Surrogating stochastic simulators using sparse polynomial chaos expansion and extended Karhunen-Loève decomposition

N. Lüthen
ETH Zürich, Switzerland

Supervisor(s): Prof. Bruno Sudret (ETH Zürich)

Ph.D. expected duration: Apr. 2018 - Mar. 2022

Adress: Stefano-Franscini-Platz 5, 8093 Zürich

Email: luethen@ibk.baug.ethz.ch

Abstract:

Surrogate modeling for computational models is a standard tool in uncertainty quantification: instead of the complex engineering model, an approximation is constructed that reproduces the input-output relationship and can be evaluated at a much smaller cost. Much of the research effort has concentrated on models for which the input-output relationship is deterministic and the uncertainty in the output is induced only by the input uncertainty.

A different class of computational models is the class of *stochastic simulators*, which are non-deterministic models: evaluating the model several times for the same input parameters will return a different output each time, so that for a fixed input vector, the model output is a random variable. The stochastic behavior is due to unknown or uncontrollable latent variables inside the model. Some stochastic simulators, such as agent-based disease propagation models (SIR model), feature inherent randomness. The individuals or agents have a certain probability of meeting each other, of getting infected, and of recovering. For the same initial conditions, repeated runs of the model have different realizations of agent meetings, infections, and recoveries, which results in a variety of epidemic outcomes.

Another example for stochastic simulators are wind turbine simulations (TurbSim-OpenFAST), which compute the structural response of a turbine to incoming wind. The time- and space-dependent wind function is high-dimensional, and it can be challenging to specify its exact distribution. Furthermore, just as for real-life experiments with wind turbines, the investigators are typically not interested in the turbine's structural response to a specific realization of the wind field; rather, they are interested in its response for a class of wind fields characterized by certain easily measurable values. The input to a stochastic wind turbine model is therefore a vector of values characterizing the wind field, such as mean and standard deviation of the wind speed, inflow angle, etc. For a given vector of characteristic values, a corresponding time series, which will be different in every run, is generated inside the simulator.

Often, the quantity of interest is the probability density function (PDF) of the model output for a given input vector. One line of research aims at constructing a model for the output PDF over the input domain, using a number of independent evaluations of the stochastic simulator across the input domain [MIDVR12, ZS19b, ZS19a]. These approaches do not require repeated evaluations of the stochastic simulator at the same point (replications). In this contribution, we take a different view: We interpret the stochastic simulator as a stochastic process $\{H(\mathbf{x}, \omega) : \mathbf{x} \in \mathcal{D}\}$, indexed over the input domain $\mathcal{D}$. The random vector $\mathbf{X} \in \mathcal{D}$ denotes the (explicitly modeled) probabilistic input to the simulator, while ω stands for the latent variables responsible for the stochastic behavior of the simulator. We assume that the trajectories $\{H(\mathbf{x}, \hat{\omega}) : \mathbf{x} \in \mathcal{D}\}$ are continuous for any given $\hat{\omega}$. Our goal is to infer the distribution of the underlying stochastic process from data, which is given in the form of discrete trajectories, i.e., evaluations

$$\{\mathcal{H}(\mathbf{x}^{(j,i)}, \omega^{(j)}) : j = 1, \dots, M, i = 1, \dots, N_j\}$$

at a set of experimental design points $\{\mathbf{x}^{(i,j)}\}$ for a number of fixed latent variables $\omega^{(j)}$. The experimental design may be different for each $\omega^{(j)}$. Note that conceptually, the latent variables are uncontrollable or unknown; however in the case of computational models, it is possible to control them through fixing the seed of the random number generator.

The simulation of non-Gaussian, non-stationary second-order stochastic processes from data is a topic that has received considerable attention, e.g., [PHQ05, PZ14, AHSW19]. In most of these simulation approaches, a central ingredient is Karhunen-Loève expansion (KLE). KLE is a spectral decomposition technique that uses the eigendecomposition of the covariance operator of the stochastic process to separate the process into spatial and random contributions. If the given discrete trajectories are interpolated by certain basis functions, the KLE integral eigenvalue problem can be computed analytically [PZ14]. We propose a similar approach: Instead of interpolation, we approximate the discrete trajectories by sparse regression-based polynomial chaos expansions (PCE). Using *extended KLE* on the space $L^2_{f_{\mathbf{X}}}$ [IDM06], where $f_{\mathbf{X}}$ is the PDF of the input random vector, the integral eigenvalue problem can be solved analytically in terms of the PCE coefficients. PCE acts as a dimension reduction technique and is efficient if the model can be represented sparsely in the PCE basis, which is often the case for engineering models (sparsity-of-effects heuristic). The joint distribution of the random KLE coefficients, which governs the non-Gaussianity of the simulated stochastic process, is estimated using state-of-the-art probabilistic methods such as inference of marginal distributions and vine copulas [TMES19, TMS19]. By resampling the estimated joint distribution, new realizations of the stochastic process can be generated.

We demonstrate the performance of our approach on several benchmark problems and compare it to other recent developments for stochastic simulators [ZS19b, AHSW19].

References

[AHSW19] S. Azzi, Y. Huang, B. Sudret, and J. Wiart. Surrogate modeling of stochastic functions – Application to computational electromagnetic dosimetry. *Int. J. Unc. Quantif.*, 9(4), 2019.

[IDM06] U. Iemma, M. Diez, and L. Morino. An extended Karhunen-Loève decomposition for modal identification of inhomogeneous structures. *J. Vib. Acoust.*, 128(3):357–365, 2006.

[MIDVR12] A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Stat. Comput.*, 22:833–847, 2012.

[PHQ05] K. K. Phoon, H. W. Huang, and S. T. Quek. Simulation of strongly non-Gaussian processes using Karhunen–Loève expansion. *Prob. Eng. Mech.*, 20(2):188–198, 2005.

[PZ14] F. Poirion and I. Zentner. Stochastic model construction of observed random phenomena. *Prob. Eng. Mech.*, 36:63–71, 2014.

[TMES19] E. Torre, S. Marelli, P. Embrechts, and B. Sudret. Data-driven polynomial chaos expansion for machine learning regression. *J. Comput. Phys*, 388:601–623, 2019.

[TMS19] E. Torre, S. Marelli, and B. Sudret. UQLab user manual – Statistical inference. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019. Report# UQLab-V1.3-114.

[ZS19a] X. Zhu and B. Sudret. Emulating the response PDF of stochastic simulators using sparse generalized lambda models. In *MascotNum Workshop, IFPEN (Rueil-Malmaison, France), March 18-20 (poster)*, 2019.

[ZS19b] X. Zhu and B. Sudret. Surrogate models of stochastic simulators using generalized lambda distributions. *Int. J. Unc. Quantif.*, 2019. Submitted.

Short biography – Nora Lüthen has studied mathematics with an emphasis on numerical mathematics at the University of Bonn, Germany. Since 2018, she is a PhD student in Prof. Sudret's Chair of Risk, Safety and Uncertainty Quantification at ETH Zürich in Switzerland. Her PhD is part of the project "Surrogate modelling for stochastic simulators" funded by the Swiss National Science Foundation.