Surrogating stochastic simulators using Karhunen-Loève expansion, sparse PCE and advanced statistical modelling

Presentation · June 2021

DOI: 10.3929/ethz-b-000493099

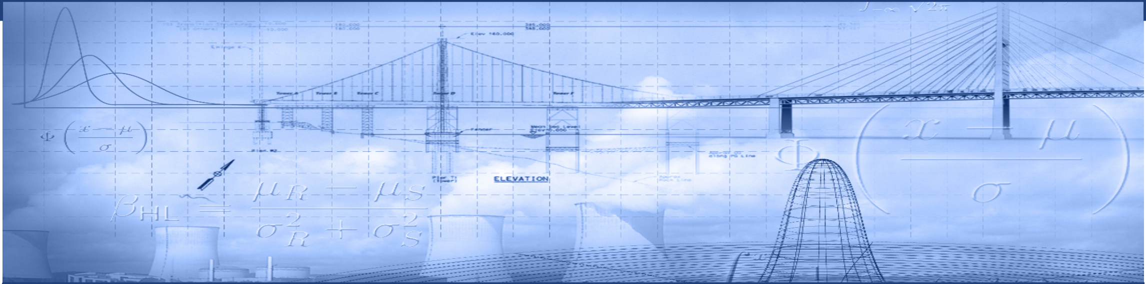
3 authors:

Nora Lüthen
ETH Zurich

Stefano Marelli
ETH Zurich

Bruno Sudret
ETH Zurich

ETH zürich



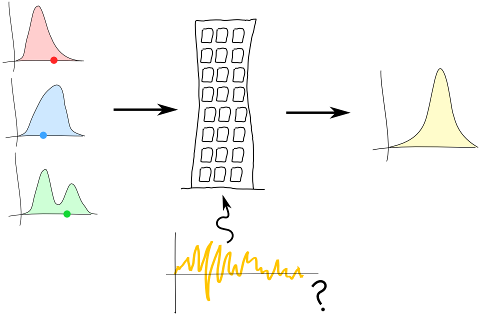
Surrogating stochastic simulators using Karhunen-Loève expansion, sparse PCE and advanced statistical modelling

UNCECOMP 2021

N. Lüthen, S. Marelli, B. Sudret

June 29, 2021

Stochastic simulators

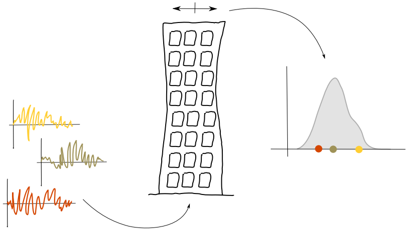


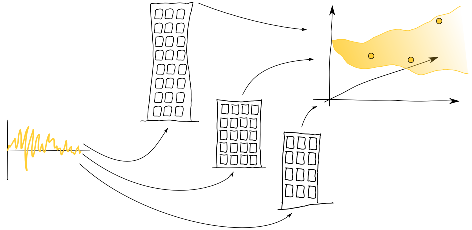
Stochastic simulator

What if it is not feasible to model all of the uncertainty?

- uncontrollable environmental variables (e.g., wind field, earthquake)
- intrinsic randomness (e.g., SIR epidemics model)

Stochastic simulators





Notation: Real-valued stochastic simulator $\mathcal{M} : \mathcal{D} \times \Omega \to \mathbb{R}$

- Input random vector $\boldsymbol{X}$ with values in $\mathcal{D} \subset \mathbb{R}^d$, finite variance, distribution $f_{\boldsymbol{X}}$
- Abstract random event $\omega \in \Omega$ denoting the stochasticity
- $Y_{\boldsymbol{x}} = \mathcal{M}(\boldsymbol{x}, \cdot)$ is a random variable
- $\mathcal{M}(\cdot, \omega) : \mathcal{D} \to \mathbb{R}$ is a function (trajectory)

$\to \mathcal{M}$ can be seen as a random field $\{Y_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{D}}$!

Karhunen-Loève expansion

Kosambi (1943); Karhunen (1947); Loève (1948); Ghanem & Spanos (1991)

Denote by $\mu(\boldsymbol{x}) = \mathbb{E}\left[\mathcal{M}(\boldsymbol{x}, \cdot)\right]$ the mean function of the random field, and by $C(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\left[\mathcal{M}(\boldsymbol{x}, \cdot), \mathcal{M}(\boldsymbol{x}', \cdot)\right]$ its covariance function.

The random field $\mathcal{M}$ can be represented as follows:[1]

Karhunen-Loève expansion (KLE)

$$\mathcal{M}(\boldsymbol{x}, \omega) = \approx \mu(\boldsymbol{x}) + \sum_{k=1}^{\infty\, M} \sqrt{\lambda_k}\, \xi_k(\omega)\, \phi_k(\boldsymbol{x})$$

with

1. an orthonormal basis $\{\phi_k\}_{k=1,2,\dots}$ of $\mathcal{L}^2(\mathcal{D})$
2. a decreasing sequence of real numbers $\lambda_1 \geq \lambda_2 \geq \dots \to 0$
   (λ_k, ϕ_k) are solutions to the integral eigenvalue problem $\int_{\mathcal{D}} C(\boldsymbol{x}, \boldsymbol{x}')\phi_k(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' = \lambda_k \phi_k(\boldsymbol{x})$
3. a countable family of zero mean, unit variance, uncorrelated random variables $\{\xi_k\}_{k=1,2,\dots}$
   ξ_k is the result of the projection of $\mathcal{M}$ onto the basis: $\xi_k(\omega) = \frac{1}{\sqrt{\lambda_k}} \int_{\mathcal{D}} \mathcal{M}(\boldsymbol{x}, \omega)\phi_k(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$

[1] Assumptions: $\mathcal{D}$ closed and bounded, C continuous, $\mathcal{M}(\boldsymbol{x}, \cdot)$ has finite variance $\forall \boldsymbol{x}$.

KLE – Challenges

1. KLE relies on the covariance *function* – what if we only have discrete data?

   Data: discrete evaluations of the stochastic simulator on R trajectories:

$$\mathcal{T}_r = \left\{ \mathcal{M}(\boldsymbol{x}^{(r,i)}, \omega^{(r)}) : i = 1, \ldots, N_r \right\}, \quad r = 1, \ldots, R$$

   where for every r, $\{\boldsymbol{x}^{(r,i)}\}$ is an i.i.d. sample from the input distribution $f_{\boldsymbol{X}}$

2. Once we have computed the expansion – how to deal with the random variables ξ_k of the KLE?

Approaches using KLE for discrete data ...

	Challenge 1: Discrete data	Challenge 2: Inference
Azzi et al. (2019) I	• First KLE on the discrete data • Then cont. approximation of the eigenvectors	• Use given realizations
Azzi et al. (2019) II	• First cont. approximation of the discrete covariance matrix • Then KLE	• Use given realizations
Poirion & Zentner (2014)	• Interpolate discrete trajectories using piecewise linear functions • Then KLE	• Assume independence • Kernel density estimation for marginals
Our approach	• Approximate the discrete trajectories using sparse PCE – orthonormal wrt input distribution $f_{\boldsymbol{X}}$ • Then apply KLE – not in $[0,1]^d$ but in $L^2_{f_{\boldsymbol{X}}}(\mathcal{D})$ (extended KLE) $\rightarrow$ equivalent to PCA on the PCE coefficients!	• Statistical modelling of the full multivariate distribution of KLE-RV using parametric inference of marginals and copulas

Toy example: stochastic differential equation

Navarro Jimenez et al. (2017), Zhu & Sudret (2020)

Stochastic differential equation

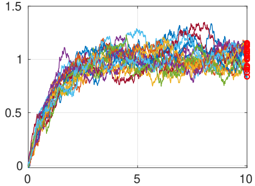
$$dU_t = (X_1 - U_t)\mathrm{d}t + (\nu U_t + 1)X_2 \mathrm{d}W_t$$

with initial condition $U_0 = 0$ (a.s.) and $\nu = 0.2$.

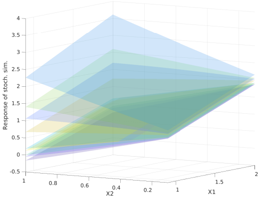
- $X_1 \sim \mathcal{U}([0.9, 2])$ and $X_2 \sim \mathcal{U}([0.1, 1])$ are the input parameters
- W_t is a standard Wiener process (source of stochasticity)
- QoI: $Y_{\boldsymbol{x}} = U_{t=10}(X_1 = x_1, X_2 = x_2)$

Trajectories: solve the SDE for several input parameters using the same process W_t

For the surrogate, we choose $N = 10$ and $p = 2$.



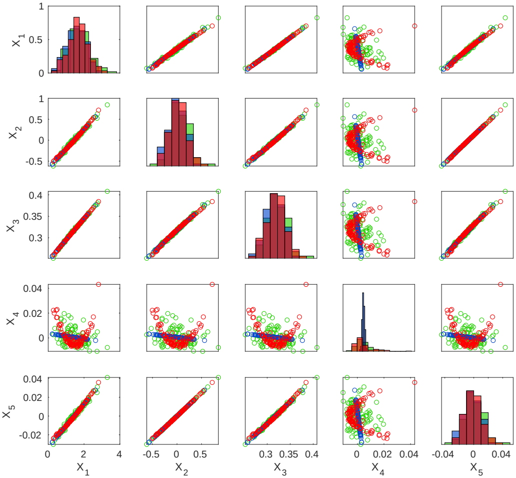
Time evolution of SDE (fixed input)



Several trajectories

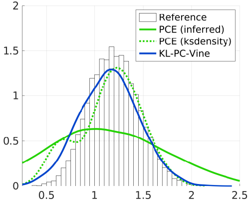
Toy example: stochastic differential equation

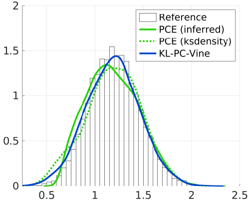
1. Sample the stochastic simulator: discrete trajectories
2. Approximate with sparse PCE (= polynomial regression using functions that are orthonormal wrt the distribution of the samples)
   → PCE coefficient sample
3. Possibility 1: directly model the distribution of PCE coefficients
4. Possibility 2: apply KLE to the PCE trajectories (= PCA on the PCE coefficients) and infer the distribution of random KL coefficients
5. *Possibility 3+: identify (nonlinear) functional dependence ...*



Stochastic differential equation: Results

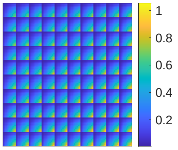
Marginal approximation at $\boldsymbol{x}_1 = (1.2, 0.3)$ for $R \in \{20, 100\}$



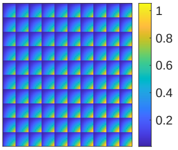


Covariance approximation for $R = 100$

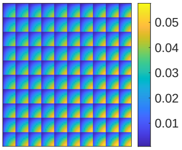
Empirical covariance



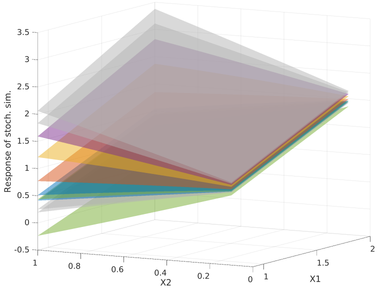
KLE-based covariance



Difference (EmpCov - KLECov)



Generating new samples by sampling from the inferred stochastic model



Conclusion

- We are developing a stochastic emulator for modelling and resampling a stochastic simulator (viewed as random field)
- Combination of Karhunen-Loève expansion, sparse PCE, and statistical inference using copulas

Outlook

- Develop ways to deal with the functional dependence between coefficients
- Improve parametric inference: find suitable marginal distributions and copulas (dependent yet uncorrelated)
- Understand theoretical implications of PCE approximation (orthogonal series expansion)
- Apply the method to real-world problems, e.g.
  - Wind turbine simulation
  - Impact of earthquake on family of buildings

Thank you for your attention!

Chair of Risk, Safety & Uncertainty Quantification

www.rsuq.ethz.ch

nluethen@ethz.ch

The Uncertainty Quantification Software

www.uqlab.com

The Uncertainty Quantification Forum

www.uqworld.com