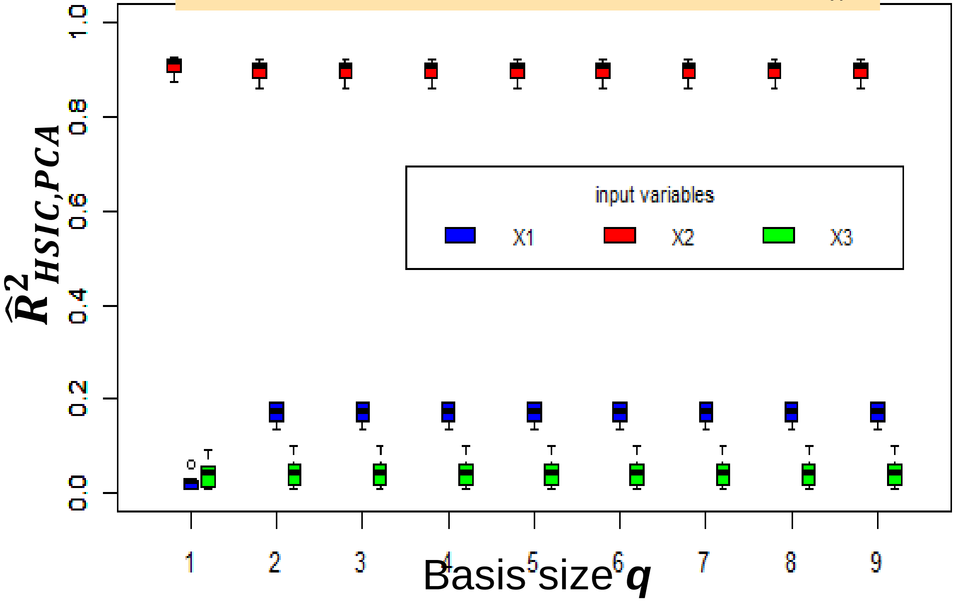
HSIC for functional output

▶ **Illustration on analytical example:** arctangent-based temporal function Y_{atan}

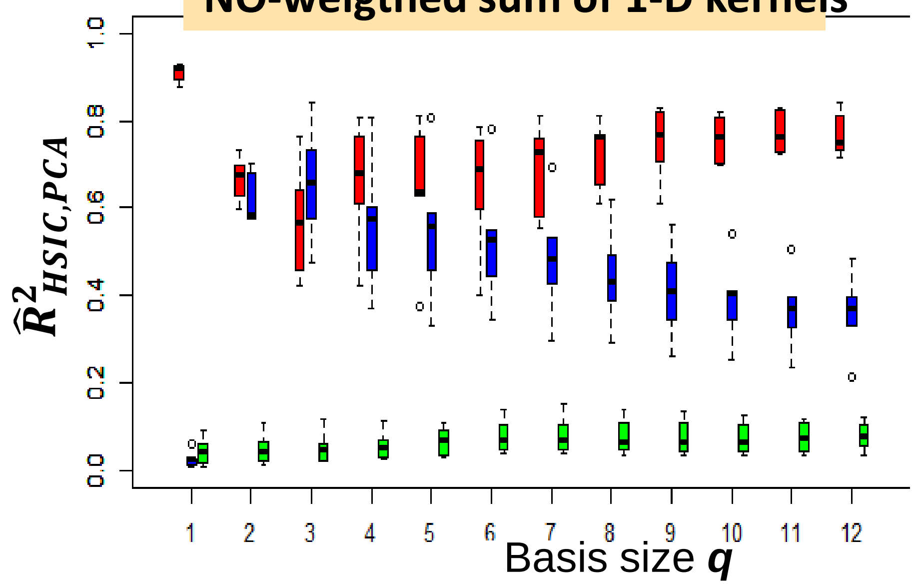
Computation of $\hat{R}^2_{HSIC,PCA}$ from *n=100* Monte-Carlo sample

50 sample replicates and trajectories of Y_{atan} *discretized on* $N_T = 100$ *equally-spaced points*

Weigthed sum of 1-D kernels k_{Σ_w}



NO-weigthed sum of 1-D kernels



HSIC for functional output

Illustration on compartmental epidemiologic model on COVID-19 :

Modified SIR model (Susceptible – Infected – Recovered)

Magal & Webb [2020]

- **2 temporal outputs** of interest

$I(t)$: number of asymptomatic infectious individuals at time t
$R(t)$: number of reported symptomatic infectious individuals at time t

- Model defined by **system of o.d.e**

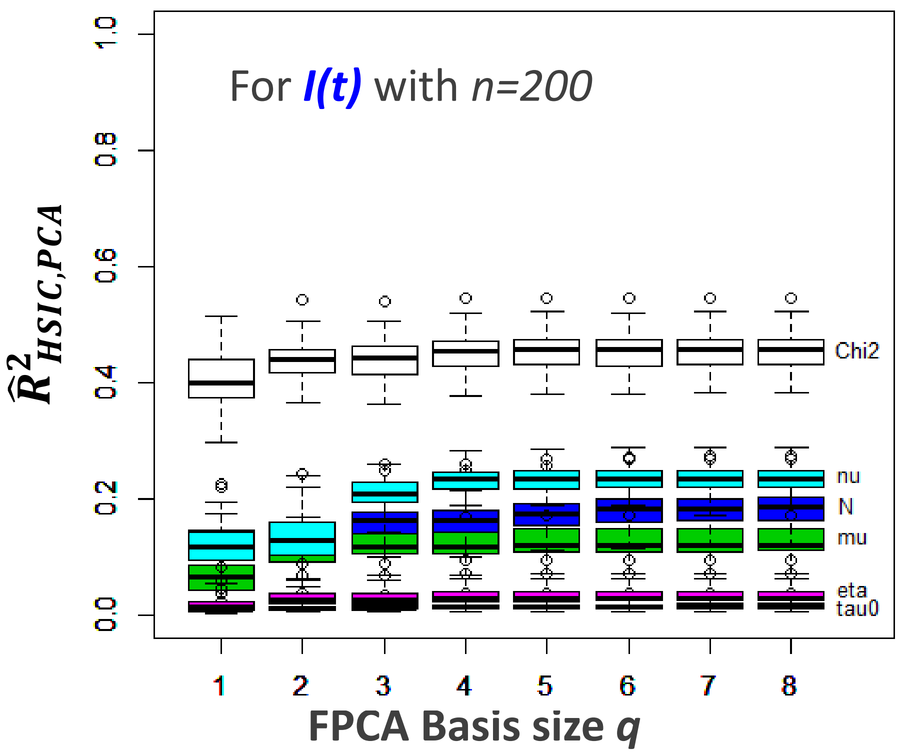
$$\begin{cases} S'(t) = -\tau(t)S(t)\big(I(t)+U(t)\big) \\ I'(t) = \tau(t)S(t)\big(I(t)+U(t)\big) - \nu I(t) \\ R'(t) = f\nu I(t) - \eta R \\ U'(t) = (1-f)\nu I(t) - \eta U. \end{cases}$$

- Depending on **6 uncertain scalar inputs**, assumed independent and uniform
  - Initial conditions, involved in $I(t_0)$
  - Transmission rate before the epidemic outbreak
  - Rate of the exponential decay of transmission as soon as social distancing and lockdown start
  - Date of effect of government measures
  - Average periods during which $I(t)$ are infectious
  - Average periods during which $R(t)$ are infectious

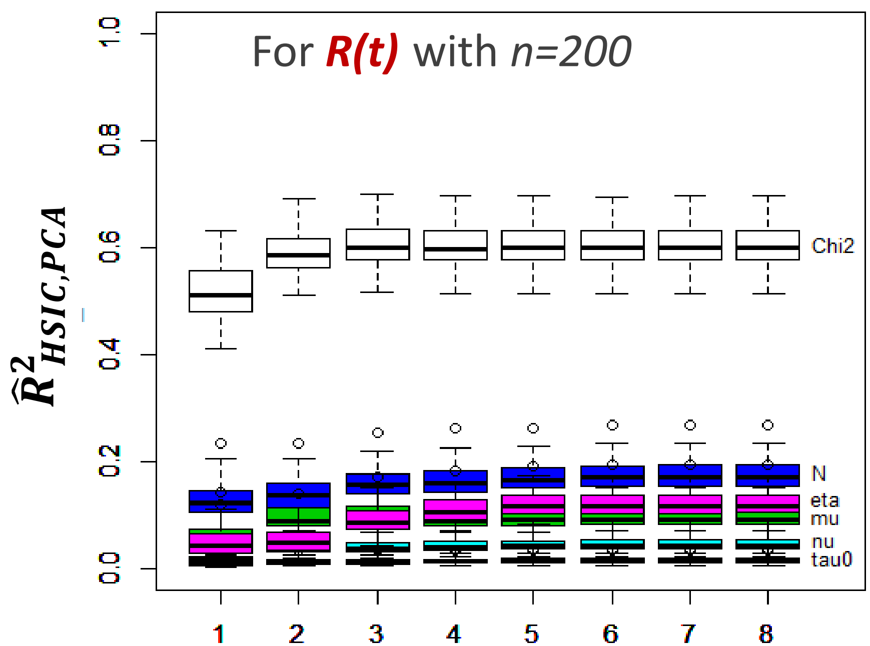
HSIC for functional output

▶ Illustration on compartmental epidemiologic model on COVID-19 :

I(t) and *R(t)*: number of asymptomatic and reported symptomatic infectious individuals at time *t*



- Initial conditions, involved in $I(t_0)$
- Transmission rate before the epidemic outbreak
- Rate of the exponential decay of transmission as soon as social distancing and lockdown start
- Date of effect of government measures
- Average periods during which *I(t)* are infectious
- Average periods during which *R(t)* are infectious



Conclusion and prospects

IRESNE | DER | SESI | LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

Conclusions

HSIC as indices of Sensitivity Analysis

- Focus the SA analysis on the difference between $P_{X,Y}$ with $P_X \otimes P_Y$
- Power of RKHS → HSIC can capture a large spectrum of relationships and characterize independence

  → Efficient for **ranking** and **screening** of uncertain inputs, even from **small samples**

Extension to functional output

- *Aggregated* kernels based on **FPCA decomposition + weighted sum** of usual 1-D kernels
- Consistent results on analytical example and real applications
- Can be also applied to **multivariate outputs**
- Statistical test of nullity of the indices can be built (not equivalent to independence)

Prospects

- Extension with other metrics for functional data like ***Global Alignment Kernel***
- Decomposition into main effects & interactions must be investigated

  ⇒ Assess the use of **HSIC with *ANOVA-like kernels***

Keynote 4 - Sébastien DA VEIGA

References

- Da Veiga, Global sensitivity analysis with dependence measures, *Journal of Statistical Computation and Simulation*, 85:1283-1305, 2015.
- Da Veiga., Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. arXiv preprint arXiv:2101.05487. 2021.
- De Lozzo and Marrel, New improvements in the use of dependence measures for sensitivity analysis and screening, *Journal of Statistical Computation and Simulation*, 86:3038-3058, 2016
- El Amri and Marrel, More powerful HSIC-based independence tests, extension to space-filling designs and functional data, https://hal-cea.archives-ouvertes.fr/cea-03406956, 2021b.
- Gretton, Fukumizu, Teo, Le Song, Schölkopf and Smola. A kernel statistical test of independence. In Advances in neural information processing systems, 585-592, 2008.
- Gretton, Bousquet, Smola and Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. *In: Proceedings Algorithmic Learning Theory*, 2005.
- Iooss B. and Lemaître P. A review on global sensitivity analysis methods. In Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer, 2015.
- Müller, A. Integral probability metrics and their generating classes of functions, Advances in Applied Probability 29(2), 429–443, 1997.
- Sriperumbudur B.K., Gretton A., and Fukumizu K. Hilbert space embeddings and metrics on probability measures. Journal of Machine Learning Research, (11):1517-1561, 2010.
- Magal P. and Webb G. 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.