# Surrogate modeling of stochastic simulators using Karhunen-Loève expansions

S. Azzi
*Télécom ParisTech, LTCI, Université Paris-Saclay*

**Supervisor(s):** J. Wiart (Télécom ParisTech), B. Sudret (ETH Zürich)

**Ph.D. expected duration:** Oct. 2016 - Oct. 2019

**Adress**: Télécom ParisTech, 46 Rue Barrault 75013 Paris

**Email**: soumaya.azzi@telecom-paristech.fr

**Abstract:** In engineering problems, simulators commonly contain sources of uncertainty due to measurements for example. They are called stochastic simulators because they yield a probability density function (PDF) with respect to every input. Even though through numerical computations, stochastic simulators can be investigated, they remain computationally expensive. Metamodels are mathematical functions that mimic the behavior of simulators and are used to overcome the pricey calls to the simulators. The abstract introduces a metamodeling approach for stochastic simulators based on Karhunen-Loève (KL) expansion [4].

Let $H(x, \omega) \in \mathbb{R}$ be a stochastic process on $D \times \Omega$, where $x \in D \subset \mathbb{R}^n$ and $\omega$ in the sample space $\Omega$. The stochastic simulator is modeled as a stochastic simulator, and its surrogate is a stochastic process as well, noted $\hat{H}(x, \omega)$. Let the stochastic process $H(x, \omega)$ be a zero mean second order process. Its covariance operator is denoted $C(x, y)$ and let $\lambda_i$ and $\phi_i$ be respectively its eigenvalues and eigenvectors. Then the KL expansion [4] reads as follow :

$$H(x, \omega) = \lim_{p \to \inf} \sum_{i=1}^{p} \sqrt{\lambda_i} \xi_i(\omega) \phi_i(x) \tag{1}$$

In practice, several calls to the stochastic simulator are made, let $M$ be the size of the design of experiment set (DoE) and $M$ the number of realization on each point from the DoE. The simulated process is then a matrix with $M$ rows and $N$ columns, each row represents the $N$ realizations made over a point from the DoE. Each column represents a trajectory, meaning that for all $x \in DoE$, simulations were carried with a same seed. Based on the data from the simulation, the empirical covariance matrix is evaluated, $\phi_i$ and $\lambda_i$ are calculated. The aim is to predict the PDF of a new point $x^* \in D$. Based on Eq.1, the predicted response reads as $\hat{H}(x^*, \omega) = \sum_{i=1}^{M} \sqrt{\lambda_i} \hat{\xi}_i(\omega) \hat{\phi}_i(x)$.

- $\phi_i(x^*)$ is unknown, the eigenvectors are only computed for the DoE set. To overcome this limitation, a $\phi_i$ originally known only over the $M$ point of the DoE can be interpolated to predict $\phi_i(x^*)$.

- Alternatively, a metamodel of the covariance is build using polynomial chaos expansions [1], the eigendecomposition is then performed to get $\hat{\phi}_i$. Both ways are used and compared.

- Concerning the random variables, they are the projection of $H$ onto the base of the eigenvectors $\hat{\phi}_i,\ i \in \{1 \dots M\}$ [4]

$$\hat{\xi}_i(\omega_k) = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^{M} H(x^{(j)}, \omega_k) \hat{\phi}_i(x^{(j)}) \tag{2}$$

This approach was applied to a radio frequency exposure (RF) computational simulator, the simulator evaluates the exposure to RF waves of a population living in cities [3]. The inputs of this simulator are the parameters of the city, and the output is the RF exposure of the population. The simulator is stochastic, given a set of city parameters, numerous cities can be generated [2] (Figure 1) hence numerous exposure rates can be evaluated (for the same city parameters).

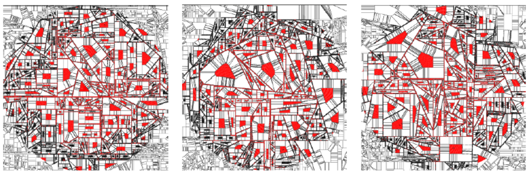

Figure 1: Three realizations of virtual cities showing the same parameters (street width, anisotropy, building height and length).

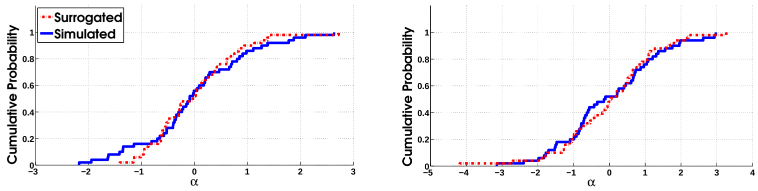


Figure 2: Surrogated and simulated CDF's plotted when the eigenvectors are linearly interpolated (right) and when the covariance is interpolated using PCE (left).

The stochastic city simulator was simulated on 30 points, 50 times. A cross validation was applied with 10% of the data left to the test (Figure 2). The accuracy of the model is evaluated by comparing the 'real' PDF and the surrogated one using different metrics : histogram intersection, Hellinger distance and Jensen Shannon divergence.

For the sensitivity analysis (SA) and based on the simulations, the entropy on each point of the DoE, noted $\mathcal{E}(x)$ is evaluated, a metamodel $\hat{\mathcal{E}}(x)$, $x \in D$ is build. The SA approaches can be evaluated on $\hat{\mathcal{E}}$.

## References

[1] G Blatman and B Sudret. Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys*, 230:2345–2367, 2011.

[2] T. Courtat, L. Decreusefond, and P. Martins. Stochastic simulation of urban environments. application to path-loss in wireless systems. *arXiv:1604.00688,*, 2016. [Online].

[3] Y. Huang and J. Wiart. Simplified assessment method for population RF exposure induced by a 4G network. *IEEE Journal of Electromagnetics, RF, and Microwaves in Medicine and Biology*, 1:34–40, 2017.

[4] Roger G. Ghanem Pol Spanos. *Stochastic Finite Elements: A Spectral Approach.*

**Short biography** – Azzi Soumaya is a Ph.D. student within Chaire C2M, LTCI, Télécom ParisTech. Her research interests include stochastic computation, surrogate modeling and uncertainty quantification. She received the M.Sc. degree in applied mathematics from Blaise Pascal University, Clermont F$^{rd}$, France.