

THÈSE DE DOCTORAT

UNCERTAINTY QUANTIFICATION IN MULTI-PHYSICS
MODEL FOR WIND TURBINE ASSET MANAGEMENT

Elias FEKHARI

EDF R&D, Université Nice Côte d'Azur

Présentée en vue de l'obtention du grade de docteur en Automatique, traitement du signal et des images d'Université Côte d'Azur, dirigée par Bertrand Iooss, soutenue le 12 mars 2024.

Devant le jury composé de :

Directeur	Bertrand IOOSS	Chercheur senior	EDF R&D, Chatou
Co-encadrant	Vincent CHABRIDON	Ingénieur-chercheur	EDF R&D, Chatou
Rapporteurs	Franck SCHOEFS	Professeur des universités	Nantes Université, Nantes
	Daniel STRAUB	Professeur des universités	TUM, Munich
Examinateurs	Mireille BOSSY	Directrice de recherche	INRIA, Sophia-Antipolis
	Sébastien DA VEIGA	Professeur associé	ENSAI, Rennes
	Bruno SUDRET	Professeur des universités	ETH, Zürich
Invités	Joseph MURÉ	Ingénieur-chercheur	EDF R&D, Chatou
	Anaïs LOVERA	Ingénierie-chercheur	EDF R&D, Saclay





UNCERTAINTY QUANTIFICATION IN MULTI-PHYSICS MODEL FOR WIND TURBINE ASSET MANAGEMENT

Elias FEKHARI

ÉLECTRICITÉ DE FRANCE R&D

Chatou, France

&

CÔTE D'AZUR UNIVERSITY

Nice, France

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
(in Computer Science)

publicly defended on March 12, 2024 in front of the following jury:

Director	Dr. Bertrand IOOSS	Senior Researcher	EDF R&D, Chatou
Co-advisors	Dr. Vincent CHABRIDON Dr. Joseph MURÉ	Research Engineer Research Engineer	EDF R&D, Chatou
Reviewers	Pr. Franck SCHOEFS Pr. Daniel STRAUB	Professor Professor	Nantes Université, Nantes TUM, Munich
Examiners	Pr. Mireille BOSSY Dr. Sébastien DA VEIGA Pr. Bruno SUDRET	Research Director Associate Professor Professor	INRIA, Sophia-Antipolis ENSAI, Rennes ETH, Zürich
Invite	Dr. Anaïs LOVERA	Research Engineer	EDF R&D, Saclay

Contents

List of Figures	vii
List of Tables	ix
Introduction	1
I Introduction to uncertainty quantification and wind energy	11
1 Uncertainty quantification in computer experiments	13
1.1 Introduction	14
1.2 Black-box model specification	14
1.3 Enumerating and modeling the uncertain inputs	15
1.3.1 Sources of the input uncertainties	15
1.3.2 Modeling uncertain inputs with the probabilistic framework	16
1.3.3 Joint input probability distribution	17
1.4 Uncertainty propagation for central tendency study	20
1.4.1 Numerical integration	20
1.4.2 Numerical design of experiments	26
1.4.3 Summary and discussion	30
1.5 Uncertainty propagation for rare event estimation	31
1.5.1 Problem statement	32
1.5.2 Rare event estimation methods	34
1.5.3 Summary and discussion	43
1.6 Global sensitivity analysis	45
1.6.1 Screening methods	46
1.6.2 Variance-based importance measures	47
1.6.3 Moment-independent importance measures	50
1.6.4 Summary and discussion	52
1.7 Surrogate modeling	52
1.7.1 Common framework	52

1.7.2	General purposes surrogate model	54
1.7.3	Goal-oriented active surrogate model	57
1.7.4	Summary and discussion	59
1.8	Conclusion	60
II	Contributions to uncertainty quantification and propagation	63
III	Contributions to rare event estimation	65
2	Adaptive rare event estimation using Bernstein copula	67
2.1	Introduction	68
2.2	Bernstein adaptive nonparametric conditional sampling	69
2.3	Numerical experiments	72
2.3.1	Analytical toy-cases	72
2.3.2	Benchmark results and analysis	74
2.4	Reliability-oriented sensitivity analysis	77
2.4.1	Target and conditional HSIC	77
2.4.2	ROSA as a post-treatment of BANCS reliability analysis	79
2.5	Conclusion	79
3	Application to wind turbine fatigue reliability and robustness	83
3.1	Introduction	84
3.2	Surrogate modeling for reliability analysis	84
3.2.1	High-performance computer evaluation	84
3.2.2	Design of experiments	84
3.2.3	Gaussian process regression	86
3.3	Reliability and robustness analysis	86
3.3.1	Nominal reliability analysis	86
3.3.2	Robustness analysis using the perturbed-law sensitivity indices	86
3.4	Conclusion	88
Conclusion and perspectives		91
Bibliography		93
Appendix A	Univariate distribution fitting	103
Appendix B	Dissimilarity measures between probability distributions	107
Appendix C	Advanced rare event estimation algorithms	113
Appendix D	Uncertainty quantification practice with OpenTURNS	115

Appendix E Résumé étendu de la thèse

119

List of Figures

1	General uncertainty quantification framework (De Rocquigny et al. (2008), adapted by Ajenjo (2023))	3
1.1	Samples of three joint distributions with identical marginals and different dependence structures	18
1.2	Ranked samples represented in the Fig. 1.1	18
1.3	Univariate quadratures nodes for increasing sizes ($1 \leq n \leq 15$)	22
1.4	Two identical univariate Gauss-Legendre quadratures combined as a tensor product (left) and a Smolyak sparse grid (right).	23
1.5	Nested Monte Carlo and quasi-Monte Carlo designs ($n = 256$)	26
1.6	Latin hypercube designs with poor and optimized space-filling properties ($n = 8$)	29
1.7	One-dimensional reliability analysis example	32
1.8	FORM and SORM approximation on a two-dimensional reliability problem .	35
1.9	Multi-FORM approximation on an example with two MPFPs	36
1.10	Illustration of a rare event estimation.	39
1.11	Illustration of a rare event estimation by subset simulation ($n = 4 \cdot 10^4, p_0 = 0.1$). .	44
1.12	Illustration of a k -fold cross-validation (with $k = 4$)	54
1.13	Illustration of an ordinary kriging model fitted on a limited set of observations ($n = 7$). The predictor is represented in and several trajectories of the conditioned Gaussian process are drawn and represented in purple.	56
1.14	Illustration of the expected improvement learning criterion	58
1.15	Illustration of the deviation number learning criterion	60
2.1	BANCS algorithm applied to toy-case #1: illustration of conditional sampling and nonparametric fit at the first and second iterations.	71
2.2	BANCS iterations on the two-dimensional reliability problems (for $N = 10^4$ and $p_0 = 0.1$). Only the samples exceeding the intermediary thresholds are represented.	75
2.3	Cross-cut visualization of the function of test-case #5 through FORM's most-probable failure point P^* . The limit-state function (full line) delimits the safe domain (in blue) and the failure domain (in red). FORM produces the Taylor approximation around P^* (dashed lines).	75

2.4	Todo.	76
2.5	Target and conditional HSIC as a post-treatment of BANCS reliability analysis of test-case #3. The consecutive samples from BANCS are denoted by $\{S_k\}_{k=1}^{k_{\#}}$ (each with size $N = 5 \times 10^3$).	80
2.6	Target and conditional HSIC as a post-treatment of BANCS reliability analysis of test-case #5. Applied to the nested samples from BANCS (denoted by $\{S_k\}_{k=1}^{k_{\#}}$), each with size $N = 5 \times 10^3$	80
3.1	Todo.	85
3.2	Todo.	85
3.3	Todo.	86
3.4	Todo.	87
3.5	Todo.	87
3.6	Todo.	88
3.7	Todo.	89
A.1	Adequation of two different Weibull models using their likelihood with a sample of observations (black crosses).	104
A.2	Fit of a bimodal density by KDE using different tuning parameters.	105
A.3	QQ-plot between the data from Example 2 and a KDE model.	106
B.1	Kernel mean embedding of a continuous and discrete probability distribution .	109
E.1	Schéma générique de la quantification des incertitudes (De Rocquigny et al. (2008), adapté par Ajenjo (2023))	122

List of Tables

2.1	Random inputs from the toy-case #5.	73
3.1	Results of the numerical experiments (subset sample size $N = 10^4$, $p_0 = 0.1$). . .	88

Introduction

Industrial context and motivation

The current challenge of energy transition involves, among other things, reducing the share of fossil fuels in the global electricity mix. In this context, offshore wind energy offers several advantages ([Beauregard et al., 2022](#)). Offshore energy benefits from more consistent winds than onshore, mainly due to the absence of terrain roughness, it also makes possible the installation of larger and more powerful wind turbines. Since the construction of the first offshore wind farm in Vindeby, Denmark, in 1991, the industry has experienced rapid growth, with a total capacity of 56 GW in operation worldwide in 2021. Over time, offshore wind technology has matured, resulting in significant achievements such as securing projects in Europe through “zero-subsidy bids” where the electricity produced is directly sold on the wholesale market ([Beauregard et al., 2022](#)).

However, despite the progress of this sector, scaling limitations and numerous scientific challenges emerge. To meet ambitious national and regional development targets, the wind energy industry must address various scaling issues, including port logistics, the demand for critical natural resources, and sustainable end-of-life processes. Furthermore, the field presents several scientific challenges that often involve coupling data with numerical simulations of physical systems and their surrounding environment. The wind energy community is focused on different objectives, including enhancing the design of floating offshore wind turbines, refining wind resource estimation techniques, and optimizing maintenance operations. In general, several decisions are made throughout the lifespan of a wind turbine by its designer, installer, and operator, all while having only partial knowledge of certain physical phenomena. Therefore, modeling and controlling the various sources of uncertainties associated with offshore wind energy proved to be a key success factor in this highly competitive industry.

Overall, the offshore wind industry needs methods for uncertainty management regarding safety margins and industrial asset management (at the component, wind turbine, and overall wind farm levels) ([Van Kuik et al., 2016](#)). For wind project developers, the primary focus is on improving the wind potential assessment of candidate sites by combining various sources of information and modeling the multivariate distribution of environmental conditions. In the case of floating wind projects, the goal is to incorporate a probabilistic aspect from the design phase

(e.g., of the floaters) to define safer, more robust, and more cost-effective solutions. For wind farm owners, end-of-life management is another significant concern. An owner of a wind farm at the end of its life has three options: extend the operational life of assets, replace current wind turbines with newer models, or decommission and sell the wind farm. The first two options require evaluating the structural reliability and residual lifespan, with quantitative assessments reviewed by certification bodies and insurers to issue operating permits. To provide rigorous risk assessments, the generic methodology of *uncertainty quantification methodology* is a widely accepted approach in industrial sectors facing these types of issues (De Rocquigny et al., 2008; Blanchard et al., 2023).

Generic methodology for uncertainty quantification

Computer experiment is a discipline that emerged with the advent of informatics. This practice produces numerical models that allow the simulation of complex system behavior based on initial conditions defined by the analyst. Numerical models quickly became essential for the analysis, design, and certification of complex systems in cases where experiments or physical measurements are too costly or even unfeasible. However, such numerical models are mostly deterministic: the reproducible result of a simulation is associated with a fixed input set of parameters. The issue of managing uncertainties associated with these inputs arises when performing analysis with numerical models.

Uncertainty quantification aims at modeling and controlling uncertainties around a numerical model. To do so, a generic methodology has been proposed to quantify and analyze uncertainties between input and output variables of a numerical model (De Rocquigny et al., 2008; Blanchard et al., 2023). An overview of the mathematical tools used in this field is provided by Sullivan (2015). This approach improves the understanding of a system, ultimately contributing to more robust decision-making. Figure 1 illustrates the main steps of the generic uncertainty quantification method, which are briefly summarized hereafter:

- **Step A – Problem specification:** This step involves identifying the system under study and constructing a numerical model capable of precisely simulating its behavior. Specifying the problem also involves the definition of a set of parameters inherent to the numerical model. These parameters include both the input variables and the output variables generated by the simulation. In this document, the numerical model is considered a black box, in contrast to approaches that are integrated within the numerical solution schemes for the system's behavioral equations (referred to as intrusive approaches (Le Maître and Knio, 2010)). Generally, these numerical models are first calibrated against measured data and pass a process of validation and verification to reduce modeling errors (see e.g., (Oberkampf and Roy, 2010; Damblin, 2015; Carmassi, 2018)).

- **Step B – Uncertainty modeling:** The objective of the second step is to identify and model all the sources of uncertainty related to the input variables. Most of the time the uncertainty modeling is done in the probabilistic framework.
- **Step C – Uncertainty propagation:** This step consists of propagating the uncertain inputs through the computer model. Consequently, the output of the numerical model (commonly scalar) also becomes uncertain. The goal is to estimate a quantity of interest, which is a statistic related to the studied random output variable. The uncertainty propagation method may differ depending on the quantity of interest targeted (e.g., central tendency, a quantile, a rare event probability, etc.).
- **Step C' – Inverse analysis:** In this additional step, a sensitivity analysis can be performed to study the role allocated to each uncertain input leading to the uncertain output.
- **Metamodeling:** Considering the high computational cost associated with some simulations, statistical approaches emulate these expensive simulators with a limited number of simulations. Uncertainty quantification can then be carried out using a “surrogate model” (or metamodel) for a reduced computational cost. This optional step of statistical learning is not strictly a part of uncertainty quantification, but it often proves to be essential for enabling its practical implementation.

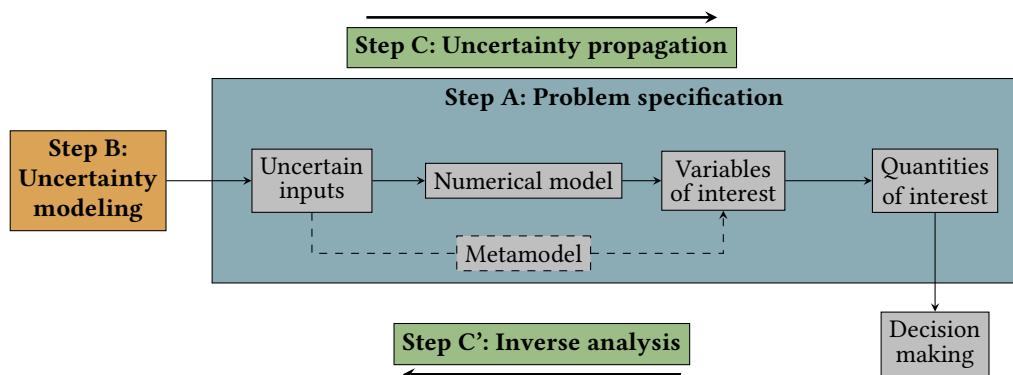


Figure 1 General uncertainty quantification framework ([De Rocquigny et al. \(2008\)](#), adapted by [Ajenjo \(2023\)](#))

Problem statement and outline of the thesis

Risk and uncertainty management in the field of wind energy is a significant concern for the electric utility Électricité de France (EDF). This thesis aims to adapt and apply the generic uncertainty quantification methodology to industrial offshore wind energy studies. As such, this use case raises scientific challenges related to its specific characteristics, described in the following:

- The numerical model exploited in the present work consists of a series of numerical models executed sequentially. This chain is divided into three parts: first, a temporal and stochastic generation of wind and wave velocity fields, followed by the simulation of the coupled hydro-aero-servo-elastic behavior of the wind turbine, and finally a post-processing phase to obtain scalar quantities of interest, aggregated over the temporal outputs.
- The complexity of this simulator, along with the high unit computational cost (about 40 minutes per simulation), requires the use of efficient sampling methods and high-performance computing systems. In addition to the complexity associated with the numerical model, modeling the input uncertainties also represents a challenge. Indeed, the joint distribution associated with environmental conditions presents a complex dependence structure. The quality of the inference step is critical as it directly impacts the conclusions of uncertainty propagation.

In order to apply the generic methodology for uncertainty quantification to the offshore wind turbine case, this thesis aims to answer the following questions:

- Q1.** *How to accurately model the dependence structure associated with the joint environmental distribution?* (⇒ Step B)
- Q2.** *How to perform uncertainty propagation through a computationally expensive numerical chain uniquely based on an empirical description (measured data) of input uncertainties?* (⇒ Step C)
- Q3.** *How to estimate rare event probabilities related to the fatigue failure of offshore wind turbine structures?* (⇒ Step C)
- Q4.** *How to assess and analyze the sensitivity of uncertain inputs regarding quantities of interest resulting from structural reliability (i.e., reliability-oriented sensitivity analysis)?* (⇒ Step C')

To propose an answer to these questions, this manuscript is divided into three parts. The first part offers an introduction to uncertainty quantification methods and offshore wind turbine numerical modeling. The second part presents the contributions of this thesis to uncertainty quantification and propagation while the third part describes the contributions to rare event estimation. This manuscript is divided into seven chapters, which are summarized hereafter:

Chapter 1 – Introduction to uncertainty quantification. This chapter gives a brief overview of various topics in uncertainty quantification ([Sullivan, 2015](#)). After a reminder of some mathematical concepts, the model specification step is described, considering a black box and its input and output variables. The different types and sources of uncertainties are then presented, along with their modeling within a probabilistic framework. Uncertainty propagation depends on the estimated quantities of interest, therefore, one section addresses propagation methods for

central tendency studies, and another focuses on rare event probability estimation (a statistic related to the tails of output distributions). The section dedicated to central tendency presents numerical integration, sampling, and design of experiment methods (Fang et al., 2018). The one about rare event probabilities introduces usual methods from the field of structural reliability (Lemaire et al., 2009; Morio and Balesdent, 2015).

This chapter also covers the main methods for global sensitivity analysis (Da Veiga et al., 2021). This field divides its methods into two major classes: screening methods and importance measures. Screening techniques typically applied in high-dimensional problems, aim to identify variables with low impact on the variability of the output of interest. Importance measures, on the other hand, quantitatively allocate, for each input variable, a share of the output variability, enabling the ranking of variables based on their influence.

Finally, this chapter presents an overview of the families of metamodels commonly used in uncertainty quantification (Forrester et al., 2008). Special attention is given to the Gaussian process regression, which involves conditioning a Gaussian process on a set of observations from the numerical model. Once conditioned, the Gaussian process provides richer information than other types of metamodels. This method simultaneously offers a surrogate model (mean of the Gaussian process, also called predictor) and an error function (variance of the process). Some iterative methods (called “active”) use this additional information to progressively enrich the metamodel and improve its predictability. These techniques were quite successful in the 1990s for solving optimization problems with expensive functions (Jones et al., 1998). Since then, their use has expanded to solve problems in structural reliability Echard et al. (2011).

Chapter 2 – Introduction to wind turbine modeling and design. The simulation of an offshore wind turbine involves modeling multiple physical aspects interacting with random environmental conditions. This chapter first introduces spectral methods used to generate wind and wave velocity fields by applying inverse Fourier transforms (e.g., as implemented in the TurbSim tool (Jonkman, 2009)). These simulated wind velocity fields then become the inputs of a multi-physics wind turbine numerical model. Such simulation includes simplified modeling of the interactions between fluids and structures (using the blade element momentum theory), dynamic modeling of the structure using flexible multibody methods, and modeling of wind turbine control systems [source]. The numerical code studied generates a time series of several physical quantities describing the system’s behavior.

This thesis particularly focuses on the probabilistic evaluation of fatigue damage in wind turbine structures. Fatigue damage is a phenomenon that deteriorates the mechanical properties of a material as a result of exposure to many cyclic low-amplitude stresses. Currently, standards recommend the use of deterministic safety factors to address this failure mode (IEC-61400-1, 2019; DNV-ST-0437, 2016). A probabilistic approach enhances the analysis and can sometimes reveal conservative safety margins. Several recent studies have addressed this topic from different methodological perspectives (Huchet, 2019; Lataniotis, 2019; Cousin, 2021; Hirvoas, 2021; Petrovska, 2022).

In this context, this chapter enumerates the input parameters of the calculation chain that are considered uncertain. These random variables are grouped into two groups: the random vector related to the environment (e.g., average wind speed, wind speed standard deviation, wind direction, significant wave height, wave period, and wave direction), and the random vector related to the system (e.g., controller wind misalignment error, soil stiffness, fatigue calculation curve parameters).

Chapter 3 – Kernel-based uncertainty quantification. This chapter examines perturbations in environmental conditions within an offshore wind farm induced by wake effects [Larsen et al. \(2008\)](#). A theoretical offshore wind farm off the southern coast of Brittany is considered as a use case, and a simplified numerical model of wake in this wind farm is used. This model provides an analytical prediction of the wind speed deficit and turbulence created by the wake, taking into account the influence of the floaters' positions due to rigid body dynamics.

In the second phase, uncertainty propagation is carried out through the wake model, considering the joint distribution of ambient environmental conditions as inputs. In the end, an environmental distribution perturbed by the wake is simulated for each wind turbine. A dissimilarity measure between distributions, based on kernels and named the *maximum mean discrepancy* (MMD), is used to compare the distributions perceived by each wind turbine. This measure allows the clustering of wind turbines exposed to similar environmental conditions, resulting in identical structural responses. Given the high computational cost of aero-servo-hydro-elastic simulations for offshore wind turbines, this preliminary study enables reliability analysis at the wind farm scale without repeating the analysis for each turbine. Ultimately, only four classes are selected to represent a wind farm of 25 turbines.

Chapter 4 – Kernel-based central tendency estimation. Chapter four presents the use of the kernel-based dissimilarity measure (MMD) in the context of probability distribution sampling, a method known as "kernel herding" introduced by [Chen et al. \(2010\)](#). This quadrature technique belongs to the family of "Bayesian quadratures" [Briol et al. \(2019\)](#), which can be viewed as a generalization of quasi-Monte Carlo methods [Li et al. \(2020\)](#).

The properties of this method are highlighted through an industrial application dedicated to estimating the mean fatigue damage of a wind turbine structure. Although this quantity is crucial in the design and certification of wind turbines, the methods used to estimate it are known to be suboptimal (i.e., regular grids). The study is conducted on a model of a fixed offshore wind turbine belonging to a farm in the North Sea. Uncertainties in input environmental conditions are inferred from in-situ measured data.

Finally, a numerical comparison with Monte Carlo and quasi-Monte Carlo sampling reveals the performance and practical advantages of kernel herding. This method allows for direct subsampling from a large environmental database without the need for inference (step B).

Chapter 5 – Kernel-based metamodel validation. This chapter proposes the use of kernel-based sampling methods in the context of model validation for machine learning (or surrogate

models). Estimating the predictivity of supervised learning models requires an evaluation of the learned surrogate model on a set of test points that were not used during training. The quality of the validation naturally depends on the properties of the test set and the metric used to summarize the prediction error. This contribution first suggests using space-filling sampling methods to “optimally” select a test set, then, it introduces a new predictivity coefficient that weights the observed errors to improve the global error estimation. A numerical comparison between several sampling methods based on geometric approaches (Shang and Apley, 2020) or kernel methods Chen et al. (2010); Mak and Joseph (2018) is carried out. Our results show that weighted versions of kernel methods offer superior performance. An application to simulated mechanical loads in an offshore wind turbine model is also presented. This experiment illustrates the practical relevance of this technique as an effective alternative to costly cross-validation techniques.

Chapter 6 – Nonparametric rare event estimation. Estimating rare event probabilities is a common issue in industrial risk management, especially in the field of structural reliability (Chabridon, 2018). To address this, several techniques have been proposed to overcome the known limitations of the Monte Carlo method. Among them, “subset simulation” (Au and Beck, 2001) is a technique based on the split of a rare probability into a product of less rare (and thus easier to estimate) conditional probabilities associated with nested failure events. However, this technique relies on conditional simulation using Markov chain Monte Carlo (MCMC) methods. These algorithms, while converging, often produce samples that are not independent and identically distributed (i.i.d.) due to the correlation between the Markov chains. In this chapter, another conditional sampling method is proposed, with the advantage of preserving the i.i.d. property. Independent sampling is particularly relevant for reusing these samples in a posterior reliability-oriented sensitivity analysis. The algorithm introduced is based on the non-parametric inference of the conditional joint distribution using kernel density estimation of marginals combined with dependence inference using the empirical Bernstein copula (Sancetta and Satchell, 2004). The so-called “Bernstein adaptive nonparametric conditional sampling” (BANCS), is compared to the subset simulation method for several structural reliability problems. The initial results are promising, but further investigation is needed to control the estimator’s bias.

Chapter 7 – Sequential reliability oriented sensitivity analysis. This chapter deals with sensitivity analysis for risk measures (e.g., rare event probabilities). Global sensitivity analysis (Da Veiga et al., 2021) assigns a portion of the global output variability to each variable (or group of variables), often using a functional decomposition of the output variance. However, when studying risk measures (often located in the distributions’ tails), the global sensitivity might be very different from the sensitivity to the risk measure. “Reliability-oriented sensitivity analysis” (ROSA), studies the impact of the inputs in regard to a risk measure such as a rare event probability (see e.g., Chabridon (2018)). Using the nested subsets obtained with the BANCS algorithm (presented in Chapter 6), the idea of this chapter is to study the ROSA evolution

as the subsets get closer to the failure domain. For each subset, a ROSA is carried out with a kernel-based importance measure called the “Hilbert-Schmidt Independence Criterion” adapted to this context ([Marrel and Chabridon, 2021](#)).

Numerical developments

Several implementations developed in this thesis are available on different platforms, allowing the reader to reproduce some numerical results in an open-data approach:

- This Python package generates designs of experiments based on kernel methods such as Kernel Herding and Support Points. A tensorized implementation of the algorithms was proposed, significantly increasing their performances. Additionally, optimal weights for Bayesian quadrature are provided.

- This Python package, developed in collaboration with J.Muré, is available on the platform Pypi and fully documented.

-
- This Python package proposes an implementation of the “Bernstein Adaptive Non-parametric Conditional Sampling” method for rare event estimation.

- This Python package is available on the PyPI platform and is illustrated with examples and analytical benchmarks.

-
- This Python package presents a standardized process to benchmark different sampling methods for central tendency estimation.

- This Python package is available on a GitHub repository with analytical benchmarks.

-
- This Python package proposes an implementation of a synthetic visualization tool for multivariate distributions.

- This Python package, developed in collaboration with V.Chabridon, is available on the Pypi platform.

¹Documentation: <https://efekhari27.github.io/otkerneldesign/master/>

²Repository: <https://github.com/efekhari27/bancs>

³Repository: <https://github.com/efekhari27/ctbenchmark>

⁴Repository: <https://github.com/efekhari27/copulogram>

Publications and communications

The research contributions in this manuscript are based on the following publications:

Book Chap.	E. Fekhari, B. Iooss, J. Muré, L. Pronzato and M.J. Rendas (2023). “Model predictivity assessment: incremental test-set selection and accuracy evaluation”. In: <i>Studies in Theoretical and Applied Statistics</i> , pages 315–347. Springer.
Jour. Pap.	E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature”. In: <i>Data-Centric Engineering</i> .
Int. Conf. Pap.	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). “Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool”. In: <i>Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)</i> , Lisbon, Portugal. (Paper & Talk)
	E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Bernstein adaptive nonparametric conditional sampling: a new method for rare event probability estimation” ⁵ . In: <i>Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)</i> , Dublin, Ireland. (Paper & Talk)
	E. Vanem, E. Fekhari, N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2023). “A joint probability distribution model for multivariate wind and wave conditions”. In: <i>Proceedings of the ASME 2023 42th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2023)</i> , Melbourne, Australia. (Paper)
	A. Lovera, E. Fekhari, B. Jézéquel, M. Dupoirion, M. Guiton and E. Ardillon (2023). “Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis”. In: <i>Journal of Physics: Conference Series (WAKE 2023)</i> , Visby, Sweden (Paper)
Int. Conf. Short Abs.	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). “Numerical Studies of Bayesian Quadrature Applied to Offshore Wind Turbine Load Estimation”. In: <i>SIAM Conference on Uncertainty Quantification (SIAM UQ22)</i> , Atlanta, USA. (Talk)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). “Model predictivity assessment: incremental test-set selection and accuracy evaluation”. In: <i>22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)</i> , Trondheim, Norway. (Talk)
Nat. Conf.	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). “Kernel-based quadrature applied to offshore wind turbine damage estimation”. In: <i>Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)</i> , Clermont-Ferrand, France (Poster)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2023). “Rare event estimation using nonparametric Bernstein adaptive sampling”. In: <i>Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)</i> , Le Croisic, France (Talk)
Invited Lec.	Le Printemps de la Recherche 2022, Nantes, France. “Traitement des incertitudes pour la gestion d’actifs éoliens”. (Talk)
	Journées Scientifiques de l’Eolien 2024, Saint-Malo, France. “Evaluation probabiliste de la fiabilité en fatigue des structures éoliennes en mer”. (Talk)

⁵This contribution was rewarded by the “CERRA Student Recognition Award”

PART I:

INTRODUCTION TO UNCERTAINTY QUANTIFICATION AND WIND ENERGY

Toute pensée émet un coup de dé.

S. MALLARMÉ

Chapter **1**

Uncertainty quantification in computer experiments

1.1	Introduction	14
1.2	Black-box model specification	14
1.3	Enumerating and modeling the uncertain inputs	15
1.3.1	Sources of the input uncertainties	15
1.3.2	Modeling uncertain inputs with the probabilistic framework	16
1.3.3	Joint input probability distribution	17
1.4	Uncertainty propagation for central tendency study	20
1.4.1	Numerical integration	20
1.4.2	Numerical design of experiments	26
1.4.3	Summary and discussion	30
1.5	Uncertainty propagation for rare event estimation	31
1.5.1	Problem statement	32
1.5.2	Rare event estimation methods	34
1.5.3	Summary and discussion	43
1.6	Global sensitivity analysis	45
1.6.1	Screening methods	46
1.6.2	Variance-based importance measures	47
1.6.3	Moment-independent importance measures	50
1.6.4	Summary and discussion	52
1.7	Surrogate modeling	52
1.7.1	Common framework	52
1.7.2	General purposes surrogate model	54
1.7.3	Goal-oriented active surrogate model	57
1.7.4	Summary and discussion	59
1.8	Conclusion	60

1.1 Introduction

The progress of computer simulation gradually allows the virtual resolution of more complex problems in scientific fields such as physics, astrophysics, engineering, climatology, chemistry, or biology. This domain often provides a deterministic solution to complex problems depending on several inputs. Associating an uncertainty quantification (UQ) analysis with these numerical models is a key element in improving the understanding of the phenomena studied. A wide panel of UQ methods has been developed over the years to pursue these studies for a reasonable computational cost.

This chapter presents the essential tools and methods from the generic UQ framework, including elements partially inspired from [Sullivan \(2015\)](#) and [Chabridon \(2018\)](#). It is structured as follows: Section 1.2 describes the context of the model specification step; Section 1.3 presents a classification of the input uncertainties and the probabilistic framework to model them; Section 1.4 and 1.5 introduce various methods to propagate the input uncertainties through the numerical model for different purposes; Section 1.6 presents the main inverse methods to perform sensitivity analysis in our framework; Finally, Section 1.7 introduces the concept of surrogate models to emulate a model by realizing statistical learning on a limited dataset. Additionally, numerical examples of the important methods are implemented during this chapter using [OpenTURNS](#), a Python package for uncertainty quantification.

OpenTURNS¹. Is a high-performance Python library dedicated to UQ ([Baudin et al., 2017](#)). OpenTURNS (“Open source initiative for the Treatment of Uncertainties, Risks’N Statistics”) is developed by industrial researchers from EDF R&D, Airbus Group, PHIMECA Engineering, IMACS and ONERA. It combines high performance using C++ programming with high accessibility through a Python API. Overall, this open-source library provides tools for various steps of the UQ framework (e.g., uncertainty quantification, uncertainty propagation, surrogate modeling, reliability, sensitivity analysis and calibration). To guarantee software quality, the development follows robust processes such as exhaustive unit testing and multiplatform continuous integration. A dedicated forum hosts an active community, which helps new users and discusses future developments. Finally, no-code users can benefit from OpenTURNS’s free-download Graphical User Interface software, named [Persalys²](#). In this chapter, the methodological concepts introduced are linked to minimal OpenTURNS implementations examples, available in Appendix D.

1.2 Black-box model specification

In our computer experiments context, uncertainty quantification is performed around an input-output numerical simulation model. This numerical model, or code, is hereafter considered

¹OpenTURNS installation guide and documentation are available at <https://openturns.github.io/www/>

²Persalys, a free-download graphical user interface available at <https://www.persalys.fr/obtenir.php>

as *black-box* since the knowledge of the underlying physics doesn't inform the UQ methods. Alternatively, one could consider *intrusive* UQ methods, introducing uncertainties within the resolution of the equations of the physics (see e.g., [Le Maître and Knio 2010](#)). In practice, numerical models might be a sequence of codes executed in series to obtain a variable of interest.

While simulation models are in most cases deterministic, they may also be qualified as intrinsically stochastic (i.e., two runs of the same model taking the same inputs return different outputs). Additionally, numerical simulation always presents modeling errors. In the following, it will be assumed that the numerical models passed a *validation & verification* phase, to quantify their confidence and predictive accuracy.

Formally, part of the problem specification is the definition of the set of d input variables $\mathbf{x} = (x_1, \dots, x_d)^\top$ considered as uncertain (e.g., wind speed, wave period, etc.). The outputs studied are also defined at this stage, which will only be of scalar type in the present work. UQ methods suited to other types of outputs exist (see e.g., for time series outputs [Lataniotis 2019](#), for functional outputs [Audet et al. 2012; Rollón de Pinedo et al. 2021](#)). Let us then define the following numerical model \mathcal{M} , with its respective input \mathcal{D}_x and output domains \mathcal{D}_y :

$$\mathcal{M} : \begin{cases} \mathcal{D}_x \subseteq \mathbb{R}^d & \longrightarrow \mathcal{D}_y \subseteq \mathbb{R} \\ \mathbf{x} & \longmapsto y. \end{cases} \quad (1.1)$$

Unlike the typical machine learning input-output dataset framework, the UQ analyst can simulate the output image of any inputs (in the input domain), using a numerical model. However, numerical simulations often come with an important computational cost. Therefore, UQ methods should be efficient and require as few simulations as possible. In this context, surrogate models (or metamodels) are statistical approximations of the costly numerical model, that can be used to perform tractable UQ. Surrogate models are built and validated on a limited number of simulations (in a *supervised learning* framework). In practice, note that the model specification step is often associated with the development of a *wrapper* of the code. It is an overlay of code allowing its execution in a parametric way, which is often associated with *high-performance computer* (HPC) deployment. Once the model is specified, a critical step in uncertainty quantification is enumerating the input uncertainties and building their associated mathematical model.

1.3 Enumerating and modeling the uncertain inputs

1.3.1 Sources of the input uncertainties

The analyst should construct a list of uncertain inputs as exhaustive as possible, to ensure a complete risk assessment (e.g., associated with the exploitation of a wind energy asset). Even if these uncertainties might have different origins, they should all be considered jointly in the UQ study. Numerous authors proposed to classify them for practical purposes into two groups:

- **aleatory uncertainty** regroups the uncertainties arising from natural randomness (e.g., wind turbulence). From a risk management point of view, these uncertainties are qualified as *irreducible* since the industrials facing them will not be able to acquire additional information to reduce them (e.g., additional measures).
- **epistemic uncertainty** gathers the uncertainties resulting from a lack of knowledge (e.g., material properties). Contrarily to the aleatory ones, epistemic uncertainties might be reduced by investigating their origin (often at a certain cost).

Der Kiureghian and Ditlevsen (2009) discuss the relevance of this classification. They affirm that this split is practical for decision-makers to identify possible ways to reduce their uncertainties. However, it should not affect the way of modeling or propagating uncertainties. In the following, the probabilistic framework is introduced to deal with uncertainties.

1.3.2 Modeling uncertain inputs with the probabilistic framework

Uncertainties are traditionally modeled with objects from the probability theory. In this thesis, the *probabilistic framework* is adopted. Alternative theories exist to mathematically model uncertainties. For example, imprecise probability theory allows more general modeling of the uncertainties (Beer et al., 2013; Schöbi and Sudret, 2017). It becomes useful when dealing with very limited and possibly contradictory information (e.g., expert elicitation). The core probabilistic tools and objects are introduced hereafter.

The *probability space* is a measure space with total measure summing to one, also called probability triple and denoted $(\Omega, \mathcal{A}, \mathbb{P})$. This mathematical concept first includes a sample space Ω , which contains a set of outcomes $\omega \in \Omega$. Note that an *event* is defined as a set of outcomes in the sample space. Then, a σ -algebra \mathcal{A} , also called event space, is a set of events. Finally, a probability function $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, is a positive probability measure associated with an event. Most often, the choice of the probability space will not be specified. The main object will be functions defined over this probability space: random variables.

The *random vector* \mathbf{X} (i.e., multivariate random variable) is a measurable function defined as:

$$\mathbf{X} : \begin{cases} \Omega & \longrightarrow \mathcal{D}_{\mathbf{x}} \subseteq \mathbb{R}^d \\ \omega & \longmapsto \mathbf{X}(\omega) = \mathbf{x}. \end{cases} \quad (1.2)$$

In the following, the random vector \mathbf{X} will be considered to be a squared-integrable function against the measure \mathbb{P} (i.e., $\int_{\Omega} |\mathbf{X}(\omega)|^2 d\mathbb{P}(\omega) < \infty$). Moreover, the present thesis deals with continuous random variables.

The *probability distribution* of the random vector \mathbf{X} is the pushforward measure of \mathbb{P} by \mathbf{X} . Which is a probability measure on $(\mathcal{D}_{\mathbf{x}}, \mathcal{A})$, denoted $\mathbb{P}_{\mathbf{X}}$ and defined by:

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \mathbb{P}(\omega \in \Omega : \mathbf{X}(\omega) \in B), \quad \forall B \in \mathcal{A}. \quad (1.3)$$

The *cumulative distribution function* (CDF) is a common tool to manipulate random variables. It is a function $F_{\mathbf{X}} : \mathcal{D}_{\mathbf{x}} \rightarrow [0, 1]$ defined for all $\mathbf{x} \in \mathcal{D}_{\mathbf{x}}$ as:

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbb{P}_{\mathbf{X}}([-\infty, x_1] \times \dots \times [-\infty, x_d]). \quad (1.4)$$

The CDF is a positive, increasing, right-continuous function, which tends to 0 as \mathbf{x} tends to $-\infty$ and to 1 as \mathbf{x} tends to $+\infty$. In the continuous case, one can also define a corresponding *probability density function* (PDF) $f_{\mathbf{X}} : \mathcal{D}_{\mathbf{x}} \rightarrow \mathbb{R}_+$ with $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^d F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_d}$.

The expected value of a random vector $\mathbb{E}[\mathbf{X}]$, also called the first moment, is a vector defined as:

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \int_{\Omega} \mathbf{X}(\omega) d\mathbb{P}(\omega) = \int_{\mathcal{D}_{\mathbf{x}}} \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^\top. \quad (1.5)$$

In addition, considering two random variables X_i and X_j , with $i, j \in \{1, \dots, d\}$, one can write their respective variance:

$$\text{Var}(X_i) = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2], \quad (1.6)$$

and a covariance describing their joint variability:

$$\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]. \quad (1.7)$$

The *standard deviation* $\sigma_{X_j} = \sqrt{\text{Var}(X_j)}$ and *coefficient of variation* $\delta_{X_j} = \frac{\mathbb{E}[X_j]}{\sqrt{\text{Var}(X_j)}}$ are two quantities directly associated to the two first moments.

1.3.3 Joint input probability distribution

This section introduces various techniques to model and infer a joint probability distribution (or multivariate distribution). It will first define the *copula*, a mathematical tool used to model the dependence structure of a joint distribution. Then, a few methods to fit a joint distribution over a dataset will be mentioned. Finally, a panel of tools to evaluate the goodness of fit between a probabilistic model and a dataset will be recalled.

In general, the single effects of multivariate distributions tend to be well modeled. However, modeling the dependence structure underlying a joint distribution is often overlooked. To illustrate the importance of this step, Fig. 1.1 represents three i.i.d samples from three bivariate distributions sharing the same single effects (e.g., here two exponential distributions) but different dependence structures. Judging from this example, one can assume that the joint distribution results from the composition of the single effects, also called marginals, and an application governing the dependence between them.

An empirical way of isolating the dependence structures from this example is to transform the samples in the ranked space. Let us consider an n -sized sample $\mathbf{X}_n = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \in \mathcal{D}_{\mathbf{x}}^n$. The corresponding ranked sample is defined as: $\mathbf{R}_n = \{\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(n)}\}$, where³ $r_j^{(i)} = \sum_{l=1}^n \mathbb{1}_{\{x_j^{(l)} \leq x_j^{(i)}\}}$, $\forall j \in \{1, \dots, d\}, i \in \{1, \dots, n\}$. Ranking a multivariate dataset allows us to isolate the dependence

³The *indicator function* is defined such that $\mathbb{1}_{\{\mathcal{A}\}}(x) = 1$ if $x \in \mathcal{A}$ and is equal to zero otherwise.

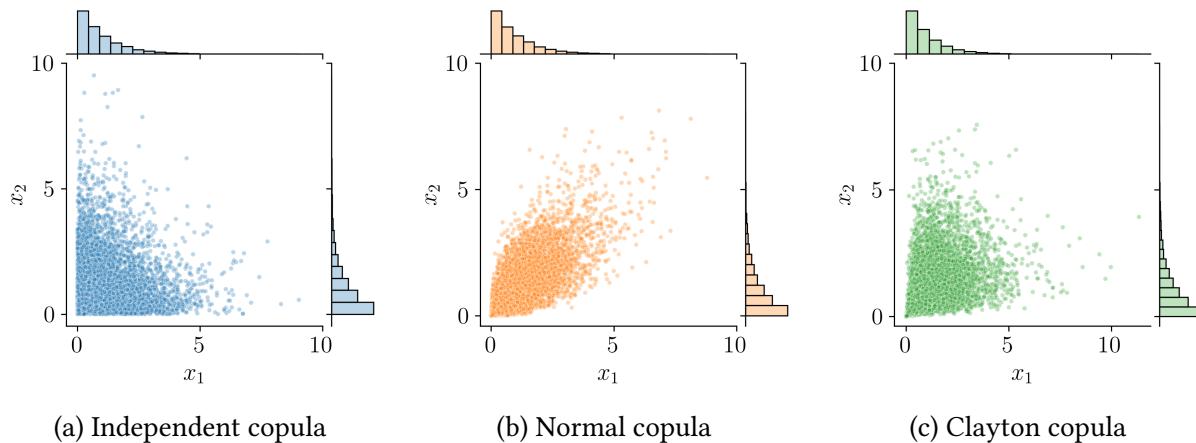


Figure 1.1 Samples of three joint distributions with identical marginals and different dependence structures

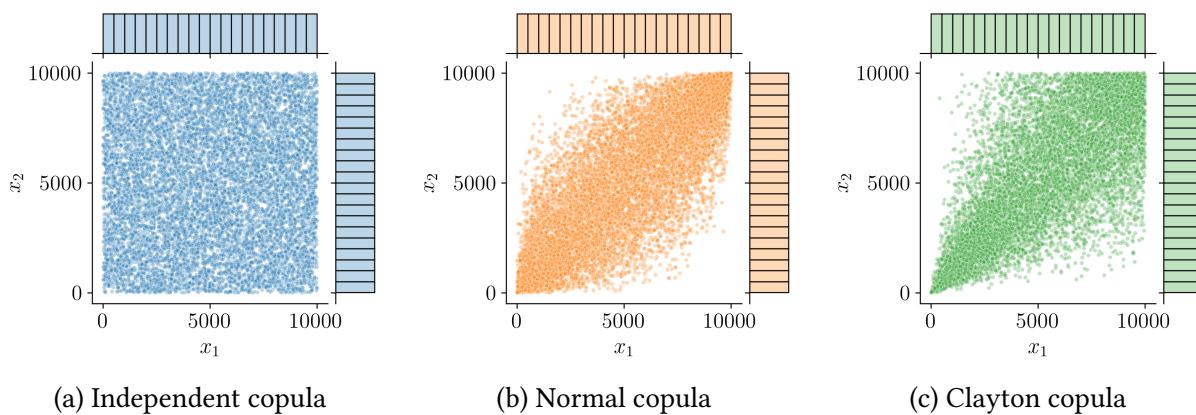


Figure 1.2 Ranked samples represented in the Fig. 1.1

structure witnessed empirically (Saporta, 2006). Fig. 1.2 shows the same three samples from Fig. 1.1 in the ranked space. One can first notice that the marginals are uniform since each rank is uniformly distributed. Then, the scatter plot from the distribution with independent copula (left plot) is uniform while the two others present different patterns.

A theorem states that the multivariate distribution of any random vector can be broken down into two objects (Joe, 1997). First, a set of univariate marginal distributions describing the behavior of the individual variables; Second, a function describing the dependence structure between all variables: a copula.

Theorem 1 (Sklar's theorem). *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector and its joint CDF $F_{\mathbf{X}}$ with marginals $\{F_{X_j}\}_{j=1}^d$, there exists a copula $C : [0, 1]^d \rightarrow [0, 1]$, such that:*

$$F_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = C(F_{X_1}(x_1), \dots, F_{X_d}(x_d)). \quad (1.8)$$

If the marginals F_{X_i} are continuous, then this copula is unique. If the multivariate distribution has a PDF $f_{\mathbf{X}}$, it can also be expressed:

$$f_{\mathbf{X}}(x_1, \dots, x_d) = c(F_{X_1}(x_1), \dots, F_{X_d}(x_d)) \times f_{X_1}(x_1) \times \dots \times f_{X_d}(x_d), \quad (1.9)$$

where c is the density of the copula, sometimes also called copula by misuse of language. The reader might refer to [Durante and Sempi \(2015\)](#) for three different mathematical proofs.

Theorem 1 expresses the joint CDF by combining marginal CDFs and a copula, which is practical for sampling joint distributions. Conversely, the copula can be defined by using the joint CDF and the marginal CDFs:

$$C(u_1, \dots, u_d) = F_{\mathbf{X}}(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)). \quad (1.10)$$

This equation allows us to extract a copula from a joint distribution by knowing its marginals. Additionally, copulas are invariant under increasing transformations. This property is important to understand the use of rank transformation to display the copula without the marginal effects.

Identically to the univariate continuous distributions, a large catalog of families of copulas exists (e.g., independent, Normal, Clayton, Frank, Gumbel copula, etc.). Note that the independent copula Π implies that the distribution is defined as the product of its marginals $\Pi = \prod_{j=1}^d u_j$. In an inference context, this theorem divides the fitting problem into two independent problems: fitting the marginals and fitting the copula. Provided a dataset, this framework allows the potential combination of a parametric (or nonparametric) fit of marginals with a parametric (or nonparametric) fit of the copula.

To infer a joint distribution over a dataset, the analyst should determine a fitting strategy. Appropriate data visualization helps to choose the fitting methods susceptible to be relevant to the problem. In practice, the following points can be asked at this early stage:

- Is the distribution unimodal? If not, mixture methods or nonparametric models might be required;
- Is the validity domain restrictive? If so, specific families of parametric distributions can be chosen or truncation can be applied;
- Is there a dependence structure? Does it concern all the variables together or only some groups of variables?
- Is the dependence structure complex? Transforming the dataset in the ranked space gives an empirical description of the dependence.

To ease the reading, a few techniques for estimating marginal distributions are available in Appendix A. In addition, two nonparametric methods are introduced in Chapter ?? to infer a copula: the “empirical Bernstein copula” and the “Beta copula”. The adequation between a fitted probabilistic model and a dataset should be validated, therefore, Appendices A recall visual and quantitative tools for univariate goodness-of-fit evaluation.

OpenTURNS 1 (Bivariate distribution). The Python code available in Appendix D proposes a minimalistic OpenTURNS example of a probabilistic uncertainty modeling. Figures illustrating the present section may be reproduced, using the OpenTURNS scripts available on GitHub⁴.

1.4 Uncertainty propagation for central tendency study

The previous section aimed at building a probabilistic model of the uncertainties considering the knowledge available. This one introduces diverse methods for forward propagation of the input uncertainties through a numerical model. In the present section, uncertainty propagation is dedicated to the “central tendency” as its goal is to study the mean and variance of the output distribution. This approach contrasts with the uncertainty propagation committed to rare event probability estimation, which will be introduced in Section 1.5 (e.g., used to assess reliability).

The difficulties related to any uncertainty propagation mostly arise from the practical properties of the numerical model. Its potential high dimension, irregularity and nonlinearity each represent a challenge. Such studies rely on a finite number of observations of the numerical model, depending on the computational budget affordable. Uncertainty propagation is at the end of the generic UQ approach (step C), however, it is affected by the “garbage in, garbage out” concept. Meaning that its conclusions depend on the accuracy of the inputs’ uncertainty modeling.

This section introduces the main methods of global uncertainty propagation, outlining the links between numerical integration (i.e., Lebesgue integration or central tendency estimation) and the numerical design of experiments.

1.4.1 Numerical integration

Forward uncertainty propagation aims at integrating a measurable function $g : \mathcal{D}_X \rightarrow \mathbb{R}$ with respect to a probability measure \mathbb{P}_X . Numerical integration provides algorithmic tools to help the resolution of this probabilistic integration (i.e., Lebesgue integration). Note that the measurable function g , in the context of computer experiments, becomes the numerical model \mathcal{M} introduced in Eq. (1.1).

In practice, this integral is approximated by summing a finite n -sized set of realizations $\mathbf{y}_n = \{g(\mathbf{x}^{(1)}), \dots, g(\mathbf{x}^{(n)})\}$ from a set of input samples $\mathbf{X}_n = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$. A *quadrature* establishes

⁴https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/copulas.ipynb

a rule selecting the input samples \mathbf{X}_n (also called nodes), and an associated set of weights $\mathbf{w}_n = \{w_1, \dots, w_n\} \in \mathbb{R}^n$. The approximation given by a quadrature rule is defined as a weighted arithmetic mean of the realizations:

$$I_{\mathbb{P}_X}(g) := \int_{\mathcal{D}_X} g(\mathbf{x}) d\mathbb{P}_X(\mathbf{x}) \approx \sum_{i=1}^n w_i g(\mathbf{x}^{(i)}). \quad (1.11)$$

For a given sample size n , the goal is to find a set of tuples $\{\mathbf{x}^{(i)}, w_i\}_{i=1}^n$ (i.e., quadrature rule), giving the best approximation of our quantity. Ideally, the approximation quality should be fulfilled for a wide class of integrands. Most quadrature rules only depend on the measure space $(\Omega, \mathcal{A}, \mathbb{P}_X)$, regardless of the integrand values. In the context of a costly numerical model, this property allows the analyst to massively distribute the calls to the numerical model.

This section aims to present the main multivariate integration techniques. These methods encompass different properties: some are deterministic and some are aleatory; some are sequential (i.e., nested) some are not; some are victims of the curse of dimensionality and some are not.

Classical multivariate deterministic quadrature

Historically, quadrature methods have been developed for univariate integrals. The Gaussian rule and the Fejér-Clebschaw-Curtis rule are two univariate deterministic quadratures that will be briefly introduced (see [Sullivan 2015](#) for further elements).

Gaussian quadrature is a powerful univariate quadrature building together a set of irregular nodes and a set of weights. The computed weights are positive, which ensures a numerically stable rule even for large sample sizes.

Different variants of Gaussian rules exist, the most common being the Gauss-Legendre quadrature. In this case, the function g to be integrated with respect to the uniform measure on $[-1, 1]$ is approximated by Legendre polynomials. Considering the Legendre polynomial of order n , denoted l_n , the quadrature nodes $x^{(i)}_{i=1}^n$ are given by the polynomial roots. The respective weights are given by the following formula:

$$w_i = \frac{2}{\left(1 - (x^{(i)})^2\right) (l'_n(x^{(i)}))^2}. \quad (1.12)$$

Gauss-Legendre quadrature guarantees a very precise approximation provided that the integrand is well-approximated by a polynomial of degree $2n - 1$ or less on $[-1, 1]$. This rule is deterministic but not sequential, meaning that two rules with sizes n_1 and n_2 , $n_1 < n_2$ will not be nested. However, a sequential extension is proposed by the Gauss-Kronrod rule ([Laurie, 1997](#)), at the expense of a slightly lower accuracy.

To overcome this practical drawback, Fejér then Clebschaw with Curtis proposed a nested rule with mostly equivalent accuracy as Gaussian quadratures. This method is usually presented to integrate a function with respect to the uniform measure on $[-1, 1]$ and starts with a change

of variables:

$$\int_{-1}^1 g(x) dx = \int_0^\pi g(\cos(\theta)) \sin(\theta) d\theta. \quad (1.13)$$

This expression can be written as an expansion of the integrand using cosine series. Therefore, knowing that cosine series are closely related to the Chebyshev polynomials of the first kind. Fejér's "first rule" (Trefethen, 2008) uses the Chebyshev polynomials roots as nodes $x^{(i)} = \cos(\theta^{(i+1/2)})$, associated with the following weights:

$$w_i = \frac{2}{n} \left(1 - 2 \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{1}{4j^2 - 1} \cos(j\theta^{(2i+1)}) \right). \quad (1.14)$$

These two univariate integration schemes are both very efficient on a wide panel of functions. Yet, Fejér-Clebschaw-Curtis is sequential and offers easy implementations, benefitting from powerful algorithms such as the *fast Fourier transform*. Fig. 1.3 illustrates the nested properties of Fejér-Clebschaw-Curtis quadrature by representing the nodes of quadrature rules with increasing size.

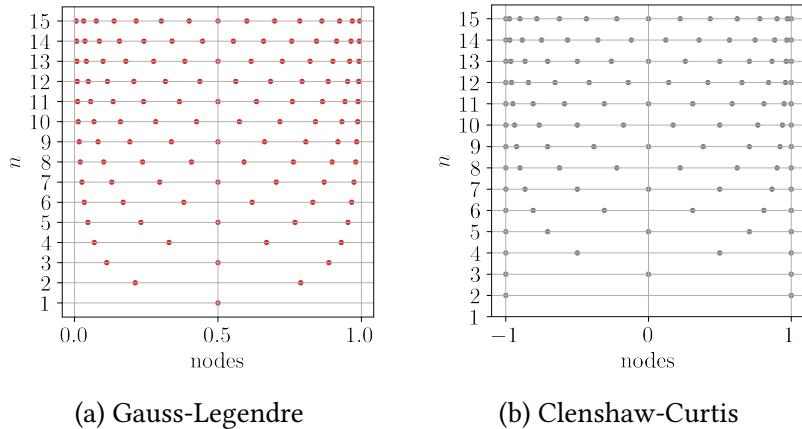


Figure 1.3 Univariate quadrature nodes for increasing sizes ($1 \leq n \leq 15$)

Uncertainty quantification problems are rarely unidimensional, but one can directly build a multivariate quadrature rule by defining the tensor product (also called full grids) of univariate rules. This exhaustive approach quickly shows its practical limits as the problem's dimension increases. In Fig. 1.4, the left plot represents a two-dimensional tensor product of identical Gauss-Legendre quadratures. Alternatively, sparse multivariate quadratures (i.e., Smolyak sparse grid) explore the joint domain more efficiently. Using the Smolyak recurrent formula (see e.g., Sullivan 2015), two univariate quadratures can be combined as illustrated on the right of Fig. 1.4.

Monte Carlo methods

Monte Carlo methods were initially developed in the 1940s to solve problems in neutronics. Ever since, this frequentist technique has been applied to the resolution of the Lebesgue integral.

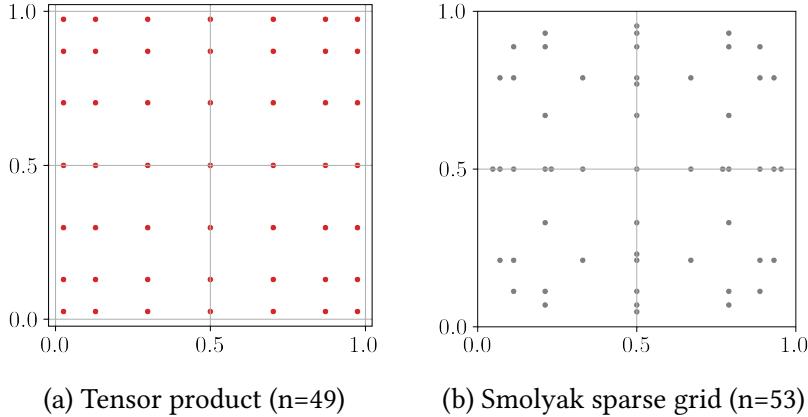


Figure 1.4 Two identical univariate Gauss-Legendre quadratures combined as a tensor product (left) and a Smolyak sparse grid (right).

To integrate a function g against a measure \mathbb{P}_X , it randomly generates points following the input measure. The integral is estimated by taking the uniform arithmetic mean of the nodes' images obtained by this random process.

This aleatory method requires to be able to generate points following a given distribution. To do so, the most common approach is to first uniformly generate a sequence of random points on $[0, 1]$. These sequences mimic actual randomness but are in fact generated by deterministic algorithms, also called pseudorandom number generators. Pseudorandom algorithms generate a sequence of numbers with a very large, but finite length. This sequence can be exactly repeated by fixing the same initial point, also called *pseudorandom seed*. Most programming languages use the Mersenne-Twister pseudorandom generator (Matsumoto and Nishimura, 1998), offering a very long period (around 4.3×10^{6001} iterations).

Formally, the “Vanilla” Monte Carlo (sometimes called “crude” Monte Carlo) method uses a set of i.i.d samples $\mathbf{X}_n = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ following the joint distribution of \mathbb{P}_X . The Monte Carlo estimator of the integral is given by:

$$I_{\mathbb{P}_X}(g) \approx \bar{y}_n^{\text{MC}} = \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}^{(i)}). \quad (1.15)$$

By construction, the law of large numbers makes this estimator unbiased, however, it converges relatively slowly. Considering the images of the sample \mathbf{X}_n , one can also estimate the variance of the output random variable $\hat{\sigma}_Y^2$. The variance of the Monte Carlo estimator results from a manipulation of the central limit theorem:

$$\text{Var}(\bar{y}_n^{\text{MC}}) = \frac{1}{n} \text{Var}(g(\mathbf{X})). \quad (1.16)$$

This estimator also comes with theoretical confidence intervals at $\alpha\%$, regardless of the output distribution:

$$I_{\mathbb{P}_X}(g) \in \left[\bar{y}_n^{\text{MC}} - q_\alpha \sqrt{\frac{\text{Var}(g(\mathbf{X}))}{n}}, \bar{y}_n^{\text{MC}} + q_\alpha \sqrt{\frac{\text{Var}(g(\mathbf{X}))}{n}} \right], \quad (1.17)$$

where q_α is the α -quantile of the standard normal distribution. Monte Carlo presents the advantage of being a universal method, with no bias and strong convergence guarantees. Moreover, it is worth noting that its convergence properties do not depend on the dimension of the input domain. Unlike the previous multivariate deterministic quadrature, it doesn't suffer from the curse of dimensionality. The main limit of crude Monte Carlo is its convergence speed, making it intractable for most practical cases. More recent methods aim at keeping the interesting properties of this technique while making it more efficient. Among the *variance reduction* methods, let us mention importance sampling, stratified sampling (e.g., Latin hypercube sampling), control variates and multi-level Monte Carlo. For further details, the reader may refer to Chapters 8, 9 and 10 from [Owen \(2013\)](#) and [\(Giles, 2008\)](#).

Quasi-Monte Carlo and Koksma-Hlawka inequality

Among the methods presented so far, classical deterministic quadratures are subject to the curse of dimensionality while Monte Carlo methods deliver contrasted performances. Quasi-Monte Carlo is a deterministic family of numerical integration schemes with respect to the uniform measure on $[0, 1]$. It offers powerful performances with strong guarantees by choosing nodes according to *low discrepancy* sequences.

The discrepancy of a set of nodes (or a design) can be seen as a metric of its uniformity. The lowest the discrepancy of a design is, the “closest” it is to uniformity.

The Koksma-Hlawka theorem ([Morokoff and Caflisch, 1995](#); [Leobacher and Pillichshammer, 2014](#)) is a fundamental result for understanding the role of the discrepancy in numerical integration.

Theorem 2 (Koksma-Hlawka). *If $g : [0, 1]^d \rightarrow \mathbb{R}$ has a bounded variation (i.e., its total variation is finite), then for any design $\mathbf{X}_n = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \in [0, 1]^d$:*

$$\left| \int_{[0,1]^d} g(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}^{(i)}) \right| \leq V(g) D^*(\mathbf{X}_n). \quad (1.18)$$

Where $D^*(\mathbf{X}_n)$ is the star discrepancy of the design \mathbf{X}_n , while $V(g)$ quantifies the complexity of the integrand, which is related to its total variation. The reader might refer to [Leobacher and Pillichshammer \(2014\)](#) Section 3.4 for further mathematical proof.

The function's variation $V(g)$ in Eq. (??) can be formally defined as the Hardy-Klause variation:

$$V(g) = \sum_{u \subseteq \{1, \dots, p\}} \int_{[0,1]^u} \left| \frac{\partial^u g}{\partial \mathbf{x}^u}(\mathbf{x}_u, 1) \right| d\mathbf{x}_u. \quad (1.19)$$

The star discrepancy $D^*(\mathbf{X}_n)$ can be defined from a geometric point of view. Let us first consider the number of elements from a design \mathbf{X}_n , falling in a subdomain $[\mathbf{0}, \mathbf{x}]$ as $\#(\mathbf{X}_n \cap [\mathbf{0}, \mathbf{x}])$. Then, if this empirical quantification is compared with the volume of the rectangle $[\mathbf{0}, \mathbf{x}]$, denoted

by $\text{vol}([\mathbf{0}, \mathbf{x})]$, the star discrepancy is expressed as:

$$D^*(\mathbf{X}_n) = \sup_{\mathbf{x} \in [0,1]^d} \left| \frac{\#(\mathbf{X}_n \cap [\mathbf{0}, \mathbf{x}))}{n} - \text{vol}([\mathbf{0}, \mathbf{x})] \right|. \quad (1.20)$$

The star-discrepancy $D^*(\mathbf{X}_n)$ is actually a particular case of the L_p *star discrepancy* denoted by $D_p^*(\mathbf{X}_n)$, for which $p = \infty$. In the general case, $D_p^*(\mathbf{X}_n)$ is defined as the L_p -norm of the difference between the empirical CDF of the design $\widehat{F}_{\mathbf{X}_n}$ and the CDF of the uniform distribution F_U :

$$D_p^*(\mathbf{X}_n) = \|\widehat{F}_{\mathbf{X}_n} - F_U\|_p = \left(\int_{[0,1]^d} |\widehat{F}_{\mathbf{X}_n}(\mathbf{x}) - F_U(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}. \quad (1.21)$$

Let us point out that this star discrepancy $D^*(\mathbf{X}_n)$ is equivalent to the Kolmogorov-Smirnov statistic, verifying whether the design follows a uniform distribution (Fang et al., 2018).

In Theorem 2, one can notice how the Koksma-Hlawka inequality dissociates the quadrature performance into a contribution from the function complexity and one from the repartition of the quadrature nodes. Knowing that the complexity of the studied integrand is fixed, this property explains the motivation to generate low-discrepancy quadratures in numerical integration.

Note that the design can also be considered as a discrete distribution (uniform sum of Dirac distributions). The discrepancy can then be expressed as a probabilistic distance between this discrete distribution and the uniform distribution. A generalized discrepancy between distributions called *maximum mean discrepancy* is introduced in Appendix D and used for efficient sampling in Chapter ?? of this manuscript.

Some famous low-discrepancy sequences (e.g., van der Corput, Halton, Sobol', Faure, etc.) can offer a bounded star discrepancy $D^*(\mathbf{X}_n) \leq \frac{C \log(n)^d}{n}$, with the constant C depending on the sequence. Therefore, using these sequences as a quadrature rule with uniform weights provides the following absolute error upper bound:

$$\left| \int_{[0,1]^d} g(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n g\left(\mathbf{x}^{(i)}\right) \right| \leq \frac{V(g) \log(n)^d}{n}. \quad (1.22)$$

The generation of such sequences does not necessarily require more effort than pseudo-random sampling. Chapter 15 in Owen (2013) offers an extended presentation of the ways to generate different low-discrepancy sequences. For example, the van der Corput and Halton sequences rely on congruential generators.

Halton sequences in medium dimension, unfortunately, introduce pathological patterns when looking at their subprojections. To overcome these limits, digital nets such as the famous Sobol' or Faure sequences were developed. Sobol' sequences are in base two and have the advantage of being extensible in dimension. Note that by construction, these sequences offer significantly lower discrepancies for specific size values. Typically, designs with sizes equal to powers of two or power of prime numbers will be favorable. To illustrate the different repartition and properties of the methods, Fig. 1.5 represents the three Monte Carlo and quasi-Monte Carlo

designs (with size $n = 256$). Each is split into the first 128 points (in red) and the following 128 points (in black) to show the nested properties of the QMC sequences.

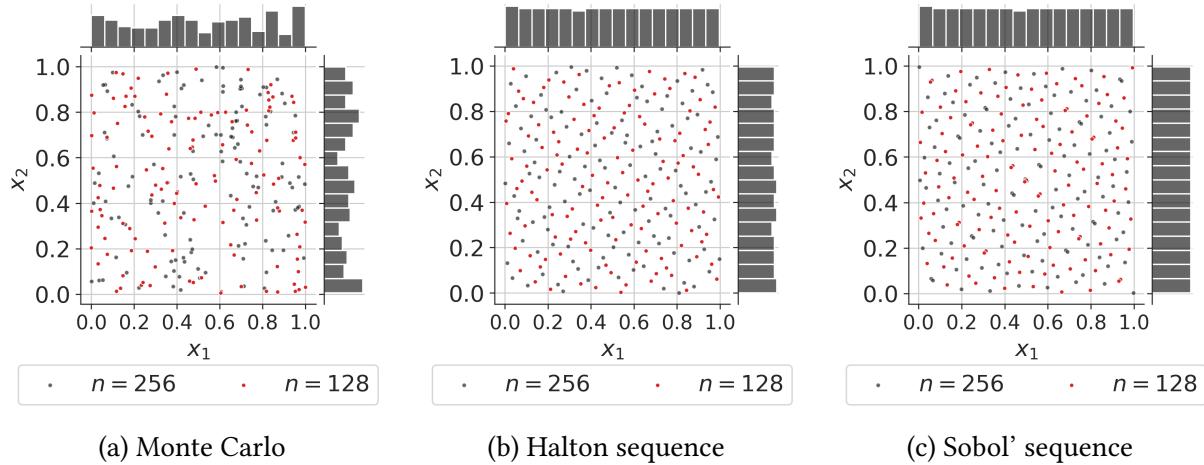


Figure 1.5 Nested Monte Carlo and quasi-Monte Carlo designs ($n = 256$)

Crude Monte Carlo estimators provide guarantees associated with the estimate. This complementary information is essential to deliver an end-to-end uncertainty quantification, which is missed in deterministic QMC methods. *Randomized quasi-Monte Carlo* (RQMC) is a method introducing some randomness in QMC in order to compute confidence intervals while benefiting from a low variance. A specific review of the randomized (also called “scrambled”) QMC is proposed by L’Ecuyer (2018). Various authors recommend the use of RQMC by default instead of QMC as a good practice. Recent works aim at exploring the use of these methods to estimate different quantities of interest, such as an expected value (Gobet et al., 2022) or a quantile (Kaplan et al., 2019).

Ultimately, quasi-Monte Carlo methods generate powerful integration schemes. The Koksma-Hlawka inequality associates an upper bound and a convergence rate to most integrals. A randomization overlay fades the deterministic property of these designs, allowing us to compute confidence intervals. In the following, sampling techniques are presented from the numerical *design of experiments* point of view. Even if the goal might look different from numerical integration, these two topics share many methods and concepts.

OpenTURNS 2 (Numerical integration). The Python code available in Appendix D proposes a minimalistic OpenTURNS example to build multivariate quadrature rules. Figures illustrating the present section may be reproduced, using the OpenTURNS scripts available on GitHub⁵.

1.4.2 Numerical design of experiments

The numerical design of experiments aims at uniformly exploring the input domain, e.g., to build a learning set for a regression model, or to initialize a multi-start optimization strategy. A

⁵https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/integration.ipynb

design of experiment (also simply called design) is qualified as *space-filling* when it uniformly covers a domain. As well as in integration, a design is used to propagate uncertainties through a numerical model (or a physical experiment). However, a difference comes from the fact that this community often works with designs of very limited size. Users of designs of experiments also consider various properties.

- Some might be interested in the sequentiality of a sampling method, to eventually add new points as they get their computational budget extended.
- Some might request a sampling method conserving its properties in any subdomains (i.e., subprojections of the inputs' domain). This second property can be useful to reduce the problem's dimension by dropping a few unimportant variables (see the following Section 1.6 on global sensitivity analysis).

Different metrics are commonly used to quantify how space-filling a design of experiments is. The previously introduced discrepancies are an example of space-filling metrics. Other types of space-filling metrics rely on purely geometrical considerations.

This section will first define a few space-filling metrics. Secondly, the *Latin hypercube sampling* (LHS) will be introduced as a variance-reduction method that became popular in the UQ community. Finally, a general discussion on uncertainty propagation with respect to non-uniform measures will be presented.

Space-filling metrics and properties

Space-filling criteria are key to evaluating designs and are often used to optimize their performances. In the previous section, the star discrepancy was introduced as a distance of a finite design to uniformity. However, the L_∞ star discrepancy is hard to estimate, fortunately, Warnock (1972) elaborated an explicit expression specific to the L_2 star discrepancy:

$$[D_2^*(\mathbf{X}_n)]^2 = \frac{1}{9} - \frac{2}{n} \sum_{i=1}^n \prod_{l=1}^d \frac{(1-x_l^{(i)})}{2} + \frac{1}{n^2} \sum_{i,j=1}^n \prod_{l=1}^d \left[1 - \max(x_l^{(i)}, x_l^{(j)}) \right]. \quad (1.23)$$

One can notice that this expression is similar to the Cramér-von Mises test statistic. Even if this expression is tractable, Fang et al. (2018) detailed its limits: the star L_2 discrepancy generates designs that are not robust to projections in sub-spaces; it is not an invariant metric by rotation and reflection; and finally, by construction, L_p discrepancies give a disproportionate role to the point $\mathbf{0}$ by anchoring the box $[\mathbf{0}, \mathbf{x}]$.

Two improved criteria were proposed by Hickernell (1998) with the *centered L_2 discrepancy* and the *wrap-around L_2 discrepancy*. Those are widely used in practice since they solve the previous limits while satisfying the Koksma-Hlawka inequality with a modification of the total

variation. Let us introduce the explicit formula of the centered L_2 discrepancy:

$$\begin{aligned} CD_2^*(\mathbf{X}_n) = & \left(\frac{13}{12} \right)^d - \frac{2}{n} \sum_{i=1}^n \prod_{l=1}^d \left(1 + \frac{1}{2} |x_l^{(i)} - 0.5| - \frac{1}{2} |x_l^{(i)} - 0.5|^2 \right) \\ & + \frac{1}{n^2} \sum_{i,j=1}^n \prod_{l=1}^d \left(1 + \frac{1}{2} |x_l^{(i)} - 0.5| + \frac{1}{2} |x_l^{(j)} - 0.5| - \frac{1}{2} |x_l^{(i)} - x_l^{(j)}| \right). \end{aligned} \quad (1.24)$$

As an alternative to discrepancies, many geometrical criteria exist to assess a space-filling design. The most common way to do so is to maximize the minimal distance among the pairs of Euclidian distances between the points of a design. The criterion to maximize is then simply called the *minimal distance* of a design (denoted ϕ_{min}). For numerical reasons, the ϕ_p criterion is often used instead of the minimal distance. The following ϕ_p criterion converges towards the minimum distance as $p \geq 1$ tends to infinity:

$$\phi_{min}(\mathbf{X}_n) = \min_{i \neq j} \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2, \quad \phi_p(\mathbf{X}_n) = \sum_{i=1}^j \sum_{j=1}^n \left(|x^{(i)} - x^{(j)}|^{-p} \right)^{\frac{1}{p}}. \quad (1.25)$$

More space-filling criteria are reviewed in [Abtini \(2018\)](#) and in Appendix A from [Da Veiga et al. \(2021\)](#). Further relations between some mathematical objects related to space-filling are developed in [Pronzato and Müller \(2012\)](#). These space-filling metrics are widely used to optimize different sampling techniques.

Latin hypercube sampling

Latin hypercube sampling is a method initially introduced in the 70s for numerical integration ([Mckay et al., 1979](#)). In a bounded domain, this stratified sampling technique forces the distribution of each sub-projection to be as uniform as possible. To do so, for an n -sized design, each marginal domain is divided into n identical segments. This creates a regular grid of n^d squared cells over the domain.

Then, a Latin hypercube design (LHD) does not allow more than one point within a segment. That way, new LHDs can be built as a permutation of the marginals of an existing LHD. Inside each selected cell from the grid, the point can be placed at the center or randomly.

Various contributions proposed a variance, and a central limit theorem to LHS ([Owen, 1992](#)). Under monotony hypothesis, the LHS variance can be expressed, when $n \rightarrow \infty$, as:

$$\text{Var} \left(\bar{y}_n^{\text{LHS}} \right) = \frac{1}{n} \text{Var}(g(\mathbf{X})) - \frac{C}{n} + o \left(\frac{1}{n} \right). \quad (1.26)$$

Where C is a positive constant, showing that the LHS usually reduces the variance for numerical integration (proof proposed in [Stein, 1987](#)). Because of its stratified structure, LHS can generate poor designs from a space-filling point of view (see e.g., the illustration in Fig. 1.6a). The following section presents various methods aiming at optimizing LHDs.

Optimized Latin hypercube sampling

To improve the space-filling property of LHD, it is common to add an optimization step. The goal of this optimization is to improve a space-filling criterion by generating LHD from permutations of an initial LHD. [Damblin et al. \(2013\)](#) reviews LHS optimization using different discrepancy criteria and subprojection properties. This optimization can be performed by different algorithms, such as the stochastic evolutionary algorithm or simulated annealing. The results from this work show that LHD optimized by L_2 centered or wrap-around discrepancies offer strong robustness to two-dimensional projections. It also shows that these designs keep this property for dimensions larger than 10, while scrambled Sobol' sequences lose it. Fig. 1.6 illustrates two LHD, optimized by the L_2 centered discrepancy and the geometrical ϕ_p . The space-filling difference is not obvious in two-dimensional problems, and they both spread uniformly.

More recent work developed different ways to optimize LHD. Among them, let us mention the maximum projection designs from [Joseph et al. \(2015\)](#) which rely on the optimization of a geometrical criterion and deliver interesting performances. In the same vein, the uniform projection designs from [Sun et al. \(2019\)](#) is also a method to optimize LHS, this time based on a criterion averaging discrepancies between each pair of marginals.

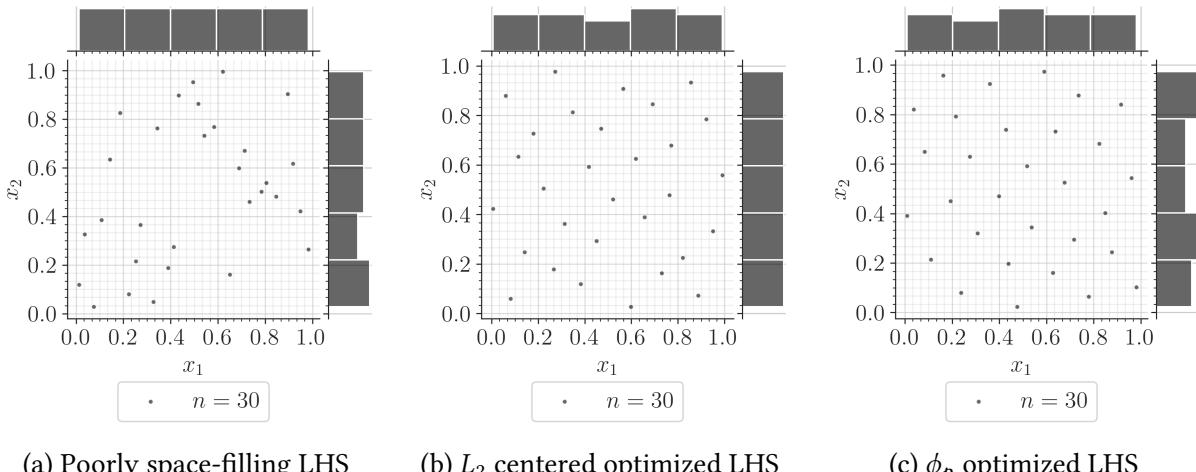


Figure 1.6 Latin hypercube designs with poor and optimized space-filling properties ($n = 8$)

OpenTURNS 3 (Design of experiments). The Python code available in Appendix D proposes a minimalistic OpenTURNS example to build an LHS and an LHS optimized w.r.t. to a space-filling metric (here the L2-centered discrepancy) using the simulated annealing algorithm. Figures illustrating the present section may be reproduced, using the OpenTURNS scripts available on GitHub⁶.

⁶https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/designofexperiments.ipynb

1.4.3 Summary and discussion

A wide panel of sampling techniques exists for numerical integration or design of experiments purposes. In both cases, the studied domain is bounded and the targeted measure is uniform. However, uncertainty propagation is often performed on complex input distributions, with possibly unbounded domains. In uncertainty quantification, this step might be referred to as the estimation of the output random variable's central tendency (i.e., its mean and variance). Central tendency estimation is a numerical integration with respect to any input distribution, sometimes called *probabilistic integration* (Briol et al., 2019) as part of *probabilistic numerics* (Oates and Sullivan, 2019).

To generate i.i.d samples following any distribution (i.e., non-uniform), one may use *inverse transform* sampling. After generating samples in the unit hypercube, the inverse CDF function (i.e., quantile function) is applied to the marginals. Finally, possible dependence effects may be added using the Sklar theorem Eq. (1).

One may wonder if the properties from the uniform design are conserved after this nonlinear transformation. Li et al. (2020) explores this question from a discrepancy point of view. The authors find correspondences between discrepancies with respect to uniformity and discrepancies with respect to the target distribution. However, this result shows practical limits, sometimes making the interpretation of the last discrepancy easier. This question will be further discussed in Chapter ??, using a more general framework.

Let us also remark that, depending on the distribution, defining the inverse CDF is not always possible. For example, samples following truncated distributions or mixture distributions might sometimes be generated with a different technique. The *acceptance-rejection* method offers a versatile generation only based on the PDF f_x . Assuming that a well-known proposal PDF f_x^* exists such that $f_x \leq c \times f_x^*$, $c \in [1, +\infty]$. One may generate a sample according to $c \times f_x^*$ and only retain from this sample the points under the PDF f_x . Note that some sampling methods, such as QMC, are not well suited to acceptance-rejection since their structure gets perturbed.

In this section, many methods were presented to propagate input uncertainties against a deterministic function. The propagation with the three following goals and contexts was introduced:

- building a quadrature rule for numerical integration against a uniform distribution,
- creating a space-filling design of experiments to uniformly explore the space, often in a small data context (e.g., to build the learning set of a surrogate model),
- generating a design for central-tendency estimation, which is simply a numerical integration against a nonuniform density.

These three objectives have been explored in different communities, but they actually share similar methods. They all have in common the general analysis (i.e., global behavior) of the output random variable. However, some studies require to shift the focus toward specific areas of the output random variables. When using uncertainty propagation to perform risk

analysis, the events studied are often contained in the tails of the output distribution. In this case, dedicated uncertainty propagation methods will significantly improve the estimation of the associated statistical quantities.

1.5 Uncertainty propagation for rare event estimation

This section aims to present another type of uncertainty propagation. In the context of a risk analysis applied to the engineering field, the reliability of a system needs to be assessed. Most often, a risk measure associated with a failure mode of the studied system is estimated.

Since most systems studied in risk analysis must be highly reliable, the occurrence of such an event is qualified as rare. Only a small amount of extreme input conditions or an unlikely unfavorable combination of inputs leads to the failure of the system. Hence, the usage of the equivalent terms *reliability analysis* and *rare event estimation*. The notion of risk associated with an event is often decomposed as a product of its likelihood and its consequences. The failure of a system might be very rare, but its consequences can be severe (e.g., civil engineering structures, nuclear infrastructure, telecommunication networks, electrical grid, railway signaling, etc.).

Different risk measures (i.e., quantities of interest related to the tail of the distributions) can be studied depending on the type of risk analysis. Quantiles are a first conservative measure, widely used for risk analysis. The α -quantile q_α of the output random variable Y is defined as:

$$q_\alpha = \inf_{y \in \mathbb{R}} \{F_Y(y) \geq \alpha\}, \quad \alpha \in [0, 1]. \quad (1.27)$$

As an alternative, one can define a scalar safety threshold y_{th} that should not be exceeded to keep the system safe. Then, a second risk measure is the probability of exceeding this safety threshold, also called *failure probability*:

$$p_f = \mathbb{P}(Y \geq y_{\text{th}}), \quad y_{\text{th}} \in \mathbb{R}. \quad (1.28)$$

To illustrate this quantity, Fig. 1.7 shows the one-dimensional propagation of a normal distribution (represented by the PDF on the left), through a function $g(\cdot)$. The probability of exceeding a given threshold y_{th} is represented by the area in red under the output PDF on top. An interesting reflection on the use and the interpretation of risk measures including measures from the finance domain such as the *conditional value-at-risk* (also called superquantile) is presented in [Rockafellar and Royset \(2015\)](#).

In the following section, the formalism for reliability analysis problems will be first presented, then the main methods for solving this specific problem will be introduced. Note however that the present work will not address the problems of time-dependent reliability analysis tackled in [Hawchar et al. \(2017\)](#).

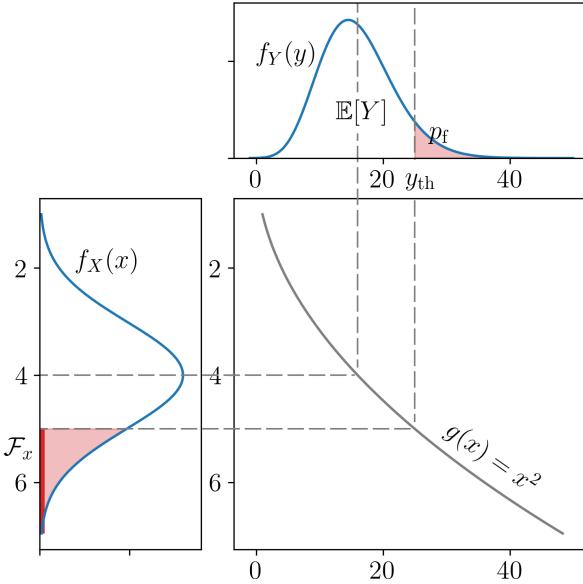


Figure 1.7 One-dimensional reliability analysis example

1.5.1 Problem statement

Following the UQ methodology, the behavior of the system is modeled by $\mathcal{M}(\cdot)$. Considering the problem of exceeding a safety threshold in Eq. (1.28), the system's performance is commonly defined as the difference between the model's output and the safety threshold $y_{\text{th}} \in \mathbb{R}$. Formally, the *limit-state function* (LSF) is a deterministic function $g : \mathbb{R} \rightarrow \mathbb{R}$ quantifying this performance:

$$g(\mathbf{x}) = y_{\text{th}} - \mathcal{M}(\mathbf{x}). \quad (1.29)$$

Depending on the sign of its images, this function splits the input space into two disjoint and complementary domains called the *failure domain* \mathcal{F}_x , and the *safe domain* \mathcal{S}_x which are defined as:

$$\mathcal{F}_x = \{\mathbf{x} \in \mathcal{D}_x \mid g(\mathbf{x}) \leq y_{\text{th}}\}, \quad \mathcal{S}_x := \{\mathbf{x} \in \mathcal{D}_x \mid g(\mathbf{x}) > y_{\text{th}}\}. \quad (1.30)$$

The border between these two domains is a hypersurface called *limit-state surface*, defined by $\mathcal{F}_x^0 := \{\mathbf{x} \in \mathcal{D}_x \mid g(\mathbf{x}) = 0\}$. Similarly to any UQ study using a numerical model, this problem may require to be resolved using a limited number of calls to a black-box simulator. The difficulties in a reliability problem come from the properties of the LSF: nonlinear, costly to evaluate or with a multimodal failure domain. Additionally, note that the reliability problem can be the composition of multiple reliability problems, often modeled as a system of problems in series and parallel.

A rare event estimation results from a particular uncertainty propagation through the LSF. Considering the output variable of interest $g(\mathbf{X})$, its probability of being negative (i.e., in the failure domain) is a common risk measure. The so-called *failure probability*, denoted by p_f , is the quantity of interest for reliability analysis considered in this work. This quantity is formally

written⁷:

$$p_f = \mathbb{P}(Y \geq y_{\text{th}}) = \mathbb{P}(g(\mathbf{X}) \leq 0) = \int_{\mathcal{F}_X} f_X(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{D}_X} \mathbb{1}_{\mathcal{F}_X}(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x}, \quad (1.31)$$

where the indicator function applied to the failure domain returns $\mathbb{1}_{\{\mathcal{F}_X\}}(x) = 1$ if $x \in \mathcal{F}_X$ and $\mathbb{1}_{\{\mathcal{F}_X\}}(x) = 0$ otherwise. Rare event estimation implies both contour finding (i.e., characterizing the LSF) and an estimation strategy targeting the failure domain (often with a limited number of simulations). Note that failure events are qualified as rare when their failure probability has an order of magnitude between $10^{-2} \leq p_f \leq 10^{-9}$ (see e.g., [Lemaire et al. 2009](#)).

Instead of directly performing a reliability analysis in the physical space (i.e., \mathbf{x} -space), these problems are usually solved in the *standard normal space* (i.e., \mathbf{u} -space). Working in the standard space reduces numerical issues caused by potentially unscaled or asymmetric marginals. Moreover, a larger panel of methods can be applied in the standard space since the random inputs become independent. The bijective mapping between these two spaces is called an “iso-probabilistic transformation”, denoted by $T : \mathcal{D}_X \subseteq \mathbb{R}^d \rightarrow \mathbb{R}^d, \mathbf{x} \mapsto T(\mathbf{X}) = \mathbf{u} = (u_1, \dots, u_d)^\top$. When considering any random vector $\mathbf{X} = (X_1, \dots, X_d)^\top$ and the independent standard Gaussian vector $\mathbf{U} = (U_1, \dots, U_d)^\top$, the following equalities hold:

$$\mathbf{U} = T(\mathbf{X}) \Leftrightarrow \mathbf{X} = T^{-1}(\mathbf{U}). \quad (1.32)$$

A reliability problem can be expressed in the standard normal space. Let us first consider the transformed limit-state function \check{g} defined as:

$$\begin{aligned} \check{g} : & \left| \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ \mathbf{u} & \longmapsto & \check{g}(\mathbf{u}) = (g \circ T^{-1})(\mathbf{u}). \end{array} \right. \end{aligned} \quad (1.33)$$

Since this transformation is a diffeomorphism⁸, one can apply the change of variable $\mathbf{x} = T(\mathbf{u})$ to express the reliability problem from Eq. (1.31) in the standard space:

$$p_f = \mathbb{P}(\check{g}(\mathbf{U}) \leq 0) = \int_{\mathcal{F}_u} \varphi_d(\mathbf{u}) d\mathbf{u} = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_u}(\mathbf{u}) \varphi_d(\mathbf{u}) d\mathbf{u}, \quad (1.34)$$

with the transformed failure domain denoted by $\mathcal{F}_u = \{\mathbf{u} \in \mathbb{R}^d \mid \check{g}(\mathbf{u}) \leq 0\}$, and the d -dimensional standard Gaussian PDF $\varphi_d(\mathbf{u}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|\mathbf{u}\|_2^2}{2}\right)$. The fact that the failure probability is invariant by this transformation allows the analyst to estimate this quantity in both spaces.

Different types of transformations exist, such as the Rosenblatt or the generalized Nataf transformation introduced by [Lebrun \(2013\)](#). In practice, the transformation choice depends on the properties of the input distribution studied. For example in OpenTURNS, depending on the three following cases, different types of transformations are applied:

⁷Note that this probabilistic integration is usually written using the PDF $f_X(\cdot)$, but it could identically be expressed in terms of probability measure by taking $f_X(\mathbf{x}) d\mathbf{x} = d\mathbb{P}_X(\mathbf{x}), \forall \mathbf{x} \in \mathcal{D}_X$.

⁸Considering two manifolds A and B , a transformation $T : A \rightarrow B$ is called a diffeomorphism if it is a differentiable bijection with a differentiable inverse $T^{-1} : B \rightarrow A$.

- for elliptical distributions, a linear Nataf transformation is applied;
- for distributions with an elliptical copula, the generalized Nataf transformation is used;
- otherwise, the Rosenblatt transformation is used.

1.5.2 Rare event estimation methods

The main risk measure chosen for rare event estimation in this work is the previously introduced failure probability. Therefore, let us recall that the goal is to build an efficient estimation (or approximation) of the following d -dimensional integral:

$$p_f = \int_{\mathcal{D}_x} \mathbb{1}_{\mathcal{F}_x}(x) f_x(x) dx. \quad (1.35)$$

In the context of rare event estimation using costly to evaluate numerical models, the simulation budget is often limited to n runs with $p_f \ll \frac{1}{n}$. This explains the need for specific methods offering approximations or simulations targeting the unknown failure domain. Two types of rare event estimation methods are classically presented: first, using approximation approaches, and second, using sampling techniques. This section introduced the commonly used rare event methods, see [Morio and Balesdent \(2015\)](#) for a more exhaustive review.

First and second order reliability methods (FORM/SORM)

The well-known first and second-order reliability methods (FORM and SORM) both rely on a geometric approximation to estimate a failure probability ([Lemaire et al., 2009](#)). They extrapolate a local approximation of the LSF built in the vicinity of a *most-probable-failure-point* (MPFP), also called *design point*.

Working in the standard space, the methods first look for this MPFP, denoted P^* , with coordinates \mathbf{u}^* . To find it, one can solve the following quadratic optimization problem:

$$\mathbf{u}^* = \arg \max_{\mathbf{u} \in \mathbb{R}^d} (\mathbb{1}_{\mathcal{F}_u}(\mathbf{u}) \varphi_d(\mathbf{u})). \quad (1.36)$$

Using the properties of the standard space allows us to rewrite it as:

$$\mathbf{u}^* = \arg \max_{\mathbf{u} \in \mathbb{R}^d} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\mathbf{u}^\top \mathbf{u}}{2}\right) \quad \text{s.t. } \mathbf{u} \in \mathcal{F}_u \quad (1.37)$$

$$= \arg \min_{\mathbf{u} \in \mathbb{R}^d} \mathbf{u}^\top \mathbf{u} \quad \text{s.t. } \check{g}(\mathbf{u}) \leq 0. \quad (1.38)$$

This problem then becomes a quadratic optimization under nonlinear constraint. It is classically solved by gradient descent algorithms (e.g., Abdo-Rackwitz algorithm [Abdo and Rackwitz, 1991](#)) but can also use gradient-free techniques (e.g., Cobyla algorithm [Powell, 1994](#)). This point defines the smallest Euclidian distance between the limit-state surface and the origin of the

standard space. To understand its role in the reliability problem, let us recall that the density of the standard normal presents an exponential decay in its radial and tangential directions. Then, P^* is the point with the biggest contribution to the failure probability (see the illustration in Fig. 1.8).

This distance between the origin and P^* is another risk measure, defined as the *Hasofer-Lind reliability index* (Lemaire et al., 2009), $\beta \in \mathbb{R}$ such that:

$$\beta = \|\mathbf{u}^*\|_2 = \boldsymbol{\alpha}^\top \mathbf{u}^*, \quad \text{s.t. } \boldsymbol{\alpha} = \frac{\nabla_{\mathbf{u}} \check{g}(\mathbf{u})}{\|\nabla_{\mathbf{u}} \check{g}(\mathbf{u})\|_2}. \quad (1.39)$$

The vector $\boldsymbol{\alpha}$ is the unit vector pointing at P^* from the origin point.

Then, FORM aims at approximating the limit-state function $\check{g}(\cdot)$ by its first-order Taylor expansion around the MPFP, denoted $\check{g}_1(\mathbf{u}^*)$:

$$\begin{aligned} \check{g}(\mathbf{u}) &= \check{g}_1(\mathbf{u}^*) + o(\|\mathbf{u} - \mathbf{u}^*\|_2^2) \\ &= \check{g}(\mathbf{u}^*) + \nabla_{\mathbf{u}} \check{g}(\mathbf{u}^*)^\top (\mathbf{u} - \mathbf{u}^*) + o(\|\mathbf{u} - \mathbf{u}^*\|_2^2) \\ &= \|\nabla_{\mathbf{u}} \check{g}(\mathbf{u})\|_2 (\boldsymbol{\alpha}^\top \mathbf{u}^* - \boldsymbol{\alpha}^\top \mathbf{u}) + o(\|\mathbf{u} - \mathbf{u}^*\|_2^2). \end{aligned} \quad (1.40)$$

Using $\check{g}_1(\cdot)$ as an approximation of the LSF, the failure probability can be approximated as:

$$p_f \approx p_f^{\text{FORM}} = \mathbb{P}(-\boldsymbol{\alpha}^\top \mathbf{u} \leq -\beta) = \Phi(-\beta), \quad (1.41)$$

with $\Phi(\cdot)$ the CDF of the standard Gaussian. Depending on the properties of the LFS, this approximation will be more or less accurate. Note that for a purely linear LFS, $p_f = p_f^{\text{FORM}}$. When the function is nonlinear, adding a quadratic term to the Taylor expansion can help the approximation. The approximation method is then called SORM for *second order reliability method*. However, this added complexity implies the computation of Hessian matrices, which can be complicated (see Chapter 1 from Bourinet 2018 for their estimation).

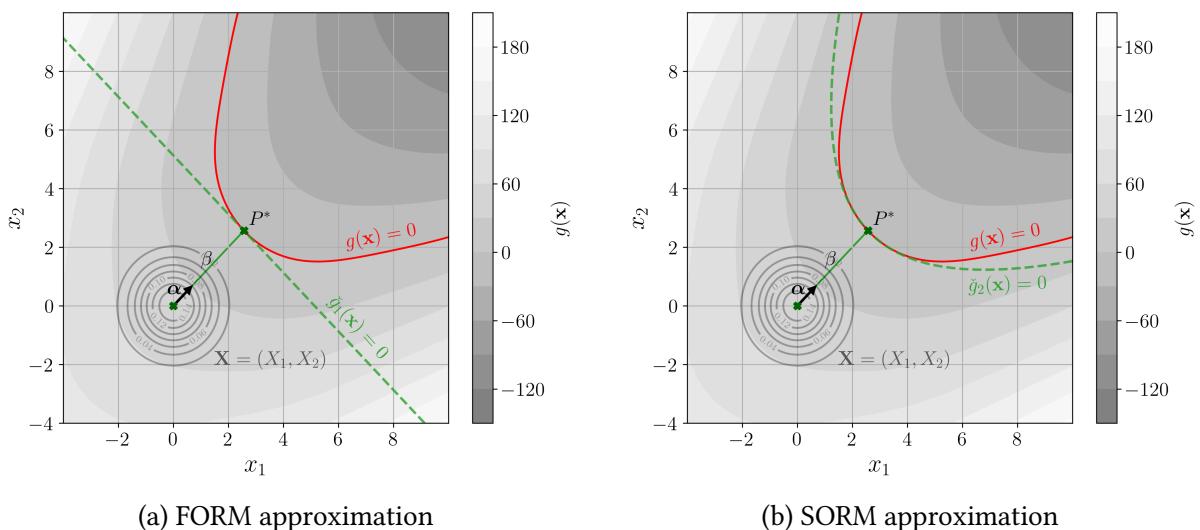


Figure 1.8 FORM and SORM approximation on a two-dimensional reliability problem

When the MPFP is not unique, the application of these methods might lead to important errors. From a geometrical point of view, having more than one MPFP means that more than one failure zone is at the same Euclidean distance of the origin. Applying a FORM or SORM resolution in this particular case leads to the estimation of only one of the failure areas. The *muti-FORM* algorithm (see [Der Kiureghian and Dakessian \(1998\)](#)) prevents this situation by applying successive FORM. Once the first MPFP $P^{*(1)}$ is found, the limit-state surface is modified by removing a nudge to find to following MPFP $P^{*(2)}$, positioned at a similar distance but in a different direction.

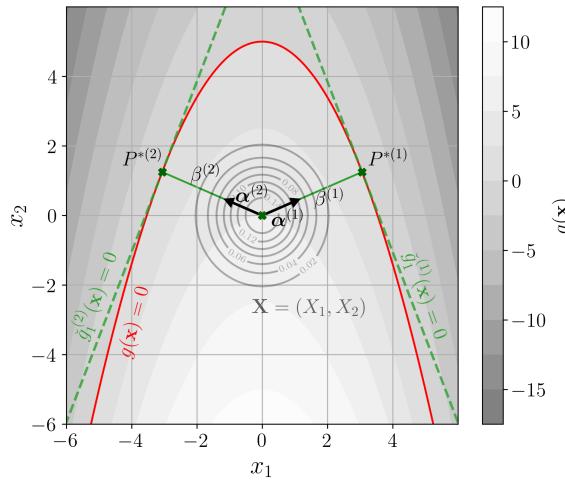


Figure 1.9 Multi-FORM approximation on an example with two MPFPs

Overall, FORM and SORM methods deliver a very efficient approximation of small probabilities for relatively simple problems (in terms of linearity and dimension). For this reason, they have been widely used in the practical context of limited simulation budgets. [Straub \(2014\)](#) illustrates the efficiency of FORM approaches on industrial cases such as probabilistic fatigue damage. However, these methods present serious limits as the dimension increases (see the discussion in Chapter 1 from [Chabridon 2018](#)). Additionally, their main drawback is the lack of complementary information concerning the confidence of the results. The textbook example illustrated in Fig. 1.9 shows that the method might miss some important areas of the failure domain, leading to poor estimations. As an alternative to approximation methods, simulation-based methods often provide the analyst with an assessment of the estimation's confidence.

Monte Carlo

Crude Monte Carlo sampling is a universal and empirical method for uncertainty propagation. As introduced earlier, it relies on the pseudo-random generation of i.i.d. samples $\{\mathbf{x}^{(i)}\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}}$. Only the estimator is now written using the indicator function applied to the LSF:

$$p_f \approx \hat{p}_f^{\text{MC}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\mathcal{F}_x}(\mathbf{x}^{(i)}). \quad (1.42)$$

Provided that the failure probability is bounded, this estimator converges toward it almost surely according to the law of large numbers. Once again, Monte Carlo offers an unbiased estimator, regardless of the problem's dimension or the regularity of the function $g(\cdot)$. Additionally, the variance of this estimator is fully known:

$$\text{Var}(\widehat{p}_f^{\text{MC}}) = \frac{1}{n} p_f (1 - p_f). \quad (1.43)$$

The variance of this estimator can be used to build its confidence interval according to the central limit theorem (similarly to the ones from Eq. (1.17)). Because of the small scale of the quantities manipulated in rare event estimation, the estimator's coefficient of variation is also widely used:

$$\delta_{\widehat{p}_f^{\text{MC}}} = \frac{\sqrt{\text{Var}(\widehat{p}_f^{\text{MC}})}}{\mathbb{E}[\widehat{p}_f^{\text{MC}}]} = \sqrt{\frac{1 - p_f}{np_f}}. \quad (1.44)$$

In theory, Monte Carlo estimation presents multiple advantages for rare event estimation. First, this method can be applied directly in the physical space, without transformation (which is practical for complex input distributions). Second, it does not suffer from the curse of dimensionality. Third, it is qualified as an embarrassingly parallel method since each of the numerical simulations is independent. Finally, it offers strong convergence guarantees and complementary information on the estimation confidence. These properties often make Monte Carlo the reference method in rare event estimation benchmarks.

However, the advantages of this estimator are shadowed by its slow convergence. To estimate a target failure probability $p_f = 10^{-\alpha}$, a Monte Carlo estimation with a convergence level $\delta_{\widehat{p}_f^{\text{MC}}} = 0.1$ famously requires $n = 10^{\alpha+2}$ simulations.

In the context of rare event estimation, Monte Carlo needs a number of simulation that is often prohibitive in practice. This excessive simulation budget comes from the fact that the vast majority of the samples drawn from the input distribution are not in the failure domain.

Importance sampling

Importance sampling (IS) is a variance reduction method, aiming at improving the performances of crude Monte Carlo sampling. In the context of rare event estimation, the main idea is to deliberately introduce a bias in the sampled density, shifting it towards the failure domain. If this shift actually goes towards the failure domain, it allows drawing more points in it, leading to a better estimate of our quantity.

The challenge in importance sampling is to pick a relevant *instrumental* distribution h_X (also called *auxiliary* distribution) to replace the distribution f_X . Then, by introducing the fully known *likelihood ratio* $w_X(x) = \frac{f_X(x)}{h_X(x)}$, one can rewrite $f_X(x) = w_X(x)h_X(x)$ and inject it in the failure probability expression:

$$p_f = \int_{\mathcal{D}_x} \mathbb{1}_{\mathcal{F}_x}(x) f_X(x) dx = \int_{\mathcal{D}_x} \mathbb{1}_{\mathcal{F}_x}(x) w_X(x) h_X(x) dx. \quad (1.45)$$

This simple writing trick allows us to integrate against the auxiliary distribution. With a Monte Carlo method, this task should be easier than integrating directly against the initial distribution.

The importance sampling estimator of the failure probability is defined for a sample drawn on the auxiliary distribution $\{\mathbf{x}^{(i)}\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} h_{\mathbf{X}}$:

$$\hat{p}_f^{\text{IS}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}^{(i)}) w_{\mathbf{X}}(\mathbf{x}^{(i)}). \quad (1.46)$$

Similarly to Monte Carlo, this estimator is unbiased, however, its variance is defined as:

$$\text{Var}(\hat{p}_f^{\text{IS}}) = \frac{1}{n} \left(\mathbb{E}_{h_{\mathbf{X}}} \left[\left(\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{X}) \frac{f_{\mathbf{X}}(\mathbf{X})}{h_{\mathbf{X}}(\mathbf{X})} \right)^2 \right] - p_f^2 \right). \quad (1.47)$$

The quality of the variance reduction associated to this technique fully depends on the choice of the instrumental distribution. In fact, importance sampling can lead to higher variance than crude Monte Carlo when the instrumental distribution is poorly chosen (Owen and Zhou, 2000). However, an optimal instrumental distribution h_{opt} theoretically gives the smallest variance by setting it equal to zero in Eq. (1.47):

$$h_{\text{opt}}(\mathbf{x}) = \frac{\mathbb{1}_{\mathcal{F}_{\mathbf{x}}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_f}. \quad (1.48)$$

The optimal expression above is unfortunately not usable in practice since it includes the targeted quantity p_f . Considering this framework, various techniques intend to define instrumental distributions as close as possible to this theoretical result. An important review of the use of importance sampling in the context of reliability analysis was proposed by Tabandeh et al. (2022).

The most immediate solution is to combine the information provided by the results of FORM with importance sampling, simply called FORM-IS. In practice, the instrumental distribution is defined as the initial distribution centered on the design point resulting from FORM. Fig. 1.10 illustrates in the same two-dimensional case, the estimation by Monte Carlo and importance sampling centered on the design point. The points in red reached the failure domain and their number seems insufficient for Monte Carlo. Note that comparing the results from FORM and FORM-IS allows us to assess the nonlinearity of the LSF in the vicinity of the design point. This strategy is simple to implement, but it inherits the main drawbacks of FORM, such as the limits related to multiple failure areas (see the example illustrated in Fig. 1.9). Finally, other importance sampling schemes integrate adaptive mechanisms, progressively leading the sampling towards the failure domain (Bugallo et al., 2017).

Adaptive importance sampling by cross-entropy

The *cross-entropy-based adaptive importance sampling* (CE-AIS) is an adaptive strategy, optimizing the IS variance reduction by searching for the best instrumental distribution within a

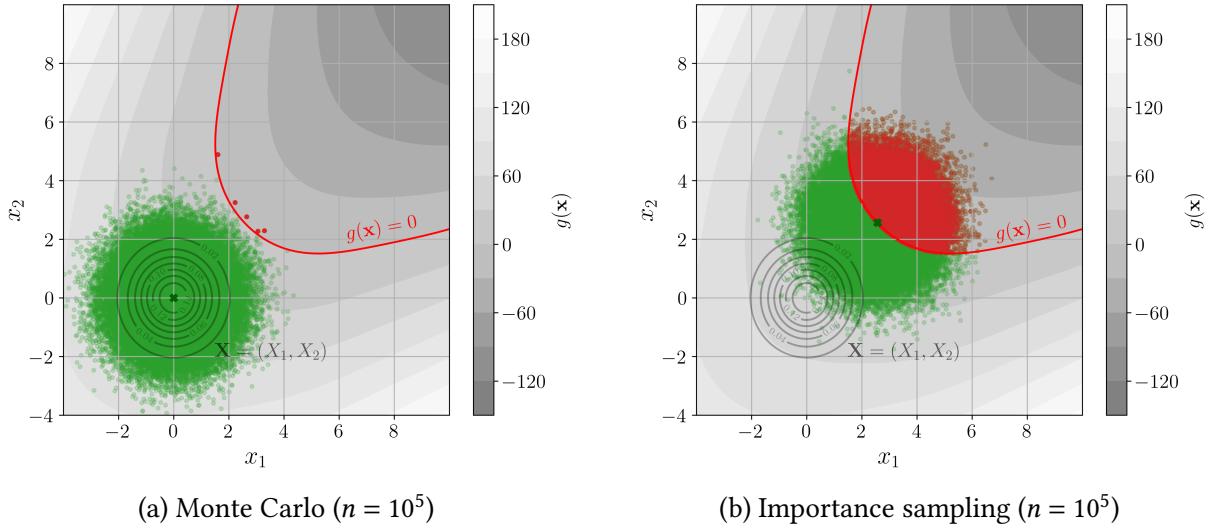


Figure 1.10 Illustration of a rare event estimation.

parametric family. Let us consider the distribution h_λ , belonging to the parametric family \mathcal{H}_λ , defined for a set of parameters λ in a parameter space $\mathcal{D}_\lambda \subseteq \mathbb{R}^p$ as:

$$\mathcal{H}_\lambda = \{\mathbf{x} \mapsto h_X(\mathbf{x}|\lambda) = h_\lambda(\mathbf{x}), \quad \lambda = (\lambda_1, \dots, \lambda_p) \in \mathcal{D}_\lambda \subseteq \mathbb{R}^p\}. \quad (1.49)$$

The early work of [Bucher \(1988\)](#) only included normal distributions to minimize the IS variance w.r.t. the parameter λ , using Eq. (1.47) the optimization simplifies as:

$$\lambda^* = \arg \min_{\lambda \in \mathcal{D}_\lambda} \mathbb{E}_{h_\lambda} \left[\mathbb{1}_{\mathcal{F}_x}(\mathbf{X}) \left(\frac{f_X(\mathbf{x})}{h_\lambda(\mathbf{x})} \right)^2 \right]. \quad (1.50)$$

However, this optimization strategy requires sampling with respect to the instrumental distribution at each optimization iteration, which was overcome by a different approach.

The “cross-entropy” (CE) method uses Kullback-Leibler (KL) divergence to optimize importance sampling. KL divergence is a dissimilarity measure between distributions, expressed between the parametric instrumental distribution h_λ and the optimal one h_{opt} :

$$D_{\text{KL}}(h_{\text{opt}} || h_\lambda) = \int_{\mathcal{D}_x} \log \left(\frac{h_{\text{opt}}(\mathbf{x})}{h_\lambda(\mathbf{x})} \right) h_{\text{opt}}(\mathbf{x}) d\mathbf{x} \quad (1.51a)$$

$$= \int_{\mathcal{D}_x} \log(h_{\text{opt}}(\mathbf{x})) h_{\text{opt}}(\mathbf{x}) d\mathbf{x} - \int_{\mathcal{D}_x} \log(h_\lambda(\mathbf{x})) h_{\text{opt}}(\mathbf{x}) d\mathbf{x}. \quad (1.51b)$$

[Rubinstein and Kroese \(2004\)](#) symplify the expression of the optimization problem minimizing the KL divergence, which is most often convex and differentiable w.r.t. λ :

$$\lambda^* = \arg \min_{\lambda \in \mathcal{D}_\lambda} D_{\text{KL}}(h_{\text{opt}} || h_\lambda). \quad (1.52)$$

By injecting the expression in Eq. (1.51b), the optimization problem simply becomes a function of an expected value over the initial density f_X :

$$\lambda^* = \arg \max_{\lambda \in \mathcal{D}_\lambda} \int_{\mathcal{D}_x} \log(h_\lambda(\mathbf{x})) h_{\text{opt}}(\mathbf{x}) d\mathbf{x} = \arg \max_{\lambda \in \mathcal{D}_\lambda} \mathbb{E}_{f_X} [\mathbb{1}_{\mathcal{F}_x}(\mathbf{X}) \log(h_\lambda(\mathbf{X}))]. \quad (1.53)$$

To directly estimate this expected value, the failure probability p_f should not be too rare, which allows to use of an empirical estimator of the expected value:

$$\lambda^* = \arg \max_{\lambda \in \mathcal{D}_\lambda} \sum_{i=1}^n \mathbb{1}_{\mathcal{F}_x}(\mathbf{x}^{(i)}) \log(h_\lambda(\mathbf{x}^{(i)})), \quad \{\mathbf{x}^{(i)}\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} f_X. \quad (1.54)$$

Eventually, this optimization can be solved by canceling the gradient:

$$\sum_{i=1}^n \mathbb{1}_{\mathcal{F}_x}(\mathbf{x}^{(i)}) [\nabla \log(h_{\lambda^*})](\mathbf{x}^{(i)}) = \mathbf{0}. \quad (1.55)$$

According to Rubinstein and Kroese (2004), this system of equations has a unique analytical solution when assuming that the instrumental distribution belongs to the “natural exponential family”.

However, when dealing with rare probabilities, the empirical estimation does not draw enough points in the failure domain to get an accurate estimate. The adaptive version of this technique, called *multilevel cross-entropy*, gradually builds a set of intermediate levels, decreasing towards the failure level (equal to zero). By working on a set of individually less rare events, the empirical estimation in Eq. (1.53) is made possible.

The algorithm starts by generating and evaluating an initial sample $\{g(\mathbf{X}_{[1]}^{(i)})\}_{i=1}^n$, on which a threshold level $q_{[1]}^{p_0}$ is computed as the empirical p_0 -quantile. Using the samples below the first threshold $q_{[1]}^{p_0}$, a first instrumental distribution $h_{\lambda_{[1]}^*}$ is optimized. At the next steps $k \in \{1, \dots, k_{\#}\}$, the sample $\{\mathbf{X}_{[k]}^{(i)}\}_{i=1}^n$ is generated from the density $h_{\lambda_{[k-1]}^*}$ and the rest of the process repeats until the estimated threshold level becomes negative, $q_{[k_{\#}]}^{p_0} \leq 0$.

The final instrumental density $h_{\lambda_{[k_{\#}]}^*}$ is then considered for importance sampling as defined in Eq. (1.46):

$$\hat{p}_f^{\text{CE-AIS}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \frac{f_X(\mathbf{x}_{[k_{\#}]}^{(i)})}{h_{\lambda_{[k_{\#}]}^*}(\mathbf{x}_{[k_{\#}]}^{(i)})}, \quad \{\mathbf{X}_{[k_{\#}]}^{(i)}\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} h_{\lambda_{[k_{\#}]}^*}. \quad (1.56)$$

CE-AIS is widely used in rare event estimation, as it develops an adaptive technique while conserving the explicit IS variance given in Eq. (1.47). According to Rubinstein and Kroese (2004), the successive instrumental distributions $h_{\lambda_{[k]}^*}$ converge towards h_{opt} under a few hypotheses. The most important one is that the optimal density must belong to the parametric family considered, which should offer enough flexibility to describe a wide range of distributions.

When the failure domain is composed of multiple regions, different improvements of the CE-AIS were proposed. [Kurtz and Song \(2013\)](#) proposed to optimize h_{λ^*} among a mixture of Gaussian distributions. This method was further studied by [Wang and Song \(2016\)](#) and [Papaioannou et al. \(2019\)](#) using advanced mixtures in the standard space. However, when using mixtures, the optimization problem does not have an analytical expression anymore ([Geyer et al., 2019](#)).

In the parametric framework, the family choice leads to a complicated trade-off between optimization complexity and flexibility allowed by the family. A similar mechanism is used by other importance sampling methods, inferring the optimal instrumental density by applying kernel density estimation to the points in the failure domain.

Nonparametric adaptive importance sampling

The use of multivariate kernel density estimation (KDE) to approximate the importance sampling optimal density h_{opt} was introduced in the context of structural reliability by [Ang et al. \(1992\)](#), latter followed by [Zhang \(1996\)](#). Let us first present the nonparametric importance sampling from [Zhang \(1996\)](#), considering the instrumental density $h_{[0]}$ (for now, $h_{[0]} \neq f_X$), on which a sample $\{\mathbf{X}_{[1]}^{(i)}\}_{i=1}^n$ is generated. A first failure probability can be roughly estimated, assuming that enough samples lead to the failure domain:

$$\widehat{p}_{f[1]} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \left(\mathbf{x}_{[1]}^{(i)} \right) \frac{f_X \left(\mathbf{x}_{[1]}^{(i)} \right)}{h_{[0]} \left(\mathbf{x}_{[1]}^{(i)} \right)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \left(\mathbf{x}_{[1]}^{(i)} \right) w_{[1]}^{(i)}. \quad (1.57)$$

On this biased sample, another density can be fitted using KDE, using the previously defined $\widehat{p}_{f[1]}$ as a normalization term:

$$\widehat{h}_{[1]}(\mathbf{x}) = \frac{\det(\mathbf{H}_{[1]})^{-1/2}}{n \widehat{p}_{f[1]}} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \left(\mathbf{x}_{[1]}^{(i)} \right) w_{[1]}^{(i)} K \left(\mathbf{H}_{[1]}^{-1/2} (\mathbf{x} - \mathbf{x}_{[1]}^{(i)}) \right). \quad (1.58)$$

Where the kernel K is commonly taken as the multivariate Gaussian-centered density with the covariance matrix \mathbf{H} . The tuning of \mathbf{H} is usually done by minimizing an asymptotic mean integrated squared error (AMISE) criterion ([Glad et al., 2007](#)). In the previous expression, the normalization constant ensures building a probability density while the weights $w_{[1]}^{(i)}$, defined above, reflect the contribution of each point to $\widehat{p}_{f[1]}$. After performing this KDE, the estimated density can be used as instrumental density in Eq. (1.46).

As for the CE-IS methods, the risk is that barely any points sampled from the instrumental density $h_{[0]}$ hit the failure domain, leading to poor estimates. [Zhang \(1996\)](#) proposed to couple an adaptive mechanism with a nonparametric inference of the optimal density. This method is further referred to as NAIS for *nonparametric adaptive importance sampling*. Later, the NAIS method was adapted by [Morio \(2011\)](#) to the reliability analysis problem, using a similar mechanism to the CE-AIS method.

In this framework, a series of intermediate thresholds are computed as empirical p_0 -quantiles $q_{[1]}^{p_0} > \dots > q_{[k_\#]}^{p_0}$ of the successive importance sampling steps. This algorithm is initiated by setting $h_{[0]} = f_X$ and stops at the step $k_\#$, when $q_{[k_\#]}^{p_0} < 0$.

At the step k , the intermediate normalization constant is written as:

$$\widehat{p}_f[k] = \frac{1}{kn} \sum_{j=1}^k \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq q_{[j]}^{p_0}\}} \left(\mathbf{x}_{[j]}^{(i)} \right) \frac{f_X \left(\mathbf{x}_{[j]}^{(i)} \right)}{\widehat{h}_{[j-1]} \left(\mathbf{x}_{[j]}^{(i)} \right)} = \frac{1}{kn} \sum_{j=1}^k \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq q_{[j]}^{p_0}\}} \left(\mathbf{x}_{[j]}^{(i)} \right) w_{[j]}^{(i)}, \quad (1.59)$$

with $\left\{ \mathbf{x}_{[j]}^{(i)} \right\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} h_{[j-1]}$. Then, an intermediate instrumental density is inferred by KDE on the samples exceeding the threshold $q_{[k]}^{p_0}$ such that:

$$\widehat{h}_{[k+1]}(\mathbf{x}) = \frac{\det(\mathbf{H}_{[k]})^{-1/2}}{kn \widehat{p}_f[k]} \sum_{j=1}^k \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq q_{[j]}^{p_0}\}} \left(\mathbf{x}_{[j]}^{(i)} \right) w_{[j]}^{(i)} K \left(\mathbf{H}_{[k]}^{-1/2} (\mathbf{x} - \mathbf{x}_{[j]}^{(i)}) \right). \quad (1.60)$$

The last instrumental density $\widehat{h}_{[k_\#]}$ is finally considered as an approximation of the optimal density for importance sampling introduced in Eq. (1.46):

$$\widehat{p}_f^{\text{NAIS}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \frac{f_X \left(\mathbf{x}_{[k_\#]}^{(i)} \right)}{\widehat{h}_{[k_\#]} \left(\mathbf{x}_{[k_\#]}^{(i)} \right)}, \quad \left\{ \mathbf{x}_{[k_\#]}^{(i)} \right\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} h_{k_\#}. \quad (1.61)$$

Overall, the NAIS offers more flexibility to infer the optimal importance sampling density. This property might suit problems presenting a highly nonlinear limit state function. Then, relying on importance sampling still provides an expression of the estimator's variance, by adapting Eq. (1.47) to the recurrent mechanism in NAIS. Because this approach depends on KDE, it inherits its drawbacks. As discussed in Morio (2011), tuning the KDE can create numerical issues and KDE famously suffers from the curse of dimensionality. In practice, the performances of NAIS seriously decrease for problems in dimensions larger than ten.

Subset simulation

Although the concept of splitting already existed, the name of *subset simulation* (SS) was first introduced by Au and Beck (2001) in the structural reliability community. This concept was generalized as a sequential Monte Carlo method under the name of “adaptive multilevel splitting”, as reviewed by Cérou et al. (2019).

Subset simulation splits the failure event \mathcal{F}_x into an intersection of $k_\#$ intermediary events $\mathcal{F}_x = \cap_{k=1}^{k_\#} \mathcal{F}_{[k]}$. Each are nested such that $\mathcal{F}_{[1]} \supset \dots \supset \mathcal{F}_{[k_\#]} = \mathcal{F}_x$. The failure probability is then expressed as a product of conditional probabilities:

$$p_f = \mathbb{P}(\mathcal{F}_x) = \mathbb{P}(\cap_{k=1}^{k_\#} \mathcal{F}_{[k]}) = \prod_{k=1}^{k_\#} \mathbb{P}(\mathcal{F}_{[k]} | \mathcal{F}_{[k-1]}). \quad (1.62)$$

From a practical point of view, the analyst tunes the algorithm⁹ by setting the intermediary probabilities $\mathbb{P}(\mathcal{F}_{[k]} | \mathcal{F}_{[k-1]}) = p_0, \forall k \in \{1, \dots, k_{\#}\}$. Then, the corresponding quantiles $q_{[1]}^{p_0} > \dots > q_{[k_{\#}]}^{p_0}$ are estimated for each conditional subset samples $\mathbf{X}_{[k],N}$ of size N . Note that the initial quantile is estimated by crude Monte Carlo sampling on the input PDF f_X . Following conditional subset samples are generated by *Monte Carlo Markov Chain* (MCMC) sampling of $f_X(\mathbf{x} | \mathcal{F}_{[k-1]})$, using as seeds initialization points the $n = Np_0$ samples given by $\mathbf{A}_{[k],n} = \{\mathbf{X}_{[k-1]}^{(i)} \subset \mathbf{X}_{[k-1],N} | g(\mathbf{X}_{[k-1]}^{(i)}) > \hat{q}_{[k-1]}^{p_0}\}_{i=1}^n$. This process is repeated until an intermediary quantile becomes negative: $\hat{q}_{[k_{\#}]}^{p_0} < 0$. Finally, the failure probability is estimated by:

$$\hat{p}_f^{\text{SS}} = p_0^{k_{\#}-1} \frac{1}{N} \sum_{i=1}^n \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}}(\mathbf{X}_{[k_{\#}],N}^{(i)}). \quad (1.63)$$

In practice, the subset sample size should be large enough to properly estimate intermediary quantiles, leading to the usual recommendation of $p_0 = 0.1$. Fig. 1.11 illustrates the consecutive subset samples moving towards the failure domain. At each step of the algorithm (corresponding to a color), a subset is generated and an intermediate quantile is estimated.

[Au and Beck \(2001\)](#) also provide bounds to the coefficient of variation of \hat{p}_f^{SS} . The first one results from a first-order Taylor expansion of Eq. (1.63) and is often considered as an upper bound. The second assumes the estimations of the conditional probabilities to be independent and tends to underestimate the coefficient of variation.

As discussed in ([Papaioannou et al., 2015](#)), the efficiency of the SS method depends on the choice and tuning of the MCMC algorithm. The Metropolis–Hastings (MH) algorithm is widely used as MCMC algorithm for subset simulation, however, it quickly becomes inefficient as the dimension increases. Different improvements of the MH are made possible by working in the standard space, such as the “component-wise” (or “modified M–H”). More recently, alternative MCMC methods including physical system dynamics (e.g., Hamiltonian MCMC) showed promising results in high-dimension reliability problems ([Papakonstantinou et al., 2023](#)).

The subset simulation is a versatile method, presenting consistent performances even for rare probabilities. Its flexibility allows it to deal with highly nonlinear LSF, but its drawbacks arise from the use of MCMC sampling. The convergence of MCMC is complex to control and depends on its tuning, in addition, the MCMC samples are dependent. Unlike the methods derived from importance sampling, the variance of \hat{p}_f^{SS} is only approximated.

1.5.3 Summary and discussion

This section introduced the generic formulation and the main methods for rare event estimation. Even if the problem is generic rare event estimation requires tailored solutions. Depending on the properties of the problem tackled some methods might outperform others. Beyond the one introduced previously, many more methods are worth mentioning in the field of reliability analysis, such as the *directional sampling* [Bjerager \(1988\)](#), or *line sampling* ([Koutsourelakis,](#)

⁹An algorithmic presentation of the generic subset simulation method is given in Appendix C.

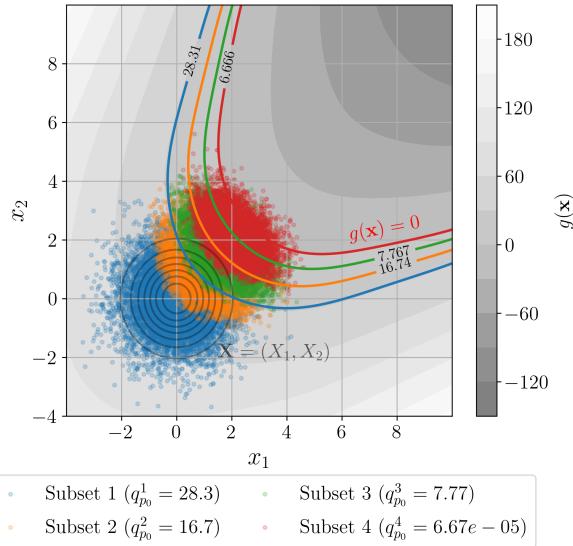


Figure 1.11 Illustration of a rare event estimation by subset simulation ($n = 4 \cdot 10^4, p_0 = 0.1$).

2004), moving particles (Walter, 2015). Morio and Balesdent (2015) compares the advantages and drawbacks of the most commonly used one, with corresponding algorithmic descriptions and numerical benchmarks.

Overall, the main properties increasing the complexity of reliability problems are related to:

- the computational cost of the limit-state function evaluation;
- the strong nonlinearity of the limit-state function;
- the rareness of the failure event.

In regard to the methods, the estimation is made easier by algorithms with simple tuning or allowing to work in the physical space (avoiding a possibly complex iso-probabilistic transform). Considering all these elements the analyst may set up a sampling strategy, possibly coupled with the use of a surrogate model (further discussed in Section 1.7).

Nevertheless, the unified formulation of reliability analysis problems (see 1.31) is an opportunity for the community to share standardized benchmark problems. Following the well-accepted benchmark platform for optimization “Comparing Continuous Optimizers” (COCO) (Hansen et al., 2021), an equivalent initiative was proposed for structural reliability. In 2019, the “black-box reliability challenge”, was organized as a hackathon by the Dutch organization for applied scientific research (TNO) (Rozsas and Slobbe, 2019). This platform proposed a large catalog of reliability problems with their respective solutions. Most of them were encapsulated as a Python package called `otbenchmark`¹⁰ (Fekhari et al., 2021), based on core OpenTURNS objects.

When working with computationally expensive numerical models, the direct use of rare event estimation methods is most often intractable. Many contributions were dedicated to the coupling of surrogate models with sampling methods for rare event estimation. Moustapha et al.

¹⁰<https://github.com/mbaudin47/otbenchmark/>

(2022) presented the results of a wide benchmark on the challenge from TNO, obtained by using surrogate models for reliability developed in the UQLab software (Marelli and Sudret, 2014).

In any case, risk assessment analysts should favor the methods offering convergence guarantees over punctual performance demonstrations. Finally, the robustness of the failure probability to the input uncertainty model is a major question, which was studied from probabilistic (Lemaître et al., 2015) and extra-probabilistic (Ajenjo et al., 2022) frameworks.

OpenTURNS 4 (Rare event estimation). The Python code available in Appendix D proposes a minimalistic OpenTURNS example to estimate rare event probabilities. Figures illustrating the present section may be reproduced, using the OpenTURNS scripts available on GitHub¹¹.

1.6 Global sensitivity analysis

The aim of sensitivity analysis (SA) is to determine the impact of a single (or a group of random inputs) on a random output(s). As described earlier, this step is qualified as an inverse analysis in the general UQ framework (illustrated in Fig. 1), in opposition to the forward uncertainty propagation step. In fact, the analyst studies the effect of the inputs at different scales, hence the distinction between “local” and “global” SA. Local SA focuses on the impact of small perturbations around nominal values of the inputs (i.e., derivative-based approaches), while global sensitivity analysis (GSA), typically studies the general variability (e.g., the variance) of the output. Two types of GSA methods exist in the literature, either proposing qualitative or quantitative approaches:

- *screening methods*: determines the non-influential variables in a UQ study (qualitative);
- *importance measures*: assess the contribution of inputs in the global variability of the output (quantitative).

Screening methods are typically used in a statistical learning process, to drop the irrelevant variables to the learning. In this context, *feature selection* serves the same purpose with a slight difference. Screening methods usually assume the inputs to be independent while feature selection does not. Moreover, feature selection not only looks for the irrelevant features to the learning but also the redundant features (Fan and Lv, 2010).

The global sensitivity of an output can be explained by different elements: the marginal effects of the inputs, their dependence, and their interactions. Two variables present interactions when their simultaneous effect on an output is not additive. Note that SA on dependent inputs is an active field of research.

¹¹https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/reliability.ipynb

1.6.1 Screening methods

Many UQ methods suffer from the curse of dimensionality, thankfully, high-dimensional problems often only depend on a few variables. This observation was formalized with the concept of *effective dimension* introduced by [Owen \(2003\)](#). Screening methods allow discriminating the non-influential variables, which can be considered afterward as determinist to simplify the problem.

Morris method

The Morris method ([Morris, 1991](#)) is a screening method historically used in engineering applications. It starts by mapping the input domain \mathcal{D}_X into a unit hypercube $[0, 1]^d$, which is discretized as a regular grid with step $\Delta \in \mathbb{R}$. The algorithm computes local elementary sensitivity by building “one at a time” (OAT) local trajectories over the regular grid. Each OAT design starts at a random node $\mathbf{x}^{(t)} = (x_1^{(t)}, \dots, x_j^{(t)}, \dots, x_d^{(t)})$ of the grid, and moves only in one direction by an increment equal to the elementary step such that: $\mathbf{x}^{(t)} + \Delta_j = (x_1^{(t)}, \dots, x_j^{(t)} + \Delta, \dots, x_d^{(t)})$. The elementary effect in the direction of the variable i from an OAT trajectory t is expressed as a finite difference:

$$\text{EE}_j^{(t)} = \frac{g(\mathbf{x}^{(t)}) - g(\mathbf{x}^{(t)} + \Delta_j)}{\Delta}. \quad (1.64)$$

The Morris method generates $T \in \mathbb{N}$ OAT trajectories and computes theirs respective elementary effects in each direction i . To assess the global sensitivity of the function, the mean $\overline{\text{EE}}_j$ and variance $\widehat{\text{Var}}(\text{EE}_j)$ of the elementary effects are computed:

$$\overline{\text{EE}}_j = \frac{1}{n} \sum_{t=1}^T |\text{EE}_j^{(t)}|, \quad \widehat{\text{Var}}(\text{EE}_j) = \frac{1}{n-1} \sum_{t=1}^T \left(\text{EE}_j^{(t)} - \overline{\text{EE}}_j \right)^2. \quad (1.65)$$

It allows to divide the variables into three categories, regardless of any regularity hypothesis on the function: (i) negligible effects; (ii) linear effects without interaction; and (iii) nonlinear effects with possible interactions. This method is very intuitive but quickly shows its limits as the dimension increases since it relies on a discretization of the space by a regular grid. Another disadvantage of this method is that it does not distinguish interactions and nonlinear effects of inputs.

Derivative-based global sensitivity measures

The Derivative-based global sensitivity measures (DGSM) are a GSA method introduced in [Sobol and Gresham \(1995\)](#) and further studied in [Kucherenko et al. \(2009\)](#). As the Morris method, they study the mean value of local derivatives of the model output with regard to the inputs:

$$v_j = \int_{\mathcal{D}_X} \left(\frac{\partial g(\mathbf{x})}{\partial x_j} \right)^2 f_X(\mathbf{x}) d\mathbf{x} = \mathbb{E} \left[\left(\frac{\partial g(\mathbf{X})}{\partial X_j} \right)^2 \right]. \quad (1.66)$$

This continuous formulation does not require using OAT designs, which was proven to be more efficient when exploiting sampling methods such as quasi-Monte Carlo. The efficiency of the DGSMs for screening purposes was outlined in many papers (e.g., [Kucherenko and Iooss \(2017\)](#)). Since their value depends on the probability distribution of the input, a normalized version was developed. The connections between DGSM and variance-based GSA measures (i.e., Sobol' indices introduced hereafter), revealed bounding properties between DGSMs and Sobol' total indices ([Roustant et al., 2017](#)).

1.6.2 Variance-based importance measures

Screening methods determine the non-influential variables in a UQ problem. Beyond this information, importance measures quantify the influence of inputs, allowing us to rank the inputs according to their contribution to the output variability.

Functional variance decomposition and Sobol' indices

Sobol' indices are the most popular importance measure in GSA. Their universality comes from the functional decomposition of the output's variance, attributing variance share to the inputs. Considering a squared-integrable and measurable function $g(\cdot)$ and an independent random vector \mathbf{X} . The output random variable $Y = g(\mathbf{X})$ can be decomposed, according to [Hoeffding \(1948\)](#), as:

$$Y = g(\mathbf{X}) = g_0 + \sum_{j=1}^d g_j(X_j) + \sum_{j < l}^d g_{jl}(X_j, X_l) + \dots + g_{1\dots d}(\mathbf{X}), \quad (1.67)$$

with the previous terms defined according to this recurrence:

$$g_0 = \mathbb{E}[g(\mathbf{X})] \quad (1.68a)$$

$$g_j(X_j) = \mathbb{E}[g(\mathbf{X})|X_j] - g_0 \quad (1.68b)$$

$$g_{jl}(X_j, X_l) = \mathbb{E}[g(\mathbf{X})|X_j, X_l] - g_j(X_j) - g_l(X_l) - g_0 \quad (1.68c)$$

$$\dots \quad (1.68d)$$

Different authors proved that this decomposition is unique by exploiting the orthogonality of the terms of the decomposition ([Efron and Stein, 1981](#); [Sobol', 1993](#)). Therefore, this decomposition can be transposed in terms of functional decomposition of variance (also called functional analysis of variance or FANOVA):

$$\text{Var}(Y) = \sum_{j=1}^d V_j(Y) + \sum_{j < l}^d V_{jl}(Y) + \dots + V_{1\dots d}(Y), \quad (1.69)$$

where the previous terms are defined in a recurrent way, in the same fashion as Eq. (1.68): $V_j(Y) = \text{Var}(\mathbb{E}[Y|X_j])$, $V_{jl}(Y) = \text{Var}(\mathbb{E}[Y|X_j, X_l]) - V_j(Y) - V_l(Y)$, and so on for higher order interaction terms. The Sobol' indices of different order are defined as normalized shares of variance. The *first-order Sobol' index* S_j quantifies the share of variance of the output only explained by the marginal X_j (also called main effect). Second order S_{jl} (or higher order) Sobol' indices quantify the effect of the interactions between a group of marginals.

$$S_j = \frac{V_j(Y)}{\text{Var}(Y)} = \frac{\text{Var}(\mathbb{E}[Y|X_j])}{\text{Var}(Y)} \quad (1.70a)$$

$$S_{jl} = \frac{V_{jl}(Y)}{\text{Var}(Y)} = \frac{\text{Var}(\mathbb{E}[Y|X_j, X_l]) - V_j(Y) - V_l(Y)}{\text{Var}(Y)} \quad (1.70b)$$

$$\dots \quad (1.70c)$$

The generic definition of the Sobol' sensitivity indices associated with a subset of inputs $A \in \mathcal{P}_d$, with \mathcal{P}_d the set of all possible subsets of $\{1, \dots, d\}$, is given by:

$$S_A = \frac{V_A(Y)}{\text{Var}(Y)} = \frac{\sum_{B \subset A} (-1)^{|A|-|B|} \text{Var}(\mathbb{E}[Y|X_B])}{\text{Var}(Y)}. \quad (1.71)$$

By using the functional decomposition of variance in Eq. (1.69), one can show that the Sobol' indices add up to one:

$$\sum_{A \in \mathcal{P}_d} S_A = 1. \quad (1.72)$$

The so-called *closed Sobol' index* associated to a subset of inputs $A \in \mathcal{P}_d$ (equivalent to the first-order Sobol' index of A) is defined as:

$$S_A^{\text{clos}} = \sum_{A' \subset A} S_{A'} = \frac{\text{Var}(\mathbb{E}[Y|X_A])}{\text{Var}(Y)}. \quad (1.73)$$

Assessing Sobol' indices for every order becomes complex in medium to high dimensions. The *total Sobol' index* S_j^T associated with the variable j , see Homma and Saltelli (1996), quantifies the share of output variance which is explained by all the interactions of the variable X_j :

$$S_j^T = 1 - \frac{\text{Var}(\mathbb{E}[Y|X_{-j}])}{\text{Var}(Y)} = \frac{\mathbb{E}[\text{Var}(Y|X_{-j})]}{\text{Var}(Y)}, \quad (1.74)$$

where X_{-j} represents all the marginals from X but X_j . This definition can also be generalized for a subset of inputs $A \in \mathcal{P}_d$, such that:

$$S_A^T = 1 - S_{A^C}^{\text{clos}} = 1 - \frac{\text{Var}(\mathbb{E}[Y|X_{A^C}])}{\text{Var}(Y)}, \quad A^C = \mathcal{P}_d \setminus A \quad (1.75)$$

By analyzing jointly the first and total Sobol' indices, one can get an indication about the decomposition between the marginal and interaction effects. Note that the total indices are

only equal to the first indices when the model does not present interactions (i.e., purely additive model).

Estimating Sobol' indices can be achieved in various ways, even if historically the *pick-freeze* scheme was the most popular. This method is based on two samples, but it often requires a prohibitive number of evaluations of the function. Many estimators using the pick-freeze generic scheme were developed to estimate Sobol' indices (e.g., Saltelli's, Jansen's, Martinez's etc.), see further details in Chapter 3 of [Da Veiga et al. \(2021\)](#). Alternatively, the surrogate models were exploited to estimate such sensitivity measures. Using an input-output dataset, the analyst may build a *polynomial chaos expansion* (PCE) surrogate model, which gives an explicit expression of the Sobol' indices ([Sudret, 2008](#)). Authors such as [Marrel et al. \(2009\)](#) also studied the use of Gaussian processes for this purpose.

In the case of independent inputs, the first and total Sobol' indices are a complete tool for GSA. The main advantage of this approach is the quantitative nature of its results, allowing to objectively compare the effect of input variables. When the inputs present a dependence structure, it becomes complicated to distinguish its effects from possible interactions. However, many authors tried to adapt Sobol' indices to this context. Chapter 5 of [Da Veiga et al. \(2021\)](#) reviews four of these approaches. For example, [Mara and Tarantola \(2012\)](#) proposed two extra Sobol' indices, called "full indices", detecting the contributions associated with the inputs' dependence. Note that the interpretation and estimation of this solution becomes complicated. Moreover, unlike the independent case, the four Sobol' indices do not divide the output variance between the inputs. Beyond Sobol' indices, another important GSA method was adapted from the theory of Shapley values by [Owen \(2014\)](#), allowing to work with dependent inputs.

OpenTURNS 5 (Sobol' indices). The Python code available in Appendix D gives a minimalist OpenTURNS implementation of the Sobol' indices to assess global sensitivity analysis on the Ishigami analytical problem. Further scripts are also available on GitHub¹².

Shapley effects

Shapley effects are an adaptation to GSA by [Owen \(2014\)](#) of the Shapley values from the cooperative games' theory ([Shapley et al., 1953](#)). This method is an alternative to the Sobol' indices in the case of dependent inputs, for which the natural interpretation of single interaction effects no longer holds. In the game theory, Shapley values act as a rule on how to share the value created by a team between its members (players). The Shapley value allocated to the player X_j is given considering the indices $\{j\} = \{1, \dots, d\} \setminus \{j\}$:

$$\varsigma_j = \sum_{A \subset -\{j\}} \binom{d-1}{\text{card}(A)}^{-1} (\text{val}(A \cup \{j\}) - \text{val}(A)), \quad (1.76)$$

¹²https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/sensitivity_analysis.ipynb

where the value (or cost) function is denoted by $\text{val}(A)$, and A is a subset of $\{1, \dots, d\}$ with cardinality $\text{card}(A)$. The Shapley effects adapted this concept to perform a GSA by considering the variables as players and the closed Sobol' indices for the value function:

$$Sh_j = \sum_{A \subset -\{j\}} \binom{d-1}{|A|}^{-1} \left(S_{A \cup \{j\}}^{\text{clos}} - S_A^{\text{clos}} \right). \quad (1.77)$$

Conceptually, this expression compares a performance defined by a cost function with or without the variable X_j , and averages it over all the possible combinations of inputs. This importance measure offers the following decomposition:

$$\sum_{j=1}^d Sh_j = 1. \quad (1.78)$$

In the case of independent inputs, the Shapley effects present properties related to the Sobol' indices. The following equation (see proof in [Owen \(2014\)](#)) reveal that the Shapley effects equally divide the interaction effects between the implicated variable:

$$S_j \leq Sh_j \leq S_j^T, \quad Sh_j = \sum_{A \in \mathcal{P}_{d,j} \in A} \frac{S_A}{\text{card}(A)}. \quad (1.79)$$

Unlike the Sobol' indices, Shapley effects are a nonnegative allocation of output variance with equitable division of the interaction effects. This method presents an interesting alternative in the dependent case, however, estimating Shapley effects creates computational difficulties. The reader may refer to the permutation-based algorithm from [Song et al. \(2016\)](#). Surrogate models were also coupled to estimate Shapley effects, using Gaussian processes in [Benoumechiara and Elie-Dit-Cosaque \(2019\)](#) and random forests in [Bénard et al. \(2022\)](#).

Shapley effects are a promising importance measure based on variance allocation. However, in some cases, the variance of the output distribution does not represent well its variability (e.g., multimodal distribution). The following section introduces another family of GSA methods based on distances between distributions.

1.6.3 Moment-independent importance measures

Beyond variance-based GSA, many types of distances between distributions have been used to evaluate the dependence between the input and output distributions. Comparing the entire distributions instead of their moments might be more robust in some cases (e.g., when the variance is a poor indicator of the variability). The tools used to do so are generally called *dissimilarity measures* between distributions. Appendix D briefly introduces two families of dissimilarity measures: the class of f -Csiszár divergences (e.g., the Kullback-Leibler divergence, total variation distance) and the class of integral probability metrics (e.g., Wasserstein distance, total variation distance, maximum mean discrepancy).

Considering the probability measures \mathbb{P}_{X_j} and \mathbb{P}_Y (associated with the random variables X_j and Y) and a dissimilarity measure $\Delta(\cdot, \cdot)$, one can define two formulations for GSA:

- directly using a dissimilarity measure to assess $\Delta(\mathbb{P}_Y, \mathbb{P}_{Y|X_j})$;
- building a *dependence measures* evaluating $\Delta(\mathbb{P}_{(X_j, Y)}, \mathbb{P}_{X_j} \otimes \mathbb{P}_Y)$.

The first approach was studied in association with f -divergences in Da Veiga (2015); Rahman (2016). However, some f -divergences introduce estimation issues, and the resulting importance measures do not propose a functional decomposition of variance (also called FANOVA). Using kernel-based integral probability metrics such as the maximum mean discrepancy (MMD), an alternative importance measure was proposed. The following section presents the *Hilbert-Schmidt Independence Criterion* (HSIC), which was initially introduced by Gretton et al. (2006) for dependence testing, and later adapted as a dependence measure in GSA by Da Veiga (2015).

Hilbert-Schmidt independence criterion

Let us first recall the definition of the maximum mean discrepancy (further discussed in Appendix D). This distance between two probability distributions π and ζ can be defined as the worst-case error for any function within a unit ball of a function space \mathcal{H} :

$$\text{MMD}(\pi, \zeta) := \sup_{\|g\|_{\mathcal{H}(k)} \leq 1} \left| \int_{\mathcal{D}_X} g(\mathbf{x}) d\pi(\mathbf{x}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\zeta(\mathbf{x}) \right| \quad (1.80)$$

This quantity is a distance in the Reproducing kernel Hilbert space (RKHS) by taking a characteristic kernel (e.g., the Gaussian or Matérn kernel). After a calculation developed in Appendix D, an unbiased one-sample estimator of the squared-MMD was proposed by Gretton et al. (2006), with a convergence rate of $O(n^{-1/2})$ in probability. Considering the two-samples $\{\pi^{(i)}\}_{i=1}^n \sim \pi$ and $\{\zeta^{(j)}\}_{j=1}^n \sim \zeta$:

$$\widehat{\text{MMD}}^2(\pi, \zeta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n k(\pi^{(i)}, \pi^{(j)}) - k(\pi^{(i)}, \zeta^{(j)}) - k(\zeta^{(i)}, \pi^{(j)}) + k(\zeta^{(i)}, \zeta^{(j)}), \quad (1.81)$$

In the context of GSA, the first option is to directly use this dissimilarity measure to define the unnormalized index:

$$S_j^{\text{MMD}} = \text{MMD}(\mathbb{P}_Y, \mathbb{P}_{Y|X_j}). \quad (1.82)$$

Da Veiga (2021) remarked that the unnormalized first order Sobol' indices are recovered by taking the linear kernel on the output $k_Y(y, y') = yy'$. Using this non-characteristic kernel (see the definition in Appendix D) brings us back to a moment-dependent importance measure.

Alternatively, the second option considers a couple of random variables (X_j, Y) , with probability distributions \mathbb{P}_{X_j} and \mathbb{P}_Y , and assumes the RKHS \mathcal{H} induced by the tensor product kernel $k((x_j, y), (x'_j, y')) = k_{X_j}(x_j, x'_j)k_Y(y_j, y'_j)$. The *Hilbert-Schmidt independence criterion* (HSIC) measures the dependence between \mathbb{P}_{X_j} and \mathbb{P}_Y by expressing the MMD between $\mathbb{P}_{(X_j, Y)}$ and

$\mathbb{P}_{X_j} \otimes \mathbb{P}_Y$:

$$\text{HSIC}(X_j, Y) = \text{MMD}^2(\mathbb{P}_{(X_j, Y)}, \mathbb{P}_{X_j} \otimes \mathbb{P}_Y). \quad (1.83)$$

This technique showed very good results for screening, and corresponding independence tests were studied for screening in [De Lozzo and Marrel \(2016\)](#).

[Da Veiga \(2021\)](#) proposed the functional decomposition of the two indices defined in Eq. (1.82) and Eq. (1.83), allowing to develop their respective normalized versions. Note that the HSIC decomposition requires a specific hypothesis on the structure of the kernel associated with the inputs.

1.6.4 Summary and discussion

This section introduced the GSA methods commonly used in uncertainty quantification. Either to reduce the dimension of a problem (screening) or to quantify the influence of inputs (with importance measures), GSA improves the understanding of an uncertainty quantification study. As for other steps of the generic UQ methodology, SA is made more complicated for computationally costly simulation models, hence the use of surrogate models. Additionally, the dependence between inputs still represents an important limit to interpreting GSA results.

Alongside rare event estimation, some literature is dedicated to the influence of random inputs on such tail statistics. The sensitivity is no longer qualified as “global” but becomes “goal-oriented”. In the field of structural reliability, an overview of the reliability-oriented sensitivity analysis methods is presented in [Chabridon \(2018\)](#). Several techniques derive from rare event estimation (e.g., the FORM importance factors [Papaioannou and Straub \(2021\)](#)), or were adapted from GSA, like Sobol’ indices ([Ehre et al., 2020](#)), Target-HSIC ([Marrel and Chabridon, 2021](#)), or Shapley effects ([Demange-Chryst et al., 2023](#)).

Finally, sensitivity analysis may describe the effects of random inputs on the variation of the output, however, this study is done by assuming a model on the input uncertainties. The role of a regulatory agency auditing an uncertainty quantification approach for certification (i.e., a nuclear safety authority), might be to challenge the way to model the uncertainties on the inputs. In this case, various tools for *robustness analysis* exist to quantify the impact of mispecifying the random inputs on the quantity of interest studied. Among the methods to perturbate uncertainty models, some remain in the probabilistic framework, such as the “perturbed-law based indices” (PLI) ([Lemaître et al., 2015; Iooss et al., 2022](#)), or on extra-probabilistic methods ([Ajenjo et al., 2022](#)).

1.7 Surrogate modeling

1.7.1 Common framework

The aim of *surrogate modeling* (or metamodeling) is to build a cheap-to-call statistical model, denoted by $\widehat{g}_n(\cdot)$, replacing a costly numerical model $g(\cdot)$ over the input domain \mathcal{D}_X . To do

so, statistical learning is performed on a finite number of observations of the costly function g . When manipulating computationally expensive simulations, its size can be limited (i.e., small-data context). This n -sized set is usually called *learning set* written:

$$\{\mathbf{X}_n, \mathbf{y}_n\} = \left\{ \mathbf{x}^{(i)}, y^{(i)} \right\}_{i=1}^n = \left\{ \mathbf{x}^{(i)}, g(\mathbf{x}^{(i)}) \right\}_{i=1}^n. \quad (1.84)$$

A very large catalog of regression methods exists, here is a list of the most encountered ones in the field of UQ: generalized linear regression, polynomial chaos expansion (PCE) (Soize and Ghanem, 2004; Blatman and Sudret, 2011), support vector machine (SVM) (Cortes and Vapnik, 1995), Gaussian processes (GP) (Rasmussen and Williams, 2006), low-rank tensor approximations (Grasedyck et al., 2013), and artificial neural network (ANN) (Hastie et al., 2009). The following section will provide a short focus on Gaussian process regression.

Validating the accuracy and precision of a surrogate model is an important step to guarantee its fidelity with regard to the numerical model. When an m -sized input-output set is dedicated to validating the surrogate model, independently of the learning set, it is called *test set* and denoted by $\{\mathbf{X}_m, \mathbf{y}_m\} = \left\{ \mathbf{x}^{(i)}, g(\mathbf{x}^{(i)}) \right\}_{i=1}^m$. Note that the analyst may work in two different frameworks, affecting the regression and validation method's choice:

- Given-data context: only using a fixed input-output dataset to build and validate the surrogate model.
- Computer experiment context: allowing to generate simulated data points (often at a certain cost).

Validating surrogate models in a small-data context appears to be an important challenge. Different validation criteria and techniques exist. The *coefficient of validation*, denoted by R^2 , is the first validation metric that can be directly computed on the learning set:

$$R^2(\hat{g}_n) = 1 - \frac{\sum_{i=1}^n (y(\mathbf{x}^{(i)}) - \hat{g}(\mathbf{x}^{(i)}))^2}{\sum_{i=1}^n (y(\mathbf{x}^{(i)}) - \bar{y}_n)^2}, \quad (1.85)$$

where $\bar{y}_n = (1/n) \sum_{i=1}^n y^{(i)}$ denotes the empirical mean of the observations in the test sample. However, such metrics are not relevant for every regression method (typically, the interpolant methods have an $R^2 = 1$). The *predictivity coefficient* is an alternative defined as a normalized *integrated square error* (ISE):

$$Q^2(\hat{g}_n) = 1 - \frac{\text{ISE}(\hat{g}_n)}{\text{Var}(g(\mathbf{X}))}, \quad (1.86)$$

where

$$\text{ISE}(\hat{g}_n) = \int_{\mathcal{D}_X} (g(\mathbf{x}) - \hat{g}(\mathbf{x}))^2 d\mathbf{x}, \quad \text{Var}(g(\mathbf{X})) = \int_{\mathcal{D}_X} \left(g(\mathbf{x}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\mathbf{x}' \right)^2 d\mathbf{x}. \quad (1.87)$$

This quantity can be estimated on a test set $\{\mathbf{X}_m, \mathbf{y}_m\}$:

$$\widehat{Q}^2(\widehat{g}_n) = 1 - \frac{\sum_{i=1}^m (y(\mathbf{x}^{(i)}) - \widehat{g}(\mathbf{x}^{(i)}))^2}{\sum_{i=1}^m (y(\mathbf{x}^{(i)}) - \bar{y}_m)^2}. \quad (1.88)$$

Note that for either criterion, the higher the value, the better the quality of the fit.

Validating a surrogate model with an independent test set is sometimes called *holdout* validation. In a small-data context, dedicating an independent test set to validation might be impossible. Then, *cross-validation* is a generic estimation strategy allowing one to learn and test on the same sample. The most common cross-validation method is the *k-fold* validation, illustrated in Fig. 1.12. The idea is first to split the n -sized dataset into several equal parts, called folds. A first surrogate can be fitted on the entire datasets but the first fold, on which a validation criterion is estimated (i.e., performance metric). The operation is repeated for each fold, providing a virtual validation of the entire dataset. Leave-One-Out validation (LOO) is an extreme case of *k*-fold cross-validation, for which $k = n - 1$. Note that multiple variations of these methods exist, for example by adding a permutation or shuffling step. The “bagging” validation method (for “bootstrap aggregating”) consists of a shuffled cross-validation repeated many times (Breiman, 1996).

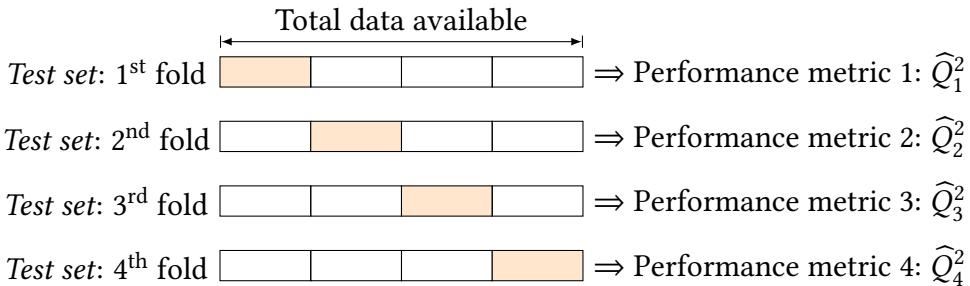


Figure 1.12 Illustration of a k -fold cross-validation (with $k = 4$)

1.7.2 General purposes surrogate model

In this section, a particular focus is dedicated to Gaussian process (GP) regression which will be useful in the following chapters (GP is also called kriging after the geostatistician D.G. Krige). Gaussian processes are a widely used regression method in UQ for their performance, flexibility and their associated confidence model. In a small-data context, the way of placing the few points forming the surrogate’s learning set is critical. Intuitively, to build a versatile surrogate model, the learning set should collect information over the entire domain uniformly. This is why space-filling designs of experiments are commonly used to build learning sets. In practice, QMC and optimized LHS design introduced in Section 1.4 are widely used.

Gaussian process regression

Considering a learning set \mathbf{X}_n , the goal is to approximate the function $g(\cdot)$ by a scalar Gaussian process conditioned on a set of observations $\mathbf{y}_n = \left\{ g(\mathbf{x}^{(i)}) \right\}_{i=1}^n$. Let us first define a prior structure G on the function approximating $g(\cdot)$, taken as a Gaussian process with a mean function $m(\cdot)$ and covariance function $k(\cdot, \cdot)$:

$$G \sim \text{GP}(m(\cdot), k(\cdot, \cdot)), \quad (1.89)$$

with a:

- *trend model*: $m(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$, composed of a functional basis $\mathbf{f} = (f_1, \dots, f_d)^\top$ and a vector of coefficients $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top$,
- *covariance model*: $k(\mathbf{x}, \mathbf{x}')$, usually taken stationary, such that $k(\mathbf{x}, \mathbf{x}') = \sigma^2 k_s(\mathbf{x} - \mathbf{x}', \boldsymbol{\theta})$ with $\sigma^2 > 0$ and $\boldsymbol{\theta} \in \mathbb{R}_+^d$.

The trend model of a GP defines its general tendency, while the covariance model influences its regularity. Gaussian process regression takes different names depending on the knowledge of the trend model. It is called “simple kriging” when the trend is fully known, “ordinary kriging” when the trend is unknown but supposed constant and “universal kriging” otherwise. Note that [Schobi et al. \(2015\)](#) introduced a hybrid method named PC-Kriging setting a PCE as the trend of a kriging model.

To ease the presentation, let us first consider the hyperparameters $\sigma, \boldsymbol{\theta}$ fully known and a zero trend $\boldsymbol{\beta} = \mathbf{0}$. At a given point $\mathbf{x} \in \mathcal{D}_X$ the realization of the GP is a Gaussian random variable $G(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}))$. Working with Gaussian variables allows us to easily write conditioning formulas between $G(\mathbf{x})$ and the observations \mathbf{y}_n . This Gaussian variable $G(\mathbf{x})$ conditioned on the observations \mathbf{y}_n is sometimes called conditional posterior $G_n(\mathbf{x}) := (G(\mathbf{x}) | \mathbf{y}_n) \sim \mathcal{N}(\eta_n(\mathbf{x}), s_n^2(\mathbf{x}))$. The well-known “Kriging equations” (see e.g., [Rasmussen and Williams 2006](#)) offer its explicit expression:

$$\begin{cases} \eta_n(\mathbf{x}) &:= \mathbf{k}^\top(\mathbf{x}) \mathbf{K}^{-1} \mathbf{y}_n \\ s_n^2(\mathbf{x}) &:= k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) \end{cases} \quad (1.90)$$

where $\mathbf{k}(\mathbf{x})$ is the column vector of the covariance kernel evaluations $[k(\mathbf{x}, \mathbf{x}^{(1)}), \dots, k(\mathbf{x}, \mathbf{x}^{(n)})]$ and \mathbf{K} is the $(n \times n)$ variance-covariance matrix such that the (i, j) -element is $\{\mathbf{K}\}_{i,j} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$.

In practice, the surrogate model is defined by the *predictor* function $\eta_n(\cdot)$. This regression model provides important complementary information with the *kriging variance* $s_n^2(\mathbf{x})$, reaching zero at the learning points. Let us remark that the kriging variance fully depends on the covariance model (defined by its parametric structure and hyperparameters). In practice, the hyperparameters are unknown, therefore, their estimation is a key step in the construction of a kriging model. This estimation can be done using different approaches, most commonly using maximum likelihood estimation or cross-validation.

The illustration in Fig. 1.13 is a typical one-dimensional representation of an ordinary kriging model. The mean of the conditioned process is plotted in red while its variability is represented

by the many trajectories drawn on the process. In the simplest framework, the kriging model exactly interpolates the observations (black crosses).

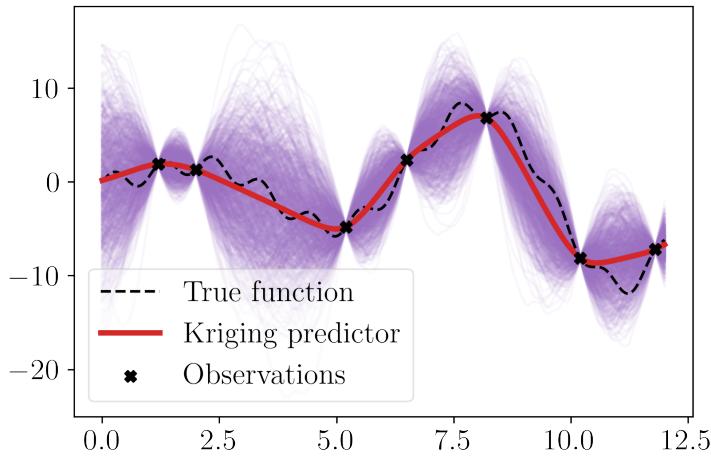


Figure 1.13 Illustration of an ordinary kriging model fitted on a limited set of observations ($n = 7$). The predictor is represented in and several trajectories of the conditioned Gaussian process are drawn and represented in purple.

Associated with kriging models, another validation criterion is relevant to evaluate the kriging variance $s_n^2(\mathbf{x})$. The predictive variance adequation (PVA) has been introduced to confirm that the kriging variance is reliable (see e.g., [Bachoc 2013](#)). For a validation performed by holdout, and using an independent m -sized test set, the PVA is defined as:

$$\text{PVA} = \left| \log \left(\frac{1}{m} \sum_{i=1}^n \frac{(y(\mathbf{x}^{(i)}) - \hat{g}(\mathbf{x}^{(i)}))^2}{s_n^2(\mathbf{x}^{(i)})} \right) \right|. \quad (1.91)$$

The smaller this quantity gets, the better the quality of the kriging variance.

Gaussian process regression is an elegant solution, offering a lot of flexibility and an associated error model (i.e., the kriging variance). However, well-known numerical issues appear during the estimation of the hyperparameters, especially as the learning size increases. More specifically, the computation and memory allocation for the variance-covariance matrix is a recurrent issue. Multiple techniques solve this issue by applying compression schemes on this matrix, e.g., based on sparse approximations (e.g., Hierarchical Matrices [Geoga et al. 2020](#)).

This section introduced a general-purpose surrogate model, uniformly approximating a function on a domain, however, surrogates are often used for specific purposes (e.g., contour finding for reliability analysis).

OpenTURNS 6 (Gaussian process regression). The Python code available in Appendix D proposes a minimalistic OpenTURNS example to fit an ordinary kriging model and active learning models. Figures illustrating the present section may be reproduced, using the OpenTURNS scripts available on GitHub^{13,14}.

1.7.3 Goal-oriented active surrogate model

Surrogates are often fitted for a specific purpose, requiring an accurate approximation over a limited subdomain only. In these cases, a more efficient approach might be to circumscribe the learning to this subdomain (i.e., *goal-oriented learning*), rather than uniformly over the entire domain. For example, to fit a surrogate model for contour finding in reliability analysis, one should concentrate the learning set around the limit-state function. Similarly, to build a surrogate for a global optimization problem, one should focus the learning set around the optimum(s). Unfortunately, the area(s) of interest is usually unknown before evaluating the true function. *Active learning* is a general concept, aiming at iteratively increasing the learning set w.r.t. a *learning criterion* (also called “acquisition function”) depending on the surrogate’s goal to enhance the surrogate in the area(s) of interest. Note that the expression “adaptive design” is also widely used in the computer experiments community to designate this concept. An exploration-exploitation trade-off arises in active learning, mostly sorted by the learning criterion.

Remark 1. This section introduces active learning methods in the computer experiment context, where the true function can be evaluated anywhere for a given computational cost. However, the “active learning” term is also used to handle big data frameworks in the machine learning community (Qiu et al., 2016). When datasets become so large that learning methods do not scale in practice, the analyst needs to select a relevant subset on which the learning is performed.

Active kriging for optimization

In the field of black-box optimization, many methods rely on approximating the function by a surrogate. The use of Gaussian processes as probabilistic surrogates for optimization was popularized by the *efficient global optimization* (EGO) algorithm (Jones et al., 1998). Ever since, many related methods were developed under the generic name of *Bayesian optimization*. The main idea is to exploit the uncertainty model from the GP to direct the point selection. Factually, the learning criterion depends on the Gaussian process variance model. Numerous reviews of this field were proposed by Shahriari et al. (2015); Gramacy (2020) and numerical benchmarks presented in Le Riche and Picheny (2021).

The generic black-box optimization problem tackled is defined as:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{D}_X} g(\mathbf{x}). \quad (1.92)$$

To illustrate Bayesian optimization, let us present the EGO algorithm, defined by its specific learning criterion: the “expected improvement”. Considering an initial learning set $\{\mathbf{X}_n, \mathbf{y}_n\}$ built on a space-filling input design \mathbf{X}_n to explore the domain. A first surrogate $G_n(\mathbf{x}) \sim \mathcal{N}(\eta_n(\mathbf{x}), s_n^2(\mathbf{x}))$

¹³https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/surrogates.ipynb

¹⁴https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/active_learning.ipynb

is fitted using Eq. (1.90). The expected improvement, to be maximized, is then written as:

$$\mathcal{A}^{\text{EI}}(\mathbf{x}; \mathbf{y}_n) = \mathbb{E}[\max(g_{\min} - G_n(\mathbf{x}))] \quad (1.93)$$

$$= (g_{\min} - \eta_n(\mathbf{x})) \Phi\left(\frac{g_{\min} - \eta_n(\mathbf{x})}{s_n(\mathbf{x})}\right) + s_n(\mathbf{x}) \phi\left(\frac{g_{\min} - \eta_n(\mathbf{x})}{s_n(\mathbf{x})}\right), \quad (1.94)$$

where $g_{\min} = \min(\mathbf{y}_n)$, ϕ and Φ respectively stand for the PDF and the CDF of the standard Gaussian distribution. This learning criterion is relatively inexpensive and allows a progressive enhancement of the Gaussian process to solve the optimization problem with a limited number of calls to the true function.

Three iterations of the EGO algorithm are represented in Fig. 1.14 to minimize a function (dashed line), knowing a few observations (black crosses). After fitting an initial kriging model (in red), the corresponding expected improvement function is represented underneath it (green line). This learning criterion determines the location of the observation to be added to the learning set to enhance the surrogate w.r.t. to the optimization problem.

Bayesian optimization is an active research field, with different open problems such as constrained Bayesian optimization (Petit, 2022), or Bayesian optimization on stochastic functions (Gramacy, 2020). Similarly, active learning was also adapted for structural reliability problems.

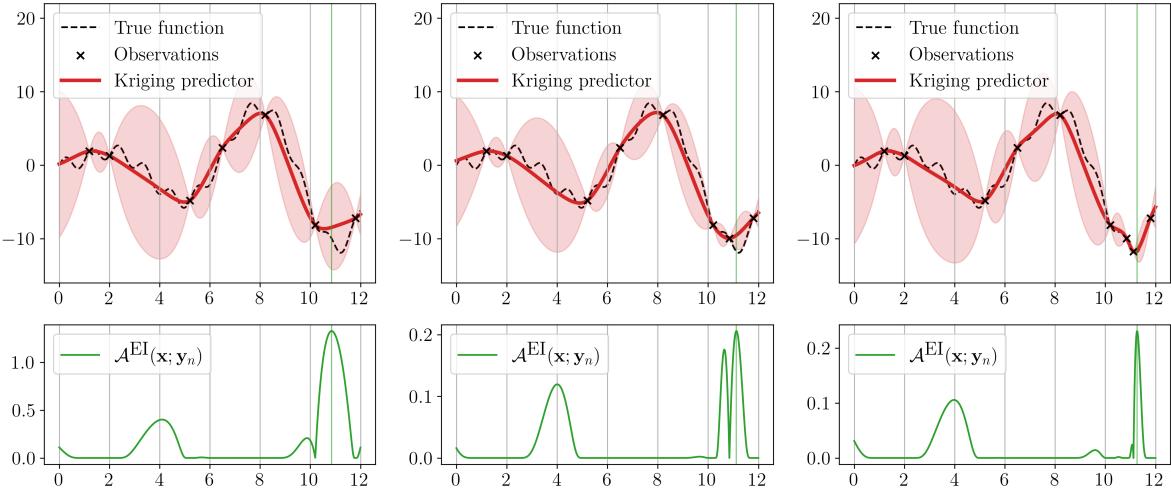


Figure 1.14 Illustration of the expected improvement learning criterion

Active kriging for reliability analysis

Rare event estimation often requires large amounts of evaluations of the limit-state function (becoming intractable for costly numerical models). Emulating this function by a surrogate model can drastically limit the number of calls to the LSF. This surrogate approximates the contour (i.e., border) of the failure domain. However, in most cases, the failure domain represents a very restricted area of the input domain. Active learning methods were proposed to iteratively concentrate the learning set around this border.

For rare event estimation, the surrogate only needs to be accurate near the limit state function. In other words, it should accurately discriminate the points leading to the safe domain from those leading to the failure domain. In fact, this problem can be seen as a binary classification. For example, active learning procedures using SVM classifiers were adapted to this specific goal (Bourinet, 2018).

The following paragraph introduces the most popular kriging-based learning criterion: the “deviation number” U (Echard et al., 2011). The reader may refer to Morio and Balesdent (2015) for further active learning techniques dedicated to rare event estimation. More recently, Teixeira et al. (2021) and Moustapha et al. (2022) reviewed this topic with the presentation of wide numerical benchmarks.

Considering an initial learning set $\{\mathbf{X}_n, \mathbf{y}_n\}$ built on a space-filling input design \mathbf{X}_n to explore the domain. A first Gaussian process $G_n(\mathbf{x}) \sim \mathcal{N}(\eta_n(\mathbf{x}), s_n^2(\mathbf{x}))$ is fitted using Eq. (1.90). The deviation number U is looking for points close to the limit-state function while presenting a high kriging variance. This criterion to be minimized is defined as:

$$\mathcal{A}^U(\mathbf{x}; \mathbf{y}_n) = \frac{|y_{\text{th}} - \eta_n(\mathbf{x})|}{s_n^2(\mathbf{x})}, \quad (1.95)$$

where $y_{\text{th}} \in \mathbb{R}$ is a threshold defining the failure domain.

Fig. 1.15 reuses the same one-dimensional function as in Fig. 1.14 to create a rare event problem. In this case, the failure domain is defined for output values below the threshold y_{th} . Once again, three iterations of the active kriging (AK) algorithm are illustrated, with the corresponding learning criterion U (to minimize). In this simple case, the LSF is defined by the two intersections of the function with the threshold. Therefore, the AK method selects points near these intersections.

Unlike optimization problems, the surrogate is used for uncertainty propagation, meaning that the rare event estimation is the result of the approximation of the LSF (i.e., contour finding) and a sampling techniques. AK methods were coupled with most sampling techniques introduced in Section 1.5 (e.g., AK-MCS, AK-IS, AK-SS, etc.). As an agnostic strategy, Moustapha et al. (2022) recommend starting by applying an AK method (using the learning function U) paired with a subset simulation (taking an intermediary probability $p_0 = 0.2$).

The AK methods present the advantages of being easily implemented and interpreted, however, their learning criterion relies on a local approach. Alternatively, the *stepwise uncertainty reduction* (SUR) chooses iterative points by reducing the future expected uncertainty related to the quantity of interest (Bect et al., 2012). If this method was proven to be theoretically more consistent (Bect et al., 2019), its scaling ability is still a bottleneck.

1.7.4 Summary and discussion

This section brought attention to surrogate modeling in the context of computer experiments. Statistical learning in this framework is made specific by the capacity of the analyst to choose the repartition of the learning set and the small data constraint (mostly due to the costly numerical

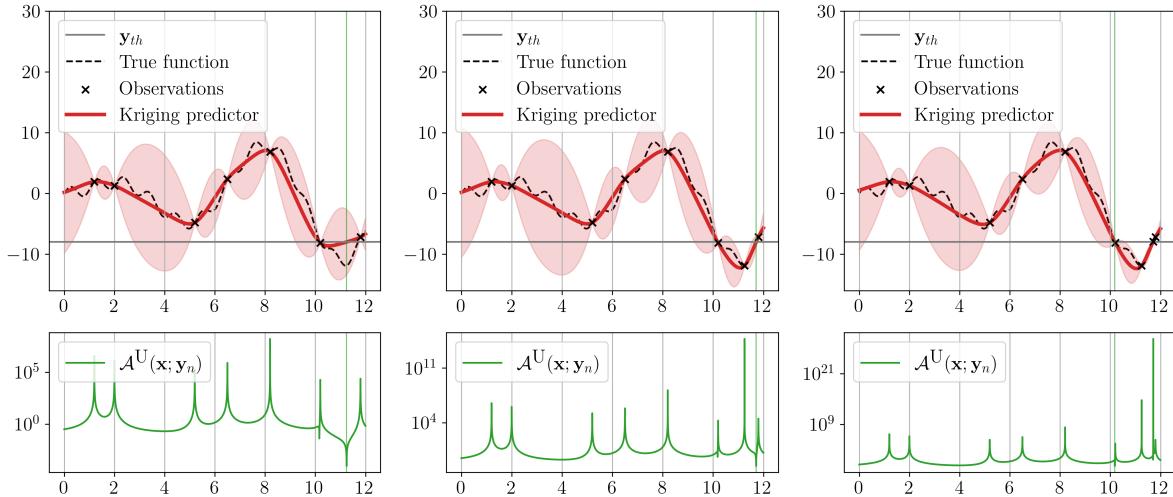


Figure 1.15 Illustration of the deviation number learning criterion

models manipulated). In this context, many methods are used, however, Gaussian processes became popular in UQ as they consider a prior structure of uncertainty that is conditioned by observations (at the edge between a Bayesian and a frequentist approach). To enhance the learning for specific purposes (e.g., optimization or contour finding), active learning methods iteratively add learning points in the subdomain of interest. For some applications, the system studied might be modeled for different fidelities (each presenting different computational costs). Multi-fidelity surrogate modeling is an active field of research, associating observations from different fidelities to improve the learning (Fernández-Godino et al., 2016). Such methods are relevant for models with a very high computational cost (typically in computer fluid dynamics).

In UQ, surrogate models are used for uncertainty propagation (step C) and inverse analysis (step C'). Surrogate modeling is made difficult when the functions present discontinuities (or strong nonlinearities), high dimension, stochasticity, or nonscalar inputs or outputs. To deal with high dimensional problems, unimportant inputs can be screened using sensitivity analysis (see Section 1.6.1), otherwise, model order reduction methods might be used (Schilders et al., 2008). When the function is stochastic, different approaches allow fitting the function and its intrinsic variability (Binois et al., 2019; Baker et al., 2022; Zhu, 2022).

Provided a strict validation process, surrogate models are a great opportunity for uncertainty quantification. However, many regulatory authorities are still reluctant to use surrogates, although their error is often much smaller than the modeling error (i.e., the error between the actual physical behavior and its numerical modeling).

1.8 Conclusion

This section gives a literature overview of the main steps in uncertainty quantification. From uncertainty modeling, uncertainty propagation, and sensitivity analysis to surrogate modeling. To ease the methodological presentation, all the illustrations from this section are reproducible using the Python/OpenTURNS scripts available on the GitHub repository mentioned earlier.

Finally in this work, the numerical models exploited are supposed to be accurate, but they obviously carry some modeling uncertainty (Oberkampf and Roy, 2010). In fact, prior to uncertainty quantification, the model should be calibrated to make it match some physical information (e.g., measurements). Numerical model calibration is also called data assimilation when a stream of measured data is available.

The aim of this work is to apply the tools presented in this chapter to offshore wind turbine models, therefore the next chapter introduces the numerical models manipulated in this thesis.

PART II:

CONTRIBUTIONS TO UNCERTAINTY QUANTIFICATION AND PROPAGATION

*Le doute est un état mental désagréable,
mais la certitude est ridicule.*

VOLTAIRE

PART III:

CONTRIBUTIONS TO RARE EVENT ESTIMATION

La résignation est un suicide quotidien.

H. BALZAC

Chapter **2**

Adaptive rare event estimation using Bernstein copula

2.1	Introduction	68
2.2	Bernstein adaptive nonparametric conditional sampling	69
2.3	Numerical experiments	72
2.3.1	Analytical toy-cases	72
2.3.2	Benchmark results and analysis	74
2.4	Reliability-oriented sensitivity analysis	77
2.4.1	Target and conditional HSIC	77
2.4.2	ROSA as a post-treatment of BANCS reliability analysis	79
2.5	Conclusion	79

2.1 Introduction

Assessing the reliability of systems' such as offshore wind turbines, often involves the estimation of rare event probabilities. In the reliability analysis framework introduced in Section 1.5, the performance of a system is typically modeled by a deterministic scalar function, denoted by $g : \mathcal{D}_x \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$, and referred to as the *limit-state function*. A critical threshold on the system's output, denoted as $y_{\text{th}} \in \mathbb{R}$, then defines the *failure domain*, expressed as $\mathcal{F}_x := \{\mathbf{x} \in \mathcal{D}_x | g(\mathbf{x}) \leq y_{\text{th}}\}$. Considering a probabilistic framework, the input uncertainties are modeled by a continuous random vector $\mathbf{X} \in \mathcal{D}_x$, with a joint probability density function (PDF) denoted by f_X . In this scenario, uncertainty propagation consists in composing the random vector \mathbf{X} with the function g to obtain the output variable of interest $Y = g(\mathbf{X}) \in \mathbb{R}$. Then, a common risk measure in reliability analysis is the *failure probability*, denoted by p_f , representing the probability of the system exceeding the threshold y_{th} :

$$p_f := \mathbb{P}(g(\mathbf{X}) \leq y_{\text{th}}) = \int_{\mathcal{D}_x} \mathbb{1}_{\mathcal{F}_x}(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x}. \quad (2.1)$$

The usual convention presented in Section 1.5 allows to modify the limit-state function in order to set the threshold to zero.

The main methods for rare event estimation (see Section 1.5) can be divided into two groups (Morio and Balesdent, 2015): (i) geometric approaches, such as the *first-/second-order reliability method* (FORM/SORM) whose aim is to approximate the limit-state function by a first-/second-order Taylor expansion at the most probable failure point; (ii) simulation-based techniques such as the *crude Monte Carlo* method. Unfortunately, FORM/SORM methods do not provide a lot of statistical information as they are purely geometric approaches. Meanwhile, estimating a rare event probability by crude Monte Carlo becomes rapidly intractable for engineering applications. To overcome this limit, advanced simulation techniques have been developed: among others, one can mention several "variance reduction" methods such as the non-adaptive and adaptive versions of the *Importance Sampling* (Rubinstein and Kroese, 2008) and splitting techniques (Cérou et al., 2012) such as the *Subset Simulation* (SS) (Au and Beck, 2001).

In subset simulation, the idea is to write the rare event p_f as a product of larger conditional probabilities, each one of them being easier to estimate. To generate intermediary conditional samples, this method uses Markov chain Monte Carlo (MCMC) sampling, which presents numerous versions (Papaioannou et al., 2015). However, MCMC algorithms are known to be highly tunable algorithms which produce non-i.i.d. samples, which consequently, cannot always be used for direct statistical estimation (e.g., failure probability or sensitivity indices Da Veiga et al., 2021).

Adaptive importance sampling infers conditional distributions before using an importance sampling estimator of the failure probability. The set of auxiliary distributions converging towards the failure domain is either fitted by parametric approaches (e.g., using the cross-entropy method Rubinstein and Kroese 2004), or nonparametric methods (e.g., using multivariate KDE

(Morio 2011). The main drawback of the cross-entropy adaptive importance sampling (AIS-CE) is its limited flexibility, not allowing it to perform well in the case of multimodal failure domains. On the other hand, nonparametric adaptive importance sampling (NAIS) inherits its drawbacks from the multivariate KDE (i.e., significant performance drop for medium to high dimension).

The present work proposes a new rare event estimation method, “Bernstein adaptive nonparametric conditional sampling” (BANCS), adopting the same adaptive importance sampling structure as AIS-CE or NAIS while using a different mechanism to fit conditional distributions. This algorithm decomposes the fit of the intermediary conditional distributions into: a fit of their marginals by univariate KDE, and a fit of their copula with the *Empirical Bernstein Copula* (introduced in Section ??). Compared to direct multivariate KDE in NAIS, this decomposition significantly simplifies the inference in medium to high dimension. Additionally, the method proposed generates i.i.d. samples of the intermediary conditional distributions, unlike SS.

In practice, such i.i.d. samples may also be used to estimate dedicated reliability-oriented sensitivity indices (see, e.g., Chabridon et al., 2021; Marrel and Chabridon, 2021). The present chapter introduces kernel-based reliability-oriented sensitivity indices and their direct estimation as a post-processing of the BANCS algortihm.

In this chapter, Section [xx] will introduce the BANCS algorithm for rare event estimation. Then, Section [xx] will apply this method to [three] toy-cases and analyze the results with respect to NAIS and SS performances. Section [xx] will present a kernel-based reliability-oriented sensitivity index and illustrate their estimation on samples generated by the BANCS algorithm. Then, the last section present some conclusions and research perspectives.

Remark 2. This contribution was initially presented in Fekhari et al. (2023a), using the failure probability estimator proposed in the subset simulation algorithm (see Eq. (1.63)). Switching to an adaptive importance sampling estimator (see e.g., Eq. (1.61)) notably improved the method.

2.2 Bernstein adaptive nonparametric conditional sampling

This section presents a new method for rare event estimation named Bernstein adaptive nonparametric conditional sampling (BANCS) and illustrates its mechanism on a simple two-dimensional case.

The BANCS algorithm uses the same structure as other adaptive importance sampling methods (e.g., NAIS or AIS-CE) while employing a different approach to fit the intermediate conditional distributions. As described in Algorithm 1, at iteration k , after estimating the intermediary quantile $\hat{q}_{[k]}^{p_0}$, a nonparametric model is fitted on the set $A_{[k+1]}$ of all samples leading to values below $\hat{q}_{[k]}^{p_0}$. This inference is done by coupling a set of marginals each fitted by KDE, with a copula fitted by EBC. The generation of the next i.i.d. N -sized sample $X_{[k+1],N}$ on the conditional distributions is straightforward and does not require MCMC sampling like in SS. Note that the BANCS method does not require iso-probabilistic transform.

Nonparametric inference implies tuning some parameters, either a scaling parameter for the KDE or a polynomial order m (considered equal for all dimension) for the EBC. In BANCS

Algorithm 1 Bernstein adaptive nonparametric conditional sampling (BANCS).

1: » Inputs:

2: f_X , joint PDF of the inputs
 3: $g(\cdot)$, limit-state function
 4: $y_{\text{th}} \in \mathbb{R}$, threshold defining the failure event
 5: N , number of samples per iteration
 6: $m \in \mathbb{N}$, parameter of the EBC fitting
 7: $p_0 \in]0, 1[$, empirical quantile order (rarity parameter)
 8: » Algorithm:

9: Set $k = 0$ and $f_{[0]} = f_X$
 10: Sample $\mathbf{X}_{[0],N} = \{\mathbf{x}_{[0]}^{(i)}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} f_{[0]}$
 11: Evaluate $Y_{[0],N} = \{g(\mathbf{x}_{[0]}^{(i)})\}_{i=1}^N$
 12: Estimate the empirical p_0 -quantile $\hat{q}_{[0]}^{p_0}$ of the set $Y_{[0],N}$
 13: **while** $\hat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**
 14: Compute IS weights $\{w_{[k]}^{(i)}\}_{i=1}^N = \left\{ \frac{f_X(\mathbf{x}_{[k]}^{(i)})}{\hat{f}_{[k]}(\mathbf{x}_{[k]}^{(i)})} \right\}_{j=1}^N$
 15: Build a weighted elite-set $\mathbf{A}_{[k+1]} = \sum_{l=1}^k \sum_{i=1}^N w_{[k]}^{(j)} \mathbb{1}_{\{g(\mathbf{x}_{[l]}^{(i)}) \leq \hat{q}_{[k]}\}} (\mathbf{x}_{[k]}^{(i)})$
 16: Fit marginals of the set $\mathbf{A}_{[k+1]}$ by KDE $\{\hat{F}_j\}_{j=1}^d$
 17: Fit the copula of the set $\mathbf{A}_{[k+1]}$ by EBC $B_m(C_n)$
 18: Build a CDF $\hat{F}_{[k+1]}(\mathbf{x}) = B_m(C_n)(\hat{F}_1(x_1), \dots, \hat{F}_d(x_d))$
 19: Sample $\mathbf{X}_{[k+1],N} = \{\mathbf{x}_{[k+1]}^{(i)}\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \hat{f}_{[k+1]}$
 20: Evaluate $Y_{[k+1],N} = \{g(\mathbf{x}_{[k+1]}^{(i)})\}_{i=1}^N$
 21: Estimate the empirical p_0 -quantile $\hat{q}_{[k+1]}^{p_0}$ of $Y_{[k+1],N}$
 22: Set $k = k + 1$
 23: **end while**
 24: Set $k_{\#} = k$
 25: Estimate $\hat{p}_f^{\text{BANCS}} = \frac{1}{N} \sum_{i=1}^N w_{[k_{\#}]}^{(i)} \mathbb{1}_{\{g(\mathbf{x}_{[k_{\#}]}^{(i)}) \leq y_{\text{th}}\}} (\mathbf{x}_{[k_{\#}]}^{(i)})$
 26: Estimate $\text{Var}(\hat{p}_f^{\text{BANCS}}) = \frac{1}{N-1} \left[\frac{1}{N} \sum_{i=1}^N (w_{[k_{\#}]}^{(i)})^2 \mathbb{1}_{\{g(\mathbf{x}_{[k_{\#}]}^{(i)}) \leq y_{\text{th}}\}} (\mathbf{x}_{[k_{\#}]}^{(i)}) - (\hat{p}_f^{\text{BANCS}})^2 \right]$
 27: » Outputs:
 28: \hat{p}_f^{BANCS} , estimate of p_f

algorithm, the KDE is tuned using Silverman's rule [add ref] while the EBC is tuned according to the results minimizing the asymptotic mean integrated squared error (AMISE). As discussed in Chapter ??, for a dataset with size n and dimension d , the AMISE tuning for the EBC defined by Sancetta and Satchell (2004) as:

$$m_{\text{AIMSE}} = 1 + n^{2/(d+4)} \quad (2.2)$$

In our experience, EBC tuning in Eq. (2.2) worked best for the BANCS algorithm and is systematically used in the following. This tuning sets a rather low polynomial order to the EBC, which avoids overfitting. For small sample sizes, e.g., $n < 100$, Segers et al. (2017) showed the limits of this tuning. However, the typical sample sizes used for rare event simulation should be appropriate for the AMISE tuning.

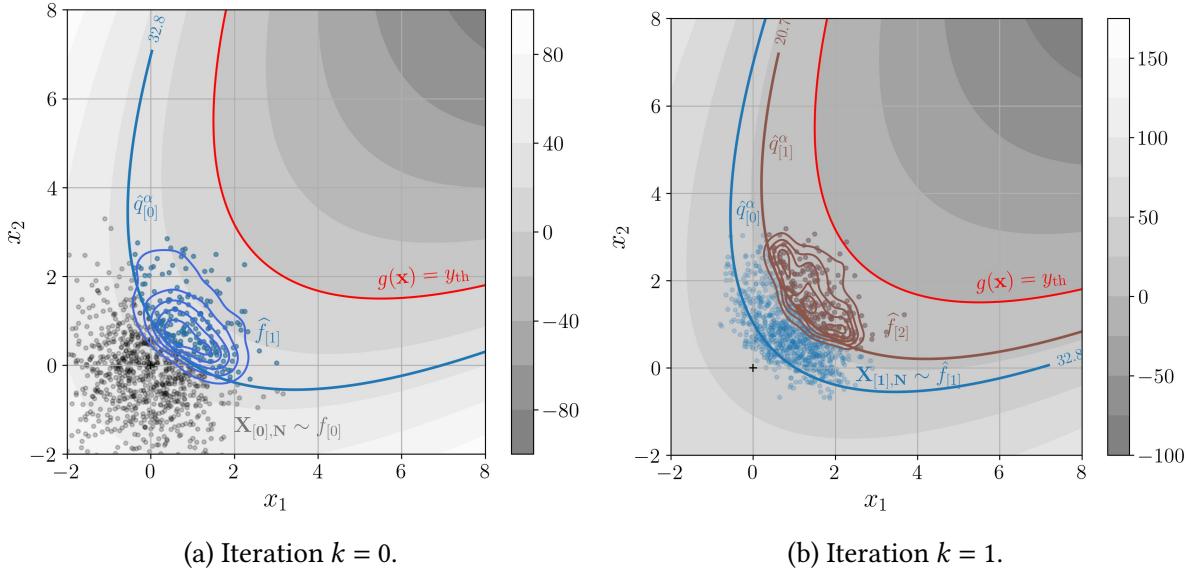


Figure 2.1 BANCS algorithm applied to toy-case #1: illustration of conditional sampling and nonparametric fit at the first and second iterations.

Ultimately, the estimator of the probability from Eq. (2.1) is written as a simple IS estimator on the last conditional distribution with PDF $\hat{f}_{[k_\#]}$:

$$\widehat{p}_f^{\text{BANCS}} = \frac{1}{N} \sum_{i=1}^N \frac{f_X(\mathbf{x}_{[k_\#]}^{(i)})}{\widehat{f}_{[k_\#]}(\mathbf{x}_{[k_\#]}^{(i)})} \mathbb{1}_{\{g(\mathbf{x}_{[k_\#]}^{(i)}) \leq y_{\text{th}}\}}(\mathbf{x}_{[k_\#]}^{(i)}), \quad \left\{ \mathbf{x}_{[k_\#]}^{(i)} \right\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \widehat{f}_{[k_\#]}. \quad (2.3)$$

This estimator also benefits from an IS variance, which can also be estimated, using the same sample as previously, by:

$$\text{Var}(\widehat{p}_f^{\text{BANCS}}) = \frac{1}{N-1} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{f_X(\mathbf{x}_{[k_\#]}^{(i)})}{\widehat{f}_{[k_\#]}(\mathbf{x}_{[k_\#]}^{(i)})} \right)^2 \mathbb{1}_{\{g(\mathbf{x}_{[k_\#]}^{(i)}) \leq y_{\text{th}}\}}(\mathbf{x}_{[k_\#]}^{(i)}) - \left(\widehat{p}_f^{\text{BANCS}} \right)^2 \right] \quad (2.4)$$

Fig. 2.1 illustrates the nonparametric fit and conditional sampling in BANCS method on a two-dimensional reliability problem (later introduced as “toy-case #1”). At the iteration $k = 0$, the conditional distribution fitted (with PDF represented by the blue isolines) slightly overlays below the quantile border $\widehat{q}_{[0]}^{p_0}$. Ideally, the conditional distribution fitted should be sharp around this border without overfitting the data. At the second iteration, the PDF of the second conditional distribution fitted is represented by the brown isolines. This second fit is realized on all the samples above the second quantile $\widehat{q}_{[1]}^{p_0}$, weighted by their respective distribution (see lines 14 and 15 from Algorithm 1). Including the samples from the previous iterations tends to make the fit sharper around the quantile border.

2.3 Numerical experiments

In the present section, the performances of BANCS algorithm are compared with the ones from subset simulation (SS) and the nonparametric adaptive importance sampling (NAIS). SS efficiency depends on the choice and tuning of the MCMC algorithm (Papaioannou et al., 2015). Our work uses the OpenTURNS implementations of the SS¹(integrating a component-wise Metropolis-Hastings algorithm), and the OpenTURNS implementations of NAIS². An implementation of the BANCS method and the following numerical experiments are available in a Git repository³.

In the following analytical numerical experiments, the intermediary probabilities were set to $p_0 = 0.1$ (following the recommendations from Au and Beck (2001)), allowing a fair comparison with subset simulation. The size N of each intermediate samples (i.e., subset) evolves in the following set $N \in \{300, 500, 700, 1000, 2000, 5000, 10000\}$, in order to get a reasonable number of points to perform the nonparametric fitting. Let us reming that the EBC tuning is setup to minimize the AMISE, such that $m = 1 + n^{\frac{2}{d+4}}$. In order to take into account the variability of the method's results, each experiment is repeated 100 times, allowing the computation of a coefficient of variation $\widehat{\delta} = \frac{\sigma_{\widehat{p}_f}}{\mu_{\widehat{p}_f}}$.

2.3.1 Analytical toy-cases

The reference values of the failure probabilities of each problem studied hereafter are obtained by Monte Carlo estimation on very large samples (typically with size $N = 10^9$).

Toy-case #1: Parabolic reliability problem. Let us define the parabolic reliability problem, considering the function $g_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$g_1(\mathbf{x}) = (x_1 - x_2)^2 - 8(x_1 + x_2 - 5), \quad (2.5)$$

with the input random vector $\mathbf{X} = (X_1, X_2)$ following a standard 2-dimensional normal distribution. The reliability problem consists in evaluating: $p_{f,1} = \mathbb{P}(g_1(\mathbf{X}) \leq 0) = 1.31 \times 10^{-4}$.

Toy-case #2: Four-branch reliability problem. Let us define the four-branch reliability problem (originally proposed by Waarts (2000)), considering the following function $g_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$g_2(\mathbf{x}) = \min \left(\begin{array}{l} 3 + 0.1(x_1 - x_2)^2 - \frac{(x_1 + x_2)}{\sqrt{2}} \\ 3 + 0.1(x_1 - x_2)^2 + \frac{(x_1 + x_2)}{\sqrt{2}} \\ (x_1 - x_2) + \frac{7}{\sqrt{2}} \\ (x_2 - x_1) + \frac{7}{\sqrt{2}} \end{array} \right), \quad (2.6)$$

¹SS: https://openturns.github.io/openturns/latest/user_manual/_generated/openturns.SubsetSampling.html

²NAIS: https://openturns.github.io/openturns/latest/user_manual/_generated/openturns.NAIS.html

³BANCS: <https://github.com/efekhari27/bancs>

with the input random vector $\mathbf{X} = (X_1, X_2)$ following a standard 2-dimensional normal distribution. The reliability problem consists in evaluating: $p_{f,2} = \mathbb{P}(g_2(\mathbf{X}) \leq 0) = 2.22 \times 10^{-3}$.

Toy-case #3: Modified Ishigami reliability problem. Let us define the modified Ishigami reliability problem (inspired by the manuscript of [cite thèse P.Lemaitre]), considering the following function $g_3 : \mathbb{R}^3 \rightarrow \mathbb{R}$:

$$g_3(\mathbf{x}) = \sin(x_1) + 7 \sin(x_2)^2 + \frac{x_3^4 \sin(x_1)}{10} - 10.5. \quad (2.7)$$

with the input random vector $\mathbf{X} = (X_1, X_2, X_3)$ following a standard 3-dimensional normal distribution. The reliability problem consists in evaluating: $p_{f,3} = \mathbb{P}(g_3(\mathbf{X}) \leq 0) = 1.94 \times 10^{-5}$.

Toy-case #4: Medium-dimensional reliability problem. Let us define the medium-dimensional reliability problem (proposed by Yun et al. (2018)), considering the following function $g_4 : \mathbb{R}^7 \rightarrow \mathbb{R}$:

$$g_4(\mathbf{x}) = 15.59 \times 10^4 - \frac{x_1 x_3^2}{2x_3^2} \frac{x_2^4 - 4x_5 x_6 x_7^2 + x_4(x_6 + 4x_5 + 2x_6 x_7)}{x_4 x_5 (x_4 + x_6 + 2x_6 x_7)}, \quad (2.8)$$

with the input random vector $\mathbf{X} = (X_1, \dots, X_7)$, following a product of normal distributions defined in Yun et al. (2018). The reliability problem consists in evaluating: $p_{f,4} = \mathbb{P}(g_4(\mathbf{X}) \leq 0) = 8.10 \times 10^{-3}$.

Toy-case #5: Medium-dimensional oscillator problem. Let us define the higher-dimensional oscillator reliability problem (adapted from [add ref]), considering the following function $g_5 : \mathbb{R}^8 \rightarrow \mathbb{R}$:

$$g_5(\mathbf{x}) = F_s - 3k_s \sqrt{\frac{\pi S_0}{4\zeta_s \omega_s^3}} \frac{\zeta_a \zeta_s}{\zeta_p \zeta_s (4\zeta_a^2 + \theta^2) + \gamma \zeta_a^2} \frac{\omega_p (\zeta_p \omega_p^3 + \zeta_s \omega_s^3)}{4\zeta_a \omega_a^4}, \quad (2.9)$$

where $\mathbf{x} = (m_p, m_s, k_p, k_s, \zeta_p, \zeta_s, F_s, S_0)$, $\omega_p = \sqrt{k_p/m_p}$, $\omega_s = \sqrt{k_s/m_s}$, $\omega_a = (\omega_p + \omega_s)/2$, $\zeta_a = (\zeta_p + \zeta_s)/2$, $\gamma = m_s/m_p$, $\theta = (\omega_p - \omega_s)/\omega_a$. The input random vector \mathbf{X} , defined as a product of marginals following the distributions defined in Table 2.1. The reliability problem consists in evaluating: $p_{f,5} = \mathbb{P}(g_5(\mathbf{X}) \leq 0) = 3.78 \times 10^{-7}$. [mention the cross-cut]

Variable	m_p	m_s	k_p	k_s	ζ_p	ζ_s	F_s	S_0
Distribution	Lognormal							
Mean	1.5	0.01	1.0	0.01	0.05	0.02	27.5	100.0
Coeff. of variation	0.1	0.1	0.2	0.2	0.4	0.5	0.1	0.1

Table 2.1 Random inputs from the toy-case #5.

2.3.2 Benchmark results and analysis

In the present section, the BANCS algorithm is applied to four analytical toy-cases and its performances are compared with SS and NAIS. After representing the two first iterations of BANCS on toy-case #1 in Fig. 2.1, all BANCS iterations from this case are illustrated in Fig. 2.2. This figure shows the intermediate quantiles $\{\tilde{q}_{[k]}^{p_0}\}_{k=1}^{k_\#}$ which are estimated over conditional samples of size $N = 10^4$. The points of each sample exceeding their respective p_0 -quantile are also represented in the same color as their p_0 -quantile border. Fig. 2.2 right provides the same kind of illustration for BANCS applied to test-case #2, a well-known system problem. On these two-dimensional cases, one can visualize the empirical quantiles driving the conditional distributions towards the failure domain(s) and the nonparametric inference capturing multi-modal patterns.

After this first illustration of one execution of BANCS, a more extensive benchmark is proposed. To present a fair assessment of the dispersion of the methods compared, each experiment is repeated 100 times to obtain a set of independent estimates of failure probabilities. Then, a Bootstrap procedure on this set allows to compute a confidence interval of the mean failure probability. An empirical variance and coefficient of variation (COV) of each probability estimator are also computed on the set of repetitions. This way, the SS coefficient of variation is not the result of an approximation (tending to underestimate the true SS coefficient of variation Papaioannou et al., 2015). Fig. 2.4 summarizes the results of this benchmark comparing BANCS with SS and NAIS for different sample sizes $N \in \{300, 500, 700, 10^3, 2 \cdot 10^3, 5 \cdot 10^3, 10^4\}$. On the left side, the estimates of failure probabilities repeated 100 times are displayed by their average over the repetitions (full lines) and their Bootstrap confidence intervals. The reference value for each problem is also represented by a horizontal black line. On the right, an estimate of the coefficient of variation provides a normalized information on the dispersion of the estimators. Note that all the estimates are plotted against their total number of samples (corresponding to the number of evaluations to the function).

To present a diverse panel of problems, the benchmark is conducted on test-case #2 (four-branch problem with $d = 2$), test-case #4 (with $d = 7$), test-case #5 (oscillator problem with $d = 8$). BANCS consistently showed good results on all three cases. The variance of its estimator is always smaller or equivalent to the SS variance (see their respective COV). In test-cases #2 and #3, BANCS estimation converges faster than SS and as fast as NAIS. For test-case #4, the NAIS implementation used does not support input distributions with bounded domains and was therefore removed. BANCS did not encounter the same difficulty as it fits separately the copula and the marginals (which can easily be truncated). On this rare and complex case, SS is more accurate for small-sized samples (i.e., $N \leq 10^3$) but becomes equivalent to BANCS for larger sample sizes. Finally, BANCS showed equivalent or better performances than SS and NAIS (on these first cases), while providing i.i.d. sampling and the flexibility of nonparametric inference.

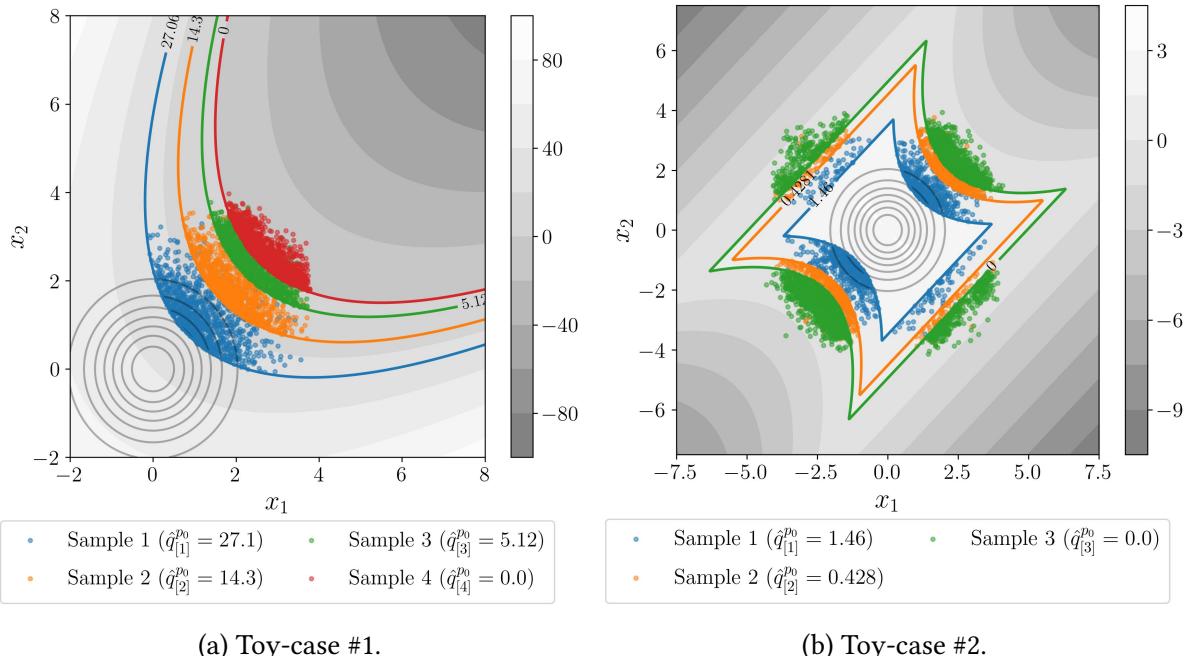


Figure 2.2 BANCS iterations on the two-dimensional reliability problems (for $N = 10^4$ and $p_0 = 0.1$). Only the samples exceeding the intermediary thresholds are represented.

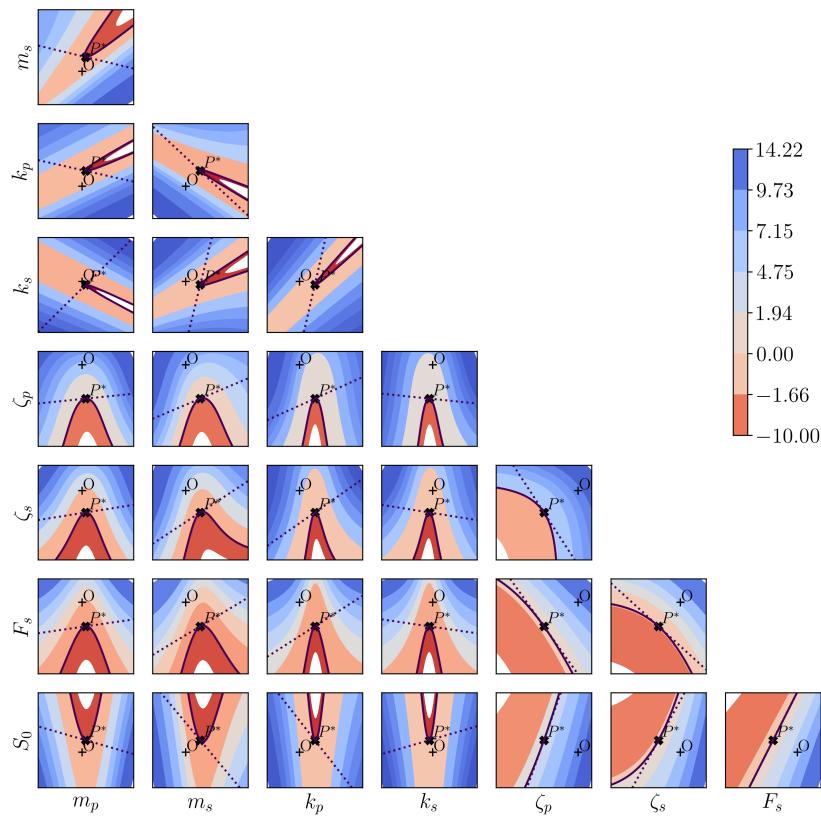


Figure 2.3 Cross-cut visualization of the function of test-case #5 through FORM's most-probable failure point P^* . The limit-state function (full line) delimits the safe domain (in blue) and the failure domain (in red). FORM produces the Taylor approximation around P^* (dashed lines).

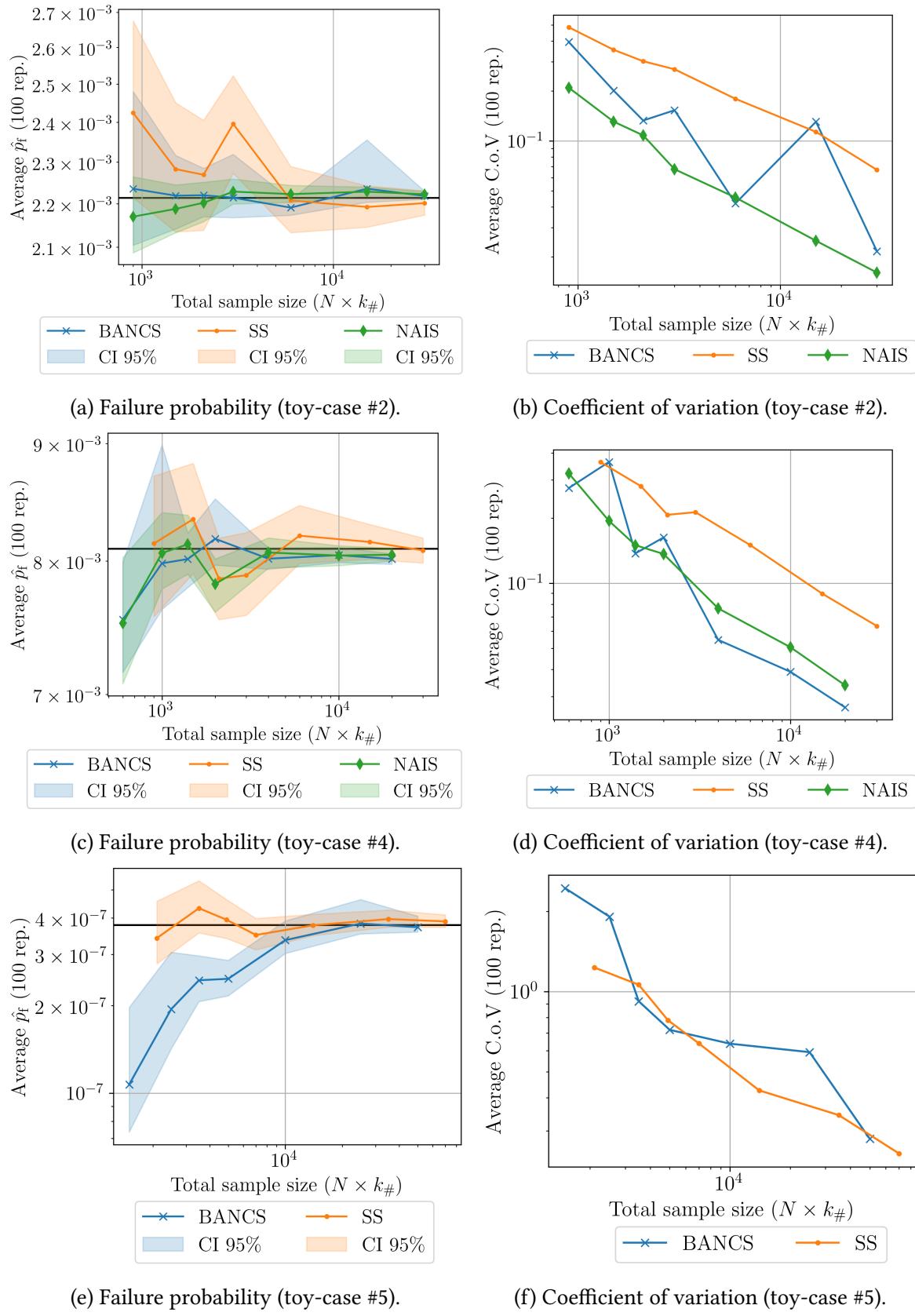


Figure 2.4 Todo.

2.4 Reliability-oriented sensitivity analysis

Global sensitivity analysis was introduced in Section 1.6 as a tool to study the general impact of the input random variables on the output random variable. When estimating quantities of interest related to the output distribution's tail (e.g., failure probabilities), the SA should be dedicated to the subdomain of interest (i.e., failure domain). In other words, the global influence of the inputs (e.g., on the output's variance) may be different from their reliability-oriented influence (e.g., their influence on a rare event probability).

To address the topic of *reliability-oriented sensitivity analysis* (ROSA), different SA methods were adapted to the analysis of risk measures. The importance factors are a first local ROSA measure, near the most-probable failure point, obtained as a post-treatment of FORM. Later on, global sensitivities such as the Sobol' indices were adapted to the indicator [Wei et al. \(2012\)](#); [Chabridon \(2018\)](#); [Perrin and Defaux \(2019\)](#). Recently, [Papaioannou and Straub \(2021\)](#) discussed the link between the importance factors and the Sobol' indices on the indicator. Moment-independent importance measures were also applied to study the sensitivity of reliability. For example, [Da Veiga \(2015\)](#) suggested the use of Hilbert-Schmidt independence criterion (HSIC) for reliability which was further explored by [Marrel and Chabridon \(2021\)](#) who distinguished two categories of ROSA with different purposes:

- *Target sensitivity analysis* (TSA): measuring the influence of input variables to reach a restricted domain defined by the output (e.g., failure domain);
- *Conditional sensitivity analysis* (CSA): studies the impact of input variables within the restricted domain (i.e., conditionally to this domain).

In the context of dependent inputs, the Shapley effects were adapted to the indicator function ([Il Idrissi et al., 2021](#); [Demange-Chryst et al., 2023](#)).

This section aims at demonstrating the estimation of TSA and CSA indices as a simple post-processing of BANCS. After estimating a rare event probability with BANCS, a set of i.i.d. samples (each with size N) are available. These consecutive samples gradually move closer to the failure domain and become rarer from the perspective of the initial input distribution. This section first recalls the TSA and CSA adaptation from [Marrel and Chabridon \(2021\)](#) of the Hilbert-Schmidt independence criterion (HSIC), a kernel-based sensitivity measure introduced in Section 1.6 for GSA. Then, the computation of the target-HSIC (respectively conditional-HSIC) is presented as a post-precessing of the BANCS reliability analysis of test-cases #3 and #5.

2.4.1 Target and conditional HSIC

A first approach for TSA is to directly apply any SA measure to the binary variable $\mathbb{1}_{\{g(\mathbf{X}) \leq y_{\text{th}}\}} = \mathbb{1}_{\{Y \leq y_{\text{th}}\}}$. This strategy does not distinguish points in the vicinity of the limit-state function from points far from this border. In the present work, a threshold relaxation using the weight transformation from [Marrel and Chabridon \(2021\)](#) is applied to gather more information. Let

us consider the weight function $w_{\mathcal{F}} : \mathbb{R} \rightarrow [0, 1]$ such that $w_{\mathcal{F}}(y) = \exp(-d_{\mathcal{F}}(y)/s)$, where $d_{\mathcal{F}} = \inf_{y' \in \mathcal{F}} \|y - y'\|$ is a distance to the border and $s \in \mathbb{R}$ is a smoothing parameter.

Then, a TSA measure can be defined as the result of any sensitivity measure applied between the random inputs \mathbf{X} and $w_{\mathcal{F}}(Y)$. In our case, the target-HSIC (T-HSIC) and their respective normalized version the target R^2_{HSIC} are defined as:

$$T\text{-HSIC}(X_j, Y) = \text{HSIC}(X_j, w_{\mathcal{F}}(Y)) = \text{MMD}^2(\mathbb{P}_{(X_j, w_{\mathcal{F}}(Y))}, \mathbb{P}_{X_j} \otimes \mathbb{P}_{w_{\mathcal{F}}(Y)}), \quad (2.10a)$$

$$T\text{-}R^2_{\text{HSIC}}(X_j, Y) = \frac{T\text{-HSIC}(X_j, w_{\mathcal{F}}(Y))}{\sqrt{T\text{-HSIC}(X_j, X_j) T\text{-HSIC}(w_{\mathcal{F}}(Y), w_{\mathcal{F}}(Y))}}. \quad (2.10b)$$

The stronger the dependence, the more X_j is influent on the occurrence of the rare event $\{Y \in \mathcal{F}\}$. Another advantage of applying the threshold relaxation is that any real-value kernel can be used to define the HSIC.

To define the CSA indices, let us define the probability of Y conditionally to the rare event $\{Y \in \mathcal{F}\}$, considering the measurable space (Ω, \mathcal{A}) :

$$\mathbb{P}_{Y|\{Y \in \mathcal{F}\}}(A) = \frac{\int_A \mathbb{1}_{g(\mathbf{X}) \leq 0}(y) d\mathbb{P}_Y}{p_f}, \quad \forall A \in \mathcal{A}. \quad (2.11)$$

After applying the same threshold relaxation as earlier, one can write:

$$\mathbb{P}_Y^{w_{\mathcal{F}}}(A) = \frac{\int_A w_{\mathcal{F}}(y) d\mathbb{P}_Y}{\int_{\Omega} w_{\mathcal{F}}(y) d\mathbb{P}_Y}, \quad (2.12)$$

and the conditional expectation [why is the “y” is in capital in the paper?]:

$$\mathbb{E}[Y|Y \in \mathcal{F}] = \mathbb{E}_{Y \sim \mathbb{P}_Y^{w_{\mathcal{F}}}}[Y] = \frac{\int_{\Omega} y w_{\mathcal{F}}(y) d\mathbb{P}_Y}{\int_{\Omega} w_{\mathcal{F}}(y) d\mathbb{P}_Y}. \quad (2.13)$$

Using this expression, the following conditional HSIC is developped in [Marrel and Chabridon \(2021\)](#):

$$C\text{-HSIC}(X_j, Y) = \text{HSIC}_{(X_j, Y) \sim \mathbb{P}_{(X_j, Y)}^{w_{\mathcal{F}}}}(X_j, w_{\mathcal{F}}(Y)), \quad (2.14a)$$

$$C\text{-}R^2_{\text{HSIC}}(X_j, Y) = \frac{C\text{-HSIC}(X_j, w_{\mathcal{F}}(Y))}{\sqrt{C\text{-HSIC}(X_j, X_j) C\text{-HSIC}(w_{\mathcal{F}}(Y), w_{\mathcal{F}}(Y))}}. \quad (2.14b)$$

The R2-HSIC is a normalized index which can be used to rank variables by influence but its value depends on the choice of kernel. Statistical tests are the only robust way of screening variables with respect to our quantity of interest. [De Lozzo and Marrel \(2016\)](#) proposed a permutation-based statistical test of independence based HSIC. Note that the HSIC generally do not require independence of the samples but the permutation tests do.

The respective estimators of all indices are provided in [Marrel and Chabridon \(2021\)](#) and their implementation is available in `OpenTURNS`. Unlike Sobol’ indices, the HSIC estimation

is realized on the same sample for every variable X_j . This property allows us to apply HSIC TSA and CSA to the samples evaluated during the BANCS reliability analysis.

2.4.2 ROSA as a post-treatment of BANCS reliability analysis

After assessing a rare event probability using BANCS, a nested set of samples moving towards the failure domain is available. This setup is the opportunity to study the evolution of the ROSA of all variables as the problem gets rarer. At iteration k , a TSA is conducted on the sample S_k with the intermediary threshold $\widehat{q}_k^{p_0}$ to study which random variables lead below this threshold. Then the CSA is assessed on the sample S_{k+1} , drawn according to distribution $\widehat{f}_{[k+1]}$ conditional to the threshold $\widehat{q}_k^{p_0}$. The target and conditional sensitivity analysis are presented via the R2-HSIC and the p-value of permutation tests. Fig. 2.5 represents these results for the test-case #3 while Fig. ?? shows the results for test-case #5.

Test-case #3: On the Ishigami problem, the variable X_2 has the least influence on the TSA while the variable X_3 has the least influence on the CSA. As the samples get closer to the failure domain the variable X_1 has less impact on the TSA and more impact on the CSA. [Add a sentence to comment the permutation tests.]

Test-case #5: On the oscillator problem, the variable F_s is of prime importance on the CSA. For the TSA, F_s has less influence as the samples get rarer while ζ_p and ζ_s become more important. [Add a sentence to comment the permutation tests.]

2.5 Conclusion

- Learn from a weighted sample
- Compare the learning-time with NAIS
- Optimize the EBC tuning with a test set (possibly build by KH)
- What if the sampling is not iid?
- Advantage of working directly in the physical space directly
- Perspective learning the entire distribution with Bernstein polynomials
- We could fit the marginals with parametric models and compare
- EBC by blocs in high dimension
- Estimate the quantiles with IS on all previous samples

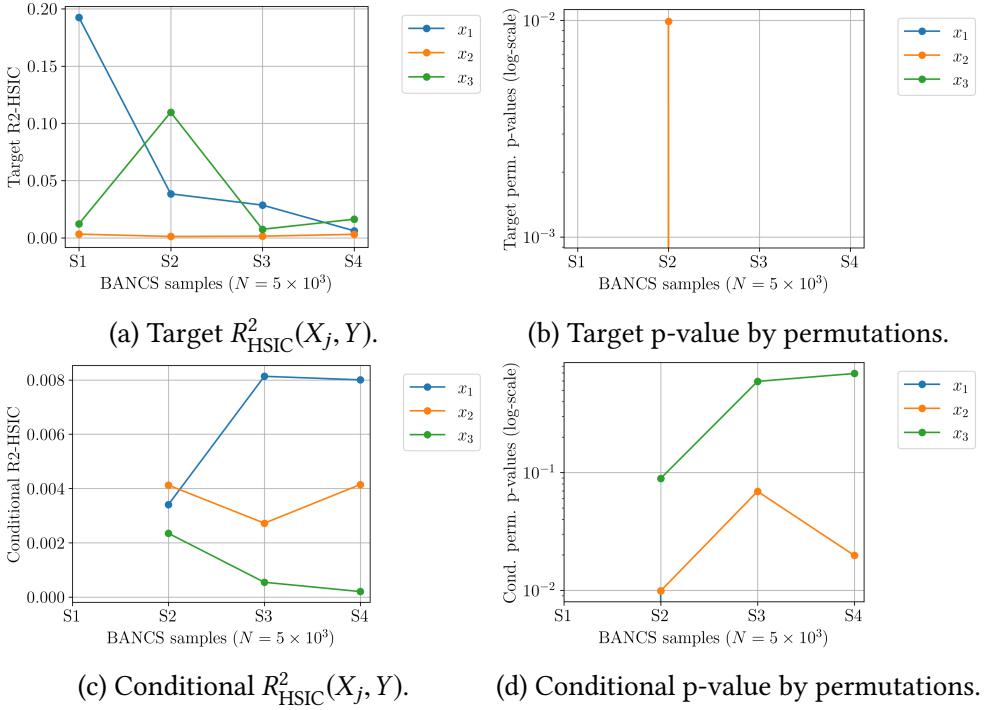


Figure 2.5 Target and conditional HSIC as a post-treatment of BANCS reliability analysis of test-case #3. The consecutive samples from BANCS are denoted by $\{S_k\}_{k=1}^{k_\#}$ (each with size $N = 5 \times 10^3$).

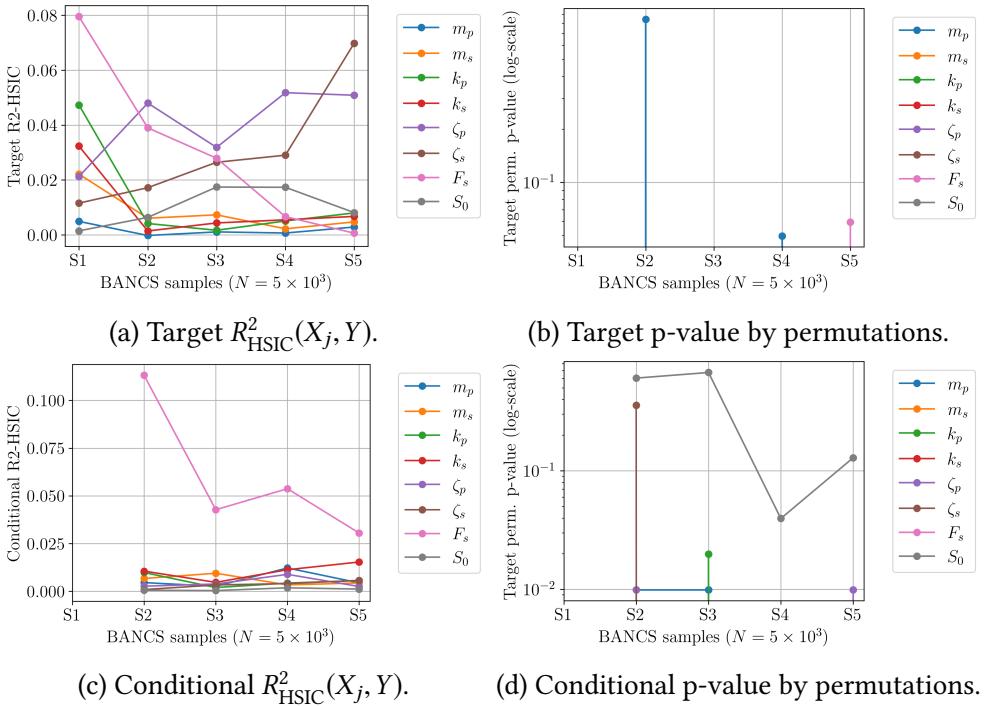


Figure 2.6 Target and conditional HSIC as a post-treatment of BANCS reliability analysis of test-case #5. Applied to the nested samples from BANCS (denoted by $\{S_k\}_{k=1}^{k_\#}$), each with size $N = 5 \times 10^3$.

Subset Simulation uses MCMC sampling to generate its intermediary conditional samples. However, MCMC algorithms tends to be complex to tune and does not generate i.i.d. conditional samples. In this work, a new method is proposed, replacing MCMC sampling with a simpler procedure. An intermediary conditional distribution is first fitted by a nonparametric approach, mixing kernel density estimation for fitting the marginals and Empirical Bernstein Copula (EBC) for fitting the copula. Then, the resulting allows to perform direct Monte Carlo sampling. This method is named “Bernstein adaptive nonparametric conditional sampling” (BANCS) and is applied to three toy-cases (two 2-dimensional and one 7-dimensional) and compared with SS.

The method shows promising results, even though a small positive bias consistently appears. This issue results from EBC tuning, creating a bias-variance tradeoff in the copula fit. Theoretical works offer optimal tuning, allowing us to find the optimal compromise. In our numerical experiments, an empirical estimation of BANCS variance is computed over a set of repetitions. BANCS estimated coefficient of variation is higher than SS approximated coefficient of variation. This work can be further explored by building an approximation of BANCS variance and confidence interval. One major advantage remains that the samples generated at each iteration are i.i.d. leading to a possible use of these samples to perform global reliability-oriented sensitivity analysis ([Marrel and Chabridon, 2021](#)) in order to detect and analyze the most influential input variables leading to failure.

Chapter **3**

Application to wind turbine fatigue reliability and robustness

3.1	Introduction	84
3.2	Surrogate modeling for reliability analysis	84
3.2.1	High-performance computer evaluation	84
3.2.2	Design of experiments	84
3.2.3	Gaussian process regression	86
3.3	Reliability and robustness analysis	86
3.3.1	Nominal reliability analysis	86
3.3.2	Robustness analysis using the perturbed-law sensitivity indices	86
3.4	Conclusion	88

On risk levels for OWT: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/we.2610>

3.1 Introduction

- Recall the system random variables
- Define our quantity of interest
- recall the calculation on the damage and the linearity of Miner's
- Expression of the failure probability (all in integral form)

3.2 Surrogate modeling for reliability analysis

3.2.1 High-performance computer evaluation

- A specific wrapper was built with a double loop structure
- The design of experience in the X space is with a size 200
- This design is repeated for 11 identical seeds
- Altogether : 2200 DIEGO simulations for each evaluation in the Z space
- The current setup and the HPC facility CRONOS allows us to compute these 2200 simulations fully in parallel.

3.2.2 Design of experiments

- The DoE concerns only two variables. The others intervene in the post-processing of DIEGO.
- A Halton sequence with size 30 was first built. In 2D, it was the iterative design of experiments with the lowest discrepancy (among Halton, Sobol, KH, SP).
- After visualizing the first 30 points, an enrichment by KH is realized on an zone defined apriori.
- This composite DoE has a final size of 50 points which represents in total more than 100k DIEGO simulations (each representing 45min of CPU time with this discretization timestep).

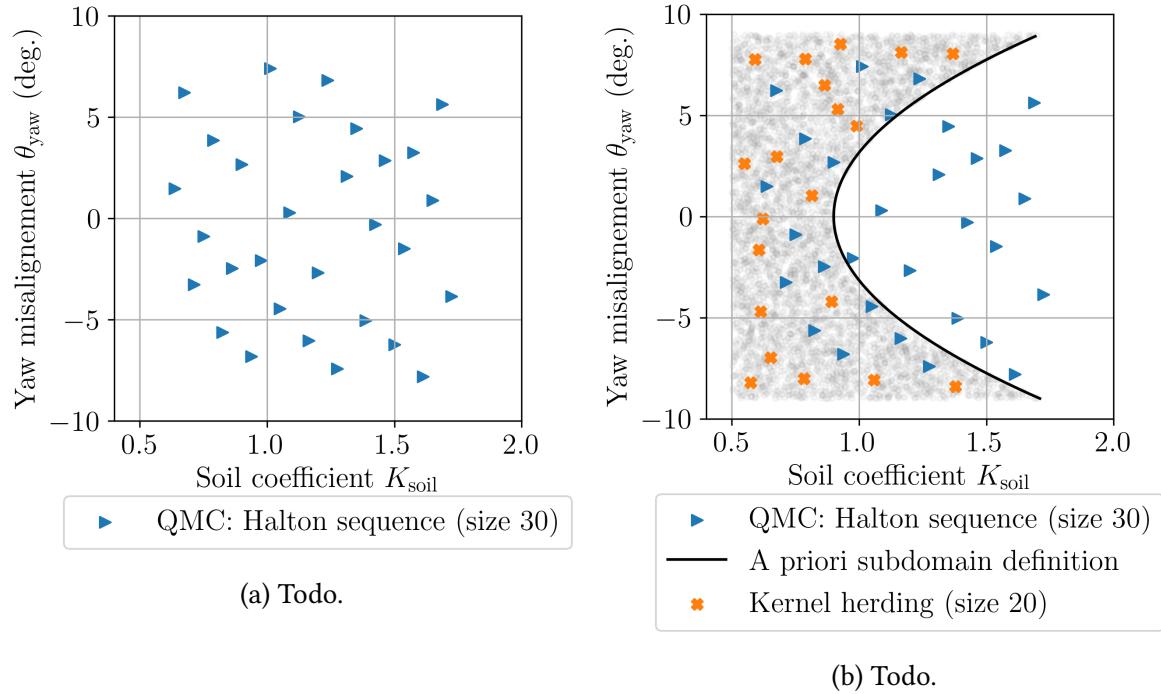


Figure 3.1 Todo.

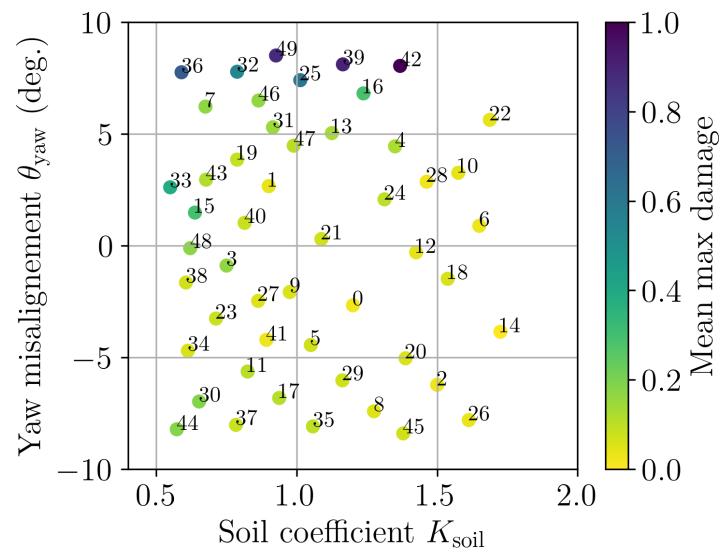


Figure 3.2 Todo.

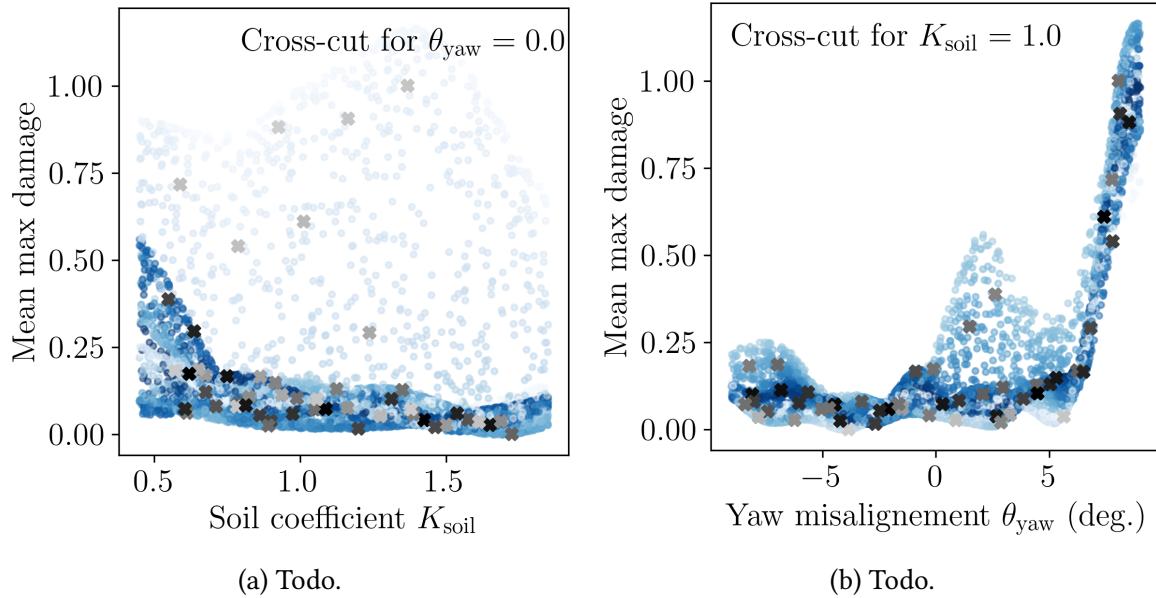


Figure 3.3 Todo.

3.2.3 Gaussian process regression

- Optimization of the hyperparameters by multi-start process in OT
- Choice of kernel Matérn 5/2
- Why not adaptive? Because the reliability analysis happens at the borders of the domain. The quality of the GP in these areas is guaranteed by the enrichment. Also: the enrichment might not be as robust to noisy functions.
- LOO validation
- Possible lack of trust in the mean estimation

3.3 Reliability and robustness analysis

3.3.1 Nominal reliability analysis

- Nominal reliability analysis for 3 possible hypothesis of $D_c r$ and using multiple RA methods including BANCS

3.3.2 Robustness analysis using the perturbed-law sensitivity indices

- Definition of the PLI
- Perturbation protocol (only on the variance)

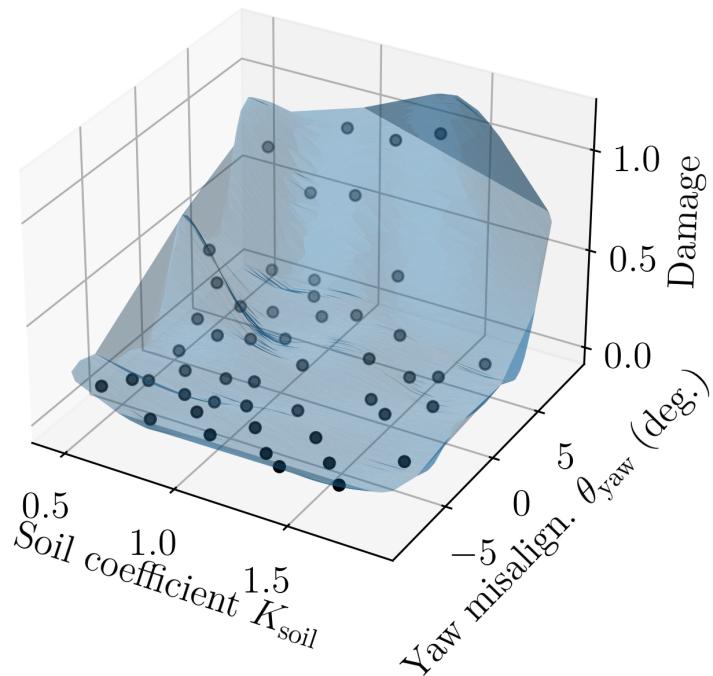


Figure 3.4 Todo.

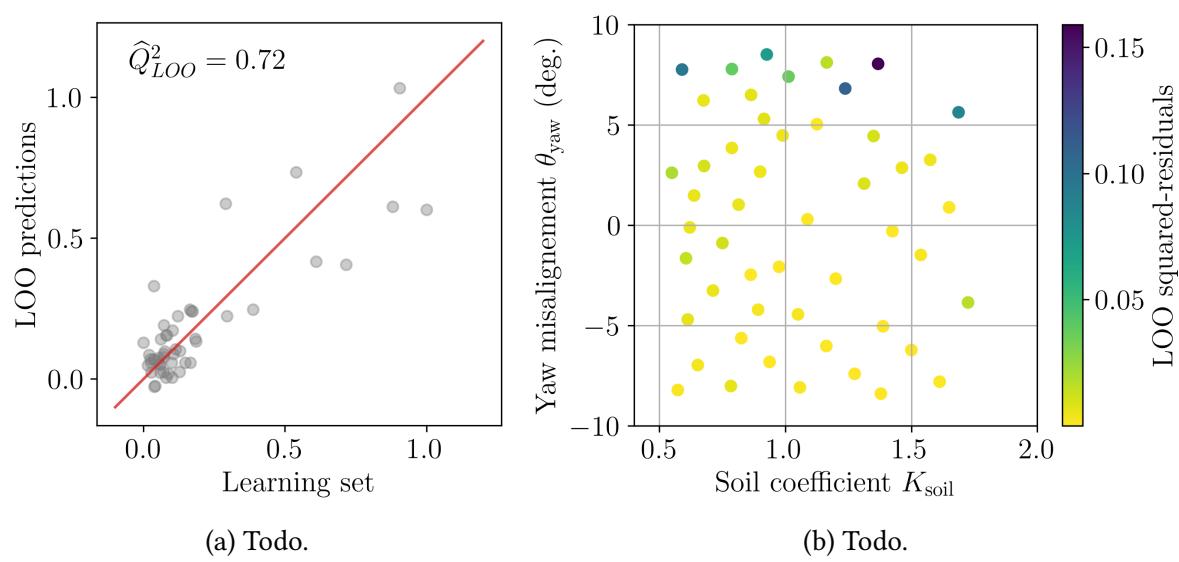


Figure 3.5 Todo.

Table 3.1 Results of the numerical experiments (subset sample size $N = 10^4$, $p_0 = 0.1$).

RA method	$D_{cr} \sim \text{Lognormal}$		$D_{cr} \sim \text{Normal}$	
	\hat{p}_f	$\widehat{\text{cov}}$	\hat{p}_f	$\widehat{\text{cov}}$
FORM	9.87×10^{-13}	—	3.35×10^{-6}	—
FORM-IS	9.84×10^{-13}	1%	3.36×10^{-6}	1%
SS	9.46×10^{-13}	7%	3.50×10^{-6}	4%

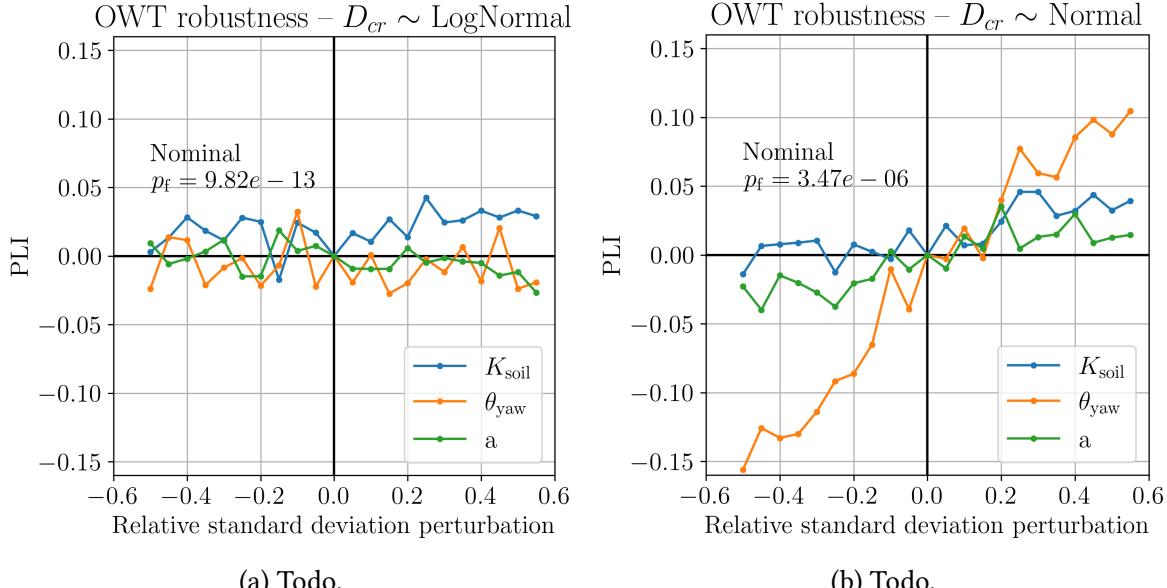


Figure 3.6 Todo.

- Numerical results and discussion on the relevance of the study
- Turbines are stopped about 6pc of their lifetime, which reduces the aerodynamic amortissement and therefore increases fatigue
- Starts and stops increase the damage
- Installation of the foundation can create early damage

3.4 Conclusion

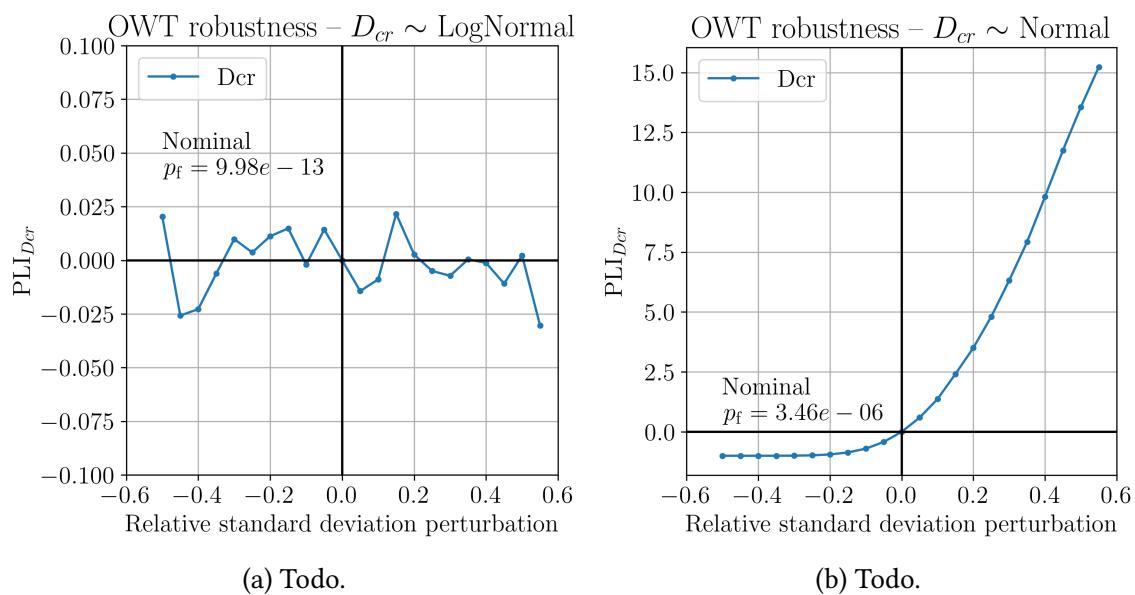


Figure 3.7 Todo.

Conclusion and perspectives

Bibliography

- Abdo, T. and Rackwitz, R. (1991). A new beta-point algorithm for large time-invariant and time-variant reliability problems. In *Reliability and Optimization of Structural Systems' 90: Proceedings of the 3rd IFIP WG 7.5 Conference*, pages 1–12.
- Abtini, M. (2018). *Plans prédictifs à taille fixe et séquentiels pour le krigeage*. PhD thesis, Ecole Centrale Lyon.
- Ajenjo, A. (2023). *Info-gap robustness assessment of reliability evaluations for the safety of critical industrial systems*. PhD thesis, Université Bourgogne Franche-Comté.
- Ajenjo, A., Ardillon, E., Chabridon, V., Iooss, B., Cogan, S., and Sadoulet-Reboul, E. (2022). An info-gap framework for robustness assessment of epistemic uncertainty models in hybrid structural reliability analysis. *Structural Safety*, 96:102196.
- Ang, G., Ang, A. H.-S., and Tang, W. (1992). Optimal importance-sampling density estimator. *Journal of engineering mechanics*, 118(6):1146–1163.
- Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277.
- Auder, B., De Crecy, A., Iooss, B., and Marques, M. (2012). Screening and metamodeling of computer experiments with functional outputs. Application to thermal–hydraulic computations. *Reliability Engineering & System Safety*, 107:122–131.
- Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55–69.
- Baker, E., Barbillon, P., Fadikar, A., Gramacy, R., Herbei, R., Higdon, D., Huang, J., Johnson, L., Mondal, A., Pires, B., et al. (2022). Analyzing Stochastic Computer Models: A Review with Opportunities. *Statistical Science*, 37(1):64 – 89.
- Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A. (2017). Open TURNS: An industrial software for uncertainty quantification in simulation. In Ghanem, R., Higdon, D., and Owhadi, H., editors, *Springer Handbook on Uncertainty Quantification*, pages 2001–2038. Springer.
- Beauregard, E., Bérille, E., Berrabah, N., Berthelot, M., Burrows, J., Capaldo, M., Cornet, S., Costan, V., Duchet, M., Dufossé, E., Dupont, E., Franchet, M., Gouze, E., Grau, A., Joly, A., Kell, N., de Laleu, V., Latraube, F., Lovera, A., de Bazelaire, A., Monnot, E., Nogaro, G., Pagot, J., Pérony, R., Peyrard, C., Piguet, C., Régnier, A., Santibanez, E., Senn, C., Smith, C., Soriano, F., Stephan, P., Terte, N., Veyan, P., Vizireanu, D., and Yeow, L. (2022). *L'éolien en mer : un défi pour la transition énergétique*. Lavoisier.

- Bect, J., Bachoc, F., and Ginsbourger, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883 – 2919.
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22:773–793.
- Beer, M., Ferson, S., and Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. *Mechanical systems and signal processing*, 37(1-2):4–29.
- Bénard, C., Biau, G., Da Veiga, S., and Scornet, E. (2022). SHAFF: Fast and consistent SHAPley eFficit estimates via random Forests. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151, pages 5563–5582.
- Benoumechiara, N. and Elie-Dit-Cosaque, K. (2019). Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: Proceedings and Surveys*, 65:266–293.
- Binois, M., Huang, J., Gramacy, R., and Ludkovski, M. (2019). Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics*, 61(1):7–23.
- Bjerager, P. (1988). Probability integration by directional simulation. *Journal of Engineering Mechanics*, 114(8):1285–1302.
- Blanchard, J.-B., Chocat, R., Damblin, G., Baudin, M., Bousquet, N., Chabridon, V., Iooss, B., Keller, M., Pelamatti, J., and Sueur, R. (2023). Fiches pédagogiques sur le traitement des incertitudes dans les codes de calcul. Technical report, EDF.
- Blatman, G. and Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6):2345–2367.
- Bourinet, J.-M. (2018). *Reliability analysis and optimal design under uncertainty-Focus on adaptive surrogate-based approaches*. HDR, Université Clermont Auvergne.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24:123–140.
- Briol, F., Oates, C., Girolami, M., Osborne, M., and Sejdinovic, D. (2019). Probabilistic Integration: A Role in Statistical Computation? *Statistical Science*, 34(1):1 – 22.
- Briol, F.-X. (2019). *Statistical computation with kernels*. PhD thesis, University of Warwick.
- Bucher, C. (1988). Adaptive sampling—an iterative fast Monte Carlo procedure. *Structural safety*, 5(2):119–126.
- Bugallo, M., Elvira, V., Martino, L., Luengo, D., Miguez, J., and Djuric, P. (2017). Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, 34(4):60–79.
- Carmassi, M. (2018). *Uncertainty quantification and calibration of a photovoltaic plant model: warranty of performance and robust estimation of the long-term production*. PhD thesis, Université Paris Saclay.
- Cérou, F., Del Moral, P., Furon, T., and Guyader, A. (2012). Sequential monte carlo for rare event estimation. *Statistics and computing*, 22:795–808.
- Cérou, F., Guyader, A., and Rousset, M. (2019). Adaptive multilevel splitting: Historical perspective and recent results. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(4):043108.
- Chabridon, V. (2018). *Reliability-oriented sensitivity analysis under probabilistic model uncertainty—Application to aerospace systems*. PhD thesis, Université Clermont Auvergne.

- Chabridon, V., Balesdent, M., Perrin, G., Morio, J., Bourinet, J.-M., and Gayton, N. (2021). Global reliability-oriented sensitivity analysis under distribution parameter uncertainty. *Mechanical Engineering under Uncertainties: From Classical Approaches to Some Recent Developments*, pages 237–277.
- Chen, Y., Welling, M., and Smola, A. (2010). Super-samples from kernel herding. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 109 – 116. AUAI Press.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.
- Cousin, A. (2021). *Optimisation sous contraintes probabilistes d'un système complexe : Application au dimensionnement d'une éolienne offshore flottante*. PhD thesis, Institut Polytechnique de Paris.
- Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85:1283 – 1305.
- Da Veiga, S. (2021). Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis. Preprint <https://hal.science/hal-03108628>.
- Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. Society for Industrial and Applied Mathematics.
- Damblin, G. (2015). *Contributions statistiques au calage et à la validation des codes de calcul*. PhD thesis, Paris, AgroParisTech.
- Damblin, G., Couplet, M., and Iooss, B. (2013). Numerical studies of space-filling designs: Optimization of Latin Hypercube Samples and subprojection properties. *Journal of Simulation*, 7(4):276–289.
- De Lozzo, M. and Marrel, A. (2016). New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, 86(15):3038–3058.
- De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley & Sons.
- Demange-Chrst, J., Bachoc, F., and Morio, J. (2023). Shapley effect estimation in reliability-oriented sensitivity analysis with correlated inputs by importance sampling. *International Journal for Uncertainty Quantification*, 13(3).
- Der Kiureghian, A. and Dakessian, T. (1998). Multiple design points in first and second-order reliability. *Structural Safety*, 20:37–49.
- Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112.
- DNV-ST-0437 (2016). DNV-ST-0437: Loads and site conditions for wind turbines. Technical report, Det Norske Veritas.
- Durante, F. and Sempi, C. (2015). *Principles of copula theory*. CRC press.
- Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, 33:145–154.
- Efron, B. and Stein, C. (1981). The Jackknife Estimate of Variance. *The Annals of Statistics*, 9(3):586–596.

- Ehre, M., Papaioannou, I., and Straub, D. (2020). A framework for global reliability sensitivity analysis in the presence of multi-uncertainty. *Reliability Engineering & System Safety*, 195:106726.
- Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101.
- Fang, K., Liu, M.-Q., Qin, H., and Zhou, Y.-D. (2018). *Theory and application of uniform experimental designs*, volume 221. Springer.
- Fekhari, E., Baudin, M., Chabridon, V., and Jebroun, Y. (2021). otbenchmark: an open source Python package for benchmarking and validating uncertainty quantification algorithms. In *Proceedings of the 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering*, pages 218 – 231.
- Fekhari, E., Chabridon, V., Mure, J., and Iooss, B. (2023a). Bernstein adaptive nonparametric conditional sampling: a new method for rare event probability estimation. In *14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP14)*.
- Fekhari, E., Iooss, B., Muré, J., Pronzato, L., and Rendas, J. (2023b). Model predictivity assessment: incremental test-set selection and accuracy evaluation. In Salvati, N., Perna, C., Marchetti, S., and Chambers, R., editors, *Studies in Theoretical and Applied Statistics*, pages 315–347. Springer.
- Fernández-Godino, M., Park, C., Kim, N.-H., and Haftka, R. (2016). Review of multi-fidelity models. arXiv preprint arXiv:1609.07196.
- Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.
- Geoga, C., Anitescu, M., and Stein, M. (2020). Scalable Gaussian process computations using hierarchical matrices. *Journal of Computational and Graphical Statistics*, 29(2):227–237.
- Geyer, S., Papaioannou, I., and Straub, D. (2019). Cross entropy-based importance sampling using Gaussian densities revisited. *Structural Safety*, 76:15–27.
- Giles, M. (2008). Multilevel Monte Carlo Path Simulation. *Operations Research*, 56:607–617.
- Glad, I., Hjort, N., and Ushakov, N. (2007). Mean-squared error of kernel estimators for finite values of the sample size. *Journal of Mathematical Sciences*, 146(4):5977–5983.
- Gobet, E., Lerasle, M., and Métivier, D. (2022). Mean estimation for randomized quasi Monte Carlo method. hal-03631879v3.
- Gramacy, R. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC press.
- Grasedyck, L., Kressner, D., and Tobler, C. (2013). A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78.
- Gretton, A., Borgwardt, K. M., Rasch, M., Schölkopf, B., and Smola, A. (2006). A Kernel Method for the Two-Sample-Problem. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, pages 513–520.
- Hansen, N., Auger, A., Ros, R., Mersmann, O., Tušar, T., and Brockhoff, D. (2021). COCO: A platform for comparing continuous optimizers in a black-box setting. *Optimization Methods and Software*, 36(1):114–144.

- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Hawchar, L., El Soueid, C.-P., and Schoefs, F. (2017). Principal component analysis and polynomial chaos expansion for time-variant reliability problems. *Reliability Engineering & System Safety*, 167:406–416.
- Hickernell, F. (1998). A generalized discrepancy and quadrature error bound. *Mathematics of computation*, 67(221):299–322.
- Hirvoas, A. (2021). *Development of a data assimilation method for the calibration and continuous update of wind turbines digital twins*. PhD thesis, Université Grenoble Alpes.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19(3):293–325.
- Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17.
- Huchet, Q. (2019). *Kriging based methods for the structural damage assessment of offshore wind turbines*. PhD thesis, Université Blaise Pascal.
- IEC-61400-1 (2019). IEC 61400-1: Wind energy generation systems - Part 1: Design requirements. Technical report, International Electrotechnical Commission (IEC).
- Il Idrissi, M., Chabridon, V., and Iooss, B. (2021). Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs. *Environmental Modelling & Software*, 143:105115.
- Iooss, B., Vergès, V., and Larget, V. (2022). BEPU robustness analysis via perturbed law-based sensitivity indices. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(5):855–865.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall.
- Jones, D., Schonlau, M., and Welch, W. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13:455–492.
- Jonkman, B. (2009). Turbsim User's Guide: Version 1.50. Technical report, NREL.
- Joseph, V., Gul, E., and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380.
- Kaplan, Z., Li, Y., Nakayama, M., and Tuffin, B. (2019). Randomized quasi-Monte Carlo for quantile estimation. In *2019 Winter Simulation Conference (WSC)*, pages 428–439.
- Koutsourelakis, P. (2004). Reliability of structures in high dimensions. Part II. Theoretical validation. *Probabilistic engineering mechanics*, 19(4):419–423.
- Kucherenko, S. and Iooss, B. (2017). *Derivative-Based Global Sensitivity Measures*, pages 1241–1263. Springer International Publishing, Cham.
- Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N. (2009). Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94:1135–1148. Special Issue on Sensitivity Analysis.
- Kurtz, N. and Song, J. (2013). Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety*, 42:35–44.

- Larsen, G. C., Madsen, H., Thomsen, K., and Larsen, T. (2008). Wake meandering: a pragmatic approach. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 11(4):377–395.
- Lataniotis, C. (2019). *Data-driven uncertainty quantification for high-dimensional engineering problems*. PhD thesis, ETH Zürich.
- Laurie, D. (1997). Calculation of Gauss-Kronrod quadrature rules. *Mathematics of Computation*, 66(219):1133–1145.
- Le Maître, O. and Knio, O. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media.
- Le Riche, R. and Picheny, V. (2021). Revisiting Bayesian optimization in the light of the COCO benchmark. *Structural and Multidisciplinary Optimization*, 64(5):3063–3087.
- Lebrun, R. (2013). *Contributions à la modélisation de la dépendance stochastique*. PhD thesis, Université Paris-Diderot – Paris VII. (in English).
- L’Ecuyer, P. (2018). Randomized Quasi-Monte Carlo: An Introduction for Practitioners. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 29–52, Cham. Springer International Publishing.
- Lemaire, M., Chateauneuf, A., and Mitteau, J.-C. (2009). *Structural reliability*. John Wiley & Sons.
- Lemaître, P., Sergienko, E., Arnaud, A., Bousquet, N., Gamboa, F., and Iooss, B. (2015). Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85(6):1200–1223.
- Leobacher, G. and Pillichshammer, F. (2014). *Introduction to quasi-Monte Carlo integration and applications*. Springer.
- Li, Y., Kang, L., and Hickernell, F. (2020). Is a transformed low discrepancy design also low discrepancy? *Contemporary Experimental Design, Multivariate Analysis and Data Mining: Festschrift in Honour of Professor Kai-Tai Fang*, pages 69–92.
- Mak, S. and Joseph, V. (2018). Support points. *The Annals of Statistics*, 46:2562 – 2592.
- Mara, T. A. and Tarantola, S. (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121.
- Marelli, S. and Sudret, B. (2014). UQLab: A framework for uncertainty quantification in Matlab. In *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, pages 2554–2563.
- Marrel, A. and Chabridon, V. (2021). Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214:107711.
- Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of Sobol indices for the Gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751.
- Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30.
- Mckay, M., Beckman, R., and Conover, W. (1979). A Comparison of Three Methods for Selecting Vales of Input Variables in the Analysis of Output From a Computer Code. *Technometrics*, 21:239 – 245.

- Milano, D. (2021). *Numerical prototype for floating offshore wind turbines*. PhD thesis, The University of Edinburgh.
- Morio, J. (2011). Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position. *Reliability Engineering and System Safety*, 96(1):178–183.
- Morio, J. and Balesdent, M. (2015). *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach*. Woodhead Publishing, Elsevier.
- Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-Monte Carlo Integration. *Journal of Computational Physics*, 122(2):218–230.
- Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174.
- Moustapha, M., Marelli, S., and Sudret, B. (2022). Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174.
- Oates, C. and Sullivan, T. (2019). A modern retrospective on probabilistic numerics. *Statistics and computing*, 29(6):1335–1351.
- Oates, C. J. (2021). Minimum Discrepancy Methods in Uncertainty Quantification. Lecture Notes at École Thématique sur les Incertitudes en Calcul Scientifique (ETICS21), <https://www.gdr-mascotnum.fr/etics.html>.
- Oberkampf, W. and Roy, C. (2010). *Verification and validation in scientific computing*. Cambridge university press.
- Owen, A. (1992). A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(2):541–551.
- Owen, A. (2003). The dimension distribution and quadrature test functions. *Statistica Sinica*, 13:1–17.
- Owen, A. (2013). Monte Carlo theory, methods and examples. Stanford University.
- Owen, A. (2014). Sobol' indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251.
- Owen, A. and Zhou, Y. (2000). Safe and effective importance sampling. *Journal of the American Statistical Association*, 95(449):135–143.
- Papaioannou, I., Betz, W., Zwirglmaier, K., and Straub, D. (2015). MCMC algorithms for Subset Simulation. *Probabilistic Engineering Mechanics*, 41:89–103.
- Papaioannou, I., Geyer, S., and Straub, D. (2019). Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety*, 191:106564.
- Papaioannou, I. and Straub, D. (2021). Variance-based reliability sensitivity analysis and the FORM α -factors. *Reliability Engineering & System Safety*, 210:107496.
- Papakonstantinou, K., N., H., and Eshra, E. (2023). Hamiltonian MCMC methods for estimating rare events probabilities in high-dimensional problems. *Probabilistic Engineering Mechanics*, 74:103485.
- Perrin, G. and Defaux, G. (2019). Efficient evaluation of reliability-oriented sensitivity indices. *Journal of Scientific Computing*, 79(3):1433–1455.

- Petit, S. (2022). *Improved Gaussian process modeling : Application to Bayesian optimization*. PhD thesis, Université Paris-Saclay.
- Petrovska, E. (2022). *Fatigue life reassessment of monopile-supported offshore wind turbine structures*. PhD thesis, University of Edinburgh.
- Powell, M. (1994). *A direct search optimization method that models the objective and constraint functions by linear interpolation*. Springer.
- Pronzato, L. and Müller, W. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22:681–701.
- Pronzato, L. and Zhigljavsky, A. (2020). Bayesian quadrature and energy minimization for space-filling design. *SIAM/ASA Journal on Uncertainty Quantification*, 8:959 – 1011.
- Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, 2016:1–16.
- Rahman, S. (2016). The f-sensitivity index. *SIAM/ASA journal on uncertainty quantification*, 4(1):130–162.
- Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*, volume 1. Springer.
- Rockafellar, R. and Royset, J. (2015). Engineering Decisions under Risk Averseness. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1.
- Rollón de Pinedo, A., Couplet, M., Iooss, B., Marie, N., Marrel, A., Merle, E., and Sueur, R. (2021). Functional outlier detection by means of h-mode depth and dynamic time warping. *Applied Sciences*, 11(23):11475.
- Roustant, O., Barthe, F., and Iooss, B. (2017). Poincaré inequalities on intervals – application to sensitivity analysis. *Electronic Journal of Statistics*, 11(2):3081 – 3119.
- Rozsas, A. and Slobbe, A. (2019). Repository and black-box reliability challenge 2019.
- Rubinstein, R. Y. and Kroese, D. P. (2004). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*, volume 133. Springer.
- Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo Method*. Wiley, Second ed. edition.
- Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(3):535–562.
- Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.
- Schilders, W., Van der Vorst, H. A., and Rommes, J. (2008). *Model order reduction: theory, research aspects and applications*, volume 13. Springer.
- Schöbi, R. and Sudret, B. (2017). Structural reliability analysis for p-boxes using multi-level meta-models. *Probabilistic Engineering Mechanics*, 48:27–38.
- Schobi, R., Sudret, B., and Wiart, J. (2015). Polynomial-chaos-based Kriging. *International Journal for Uncertainty Quantification*, 5(2).
- Segers, J., Sibuya, M., and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51.

- Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41:2263–2291.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R., and De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.
- Shang, B. and Apley, D. (2020). Fully-sequential space-filling design algorithms for computer experiments. *Journal of Quality Technology*, 53:1 – 24.
- Shapley, L. S. et al. (1953). A value for n-person games.
- Sobol̄, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1:407.
- Sobol, I. and Gresham, A. (1995). On an alternative global sensitivity estimators. *Proceedings of SAMO*, pages 40–42.
- Soize, C. and Ghanem, R. (2004). Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26:395–410.
- Song, E., Nelson, B. L., and Staum, J. (2016). Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083.
- Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. (2010). Hilbert Space Embeddings and Metrics on Probability Measures. *Journal of Machine Learning Research*, 11:1517–1561.
- Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151.
- Straub, D. (2014). *Engineering risk assessment*, pages 333–362. Springer.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979.
- Sullivan, T. (2015). *Introduction to uncertainty quantification*, volume 63. Springer.
- Sun, F., Wang, Y., and Xu, H. (2019). Uniform projection designs. *The Annals of Statistics*, 47:641 – 661.
- Tabandeh, A., Jia, G., and Gardoni, P. (2022). A review and assessment of importance sampling methods for reliability analysis. *Structural Safety*, 97:102216.
- Teixeira, R., Nogal, M., and O'Connor, A. (2021). Adaptive approaches in metamodel-based reliability analysis: A review. *Structural Safety*, 89:102019.
- Trefethen, L. (2008). Is Gauss quadrature better than Clenshaw–Curtis? *SIAM review*, 50(1):67–87.
- Van Kuik, G., Peinke, J., Nijssen, R., Lekou, D., Mann, J., Sørensen, J., Ferreira, C., van Wingerden, J., Schlipf, D., Gebraad, P., et al. (2016). Long-term research challenges in wind energy—a research agenda by the European Academy of Wind Energy. *Wind energy science*, 1(1):1–39.
- Waarts, P. (2000). *Structural reliability using finite element methods: an appraisal of directional adaptive response surface sampling (DARS)*. PhD thesis, Technical University of Delft, The Netherlands.
- Walter, C. (2015). Moving particles: A parallel optimal multilevel splitting method with application in quantiles estimation and meta-model based algorithms. *Structural Safety*, 55:10–25.

- Wand, M. and Jones, M. (1994). *Kernel smoothing*. CRC press.
- Wang, Z. and Song, J. (2016). Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis. *Structural Safety*, 59:42–52.
- Warnock, T. (1972). Computational investigations of low-discrepancy point sets. In *Applications of number theory to numerical analysis*, pages 319–343. Elsevier.
- Wei, P., Lu, Z., Hao, W., Feng, J., and Wang, B. (2012). Efficient sampling methods for global reliability sensitivity analysis. *Computer Physics Communications*, 183(8):1728–1743.
- Yun, W., Lu, Z., Zhang, Y., and Jiang, X. (2018). An efficient global reliability sensitivity analysis algorithm based on classification of model output and subset simulation. *Structural Safety*, 74:49–57.
- Zhang, P. (1996). Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(435):1245–1253.
- Zhu, X. (2022). *Surrogate Modeling for Stochastic Simulators Using Statistical Approaches*. PhD thesis, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich.

Appendix **A**

Univariate distribution fitting

This appendix recalls the main methods to infer a univariate distribution considering a n -sized i.i.d sample $X_n = \{x^{(1)}, \dots, x^{(n)}\} \in \mathbb{R}^n$. The goal is to use this finite set of observations of the random variable X to approach its underlying distribution by an estimated distribution. The inference techniques are split into two main groups, the methods assuming that the underlying distribution belongs to a family of parametric distributions are called parametric. Otherwise, the fitting method falls into the nonparametric group. Nonparametric methods often require a larger amount of data but allow more flexibility. In fact, nontrivial distributions (e.g., multimodal) might be easier to model using nonparametric approaches. To assess the quality of this estimation regarding the sample, a panel of goodness-of-fit methods are proposed [add ref], this appendix recalls a few of them. Note that the following tools can be used to estimate the marginals of a multivariate distribution.

Main parametric methods

Moments method

The moment's method aims at looking for a parametric distribution with density $f_X(\theta)$, whose first moments (e.g., $m(\theta)$ and $\sigma^2(\theta)$) match the empirical moments of the sample X_n (e.g., \widehat{m}_{X_n} and $\widehat{\sigma}^2$). After computing the empirical moments:

$$\widehat{m}_n = \frac{1}{n} \sum_{i=1}^n x^{(i)}, \quad \widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x^{(i)} - \widehat{m}_{X_n})^2, \quad (\text{A.1})$$

one can solve the system of equations ($m(\theta) = \widehat{m}_n$; $\sigma^2(\theta) = \widehat{\sigma}_n^2$) to determine the optimal set of parameters θ in this situation. Some families of distributions are more suited to this method (i.e., \mathcal{N}) because of the analytical expression of their moments. Moreover, this technique is sensitive to the possible biases in the estimation of the sample moments.

Maximum likelihood estimation

Maximum likelihood estimation (MLE) is a popular alternative to the moments method. Similarly, it aims at maximizing a given correspondence metric between the dataset X_n and a parametric distribution with density $f_X(\theta)$. This metric is the *likelihood* function, defined as:

$$\mathcal{L}(\theta|X_n) = \prod_{i=1}^n f_X(x^{(i)}; \theta), \quad (\text{A.2})$$

with the PDF taking the set of parameters θ written: $f_X(x^{(i)}; \theta)$. For numerical reasons, the optimization is often performed on the natural logarithm of the likelihood function, called *log-likelihood*. The goal is then finding the optimal vector $\hat{\theta}^*$ of parameters minimizing the following expression:

$$\hat{\theta}^* = \arg \min_{\theta \in \mathcal{D}_\theta} \left(- \sum_{i=1}^n \ln(f_X(x^{(i)}; \theta)) \right). \quad (\text{A.3})$$

Remark that the quick analytical results from the moment method can be used as a starting point of the MLE optimization. [Asymptotic behaviors of this method are described in: add ref] [This method can be applied to censored data in the field of survival analysis. Add ref]

Example 1. Considering a small set of observations $X_n = \{1, 2, 3, 4, 6\}$, the following figure xx represents

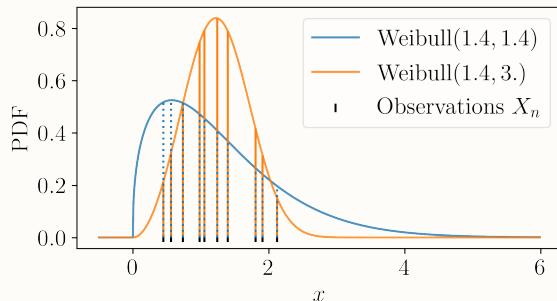


Figure A.1 Adequation of two different Weibull models using their likelihood with a sample of observations (black crosses).

Main nonparametric methods

Empirical CDF and histogram

The empirical CDF is a cumulative stair-shaped representation of the sorted sample X_n :

$$\widehat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x^{(i)} \leq x\}}. \quad (\text{A.4})$$

A histogram consists of sorting and gathering the observations in a sample X_n into a finite number of categories. These categories are called bins and each regroups the same number of observations (identical binwidth). The number of bins is the only tuning parameter of this method. Its definition has a great impact on the visual consistency of the plot, therefore, many rules exist to define it. Note that the empirical CDF can be seen as a cumulative histogram with the number of bins equal to the number of observations.

Kernel density estimation

Kernel density estimation (KDE) is a nonparametric method, it estimates a PDF by weighing a sample of observations X_n with kernels. After setting a kernel $k : \mathbb{R} \rightarrow \mathbb{R}_+$ and a scaling parameter $h > 0$, also called bandwidth, the kernel density estimator is defined as:

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x^{(i)}}{h}\right) \quad (\text{A.5})$$

Different types of kernels are used for KDE, such as the uniform, triangular, squared exponential or Epachnikov. The choice of bandwidth results in a bias-variance trade-off, that has been extensively discussed in the literature (Wand and Jones, 1994).

Example 2. Considering a small set of observations $X_n = \{1, 2, 3, 4, 6\}$, the following figure xx represents three fits obtained by.

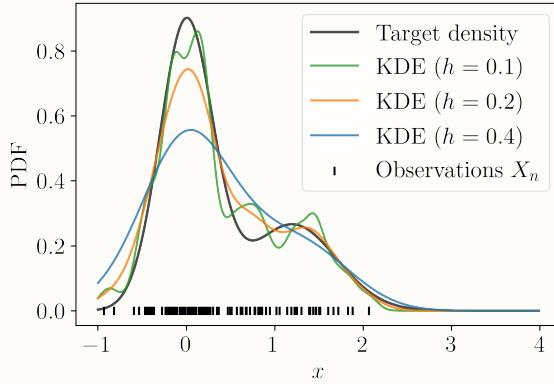


Figure A.2 Fit of a bimodal density by KDE using different tuning parameters.

Main goodness-of-fit methods

Penalized likelihood criteria

Two quantitative goodness-of-fit criteria are commonly used to assess parametric inference: the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC). The likelihood as a goodness-of-fit criterion should only be applied to the same family of distributions. Otherwise, the comparison would unfairly advantage distributions with a large number of degrees of

freedom. The two following criteria are metrics based on the likelihood with a correction related to the number of degrees of freedom of the distribution. Moreover, let us remember that more flexible models will require more data to provide a robust estimation.

The AIC and BIC are expressed as follows:

$$\text{AIC} = \frac{-2\ln(\mathcal{L}(\theta|X_n))}{n} + \frac{2q}{n}, \quad \text{BIC} = \frac{-2\ln(\mathcal{L}(\theta|X_n))}{n} + \frac{q\ln(n)}{n}, \quad (\text{A.6})$$

with the likelihood $\mathcal{L}(\theta|X_n)$ and the number of distribution's number degrees of freedom denoted q . The second term adds a penalty depending on the number of parameters. The best inference will be given by the model with the smallest AIC or BIC. Note that an additional correction can be applied in a small data context.

Kolmogorov-Smirnov adequacy test

Quantile-quantile plot

The quantile-quantile plot (also called QQ-plot) is a graphical tool providing a qualitative check of the goodness of fit. It compares the CDF of the fitted model with the empirical CDF of the sample X_n . To do so, it represents a scatterplot of the empirical quantiles (i.e., the ranked observations), against the quantiles of the fitted model at the levels $\{\alpha^{(i)}\}_{i=1}^n = \{\widehat{F}_X(x^{(i)})\}_{i=1}^n$. The following Fig. A.3 is a QQ-plot of the model fitted in [Example xx]. The closer the scatter plot gets to the first bisector line the better the fit is.

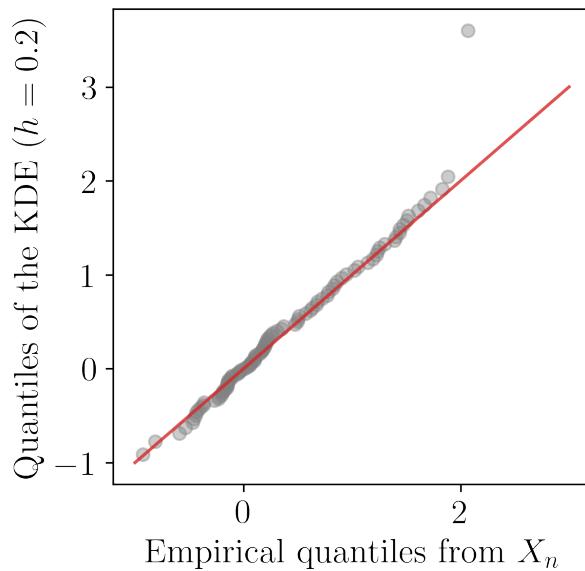


Figure A.3 QQ-plot between the data from Example 2 and a KDE model.

Appendix **B**

Dissimilarity measures between probability distributions

Beyond the discrepancy measure to the uniform distribution, this section introduces different dissimilarity measures between probability distributions.

Csizár f -divergences

[General definition]

[Numerous examples depending on the function chosen: see the book culte]

[Link between KL and mutual information] [Further inputs in the review from Rahman, maybe some in the PhD subject from A.Dufroy.] [Problems generated in the estimation]

Integral probability metrics

[general definition]

[Numerous examples see the book culte]

[No closed form expression unlike the f -divergence but the use of RKHS goes around this issue.]

Kernel discrepancy

Quasi-Monte Carlo sampling methods widely rely on a uniformity metric, called *discrepancy*. This section first presents the link between discrepancy and numerical integration. Then it introduces a kernel-based discrepancy, generalizing the concept to non-uniform measures. This tool is eventually used to build a sequential quadrature rule by subsampling from a finite dataset.

Reproducing kernel Hilbert space and kernel mean embedding To generalize the Koksma-Hlawka inequality to any probability measure, let us assume that the integrand g lives in a specific function space $\mathcal{H}(k)$. $\mathcal{H}(k)$ is a *reproducing kernel Hilbert space* (RKHS), which is an

inner product space of functions $g : \mathcal{D}_X \rightarrow \mathbb{R}$. Considering a symmetric and positive definite function $k : \mathcal{D}_X \times \mathcal{D}_X \rightarrow \mathbb{R}$, later called a “reproducing kernel” or simply a “kernel”, an RKHS verifies the following axioms:

- The “feature map” $\phi : \mathcal{D}_X \rightarrow \mathcal{H}(k); \phi(\mathbf{x}) = k(\cdot, \mathbf{x})$ belongs to the RKHS: $k(\cdot, \mathbf{x}) \in \mathcal{H}(k), \forall \mathbf{x} \in \mathcal{D}_X$;
- The “reproducing property”: $\langle g, k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(k)} = g(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{D}_X, \forall g \in \mathcal{H}(k)$.

Note that it can be shown that every positive semi-definite kernel defines a unique RKHS (and vice versa) with a feature map ϕ , such that $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}(k)}$. This framework allows us to embed a continuous or discrete probability measure in an RKHS, as illustrated in Fig. B.1. For any measure π , let us define its *kernel mean embedding* (Sejdinovic et al., 2013), also called “potential” $P_\pi(\mathbf{x})$ in Pronzato and Zhigljavsky (2020), associated with the kernel k as:

$$P_\pi(\mathbf{x}) := \int_{\mathcal{D}_X} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}'). \quad (\text{B.1})$$

Respectively, the potential $P_{\zeta_n}(\mathbf{x})$ of a discrete distribution $\zeta_n = \sum_{i=1}^n w_i \delta(\mathbf{x}^{(i)})$, $w_i \in \mathbb{R}$ associated with the kernel k can be written as:

$$P_{\zeta_n}(\mathbf{x}) = \int_{\mathcal{D}_X} k(\mathbf{x}, \mathbf{x}') d\zeta_n(\mathbf{x}') = \sum_{i=1}^n w_i k(\mathbf{x}, \mathbf{x}^{(i)}). \quad (\text{B.2})$$

The potential $P_\pi(\mathbf{x})$ of the targeted measure π will be referred to as “target potential” and the potential $P_{\zeta_n}(\mathbf{x})$ associated with the discrete distribution ζ_n called “current potential” when its support is the current design X_n . When $P_{\zeta_n}(\mathbf{x})$ is close to $P_\pi(\mathbf{x})$, it can be interpreted as ζ_n being an adequate quantization or representation of π (which leads to a good estimation of a quantity such as $I_\pi(g)$ from Eq. (??)). Potentials can be computed in closed forms for specific pairs of distribution and associated kernel. Summary tables of some of these cases are detailed in Briol (2019) (section 3.4), Pronzato and Zhigljavsky (2020) (section 4), and extended in Fekhari et al. (2023b). However, in most cases, the target potentials must be estimated on a large and representative sample, typically a large quasi-Monte Carlo sample of π .

Definition 1. The *energy* of a measure π is defined as the integral of the potential P_π against the measure, which leads to the following scalar quantity:

$$\varepsilon_\pi := \int_{\mathcal{D}_X} P_\pi(\mathbf{x}) d\pi(\mathbf{x}) = \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\pi(\mathbf{x}'). \quad (\text{B.3})$$

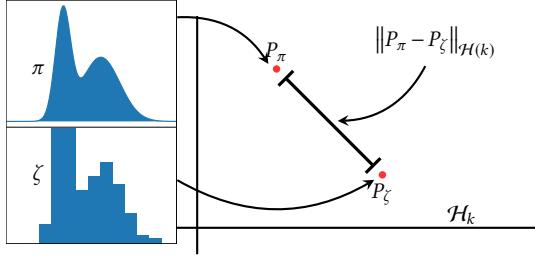


Figure B.1 Kernel mean embedding of a continuous and discrete probability distribution

Finally, using the reproducing property and writing the Cauchy-Schwarz inequality on the absolute quadrature error leads to the following inequality, similar to the Koksma-Hlawka inequality Eq. (??) (see Briol et al. (2019)):

$$\left| \sum_{i=1}^n w_i g(\mathbf{x}^{(i)}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\pi(\mathbf{x}) \right| = \left| \langle g, P_{\zeta_n}(\mathbf{x}) \rangle_{\mathcal{H}(k)} - \langle g, P_\pi(\mathbf{x}) \rangle_{\mathcal{H}(k)} \right| \quad (\text{B.4a})$$

$$= \left| \langle g, (P_{\zeta_n}(\mathbf{x}) - P_\pi(\mathbf{x})) \rangle_{\mathcal{H}(k)} \right| \quad (\text{B.4b})$$

$$\leq \|g\|_{\mathcal{H}(k)} \|P_\pi(\mathbf{x}) - P_{\zeta_n}(\mathbf{x})\|_{\mathcal{H}(k)}. \quad (\text{B.4c})$$

Maximum mean discrepancy A metric of discrepancy and quadrature error is offered by the *maximum mean discrepancy* (MMD). This distance between two probability distributions π and ζ is given by the worst-case error for any function within a unit ball of the Hilbert space $\mathcal{H}(k)$, associated with the kernel k :

$$\text{MMD}(\pi, \zeta) := \sup_{\|g\|_{\mathcal{H}(k)} \leq 1} \left| \int_{\mathcal{D}_X} g(\mathbf{x}) d\pi(\mathbf{x}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\zeta(\mathbf{x}) \right| \quad (\text{B.5})$$

According to the inequality in Eq. (B.4c), $\text{MMD}(\pi, \zeta) = \|P_\pi - P_\zeta\|_{\mathcal{H}(k)}$, meaning that the MMD fully relies on the difference of potentials. Moreover, Sriperumbudur et al. (2010) defines a kernel as “characteristic kernel” when the following equivalence is true: $\text{MMD}(\pi, \zeta) = 0 \Leftrightarrow \pi = \zeta$. This property makes the MMD a metric on \mathcal{D}_X . The squared MMD has been used for other purposes than numerical integration: e.g., statistical testing (Gretton et al., 2006), and global sensitivity analysis (Da Veiga, 2015). It can be written as follows:

$$\text{MMD}(\pi, \zeta)^2 = \|P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})\|_{\mathcal{H}(k)}^2 \quad (\text{B.6a})$$

$$= \langle (P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})), (P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})) \rangle_{\mathcal{H}(k)} \quad (\text{B.6b})$$

$$= \langle P_\pi(\mathbf{x}), P_\pi(\mathbf{x}) \rangle_{\mathcal{H}(k)} - 2 \langle P_\pi(\mathbf{x}), P_\zeta(\mathbf{x}) \rangle_{\mathcal{H}(k)} + \langle P_\zeta(\mathbf{x}), P_\zeta(\mathbf{x}) \rangle_{\mathcal{H}(k)} \quad (\text{B.6c})$$

$$= \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\pi(\mathbf{x}') - 2 \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\zeta(\mathbf{x}') + \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\zeta(\mathbf{x}) d\zeta(\mathbf{x}'). \quad (\text{B.6d})$$

Taking a discrete distribution with uniform weights $\zeta_n = \frac{1}{n} \sum_{i=1}^n \delta(\mathbf{x}^{(i)})$, the squared MMD reduces to:

$$\text{MMD}(\pi, \zeta_n)^2 = \varepsilon_\pi - \frac{2}{n} \sum_{i=1}^n P_\pi(\mathbf{x}^{(i)}) + \frac{1}{n^2} \sum_{i,j=1}^n k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}). \quad (\text{B.7})$$

Maximum discrepancy measure

A metric of discrepancy between distributions is introduced as the *maximum mean discrepancy* (MMD). This distance between two probability distributions μ and ζ is defined as the worst-case error for any function within a unit ball of a function space \mathcal{H} :

$$\text{MMD}(\mu, \zeta) := \sup_{\|g\|_{\mathcal{H}} \leq 1} \left| \int_{\mathcal{D}_X} g(\mathbf{x}) d\mu(\mathbf{x}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\zeta(\mathbf{x}) \right| = \|P_\mu(\mathbf{x}) - P_\zeta(\mathbf{x})\|_{\mathcal{H}}. \quad (\text{B.8})$$

To ease the calculation of the quantity, this metric was studied for a particular function space, offering specific properties. A *reproducing kernel Hilbert space* (RKHS), denoted $\mathcal{H}(k)$, is an inner product space $\mathcal{H}(k)$ of functions $g : \mathcal{D}_X \rightarrow \mathbb{R}$. It verifies the following axioms, considering a symmetric and positive definite function $k : \mathcal{D}_X \times \mathcal{D}_X \rightarrow \mathbb{R}$, later called a “reproducing kernel” or simply a “kernel”:

- The “feature map” $\phi : \mathcal{D}_X \rightarrow \mathcal{H}(k); \phi(\mathbf{x}) = k(\cdot, \mathbf{x}) \in \mathcal{H}(k), \forall \mathbf{x} \in \mathcal{D}_X$.
- The “reproducing property”: $\langle g, k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(k)} = g(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{D}_X, \forall g \in \mathcal{H}(k)$.

Every positive semi-definite kernel defines a unique RKHS (and vice versa) with a feature map ϕ , such that $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}(k)}$. Moreover, [Sriperumbudur et al. \(2010\)](#) defines a kernel as “characteristic kernel” when the following equivalence is true: $\text{MMD}_k(\mu, \zeta) = 0 \Leftrightarrow \mu = \zeta$. This property makes the MMD a metric on \mathcal{D}_X .

Then, a probability measure has a representation in the RKHS through its *kernel mean embedding* ([Sejdinovic et al., 2013](#)), also called “potential” $P_\mu(\mathbf{x})$ in [Pronzato and Zhigljavsky \(2020\)](#), defined as:

$$P_\mu(\mathbf{x}) := \int_{\mathcal{D}_X} k(\mathbf{x}, \mathbf{x}') d\mu(\mathbf{x}'). \quad (\text{B.9})$$

The reproducing property from the RKHS allows to express the squared MMD as expectations of kernels:

$$\text{MMD}_k(\mu, \zeta)^2 = \int_{\mathcal{D}_X} P_\mu(\mathbf{x}) d\mu(\mathbf{x}) - 2 \int_{\mathcal{D}_X} P_\mu(\mathbf{x}) d\zeta(\mathbf{x}) + \int_{\mathcal{D}_X} P_\zeta(\mathbf{x}) d\zeta(\mathbf{x}). \quad (\text{B.10})$$

[Add a sentence on estimation]

Analytical computation of potentials for Matérn kernels

As for tensor-product kernels, the potential is the product of the one-dimensional potentials, we only consider one-dimensional input spaces.

For μ the uniform distribution on $[0, 1]$ and K the Matérn kernel $K_{5/2,\theta}$ with smoothness $\nu = 5/2$ and correlation length θ , see (??), we get

$$P_{K_{5/2,\theta},\mu}(x) = \frac{16\theta}{3\sqrt{5}} - \frac{1}{15\theta}(S_\theta(x) + S_\theta(1-x)),$$

where

$$S_\theta(x) = \exp\left(-\frac{\sqrt{5}}{\theta}x\right)\left(5\sqrt{5}x^2 + 25\theta x + 8\sqrt{5}\theta^2\right).$$

The expressions $P_{K_{\nu,\theta},\mu}(x)$ for $\nu = 1/2$ and $\nu = 3/2$ can be found in ?.

When μ is the standard normal distribution $\mathcal{N}(0, 1)$, the potential $P_{K_{5/2,\theta},\mathcal{N}(0,1)}$ is $P_{K_{5/2,\theta},\mathcal{N}(0,1)}(x) = T_\theta(x) + T_\theta(-x)$, where

$$\begin{aligned} T_\theta(x) &= \frac{1}{6}\left(\frac{5}{\theta^2}x^2 + \left(3 - \frac{10}{\theta^2}\right)\frac{\sqrt{5}}{\theta}x + \frac{5}{\theta^2}\left(\frac{5}{\theta^2} - 2\right) + 3\right) \\ &\quad \times \operatorname{erfc}\left(\frac{\frac{\sqrt{5}}{\theta} - x}{\sqrt{2}}\right) \exp\left(\frac{5}{2\theta^2} - \frac{\sqrt{5}}{\theta}x\right) + \frac{1}{3\sqrt{2\pi}}\frac{\sqrt{5}}{\theta}\left(3 - \frac{5}{\theta^2}\right) \exp\left(-\frac{x^2}{2}\right). \end{aligned}$$

Appendix

C

Advanced rare event estimation algorithms

Subset simulation (SS)

Nonparametric adaptive importance sampling (NAIS)

Appendix D

Uncertainty quantification practice with OpenTURNS

[Add short introduction to the motivation]

[Should I print the results?]

OpenTURNS 7 (Bivariate distribution). The following Python code proposes a minimalist OpenTURNS example of a probabilistic uncertainty modeling.

```
1 #!/usr/bin/python3
2 import openturns as ot
3 # Build multivariate distribution from marginals and copula
4 copula=ot.GumbelCopula(2.0)
5 marginals=[ot.Uniform(1.0, 2.0), ot.Normal(2.0, 3.0)]
6 distribution=ot.ComposedDistribution(marginals, copula)
7 # Compute first moments
8 mean_vector=distribution.getMean()
9 covariance_matrix=distribution.getCovariance()
10 # Compute CDF (respectively PDF)
11 x_cdf=distribution.computeCDF([1.5, 2.5]) # x=[1.5, 2.5]
12 a_quantile=distribution.computeQuantile([0.9]) # alpha=0.9
```

OpenTURNS 8 (Numerical integration). The following Python code presents a minimalistic OpenTURNS example to build multivariate quadrature rules.

```

1  #!/usr/bin/python3
2  import openturns as ot
3  marginals=[ot.Exponential(1.0), ot.Uniform(-1.0, 1.0)]
4  distribution=ot.ComposedDistribution(marginals)
5  # Build a 2D Gaussian quadrature
6  n_marginal=[4, 4] # Number of nodes per marginal
7  g_quad=ot.GaussProductExperiment(distribution, n_marginal)
8  g_nodes, weights=g_quad.generateWithWeights()
9  # Build a Monte Carlo design
10 n=16
11 mc_nodes=distribution.getSample(n)
12 # Build a quasi-Monte Carlo design
13 sequence=ot.HaltonSequence(2) # d=2
14 qmc_experiment=ot.LowDiscrepancyExperiment(sequence, distribution, n)
15 qmc_nodes=qmc_experiment.generate()
```

OpenTURNS 9 (Design of experiments). The following Python code is a minimalistic OpenTURNS example to build an LHS and an LHS optimized w.r.t. to a space-filling metric (here the L2-centered discrepancy) using the simulated annealing algorithm.

```

1  #!/usr/bin/python3
2  import openturns as ot
3  marginals=[ot.Uniform(0.0, 1.0), ot.Uniform(0.0, 1.0)]
4  distribution=ot.ComposedDistribution(marginals)
5  # Build a LHS
6  n=10
7  LHS_exp=ot.LHSExperiment(distribution, n)
8  LHS_design=LHS_exp.generate()
9  # Build an optimized LHS using L2-centered discrepancy
10 LHS_exp=ot.LHSExperiment(distribution, n)
11 SF_metric=ot.SpaceFillingC2()
12 SA_profile=ot.GeometricProfile(10., 0.95, 20000)
13 LHS_opt=ot.SimulatedAnnealingLHS(LHS_exp, SF_metric, SA_profile)
14 LHS_opt.generate()
15 LHS_design=LHS_opt.getResult().getOptimalDesign()
```

OpenTURNS 10 (Rare event estimation). The following Python code proposes a minimalist OpenTURNS implementation of rare event estimation algorithms.

```

1  #!/usr/bin/python3
2  import openturns as ot
3  marginals=[ot.Normal(0.0, 1.0), ot.Exponential(1.0)]
4  distribution=ot.ComposedDistribution(marginals)
5  # Build a limit-state function and failure event
6  g=ot.SymbolicFunction(["x1", "x2"], ["(x1 - x2) ^ 2"])
7  X=ot.RandomVector(distribution)
8  Y=ot.CompositeRandomVector(g, X)
9  th=0.0
10 failure_event=ot.ThresholdEvent(Y, ot.LessOrEqual(), th)
11 # Estimate pf using FORM
12 starting_p=distribution.getMean()
13 FORM_algo=ot.FORM(ot.Cobyla(), failure_event, starting_p)
14 FORM_algo.run()
15 FORM_results=FORM_algo.getResult()
16 design_point=FORM_results.getStandardSpaceDesignPoint()
17 FORM_pf=FORM_results.getEventProbability()
18 # Estimate pf using Monte Carlo
19 MC_exp=ot.MonteCarloExperiment()
20 MC algo=ot.ProbabilitySimulationAlgorithm(failure_event, MC_exp)
21 MC algo.run()
22 MC_results=MC algo.getResult()
23 MC_pf=MC_results.getProbabilityEstimate()
24 MC_pf_confidence=MC_results.getConfidenceLength(0.95)
25 # Estimate pf using importance sampling
26 aux_distribution=ot.Normal(design_point, [1.0, 1.0])
27 standard_event=ot.StandardEvent(failure_event)
28 IS_exp=ot.ImportanceSamplingExperiment(aux_distribution)
29 IS algo=ot.ProbabilitySimulationAlgorithm(standard_event, IS_exp)
30 IS algo.run()
31 IS_results=IS algo.getResult()
32 IS_pf=IS_results.getProbabilityEstimate()
33 IS_pf_confidence=IS_results.getConfidenceLength(0.95)
34 # Estimate pf using subset simulation
35 SS algo=ot.SubsetSampling(failure_event)
36 SS algo.run()
37 SS_results=SS algo.getResult()
38 SS_pf=SS_results.getProbabilityEstimate()
39 SS_pf_confidence=SS_results.getConfidenceLength(0.95)
```

OpenTURNS 11 (Sobol' indices). The following Python code gives a minimalistic OpenTURNS implementation of the Sobol' indices to assess global sensitivity analysis on the Ishigami analytical problem.

```

1  #!/usr/bin/python3
2  import openturns as ot
3  g=ot.SymbolicFunction(
4      ['x1', 'x2', 'x3'],
5      ['sin(x1) + 7.0 * sin(x2)^2 + 0.1 * x3^4 * sin(x1)']
6  )
7  X=ot.ComposedDistribution([ot.Uniform(-3.14, 3.14)] * 3)
8  size=1000
9  # Generate samples and evaluate their images
10 sie=ot.SobolIndicesExperiment(im.distributionX, size)
11 input_design=sie.generate()
12 output_design=im.model(input_design)
13 # Four estimators : Saltelli, Martinez, Jansen, and Mauntz-Kucherenko
14 SA=ot.JansenSensitivityAlgorithm(input_design, output_design, size)
15 sobol_first_order=SA.getFirstOrderIndices()
16 sobol_tolal=SA.getTotalOrderIndices()
```

OpenTURNS 12 (Gaussian process regression). The following Python code gives a minimalistic OpenTURNS implementation of an ordinary kriging model fitting.

```

1  #!/usr/bin/python3
2  import openturns as ot
3  g=ot.SymbolicFunction(['x'], ['x * sin(x) + sin(6 * x)'])
4  x_train=ot.Uniform(0., 12.).getSample(7) # n=7
5  y_train=g(x_train)
6  basis=ot.ConstantBasisFactory(1).build() # d=1
7  cov_model=ot.MaternModel([1.], 1.5)
8  algo=ot.KrigingAlgorithm(x_train, y_train, cov_model, basis)
9  algo.run()
10 kriging_results=algo.getResult()
11 kriging_predictor=kriging_results.getMetaModel()
```

Appendix E

Résumé étendu de la thèse

Introduction

Contexte industriel

L'enjeu actuel de la transition énergétique implique, entre autres, de réduire la part des énergies fossiles au sein du mix électrique mondial. Dans ce contexte, l'énergie éolienne en mer présente plusieurs avantages [Beauregard et al. \(2022\)](#). L'éolien en mer bénéficie notamment de vents plus constants que l'éolien terrestre, notamment dû à l'absence de relief, et offre la possibilité d'installer des éoliennes plus grandes donc plus puissantes. Depuis l'installation de la première ferme éolienne en mer à Vindeby, au Danemark, en 1991, l'industrie a connu une croissance rapide, avec une capacité totale de 56 GW exploitée dans le monde en 2021. Au fil du temps, la technologie éolienne en mer s'est améliorée, aboutissant à des succès importants tels que la signature de projets non subventionnés en Europe (en anglais *zero-subsidy bids*), pour lesquels l'électricité produite est directement vendue sur le marché de gros ([Beauregard et al., 2022](#)).

Cependant, malgré les progrès techniques indéniables, des limites industrielles émergent vis-à-vis de ces parcs éoliens en mer, posant ainsi de nombreux défis scientifiques. Pour atteindre les ambitieux objectifs de développement au niveau national et régional, la filière de l'éolien en mer fait face à plusieurs problèmes liés à l'augmentation de la taille des turbines. Ce changement d'échelle crée notamment des tensions liées à la logistique portuaire, aux besoins en ressources primaires et à la gestion durable du démantèlement futur. Ce secteur présente plusieurs défis techniques et scientifiques, qui requièrent l'utilisation conjointe de données mesurées et de simulations numériques d'éoliennes dans leur environnement. La recherche appliquée à l'éolien en mer fait intervenir plusieurs disciplines qui étudient notamment des sujets tels que la conception d'éoliennes flottantes, l'amélioration de l'estimation des ressources éoliennes, l'optimisation des opérations de maintenance et l'augmentation de la durée de vie utile des parcs. De manière générale, plusieurs décisions sont prises durant la vie d'une éolienne par son concepteur, installateur et exploitant, tout en ayant une connaissance partielle de certains phénomènes physiques. Par conséquent, modéliser et maîtriser les diverses sources

d'incertitudes associées à l'éolien en mer s'avère être un élément déterminant dans une industrie hautement concurrentielle.

Dans l'ensemble, l'industrie de l'éolien en mer a besoin de méthodes de traitement des incertitudes pour maîtriser les marges de sûreté et la gestion des actifs industriels (à la maille des composants, de l'éolienne et du parc dans son ensemble) (Van Kuik et al., 2016). Pour un développeur de projets éoliens, l'attention est d'abord portée sur l'amélioration du potentiel éolien des sites candidats en combinant différentes sources d'information et en modélisant la distribution multivariée des conditions environnementales au sein d'un parc éolien. Dans le cas de projets en éolien flottant, l'objectif est d'intégrer un aspect probabiliste dès la phase de conception (par exemple, du flotteur) afin de définir des solutions plus sûres, plus robustes et plus rentables. Pour un propriétaire d'un parc éolien, la gestion de la fin de vie est une autre problématique importante. Un propriétaire de parc éolien en fin de vie a le choix entre trois options : prolonger la durée de vie des actifs en exploitation, remplacer les éoliennes actuelles par des modèles plus récents, ou démanteler et vendre le parc éolien. Les deux premières solutions nécessitent d'évaluer la fiabilité de la structure et sa durée de vie résiduelle. Ces évaluations quantitatives sont examinées par des organismes de certification et des assureurs pour délivrer des permis d'exploitation. Pour fournir des évaluations rigoureuses des risques, la méthodologie générique de *traitement des incertitudes* est une démarche qui fait consensus dans les secteurs industriels confrontés à ce genre de problématique (De Rocquigny et al., 2008).

Méthodologie générique de traitement des incertitudes dans les outils de calcul scientifiques

La simulation numérique est une discipline qui a émergé avec l'avènement de l'informatique. Cette pratique produit des outils de calcul scientifique (OCS) qui permettent de simuler le comportement de système complexes compte tenu de conditions initiales définies par l'analyste. Les OCS sont vite devenus indispensables pour l'analyse, la conception, et la certification de systèmes complexes dans les cas où des expériences ou des mesures physiques sont coûteuses à obtenir, voire impossibles à réaliser. Cependant, ces modèles numériques s'intègrent dans une démarche déterministe : le résultat d'une simulation est associé à un vecteur de paramètres fixé en entrée. La question de la gestion des incertitudes associées aux entrées se pose rapidement lors de l'utilisation des OCS.

Le traitement des incertitudes vise à modéliser et à traiter les incertitudes autour d'un modèle numérique. Pour ce faire, une méthodologie générique a été proposée pour quantifier et analyser les incertitudes entre les variables d'entrée et de sortie d'un OCS (De Rocquigny et al., 2008). Une présentation des outils mathématiques utilisés dans ce domaine est proposée par Sullivan (2015). Cette approche apporte une meilleure compréhension d'un système, ce qui contribue à une prise de décision plus robuste.

La Figure E.1 illustre les étapes génériques de la méthodologie de quantification des incertitudes, qui sont brièvement décrites ci-après :

- **Étape A – Spécification du problème.** Cette étape consiste à déterminer le système étudié et construire un modèle numérique capable de simuler (précisément) son comportement. La spécification du problème implique également de définir l'ensemble des paramètres inhérents au modèle numérique. Ces paramètres comprennent aussi bien les variables d'entrée que les variables de sortie générées par la simulation. Dans ce document, le modèle numérique est considéré comme une boîte-noire, par opposition à des approches qui s'intègrent à l'intérieur des schémas de résolution numérique des équations de comportement du système (approches dites intrusives ([Le Maître and Knio, 2010](#))). En général, ces modèles numériques sont au préalable calibrés par rapport à des données mesurées et suivent un processus de validation et de vérification pour réduire les erreurs de modélisation ([Oberkampf and Roy, 2010](#)).
- **Étape B – Modélisation et quantification des incertitudes.** L'objectif de la deuxième étape est d'identifier et modéliser toutes les sources d'incertitude associées aux variables d'entrée. Dans la plupart des cas, cette modélisation est effectuée dans un cadre probabiliste.
- **Étape C – Propagation des incertitudes.** Lors de cette étape, les entrées incertaines sont propagées au travers du modèle de simulation numérique. Dès lors, la sortie du modèle numérique (habituellement de type scalaire) devient également incertaine. L'objectif est alors d'estimer une quantité d'intérêt, c'est-à-dire une statistique sur la variable aléatoire de sortie étudiée. La méthode de propagation de l'incertitude peut différer en fonction de la quantité d'intérêt visée (par exemple, la tendance centrale, un quantile, une probabilité d'événement rare, etc.).
- **Étape C’ – Analyse de sensibilité.** En complément de la propagation d'incertitudes, une analyse de sensibilité peut être réalisée afin d'étudier le rôle attribué à chaque entrée incertaine dans la variabilité de la sortie d'intérêt.
- **Métamodélisation.** Compte tenu du coût de calcul élevé que représentent certaines simulations, des approches statistiques visent à émuler ces simulateurs coûteux partir d'un nombre limité de simulations. La quantification de l'incertitude peut alors être réalisée avec le modèle statique de substitution (ou métamodèle) pour un moindre coût de calcul. Cette étape optionnelle d'apprentissage statistique ne fait pas à proprement dit partie du traitement des incertitudes mais elle s'avère souvent essentielle pour permettre sa mise en œuvre pratique.

Verrous scientifiques et objectifs de la thèse

La maîtrise des risques et des incertitudes dans l'éolien est un enjeu majeur pour le groupe EDF en tant qu'exploitant. Cette thèse vise à adapter et appliquer, sur un cas d'usage issu de l'éolien en mer, une démarche globale de traitement des incertitudes. Ainsi, ce cas d'usage soulève des verrous scientifiques associés à ses particularités qui peuvent être décrites comme suit :

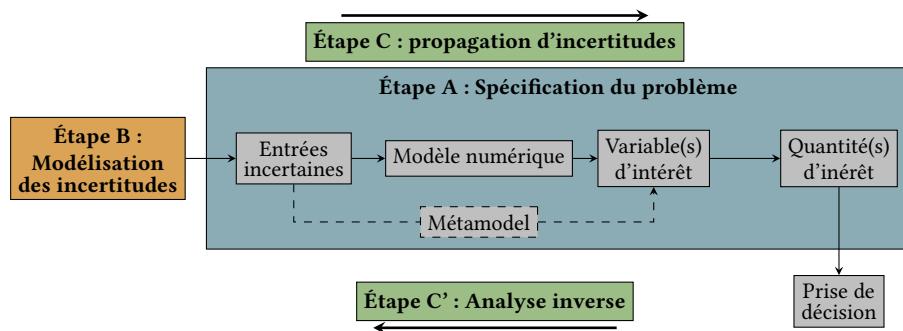


FIGURE E.1 Schéma générique de la quantification des incertitudes (De Rocquigny et al. (2008), adapté par Ajenjo (2023))

- Le code de simulation numérique autour duquel les travaux sont réalisés est constitué d'une chaîne de codes de calcul, exécutés en série. Cette chaîne s'articule en trois étapes : d'abord une génération temporelle et stochastique d'un champ de vitesse de vent et de houle, puis la simulation du comportement hydro-aéro-servo-élastique de l'éolienne et enfin une phase d'agrégation des résultats temporels pour obtenir des quantités d'intérêt scalaires ;
- La complexité de cet outil de calcul scientifique ainsi que le coût de calcul unitaire élevé (de l'ordre de 20 minutes par simulation) nécessite l'utilisation de méthodes d'échantillonnage performantes, ainsi que des systèmes de calcul haute performance. En plus de la complexité liée au modèle numérique, la modélisation des incertitudes en entrée présente, elle aussi, des difficultés. En effet, la loi conjointe des conditions environnementales liées à un site comporte une structure de dépendance complexe à capturer et à modéliser. L'étape d'inférence vis-à-vis des grandes quantités de données mesurées est d'autant plus importante que sa qualité impacte directement les conclusions de la propagation d'incertitudes.

Afin d'appliquer le schéma global de traitement des incertitudes au cas éolien, cette thèse vise à répondre aux problématiques suivantes :

- Q1.** *Comment précisément modéliser la structure de dépendance complexe associée aux lois conjointes de conditions environnementales ?* (⇒ Étape B)
- Q2.** *Comment réaliser une propagation d'incertitudes au travers d'une chaîne de simulation numérique coûteuse, uniquement basée sur une description empirique (données mesurées) des incertitudes en entrée ?* (⇒ Étape C)
- Q3.** *Comment estimer des probabilités d'événements rares associées à la ruine de structures éoliennes en mer ?* (⇒ Étape C)
- Q4.** *Comment évaluer et interpréter la sensibilité des entrées incertaines vis-à-vis des quantités d'intérêt liées à la fiabilité des structures (analyse de sensibilité fiabiliste) ?* (⇒ Étape C')

Les sections suivantes résument les travaux de thèse, tout en respectant la structure du manuscrit.

Résumés des chapitres relatifs à l'état de l'art des méthodes et outils mis en œuvre dans la thèse

Les deux premiers chapitres relateront l'état de l'art dans le domaine du traitement des incertitudes et de la modélisation numérique des systèmes éoliens.

Chapitre 1 – Traitement des incertitudes en simulation numérique

Ce chapitre vise à présenter un état de l'art concis des différentes thématiques en quantification des incertitudes (Sullivan, 2015). Après un rappel de quelques prérequis mathématiques, l'étape de spécification du modèle numérique (considéré comme étant une boîte-noire), ainsi que les variables d'entrée et de sortie est détaillée. Les différents types et sources d'incertitudes sont ensuite présentés, ainsi que leur modélisation dans un cadre probabiliste. La propagation des incertitudes dépend de la nature des quantités d'intérêt estimées, ainsi, une section aborde les méthodes de propagation pour l'étude en tendance centrale et une autre s'intéresse aux problèmes d'estimation de probabilités d'événements rares (statistiques liées aux queues de distributions). La section dédiée à la tendance centrale présente des méthodes d'intégration numérique, d'échantillonnage et de planification d'expériences (Fang et al., 2018). Celle consacrée aux probabilités d'événements rares présente des méthodes classiques issues du domaine de la fiabilité des structures (Lemaire et al., 2009; Morio and Balesdent, 2015).

Ce chapitre aborde également les principales méthodes d'analyse de sensibilité globale (Da Veiga et al., 2021). Ce domaine divise ses méthodes en deux grandes classes : les méthodes de criblage et les mesures d'importance. D'une part, les techniques de criblage, généralement mises en œuvre dans les problèmes de grande dimension, visent à identifier les variables n'ayant qu'un faible impact sur la variabilité de la sortie d'intérêt. D'autre part, les mesures d'importances visent, quant à elles, à attribuer de manière quantitative, pour chaque variable d'entrée, une part de variabilité de la sortie, permettant de proposer un classement des variables en fonction de leur influence.

Finalement, ce chapitre présente un panorama des familles de métamodèles communément utilisés en quantification des incertitudes (Forrester et al., 2008). Une attention particulière est apportée à la régression par processus gaussiens qui revient à conditionner un processus gaussien par un ensemble d'observations du code de simulation numérique. Une fois conditionné, le processus gaussien apporte une information plus riche que d'autres types de métamodèles. En effet, cette méthode propose conjointement un métamodèle (un prédicteur, ou moyenne du processus), et une fonction d'erreur (variance du processus). Certaines méthodes itératives (dites « actives ») exploitent cette information complémentaire pour enrichir progressivement le métamodèle et améliorer sa prédictivité. Ces techniques ont connu un franc succès dans les années 90 pour résoudre des problèmes d'optimisation de fonctions coûteuses (Jones et al., 1998). Depuis, leur utilisation s'est étendue à la résolution de problèmes de fiabilité des structures (Echard et al., 2011).

Chapitre 2 – Introduction à la modélisation et la conception de systèmes éoliens

La simulation d'une éolienne en mer implique la modélisation de plusieurs physiques en interaction avec des conditions environnementales de nature aléatoire. Ce chapitre introduit premièrement les méthodes spectrales utilisées pour générer des champs de vitesse de vent et de houle en appliquant des transformées de Fourier inverses (par exemple implémentées dans l'outil TurbSim ([Jonkman, 2009](#))). Ces champs de vitesses de vent simulés alimentent par la suite un outil de simulation multi-physique des éoliennes. Cette simulation intègre une modélisation simplifiée des interactions entre fluides et structures (méthode "BEMT" pour *blade element momentum theory*), une modélisation dynamique de la structure par des éléments finis de type poutre et une modélisation du contrôle-commande de l'éolienne [Milano \(2021\)](#). Ce code numérique produit en sortie des séries temporelles de plusieurs grandeurs physiques décrivant le comportement du système.

Cette thèse s'intéresse particulièrement à l'évaluation probabiliste du dommage en fatigue des structures éoliennes. Le dommage en fatigue est un phénomène qui détériore les propriétés mécaniques d'un matériau suite à sa sollicitation via un grand nombre de contraintes cycliques de faible amplitude. A l'heure actuelle, les standards [IEC-61400-1 \(2019\)](#); [DNV-ST-0437 \(2016\)](#) recommandent l'utilisation de coefficients de sécurité déterministes pour faire face à ce mode de défaillance. Une approche probabiliste permet d'enrichir l'analyse et parfois de mettre en évidence le conservatisme des marges de sûreté. Plusieurs travaux récents se sont intéressés à cette thématique en abordant des angles méthodologiques différents ([Huchet, 2019](#); [Lataniotis, 2019](#); [Cousin, 2021](#); [Hirvoas, 2021](#); [Petrovska, 2022](#)).

Dans ce contexte, ce chapitre liste les paramètre d'entrée de la chaîne de calcul considérés comme incertains par la suite. Ces variables aléatoires sont regroupées en deux groupes : le vecteur aléatoire lié à l'environnement (par exemple : la vitesse moyenne du vent, l'écart-type de la vitesse du vent, la direction du vent, la hauteur de houle, la période de houle, et la direction de houle), et le vecteur aléatoire lié au système (par exemple : l'erreur de d'alignement au vent du contrôleur, la rigidité du sol, les paramètres des courbes de calcul de fatigue).

Résumés des chapitres relatifs aux contributions méthodologiques et apports vis-à-vis des applications

Après avoir dressé l'état de l'art sur ce sujet, les prochains chapitres du manuscrit présentent les nouvelles contributions de la thèse. D'un point de vue méthodologique, un objet mathématique servira de fil conducteur au cours de ces travaux. La *maximum mean discrepancy* (MMD) [Oates \(2021\)](#) est une mesure de dissimilarité entre des lois de probabilité basée sur des noyaux qui est utilisée dans des contextes différents (tests statistiques [Gretton et al. \(2006\)](#), analyse de sensibilité [Da Veiga \(2015\)](#), échantillonnage [Pronzato and Zhigljavsky \(2020\)](#), etc.).

Chapitre 3 – Quantification des perturbations induites par les effets de sillage au sein d'un parc éolien

Ce chapitre étudie les perturbations sur les conditions environnementales à l'intérieur d'une ferme éolienne en mer induites par les effets de sillage (*wake effect* en anglais) [Larsen et al. \(2008\)](#). Un parc éolien en mer théorique au large de la côte sud de la Bretagne est considéré comme cas d'usage, et un modèle numérique simulant le sillage de ce parc est exploité. Ce modèle donne une prédition analytique du déficit en vitesse de vent et de la turbulence créés par le sillage, en tenant compte de l'influence de la position des flotteurs en raison des forces moyennes du vent. Une propagation de l'incertitude sur le modèle de sillage est réalisée, en considérant la loi conjointe des conditions environnementales ambiantes en entrée. Au final une distribution environnementale perturbée par le sillage est simulée pour chaque éolienne. Une mesure de dissimilarité (la MMD) est utilisée pour comparer les distributions perçues par chaque éolienne. Cette quantité permet de regrouper les éoliennes (phase de *clustering*) exposées à des conditions environnementales similaires, entraînant une réponse structurelle identiques. Compte tenu du coût de calcul élevé des simulations aéro-servo-hydro-élastiques des éoliennes en mer, cette étude préalable permet de réaliser une analyse de fiabilité à l'échelle d'une ferme éolienne sans répéter l'analyse pour chaque turbine. En fin de compte, seules quatre classes sont retenues pour représenter une ferme de 25 éoliennes. Ce travail a mené à la publication suivante :

☞ A. Lovera, [E. Fekhari](#), B. Jézéquel, M. Dupoiron, M. Guiton and E. Ardillon (2023). "Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis". In : *Journal of Physics : Conference Series (WAKE 2023)*, Visby, Sweden.

Chapitre 4 – Méthodes à noyaux pour l'estimation de la tendance centrale

Ce chapitre présente une utilisation d'une mesure de dissimilarité basée sur des noyaux (la MMD) pour échantillonner suivant une loi de probabilité, méthode du "*kernel herding*" introduite par [Chen et al. \(2010\)](#). Cette technique de quadrature appartient à la famille dite des « quadratures Bayésiennes » ([Briol et al., 2019](#)) qui s'interprètent comme une généralisation des méthodes de quasi-Monte Carlo ([Li et al., 2020](#)). Le *kernel herding* est présenté en détails et plusieurs expériences numériques sur des fonctions analytiques illustrent son intérêt.

Les propriétés de cette méthode sont mises en valeur via une application industrielle dédiée à l'estimation de la moyenne du dommage en fatigue d'une structure éolienne. Cette quantité est déterminante dans le dimensionnement et la certification des éoliennes. Toutefois, son estimation par le biais de simulations numériques s'avère coûteuse. L'étude est réalisée sur un modèle d'une éolienne posée appartenant à une ferme installée en mer du Nord. Les incertitudes des conditions environnementales en entrée sont inférées sur des données mesurées in-situ.

Dans ce cadre, une comparaison numérique avec un échantillonnage Monte Carlo et quasi-Monte Carlo révèle la performance et les avantages pratiques du *kernel herding*. Cette méthode

permet notamment sous-échantillonner directement depuis une base de données environnementales importante, sans effectuer d'inférence (étape B). Ce travail a mené à la publication et au développement informatique suivant :

- ☞ E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature”. In : *Data-Centric Engineering*, In press.
- ☞ Le module Python `ctbenchmark` standardise les expériences numériques liées à la quadrature Bayésienne et est disponible sur la plateforme GitHub.
- ☞ Le module Python `copulogram` propose une nouvelle représentation graphique de jeux de données multivariés et est disponible sur la plateforme de téléchargement Pypi.

Chapitre 5 – Méthodes à noyaux pour la validation de métamodèles

Ce chapitre propose une utilisation des méthodes d'échantillonage à base de noyaux dans le cadre de la validation de modèles d'apprentissage (ou métamodèles). L'estimation de la prédictivité des modèles d'apprentissage supervisé nécessite une évaluation de la fonction apprise sur un ensemble de points de test (non utilisés par lors de l'apprentissage). La qualité de l'évaluation dépend naturellement des propriétés de l'ensemble de test et de la statistique d'erreur utilisée pour estimer l'erreur de prédiction. Cette contribution propose d'une part d'utiliser des méthodes d'échantillonnage pour sélectionner de manière “optimale” un ensemble de test et d'autre part présente un nouveau critère de prédictivité qui pondère les erreurs observées pour obtenir une estimation globale de l'erreur. Une comparaison numérique entre plusieurs méthodes d'échantillonnage basées sur des approches géométriques ([Shang and Apley, 2020](#)) ou sur des méthodes à noyaux ([Chen et al., 2010](#); [Mak and Joseph, 2018](#)) est effectuée. Nos résultats montrent que les versions pondérées des méthodes à noyau offrent des performances supérieures. Une application aux efforts mécaniques simulées par un modèle éolien en mer est également présentée. Cette expérience illustre la pertinence pratique de cette technique comme alternative efficace aux techniques coûteuses de validation croisée. Ce travail a mené à la publication et au développement informatique suivant :

- ☞ E. Fekhari, B. Iooss, J. Muré, L. Pronzato and M.J. Rendas (2023). “Model predictivity assessment : incremental test-set selection and accuracy evaluation”. In : *Studies in Theoretical and Applied Statistics*, pages 315–347. Springer.
- ☞ Le module Python `otkerneldesign` est développé en collaboration avec J.Muré. Ce module dédié à la quadrature Bayésienne est documenté et disponible sur la plateforme de téléchargement Pypi.

Chapitre 6 – Estimation non-paramétrique de probabilités d'événements rares

L'estimation de probabilités d'événements rares est un problème courant dans la gestion des risques industriels, notamment dans le domaine de la fiabilité des structures Chabridon (2018). Pour ce faire, plusieurs techniques ont été proposées pour surmonter les limites connues de la méthode de Monte Carlo. Parmi elles, la méthode de “*subset simulation*” Au and Beck (2001) est une technique qui repose sur la décomposition de la probabilité de l'événement rare en un produit de probabilités conditionnelles moins rares (donc plus simples à estimer) associées à des événements de défaillance imbriqués. Cependant, cette technique repose sur la simulation conditionnelle à base de méthodes de Monte Carlo par chaînes de Markov (MCMC). Ces algorithmes permettent, à la convergence, de simuler selon la densité cible. Cependant, en pratique, ils produisent souvent des échantillons non indépendants et identiquement distribués (i.i.d.) en raison de la corrélation entre les chaînes de Markov. Ce chapitre propose une autre méthode pour échantillonner conditionnellement aux événements de défaillance imbriqués afin d'obtenir des échantillons dont la propriété d'être i.i.d. est préservée. La propriété d'indépendance des échantillons est particulièrement pertinente pour exploiter ces mêmes échantillons pour une analyse de sensibilité fiabiliste. L'algorithme proposé repose sur l'inférence non-paramétrique de la distribution conjointe conditionnelle en utilisant une estimation par noyau des marginales combinée à une inférence de la dépendance à l'aide de la copule empirique de Bernstein Sanctetta and Satchell (2004). L'algorithme appelé “*Bernstein adaptive nonparametric conditional sampling*” (BANCS) est comparée à la méthode du *subset simulation* pour plusieurs problèmes de fiabilité des structures. Les premiers résultats sont encourageants, mais le contrôle du biais de l'estimateur doit être plus amplement investigué. Ce travail a mené à la publication et au développement informatique suivant :

- ☞ E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Bernstein adaptive nonparametric conditional sampling : a new method for rare event probability estimation”. In : *Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)*, Dublin, Ireland.
- ☞ Le module Python `bancs` propose une implémentation de la méthode BANCS et est disponible sur la plateforme GitHub.

Chapitre 7 – Analyse de sensibilité fiabiliste adaptative

Ce chapitre traite d'analyse de sensibilité pour des mesures de risque (par exemple, une probabilité d'événement rare). L'analyse de sensibilité globale Da Veiga et al. (2021) attribue à chaque variable (ou groupe de variable) une part de variabilité globale de la sortie (le plus souvent à l'aide d'une décomposition fonctionnelle de la variance de la sortie). Cependant, les variables ayant un impact sur des quantités liées à une queue de distribution peuvent être très

différentes que celles ayant un impact sur la variabilité globale (pondérée par le poids associé au centre de la distribution). L’analyse de sensibilité fiabiliste (en anglais “*reliability-oriented sensitivity analysis*”, Chabridon (2018)) permet d’expliquer le rôle des entrées vis-à-vis de probabilités d’événements rares. L’idée de ce chapitre est d’étudier l’évolution de la sensibilité au fur et à mesure que l’échantillonnage se rapproche de l’événement rare. Cette analyse permet ainsi d’exploiter les paquets successifs d’échantillons conditionnels générés par l’algorithme BANCS (présenté dans le Chapitre 6). En post-traitement de l’estimation de la probabilité d’un événement rare, cette approche utilise une mesure d’importance à base de noyaux, nommée *Hilbert-Schmidt Independence Criterion*, pour évaluer la dynamique de la sensibilité fiabiliste Marrel and Chabridon (2021).

Conclusion

En résumé, cette thèse aborde plusieurs aspects du traitement des incertitudes à l’aide d’outils mathématiques à base de noyaux et présente un débouché industriel lié à l’enjeu de la maîtrise des risques des actifs éoliens en mer. Les contributions de cette thèse ont été principalement réalisées dans le cadre du projet européen HIPERWIND (*Highly advanced Probabilistic design and Enhanced Reliability methods for high-value, cost-efficient offshore wind.*), et de l’ANR INDEX (INcremental Design of EXperiments). Le sous-sections ci-après résument les communications, les publications dans revue à comité de lecture et les développements informatiques.

Communications et publications dans revues à comité de lecture

Book Chap.	<u>E. Fekhari</u> , B. Iooss, J. Muré, L. Pronzato and M.J. Rendas (2023). "Model predictivity assessment : incremental test-set selection and accuracy evaluation". In : <i>Studies in Theoretical and Applied Statistics</i> , pages 315–347. Springer.
Jour. Pap.	<u>E. Fekhari</u> , V. Chabridon, J. Muré and B. Iooss (2023). "Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature". In : <i>Data-Centric Engineering</i> .
Int. Conf. Pap.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool". In : <i>Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)</i> , Lisbon, Portugal. (Paper & Talk)
	<u>E. Fekhari</u> , V. Chabridon, J. Muré and B. Iooss (2023). "Bernstein adaptive nonparametric conditional sampling : a new method for rare event probability estimation" ¹ . In : <i>Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)</i> , Dublin, Ireland. (Paper & Talk)
	E. Vanem, <u>E. Fekhari</u> , N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2023). "A joint probability distribution model for multivariate wind and wave conditions". In : <i>Proceedings of the ASME 2023 42th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2023)</i> , Melbourne, Australia. (Paper)
	A. Lovera, <u>E. Fekhari</u> , B. Jézéquel, M. Dupoiron, M. Guiton and E. Ardillon (2023). "Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis". In : <i>Journal of Physics : Conference Series (WAKE 2023)</i> , Visby, Sweden (Paper)
Int. Conf. Short Abs.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Numerical Studies of Bayesian Quadrature Applied to Offshore Wind Turbine Load Estimation". In : <i>SIAM Conference on Uncertainty Quantification (SIAM UQ22)</i> , Atlanta, USA. (Talk)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Model predictivity assessment : incremental test-set selection and accuracy evaluation". In : <i>22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)</i> , Trondheim, Norway. (Talk)
Nat. Conf.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Kernel-based quadrature applied to offshore wind turbine damage estimation". In : <i>Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)</i> , Clermont-Ferrand, France (Poster)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2023). "Rare event estimation using nonparametric Bernstein adaptive sampling". In : <i>Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)</i> , Le Croisic, France (Talk)
Invited Lec.	Le Printemps de la Recherche 2022, Nantes, France. "Traitement des incertitudes pour la gestion d'actifs éoliens". (Talk)
	Journées Scientifiques de l'Eolien 2024, Saint-Malo, France. "Evaluation probabiliste de la fiabilité en fatigue des structures éoliennes en mer". (Talk)

¹This contribution was rewarded by the "CERRA Student Recognition Award"

Développements informatiques open source

`otkerneldesign`²

- Ce module Python génère des échantillons (aussi appelés plans d’expérience) en utilisant des méthodes à base de noyaux comme le *kernel herding* et les *support points*. Une implementation tensorisée qui améliore grandement les performances est également proposée. En complément, une méthode de pondération “optimale” à l’aide de quadrature Bayésienne est proposée.
- Ce module est développé en collaboration avec J. Muré, est documenté et disponible sur la plateforme de téléchargement Pypi.

`bancs`³

- Ce module Python offre une implémentation de la méthode “*Bernstein Adaptive Nonparametric Conditional Sampling*” mentionnée en Section E.
- Ce module est disponible sur la plateforme de GitHub et son utilisation est illustrée par des exemples analytiques.

`ctbenchmark`⁴

- Ce module Python standardise les comparaisons numériques réalisés pour étudier les méthodes de quadrature Bayésiennes.
- Le module et les expériences numériques sont disponibles sur un dépôt GitHub.

`copulogram`⁵

- Ce module Python propose une nouvelle représentation graphique de jeux de données multivariés appelée *copulogram*.
- Ce module, développé en collaboration avec V. Chabridon, est disponible sur la plateforme de téléchargement Pypi.

²Documentation :<https://efekhari27.github.io/otkerneldesign/master/>

³Dépôt: <https://github.com/efekhari27/bancs>

⁴Repository: <https://github.com/efekhari27/ctbenchmark>

⁵Repository: <https://github.com/efekhari27/copulogram>

