Emulating stochastic simulators

MascotNum Annual Conference
March 21-23 2018
Nantes, France

Auteurs

Soumaya AZZI [1]

Yuanyuan HUANG [1]

Bruno SUDRET [2]

Joe WIART [1]

Affiliations

1 Télécom ParisTech , LTCI, Chair C2M
*2 ETH Z*ürich, Chair of Risk, Safety & Uncertainty Quantification

Summary

Stochastic city models were applied to assess the global Electromagnetic Field (EMF) exposure of a population. These stochastic models were constructed in terms of the main structures of a city. With these models, the same input leads to different outputs. Because of the computational cost, surrogate model of such random function is needed.
Today, large efforts have been dedicated to surrogate deterministic simulators, it is therefore in our interest to develop methods dedicated to the stochastic ones. This paper proposes an innovative approach based on Karhunen–Loève decomposition to built and evaluate a surrogate model for stochastic functions.

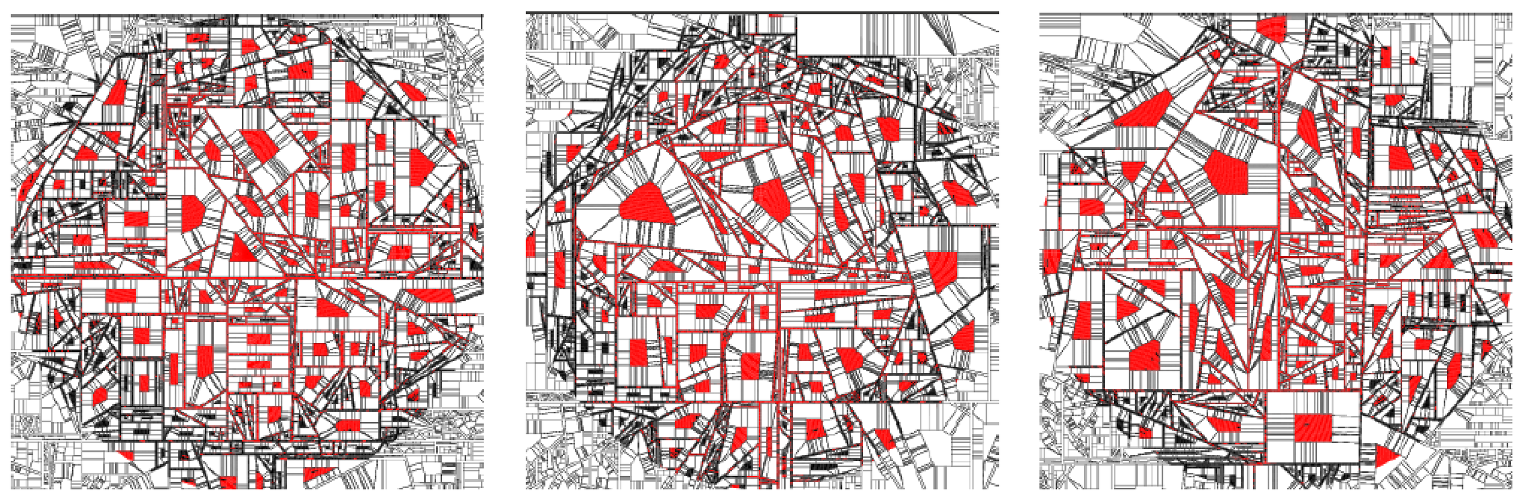


Figure 3: Three realizations of virtual cities showing the same macroscopic parameters (street width, anisotropy, building length and height).

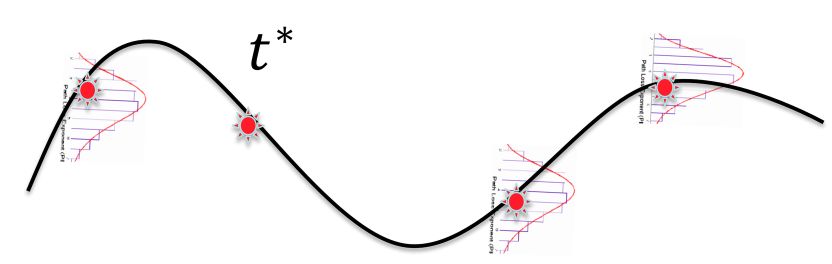
Output variable of interest: EMF exposure α

The model will be treated as a stochastic process $H(t,\omega)$ where t denotes the deterministic inputs (street width, anisotropy, building length and height), and ω all the randomness causing the stochastic behavior

Sampling: we run the process $H(t,\omega)$ for a DoE of N points t, M time. It will take $\boldsymbol{M.N.1,5}$ ***hours***

Aim: for a new point t^* what is a predictable distribution of $H(t^*,\omega)$

Random vector of **input** parameters ***T*** → Stochastic city generator → Random model response Y = M (***T***)



Tools

Karhunen–Loève theorem

Let $H: D\times\Omega \to \mathbb{R}$ be a zero mean second order stochastic process. Its covariance function is consntinuous in the mean square sense and denoted as $C(s,t)$. Assume that $\phi_i$and λ_i satisfies the following equation $\int C(s,t)\phi_i(t)\,dt = \lambda_i\phi_i(s)$

Furthermore choose $\xi_i = \frac{1}{\sqrt{\lambda_i}}\int H(t,\omega)\phi_i(t)dt$

Then $H(t,\omega) = \lim_{p\to\infty}\sum_{i=1}^{p}\sqrt{\lambda_i}\xi_i(\omega)\phi_i(t)$

Surrogating the covariance:

$$C(s,t) \approx \sum_{j=0}^{K-1} \alpha_j\psi_j(s,t)$$

The covariance is onlu known on the N DoE points, surrogating the covariance is a tool to get the missing values. To surrogate $C(s,t)$ we use **Splines** or **Polynomial Chaos (PC) decomposition** (approximates the unknown random response of a model in a suitable finite-dimensional basis of orthogonal polynomials ($\psi_j(\boldsymbol{X})$) $_{0\le j\le P-1}$)

Method

Data matrix $H(t_i,\omega)$ → Covariance $C(t_i,t_j)$ → Cov Surrogate $\hat{C}(s,t)$ → Eigen decomposition

↓

Random variables ξ_i
If the process is ~Gaussian then ξ_i are normal i.i.d
Otherwise we compute ξ_i using the projecton formula
$\xi_i(\omega_k) = \frac{1}{\sqrt{\lambda_i}}\sum_{j=1}^{N} H(t_j,\omega_k)\phi_i(t_j)$. ω_k are 'freezed' in that case (we know that $\xi_i(\omega_k)$ and $H(t_j,\omega_k)$ are from the same realization)

↓

Using Karhunen-Loève theorem, comput
$H(t^*,\omega)=\sum_{i=1}^{p}\sqrt{\lambda_i}\xi_i(\omega)\phi_i(t^*)$

↓

Evaluate the error

Toy example

$$H(t,\omega) = \cos(2\pi\omega t)\,, \omega\sim U(0,0.2), t\in[0,1]$$

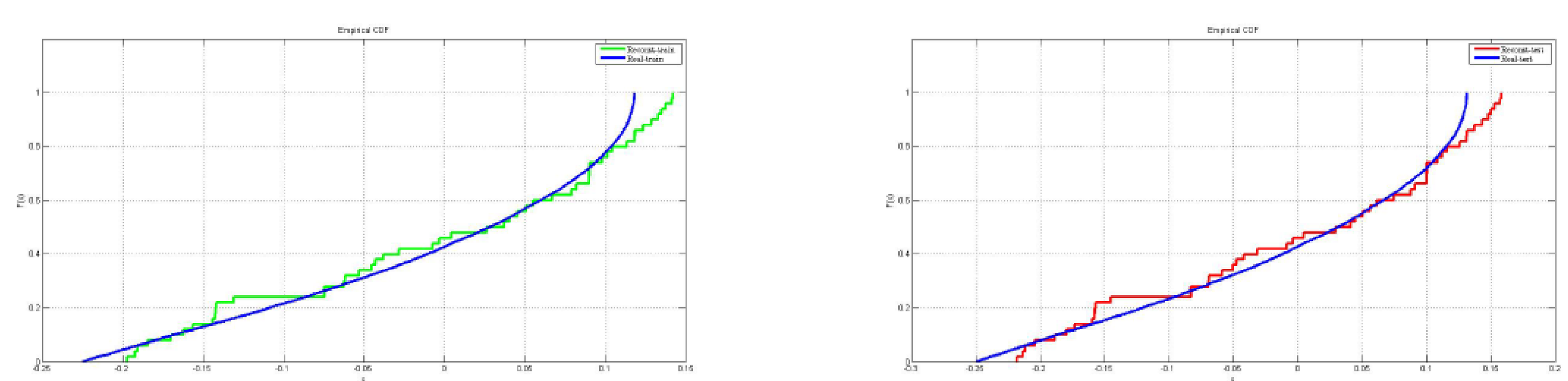
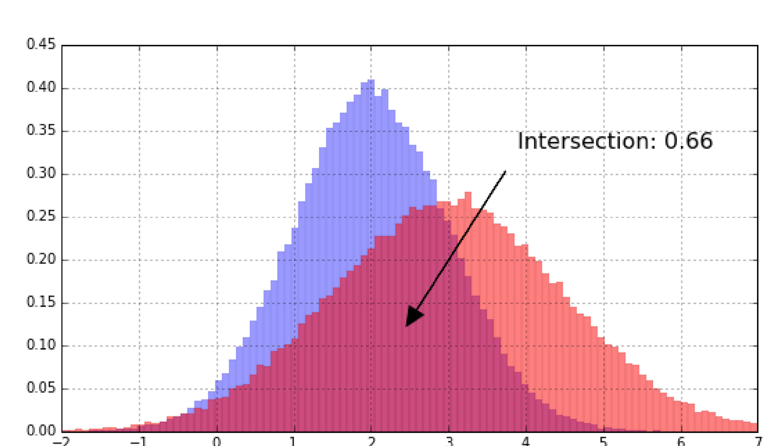


Figure 2: left: reconstructed cdf of $\hat{H}_2(x_i,\omega)$ (green) vs true cdf of $H_2(x_i,\omega)$ (blue), x_i within the training set. Right: predicted cdf of $\hat{H}_2(x^*,\omega)$ (red) vs true cdf of $H_2(x^*,\omega)$ (blue).

Error

How to compare p and q? p/$F_1$and q /F_2 are pdf/cdf of real vs test output

Histogram intersection



Kolmogorov Smirov statistical distance
$KS_{n,m} = sup_x|F_{1,n}(x)\text{-}\ F_{2,m}(x)|$

Hellinger distance
$H(p,q) = \frac{1}{\sqrt{2}}||\sqrt{p}\ \text{-}\ \sqrt{q}||_2$

Jensen-Shanon divergence $JSD(p,q) = \frac{D_{KL}(p||r)+D_{KL}(p||r)}{2}$ where $r = \frac{p+q}{2}$ and

$$D_{KL}(p||q) = -\sum_i p(i)\log(\frac{q(i)}{p(i)})$$

Discussion and perspective

-Nature of ξ_i: do all of them belong to a widder family of distributions, in which case a parametric approach would be conceiveable
-Using the covariace surrogate we can increase the number of t, not the case for the realizations of each ξi
-Minimize the product NM: with the same computational cost, a parsimonious DoE is needed (example: what is the best sampling $(20t, 50\omega)$ or $(50t, 20\omega)$)

Contact : Soumaya AZZI
C2M Télécom ParisTech
Email: soumaya.azzi@telecom-paristech.com