

DISS. ETH NO. 28965

# Surrogate Modeling for Stochastic Simulators Using Statistical Approaches

A THESIS SUBMITTED TO ATTAIN THE DEGREE OF  
DOCTOR OF SCIENCES OF ETH ZURICH  
(DR. SC. ETH ZURICH)

PRESENTED BY

XUJIA ZHU

DIPL. ING., ECOLE POLYTECHNIQUE  
M.SC., TECHNISCHE UNIVERSITÄT MÜNCHEN  
BORN ON 15.01.1991  
CITIZEN OF P.R. CHINA

ACCEPTED ON THE RECOMMENDATION OF

PROF. DR. BRUNO SUDRET, EXAMINER  
PROF. DR. JOSSELIN GARNIER, CO-EXAMINER  
PROF. DR. BOZIDAR STOJADINOVIĆ, CO-EXAMINER  
PROF. DR. MARCO BROCCARDO, CO-EXAMINER  
DR. STEFANO MARELLI, CO-EXAMINER

2023

# Abstract

Nowadays, more and more complex interdependent infrastructures and networks are developed in engineering. The design and maintenance of such systems increasingly call for advanced computational models to optimize their performance and assess their reliability under various operational conditions. Unlike many conventional simulators that are deterministic, stochastic simulators feature intrinsic stochasticity. More precisely, they produce different results when run multiple times with a given set of input parameters. Due to this random nature, repeated model evaluations of the stochastic model with the same input value, called replications, are necessary to fully characterize the probability distribution of the associated model response.

For the purpose of optimization or uncertainty quantification (e.g., uncertainty propagation or sensitivity analysis), computational models typically need to be evaluated a large number of times. The additional layer of randomness due to the intrinsic stochasticity of stochastic simulators makes it even more computationally demanding to perform these complex analyses. A common practice to alleviate the prohibitive cost associated with expensive simulators is to build surrogate models, which behave similarly to the original model but are much cheaper to evaluate.

Contrary to the deterministic case, surrogate modeling of stochastic simulators has only emerged in the past decade. The main challenge in this field is that one model evaluation yields only a single realization of the random model response associated with the given input value. In other words, one run of a stochastic simulator provides proportionally much less information than that of a deterministic one.

This thesis focuses on developing efficient and accurate surrogate models to emulate the response distribution of stochastic simulators, combining statistical methods with state-of-the-art deterministic surrogate modeling techniques.

To this end, we propose two new approaches: the generalized lambda model (GLaM) and the stochastic polynomial chaos expansion (SPCE). The first one capitalizes on the use of the generalized lambda distribution to characterize the random nature of the simulator response. The distribution parameters are functions of the input variables and are represented by polynomial chaos expansions (PCEs). We explore replication-based methods to build GLaMs and improve their performance by an additional joint optimization of the overall likelihood function. We further elaborate this idea and develop a new method that does not require replications. Using this surrogate, we investigate sensitivity analysis for stochastic simulators.

The second class of stochastic surrogates, SPCE, overcomes the main shortcoming of GLaM, which is unable to represent multimodal distributions. In this more versatile stochastic emulator, we extend PCE by introducing an artificial latent variable to the expansion and an additive noise variable to mimic the intrinsic stochasticity of the simulator. We also propose an adaptive algorithm to construct the surrogate model without the need for replications.

For both stochastic surrogate models, we investigate basic theoretical properties of the primary estimation method. Analytical examples and engineering applications, including wind turbine design and seismic fragility analysis, are used to validate and illustrate the performance of the new approaches. Furthermore, these engi-

neering case studies provide valuable insights into the applicability of the developed framework to real-world industrial problems.

# Résumé

De nos jours, de plus en plus d'infrastructures complexes et de réseaux interdépendants sont développés en ingénierie. La conception et la maintenance de ces systèmes font appel à des modèles numériques avancés afin d'optimiser leur performances et d'évaluer leur fiabilité dans diverses conditions opérationnelles. Contrairement à de nombreux simulateurs conventionnels qui sont déterministes, les simulateurs dits stochastiques possèdent une stochasticité intrinsèque. Plus précisément, ils produisent des résultats différents lorsqu'ils sont exécutés plusieurs fois avec les mêmes paramètres d'entrée. En raison de ce caractère aléatoire, des évaluations répétitives d'un modèle stochastique avec la même valeur d'entrée, appelées répliques, sont nécessaires pour caractériser entièrement la loi de probabilité de la réponse associée.

Dans un contexte d'optimisation ou de quantification des incertitudes (e.g., propagation des incertitudes ou analyse de sensibilité), les modèles numériques doivent généralement être évalués un grand nombre de fois. La couche aléatoire supplémentaire due à la stochasticité intrinsèque des simulateurs stochastiques rend ces analyses encore plus exigeantes sur le plan des ressources informatiques. Une façon courante d'alléger les coûts de calculs prohibitifs consiste à construire des métamodèles, qui se comportent de manière similaire au modèle original, mais dont l'évaluation est beaucoup moins coûteuse.

Contrairement aux métamodèles développés pour les simulateurs déterministes, la métamodélisation des simulateurs stochastiques n'est apparue qu'au cours de la dernière décennie dans diverses applications d'ingénierie. L'enjeu principal dans ce domaine est qu'une évaluation du modèle ne produit qu'une seule réalisation de la réponse aléatoire associée à la valeur d'entrée. En d'autres termes, une exécution d'un simulateur stochastique fournit beaucoup moins d'information que dans le cas d'un simulateur déterministe.

Cette thèse est dédiée au développement de métamodèles efficaces et précis pour émuler la loi de probabilité de la réponse de simulateurs stochastiques, en combinant des méthodes statistiques avec des techniques de pointe de métamodélisation déterministe.

Pour cela, nous proposons deux métamodèles : le modèle « lambda généralisé » (MLaG) et les polynômes de chaos stochastiques (PCS). Le premier modèle capitalise sur l'utilisation de la loi lambda généralisée pour caractériser la réponse aléatoire. Les paramètres de la distribution sont des fonctions des variables d'entrée et sont représentés par polynômes de chaos. Nous explorons des méthodes basées sur la réplique pour construire des MLaGs et améliorer leurs performances par une optimisation conjointe supplémentaire de la fonction de vraisemblance globale. Nous approfondissons cette idée et développons une nouvelle méthode qui ne nécessite pas de répliques. À l'aide de ce métamodèle, nous étudions l'analyse de sensibilité pour simulateurs stochastiques.

La deuxième catégorie de métamodèles stochastiques PCS permet de représenter des distributions multimodales, ce que ne permet pas l'approche MLaG. Dans cet émulateur stochastique plus flexible, nous introduisons une variable latente artificielle dans l'expansion des polynômes de chaos et une variable de bruit additive pour imiter la stochasticité intrinsèque du simulateur. Nous proposons également un algorithme adaptatif pour construire ce modèle sans avoir besoin de répliques.

Pour les deux métamodèles stochastiques proposés, nous étudions certaines propriétés théoriques sur la

méthode d'estimation principale. Des exemples analytiques et des applications d'ingénierie, y compris la conception d'éoliennes et l'analyse de fragilité sismique, valident et illustrent la performance des nouvelles approches. En outre, ces études des cas d'ingénierie fournissent des indications précieuses sur l'applicabilité des méthodes développées à des problèmes industriels réels.