UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

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, Zurich
Invités	Anaïs LOVERA	Ingénieure-chercheur	EDF R&D, Saclay
	Joseph MURÉ	Ingénieur-chercheur	EDF R&D, Chatou

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 fullfilment 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	Research Engineer	EDF R&D, Chatou
Reviewers	Pr. Franck SCHOEFS	Professor	Nantes Université, Nantes
	Pr. Daniel STRAUB	Professor	TUM, Munich
Examiners	Pr. Mireille BOSSY	Research Director	INRIA, Sophia-Antipolis
	Dr. Sébastien DA VEIGA	Associate Professor	ENSAI, Rennes
	Pr. Bruno SUDRET	Professor	ETH, Zurich
Invitees	Dr. Anaïs LOVERA	Research Engineer	EDF R&D, Saclay
	Dr. Joseph MURÉ	Research Engineer	EDF R&D, Chatou

Funding Statement. Part of this thesis was the result of the HIPERWIND project which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101006689. Part of this work was also supported by project INDEX (INcremental Design of EXperiments) ANR-18-CE91-0007 of the French National Research Agency (ANR).

Acknowledgements

[And I would like to acknowledge]

Abstract

Offshore wind energy is one of the ways to reduce the share of fossil fuels in the global electricity mix. This technology benefits from more consistent winds than the onshore one, mainly due to the absence of terrain roughness. Operating offshore also allows the installation of larger and more powerful wind turbines, which poses several scaling issues about port logistics, the demand for critical natural resources, and sustainable end-of-life processes.

Offshore wind turbines are dynamic systems interacting with a highly uncertain environment. Uncertainty quantification of the multi-physics numerical models used to simulate them is therefore essential to propose risk-informed design and operation. However, developing a dedicated uncertainty quantification strategy for these systems raises numerous questions and requires coupling data with multi-physics numerical models.

This thesis first addresses the problems related to the multivariate probabilistic modeling of offshore environmental conditions. A semiparametric approach is suggested, mixing parametric methods for marginals fitting with the empirical Bernstein copula to fit the complex dependence structure among environmental variables. After defining a probabilistic model of the ambient metocean conditions, the perturbations caused by the turbines' wake effect are studied at the farm scale by creating clusters of similarly perturbed turbines. This preliminary clustering aims to reduce the number of loading studies at the farm scale (e.g., for fatigue assessment).

The second methodological axis of this work concerns uncertainty propagation for both central study and rare event estimation. As an alternative to the design load cases recommended by international standards for the mean cumulative damage estimation, the kernel herding method proved to be an efficient and flexible solution for given-data uncertainty propagation (i.e., directly subsampling from a large dataset without inference). This Bayesian quadrature method is also well-suited for the construction of test samples for the estimation of the mean predictivity of statistical learning models.

For rare event estimation, a new method incorporating a nonparametric copula into an adaptive importance sampling mechanism is proposed. This method displays equivalent results to splitting methods as the subset simulation while avoiding any Markov Chain Monte Carlo sampling and thus generating independent and identically distributed samples. Then, the fatigue reliability of an offshore wind turbine is studied with respect to uncertain environmental variables while considering other variables related to soil stiffness, yaw misalignment, stress-number of cycles curve, and critical damage resistance. The robustness of the estimated reliability is then studied using perturbed-law based indices. Finally, to ensure the reproducibility of the numerical results, most of the developments presented in this work are open source and documented.

Contents

List of Figures

List of Tables

List of Acronyms

AK Active Kriging. 58

AMISE Asymptotic Mean Integrated Squared Error. 40

BANCS Bernstein Adaptive Nonparametric Conditional Sampling. 143

BQ Bayesian Quadrature. 101

CDF Cumulative Distribution Function. 16

CE-AIS Cross-Entropy-based Adaptive Importance Sampling. 38

COV Coefficient Of Variation. 17

CPU Central processing unit. 91, 172

CSA Conditional Sensitivity Analysis. 152

DGSM Derivative-based Global Sensitivity Measures. 45

EBC Empirical Bernstein Copula. 69

FANOVA Functional ANalysis Of VAriance. 47

FORM First-Order Reliability Method. 33

FSSF Fully-Sequential Space-Filling. 115

GSA Global Sensitivity Analysis. 44

HPC High-Performance Computer. 15

HSIC Hilbert-Schmidt Independence Criterion. 50

i.i.d. Independent and identically distributed. 17

IPM Integral probability metrics. 189

IS Importance Sampling. 36

ISE Integrated Squared Error. 53

KDE Kernel Density Estimation. 40

KH Kernel Herding. 98

KL Kullback–Leibler divergence. 38

LHS Latin Hypercube Sampling. 27

LOO Leave-One-Out. 53

LSF Limit-State Function. 31

MC Monte Carlo. 22

MCMC Markov Chain Monte Carlo. 42

MMD Maximum Mean Discrepancy. 50

MPFP Most-Probable-Failure-Point. 33

NAIS Nonparametric Adaptive Importance Sampling. 41

OAT One At a Time. 45

OpenTURNS Open source initiative for the Treatment of Uncertainties, Risks'N Statistics. 15

PDF Probability Density Function. 16

PLI Perturbed-Law based Indices. 52

PLS Partial Least Squares. 133

PVA Predictive Variance Adequation. 55

QMC Quasi-Monte Carlo. 23

RCV Random Cross-Validation. 132

RKHS Reproducing Kernel Hilbert Space. 50

ROSA Reliability-Oriented Sensitivity Analysis. 8

SORM Second-Order Reliability Method. 33

SS Subset Simulation. 41

TSA Target Sensitivity Analysis. 152

UQ Uncertainty Quantification. 14

w.r.t. With respect to. 38

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 worldwide capacity of 56 GW in operation 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 questions emerge. To meet ambitious national and regional development targets, the wind energy industry must address various scaling issues, including port logistics, the critical demand for natural resources, and the lack of sustainable end-of-life processes. Furthermore, the field presents 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, with only partial knowledge about specific physical phenomena. Therefore, modeling and controlling the various sources of uncertainty associated with offshore wind energy could become a key success factor in this highly competitive industry.

Overall, the offshore wind industry needs methods for uncertainty management (Van Kuik et al., 2016) regarding design safety margins and operational asset management (at the component, wind turbine, and overall wind farm levels). For wind project developers, the primary focus is on improving the wind potential assessment of candidate geographic locations by combining various sources of information for 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 reaching its end of life has three options: *(i)* extend the operational life of assets, *(ii)* replace current wind turbines with newer models, *(iii)* 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 agencies and insurers to issue operating permits. To provide a rigorous and quantitative risk assessment, the generic methodology know as the *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

The field of computer experiments emerged after the formidable increase of processing power that occurred in the past decades. Numerical models can simulate complex system behavior based on initial conditions defined by the analyst. They 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.

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 among the 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 challenges 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, 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 this step is to identify and model all sources of uncertainty related to the input variables. Most of the time, uncertainty is modeled in the probabilistic framework (Sullivan, 2015), but other approaches could be considered (Ajenjo, 2023).

- **Step C – Uncertainty propagation**: This step propagates 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., a moment, a quantile, a rare event probability).

- **Step C' – Inverse analysis**: In this 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 techniques can be used to emulate these expensive simulators with a limited number of simulations. Uncertainty quantification can then be carried out using a "surrogate model" (or metamodel) to reduce the computational cost. This optional statistical learning often proves to be essential in practice to perform uncertainty quantification.

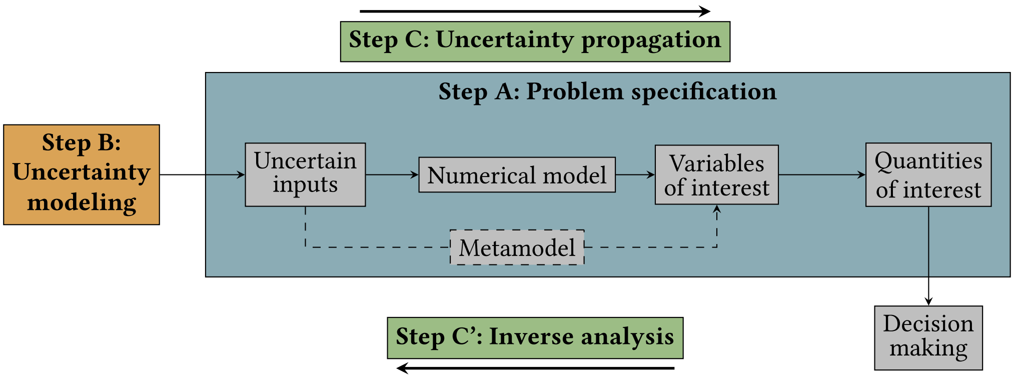


Figure 1 General uncertainty quantification framework (De Rocquigny et al., 2008, adapted by Ajenjo, 2023)

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 motivated the research and development division of EDF to participate in a EU founded project named HIPERWIND[1] (for "HIghly advanced Probabilistic design and Enhanced Reliability methods for high-value, cost-efficient offshore wind"). Collaborating with various partners from the academic sector (such as DTU, ETH Zurich, University of Bergen) and the industrial domain (such as DNV, IFPEN, EPRI) brought

[1]HIPERWIND web page: https://www.hiperwind.eu/

different perspectives on this topic. As a result, this thesis aims to adapt and apply the generic uncertainty quantification methodology to industrial offshore wind energy design and operation. In particular to assess the structural integrity with respect to fatigue damage (caused by a large number of small amplitude solicitations). 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 serie of numerical models executed sequentially. This chain is divided into three parts: first, stochastic wind and wave velocity spatio-temporal fields are generated, then the coupled hydro-aero-servo-elastic behavior of the wind turbine is simulated, and finally these data are post-processed to obtain scalar quantities of interest.
- The complexity of this simulator, along with the high unit computational cost (about 40 minutes per simulation), requires 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 (measured in situ) 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 study 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 recalls the principal tools for uncertainty quantification and introduces offshore wind turbine modeling and design. The second part presents the contributions of this thesis associated 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 – Uncertainty quantification in computer experiments. This chapter gives a brief over-view 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 risk and reliability analysis. 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). In general terms, this field divides its methods into two major classes: screening methods and importance measures. Screening techniques are typically applied in high-dimensional problems and 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 surrogate models commonly used in uncertainty quantification (Forrester et al., 2008). Special attention is given to Gaussian process regression, which consists in conditioning a Gaussian process on a set of observations from the numerical model. Once conditioned, the Gaussian process 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 surrogate model 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 structural reliability problems (Echard et al., 2011).

Chapter 2 – Introduction to wind turbine modeling and design. Simulating 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 e.g., the blade element momentum theory), dynamic modeling of the structure using flexible multibody methods, and modeling of wind turbine control systems (Burton et al., 2021). 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 low-amplitude cyclic stresses. Currently, wind energy standards recommend the use of deterministic safety factors to address this failure mode (DNV-ST-0437, 2016; IEC-61400-1, 2019). A probabilistic approach can enhance the analysis and might sometimes reveal conservative safety margins, as addressed in various methodological works (Huchet, 2019; Lataniotis, 2019; Petrovska, 2022).

In this context, this chapter enumerates the input variables of the calculation chain that are considered uncertain. These variables are grouped into two groups: those related to the environment (e.g., average wind speed, wind speed standard deviation, wind direction, significant wave height, wave period, and wave direction), and those 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 "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 allows to assess the 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. This chapter presents the use of the MMD (introduced in the previous chapter) 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 by standards 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. Overall, this method allows for direct subsampling from a large environmental database without the need for inference (step B).

Chapter 5 – Kernel-based surrogate model 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 – Adaptive rare event estimation using Bernstein copula. Estimating rare event probabilities is a common issue in industrial risk management, especially in the field of structural reliability (Morio and Balesdent, 2015). To overcome the well-known limitations of the Monte Carlo method, several techniques have been proposed 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. In practice, these algorithms produce correlated chains which converge asymptotically towards the targeted distribution. The issue in reliability analysis is that their convergence is usually not checked (e.g., Roy, 2020 reviews diagnostic tools for MCMC). In this chapter, another method using an adaptive importance sampling structure is proposed (Zhang, 1996), 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 nonparametric 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 first results are promising, and various perspectives could still improve this technique.

After proposing a new method for reliability 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 results 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 the previous chapter), the idea here 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 (Da Veiga, 2015; Marrel and Chabridon, 2021).

Chapter 7 – Application to wind turbine fatigue reliability and robustness. This chapter proposes a fully probabilistic reliability analysis of an offshore wind turbine's monopile foundation. Considering a set of variables related to the system, a surrogate model of the lifetime fatigue damage is built on a learning set gathering over 10^5 simulations of the offshore wind turbine. This huge computational effort was made possible by deploying a wrapper of the simulator on a high-performance computer. Using the surrogate model to emulate the costly wind turbine model, a nominal reliability analysis is performed. To complete this analysis, the robustness of the failure probability obtained is assessed by perturbating the inputs' distributions. This approach called the "perturbed-law based sensitivity indices" (Lemaître et al., 2015) reveals that the resistance variable has a primary role on the reliability.

Publications and communications

The research contributions in this manuscript are related to 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: *Studies in Theoretical and Applied Statistics*, pages 315–347. Springer.
Jour. Pap.	E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2024). "Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature". In: *Data-Centric Engineering*, In press.
	E. Vanem, E. Fekhari, N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2024). "A joint probability distribution for multivariate wind-wave conditions and discussions on uncertainties". In: *Journal of Offshore Mechanics and Arctic Engineering*, In press.
Int. Conf. Pap.	E. Fekhari, M. Baudin, V. Chabridon, and Y. Jebroun (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 (UNCECOMP 2021)*, Athens, Greece. (Paper & Talk)
	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: *Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)*, 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"[2]. In: *Proceedings of the 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)*, Dublin, Ireland. (Paper & Talk)
	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. (Paper)
	E. Vanem, Ø. Lande, E. Fekhari (2024). "A joint probability distribution model for multivariate wind and wave conditions". In: *Proceedings of the ASME 2024 43th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2024)*, Singapore. (Paper to appear)
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: *SIAM Conference on Uncertainty Quantification (SIAM UQ22)*, Atlanta, USA. (Talk)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). "Model predictivity assessment: incremental test-set selection and accuracy evaluation". In: *22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)*, 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: *Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)*, Clermont-Ferrand, France. (Poster)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2023). "Rare event estimation using nonparametric Bernstein adaptive sampling". In: *Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)*, 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)

Numerical developments

Several implementations developed in this thesis are available on multiple platforms, allowing the reader to reproduce most of the numerical results presented in this manuscript:

`otkerneldesign` [3]	• This Python package generates designs of experiments based on kernel methods such as kernel herding (Chen et al., 2018a) and support points (Mak and Joseph, 2018). 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.
`bancs` [4]	• This Python package proposes an implementation of the "Bernstein Adaptive Nonparametric Conditional Sampling" method for rare event estimation. • This Python package is available on the PyPI platform and is illustrated with examples and analytical benchmarks.
`ctbenchmark` [5]	• 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.
`copulogram` [6]	• 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.

[2] Rewarded by the "CERRA Student Recognition Award" (https://icasp14.com/presenter/awards/)
[3] Documentation: https://efekhari27.github.io/otkerneldesign/master/
[4] Repository: https://github.com/efekhari27/bancs
[5] Repository: https://github.com/efekhari27/ctbenchmark
[6] Repository: https://github.com/efekhari27/copulogram

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

The progress of computer simulation gradually allows the 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 at a reasonable computational cost.

This chapter presents the essential tools and methods from the generic UQ framework, including elements partially inspired from Sullivan (2015). It is structured as follows: Section 1.2 describes the model specification step; Section 1.4 presents a classification of the input uncertainties and the probabilistic framework to model them; Sections 1.5 and 1.6 introduce various methods to propagate the input uncertainties through the numerical model for different purposes; Section 1.7 presents the main inverse methods to perform sensitivity analysis in our framework; Finally, Section 1.8 introduces the concept of surrogate models to emulate a model by realizing statistical learning on a limited dataset. Additionally, numerical examples of the notable methods are implemented during this chapter using `OpenTURNS`, a Python package for uncertainty quantification.

1.2 Black-box model specification

In the computer experiments context, UQ is performed around an input-output numerical simulation model. This numerical model, or code, is hereafter viewed as *black-box* since the physics equations remain unchaged. 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 can actually be a chain of codes executed in sequence to obtain a variable of interest, making their evaluation more complex.

While simulation models are often deterministic, they can 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 *verification & validation* phase, to quantify their domain of validity and predictive accuracy (Damblin, 2015).

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 (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. Note that UQ method were adapted to other types of outputs (see e.g., for time series outputs the work of Lataniotis, 2019, and for other functional outputs the work of Auder et al., 2012; Rollón de Pinedo et al., 2021). In the following, the numerical model $\mathcal{M}$, with its respective input $\mathcal{D}_\mathbf{x}$ and output domains $\mathcal{D}_{\mathbf{y}}$ is defined as:

$$\mathcal{M} : \left| \begin{array}{rcl} \mathcal{D}_{\mathbf{x}} \subseteq \mathbb{R}^d & \longrightarrow & \mathcal{D}_{\mathbf{y}} \subseteq \mathbb{R} \\ \mathbf{x} & \longmapsto & y. \end{array} \right. \tag{1.1}$$

Unlike data-scientists who work with the typical machine learning input-output dataset framework, UQ analysts can evaluate this numerical model at any input point. However, numerical simulations often come with a large computational cost. Therefore, UQ methods should be efficient and require as few simulations as possible. To solve this issue, surrogate models (or metamodels) are statistical learning algorithms emulating the costly numerical model, which can be used to make UQ studies more tractable. 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 often necessitates 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-perfomance computer (HPC) deployment. Once the model is specified, a critical step in UQ is enumerating the input uncertainties and building their associated mathematical model.

1.3 Uncertainty quantification practice with OpenTURNS

OpenTURNS[1] ("Open source initiative for the Treatment of Uncertainties, Risks'N Statistics") is a high-performance Python library dedicated to UQ (Baudin et al., 2017). It is developed by industrial researchers from a consortium gathering EDF R&D, Airbus Group, PHIMECA Engineering, IMACS and ONERA. It combines high performance using C++ programming with high accessibility through a Python application programming interface. Overall, this open source library provides tools for every steps of the UQ framework (e.g., UQ, 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. Furthermore, no-code users can benefit from OpenTURNS's free-download graphical user interface software, named Persalys[2]. Finally, as most developments in this thesis use OpenTURNS, minimial working examples associated to the main methodological concepts are available in Appendix D.

1.4 Identifying and modeling the uncertain inputs

1.4.1 Sources of the input uncertainties

After defining a numerical model, the analyst should list the sources of uncertainties. Authors as Thunnissen (2005) proposed to classify them for practical purposes into two groups:

[1] OpenTURNS installation guide and documentation are available at: https://openturns.github.io/www/
[2] Persalys free-download graphical user interface is available at: https://www.persalys.fr/obtenir.php

- aleatory uncertainty regroups the uncertainties arising from natural randomness (e.g., soil properties). These uncertainties are qualified as irreducible since the analyst 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., geometry of a component). Unlike 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.

1.4.2 Modeling uncertain inputs with the probabilistic framework

Uncertainties are traditionally modeled with objects from the probability theory. Considering a probability space denoted by $(\Omega, \mathcal{A}, \mathbb{P})$, the *random vector* $\mathbf{X}$ (i.e., multivariate random variable) is a measurable function defined as:

$$\mathbf{X} : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{D}_{\mathbf{x}} \subseteq \mathbb{R}^d \\ \omega & \longmapsto & \mathbf{X}(\omega). \end{array} \right. \tag{1.2}$$

In the following, the random vector $\mathbf{X}$ will be assumed to be a squared-integrable function against the measure $\mathbb{P}$ (i.e., $\int_\Omega |\mathbf{X}(\omega)|^2 \, \mathrm{d}\mathbb{P}(\omega) < \infty$). 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 by $\mathbb{P}_{\mathbf{X}}$ and defined as follows:

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

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

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \mathbb{P}_{\mathbf{X}}\left(]-\infty, x_1] \times \cdots \times]-\infty, x_d]\right). \tag{1.4}$$

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) $f_{\mathbf{X}} : \mathcal{D}_{\mathbf{x}} \longrightarrow \mathbb{R}_+$. This means that $\forall B \subset \mathcal{A}$, $\int f_{\mathbf{X}}(\mathbf{x}) \mathbb{1}_{\{B\}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \mathbb{P}_{\mathbf{X}}(B)$.

The expected value of a random vector $\mathbb{E}[\mathbf{X}]$, is given by:

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

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

$$\mathrm{Var}(X_i) = \mathbb{E}\left[X_i - \mathbb{E}[X_i]\right], \tag{1.6}$$

and a covariance describing their joint variability:

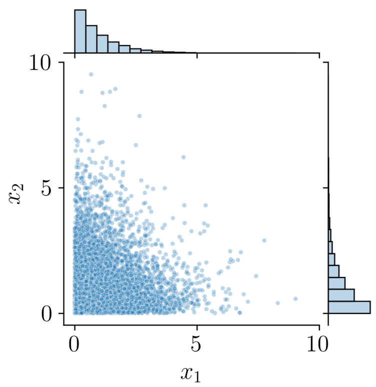
$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])\left(X_j - \mathbb{E}[X_j]\right)\right] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2. \tag{1.7}$$

The *standard deviation* $\sigma_{X_j} = \sqrt{\mathrm{Var}(X_j)}$ and *coefficient of variation* (COV) $\delta_{X_j} = \frac{\sigma_{X_j}}{\mu_{X_j}}$ are two quantities widely used in practice.

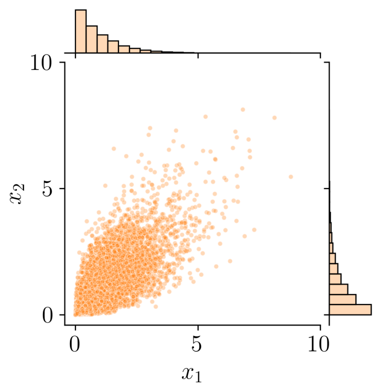
Finally, note that alternative theories exist to mathematically model uncertainties. For example, imprecise probability theory allows more general modeling of the uncertainties (Beer et al., 2013; Schöbi, 2019; Ajenjo, 2023). Such approaches become useful when dealing with very limited information (e.g., expert elicitation), but they will not be used in this thesis.

1.4.3 Joint input probability distribution

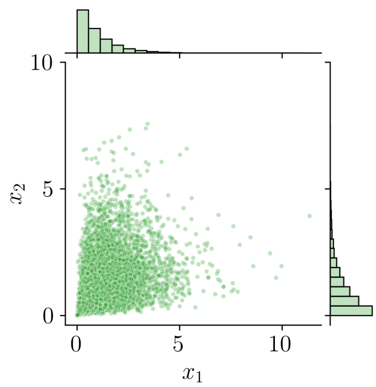
This section introduces various techniques to model and infer a joint multivariate probability distribution. To illustrate the importance of this step, Fig. 1.1 represents three i.i.d. samples from three bivariate distributions sharing the same marginals (e.g., here two exponential distributions) but different dependence structures. Judging from this example, one can assume that the joint distribution results from the composition of the marginals, and an application governing the dependence between them.



(a) Independent copula.



(b) Normal copula.

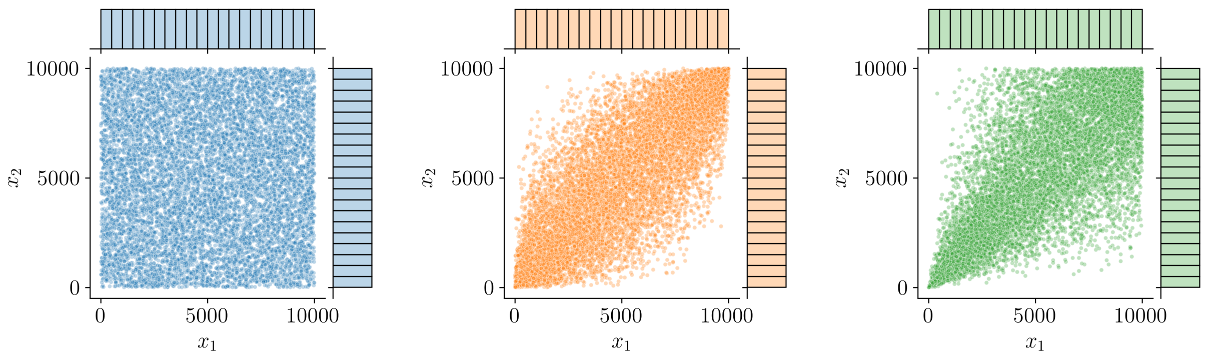


(c) Clayton copula.

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

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 $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\} \in \mathcal{D}_{\mathbf{x}}^n$. The corresponding ranked sample is defined as: $\mathbf{R}_n = \left\{\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(n)}\right\}$, where[3] $r_j^{(i)} = \sum_{l=1}^n \mathbb{1}_{\left\{x_j^{(l)} \leq x_j^{(i)}\right\}}$, $\forall j \in \{1, \ldots d\}, i \in \{1, \ldots n\}$. Ranking a multivariate dataset allows us to isolate the dependence

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



(a) Independent copula. (b) Normal copula. (c) Clayton copula.

Figure 1.2 Samples in the ranked space 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 over the whole space while the others present dependence patterns.

Sklar's theorem (Sklar, 1959) states that the multivariate distribution of any random vector can be broken down into two objects. First, a set of univariate marginal distributions describing the behavior of each individual variables; second, a function describing the dependence structure between all variables: this function is called *copula*.

Theorem 1 (Sklar's theorem, Sklar, 1959)**.** *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 funtion $C : [0,1]^d \longrightarrow [0,1]$, such that:*

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

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

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

where c is the density of the copula, sometimes also called "copula" by abuse of terminology.

The reader might refer to Durante and Sempi (2015, Chap. 2) for three different mathematical proofs of this result. 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:

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

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 essential 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 paraletric copulas exists (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 support restrictive? If so, specific families of parametric distributions with restrictive support can be chosen or truncation can be applied;
- Is there a dependence structure among the input variables? Does it concern all the variables together or only some groups of variables?
- Is the dependence structure complex (e.g., stronger dependence in the tail of the distribution)? Transforming the dataset in the ranked space gives an empirical description of the dependence.

Techniques designed to estimate marginal distributions are available in Appendix A. In addition, a nonparametric method are introduced in Chapter 2 to infer a copula: the "empirical Bernstein copula". The adequation between a fitted probabilistic model and a dataset should be validated, therefore Appendix A lists visual and quantitative tools for univariate goodness-if-fit evaluation.

> **`OpenTURNS` 1** (Bivariate distribution)**.** The Python code available in Appendix D proposes a minimalistic `OpenTURNS` example of probabilistic uncertainty modeling. Figures illustrating the present section may be reproduced, using the `OpenTURNS` scripts available on GitHub[4].

1.5 Uncertainty propagation for central tendency study

The previous section aimed at building a probabilistic model of the uncertainties considering the availableknowledge. 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

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

distribution. This approach contrasts with the uncertainty propagation committed to rare event probability estimation, which will be introduced in Section 1.6 (e.g., used to assess reliability).

The difficulties related to any uncertainty propagation task mostly arise from the practical properties of the numerical model. Its potential high dimension, irregularity and nonlinearity each represents a challenge. Such studies rely on a finite number of observations of the numerical model, depending on the computational affordable budget. 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 uncertainty propagation for central tendency estimation, outlining the links between numerical integration and the numerical design of experiments.

1.5.1 Numerical integration

Forward uncertainty propagation aims at integrating a measurable function $g : \mathcal{D}_{\mathbf{X}} \to \mathbb{R}$ w.r.t. a probability measure $\mathbb{P}_{\mathbf{X}}$. Numerical integration provides algorithmic tools to help the resolution of this probabilistic integration problem. To ease the notations, the numerical model $\mathcal{M}(\cdot)$ introduced in Eq. (1.1) is replaced by a generic measurable function $g(\cdot)$.

In practice, this integral is approximated by summing a finite n-sized set of realizations $\mathbf{y}_n = \left\{g(\mathbf{x}^{(1)}), \dots, g(\mathbf{x}^{(n)})\right\}$ from a set of input samples $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\}$. A *quadrature* selects the input samples $\mathbf{X}_n$ (also called "nodes" in classical numerical integration), 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:

$$I_{\mathbb{P}_{\mathbf{X}}}(g) = \mathbb{E}[g(\mathbf{X})] = \int_{\mathcal{D}_{\mathbf{X}}} g(\mathbf{x})\, \mathrm{d}\mathbb{P}_{\mathbf{X}}(\mathbf{x}) \approx \sum_{i=1}^{n} w_i\, g\left(\mathbf{x}^{(i)}\right). \tag{1.11}$$

For a given sample size n, the goal is to find a set of tuples $\left\{\mathbf{x}^{(i)}, w_i\right\}_{i=1}^{n}$ (i.e., a quadrature rule), which provide the best approximation of the quantity of interest $I_{\mathbb{P}_{\mathbf{X}}}(g)$. Ideally, the approximation quality should be fulfilled for a wide class of integrands (e.g., regardless of their regularity). Most quadrature rules only depend on the probability space $(\Omega, \mathcal{A}, \mathbb{P}_{\mathbf{X}})$, regardless of the integrand values. In the context of a costly numerical model, defining the quadrature ahead allows the analyst to massively distribute the evaluations of the numerical model.

Classical multivariate deterministic quadrature

Historically, quadrature methods have been developed for univariate integrals. The Gaussian rule and the Fejér-Clenshaw-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 one being the Gauss-Legendre quadrature. In this case, the function g to be integrated w.r.t. 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, while the respective weights are given by the following formula:

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

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 (i.e., its construction is not iterative). However, a sequential extension is proposed by the Gauss-Kronrod rule (Laurie, 1997), at the expense of a slightly lower accuracy regarding $I_{\mathbb{P}_X}(g)$.

To overcome this practical drawback, Fejér (Féjer, 1933) then Clenshaw and Curtis (Clenshaw and Curtis, 1960) proposed a nested rule with mostly equivalent accuracy as Gaussian quadratures. This method is usually presented to integrate a function w.r.t. the uniform measure on $[-1, 1]$ and starts with a change of variables, $x := \cos(\theta)$, $\mathrm{d}x := \sin(\theta)$:

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

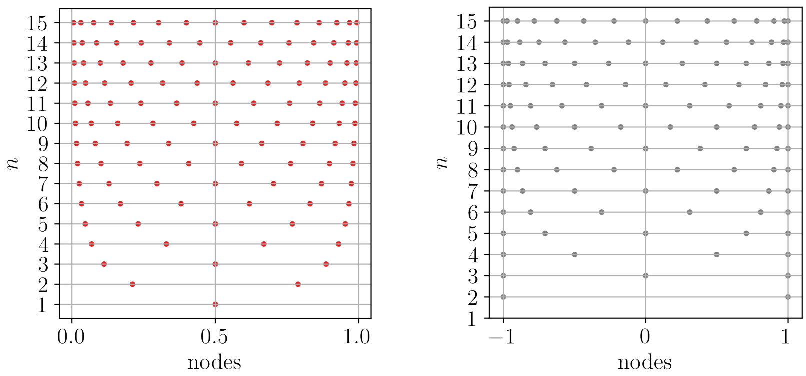
This expression can be written as an expansion of the integrand using trigonometric series. Therefore, knowing that cosine series are closely related to the Chebyshev polynomials of the first kind, Fejér's "first rule" (Trefethen, 2008) uses the Chebyshev polynomials roots as nodes $x^{(i)} = \cos(\theta^{(i+1/2)})$,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

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

where $\lfloor \cdot \rfloor$ denotes the floor function.

These two univariate integration schemes are both very efficient on a wide panel of functions. Yet, Fejér-Clenshaw-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-Clenshaw-Curtis quadrature by representing the nodes of quadrature rules with increasing size on the y-axis. For example, the Fejér-Clenshaw-Curtis quadrature with size $n = 13$ includes all the nodes from the quadratures with sizes $n = 7$ and $n = 4$.

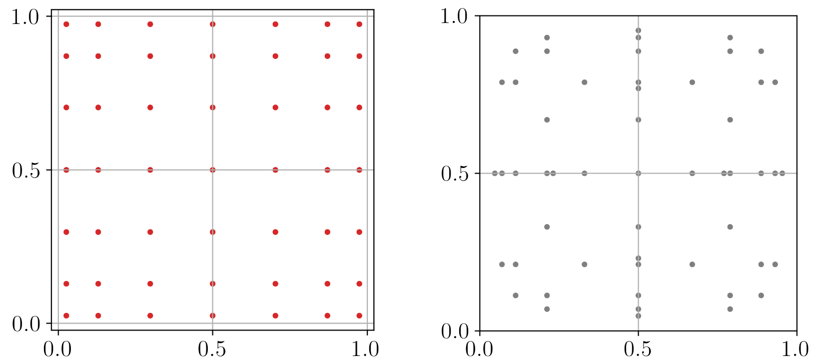
However, in practice, UQ problems are rarely unidimensional, but one can directly extend these univariate rules to multivariate ones by defining the tensor product (also called "full grids") of univariate rules. This exhaustive approach quickly shows its practical limits as the problem's dimension increases. In Fig. 1.4, the left plot represents a two-dimensional tensor product of identical Gauss-Legendre quadratures. Alternatively, sparse multivariate quadratures (i.e.,



(a) Gauss-Legendre. (b) Clenshaw-Curtis.

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

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. Finally, this technique showed great properties, but other methods offer more flexibility and guarantees.



(a) Tensor product (n=49). (b) Smolyak sparse grid (n=53).

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

Monte Carlo methods

Monte Carlo (MC) methods were initially developed in the 1940s to solve problems in neutronics. Ever since, this technique has been applied to the numerical resolution of integrals. To integrate a function g against a measure $\mathbb{P}_{\mathbf{X}}$, it randomly generates points according to the input measure. The integral is estimated by taking the uniform arithmetic mean of the nodes' images obtained.

This 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 $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\} \sim \mathbb{P}_{\mathbf{X}}$. The Monte Carlo estimator of the integral is given by:

$$I_{\mathbb{P}_{\mathbf{X}}}(g) \approx \overline{y}_n^{\mathrm{MC}} = \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{x}^{(i)}). \tag{1.15}$$

By construction, the strong law of large numbers makes this estimator unbiased and strongly consistent. However, it converges relatively slowly. The variance of the Monte Carlo estimator results from a manipulation of the central limit theorem:

$$\mathrm{Var}\left(\overline{y}_n^{\mathrm{MC}}\right) = \frac{1}{n} \mathrm{Var}\left(Y\right). \tag{1.16}$$

Considering the images of the sample $\mathbf{X}_n$, one can also express the unbiaised empirical variance of the output random variable $\widehat{\sigma}_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(g(\mathbf{x}^{(i)}) - \overline{y}_n^{\mathrm{MC}}\right)^2$. The Monte Carlo estimator also comes with theoretical confidence intervals at $\alpha\%$, given for a known value of $\mathrm{Var}(g(\mathbf{X}))$ by:

$$I_{\mathbb{P}_{\mathbf{X}}}(g) \in \left[\overline{y}_n^{\mathrm{MC}} - q_\alpha \sqrt{\frac{\mathrm{Var}\left(g(\mathbf{X})\right)}{n}}, \overline{y}_n^{\mathrm{MC}} + q_\alpha \sqrt{\frac{\mathrm{Var}\left(g(\mathbf{X})\right)}{n}}\right], \tag{1.17}$$

where $q_\alpha = \Phi(1 - \alpha/2)$, and Φ is the CDF 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 explicitly depend on the dimension of the input domain. Unlike the previous multivariate deterministic quadrature, it does not explicitly 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 Owen (2013, Chap. 8,9,10) 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 mixed performances. Quasi-Monte Carlo (QMC) is a deterministic family of numerical integration schemes w.r.t. 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 lower the discrepancy of a design is, the "closest" it is to uniformity.

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

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

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 variation.

The reader might refer to Leobacher and Pillichshammer (2014, Sec. 3.4) for further mathematical proof. The function's variation $V(g)$ in Theorem 2 can be formally defined as the Hardy-Klause variation:

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

According to Sullivan, 2015, p.188, this variation coincides with the notion of total variation defined in Appendix B for $d = 1$.

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 $\#(\mathbf{X}_n \cap [\mathbf{0}, \mathbf{x}))$, where # denotes the cardinal of a set. Then, if this empirical quantification is compared with the volume of the rectangle $[\mathbf{0}, \mathbf{x})$, denoted by $\mathrm{vol}([\mathbf{0}, \mathbf{x}))$. Thus, the star discrepancy is expressed as:

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

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}:

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

Interestingly, the 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 sequences for quadrature.

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 the *maximum mean discrepancy* is introduced in Appendix D and used for efficient sampling in Chapter 3 of this manuscript.

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

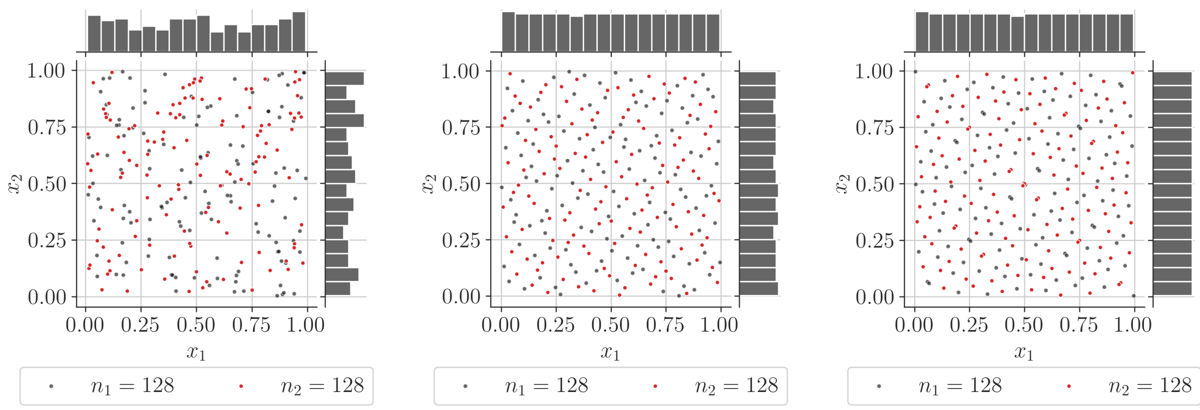
$$\left| \int_{[0,1]^d} g(\mathbf{x}) \mathrm{d}\mathbf{x} - \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{x}^{(i)}\right) \right| \leq \frac{V(g) \log(n)^d}{n}. \tag{1.22}$$

The generation of such sequences does not necessarily require more effort than pseudo-random sampling. In Owen (2013, Chap. 15), an extended presentation is offered about the various ways of generating 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, other methods such as the famous Sobol' or Faure sequences were developed (later gathered under a generic notion called "digital nets" by Dick and Pillichshammer, 2010). Sobol' sequences are in base two and have the advantage of being extensible in dimension. Note that by construction, these sequences offer particularly low 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 QMC designs (with size $n = 2^8 = 256$). Each one is then split into the first $n_1 = 128$ points (in red) and the following $n_2 = 128$ points (in black) to show to nested properties of the QMC sequences. This illustration clearly shows the way QMC sequences uniformly occupy the domain while MC sampling leaves uncovered areas.

Crude Monte Carlo estimators provide guarantees associated with the estimate. This complementary information is essential to deliver an end-to-end UQ, which is missed in deterministic QMC methods. *randomized quasi-Monte Carlo* extends the usual QMC methods by introducing some randomness 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 randomized QMC by default instead of QMC as a good practice (e.g., Owen, 2013). Recent works aim at exploring the use of these methods to estimate different quantities of interest, such as an expected value similar to $I_{\mathbb{P}_\mathbf{X}}(g)$ (Gobet et al., 2022) or a quantile (Kaplan et al., 2019).

Ultimately, QMC methods generate powerful integration schemes. The Koksma-Hlawka inequality associates an upper bound and a convergence rate to most integrals. A randomization overlay removes the deterministic property of these designs, allowing to compute confidence



(a) Monte Carlo. (b) Halton sequence. (c) Sobol' sequence.

Figure 1.5 Nested Monte Carlo and quasi-Monte Carlo designs ($n = 2^8 = 256$).

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 GitHub[5].

1.5.2 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 design of experiment (also simply called "design") is called *space-filling* when it uniformly covers a domain. Such designs can be used to propagate uncertainties for a very limited sample size. Therefore, users of designs of experiments consider various properties:

- A first desirable property is the sequentiality of the method, to eventually add new points when their initial computational budget is extended;

- A second quality is the preservation of the methods' 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 1.7 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.

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

This subsection 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 w.r.t. 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. While, the L_∞ star discrepancy is hard to estimate in practice, Warnock (1972) elaborated an explicit expression specific to the L_2 star discrepancy:

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

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 invariant 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 Hardy-Klause variation. Let us introduce the explicit formula of the centered L_2 discrepancy given by Hickernell (1998):

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

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 by 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 ϕ_{min} as $p \geq 1$ tends to infinity:

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

Further space-filling criteria are reviewed in Abtini (2018) and in Da Veiga et al. (2021, Appendix A).

Latin hypercube sampling

Latin hypercube sampling (LHS) is a method initially introduced in the 1970s as numerical integration method (Mckay et al., 1979). In a bounded domain, this technique corresponds to a stratified sampling which aims to force 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 cell. 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 either be placed at the center of the cell or randomly in the cell.

Various consecutive contributions explicited the variance, then a central limit theorem to LHS (e.g., Owen, 1992). Under monotony hypothesis, the LHS variance can be expressed, when $n \longrightarrow \infty$, as:

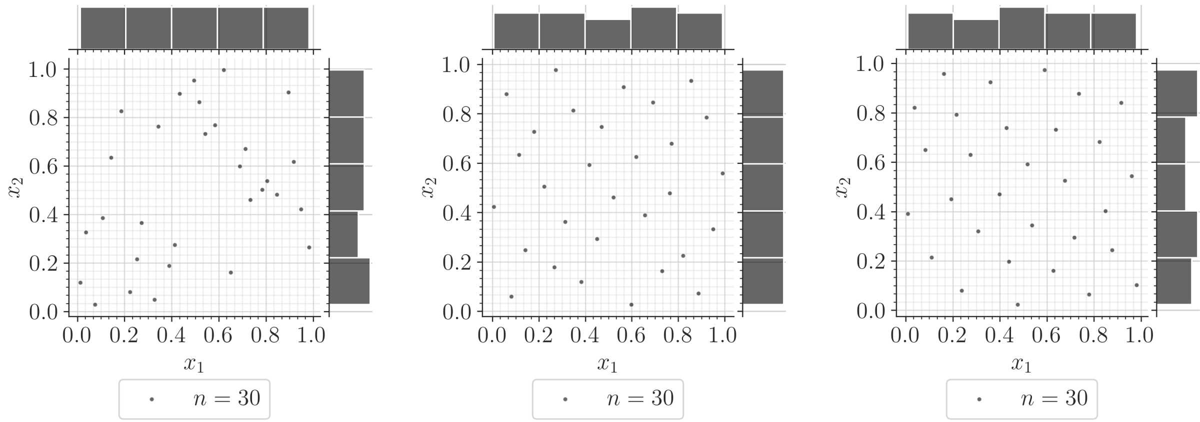
$$\mathrm{Var}\left(\overline{y}_n^{\mathrm{LHS}}\right) = \frac{1}{n}\mathrm{Var}\left(g(\mathbf{X})\right) - \frac{C}{n} + o\left(\frac{1}{n}\right), \tag{1.26}$$

where C is a positive constant, showing that the LHS usually asymptotically reduces the variance for numerical integration (a proof is proposed in Stein, 1987), and $\overline{y}_n^{\mathrm{LHS}}$ is the arithmetic mean over the LHS. However, 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 paragraph presents various methods to at optimize LHDs.

Optimized Latin hypercube sampling

To improve the space-filling property of a LHD, it is common to add an optimization step. The goal of this optimization is to improve a space-filling criterion by generating a LHD from permutations of an initial LHD. Damblin et al. (2013) reviewed 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 a LHD optimized by the 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 may 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 LHDs. Among them, let us mention the "maximum projection designs" (for MaxPro) 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) also optimizes LHS, but based on a criterion averaging discrepancies between each pair of marginals.



(a) Poorly space-filling LHS. (b) L_2 centered optimized LHS. (c) ϕ_p optimized LHS.

Figure 1.6 Latin hypercube designs with poor and optimized space-filling properties ($n = 30$).

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 GitHub[6].

1.5.3 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 UQ, this step might be referred to as central tendency estimation (i.e., estimating of the output's mean and variance). Central tendency estimation can be viewed as a numerical integration w.r.t. an input distribution, sometimes called by authors *probabilistic integration* (Briol et al., 2019), as part of what is called *probabilistic numerics* (Oates and Sullivan, 2019).

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 w.r.t. uniformity and discrepancies w.r.t. the target distribution. This question will be further discussed in Chapter 3, using a more general framework.

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

- building a quadrature rule for numerical integration against a uniform distribution;

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

- 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, i.e., to perform 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 analysis of the central behavior of the output random variable. However, some studies require to shift the focus toward specific areas of the output random variables distribution. When using uncertainty propagation to perform risk analysis, the events studied are mostly 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.

1.6 Uncertainty propagation for rare event estimation

This section aims to present another type of problem often encountered in uncertainty propagation. The goal of these problems is to assess the risk or reliability of a critical engineering system. Most often, a risk measure associated with a failure mode of the studied system has to be 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 input conditions or an unlikely unfavorable combination of inputs leads to the failure of the system. Therefore, the equivalent terms *reliability analysis* and *rare event estimation* are used. The notion of risk associated with an event is often decomposed as a product of its probability of occurrence 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 risk measure, widely used for risk analysis. The quantile of order α, also called α-quantile, denoted by $q_\alpha(\mathrm{Y})$, of the output random variable Y is defined as:

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

As an alternative, one can define a scalar safety threshold $y_{\mathrm{th}} \in \mathbb{R}$ that should not be exceeded to keep the system safe. Then, another risk measure is the probability of crossing this safety threshold, also called *failure probability*, and given by:

$$p_{\mathrm{f}} = \mathbb{P}(Y \geq y_{\mathrm{th}}). \tag{1.28}$$

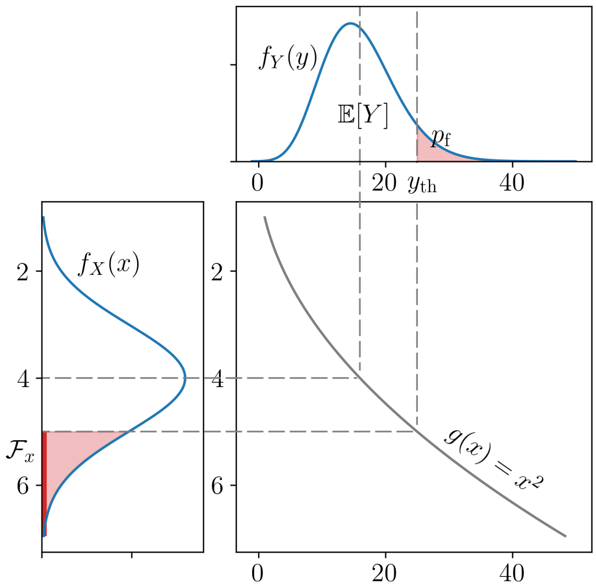


Figure 1.7 One-dimensional reliability analysis example.

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 crossing a given threshold y_{th} is represented by the area in red under the output PDF on top. More information about 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 Rackwitz (2001); Andrieu-Renaud et al. (2004); Hawchar et al. (2017).

1.6.1 Problem statement

Following the UQ methodology, the behavior of the system is modeled by $\mathcal{M}(\cdot)$. Considering the problem of crossing 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 $y_{th} \in \mathbb{R}$. Formally, the *limit-state function* (LSF) is a deterministic function $g : \mathcal{D}_X \to \mathbb{R}$ quantifying this performance:

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

Depending on the sign of the output realizations of $g(\mathbf{x})$, this function splits the input space into two disjoint and complementary domains called the *failure domain* $\mathcal{F}_\mathbf{x}$, and the *safe domain* $\mathcal{S}_\mathbf{x}$ which are defined as:

$$\mathcal{F}_\mathbf{x} = \{\mathbf{x} \in \mathcal{D}_\mathbf{x} \,|\, g(\mathbf{x}) \leq y_{th}\}, \qquad \mathcal{S}_\mathbf{x} = \{\mathbf{x} \in \mathcal{D}_\mathbf{x} \,|\, g(\mathbf{x}) > y_{th}\}. \tag{1.30}$$

The border between these two domains is a hypersurface called *limit-state surface*, and defined by $\mathcal{F}_{\mathbf{x}}^0 = \{\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \mid g(\mathbf{x}) = 0\}$. 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 partially 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 sequential and parallel problems (Lemaire et al., 2009; Der Kiureghian, 2022).

A rare event estimation results from a particular uncertainty propagation through the LSF. *failure probability*, denoted by p_{f} is formally written:

$$p_{\mathrm{f}} = \mathbb{P}(Y \geq y_{\mathrm{th}}) = \mathbb{P}(g(\mathbf{X}) \leq 0) = \int_{\mathcal{F}_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}_{\mathbf{x}}} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{1.31}$$

Rare event estimation implies both contour finding (i.e., characterizing the LSF) and an estimation strategy targeting the failure domain (with the fewest number of simulations). Note that failure events can be qualified as rare when their failure probability has an order of magnitude between $10^{-9} \leq p_{\mathrm{f}} \leq 10^{-2}$ (see e.g., Lemaire et al., 2009).

Instead of directly performing a reliability analysis in the physical space (often called "**x**-space"), these problems are usually solved in the *standard normal space* (also called "**u**-space"). Working in the standard normal space reduces numerical issues caused by potentially unscaled or asymmetric marginals. Moreover, a larger panel of methods can be applied in the standard normal space since the random inputs become independent. The bijective mapping between these two spaces is called an "iso-probabilistic transformation", denoted by $T : \mathcal{D}_{\mathbf{x}} \longrightarrow \mathbb{R}^d$, with $\mathbf{x} \mapsto T(\mathbf{x}) = \mathbf{u} = (u_1, \dots, u_d)^\top$. 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 equivalence holds:

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

Thus, a reliability problem can be expressed in the standard normal space. In the following, the transformed LSF $\breve{g}$ defined as:

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

Since this transformation is a diffeomorphism[7], one can apply the change of variable $\mathbf{x} = T(\mathbf{u})$ to express the reliability problem from Eq. (1.31) in the standard normal space:

$$p_{\mathrm{f}} = \mathbb{P}(\breve{g}(\mathbf{U}) \leq 0) = \int_{\mathcal{F}_{\mathbf{u}}} \varphi_d(\mathbf{u})\,\mathrm{d}\mathbf{u} = \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{F}_{\mathbf{u}}}(\mathbf{u}) \varphi_d(\mathbf{u})\,\mathrm{d}\mathbf{u}, \tag{1.34}$$

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

with the transformed failure domain denoted by $\mathcal{F}_{\mathbf{u}} = \{\mathbf{u} \in \mathbb{R}^d \,|\, \breve{g}(\mathbf{u}) \leq 0\}$, 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:

- for elliptical distributions, a linear Nataf transformation is applied;
- for distributions with an elliptical copula, the generalized Nataf transformation is used (Lebrun and Dutfoy, 2009a);
- otherwise, the Rosenblatt transformation is used (Lebrun and Dutfoy, 2009b).

1.6.2 Rare event estimation methods

The main risk measure chosen for rare event estimation in this work is the previously introduced failure probability. In the computer experiments context, the simulation budget is often limited to n runs with $p_{\mathrm{f}} \ll \frac{1}{n}$. 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 introduces 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 normal space, the methods first look for this MPFP, denoted by P^*, with coordinates $\mathbf{u}^*$. To find it, one can solve the following quadratic optimization problem:

$$\mathbf{u}^* = \underset{\mathbf{u}\in\mathbb{R}^d}{\arg\max} \left(\mathbb{1}_{\mathcal{F}_{\mathbf{u}}}(\mathbf{u})\varphi_d(\mathbf{u})\right). \tag{1.35}$$

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

$$\begin{aligned}
\mathbf{u}^* &= \underset{\mathbf{u}\in\mathbb{R}^d}{\arg\max} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\mathbf{u}^\top \mathbf{u}}{2}\right) \qquad \text{s.t.} \quad \mathbf{u} \in \mathcal{F}_{\mathbf{u}} &(1.36)\\
&= \underset{\mathbf{u}\in\mathbb{R}^d}{\arg\min}\, \mathbf{u}^\top \mathbf{u} \qquad \text{s.t.} \quad \breve{g}(\mathbf{u}) \leq 0. &(1.37)
\end{aligned}$$

Such a quadratic optimization under nonlinear constraint is classically solved by gradient descent algorithms (e.g., Abdo-Rackwitz algorithm from Abdo and Rackwitz, 1991) but also by gradient-free techniques (e.g., Cobyla algorithm, see Powell, 1994). The design point corresponds to the smallest Euclidian distance between the limit-state surface and the origin of the standard normal 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 failure point with the largest probability density (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:

$$\beta = \|\mathbf{u}^*\|_2 = \boldsymbol{\alpha}^\top \mathbf{u}^*, \qquad \text{s.t.} \quad \boldsymbol{\alpha} = \frac{\nabla_{\mathbf{u}} \breve{g}(\mathbf{u})}{\|\nabla_{\mathbf{u}} \breve{g}(\mathbf{u})\|_2}. \tag{1.38}$$

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

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

$$\begin{aligned} \breve{g}(\mathbf{u}) &= \breve{g}_1(\mathbf{u}^*) + o\left(\|\mathbf{u}-\mathbf{u}^*\|_2^2\right) \\ &= \cancel{\breve{g}(\mathbf{u}^*)} + \nabla_{\mathbf{u}} \breve{g}(\mathbf{u}^*)^\top (\mathbf{u}-\mathbf{u}^*) + o\left(\|\mathbf{u}-\mathbf{u}^*\|_2^2\right) \\ &= \|\nabla_{\mathbf{u}} \breve{g}(\mathbf{u})\|_2 \left(\boldsymbol{\alpha}^\top \mathbf{u}^* - \boldsymbol{\alpha}^\top \mathbf{u}\right) + o\left(\|\mathbf{u}-\mathbf{u}^*\|_2^2\right). \end{aligned} \tag{1.39}$$

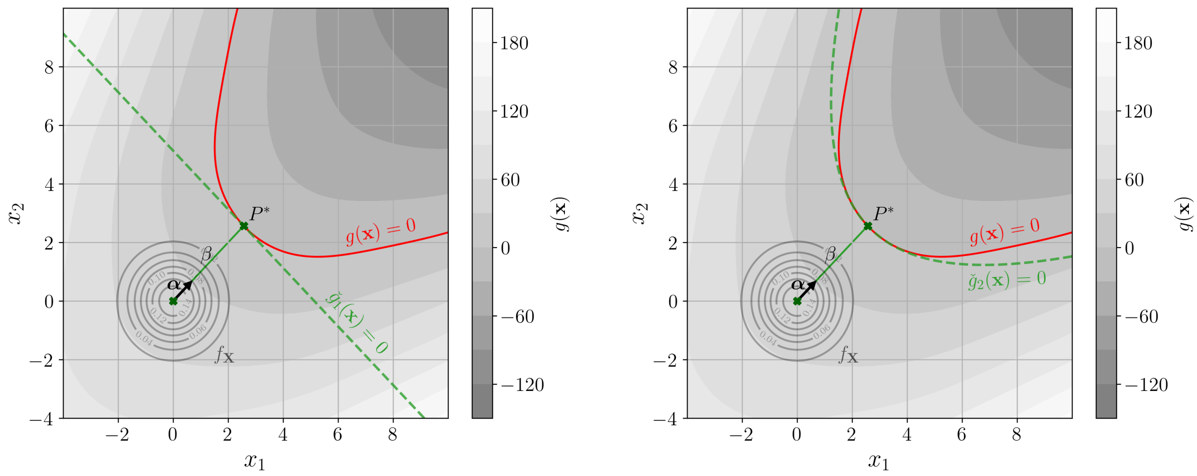
Using $\breve{g}_1(\cdot)$ as an approximation of the LSF, the failure probability can be approximated as:

$$p_{\mathrm{f}} \approx p_{\mathrm{f}}^{\mathrm{FORM}} = \mathbb{P}\left(-\boldsymbol{\alpha}^\top \mathbf{u} \leq -\beta\right) = \Phi(-\beta), \tag{1.40}$$

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, $p_{\mathrm{f}} = p_{\mathrm{f}}^{\mathrm{FORM}}$. 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).

When the MPFP is not unique, the application of these methods might lead to large 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.

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



(a) FORM approximation.

(b) SORM approximation.

Figure 1.8 FORM and SORM approximation applied to a two-dimensional problem where $g(x_1, x_2) = (x_1 - x_2)^2 - 8(x_1 + x_2 - 5) + \sin(x_1) + \sin(x_2)$.

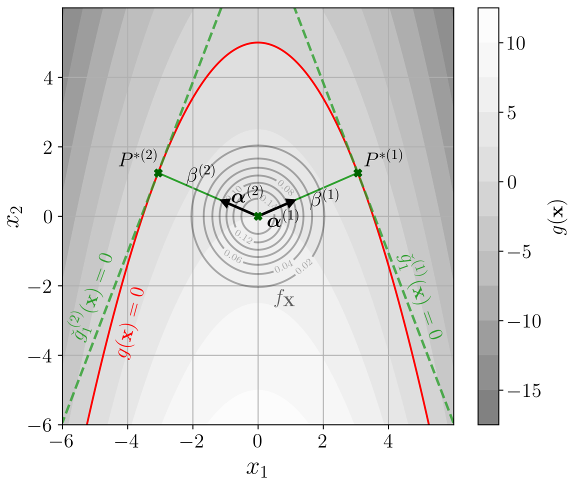


Figure 1.9 Multi-FORM approximation on an example with two MPFPs.

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 significant 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.

Crude Monte Carlo sampling

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 $\left\{\mathbf{x}^{(i)}\right\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} f_{\mathbf{X}}$.

The estimator is written using the indicator function applied to the LSF:

$$p_{\mathrm{f}} \approx \widehat{p_{\mathrm{f}}}^{\mathrm{MC}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x}^{(i)}). \tag{1.41}$$

This unbiased estimator converges towards the failure probability almost surely according to the law of large numbers. Once again, Monte Carlo offers an unbiased estimator, regardless of the intput dimension or the regularity of the function $g(\cdot)$. Additionally, the variance of this estimator is given by:

$$\mathrm{Var}\left(\widehat{p_{\mathrm{f}}}^{\mathrm{MC}}\right) = \frac{1}{n} p_{\mathrm{f}} \left(1 - p_{\mathrm{f}}\right). \tag{1.42}$$

This variance can be used to build a 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:

$$\delta_{\widehat{p_{\mathrm{f}}}^{\mathrm{MC}}} = \frac{\sqrt{\mathrm{Var}\left(\widehat{p_{\mathrm{f}}}^{\mathrm{MC}}\right)}}{p_{\mathrm{f}}} = \sqrt{\frac{1 - p_{\mathrm{f}}}{n p_{\mathrm{f}}}}. \tag{1.43}$$

In theory, the Monte Carlo technique 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 is qualified as an embarrassingly parallel method since all the numerical simulations are mutually independent. Finally, it offers strong convergence guarantees and complementary information on the quality of this estimation. These properties often make Monte Carlo sampling the reference technique in rare event estimation benchmarks.

However, the advantages of this estimator are overshadowed by its slow convergence. To estimate a target failure probability $p_{\mathrm{f}} = 10^{-\alpha}$, a Monte Carlo estimation with a convergence level $\delta_{\widehat{p_{\mathrm{f}}}^{\mathrm{MC}}} = 10\%$ famously requires $n = 10^{\alpha+2}$ simulations. Thus, 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 too far from 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 a region of interest (here the failure domain). By doing so, it allows drawing more points in the failure domain, leading to a better estimate of the quantity.

The challenge in IS is to pick a relevant *instrumental* distribution $h_{\mathbf{X}}$ (also called *auxiliary* distribution) to replace the initial distribution $f_{\mathbf{X}}$. Then, by introducing the *likelihood ratio* $w_{\mathbf{X}}(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{h_{\mathbf{X}}(\mathbf{x})}$, one can rewrite $f_{\mathbf{X}}(\mathbf{x}) = w_{\mathbf{X}}(\mathbf{x})\, h_{\mathbf{X}}(\mathbf{x})$ and inject it in the failure probability expression:

$$p_{\mathrm{f}} = \int_{\mathcal{D}_{\mathbf{x}}} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}_{\mathbf{x}}} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x})\, w_{\mathbf{X}}(\mathbf{x})\, h_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{1.44}$$

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

The IS estimator of the failure probability is defined for a sample drawn from the auxiliary distribution such that $\left\{\mathbf{x}^{(i)}\right\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} h_{\mathbf{X}}$:

$$\widehat{p_{\mathrm{f}}}^{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x}^{(i)})\, w_{\mathbf{X}}(\mathbf{x}^{(i)}). \tag{1.45}$$

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

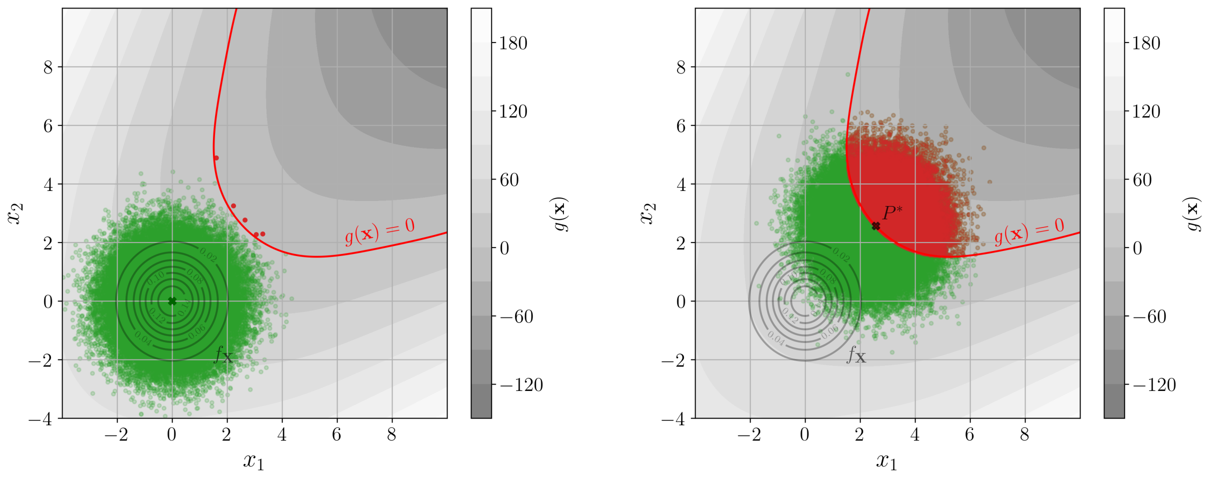
$$\mathrm{Var}\left(\widehat{p_{\mathrm{f}}}^{\mathrm{IS}}\right) = \frac{1}{n}\left(\mathbb{E}_{h_{\mathbf{X}}}\left[\left(\mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{X})\, \frac{f_{\mathbf{X}}(\mathbf{X})}{h_{\mathbf{X}}(\mathbf{X})}\right)^2\right] - p_{\mathrm{f}}^2\right). \tag{1.46}$$

The quality of the variance reduction associated to this technique fully depends on the choice of the instrumental distribution. In fact, IS 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. (1.46):

$$h_{\mathrm{opt}}(\mathbf{x}) = \frac{\mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{p_{\mathrm{f}}}. \tag{1.47}$$

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 IS 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 IS, simply called "FORM-IS" (Melchers, 1989). 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 IS centered on the design point. On the left, it can be noticed that only a few points in red reached the failure domain. Such a number their number is insufficient to estimtate the failure probability by 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 IS schemes integrate adaptive mechanisms, progressively leading the sampling towards the failure domain (Bugallo et al., 2017).



(a) Monte Carlo ($n = 10^5$).

(b) Importance sampling ($n = 10^5$).

Figure 1.10 Reliability assessment by Monte Carlo and importance sampling applied to a two-dimensional problem where $g(x_1, x_2) = (x_1 - x_2)^2 - 8(x_1 + x_2 - 5) + \sin(x_1) + \sin(x_2)$.

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

$$\mathcal{H}_{\boldsymbol{\lambda}} = \left\{ h_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\lambda}) = h_{\boldsymbol{\lambda}}(\mathbf{x}),\, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p) \in \mathcal{D}_{\boldsymbol{\lambda}} \subseteq \mathbb{R}^p \right\}. \tag{1.48}$$

The early work of Bucher (1988) only included normal distributions to minimize the IS variance w.r.t. the parameter $\boldsymbol{\lambda}$. Using Eq. (1.46) the optimization simplifies as:

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}}{\arg\min}\, \mathbb{E}_{h_{\boldsymbol{\lambda}}} \left[\mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{X}) \left(\frac{f_{\mathbf{X}}(\mathbf{x})}{h_{\boldsymbol{\lambda}}(\mathbf{x})} \right)^2 \right]. \tag{1.49}$$

However, this optimization strategy requires sampling w.r.t. the instrumental distribution at each optimization iteration, which can be avoided by using a different approach.

The "cross-entropy" (CE) method (Rubinstein and Kroese, 2004) uses Kullback-Leibler (KL) divergence (introduced in Appendix B) to optimize IS. The KL divergence is a dissimilarity measure between distributions, expressed between the optimal distribution h_{opt} and the parametric instrumental distribution $h_{\boldsymbol{\lambda}}$:

$$\begin{aligned} D_{\text{KL}}\left(h_{\text{opt}} || h_{\boldsymbol{\lambda}}\right) &= \int_{\mathcal{D}_{\mathbf{x}}} \log\left(\frac{h_{\text{opt}}(\mathbf{x})}{h_{\boldsymbol{\lambda}}(\mathbf{x})}\right) h_{\text{opt}}(\mathbf{x})\, \mathrm{d}\mathbf{x} && \text{(1.50a)} \\ &= \int_{\mathcal{D}_{\mathbf{x}}} \log\left(h_{\text{opt}}(\mathbf{x})\right) h_{\text{opt}}(\mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\mathcal{D}_{\mathbf{x}}} \log\left(h_{\boldsymbol{\lambda}}(\mathbf{x})\right) h_{\text{opt}}(\mathbf{x})\, \mathrm{d}\mathbf{x}. && \text{(1.50b)} \end{aligned}$$

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. $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda}^* = \arg\min_{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}} D_{\mathrm{KL}} \left(h_{\mathrm{opt}} || h_{\boldsymbol{\lambda}} \right). \tag{1.51}$$

By injecting the expression in Eq. (1.50b), the optimization problem simply becomes a function of an expected value over the initial density $f_{\mathbf{X}}$:

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}} \int_{\mathcal{D}_{\mathbf{x}}} \log\left(h_{\boldsymbol{\lambda}}(\mathbf{x})\right) h_{\mathrm{opt}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \arg\max_{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}} \mathbb{E}_{f_{\mathbf{X}}} \left[\mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}}(\mathbf{X}) \log\left(h_{\boldsymbol{\lambda}}(\mathbf{X})\right) \right]. \tag{1.52}$$

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

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}} \sum_{i=1}^{n} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}} \left(\mathbf{x}^{(i)} \right) \log\left(h_{\boldsymbol{\lambda}} \left(\mathbf{x}^{(i)} \right) \right), \qquad \left\{ \mathbf{x}^{(i)} \right\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} f_{\mathbf{X}}. \tag{1.53}$$

This optimization can be solved by canceling the gradient:

$$\sum_{i=1}^{n} \mathbb{1}_{\{\mathcal{F}_{\mathbf{x}}\}} \left(\mathbf{x}^{(i)} \right) \left[\nabla \log\left(h_{\boldsymbol{\lambda}^*} \right) \right] \left(\mathbf{x}^{(i)} \right) = \mathbf{0}. \tag{1.54}$$

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. (1.52) is made possible.

The algorithm starts by generating and evaluating an initial sample $\left\{ g\left(\mathbf{X}_{[1]}^{(i)} \right) \right\}_{i=1}^{n}$, 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 $h_{\boldsymbol{\lambda}_{[1]}^*}$ is optimized. At the next steps $k \in \{1, \dots, k_\#\}$, the sample $\left\{ \mathbf{X}_{[k]}^{(i)} \right\}_{i=1}^{n}$ is generated from the density $h_{\boldsymbol{\lambda}_{[k-1]}^*}$ and the rest of the process repeats until the estimated threshold level becomes negative, $q_{[k_\#]}^{p_0} \leq 0$.

The final instrumental density $h_{\boldsymbol{\lambda}_{[k_\#]}^*}$ is then used for IS as defined in Eq. (1.45):

$$\widehat{p_{\mathrm{f}}}^{\mathrm{CE-AIS}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \frac{f_{\mathbf{X}}\left(\mathbf{x}_{[k_\#]}^{(i)} \right)}{h_{\boldsymbol{\lambda}_{[k_\#]}^*}\left(\mathbf{x}_{[k_\#]}^{(i)} \right)}, \qquad \left\{ \mathbf{X}_{[k_\#]}^{(i)} \right\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} h_{\boldsymbol{\lambda}_{[k_\#]}^*}. \tag{1.55}$$

The CE-AIS algorithm is widely used in rare event estimation, as it based on an adaptive technique while conserving the explicit IS variance given in Eq. (1.46). 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 have been 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 normal 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.

Nonparametric adaptive importance sampling

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

$$\widehat{p_{\mathrm{f}}}_{[1]} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x} \leq 0)\}}\left(\mathbf{x}_{[1]}^{(i)}\right) \frac{f_{\mathbf{X}}\left(\mathbf{x}_{[1]}^{(i)}\right)}{h_{[0]}\left(\mathbf{x}_{[1]}^{(i)}\right)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}}\left(\mathbf{x}_{[1]}^{(i)}\right) w_{[1]}^{(i)}. \tag{1.56}$$

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

$$\widehat{h}_{[1]}(\mathbf{x}) = \frac{\det(\mathbf{H}_{[1]})^{-1/2}}{n\widehat{p_{\mathrm{f}}}_{[1]}} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}}\left(\mathbf{x}_{[1]}^{(i)}\right) w_{[1]}^{(i)} K\left(\mathbf{H}_{[1]}^{-1/2}\left(\mathbf{x} - \mathbf{x}_{[1]}^{(i)}\right)\right), \tag{1.57}$$

where the kernel K is commonly taken as the multivariate standard Gaussian kernel 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 $\widehat{p_{\mathrm{f}}}_{[1]}$. After performing this KDE, the estimated density can be used as instrumental density in Eq. (1.45).

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 $q^{p_0}_{[1]} > \cdots > q^{p_0}_{[k_\#]}$ of the successive importance sampling steps. This algorithm is initiated by setting $h_{[0]} = f_{\mathbf{X}}$ and stops at the step $k_\#$, when $q^{p_0}_{[k_\#]} < 0$. At the step k, the intermediate failure probability is written as:

$$\widehat{p}_{\mathrm{f}[k]} = \frac{1}{kn} \sum_{j=1}^{k} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq q^{p_0}_{[j]}\}} \left(\mathbf{x}^{(i)}_{[j]}\right) \frac{f_{\mathbf{X}}\left(\mathbf{x}^{(i)}_{[j]}\right)}{\widehat{h}_{[j-1]}\left(\mathbf{x}^{(i)}_{[j]}\right)} = \frac{1}{kn} \sum_{j=1}^{k} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq q^{p_0}_{[j]}\}} \left(\mathbf{x}^{(i)}_{[j]}\right) w^{(i)}_{[j]}, \tag{1.58}$$

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

$$\widehat{h}_{[k+1]}(\mathbf{x}) = \frac{\det(\mathbf{H}_{[k]})^{-1/2}}{kn\widehat{p}_{\mathrm{f}[k]}} \sum_{j=1}^{k} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq q^{p_0}_{[j]}\}} \left(\mathbf{x}^{(i)}_{[j]}\right) w^{(i)}_{[j]} K\left(\mathbf{H}^{-1/2}_{[k]}\left(\mathbf{x} - \mathbf{x}^{(i)}_{[j]}\right)\right). \tag{1.59}$$

The last instrumental density $\widehat{h}_{[k_\#]}$ is finally considered as an approximation of the optimal density for IS introduced in Eq. (1.45):

$$\widehat{p}_{\mathrm{f}}^{\,\mathrm{NAIS}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \frac{f_{\mathbf{X}}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)}{\widehat{h}_{[k_\#]}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)}, \qquad \left\{\mathbf{X}^{(i)}_{[k_\#]}\right\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} h_{k_\#}. \tag{1.60}$$

Overall, NAIS offers more flexibility to infer the optimal IS density. This property might suit problems presenting a highly nonlinear LSF. Then, relying on IS still provides an expression of the estimator's variance, by adapting Eq. (1.46) 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 significantly decrease for problems in dimensions larger than ten.

Subset simulation

Although the concept of "splitting" already existed (Kahn and Harris, 1951), 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 $\mathcal{F}_{\mathbf{x}}$ into an intersection of $k_\#$ intermediary events $\mathcal{F}_{\mathbf{x}} = \cap_{k=1}^{k_\#} \mathcal{F}_{[k]}$. Each are nested such that $\mathcal{F}_{[1]} \supset \cdots \supset \mathcal{F}_{[k_\#]} = \mathcal{F}_{\mathbf{x}}$. The failure probability is then expressed as a product of conditional probabilities:

$$p_{\mathrm{f}} = \mathbb{P}(\mathcal{F}_{\mathbf{x}}) = \mathbb{P}(\cap_{k=1}^{k_\#} \mathcal{F}_{[k]}) = \prod_{k=1}^{k_\#} \mathbb{P}(\mathcal{F}_{[k]} | \mathcal{F}_{[k-1]}). \tag{1.61}$$

From a practical point of view, the analyst tunes the algorithm[8] 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 $q_{[1]}^{p_0} > \cdots > q_{[k_\#]}^{p_0}$ 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 $f_{\mathbf{X}}$. Following conditional subset samples are generated by a *Monte Carlo Markov Chain* (MCMC) sampling technique of $f_{\mathbf{X}}(\mathbf{x}|\mathcal{F}_{[k-1]})$, using as initialization points the $n = Np_0$ samples given by $\mathbf{A}_{[k],n} = \left\{\mathbf{X}_{[k-1]}^{(i)} \subset \mathbf{X}_{[k-1],N} | g(\mathbf{X}_{[k-1]}^{(i)}) > \widehat{q}_{[k-1]}^{p_0}\right\}_{i=1}^{n}$. This process is repeated until an intermediary quantile becomes negative: $\widehat{q}_{[k_\#]}^{p_0} < 0$. Finally, the failure probability is estimated by:

$$\widehat{p_{\mathrm{f}}}^{\mathrm{SS}} = p_0^{k_\# - 1} \frac{1}{N} \sum_{i=1}^{n} \mathbb{1}_{\{g(\mathbf{x}) \leq 0\}} \left(\mathbf{X}_{[k_\#],N}^{(i)}\right). \tag{1.62}$$

In practice, the subset sample size should be large enough to properly estimate intermediary quantiles, leading to the usual recommendation of $p_0 = 0.1$ (Au and Beck, 2001). 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 $\widehat{p_{\mathrm{f}}}^{\mathrm{SS}}$. The first one results from a first-order Taylor expansion of Eq. (1.62) and is often considered as an upper bound. The second one 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 often depends on the choice and tuning of the MCMC algorithm. The "component-wise" Metropolis–Hastings algorithm (or "modified Metropolis–Hastings") is widely used in the standard normal space for SS. 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 SS is a versatile method, presenting consistent performances even for rare probabilities. Its flexibility allows it to deal with highly nonlinear LSFs, but its drawbacks arise from the use of MCMC sampling. For example, the convergence of MCMC is complex to control and depends on its tuning (Roy, 2020).

[8] An algorithmic presentation of the generic subset simulation method is given in Appendix C.

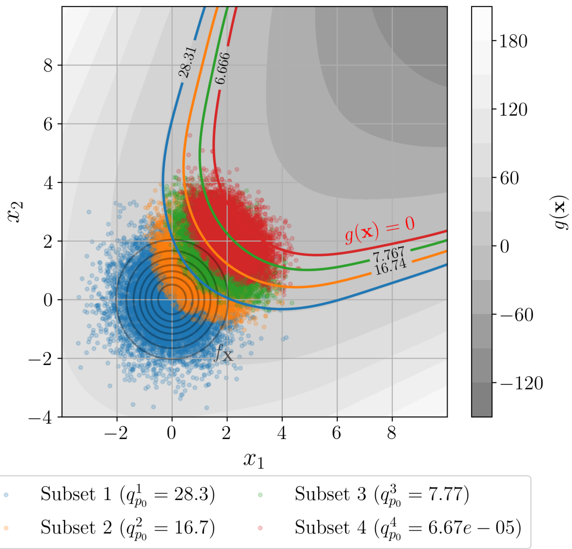


Figure 1.11 Reliability assessment by subset simulation ($n = 4 \cdot 10^4, p_0 = 0.1$) applied to a two-dimensional problem where $g(x_1, x_2) = (x_1 - x_2)^2 - 8\,(x_1 + x_2 - 5) + \sin(x_1) + \sin(x_2)$.

1.6.3 Summary and discussion

This section introduces a generic formulation and various methods for rare event estimation. However, several other methods from the field of reliability analysis are worth mentioning, such as *directional sampling* Bjerager (1988), *line sampling* (Koutsourelakis, 2004), or more recently *moving particles* (Walter, 2015). To go further on this topic the reader might refer to Morio and Balesdent (2015), which compares the advantages and drawbacks of the most common methods and presents their algorithmic structure.

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

- the computational cost of the LSF evaluation;
- the nonlinearity of the LSF;
- 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 1.8).

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`[9] (Fekhari et al., 2021), based on core `OpenTURNS` objects.

When working with computationally expensive numerical models, directly using rare event estimation methods is most often intractable. Many contributions were dedicated to coupling surrogate models with sampling methods for rare event estimation. Moustapha et al. (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, evaluating the sensitivities and assessing the robustness of the failure probability to the input uncertainty model is a major question, which was studied within probabilistic (Lemaître et al., 2015; Chabridon et al., 2017; Chabridon, 2018; Chabridon et al., 2021) and extra-probabilistic (Ajenjo et al., 2022; Ajenjo, 2023) 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 GitHub[10].

1.7 Global sensitivity analysis

After propagating uncertainties, the analyst might perform a sensitivity analysis (SA) 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 noninfluential variables in a UQ study (i.e., in a qualitative way);
- *importance measures*: assess the contribution of inputs in the global variability of the output (i.e., in a quantitative way).

The global sensitivity of an output can be explained by multiple 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.

[9] https://github.com/mbaudin47/otbenchmark/
[10] https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/reliability.ipynb

Screening methods are typically used in a statistical learning process, to drop the irrelevant variables. Remark that in the machine learning community, *feature selection* (Fan and Lv, 2010) serves the same purpose with a slight difference. On top of looking for the irrelevant features to the learning, feature selection investigates the redundant features.

1.7.1 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 to discriminate the noninfluential variables, which can afterwards be treated as deterministic 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 $\mathcal{D}_{\mathbf{X}}$ 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 j from an OAT trajectory t is expressed as a finite difference:

$$\mathrm{EE}_j^{(t)} = \frac{g(\mathbf{x}^{(t)}) - g(\mathbf{x}^{(t)} + \Delta_j)}{\Delta}. \tag{1.63}$$

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

$$\overline{\mathrm{EE}}_j = \frac{1}{n} \sum_{t=1}^{T} |\mathrm{EE}_j^{(t)}|, \qquad \widehat{\mathrm{Var}}(\mathrm{EE}_j) = \frac{1}{n-1} \sum_{t=1}^{T} \left(\mathrm{EE}_j^{(t)} - \overline{\mathrm{EE}}_j \right)^2. \tag{1.64}$$

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 very intuitive method 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, it consists in computing the mean value of local derivatives of the model output w.r.t. the inputs:

$$v_j = \int_{\mathcal{D}_{\mathbf{X}}} \left(\frac{\partial g(\mathbf{x})}{\partial x_j}\right)^2 f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbb{E}\left[\left(\frac{\partial g(\mathbf{X})}{\partial X_j}\right)^2\right]. \tag{1.65}$$

This continuous formulation can be computed with advanced numerical integration methods such as QMC. 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 (see Roustant et al., 2017 for further details).

1.7.2 Variance-based importance measures

Beyond the qualitative results from the screening methods, importance measures quantify the influence of inputs, allowing 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 shares to the inputs. Consider 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 follows:

$$Y = g_0 + \sum_{j=1}^{d} g_j(X_j) + \sum_{j<l}^{d} g_{jl}(X_j, X_l) + \cdots + g_{1\ldots d}(\mathbf{X}), \tag{1.66}$$

with the previous terms defined according to this recurrence relation:

$$g_0 = \mathbb{E}[g(\mathbf{X})] \tag{1.67a}$$

$$g_j(X_j) = \mathbb{E}[g(\mathbf{X})|X_j] - g_0 \tag{1.67b}$$

$$g_{jl}(X_j, X_l) = \mathbb{E}[g(\mathbf{X})|X_j, X_l] - g_j(X_j) - g_l(X_l) - g_0 \tag{1.67c}$$

$$\ldots \tag{1.67d}$$

Several 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 used to derive a functional decomposition of the variance (also called "functional analysis of variance" or FANOVA):

$$\mathrm{Var}(Y) = \sum_{j=1}^{d} V_j(Y) + \sum_{j<l}^{d} V_{jl}(Y) + \cdots + V_{1\ldots d}(Y), \tag{1.68}$$

where the previous terms are defined in a recurrent way, in the same fashion as Eq. (1.67): $V_j(Y) = \mathrm{Var}\left(\mathbb{E}[Y|X_j]\right), V_{jl}(Y) = \mathrm{Var}\left(\mathbb{E}[Y|X_j, X_l]\right) - V_j(Y) - V_l(Y)$, 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} (and higher order) Sobol' indices quantify the effect of the interactions between a group of marginals:

$$S_j = \frac{V_j(Y)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_j]\right)}{\mathrm{Var}(Y)} \tag{1.69a}$$

$$S_{jl} = \frac{V_{jl}(Y)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_j, X_l]\right) - V_j(Y) - V_l(Y)}{\mathrm{Var}(Y)} \tag{1.69b}$$

$$\ldots \tag{1.69c}$$

The generic definition of the Sobol' indices associated with a subset of inputs $A \in \mathcal{P}_d$ (see Da Veiga et al., 2021), with $\mathcal{P}_d$ the set of all possible subsets of $\{1, \ldots, d\}$, is given by:

$$S_A = \frac{V_A(Y)}{\mathrm{Var}(Y)} = \frac{\sum_{B\subset A} (-1)^{|A|-|B|} \mathrm{Var}\left(\mathbb{E}[Y|X_B]\right)}{\mathrm{Var}(Y)}, \tag{1.70}$$

where $|A|$ denotes here the cardinality of the set A. By using the FANOVA in Eq. (1.68), one can show that the Sobol' indices sum up to one:

$$\sum_{A\in\mathcal{P}_d} S_A = 1. \tag{1.71}$$

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:

$$S_A^{\mathrm{clos}} = \sum_{A'\subset A} S_{A'} = \frac{\mathrm{Var}\left(\mathbb{E}[Y|\mathbf{X}_A]\right)}{\mathrm{Var}(Y)}. \tag{1.72}$$

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

$$S_j^T = 1 - \frac{\mathrm{Var}\left(\mathbb{E}[Y|\mathbf{X}_{-j}]\right)}{\mathrm{Var}(Y)} = \frac{\mathbb{E}\left[\mathrm{Var}(Y|\mathbf{X}_{-j})\right]}{\mathrm{Var}(Y)}, \tag{1.73}$$

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

$$S_A^T = 1 - S_{A^\complement}^{\text{clos}} = 1 - \frac{\text{Var}\left(\mathbb{E}[Y|\mathbf{X}_{A^\complement}]\right)}{\text{Var}(Y)}, \tag{1.74}$$

where $A^\complement = \mathcal{P}_d \setminus A$. By jointly analyzing 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 any interactions, which means that the model is purely additive.

Estimating Sobol' indices can be achieved in various ways, even if historically the *pick-freeze* scheme is 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, Jansen, Martinez estimators), see further details in Chapter 3 of Da Veiga et al. (2021). Alternatively, the surrogate models can be 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 notable GSA method was adapted from the cooperative game theory by Owen (2014), allowing to work with dependent inputs.

`OpenTURNS` 5 (Sobol' indices)**.** The Python code available in Appendix D gives a minimalistic `OpenTURNS` implementation of the Sobol' indices to assess global sensitivity analysis on the Ishigami analytical problem. Further scripts are also available on GitHub[11].

Shapley effects

The Shapley effects are an adaptation to GSA by Owen (2014) of the Shapley values from the cooperative game theory (Shapley et al., 1953). This method is an alternative to Sobol' indices in the case of dependent inputs, for which the natural interpretation of single interaction effects

[11] https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/sensitivity_analysis.ipynb

no longer holds. In game theory, Shapley values act as a rule on how to share the value created by a team between its members (i.e., the players). The Shapley value allocated to the player X_j is given considering the indices $-\{j\} = \{1, \dots, d\} \backslash \{j\}$:

$$\varsigma_j = \sum_{A \subset -\{j\}} \binom{d-1}{|A|}^{-1} (\mathrm{val}(A \cup \{j\}) - \mathrm{val}(A)), \tag{1.75}$$

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

$$\mathrm{Sh}_j = \sum_{A \subset -\{j\}} \binom{d-1}{|A|}^{-1} \left(S^{\mathrm{clos}}_{A \cup \{j\}} - S^{\mathrm{clos}}_{A} \right). \tag{1.76}$$

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:

$$\sum_{j=1}^{d} \mathrm{Sh}_j = 1. \tag{1.77}$$

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

$$S_j \leq \mathrm{Sh}_j \leq S_j^T, \qquad \mathrm{Sh}_j = \sum_{A \in \mathcal{P}_d, j \in A} \frac{S_A}{|A|}. \tag{1.78}$$

Unlike Sobol' indices, Shapley effects are a nonnegative allocation of output variance with equitable division of the dependence. This method presents an interesting alternative in the dependent case, however, estimating Shapley effects creates computational difficulties. A first algorithm based on permutations was proposed by Song et al. (2016). Later on, surrogate models were also coupled to estimate Shapley effects, using Gaussian processes in Benoumechiara and Elie-Dit-Cosaque (2019)[12]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.

[12]In the vein of this work, a Python implementation using `OpenTURNS` is available in the GitHub repository: https://github.com/josephmure/otshapley.

1.7.3 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 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) to detect the dependence, 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 $\mathcal{H}$, which is a Reproducing kernel Hilbert space (RKHS):

$$\mathrm{MMD}(\pi,\zeta) = \sup_{\|g\|_{\mathcal{H}(k)}\leq 1} \left| \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x})\mathrm{d}\pi(\mathbf{x}) - \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x})\mathrm{d}\zeta(\mathbf{x}) \right| \tag{1.79}$$

This quantity is a distance in the 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 $\left\{\pi^{(i)}\right\}_{i=1}^n \sim \pi$ and $\left\{\zeta^{(j)}\right\}_{j=1}^n \sim \zeta$:

$$\widehat{\mathrm{MMD}}^2(\pi,\zeta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n k\left(\pi^{(i)},\pi^{(j)}\right) - k\left(\pi^{(i)},\zeta^{(j)}\right) - k\left(\zeta^{(i)},\pi^{(j)}\right) + k\left(\zeta^{(i)},\zeta^{(j)}\right), \tag{1.80}$$

where $k(\cdot, cdot)$ is a symmetric and positive definite kernel with specific properties (see Appendix B for further details).

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

$$S_j^{\mathrm{MMD}} = \mathrm{MMD}(\mathbb{P}_Y, \mathbb{P}_{Y|X_j}). \tag{1.81}$$

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

Alternatively, the second option consists in considering a couple of random variables (X_j, Y), with probability distributions $\mathbb{P}_{X_j}$ and $\mathbb{P}_Y$, and assuming 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 $\mathbb{P}_{X_j} \otimes \mathbb{P}_Y$ and is given by:

$$\mathrm{HSIC}(X_j, Y) = \mathrm{MMD}^2(\mathbb{P}_{(X_j,Y)}, \mathbb{P}_{X_j} \otimes \mathbb{P}_Y). \tag{1.82}$$

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. (1.81) and Eq. (1.82), 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.

1.7.4 Summary and discussion

This section introduces the GSA methods commonly used in UQ. 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 UQ study. As for other steps of the generic UQ methodology, GSA 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 reinterpretation of the FORM importance factors by Papaioannou and Straub (2021)), or were adapted from GSA, like Sobol' indices (Wei et al., 2012; Chabridon, 2018; Perrin and Defaux, 2019; Ehre et al., 2020), Target-HSIC (Marrel and Chabridon, 2021), or Shapley effects (Il Idrissi et al., 2021).

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 UQ 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 pertubate uncertainty models, some remain in the probabilistic framework, such as the "perturbed-law based indices" (PLI) (Lemaître et al., 2015; Iooss et al., 2022), while other rely on extra-probabilistic methods (Ajenjo et al., 2022).

1.8 Surrogate modeling

1.8.1 Common framework

The aim of *surrogate modeling* (or metamodeling) is to build a cheap-to-call statistical model, denoted by $\widehat{g}_n(\cdot)$, to replace a costly numerical model $g(\cdot)$ over the input domain $\mathcal{D}_{\mathbf{X}}$. To do so, a statistical learning is performed on a finite number of observations of the costly function g. When manipulating computationally expensive simulations, the sample size can be limited (i.e., in a small-data context). This n-sized sample is usually called *learning set* and written:

$$\{\mathbf{X}_n, \mathbf{y}_n\} = \left\{\mathbf{x}^{(i)}, y^{(i)}\right\}_{i=1}^{n} = \left\{\mathbf{x}^{(i)}, g\left(\mathbf{x}^{(i)}\right)\right\}_{i=1}^{n}. \tag{1.83}$$

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 (Cortes and Vapnik, 1995), Gaussian processes (GP) (Rasmussen and Williams, 2006), low-rank tensor approximations (Grasedyck et al., 2013), and artificial neural network (Hastie et al., 2009). The following section will provide a short focus on GP regression.

Validating the accuracy and precision of a surrogate model is an crutial 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 $\{\mathbf{X}_m, \mathbf{y}_m\} = \left\{\mathbf{x}^{(i)}, g\left(\mathbf{x}^{(i)}\right)\right\}_{i=1}^{m}$. Note that the analyst may work in two different frameworks, affecting the regression and validation method's choice:

- The given-data context: only using a fixed input-output dataset to build and validate the surrogate model.
- The 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 a challenge. Multiple validation criteria and techniques exist. The *regression coefficient*, denoted by R^2, is the first validation metric that can be directly computed on the learning set:

$$\mathrm{R}^2(\widehat{g}_n) = 1 - \frac{\sum_{i=1}^{n}\left(y(\mathbf{x}^{(i)}) - \widehat{g}(\mathbf{x}^{(i)})\right)^2}{\sum_{i=1}^{n}\left(y(\mathbf{x}^{(i)}) - \overline{\mathbf{y}}_n\right)^2}, \tag{1.84}$$

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

$$\mathrm{Q}^2(\widehat{g}_n) = 1 - \frac{\mathrm{ISE}(\widehat{g}_n)}{\mathrm{Var}(g(\mathbf{X}))}, \tag{1.85}$$

where

$$\mathrm{ISE}(\widehat{g}_n) = \int_{\mathcal{D}_\mathbf{X}} \left(g(\mathbf{x}) - \widehat{g}(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x}, \qquad \mathrm{Var}(g(\mathbf{X})) = \int_{\mathcal{D}_\mathbf{X}} \left(g(\mathbf{x}) - \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x})\,\mathrm{d}\mathbf{x}'\right)^2 \mathrm{d}\mathbf{x}. \tag{1.86}$$

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

$$\widehat{\mathrm{Q}}^2(\widehat{g}_n) = 1 - \frac{\sum_{i=1}^{m}\left(y(\mathbf{x}^{(i)}) - \widehat{g}(\mathbf{x}^{(i)})\right)^2}{\sum_{i=1}^{m}\left(y(\mathbf{x}^{(i)}) - \overline{\mathbf{y}}_m\right)^2}. \tag{1.87}$$

Note that for either criterion, the closer to one, 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).

1.8.2 Focus on Gaussian process regression

In this subsection, a particular focus is dedicated to GP regression which will be useful in the following chapters (GP is also called Kriging after the geostatician D.G. Krige). GPs 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

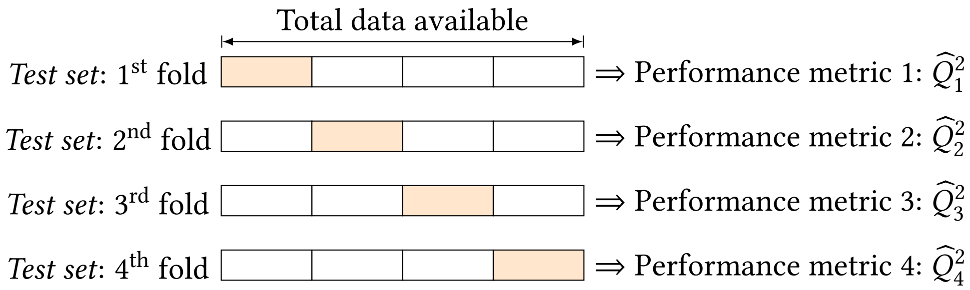


Figure 1.12 Illustration of a k-fold cross-validation (with $k = 4$).

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 1.5 are widely used.

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

$$G \sim \mathrm{GP}(m(\cdot), k(\cdot,\cdot)), \tag{1.88}$$

with:

- a *trend model*: $m(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$, composed of a functional basis $\mathbf{f} = (f_1, \ldots, f_d)^\top$ and a vector of coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$,
- a *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. GP 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.

To ease the presentation, let us first assume the hyperparameters $(\sigma^2, \boldsymbol{\theta})$ are fully known and the trend is zero: $\boldsymbol{\beta} = \mathbf{0}$. At a given point $\mathbf{x} \in \mathcal{D}_{\mathbf{X}}$ the realization of the GP is a Gaussian random variable, i.e., $G(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}))$. 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 the "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 the explicit expression of the parameters of this distribution:

$$\begin{cases} \eta_n(\mathbf{x}) & = \mathbf{k}^\top(\mathbf{x}) \mathbf{K}^{-1} \mathbf{y}_n \\ s_n^2(\mathbf{x}) & = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) \end{cases}, \tag{1.89}$$

where $\mathbf{k}(\mathbf{x})$ is the column vector of the covariance kernel evaluations $[k(\mathbf{x}, \mathbf{x}^{(1)}), \ldots, k(\mathbf{x}, \mathbf{x}^{(n)})]$ 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 remarkable 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).

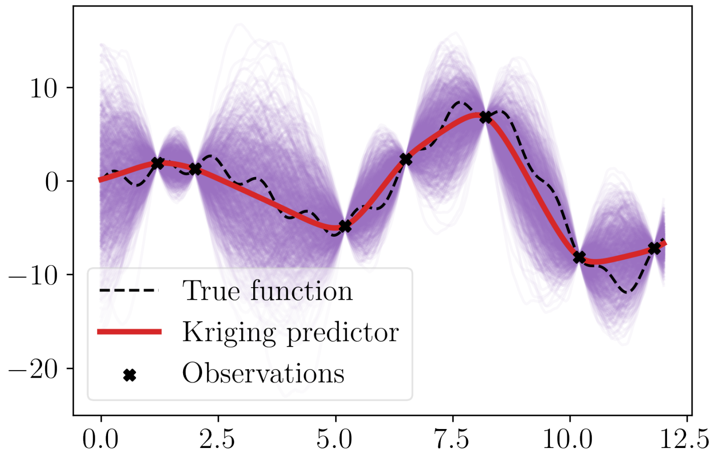


Figure 1.13 Illustration of an ordinary Kriging model fitted on a limited set of observations ($n = 7$). The predictor is represented in and several trajectories of the conditioned GP are drawn and represented in purple.

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 enough (see e.g., Bachoc, 2013). For a validation performed by holdout, and using an independent m-sized test set, the PVA is defined as:

$$\mathrm{PVA} = \left| \log \left(\frac{1}{m} \sum_{i=1}^{n} \frac{\left(y(\mathbf{x}^{(i)}) - \widehat{g}(\mathbf{x}^{(i)}) \right)}{s_n^2(\mathbf{x}^{(i)})} \right) \right| . \tag{1.90}$$

The closer to zero this quantity gets, the better the quality of the Kriging variance.

Despite its numerous advantages, GP regression reveals some numerical issues 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 as presented in Geoga et al., 2020).

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].

1.8.3 Goal-oriented active surrogate model

Surrogates are often fitted for specific purposes, 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. This approach is called a *goal-oriented learning*, rather than learning uniformly over the entire domain. For example, to fit a surrogate model for reliability analysis, one should concentrate the learning set around the LSF. Similarly, to build a surrogate for a global optimization problem, one should focus the learning set around the global optimum. Unfortunately, the areas of interest are 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"), which depends on the purpose of surrogate. 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 global optimization

In the field of black-box optimization, many methods rely on approximating the function by a surrogate. The use of GP 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 drive the selection of the next point. Factually, the learning criterion depends on the GP variance model. Numerous reviews of this field were proposed by Shahriari et al. (2015), or Gramacy (2020) and numerical benchmarks presented in Le Riche and Picheny (2021).

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

$$\mathbf{x}^* = \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{D}_{\mathbf{X}}} g(\mathbf{x}). \tag{1.91}$$

[13] https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/surrogates.ipynb
[14] https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter1/active_learning.ipynb

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 input domain. A first surrogate denoted by $G_n(\mathbf{x}) \sim \mathcal{N}(\eta_n(\mathbf{x}), s_n^2(\mathbf{x}))$ is fitted using Eq. (3.11). The expected improvement (EI), to be maximized, is then written as:

$$\mathcal{A}^{\mathrm{EI}}(\mathbf{x}; \mathbf{y}_n) = \mathbb{E}\left[\max\left(g_{\min} - G_n(\mathbf{x})\right)\right] \tag{1.92}$$

$$= \left(g_{\min} - \eta_n(\mathbf{x})\right) \Phi\left(\frac{g_{\min} - \eta_n(\mathbf{x})}{s_n(\mathbf{x})}\right) + s_n(\mathbf{x})\, \phi\left(\frac{g_{\min} - \eta_n(\mathbf{x})}{s_n(\mathbf{x})}\right), \tag{1.93}$$

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 is used to progressively enhance the GP in order 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 strategies were also adapted for structural reliability problems.

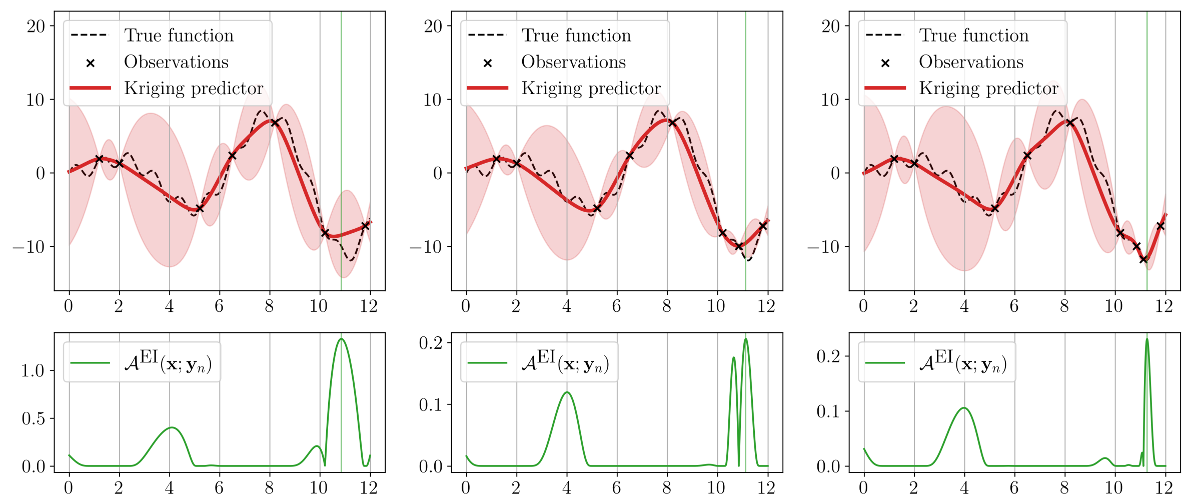


Figure 1.14 Illustration of the expected improvement (EI) learning criterion.

Active Kriging for reliability analysis

Rare event estimation often requires large amounts of evaluations of the LSF (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 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 LSF. 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 binary classification. For example, active learning procedures using SVM classifiers have been adapted to this specific goal (Bourinet, 2018).

The following paragraphs introduce a popular Kriging-based learning criterion called the "deviation number" U (Echard et al., 2011), which is used in a generic algorithm named "active Kriging" (AK). 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. (3.11). The deviation number U is looking for points close to the LSF while presenting a high Kriging variance. This criterion to be minimized is defined as:

$$\mathcal{A}^{\mathrm{U}}(\mathbf{x}; \mathbf{y}_n) = \frac{|y_{\mathrm{th}} - \eta_n(\mathbf{x})|}{s_n^2(\mathbf{x})}, \tag{1.94}$$

where $y_{\mathrm{th}} \in R$ is a threshold defining the failure domain such as $\{\mathbf{x} \in \mathcal{D}_{\mathbf{X}} \,|\, g(\mathbf{x}) \leq y_{\mathrm{th}}\}$.

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}. Here, three iterations of the active Kriging 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 active methods are used here for rare event estimation is the result of the approximation of the LSF and a sampling technique. AK methods were coupled with most sampling techniques introduced in Section 1.6 (e.g., AK-MCS, AK-IS, AK-SS). In practice, note that the agnostic strategy recommended after the wide benchmark of Moustapha et al. (2022), consists in coupling an AK method (using the learning function U) with 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, *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.

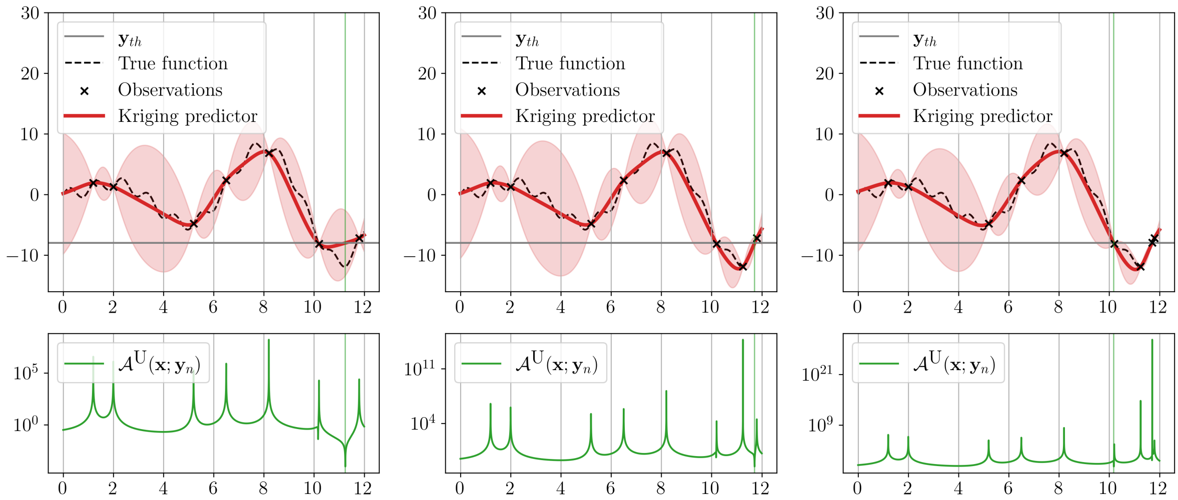


Figure 1.15 Illustration of the deviation number learning criterion

1.8.4 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 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.To enhance the learning for specific purposes (e.g., optimization or reliability analysis), 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 dedicated techniques introduced in Subsection 1.7.1, otherwise, model order reduction methods might be used (Schilders et al., 2008). When the function is stochastic, several approaches allow fitting the function and its intrinsic variability (Binois et al., 2019; Baker et al., 2022; Zhu, 2022).

1.9 Conclusion

This section gives a literature overview of the main steps in UQ. From uncertainty modeling, uncertainty propagation, and global sensitivity analysis to surrogate modeling. To ease the methodological presentation, all the illustrations from this section are reproducible using the Python/OpenTURNSscripts 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 UQ, the model should be calibrated to make it match some physical information (e.g., measurements). 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

Chapter 2

Kernel-based uncertainty quantification

Parts of this chapter are adapted from the following publications:

- E. Vanem, E. Fekhari, N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2024). “A joint probability distribution model for multivariate wind and wave conditions”. In: *Journal of Offshore Mechanics and Arctic Engineering*, In press.
- E. Vanem, Ø. Lande and E. Fekhari, (2024). “A simulation study on the usefulness of the Bernstein copula for statistical modeling of metocean variables”. In: *Proceedings of the ASME 2024 43th International Conference on Ocean, Offshore and Arctic Engineering (to appear).*
- 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).*

2.1 Introduction

The main sources of solicitation in offshore design reside in the metocean conditions. To accurately verify a structural design against the joint wind and wave conditions, these random excitations must be carefully modeled. Offshore structures are usually certified against ultimate limit states (related to the occurrence of extreme metocean conditions) and fatigue limit states (related to the average fatigue over the metocean conditions). In this context, the probabilistic framework is typically used to model the joint distribution of random variables describing the metocean conditions (listed in Section **??**).

Note that a given probabilistic model might describe well the central behavior of the environmental distribution but not its tail behavior (and vice-versa). Extreme value theory develops specific methods to model the far tails of distributions (Beirlant et al., 2006). Modeling the tails is not the priority in the present work since the focus is on mean fatigue estimation.

The environmental random variables studied present different particularities. First, an offshore wind turbine project leads to the collection of an large amount of metocean data (possibly merged with data from mesoscale simulations). Second, their dependence structure is complex, making the probabilistic modeling more complicated.

This chapter explores different aspects of the environmental conditions’ uncertainty quantification. The theory of some nonparametric copulas is introduced before their use in metocean conditions inference. A semiparametric approach is applied to the South Brittany data, mixing parametric modeling of the marginals with nonparametric modeling of the copula. To visually analyze multivariate distributions, the *copulogram* is a new tool that decomposes the marginal effects and the dependence structure of a joint distribution.

At the scale of a wind farm, each turbine perceives different metocean conditions as the wake of other turbines creates wind perturbations. To study this perturbation, an engineering wake model (see Subsection **??**) was used to obtain one perturbed environmental distribution per turbine. This work applies a kernel-based discrepancy (the maximum mean discrepancy) to compare wake-induced perturbations. In a second phase, this discrepancy is used to gather wind turbines perceiving similar perturbations. This clustering can be used to perform uncertainty propagation at the farm scale by considering a few turbines with are representative of a cluster.

2.2 Dependence modeling with nonparametric copula

In uncertainty quantification, the lack of knowledge can lead to rough assumptions regarding the dependence modeling. However, an accurate representation of the uncertain inputs is of prime importance. For example, the work of Torre et al. (2019) demonstrates the influence of the dependence model on the estimation of rare event probabilities by studying the same problem with different copula models.

When inferring a probabilistic model over a multivariate dataset, one can decompose the problem into the fit of a set of marginals and the fit of a copula (see the Sklar Theorem 1). In the case of metocean conditions, the fit of the marginals is not problematic considering the amount of data available. However, the complex dependence structure appears to be more challenging. Different strategies to model the dependence for multivariate distributions are briefly summarized hereafter:

- **Vine copulas** (also known as pair copula) decompose the joint distribution as a product of conditioned bivariate copulas organized in a tree-like structure called a vine. This approach proved to be very efficient, but it requires the definition of the vine and the bivariate parametric copulas (Joe and Kurowicka, 2011).
- **Conditional modeling** defines the joint distribution as a product of univariate conditional distributions. In practice, the parameter of a marginal is defined as a function of other marginals (see e.g., Vanem et al., 2023).
- **Multivariate KDE** is another way to capture the dependence together with marginal effects. As the dimension and the size of the dataset increase, this method becomes less tractable (Wand and Jones, 1994).
- **Nonparametric copulas** are methods uniformly approximating an empirical copula without any assumption on a dependence structure. They will be further described and used for metocean conditions inference in the present chapter.

Remark 2. The strategy referred to as "conditional modeling" can in fact be expressed as a copula (Vanem, 2016) for continuous variables. For example, in the bivariate case of a continuous random vector $\mathbf{X} = (X_1, X_2)$ with PDF $f_{\mathbf{X}}(\mathbf{x})$, CDF $F_{\mathbf{X}}(\mathbf{x})$ and density copula c:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)\, f_{X_2|X_1}(x_2|x_1) = f_{X_1}(x_1)\, f_{X_2}(x_2)\, c\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) \tag{2.1a}$$

$$\Leftrightarrow c\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) = \frac{f_{X_1}(x_1)\, f_{X_2|X_1}(x_2|x_1)}{f_{X_1}(x_1)\, f_{X_2}(x_2)} \tag{2.1b}$$

The notions related to the copula theory are further introduced in the monographs of Nelsen (2006); Joe (2014); Durante and Sempi (2015) while the key properties are introduced hereafter.

2.2.1 Preliminary definitions and properties

Let us consider a random vector $\mathbf{X} \in \mathcal{D}_{\mathbf{X}} \subseteq \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Its probability distribution $\mathbb{P}_{\mathbf{X}}$ can be represented by a CDF $F_{\mathbf{X}}$ and PDF $f_{\mathbf{X}}$. The functional definition of a *d-dimensional copula* (or simply "d-copula") is a density function $C : [0,1]^d \mapsto [0,1]$ whose marginals are uniformly distributed on $[0,1]$.

Theorem 3 (Copula). *A function $C : [0,1]^d \mapsto [0,1]$ is a d-copula if, and only if, it presents the following properties:*

- *The function C is "grounded" (also called "anchored"):* $C(u_1, \dots, u_d) = 0$ *if* $u_j = 0, \forall j \in \{1, \dots, d\}$;
- *The marginals of C are uniform, then:* $C(1, \dots, u_j, \dots, 1) = u_j, \forall j \in \{1, \dots, d\}$;
- *The function C is "d-increasing", meaning that for any hyperrectangle $A \subset [0,1]^d$, the corresponding volume induced by C is positive (see Durante and Sempi (2015) p.7).*

A copula is bounded by two functions according to the Fréchet-Hoeffding bounds.

Theorem 4 (Fréchet-Hoeffding bounds). *If a function $C : [0,1]^d \mapsto [0,1]$ is a d-copula, then it respects the following bounds for all $\mathbf{u} \in [0,1]^d$:*

$$W(\mathbf{u}) = \max(1 - d + u_1 + \cdots + u_d, 0) \leq C(\mathbf{u}) \leq M(\mathbf{u}) = \min(u_1, \dots, u_d). \tag{2.2}$$

Where the upper bound M is still a copula while the lower bound W is only one for $d = 2$.

The rank transform plays an essential role in understanding copulas. Considering a continuous random vector $\mathbf{X} \in \mathcal{D}_{\mathbf{X}}$ and the sample $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\} \sim \mathbf{X}$, its *ranks* $\mathbf{R}_n = \left\{\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(n)}\right\} \in \mathbb{N}^n$ correspond to the indexes of its order statistics:

$$r_j^{(i)} = n \widehat{F}_{X_j}(x_j^{(i)}) = \sum_{l=1}^{n} \mathbb{1}_{\left\{x_j^{(l)} \leq x_j^{(i)}\right\}}, \quad \forall j \in \{1, \dots d\}, i \in \{1, \dots n\}, \tag{2.3}$$

where $\widehat{F}_{X_j}$ stands for the marginal empirical CDF associated with the random variable X_j.

Theorem 5 (Rank-invariance). *Considering a random vector $\mathbf{X} = (X_1, \dots, X_d)$, a set of mappings $\{r_j(\cdot)\}_{j=1}^d$, and the image random vector $\mathbf{R} = (r_j \circ X_1, \dots, r_d \circ X_d)$. If the mappings are strictly increasing (which is the case for the rank transform introduced in Eq. (2.3)), then, the copula associated to $\mathbf{R}$ is invariant by transformation: $C_{\mathbf{X}} = C_{\mathbf{R}}$. A proof is presented in Durante and Sempi (2015) p. 57.*

Transforming in the ranks generally reduces the effect of outliers and ensures more robust estimates. The invariance by the rank transform of copulas allows the estimation of different *dependence measures* in the ranked space.

Spearman's rho. Is a well-known dependence measure, also called the "Spearman's rank correlation coefficient", which is defined for two random variables X_i, X_j as:

$$\rho^{\mathrm{S}}(X_i, X_j) = \frac{\mathrm{Cov}(r_i(X_i), r_j(X_j))}{\sigma_{r_i(X_i)}\,\sigma_{r_j(X_j)}}, \tag{2.4}$$

an equivalent definition exists, using the copula C between the joint distribution of X_i and X_j and the independent copula $\Pi(u_i, u_j) = u_i\,u_j$:

$$\rho^{\mathrm{S}}(X_i, X_j) = 12\int_{[0,1]^2} C(u_i, u_j)\,\mathrm{d}u_i\mathrm{d}u_j - 3 = 12\int_{[0,1]^2} (C(u_i, u_j) - \Pi(u_i, u_j))\,\mathrm{d}u_i\mathrm{d}u_j. \tag{2.5}$$

Kendall's tau. Also referred to as the "Kendall's rank correlation coefficient", is defined for a pair of random variables (X_i, X_j) and their respective independent copies (X_i', X_j') as:

$$\tau(X_i, X_j) = \mathbb{P}\left((X_i - X_i')(X_j - X_j') > 0\right) - \mathbb{P}\left((X_i - X_i')(X_j - X_j') < 0\right), \tag{2.6}$$

and can also be defined using the copula C between the joint distribution of the two random variables:

$$\tau(X_i, X_j) = 4\int_{[0,1]^2} C(u_i, u_j)\,\mathrm{d}C(u_i, u_j) - 1 = 1 - 4\int_{[0,1]^2} \frac{\partial C(u_i, u_j)}{\partial u_i}\frac{\partial C(u_i, u_j)}{\partial u_j}\,\mathrm{d}u_i\mathrm{d}u_j \tag{2.7}$$

These dependence measures fully rely on the copula and are both bounded between -1 and 1. Further properties and estimators of Spearman's rho and Kendall's tau are presented in Durante and Sempi (2015) Section 2.4.

Upper/lower tail dependence. Considering the random vector $\mathbf{X} = (X_i, X_j)$ and the copula C underlying their joint distribution. The *upper/lower tail dependence* coefficients are defined as:

$$\lambda_U(X_i, X_j) = \lim_{\substack{u \to 1 \\ u<1}} \mathbb{P}\left(X_i > F_{X_i}^{-1}(u) | X_j > F_{X_j}^{-1}(u)\right) = \lim_{\substack{u \to 1 \\ u<1}} \left(2 - \frac{1 - C(u,u)}{1-u}\right) \tag{2.8a}$$

$$\lambda_L(X_i, X_j) = \lim_{\substack{u \to 0 \\ u>0}} \mathbb{P}\left(X_i \leq F_{X_i}^{-1}(u) | X_j \leq F_{X_j}^{-1}(u)\right) = \lim_{\substack{u \to 0 \\ u>0}} \left(\frac{C(u,u)}{1-u}\right) \tag{2.8b}$$

Joe (2014) further discusses the asymptotic limit and outlines the particular case of the bivariate Gaussian copula, for which the tail dependence measures are null, $\lambda_U = \lambda_L = 0$. Note that Kendall's tau and the tail dependence coefficients both have their associated plots, allowing us to compare the dependence of two distributions [add ref].

2.2.2 Empirical and checkerboard copula

The *empirical copula* was introduced by Deheuvels (1979), as an estimator of the copula C associated with the random vector $\mathbf{X}$. Since the normalized ranks are a creasing mapping that presents uniform marginals by construction, they are a natural empirical representation of the density copula. Considering a sample $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\} \sim \mathbf{X}$ with the respective ranks $\mathbf{R}_n = \left\{\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(n)}\right\}$, a definition of the empirical copula is:

$$C_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbb{1}\left\{\frac{r_j^{(i)}}{n} \leq u_j\right\}, \quad \mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d \tag{2.9}$$

Even if this function converges uniformly towards the copula C (according to the Glivenko-Cantelli theorem), it does not fulfill the conditions to be a copula (see e.g., González-Barrios and Hoyos-Argüelles, 2021).

In this context, different methods may be applied to smooth the empirical copula into a genuine copula. This problem can be perceived as a functional approximation of the underlying copula C, which is unique for continuous variables (according to Sklar's Theorem 1). Let us consider a discretization of the unit hypercube as a grid:

$$G = \left\{\frac{0}{m_1}, \dots, \frac{m_1}{m_1}\right\} \times \dots \times \left\{\frac{0}{m_d}, \dots, \frac{m_d}{m_d}\right\}, \quad \mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d. \tag{2.10}$$

The *checkerboard copula* is a simple approximation of the empirical copula using the discretization G. This method is comparable to a multivariate histogram of the empirical density copula c_n (see the formal multivariate definition proposed by Cottin and Pfeifer, 2014). In the particular case for which $m_j = m, \forall j \in \{1, \dots, d\}$, the checkerboard copula is called the "rook" copula, and expressed by Segers et al. (2017) as:

$$C_n^{\#m}(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \min\left(\max(n u_j - r_j^{(i)} + 1, 0), 1\right). \tag{2.11}$$

This empirical copula has a low complexity (see Rose, 2015) and efficient results for large samples (González-Barrios and Hoyos-Argüelles, 2021), however, its variance is comparable to the empirical copula for small-sized samples (Segers et al., 2017). It is proven to be a genuine copula and its asymptotic behavior was studied by various authors such as Li et al. (1998); Genest et al. (2017). In the following, an approximation of the empirical copula with Bernstein polynomials is presented.

2.2.3 Empirical Bernstein and Beta copula

A few elements about Bernstein polynomials and their corresponding approximation are reminded before introducing the empirical Bernstein copula.

Bernstein polynomials and approximation

Let us first define the *Bernstein basis polynomial* of order $m \in \mathbb{N}$ as:

$$b_{m,t}(u) = \binom{m}{t} u^t (1-u)^{m-t}, \quad t \in \{0, \dots, m\}. \tag{2.12}$$

These polynomials present various interesting properties, such as their nonnegativity over $[0,1]$, being bounded by one, and offering a partition of unity on $[0,1]$ (Lasserre, 2023):

$$1 = \sum_{t=0}^{n} b_{m,t}(u)(x), \quad \forall x \in \mathbb{R}, \quad \forall n \in \mathbb{N}. \tag{2.13}$$

Bernstein's polynomials allow us to uniformly approximate any continuous and real-valued function defined on a compact set $f : [0,1]^d \mapsto \mathbb{R}$ (as they were used to demonstrate the Weierstrass approximation theorem). In the multivariate case, the *Bernstein approximation* of the function f can be written on a grid over the unit hypercube $G = \left\{\frac{0}{m_1}, \dots, \frac{m_1}{m_1}\right\} \times \dots \times \left\{\frac{0}{m_d}, \dots, \frac{m_d}{m_d}\right\}$, $\mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d$, as:

$$B_{\mathbf{m}}(f)(\mathbf{u}) = \sum_{t_1=0}^{m_1} \dots \sum_{t_d=0}^{m_d} f\left(\frac{t_1}{m_1}, \dots, \frac{t_d}{m_d}\right) \prod_{j=1}^{d} b_{m_j,t_j}(u_j), \quad \mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d. \tag{2.14}$$

The Bernstein polynomials approximate f such that $\lim_{m\to\infty} B_m(f) = f$ uniformly on $[0,1]$.

Bernstein polynomials for copula approximation

Copulas are continuous and bounded functions defined on a compact set (the unit hypercube). Therefore, they are good candidates to be approximated by Bernstein polynomials. The Bernstein approximation applied on an empirical copula C_n was introduced as *empirical Bernstein copula* (EBC) by Sancetta and Satchell (2004) for applications in economics and risk management:

$$B_{\mathbf{m}}(C_n)(\mathbf{u}) = \sum_{t_1=0}^{m_1} \dots \sum_{t_d=0}^{m_d} C_n\left(\frac{t_1}{m_1}, \dots, \frac{t_d}{m_d}\right) \prod_{j=1}^{d} b_{m_j,t_j}(u_j), \quad \mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d. \tag{2.15}$$

In this expression, the evaluations of the empirical copula on the vertices of the grid are smoothed by the product of Bernstein polynomials. A respective approximation of the copula density can be directly expressed by deriving the previous formula. The EBC delivers a genuine copula, if and only if all the polynomial degrees $\{m_j\}_{j=1}^d$ are divisors of n (see Segers et al., 2017, Proposition 2.5).

In the particular case of regular grids, $\{m_j = m\}_{j=1}^d$, the EBC can be expressed as a mixture of beta distributions (Segers et al., 2017). Let us consider an n-sized rank sample, $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_d) \in \mathbb{N}^n$, and the degree m taken as divisor of n. Note that the r$^{\text{th}}$ order statistic of an n-sized sample following a uniform $[0,1]$ is distributed according to the beta distribution $\mathcal{B}(r, n-r+1)$.

Considering these hypotheses, the EBC can be written as:

$$B_m(C_n)(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} F_{m,r_j^{(i)}}, \quad \mathbf{u} = (u_1,\ldots,u_d) \in [0,1]^d, \tag{2.16}$$

where $F_{m,r}$ is the CDF of the beta distribution $\mathcal{B}(r, m-r+1)$ (also called the "regularized incomplete beta function"):

$$F_{m,r} = \sum_{t=r}^{m}\binom{m}{t}u^t(1-u)^{m-t}, \quad u \in [0,1], \quad r \in \{1,\ldots,m\}. \tag{2.17}$$

Overall, the EBC is a very versatile tool which able to approximate complex dependence patterns. Moreover, Monte Carlo sampling on an EBC is straightforward and licit since it is a genuine copula. As a drawback, the estimation accuracy of this nonparametric method heavily relies on polynomial order tuning.

Asymptotic behavior of the empirical Bernstein copula

In practice, the choice of polynomial degree for an EBC leads to a challenging bias-variance tradeoff. For example, the particular case of $\{m = n\}$, introduced as the *empirical Beta copula* by Segers et al. (2017), tends to reduce the bias while increasing the variance. In this paper, the beta copula presents interesting results compared to the Bernstein or the checkerboard copula for small sample sizes (i.e., $n < 100$). Theoretically, the tuning of the degree was first optimized to minimize an "Asymptotic Mean Integrated Squared Error" (AMISE) of $B_{\mathbf{m}}(C_n)$:

$$\text{AMISE}(B_{\mathbf{m}}(C_n)) = \mathbb{E}\left[\|B_{\mathbf{m}}(C_n) - C\|_2^2\right] = \mathbb{E}\left[\int_{\mathbb{R}^d}(B_{\mathbf{m}}(C_n)(\mathbf{u}) - C(\mathbf{u})\mathrm{d}\mathbf{u})^2\right]. \tag{2.18}$$

The seminal work of Sancetta and Satchell (2004) proves in Theorem 3 that:

- $B_{\mathbf{m}}(C_n)(\mathbf{u}) \longrightarrow C(\mathbf{u})$ for any $u_j \in]0,1[$ if $\frac{m^{d/2}}{n} \longrightarrow 0$, when $m, n \longrightarrow \infty$.
- The optimal polynomial order in terms of AMISE is[1]: $m \lesssim m_{\text{AIMSE}} = n^{2/(d+4)}, \forall u_j \in]0,1[$.

To illustrate the previous theorem, Fig. 2.1 represents the evolution of the m_{AMISE} for different dimensions and sample sizes (adapted from Lasserre, 2022). In medium dimension, the values of m_{IMSE} tend towards one, which is equivalent to the independent copula. Therefore, high-dimensional problems should rather be divided into a product of smaller problems on which the EBC is tractable.

The polynomial order for EBC estimation is still a bottleneck that was studied over the years by different authors (see e.g., Janssen et al., 2012; Bouezmarni et al., 2013; Rose, 2015; Segers et al., 2017). Meanwhile, other nonparametric approaches such as the "penalized Bernstein" and the "penalized B-spline" estimators were compared to the EBC and vine copulas in a benchmark

[1]The sign $\lesssim$ stands for "less than or approximately".

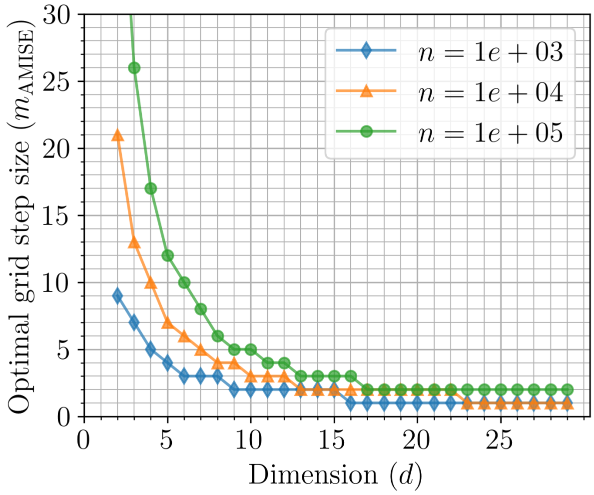


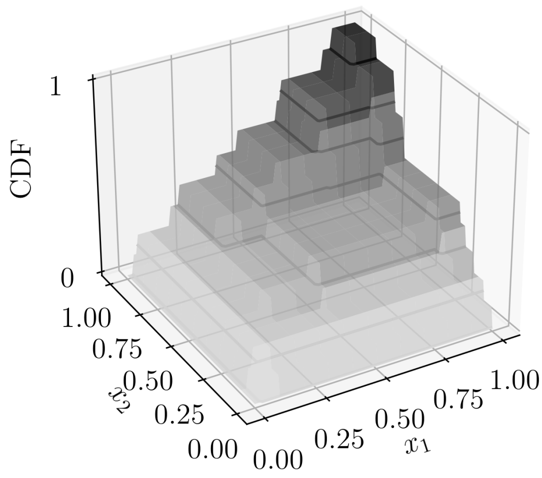
Figure 2.1 Evolution of m_{IMSE} for different dimensions and sample sizes.

realized by Nagler et al. (2017). The results showed that the most performant methods vary depending on the problem studied (for different dimensions, sample sizes, and strength of dependence). Regarding tail dependence modeling, nonparametric approaches are generally limited, but recent contributions introduced Bootstrap procedures to better this aspect (Kiriliouk et al., 2021).

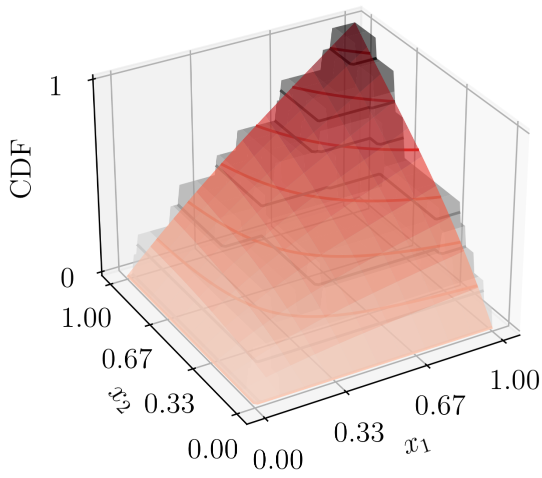
Illustrative example on a Clayton copula

Let us consider a bivariate Clayton copula C with parameter $\theta = 2.5$ (see Nelsen, 2006) A Monte Carlo sample with size $n = 10$ is generated on it, which is then used to build an empirical copula C_n as defined in Eq. (2.9). Fig. 2.2 (a) illustrates the empirical copula corresponding to the sample, with the shade of grey matching the CDF values. Then, the Bernstein approximation of the empirical copula (i.e., the EBC) is represented in Fig. 2.2 (b), (c), (d) according to the Eq. (2.15). The three versions of the EBC correspond to different polynomial orders, assuming that $(m_1 = m_2)$.

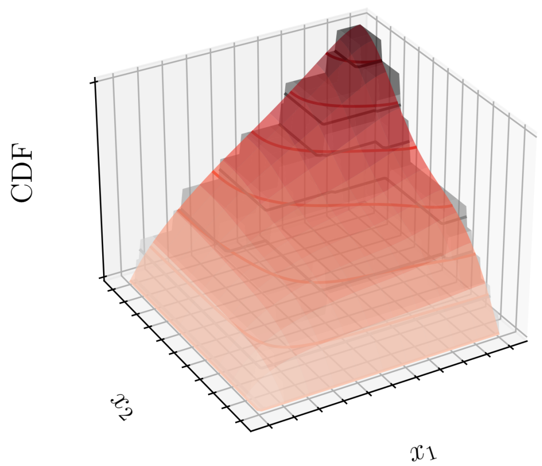
As the order increases, the bias between the EBC and the copula C tends to be reduced. Note that the second EBC in Fig. 2.2 (c) where $(m_1 = m_2 = n)$ is equivalent to the Beta copula. Moreover, increasing the order beyond the sample size definitely overfits the copula.



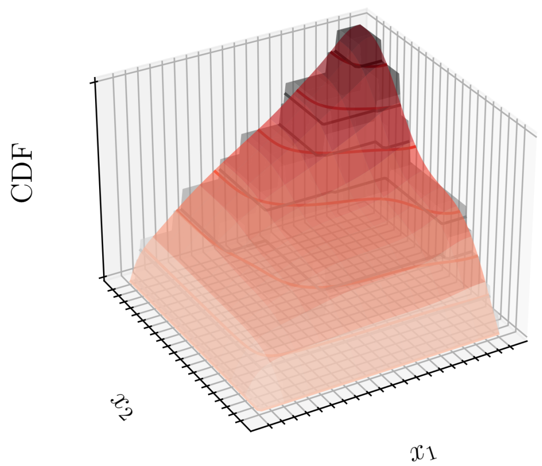
(a) Empirical copula C_n with $(n = 10)$.



(b) $B_{\mathbf{m}}(C_n)$ with $(m_1 = m_2 = 3)$.



(c) $B_{\mathbf{m}}(C_n)$ with $(m_1 = m_2 = 10)$.



(d) $B_{\mathbf{m}}(C_n)$ with $(m_1 = m_2 = 20)$.

Figure 2.2 Bernstein approximations of the empirical copula C_n (with size $n = 10$) of a Clayton copula (with parameter $\theta = 2.5$). The polynomial orders are assumed equal in the two dimensions $m_1 = m_2 \in \{3, 10, 20\}$.

2.3 *Copulogram*: a tool for multivariate data visualization

In statistics, data visualization offers a wide set of tools to analyze data. Multivariate data visualization is of great help in apprehending problems with dimensions higher than two. In the context of continuous variables, let us consider the n-sized sample $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\} \sim \mathbf{X} \in \mathcal{D}_\mathbf{x} \subseteq \mathbb{R}^d$. The marginal samples of $\mathbf{X}_n$ are denoted by $X_{n,j} = \{x_j^{(1)}, \dots, x_j^{(n)}\}, j \in \{1, \dots, d\}$.

Various techniques exist to represent multivariate data, such as the "parallel coordinate plot", also called "cobweb plot" (see e.g., Heinrich and Weiskopf, 2013). For each sample $\mathbf{x}^{(i)} \in \mathbf{X}_n$, this plot draws a line passing by the values of $\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_d^{(i)}]$. This representation was used in sensitivity analysis to illustrate the connections between a set of inputs and an output, however, it does not provide a good representation of the dependence structure between the inputs.

2.3.1 From the pairwise plot to the copulogram

Alternatively, the "pairwise plot", also named "generalized draftsman plot", was initially introduced by Hartigan (1975) to draw a matrix of scatter-plots between all the pairs of marginal samples $\{X_{n,i}, X_{n,j}\}, i \neq j \in \{1, \dots, d\}^2$. Because of the symmetry, the pairwise plot is usually represented on the lower triangle of the matrix. Later on, statisticians improved the pairwise plot by adding a histogram (or KDE) of the marginal samples $X_{n,j} j \in \{1, \dots, d\}$ on the diagonal. Additionally, the upper triangle was completed with the values of linear correlation for each pair of marginal samples $\{X_{n,i}, X_{n,j}\}, i \neq j$. This matrix of correlation coefficients is also known as "correlogram". Altogether, this matrix plot became known as the "scatter plot of matrices" (SPLOM).

However, the linear correlation coefficient is known to give a poor description of the dependence in nonlinear cases. When analyzing a continuous sample $\mathbf{X}_n \sim \mathbf{X}$, the Sklar theorem states that the dependence structure within the random vector $\mathbf{X}$ has a unique expression with its d-copula C. As mentioned in Subsection 2.2.1, the component-wise normalized ranks of the original sample $\mathbf{X}_n$ define the empirical copula density c_n (converging towards C as n increases).

To the best of our knowledge, the *copulogram* is a new multivariate data visualization tool improving the SPLOM by representing the empirical copula density c_n on the upper triangle of the matrix plot. This plot is an empirical decomposition of a multivariate sample following the Sklar theorem between marginals on the diagonal and copula on the upper triangle.

2.3.2 Implementation in a Python package

An open source implementation is proposed in the python package `copulogram`. This code mostly relies on the Python package for data visualization `seaborn` (Waskom, 2021). The developments are tracked and archived in a GitHub repository[2] and the package can be installed from the package-management system "PyPI".

Multiple visual options are offered by the `copulogram` package, as illustrated in the GitHub repository. For example, the user can represent the univariate samples on the diagonal or the bivariate samples in the triangles with kernel density estimation. Categorical variables can be used to assign different colors depending on the data class. The colors can also vary depending on a continuous variable after defining a mapping between the values of this variable and a set of colors (also called colorbar).

Example #1: Iris flower dataset

The first example illustrates the copulogram on a widely used dataset in the machine learning community. The iris flower dataset was first introduced by Fisher and became a reference dataset for classification techniques. In the following lines of Python code, the dataset is loaded and the copulogram package is used to draw the new plot. The resulting copulogram applied to the iris flower data is represented in Fig. 2.3.

```
#!/usr/bin/python3
import seaborn as sns
import copulogram as cp
data = sns.load_dataset("iris")
copulogram = cp.Copulogram(data)
copulogram.draw(hue="species")
```

Since this data mostly presents linear dependencies, the copulogram is not very instructive. In other cases, the role of the dependence in the joint distribution is more important.

Example #2: Ishigami function

The Ishigami function is commonly used as a benchmark problem for global sensitivity analysis (GSA):

$$y = g(x_1, x_2, x_3) = \sin(x_1) + 7\,\sin(x_2)^2 + \frac{x_3^4\,\sin(x_1)}{10}. \tag{2.19}$$

This uncertainty quantification problem considers an independent random input vector $\mathbf{X} = \prod_{j=1}^{3} X_j$. While the marginals in GSA benchmarks are usually assumed to be uniform, they will be considered Gaussian hereafter to distinguish the different elements of the joint distribution. Therefore, let us define $X_j \sim \mathcal{N}(0,1) \forall j \in \{1,2,3\}$. In this setup, the random inputs are independent but they each present interesting dependencies with the random output $Y = g(\mathbf{X})$.

A Monte Carlo sample with size $n = 10^3$ is generated, $\mathbf{X}_n \overset{\text{i.i.d.}}{\sim} \mathbf{X}$, and evaluated such that $Y_n = g(\mathbf{x}_n)$. The copulogram of the input-output sample $(\mathbf{X}_n, Y_n)$ is represented in Fig. 2.4. As expected, the scatter plots between the inputs in the upper triangle are uniform (representing an independent density copula).

In GSA, the work of Póczos et al. (2012) studied the discrepancy between the empirical density copula $c_n(X_j, Y), j \in \{1, \dots, d\}$ and the independent copula to qualitatively assess the importance of X_j. In the same vein, the paper of Plischke and Borgonovo (2019) attempted to formalize a link between different GSA approaches based on copulas and to quantitative approaches as the Sobol' indices.

[2] GitHub repository: https://github.com/efekhari27/copulogram

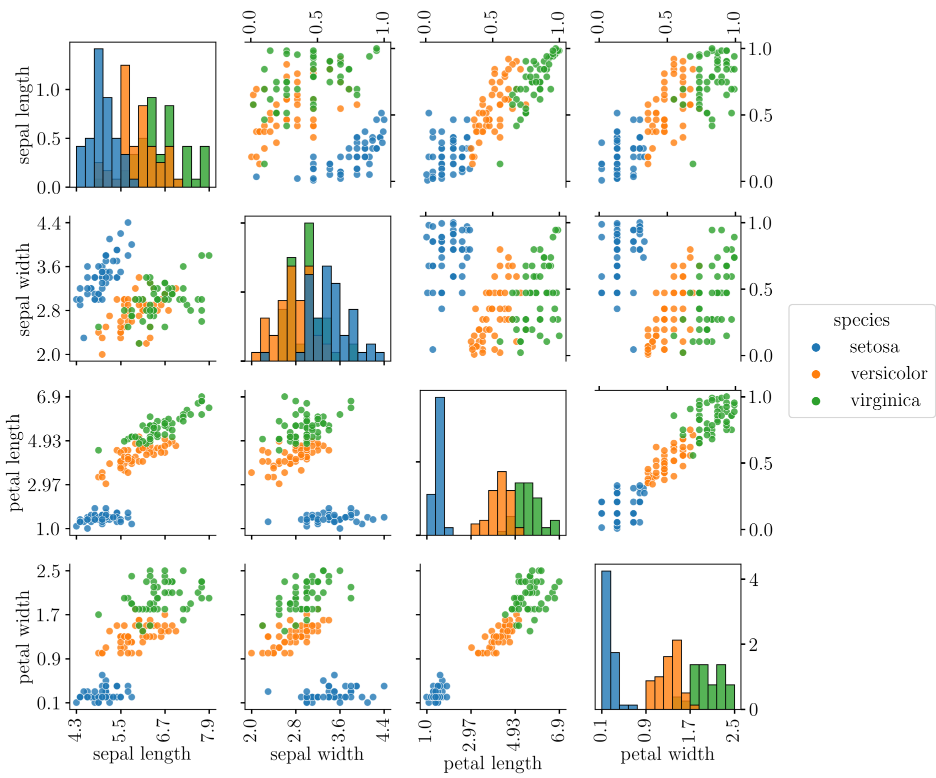


Figure 2.3 Copulogram of the iris flower dataset with colors assigned by the iris species.

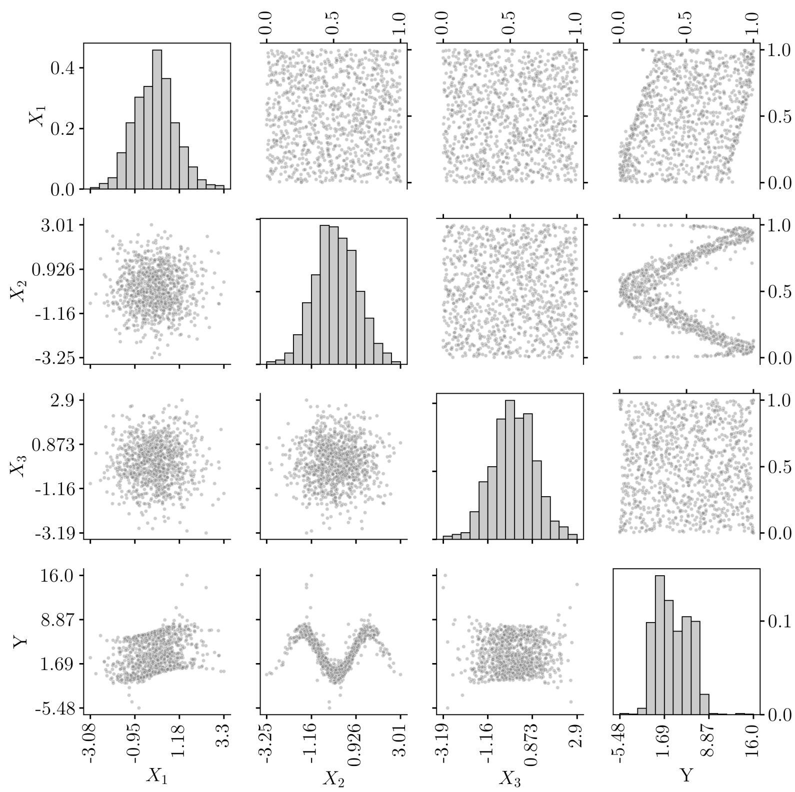


Figure 2.4 Copulogram of Monte Carlo sample (with size $n = 10^3$) of the inputs and outputs of the modified Ishigami problem.

2.4 Semiparametric inference of the South Brittany metocean conditions

Metocean conditions have been long studied in coastal and offshore engineering. Inferring multivariate probabilistic models on metocean data became essential in wind energy.

Numerous approaches are proposed in the literature to fit a model on environmental data. Among them, let us mention the use of parametric methods as the conditional modeling (e.g., Bitner-Gregersen, 2015; Vanem et al., 2023), or the construction of vine copulas (e.g., Vanem, 2016; Montes-Iturrizaga and Heredia-Zavoni, 2016; Lin and Dong, 2019). Nonparametric methods as the KDE were also applied in this context (e.g., Han et al., 2018). The nonparametric techniques generally struggle to model the distributions' tails, even if the tails are essential to qualify structures for ultimate events. However, they are highly flexible and often easier to implement than parametric methods.

In this section, a semiparametric inference strategy is presented, composing some well-known parametric models for the marginals (e.g., Weibull distribution for the wind speed), with a highly flexible dependence modeling by the EBC. A metocean dataset is used to showcase the empirical Bernstein copula and its representation by the copulogram. This dataset from the ANEMOC (Digital Atlas of Ocean and Coastal Sea States atlas, Raoult et al., 2018) gathers 32 years of preprocessed data (at an hourly resolution) from a location off the coast of South Brittany, France. A subset of 10^4 points is randomly selected among the ANEMOC data, which will be used to realize the semiparametric inference. The full code developed for this inference study is available on a GitHub repository[3].

2.4.1 Inference of the marginals

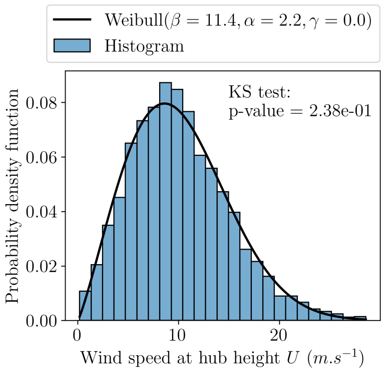
The variables studied to describe the environmental conditions match the ones defined in Table **??**. Unfortunately, the turbulence is provided by the ANEMOC database and is therefore not fitted. A straightforward inference is performed on the data, resulting in the models presented in Table 2.1. The wind and wave directions are fitted by KDE to catch their multimodal behavior while the other variables by MLE on various parametric models. Note that some variations of KDE with kernels specific to circular data could be interesting to ensure the continuity of the model at the bounds (Bai et al., 1989).

The results of the marginals' inferences, plotted in Fig. 2.5 against histograms, are visually satisfying. Statistical testing is not necessary in our case since the actual topic of discussion is related to the inference of the dependence. Considering these marginals, a study of the copula inference can be developed.

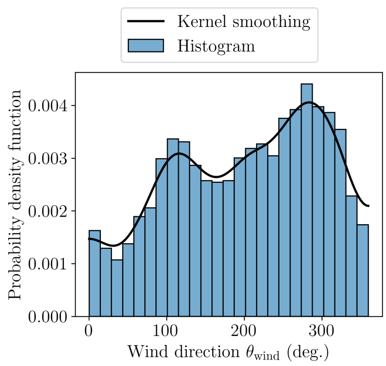
[3]GitHub repository: https://github.com/efekhari27/thesis/blob/main/numerical_experiments/chapter3/south_brittany_inference.ipynb

Name	*Notation*	*Fitted model*	*KS p-value (α= 5%)*
Wind speed	U	Weibull ($\beta = 11.4, \alpha = 2.2, \gamma = 0$)	0.238
Wind direction	θ_{wind}	KDE	–
Significant wave height	H_s	Inverse Normal ($\mu = 2.3, \lambda = 6.8$)	0.533
Wave period	T_p	Weibull ($\beta = 9.3, \alpha = 3.3, \gamma = 2$)	0.00021
Wave direction	θ_{wave}	KDE	–

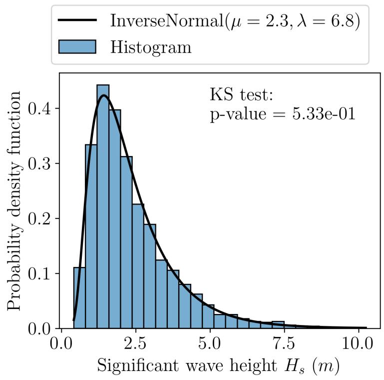
Table 2.1 Marginal inference results of the South Brittany metocean data.



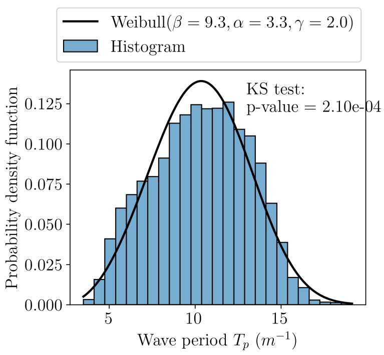
(a) Wind speed U.



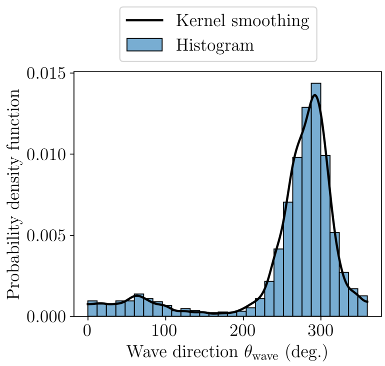
(b) Wind direction θ_{wind}.



(c) Significant wave height H_s.



(d) Wave period T_p.



(e) Wave direction θ_{wave}.

Figure 2.5 Marginal inference results of the South Brittany metocean data.

2.4.2 Nonparametric inference of the dependence

The aim of this section is to complete the set of marginals fitted previously by the inference of a copula. A nonparametric estimation using the empirical Bernstein copula is studied in this section. To validate the goodness-of-fit of the EBC, the ANEMOC dataset (with size $N = 10^4$) is randomly split into two: a learning set $\mathbf{X}_n$ and a validation set $\mathbf{X}'_n$. A joint distribution $\widehat{F}_{\mathbf{X}}$ is then build using $\mathbf{X}_n$, by combining the marginals fitted earlier with an empirical copula $B_{\mathbf{m}}(C_n)$, such that:

$$\widehat{F}_{\mathbf{X}}(\mathbf{x}) = B_{\mathbf{m}}(C_n)\left(\widehat{F}_{X_1}(x_1), \dots, \widehat{F}_{X_d}(x_d)\right). \tag{2.20}$$

Where $\widehat{F_{X_j}}$ stand for the model of the marginal j, just inferred in Subsection 2.4.1. While C_n is the empirical copula associated with the sample $\mathbf{X}_n$, and all the polynomial orders of the EBC are equal, $\mathbf{m} = \{m, \dots, m\}, m \in \mathbb{N}$.

Then, one could compare a sample $\widehat{\mathbf{X}_n}$, generated from the fitted joint distribution $\widehat{F}_{\mathbf{X}}$, with the learning set $\mathbf{X}_n$. However, to prevent an overfit from the semiparametric model, the comparison is rather done between $\mathbf{X}_n$ and the independent validation set $\mathbf{X}'_n$. The statistic used is the maximum mean discrepancy, MMD$(\widehat{\mathbf{X}_n}, \mathbf{X}'_n)$, initially introduced for multivariate two-sample testing (a specific presentation of the MMD and its estimation is developed in Appendix B). For a given fitted joint distribution, this procedure is repeated 100 times to take into account the sampling variability.

In Fig. 2.6, MMD distributions are represented for different values of the EBC polynomial order, with $m \in \{5, 10, 20, 50, 100, 1000\}$. The smaller the values of this dissimilarity measure, the closer the samples should be. Even if further developments could be implemented to improve the MMD estimation, these results are sufficient to set the EBC tuning at $m = 100$. Considering this setup, Fig. 2.7 represents the copulogram of a sample $\widehat{\mathbf{X}_n}$ (in red), side by side with the copulogram of the learning set $\mathbf{X}_n$ (in blue). This semiparametric approach offers a lot of flexibility, which is essential when inferring such complete dependence structures. As with any nonparametric method, it should be used with caution when inferring distributions' tails.

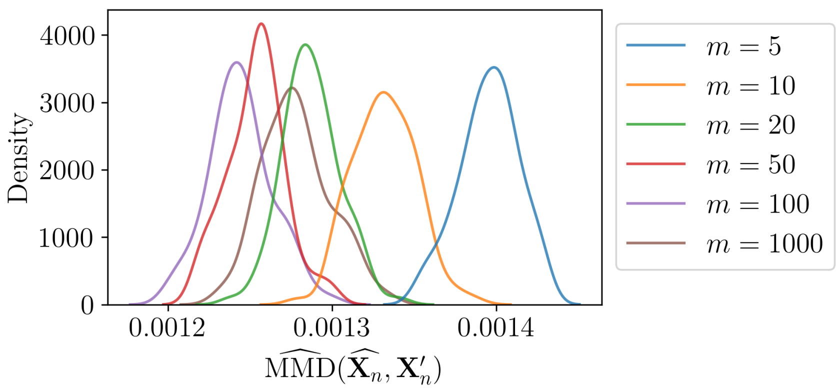
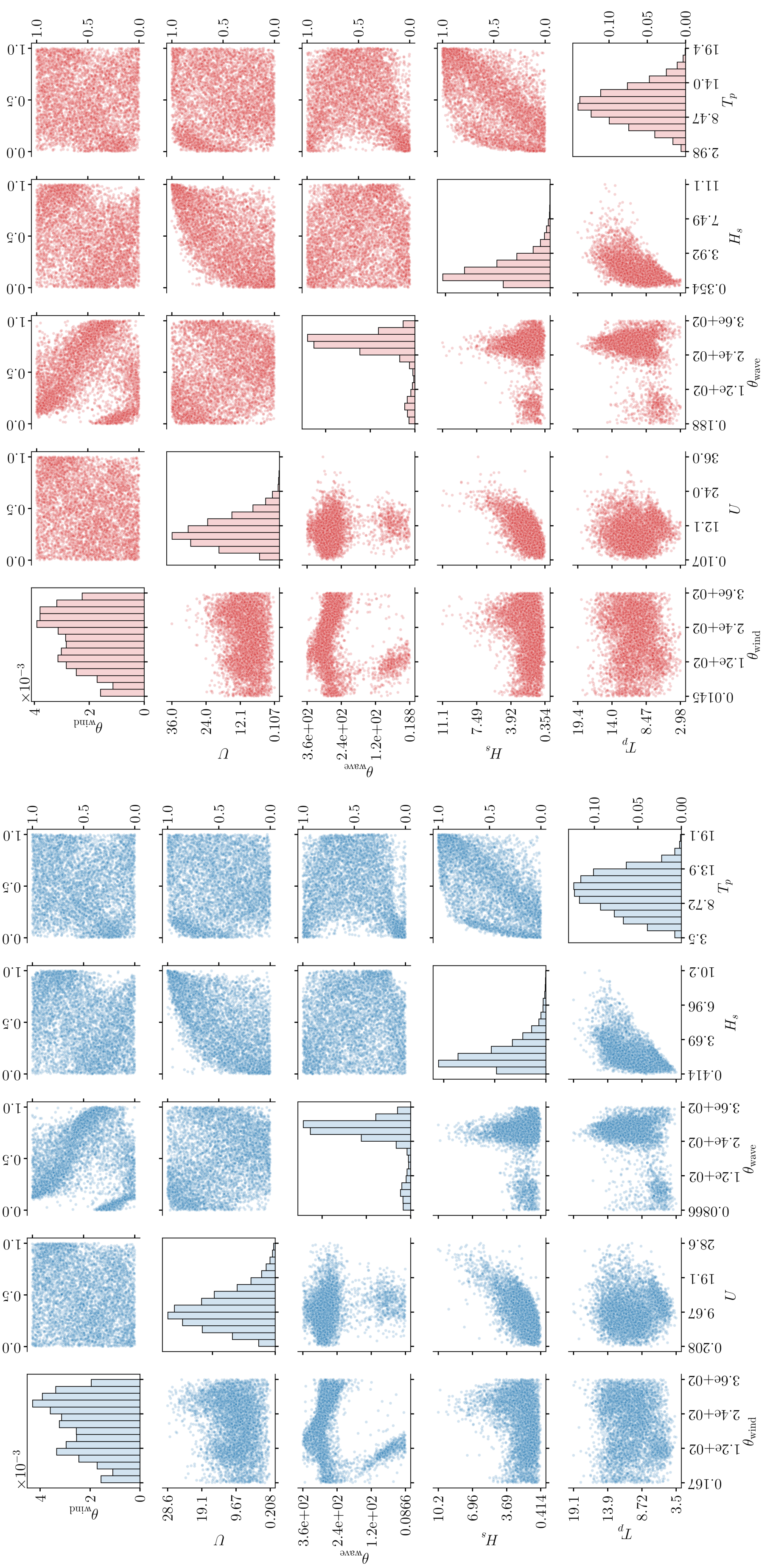


Figure 2.6 Empirical distributions of the maximum mean discrepancy between the validation sample $\mathbf{X}'$ and the sample $\widehat{\mathbf{X}_n} \overset{\text{i.i.d.}}{\sim} \widehat{F}_{\mathbf{X}}$ (repeated for 100 samples $\widehat{\mathbf{X}_n}$).



(b) Monte Carlo sample (size $n = 5000$), generated from the semiparametric model fitted on the ANEMOC data (EBC $B_{\mathbf{m}}(C_n)$ with $\{m_j = 100\}_{j=1}^{d}$).

(a) South Brittany ANEMOC data (size $n = 5000$).

Figure 2.7 Copulogram of the South Brittany metocean data.

2.4.3 Summary and discussion

In this section, a semiparametric inference strategy was illustrated on a metocean dataset from a site off the coast of South Brittany, France. This approach is pragmatic and offers a lot of flexibility. In our case, the unimodal marginals are fitted by MLE while the multimodal ones are fitted by KDE. Considering the complexity of the dependence, the EBC showed interesting results for large-size samples. Its capacity to extrapolate in the tails could be further studied (Heredia-Zavoni and Montes-Iturrizaga, 2022) but this tool is an appropriate solution for general inference (for example needed for fatigue assessment).

In the monograph of Joe and Kurowicka (2011), the use of nonparametric methods is briefly discussed p.250. The author recommends using nonparametric copulas when the marginals are well-behaved, but the dependence structure is nonlinear.

2.5 Quantifying and clustering the wake-induced perturbations within a wind farm

After defining a probabilistic model based on ambient metocean data, the present section studies the impact of the wake on the wind conditions within a farm. The wake arises from extracting kinetic energy from the wind, leading to a decrease in wind speed and an increase in turbulence downstream of the turbines. In a wind farm, the wake mostly depends on the turbines' layout, the ambient wind speed, and the ambient turbulence intensity. The resulting heterogeneous wind field in a wind farm can be simulated by numerical models with different fidelities (as discussed in Subsection **??**).

In our case, simplified wake models (sometimes called "dynamic wake meandering", or "engineering" models, see e.g., Doubrawa et al., 2020) are used to simulate the wind speed deficit and the added turbulence at each turbine. To recover a wake-perturbed wind distribution at each turbine, the ambient wind distribution is propagated through a wake model. Having different wind distributions naturally impacts the loading and should be considered during fatigue assessment at a farm scale. Such heterogeneity is mostly considered by international standards via empirical coefficients (also called "effective turbulence"). Doubrawa et al. (2023) compares the effective turbulence approach with dynamic wake meandering models and studies their impact on loading.

In practice, fatigue assessment on a turbine represents a computational effort. As a consequence of wake modeling, each turbine presents a different wind distribution, which implies repeating a fatigue assessment for every turbine. To make this computation tractable at a farm scale, the present section aims at building clusters of turbines similarly affected by the wake. Then, fatigue assessment can be computed on a few turbines, each representing a cluster of turbines facing similar wake-modified wind loading. The maximum mean discrepancy (MMD) is used as a statistical dissimilarity measure to compare the perturbed distributions induced by the wake.

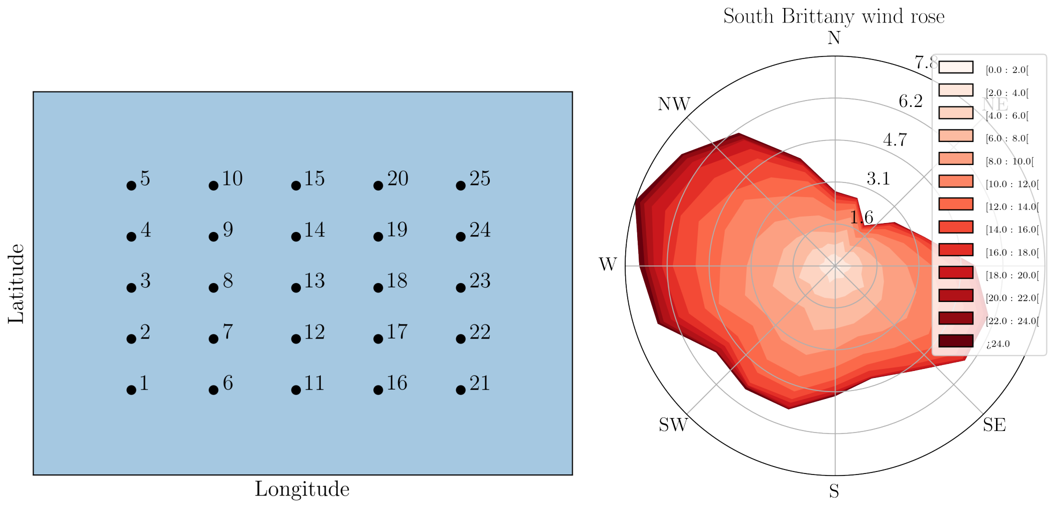


Figure 2.8 South Brittany wind farm layout, the vertical direction does not represent the exact north (left). South Brittany wind rose from the ANEMOC data (right).

This new approach is applied to a theoretical wind farm representing the recent call for tenders off the coast of South Brittany, France. The turbines considered are a modified version of the floating offshore wind turbine (FOWT) IEA-15MW (described in Kim et al., 2022). Figure 2.8 illustrates the layout of the 25 FOWTs modeled in the following, the coordinates are normalized by the rotor's diameter D. The spacing between turbines in this regular layout is equivalent to seven rotor diameters in the dominant wind direction and five rotor diameters in the orthogonal direction (i.e., crosswind).

To develop this approach on the South Brittany farm, the present section is structured as follows: first, the wake model and its corresponding uncertainty propagation are presented, then MMDs are estimated between the resulting empirical joint wind distributions, and finally, a simple clustering gathers turbines perceiving similar wakes and defines their representative turbines. In the end, the 25 FOWT are split into four groups, whose representatives can be used for the fatigue assessment of the whole group.

2.5.1 Uncertainty propagation on a wake model

Wake model definition When simulating the wake effect of floating wind turbines, different studies (either using dynamic wake meandering Wise and Bachynski (2020) or LES Johlas et al. (2020)), showed the importance of modeling the floaters' position (translation and rotation). Therefore, an engineering wake model based on the Farmshadow™ software (developed by IFPEN) was coupled with a hydro-static calculation to predict the floater's position, as well as the wind speed and turbulence intensity. In this model, the floaters are considered to be rigid, and all degrees of freedom are considered (surge, sway and heave for the three translations and roll, pitch and yaw for the three rotations).

FarmShadow™ uses engineering wake models to simulate the wind field throughout the whole farm, starting from the most upstream WT and working downwards. More precisely, the model used includes the "super-gaussian" approach for speed deficit (Blondel and Cathelain, 2020), waked-induced turbulence according to Ishihara and Qian (2018), and superimposition following the linear sum approach defined by Zong and Porté-Agel (2020). Further assumptions regarding the hydro-statics loading, the effect of the mooring lines, and the wake model are defined in Lovera et al. (2023).

Monte Carlo uncertainty propagation The wake model described earlier takes as input a set of variables describing the ambient wind conditions $\mathbf{x} \in \mathbb{R}^3$ and computes the perturbed wind conditions at each WT represented by the vector $\mathbf{x}'_l, l \in (1, .., n_{WT})$, where $n_{WT} \in \mathbb{N}$ is the total number of turbines in the farm:

$$g : \mathbb{R}^3 \longrightarrow \mathbb{R}^{3n_{WT}} \tag{2.21}$$

$$\mathbf{x} \longmapsto g(\mathbf{x}) = (\mathbf{x}'_1, .., \mathbf{x}'_{n_{WT}}) \tag{2.22}$$

The uncertainties associated with the ambient wind conditions are represented by a random vector $\mathbf{X}$ following the distribution f_0. A parametric model has been fitted in Vanem et al. (2023) using conditional probability density functions to capture the dependence structure, but the semi-parametric inference proposed in Section 2.4 could have been an alternative. Note that the missing turbulence intensity in the ANEMOC data from South Brittany was assumed to follow a lognormal distribution. The random vector $\mathbf{X}$ gathers the following random variables:

- Mean wind speed (U) is the 10-min average horizontal wind speed at hub height.
- Wind turbulence intensity (TI) is the 10-min wind speed turbulence intensity at hub height.
- Wind direction (θ_{wind}) is the 10-min average wind direction.

In the following, the wind orientation θ_{wind} is supposed to be unaffected by the wake. The uncertainty propagation through the wake models provides a set of perturbed environmental distributions $f'_l, l \in (1, .., n_{WT})$. In practice, a Monte Carlo sample $\mathbf{X}_n = \mathbf{x}^{(1)}, .., \mathbf{x}^{(n)} \sim \mathbf{X}$ (with size n=6000) is generated and evaluated by the wake model. Since the model has a low computational cost, Monte Carlo sampling was affordable while giving strong convergence guarantees.

Fig. 2.9 illustrates the perturbation of the wind distributions for three WT differently affected by the wake depending on their position in the farm (see Figure 2.8). One can notice that the distribution of WT 25 (in orange) is very close to the ambient distribution (in black), as expected since this WT is located on the edge of the farm and facing the dominant wind direction. Meanwhile, the distribution of WT 13 (in red) seems more affected by the wake, by getting higher wind turbulence with lower wind speed. This analysis can be completed with the two marginals in Fig. 2.10a and Fig. 2.10b, both describing the ambient marginal distributions (in black) and wake-disturbed distributions. In general, a small wind speed deficit is indicated by the small shifts of the probability density functions to the left on Fig. 2.10a. Also, a small added

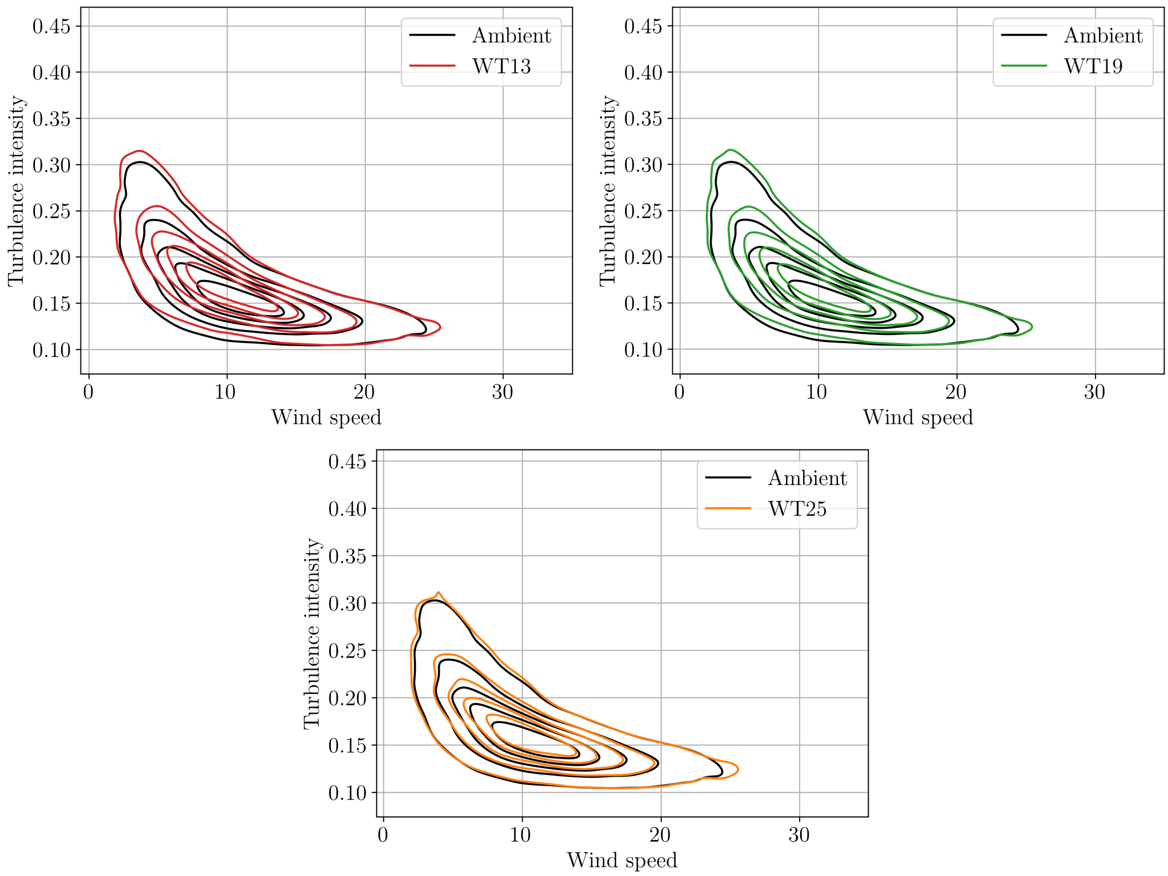


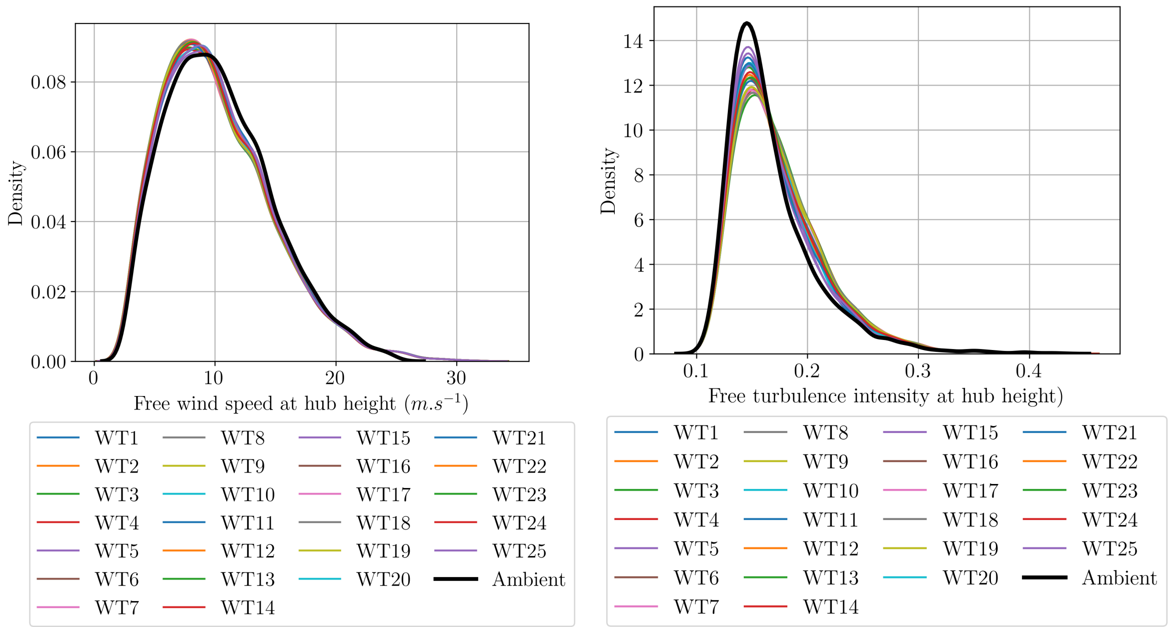
Figure 2.9 Joint distributions of the wake-perturbed wind conditions at WT 13, 19, and 25 (in color) compared with the ambient wind conditions (in black).

turbulence is noticeable with shifts of the PDFs to the right on Fig. 2.10b. A tool is needed to quantify the wind perturbations induced by the wake.

2.5.2 Statistical metric of wake-induced perturbations

The maximum mean discrepancy was introduced by Gretton et al. (2006) as a statistic for two-sample testing. After further work on this tool, authors such as Sriperumbudur et al. (2010) showed that the MMD is a distance between two distributions embedded in a specific function space. The MMD becomes a metric for specific kernels called “characteristic kernels”, which offer the following property: $\mathrm{MMD}(\pi, \zeta) = 0 \iff \pi = \zeta$. The squared MMD has been used for multiple purposes and is further presented in Appendix B In the following, the idea is to compare the ambient wind distribution f_0 to the wake-perturbed wind conditions f_l' at the WT l using the squared MMD.

Application to the South Brittany wind farm project Once the joint perturbed distributions of each WT are evaluated on a large Monte Carlo sample, their squared MMD with the ambient wind conditions can be computed. Fig. 2.11 represents the squared MMD for each WT to quantify the wake-induced perturbation. The values of squared MMD presented in this



(a) Wind speed. (b) Turbulence intensity.

Figure 2.10 Ambient (in black) and wake-perturbed (in color) distributions of wind distributions.

figure are estimated between two Monte Carlo samples with size $n = 6000$. A verification of their convergence is realized in terms of coefficient of variation.

2.5.3 Clustering the wake-induced perturbations

The aim of this section is to use the MMD as a metric to define clusters of turbines getting similar wind conditions. Instead of comparing the wake-perturbed distributions with the ambient one, let us compare all the pairs of wake-perturbed distributions. Considering all the WT $\{1, \dots, n_{WT}\}$ in the farm, and their respective wake-perturbed distributions $\{f_l'\}_{l=1}^{n_{WT}}$, let us

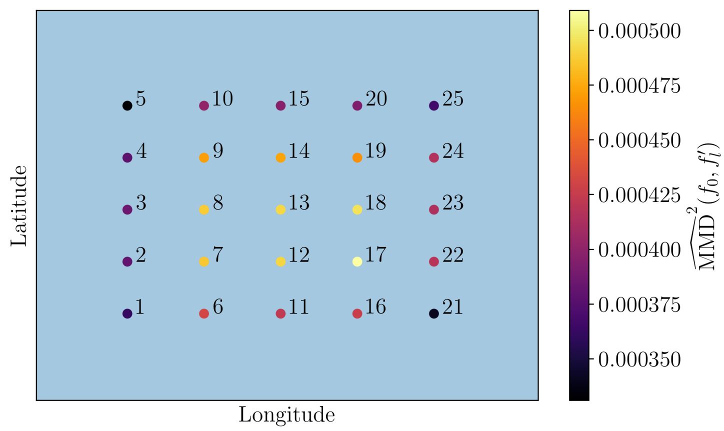


Figure 2.11 South Brittany layout and wake effects measured by the squared MMD on wind conditions. Note that the vertical direction does not represent the north direction.

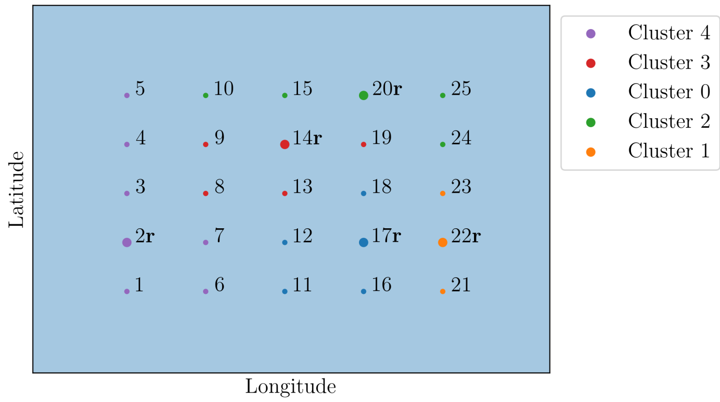


Figure 2.12 K-medoids clustering solution for five clusters. The representative elements of the clusters are tagged with the mention “r”.

define the symmetric matrix D of MMD between every pair of perturbed distributions such that $D_{i,j} = \widehat{\mathrm{MMD}}^2(f_i', f_j')$.

Different unsupervised clustering techniques, such as hierarchical or centroid-based methods were compared in Lovera et al. (2023). The “k-medoids” (Park and Jun, 2009) are a variation of the well-known k-means that selects actual data points as centers (i.e., medoids). In our case, this method not only gathers turbines with the same wind conditions but also determines a representative turbine for each cluster.

Assuming a final number of clusters equal to five, Fig. 2.12 represents the clusters defined by the k-medoid method applied to the matrix D. Let us notice that the results are rather coherent with the main wind orientation illustrated in the wind rose (see Fig. 2.8). Interestingly, the conclusion that emerged from comparing the wake-perturbed distributions to the ambient one (see Fig. 2.11) is different from the ones obtained by comparing pairs of wake-perturbed distributions. Finally, the representative turbines are tagged with the mention “r” and could be used to perform a fatigue analysis.

2.5.4 Summary and discussion

This section studied the impact of the wake on the wind conditions in a farm. The ambient wind conditions were propagated on a low-fidelity wake model of a floating wind farm. Even if higher-fidelity models could simulate this phenomenon more accurately, their huge computational cost would not allow uncertainty propagation (e.g., several days for LES certain solver, with intensive parallelization). The present model has a reasonable error and has a short execution (i.e., about a few minutes on a regular computer). The resulting wake-perturbed distributions at each turbine were compared to each other using a kernel-based dissimilarity measure for distributions. Using this scalar metric of perturbations created by the wake, clusters were built. In the context of fatigue assessment at the farm scale, the computed clusters may reduce the computational effort by assessing one turbine per cluster.

2.6 Conclusion

In this chapter, different aspects of uncertainty quantification were applied to define the metocean conditions in a wind farm. First, to infer a joint environmental distribution, a semiparametric approach was applied to a large dataset from a site in South Brittany, France. This semiparametric approach combines parametric models of the marginals with nonparametric models of the copula such as the highly flexible empirical Bernstein copula. Second, to take into account the heterogeneous wind conditions inside a farm, a dynamic wake meandering model was developed for the South Brittany farm. Then, the ambient joint environmental distribution was propagated through this model to recover a wake-perturbed environmental distribution per turbine.

The copulogram is a new data visualization tool based on the empirical copula, aiming at improving the description of nonlinear dependencies in datasets.

The maximum mean discrepancy is a kernel-based dissimilarity measure between multivariate distributions which was used throughout this chapter. Either for testing the goodness-of-fit of the semiparametric model fitted using the EBC, or as a measure of the perturbation occasioned by the wake on the wind conditions. In this context, the MMD allowed building clusters of turbines witnessing the same wind conditions. As perspectives, the robustness of this metric to the choice of kernel could be further investigated as well as the goodness-of-fit of the EBC over distributions' tails. After defining the probabilistic model of the environmental conditions, this uncertainty is ready to be propagated in a multiphysics wind turbine model as DIEGO.

Chapter 3

Kernel-based central tendency estimation

Parts of this chapter are adapted from the following publication:

> E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2024). "Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature". In: *Data-Centric Engineering*, In press.

3.1 Introduction

As a sustainable and renewable energy source, offshore wind turbines (OWT) are likely to take a growing share of the global electric mix. However, to be more cost-effective, wind farm projects tend to move further from the coast, exploiting stronger and steadier wind resources. Going further offshore, wind turbines are subject to more severe and uncertain environmental conditions (i.e., wind and waves). In such conditions, their structural integrity should be certified. To do so, numerical simulation and probabilistic tools have to be used. In fact, according to Graf et al. (2016), for new environmental conditions or new turbine models, international standards such as IEC-61400-1 (2019) from the International Electrotechnical Commission and DNV-ST-0437 (2016) from Det Norske Veritas recommend performing over 2×10^5 simulations distributed over a grid. However, numerical simulations are computed by a costly hydro-servo-aero-elastic wind turbine model, making the design process time-consuming. In the following, the simulated output cyclic loads studied are aggregated over the simulation period to assess the mechanical fatigue damage at hot spots of the structure. To compute the risks associated with wind turbines throughout their lifespan, one can follow the steps of the universal framework for the treatment of uncertainties presented in the introduction of this manuscript Fig. 1. After specifying the problem (Step A), one can quantify the uncertainties related to site-specific environmental conditions represented by the random vector $\mathbf{X} \in \mathcal{D}_{\mathbf{X}} \subset \mathbb{R}^d, d \in \mathbb{N}^*$ (Step B). Then, one can propagate them through the OWT simulation model (Step C) denoted by $g : \mathcal{D}_{\mathbf{X}} \rightarrow \mathbb{R}, \mathbf{X} \mapsto Y = g(\mathbf{X})$, and estimate a relevant quantity of interest $\psi(Y) = \psi(g(\mathbf{X}))$ (e.g., a mean, a quantile, a failure probability). An accurate estimation of the quantity of interest $\psi(Y)$ relies on both a relevant quantification of the input uncertainty and an efficient sampling method.

Regarding Step B, when dealing with uncertain environmental conditions, a specific difficulty often arises from the complex dependence structure such variables may exhibit. Here, two cases may occur: either measured data are directly available (i.e., the "given-data" context) or a theoretical parametric form for the joint input probability distribution can be postulated. Such existing parametric joint distributions often rely on prior data fitting combined with expert knowledge. For example, several parametric approaches have been proposed in the literature to derive such formulations, ranging from fitting conditional distributions Vanem et al., 2023) to using vine copulas (Li and Zhang, 2020). When a considerable amount of environmental data is available, nonparametric approaches such as the empirical Bernstein copula were studied in Chapter 2 to capture complex dependence structures. Alternatively, an idea is to directly use the data as an empirical representation of input uncertainties in order to avoid an additional inference error.

Step C usually focuses on propagating the input uncertainties in order to estimate the quantity of interest. Depending on the nature of $\psi(Y)$, one often distinguishes between two types of uncertainty propagation: a central tendency estimation (e.g., focusing on the output mean value or the variance) and a tail estimation (e.g., focusing on a high-order quantile or a failure probability). When uncertainty propagation aims at central tendency estimation, the usual methods can be split into two groups. First, those relying on sampling, i.e., mainly Monte Carlo sampling (Graf et al., 2016), quasi-Monte Carlo sampling (Müller and Cheng, 2018), geometrical subsampling (Kanner et al., 2018), or deterministic quadrature rules (Van den Bos, 2020). All these methods estimate the quantity directly on the numerical simulator's outputs. Second, those that rely on the use of surrogate models (or metamodels, see Fig. 1) to emulate the costly numerical model by a statistical model. Among a large panel of surrogates, one can mention, regarding wind energy applications, the use of polynomial chaos expansions (Dimitrov et al., 2018; Murcia et al., 2018), Gaussian process regression (Huchet, 2019; Teixeira et al., 2019a; Slot et al., 2020; Wilkie and Galasso, 2021), or artificial neural networks (Bai et al., 2023). When uncertainty propagation aims at studying the tail of the output distribution such as in risk or reliability assessment, one usually desires to estimate a quantile or a failure probability. In the wind energy literature, failure probability estimation has been largely studied, e.g., in time-independent reliability assessment (Zwick and Muskulus, 2015; Slot et al., 2020; Wilkie and Galasso, 2021) or regarding time-dependent problems (Abdallah et al., 2019; Lataniotis, 2019).

During the overall process described in Fig. 1, modelers and analysts often need to determine whether inputs are influential or not in order to prioritize their effort (in terms of experimental data collecting, simulation budget, or expert elicitation). Sometimes, they want to get a better understanding of the OWT numerical models' behavior or to enhance the input uncertainty modeling. All these questions are intimately related to the topic of sensitivity analysis (Saltelli et al., 2008; Da Veiga et al., 2021) and can be seen as an "inverse analysis" denoted by Step C' in Fig. 1. In the wind energy literature, one can mention, among others, some works related to Spearman's rank correlation analysis and the use of the Morris method in Velarde et al. (2019); Petrovska (2022). Going to variance-based analysis, the direct calculation of Sobol' indices after fitting a polynomial chaos surrogate model has been proposed in many works (e.g., in Murcia et al., 2018) while the use of distributional indices (e.g., based on the Kullback–Leibler divergence) has been investigated by Teixeira et al. (2019b).

The present chapter focuses on the problem of uncertainty propagation, and more specifically, on the mean fatigue damage estimation (i.e., $\psi(Y) = \mathbb{E}[g(\mathbf{X})]$). Such a problem is usually encountered, by engineers, during the design phase. Most of the time, current standards as well as common engineering practices make them use regular grids (Huchet, 2019). Altogether, one can describe three alternative strategies: (i) direct sampling on the numerical model (e.g., using Monte Carlo), (ii) sampling on a static surrogate model (e.g., using Gaussian process regression), or (iii) using an "active learning" strategy (i.e., progressively adding evaluations of the numerical

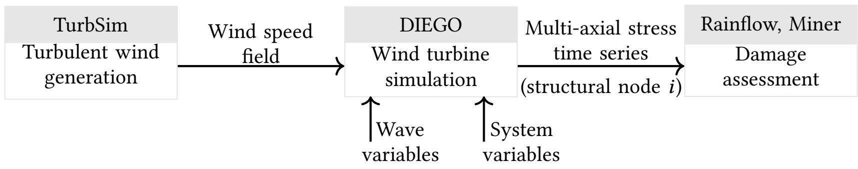


Figure 3.1 Diagram of the chained OWT simulation model.

model to enhance the surrogate model fitting process). In practice, fitting a surrogate model in the context of OWT fatigue damage can be challenging due to the nonlinearity of the code. Moreover, the surrogate model validation procedure complexifies the process. Finally, active learning strategies restrict the potential number of parallel simulations, which limits the use of HPC facilities. Thus, the main contribution of this chapter is to explore different ways to propagate uncertainties by directly evaluating the numerical model (i.e., without any surrogate model) with a relevant tradeoff between computational cost and accuracy. In the specific context of wind turbine fatigue damage, this work shows how to propagate uncertainties arising from a complex input distribution through a costly wind turbine simulator. The proposed work consists of evaluating the advantages and limits of kernel herding as a tool for given-data, fast, and fully-distributable uncertainty propagation in OWT simulators. Additionally, this sampling method is highly flexible, allowing one to complete an existing design of experiments. Such a property can be crucial in practice when the analyst is asked to include some specific points to the design (e.g., characteristic points describing the system's behavior required by experts or by standards, see Huchet, 2019).

The present chapter is organized as follows. Section 3.2 will present the industrial use case related to a wind farm operating in Teesside, UK. Then, Section 3.3 will introduce various kernel-based methods for central tendency estimation. Section 3.4 will analyze the results of numerical experiments obtained on both analytical and industrial cases. Finally, the last section will present some discussions and draw some conclusions.

3.2 Treatment of uncertainties on the Teesside wind farm

An OWT is a complex system interacting with its environment. To simulate the response of this system against a set of environmental solicitations, multi-physics numerical models are developed. In the present chapter, the considered use case consists of a chain of three numerical codes executed sequentially. As illustrated in Fig. 3.1, a simulation over a time period is the sequence of, first, a turbulent wind speed field generation, then a wind turbine simulation (computing various outputs including mechanical stress), and finally, a post-processing phase to assess the fatigue damage of the structure.

3.2.1 Numerical simulation model

This subsection generally describes the modeling hypotheses considered in the industrial use case, further details regarding wind turbines modeling a provided in Chapter **??** of this manuscript. The first block of the chain consists of a turbulent wind field simulator called "TurbSim" (developed by Jonkman, 2009 from the National Renewable Energy Laboratory, USA) that uses, as a turbulence model, a Kaimal spectrum (Kaimal et al., 1972) (as recommended by the IEC-61400-1, 2019). Moreover, to extrapolate the wind speed vertically, the shear is modeled by a power law. Since the wind field generation shows inherent stochasticity, each 10-minute long simulation is repeated with different pseudo-random seeds and one averages the estimated damage over these repetitions. This question has been widely studied by some authors, (e.g., Slot et al., 2020), who concluded that the six repetitions recommended by the IEC-61400-1 (2019) may be insufficient to properly average this stochasticity. Thus, in the following, the simulations are repeated eleven times (picking an odd number also directly provides the median value over the repetitions). This number of repetitions was chosen to suit the maximum number of simulations and the storage capacity of the generated simulations.

As a second block, one finds the "DIEGO" software (for "Dynamique Intégrée des Éoliennes et Génératrices Offshore"[1]) which is developed by EDF R&D (Kim et al., 2022) to simulate the aero-hydro-servo-elastic behavior of OWTs. It takes the turbulent wind speed field generated by TurbSim as input and computes the dynamical behavior of the system (including the multiaxial mechanical stress at different nodes of the structure). For the application of interest here, the control system is modeled by the open source DTU controller (Hansen and Henriksen, 2013), and no misalignment between the wind and the OWT is assumed. As for the waves, they are modeled in DIEGO using a JONSWAP spectrum (named after the 1975 Joint North Sea Wave Project). The considered use case here consists of a DIEGO model of a Siemens SWT 2.3MW bottom-fixed turbine on a monopile foundation (see the datasheet in Table 3.1), currently operating in Teesside, UK (see the wind farm layout and wind turbine diagram in Fig. 3.2). Although wind farms are subject to the wake effect, affecting the behavior and performance of some turbines in the farm, this phenomenon is not considered in this chapter. To avoid numerical perturbations and reach the stability of the dynamical system, our simulation period is extended to 1000 seconds and the first 400 seconds are cropped in the post-processing step. This chained OWT numerical simulation model has been deployed on an EDF R&D HPC facility to benefit from parallel computing speed up (a single simulation on one CPU takes around 20 minutes).

3.2.2 Measured environmental data

During the lifespan of a wind farm project, environmental data is collected at different phases. In order to decide on the construction of a wind farm, meteorological masts, and wave buoys are usually installed on a potential site for a few years. After its construction, each wind turbine

[1] In English, "Integrated Dynamics of Wind Turbines and Offshore Generators".

Table 3.1 Teesside Offshore Wind turbine datasheet

Siemens SWT-2.3–93	
Rated power	2.3 MW
Rotor diameter	93 m
Hub height	83 m
Cut-in, cut-out wind speed	4 m/s, 25 m/s

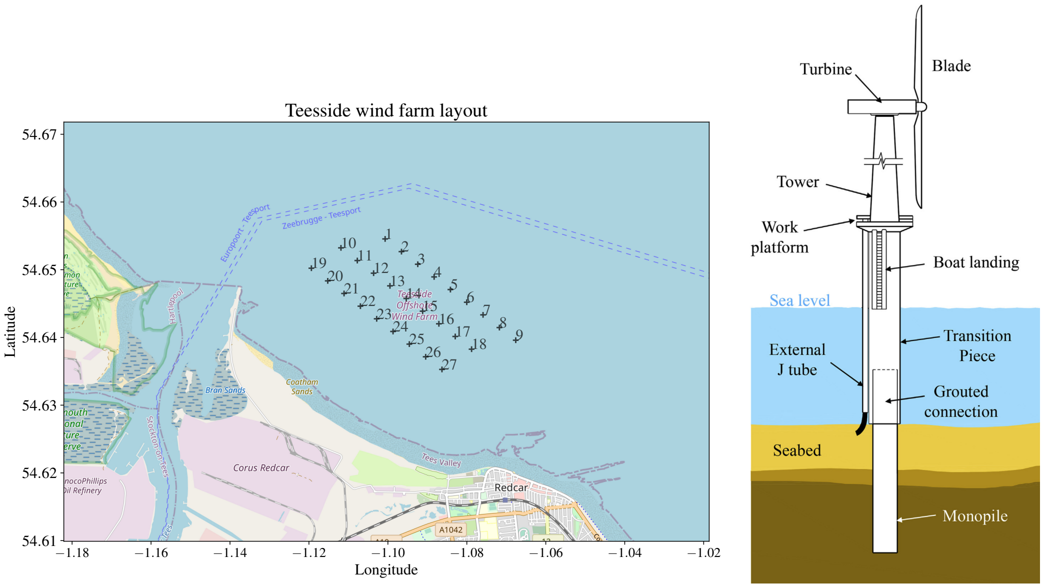


Figure 3.2 Teesside wind farm layout (left). Monopile OWT diagram (Chen et al., 2018a) (right)

Variable	Notation	Unit	Description
Mean wind speed	U	m.s^{-1}	10-min. average horizontal wind speed
Wind turbulence	σ_U	m.s^{-1}	10-min. wind speed standard deviation
Wind direction[2]	θ_{wind}	deg.	10-min. average wind direction
Significant wave height	H_s	m	Significant wave height
Peak wave period	T_p	s	Peak 1-hour spectral wave period
Wave direction	θ_{wave}	deg.	10-min. average wave direction

Table 3.3 Description of the environmental data.

is equipped with monitoring instruments (e.g., cup anemometers). In total, five years of wind data have been collected on the turbines which are not affected by the wake on this site. Their acquisition system (usually called “SCADA” for “Supervisory Control And Data Acquisition”) has a sampling period of ten minutes. The wave data arise from a buoy placed in the middle of the farm. These data describe the physical features listed in Table 3.3. A limitation of the present study is that its controller-induced uncertainty (like wind misalignment) is not considered.

The Teesside farm is located close to the coast, making the environmental conditions very different depending on the direction (see the wind farm layout in Fig. 3.2). Since measures are also subject to uncertainties, a few checks were made to ensure that the data were physically consistent. Truncation bounds were applied since this study is focused on central tendency estimation (i.e., mean behavior) rather than extreme values. In practice, this truncation only removes extreme data points (associated with storm events). In addition, a simple trigonometric transform is applied to each directional feature to take into account their cyclic structure. Finally, the remaining features are rescaled (i.e., using a min-max normalization).

Teesside’s environmental data is illustrated by its copulogram in Fig. 3.3, a graphical tool presented in Section 2.3 to visualize multivariate data. The copulogram exhibits the marginals with univariate kernel density estimation plots (in the diagonal), and the dependence structure with scatter plots in the normalized rank space (in the upper triangle). Looking at data in the rank space instead of the initial space allows one to observe the ordinal associations between variables. The scatter plots of normalized ranks are actually a representation of the empirical copula density. Two independent variables will present a uniformly distributed scatter plot in the rank space. In the lower triangular matrix, the scatter plots are set in the physical space, merging the effects of the marginals and the dependencies (as in the usual visualization offered by the matrix plot). Since the dependence structure is theoretically modeled by an underlying copula, this plot is called *copulogram*, generalizing the well-known “correlogram” to nonlinear dependencies. It gives a synthetic and empirical decomposition of the dataset that was implemented in a new open source Python package named `copulogram`[3].

[2] Note that the two directional variables could be replaced by a wind-wave misalignment variable for a bottom-fixed technology, however, our framework can be directly transposed to floating models.

[3] GitHub repository: https://github.com/efekhari27/copulogram

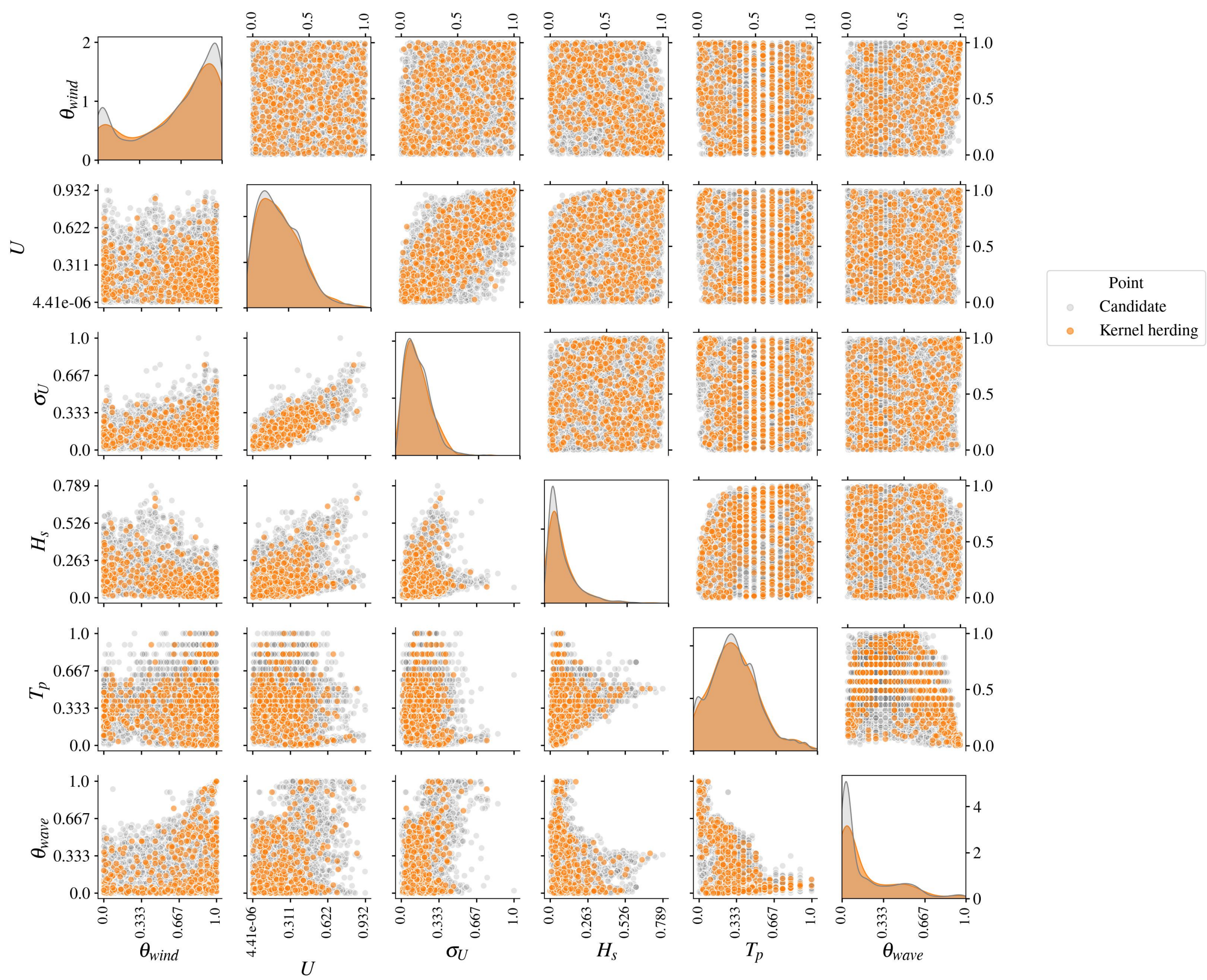


Figure 3.3 Copulogram of the Teesside measured data ($N = 10^4$ in grey), kernel herding subsample ($n = 500$ in orange). Marginals are represented by univariate kernel density estimation plots (diagonal), the dependence structure with scatter plots in the rank space (upper triangle). Scatter plots on the bottom triangle are set in the physical space.

On Fig. 3.3, a large sample $\mathcal{S} \subset \mathcal{D}_{\mathbf{X}}$ (with size $N = 10^4$) is randomly drawn from the entire Teesside data (with size $N_{\text{Teesside}} = 2 \times 10^5$), and plotted in grey. In the same figure, the orange matrix plot is a subsample of the sample $\mathcal{S}$, selected by kernel herding, a method minimizing some discrepancy measure with the sample $\mathcal{S}$ that will be presented in Section 3.3. For this example, generating the kernel herding subsample takes under one minute, which is negligible compared with the simulation time of OWT models. Visually, this orange subsample seems to be representative of the original sample both in terms of marginal distributions and dependence structure. In the following study, the large samples $\mathcal{S}$ will be considered as an empirical representation of the distribution of the random vector $\mathbf{X} \in \mathcal{D}_{\mathbf{X}}$, with probability density function $f_{\mathbf{X}}$, and called *candidate set*. Kernel herding allows direct subsampling from this large and representative dataset, instead of fitting a joint distribution and generating samples from it. Indeed, fitting a joint distribution would introduce an additional source of error in the uncertainty propagation process. Note that a proper parametric model fit would be challenging for complex dependence structures such as the one plotted on Fig. 3.3. As examples of works that followed this path, one can mention the work of Li and Zhang (2020) who built a parametric model of a similar multivariate distribution using vine copulas.

For a similar purpose and to avoid some limits imposed by the parametric framework, a nonparametric approach coupling empirical Bernstein copula fitting with kernel density estimation of the marginals is proposed in Subsection 3.2.3.

3.2.3 Non parametric fit with empirical Bernstein copula

Instead of directly subsampling from a dataset such as the one from Fig. 3.3, one could first infer a multivariate distribution and generate a sample from it. However, accurately fitting such complex multivariate distributions is challenging. The amount of available data is large enough to make nonparametric inference a viable option.

The Sklar theorem (Durante and Sempi, 2015) states that the multivariate distribution of any random vector $\mathbf{X} \in \mathbb{R}^d, d \in \mathbb{N}^*$ can be broken down into two objects:

1. A set of univariate marginal distributions to describe the behavior of the individual variables;
2. A function describing the dependence structure between all variables, called a *copula*.

This theorem states that considering a random vector $\mathbf{X} \in \mathbb{R}^d$, with its cumulative distribution function F and its marginals $\{F_i\}_{i=1}^d$, there exists a copula $C : [0,1]^d \longrightarrow [0,1]$, such that:

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right). \tag{3.1}$$

It allows us to divide the problem of fitting a joint distribution into two independent problems: fitting the marginals and fitting the copula. The empirical Bernstein copula is a nonparametric copula approximation method introduced and applied to similar data in Chapter 2. Provided a large enough learning set $\mathbf{X}_n$ (over five years in the present case), the EBC combined with kernel density estimation for the marginals fits well the environmental joint distribution related to the dataset in Fig. 3.3. Moreover, the densities of the EBC are available in an explicit form, making Monte Carlo or quasi-Monte Carlo generation easy. As discussed in Chapter 2, this method is sensitive to the chosen polynomial orders $\{m_j\}_{j=1}^d$ and the learning set size.

3.2.4 Fatigue assessment

As described in Fig. 3.1, a typical DIEGO simulation returns a 10-minute multiaxial stress time series at each node $i \in \mathbb{N}$ of the 1D meshed structure. Since classical fatigue laws are established for uniaxial stresses, the first step is to compute one equivalent Von Mises stress time series at each structural node. The present section recalls the main concepts but fatigue assessment is further discussed in Subsection **??**.

The foundation and the tower of an OWT are a succession of tubes with various sections connected by bolted or welded joints. Our work focuses on the welded joints at the mudline level, identified as a critical area for the structure. This hypothesis is confirmed in the literature by different contributions, see e.g., the results of Müller and Cheng (2018) in Figure 12, or Katsikogiannis et al. (2021). At the top of the turbine, the fatigue is commonly studied at the blade root, which was not studied here since the blades in composite material have different properties (see e.g., Dimitrov, 2013). Note that the OWT simulations provide outputs allowing us to similarly study any node along the structure (without any additional computational effort).

To compute fatigue in a welded joint, the external circle of the welding ring is discretized for a few azimuth angles $\theta \in \mathbb{R}_+$ (see the red points in the monopile cross-section on the right in Fig. 3.4). The equivalent Von Mises stress time series is then reported on the external welding ring for an azimuth θ. According to Li and Zhang (2020) and our own experience, the most critical azimuth angles are roughly aligned with the main wind and wave directions (whose distributions are illustrated in Fig. 3.4). Looking at these illustrations, the wind and wave conditions have a very dominant orientation, which is explained by the closeness of the wind farm to the shore. Then, it is assumed that azimuth angles in these directions will be more solicited, leading to higher fatigue damage. To assess fatigue damage, rainflow counting (Dowling, 1972) first identifies the stress cycles and their respective amplitudes (using the implementation of the ASTM E1049-85 rainflow cycle counting algorithm from the Python package `rainflow`[4]). For each identified stress cycle of amplitude, $s \in \mathbb{R}_+$, the so-called "Stress vs. Number of cycles" curve (also called the "S-N curve" or "Wöhler curve") allows one to estimate the number N_{c} of similar stress cycles necessary to reach fatigue ruin. The S-N curve, denoted by $W(\cdot)$ is an affine function in the log-log scale with slope $-m$ and y-intercept a:

$$N_{\mathrm{c}}(s) = a s^{-m}, \; a \in \mathbb{R}_+, \, m \in \mathbb{R}_+. \tag{3.2}$$

Finally, a usual approach to compute the damage is to aggregate the fatigue contribution of each stress cycle identified using Miner's rule. Damage occurring during a 10-minute operating time is simulated and then scaled up to the OWT lifetime. More details regarding damage assessment and the Wöhler curve used are available in DNV-RP-C203 (2016, Sec. 2.4.6). For a realization $\mathbf{x} \in \mathcal{D}_{\mathbf{X}}$ of environmental conditions, at a structural node i, an azimuth angle θ; $k \in \mathbb{N}$ stress cycles of respective amplitude $\{s_{i,\theta}^{(j)}(\mathbf{x})\}_{j=1}^{k}$ are identified. Then, Miner's rule (see Subsection ??) defines the damage function $g_{i,\theta}(\mathbf{x}) : \mathcal{D}_{\mathbf{X}} \rightarrow \mathbb{R}_+$ by:

$$g_{i,\theta}(\mathbf{x}) = \sum_{j=1}^{k} \frac{1}{N_{\mathrm{c}}\left(s_{i,\theta}^{(j)}(\mathbf{x})\right)}. \tag{3.3}$$

As defined by the DNV standards for OWT fatigue design (DNV-RP-C203, 2016), the quantity of interest in the present chapter is the "mean damage" $D_{\mathrm{c}}^{i,\theta}$ (also called "mean cumulative

[4] https://github.com/iamlikeme/rainflow

damage"), computed at a node i, for an azimuth angle θ:

$$D_{\mathrm{c}}^{i,\theta} = \mathbb{E}[g_{i,\theta}(\mathbf{X})] = \int_{\mathcal{D}_{\mathbf{X}}} g_{i,\theta}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \,. \tag{3.4}$$

To get a preview of the distribution of this output random variable $g_{i,\theta}(\mathbf{X})$, a histogram of a large Monte Carlo simulation ($N_{\mathrm{ref}} = 2000$) is represented in Fig. 3.5 (with a log scale). In this case, the log-damage histogram presents a little asymmetry but it is frequently modeled by a normal distribution (see e.g., Teixeira et al., 2019b).

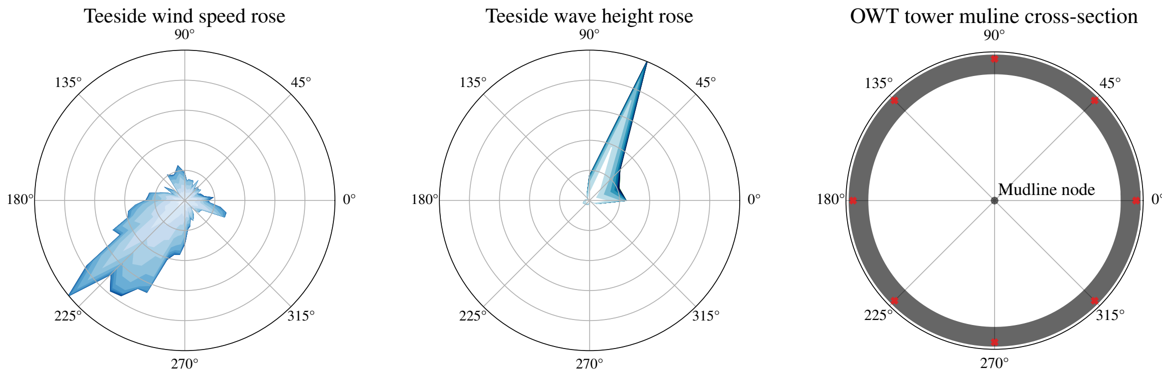


Figure 3.4 Angular distribution of the wind and waves with a horizontal cross-section of the OWT structure and the mudline. Red crosses represent the discretized azimuths for which the fatigue is computed

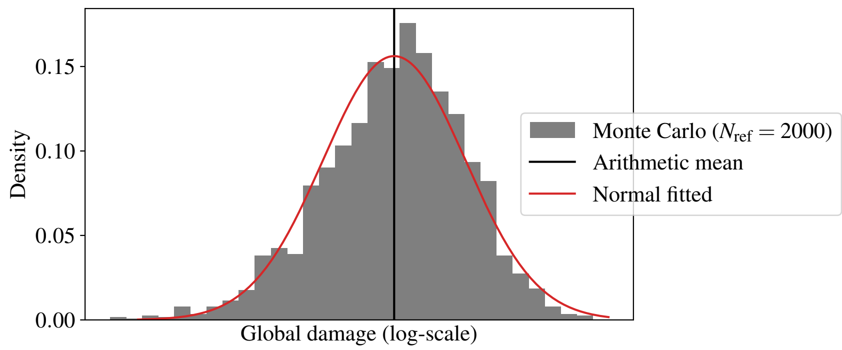


Figure 3.5 Histogram of the log-damage, at mudline, azimuth 45 deg. (Monte Carlo reference sample)

3.3 Numerical integration procedures for mean damage estimation

The present section explores different methods aiming at approximating the integral of a function against a probability measure. In the case of OWT mean damage estimation, these methods can be used for defining efficient design load cases. This problem is equivalent to the central tendency estimation of $\mathbf{Y} = g(\mathbf{X})$, the image of the environmental random variable $\mathbf{X}$ by the damage function $g(\cdot) : \mathcal{D}_\mathbf{X} \to \mathbb{R}$ (see e.g., Eq. (3.4)). Considering a measurable space $\mathcal{D}_\mathbf{X} \subset \mathbb{R}^d, d \in \mathbb{N}^*$, associated with a known probability measure π, this section studies the approximation of integrals of the form $\int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\pi(\mathbf{x})$.

3.3.1 Quadrature rules and quasi-Monte Carlo methods

Numerical integration authors may call this generic problem *probabilistic integration* (Briol et al., 2019). In practice, this quantity of interest is estimated on an n-sized set of damage realizations $\mathbf{y}_n = \left\{g(\mathbf{x}^{(1)}), \dots, g(\mathbf{x}^{(n)})\right\}$ of an input sample $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\}$. A weighted arithmetic mean of the realizations $\left\{g(\mathbf{x}^{(1)}), \dots, g(\mathbf{x}^{(n)})\right\}$ is called a *quadrature rule* with a set of unconstrained weights $\mathbf{w}_n = \{w_1, \dots, w_n\} \in \mathbb{R}^n$:

$$I_\pi(g) = \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\pi(\mathbf{x}) \approx \sum_{i=1}^{n} w_i g(\mathbf{x}^{(i)}). \tag{3.5}$$

Our numerical experiment framework often implies that the function g is costly to evaluate, making the realization number limited. For a given sample size n, our goal is to find a set of tuples $\left\{\mathbf{x}^{(i)}, w_i\right\}_{i=1}^{n}$ (i.e., quadrature rule), giving the best approximation of our quantity. In the literature, a large panel of numerical integration methods has been proposed to tackle this problem. For example, Van den Bos (2020) recently developed a quadrature rule based on arbitrary sample sets which was applied to a similar industrial OWT use case.

Alternatively, sampling methods rely on generating a set of points $\mathbf{X}_n$ drawn from the input distribution to compute the arithmetic mean of their realizations (i.e., uniform weights $\left\{w_i = \frac{1}{n}\right\}_{i=1}^{n}$). Among them, low-discrepancy sequences, also called "quasi-Monte Carlo" sampling (e.g., Sobol', Halton, Faure sequences) are known to improve the standard Monte Carlo convergence rate and will be used as a deterministic reference method in the following numerical experiments (Leobacher and Pillichshammer, 2014).

Quantization of probability measures and quadrature When dealing with probabilistic integration such as Eq. (3.5), a quadrature rule is a finite representation of a continuous measure π by a discrete measure $\zeta_n = \sum_{i=1}^{n} w_i \delta(\mathbf{x}^{(i)})$ (weighted sum of Dirac distributions at the design points $\mathbf{X}_n$). In the literature, this procedure is also called *quantization* of a continuous measure π. Overall, numerical integration is a particular case of probabilistic integration against a uniform input measure. For uniform measures, the Koksma-Hlawka inequality (Morokoff and Caflisch, 1995) provides a useful upper bound on the absolute error of a quadrature rule:

$$\left| \int_{[0,1]^d} g(\mathbf{x}) \mathrm{d}\mathbf{x} - \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{x}^{(i)}) \right| \leq V(g) D_n^*(\mathbf{X}_n). \tag{3.6}$$

As presented in Oates (2021), $V(g) = \sum_{\mathfrak{u}\subseteq\{1,\ldots,p\}} \int_{[0,1]^{\mathfrak{u}}} \left|\frac{\partial^{\mathfrak{u}} g}{\partial \mathbf{x}_{\mathfrak{u}}}(\mathbf{x}_{\mathfrak{u}}, 1)\right| \mathrm{d}\mathbf{x}$, quantifies the complexity of the integrand, while $D_n^*(\mathbf{X}_n)$ evaluates the discrepancy to uniformity of the design $\mathbf{X}_n$. Therefore, the Koksma-Hlawka inequality shows that the quadrature rule's accuracy relies on the good quantization of π by $\mathbf{X}_n$. For a uniform measure π, the star discrepancy $D_n^*(\mathbf{X}_n)$ is a metric assessing how far from uniformity a sample $\mathbf{X}_n$ is. When generalizing to a non-uniform measure, a good quantization of π should also lead to a good approximation of the quantity.

3.3.2 Kernel herding sampling

Quasi-Monte Carlo sampling methods widely rely on a metric of uniformity, called *discrepancy*. To go beyond uniform measures, Appendix B introduces a kernel-based discrepancy, generalizing the discrepancy concept to non-uniform measures. This tool, named the maximum mean discrepancy (MMD) allows comparing multivariate distributions by embedding them in a specific function space. In this manuscript, the MMD was employed as a tool for statistical testing and quantifying the perturbations of distributions in Chapter 2, and for sensitivity analysis in Section 1.7.

Herin, the MMD is used to build a quadrature rule by sampling from a known measure. In other words, to quantize a known target measure π by a design sample $\mathbf{X}_n$. For practical reasons, the design construction is done sequentially. Sequential strategies have also been used to learn and validate regression models for statistical learning (see Fekhari et al., 2023b). Moreover, since each realization is supposed to be obtained at the same unitary cost, the quadrature weights are first fixed as uniform during the construction of the design $\mathbf{X}_n$.

Kernel herding (KH), proposed by Chen et al. (2010), is a sampling method that offers a quantization of the measure π by minimizing a squared MMD when adding points iteratively. With a current design $\mathbf{X}_n$ and its corresponding discrete distribution with uniform weights $\zeta_n = \frac{1}{n}\sum_{i=1}^{n} \delta(\mathbf{x}^{(i)})$, a KH iteration is as an optimization of the following criterion, selecting the point $\mathbf{x}^{(n+1)} \in \mathcal{D}_{\mathbf{X}}$:

$$\mathbf{x}^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{D}_{\mathbf{X}}} \left\{ \mathrm{MMD}\left(\pi, \frac{1}{n+1}\left(\delta(\mathbf{x}) + \sum_{i=1}^{n} \delta(\mathbf{x}^{(i)})\right)\right)^2 \right\}. \tag{3.7}$$

In the literature, two formulations of this optimization problem can be found. The first one uses the Frank-Wolfe algorithm (or "conditional gradient algorithm") to compute a linearization of the problem under the convexity hypothesis (see Lacoste-Julien et al., 2015 and Briol et al., 2015 for more details). The second one is a straightforward greedy optimization. Due to the combinatorial complexity, the greedy formulation is tractable for sequential construction only.

Let us develop the MMD criterion from Eq. (??):

$$\mathrm{MMD}\left(\pi, \frac{1}{n+1}\left(\delta(\mathbf{x}) + \sum_{i=1}^{n}\delta(\mathbf{x}^{(i)})\right)\right)^2 = \varepsilon_\pi - \frac{2}{n+1}\sum_{i=1}^{n+1} P_\pi\left(\mathbf{x}^{(i)}\right) + \frac{1}{(n+1)^2}\sum_{i,j=1}^{n+1} k\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right) \tag{3.8a}$$

$$= \varepsilon_\pi - \frac{2}{n+1}\left(P_\pi(\mathbf{x}) + \sum_{i=1}^{n} P_\pi\left(\mathbf{x}^{(i)}\right)\right) \tag{3.8b}$$

$$+ \frac{1}{(n+1)^2}\left(\sum_{i,j=1}^{n} k\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right) + 2\sum_{i=1}^{n} k\left(\mathbf{x}^{(i)}, \mathbf{x}\right) - k(\mathbf{x}, \mathbf{x})\right). \tag{3.8c}$$

In the previously developed expression, only a few terms actually depend on the next optimal point $\mathbf{x}^{(n+1)}$ since the target energy, denoted by ε_π, and $k(\mathbf{x}, \mathbf{x}) = \sigma^2$ are constant (by taking a stationary kernel). Therefore, the greedy minimization of the MMD can be equivalently written as:

$$\mathbf{x}^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{D}_\mathbf{X}} \left\{\frac{1}{n+1}\sum_{i=1}^{n} k\left(\mathbf{x}^{(i)}, \mathbf{x}\right) - P_\pi(\mathbf{x})\right\} = \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{D}_\mathbf{X}} \left\{\frac{n}{n+1}P_{\zeta_n}(\mathbf{x}) - P_\pi(\mathbf{x})\right\}. \tag{3.9}$$

Remark 3. For the sequential and uniformly weighted case, the formulation in Eq. (3.9) is almost similar to the Frank-Wolfe formulation. Our numerical experiments showed that these two versions generate very close designs, especially as n becomes large. Pronzato and Rendas (2023) express the Frank-Wolfe formulation in the sequential and uniformly weighted case as follows:

$$\mathbf{x}^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{D}_\mathbf{X}} \left\{P_{\zeta_n}(\mathbf{x}) - P_\pi(\mathbf{x})\right\}. \tag{3.10}$$

Remark 4. In practice, the optimization problem is solved by a brute-force approach on a fairly dense finite subset $\mathcal{S} \subseteq \mathcal{D}_\mathbf{X}$ of candidate points with size $N \gg n$ that emulates the target distribution, also called the "candidate set". This sample is also used to estimate the target potential $P_\pi(\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N} k\left(\mathbf{x}^{(i)}, \mathbf{x}\right)$.

The diagram illustrated in Fig. 3.6 summarizes the main steps of a kernel herding sampling algorithm. One can notice that the initialization can either be done using a median point (maximizing the target potential) or from any existing design of experiments. This second configuration is practical when the analyst must include some characteristic points in the design (e.g., points with a physical interpretation).

As explained previously, choosing the kernel defines the function space on which the worst-case function is found (see Eq. (B.7)). Therefore, this sampling method is sensitive to the kernel's choice. A kernel is defined, both by the choice of its parametric family (e.g., Matérn, squared exponential) and the choice of its tuning. The so-called "support points" method developed by Mak and Joseph (2018) is a special case of kernel herding that uses the

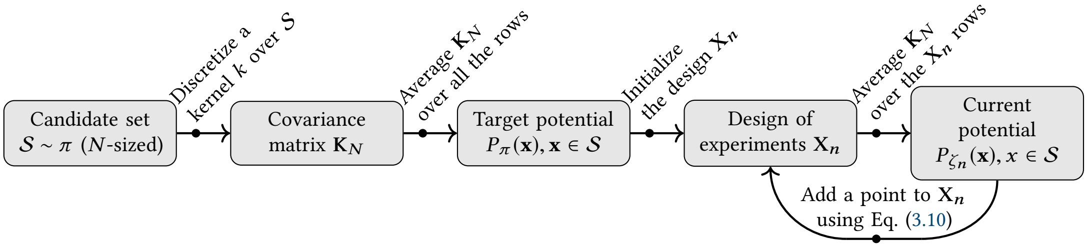


Figure 3.6 Greedy kernel herding algorithm

Energy-distance	$k_E(\mathbf{x},\mathbf{x}') = \frac{1}{2}\left(\|\mathbf{x}\| + \|\mathbf{x}'\| - \|\mathbf{x}-\mathbf{x}'\|\right)$	
Squared exponential	$k_G(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{p} k_{\theta_i}(x_i - x_i')$	$k_\theta(x-x') = \exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$
Matérn ($\nu = 5/2$)	$k_M(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{p} k_{5/2,\theta_i}(x_i - x_i')$	$k_{5/2,\theta}(x-x') = \left(1 + \frac{\sqrt{5}}{\theta}\lvert x-x'\rvert + \frac{5}{3\theta^2}(x-x')^2\right)\exp\left(-\frac{\sqrt{5}}{\theta}\lvert x-x'\rvert\right)$

Table 3.4 Kernels considered in the following numerical experiments.

characteristic and parameter-free "energy-distance" kernel (introduced by Székely and Rizzo, 2013). In the following numerical experiments, the energy-distance kernel will be compared with an isotropic tensor product of a Matérn kernel (with regularity parameter $\nu = 5/2$ and correlation lengths θ_i), or a squared exponential kernel (with correlation lengths θ_i) defined in Table 3.4. Since the Matérn and squared exponential kernels are widely used for Gaussian process regression (Rasmussen and Williams, 2006), they were naturally picked to challenge the energy-distance kernel. The correlation lengths for the squared exponential and Matérn kernels are set using the heuristic given in Fekhari et al. (2023b), $\theta_i = n^{-1/d}, i \in \{1, \ldots, d\}$.

Fig. 3.7 represents the covariance structure of the three kernels. One can notice that the squared exponential and Matérn $\nu = 5/2$ kernels are closer to one another than they are to the energy-distance. In fact, as ν tends to infinity, the Matérn kernel tends toward the squared exponential kernel (which has infinitely differentiable sample paths, see Rasmussen and Williams, 2006). For these two stationary kernels, the correlation length controls how fast the correlation between two points decreases as their distance from one another increases.

Meanwhile, the energy distance is not stationary (but still positive and semi-definite). Therefore, its value does not only depend on the distance between two points but also on the norm of each of the points. Interestingly, the energy-distance kernel is almost similar to the kernel used by Hickernell (1998) to define a widely-used space-filling metric called the centered L^2-discrepancy. A presentation of these kernel-based discrepancies from the design of experiment point of view is also provided in Chapter Two from Fang et al. (2018).

To illustrate the kernel herding sampling of a complex distribution, Fig. 3.8 shows three nested samples (orange crosses for different sizes $n \in \{10, 20, 40\}$) of a mixture of Gaussian distributions with complex nonlinear dependencies (with density represented by the black isoprobability contours). In this example, the method seems to build a parsimonious design

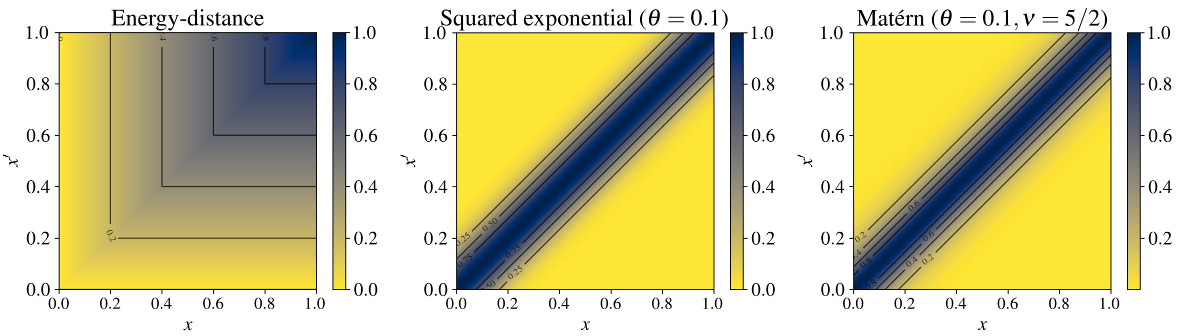


Figure 3.7 Kernel illustrations (left to right: energy-distance, squared exponential, and Matérn 5/2)

between each mode of the distribution (by subsampling directly without any transformation). The candidate set (in light grey) was generated by a large quasi-Monte sample of the underlying Gaussian mixture. In this two-dimensional case, this candidate set is sufficient to estimate the target potential P_π. However, the main bottleneck of kernel herding is the estimation of the potentials, which becomes costly in high dimensions.

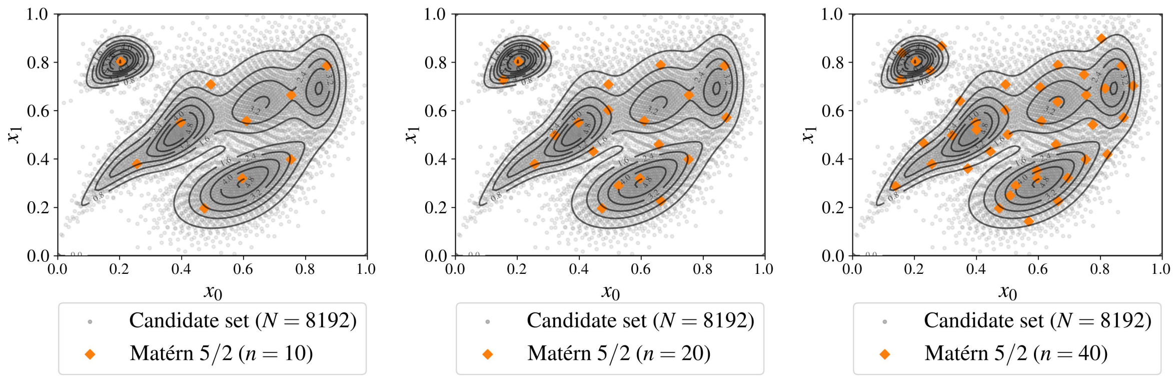


Figure 3.8 Sequential kernel herding for increasing design sizes ($n \in \{10, 20, 40\}$) built on a candidate set of $N = 8196$ points drawn from a complex Gaussian mixture π

Other approaches take advantage of the progressive knowledge acquired sequentially from the outputs to select the following points in the design. These methods are sometimes called "active learning" or "adaptive strategies" (Fuhg. et al., 2021). Many of them rely on a sequentially updated Gaussian process (or Kriging) metamodel. To solve a probabilistic integration problem, the concept of Bayesian quadrature is introduced in the following.

3.3.3 Bayesian quadrature

Gaussian processes for Bayesian quadrature Kernel methods and Gaussian processes present a lot of connections and equivalences, thoroughly reviewed by Kanagawa et al. (2018). In numerical integration, Gaussian processes have been used to build quadrature rules in the seminal paper of O'Hagan (1991), introducing the concept of *Bayesian quadrature* (BQ). Let us recall the probabilistic integration problem $I_\pi(g) = \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\pi(\mathbf{x})$ (stated in Eq. (3.5)). From a general point of view, this quantity could be generalized by composing g with another function ψ (e.g., other moments, quantiles, exceedance probabilities). The quantity of interest then becomes, $I_\pi(\psi(g))$, for example when ψ is a monomial, it gives a moment of the output distribution.

Let us assume, adopting a Bayesian point of view, that G is a stochastic process describing the uncertainty affecting the knowledge about the true function g. Let G be a Gaussian process (GP) prior with a zero trend (denoted by $\mathbf{0}$) to ease the calculation, and a stationary covariance kernel (denoted by $k(\cdot,\cdot)$). The conditional posterior $G_n = (G|\mathbf{y}_n) \sim \mathrm{GP}(\eta_n, s_n^2)$ has been conditioned on the function observations $\mathbf{y}_n = \left[g\left(\mathbf{x}^{(1)}\right), \dots, g\left(\mathbf{x}^{(n)}\right)\right]^\top$ computed from the input design $\mathbf{X}_n$ and is fully defined by the well-known "Kriging equations" (see e.g., Rasmussen and Williams, 2006):

$$\begin{cases} \eta_n(\mathbf{x}) & = \mathbf{k}_n^\top(\mathbf{x}) \mathbf{K}_n^{-1} \mathbf{y}_n \\ s_n^2(\mathbf{x}) & = k_n(\mathbf{x},\mathbf{x}) - \mathbf{k}_n^\top(\mathbf{x}) \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}) \end{cases} \tag{3.11}$$

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

In BQ, the main object is the random variable $I_\pi(G_n)$. According to Briol et al. (2019), its distribution on $\mathbb{R}$ is the pushforward of G_n through the integration operator $I_\pi(\cdot)$, sometimes called *posterior distribution*:

$$I_\pi(G_n) = \int_{\mathcal{D}_\mathbf{X}} (G(\mathbf{x})|\mathbf{y}_n) \mathrm{d}\pi(\mathbf{x}) = \int_{\mathcal{D}_\mathbf{X}} G_n(\mathbf{x}) \mathrm{d}\pi(\mathbf{x}) \,. \tag{3.12}$$

Fig. 3.9 provides a one-dimensional illustration of the Bayesian quadrature of an unknown function (dashed black curve) against a given input measure π (with corresponding grey distribution at the bottom). For an arbitrary design, one can fit a Gaussian process model, interpolating the function observations (black crosses). Then, multiple trajectories of this conditioned Gaussian process G_n are drawn (orange curves) whilst its mean function, also called "predictor", is represented by the red curve. Therefore, the input measure π is propagated through the conditioned Gaussian process to obtain the random variable $I_\pi(G_n)$, with distribution represented on the right plot (brown curve). Again on the right plot, remark how the mean of this posterior distribution (brown line) is closer to the reference output expected value (dashed black line) than the arithmetic mean of the observations (black line). This plot was inspired by the paper of Huszár and Duvenaud (2012).

Optimal weights computed by Bayesian quadrature Taking the random process G_n as Gaussian conveniently implies that its posterior distribution $a_\pi(G_n)$ is also Gaussian. This comes from the linearity of the infinite sum of realizations of a Gaussian process. The posterior distribution is described in a closed form through its mean and variance by applying Fubini's theorem (see the supplementary materials from Briol et al., 2019 for the proof regarding the

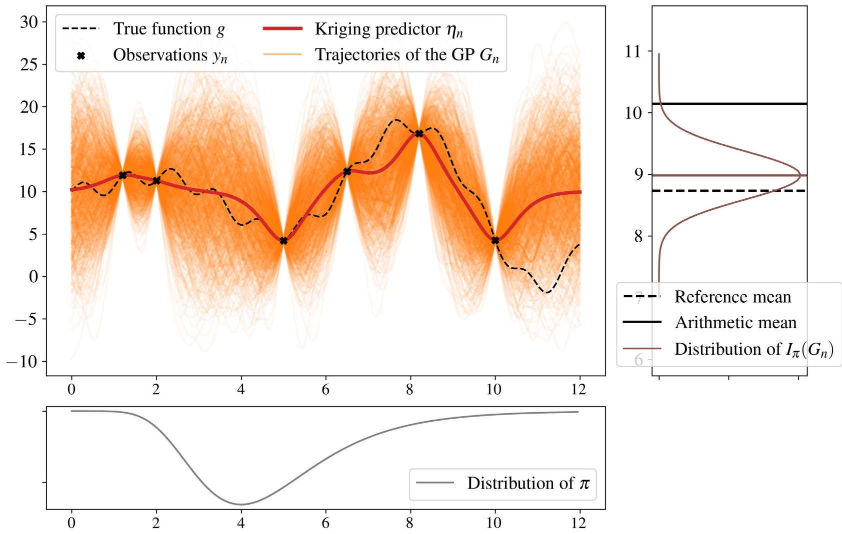


Figure 3.9 Bayesian quadrature on a one-dimensional case

variance):

$$\overline{y}_n^{\mathrm{BQ}} = \mathbb{E}\left[I_\pi(G_n)|\mathbf{y}_n\right] = \int_{\mathcal{D}_{\mathbf{X}}} \eta_n(\mathbf{x})\mathrm{d}\pi(\mathbf{x}) = \left[\int_{\mathcal{D}_{\mathbf{X}}} \mathbf{k}_n^\top(\mathbf{x})\mathrm{d}\pi(\mathbf{x})\right] \mathbf{K}_n^{-1}\mathbf{y}_n = P_\pi(\mathbf{X}_n)\mathbf{K}_n^{-1}\mathbf{y}_n, \tag{3.13}$$

$$\left(\sigma_n^{\mathrm{BQ}}\right)^2 = \mathrm{Var}\left(I_\pi(G_n)\right) = \iint_{\mathcal{D}_{\mathbf{X}^2}} k_n(\mathbf{x},\mathbf{x}')\mathrm{d}\pi(\mathbf{x})\mathrm{d}\pi(\mathbf{x}') = \varepsilon_\pi - P_\pi(\mathbf{X}_n)\mathbf{K}_n^{-1}P_\pi(\mathbf{X}_n)^\top. \tag{3.14}$$

Where $P_\pi(\mathbf{X}_n)$ is the row vector of potentials $\left[\int k_n(\mathbf{x},\mathbf{x}^{(1)})\mathrm{d}\pi(\mathbf{x}),\dots,\int k_n(\mathbf{x},\mathbf{x}^{(n)})\mathrm{d}\pi(\mathbf{x})\right]$, and ε_π is given in Eq. (B). As in the one-dimensional example presented in Fig. 3.9, the expected value of $I_\pi(G_n)$ expressed in Eq. (3.13) is a direct estimator of the quantity of interest Eq. (3.5). The so-called "Bayesian quadrature estimator" appears to be a simple linear combination of the observations by taking the row vector of "optimal weights" as:

$$\mathbf{w}_{\mathrm{BQ}} = P_\pi(\mathbf{X}_n)\mathbf{K}_n^{-1} \tag{3.15}$$

For any given sample, an optimal set of weights can be computed, leading to the mean of the posterior distribution. Remark here that this enhancement depends on the evaluation of the inverse variance-covariance matrix $\mathbf{K}_n^{-1}$, which can present numerical difficulties, either when design points are too close, making the conditioning bad. Moreover, a prediction interval on the BQ estimator can be computed since the posterior distribution is Gaussian, with a variance expressed in closed-form in Eq. (3.14). The expressions in Eq. (3.13) and Eq. (3.14) were extended to Gaussian processes in the case of constant and linear trends in Pronzato and Zhigljavsky (2020). In the following numerical experiments, the expression with a hypothesis of constant trend β_n is used, which leads to:

$$\mathbb{E}\left[I_\pi(G_n)\right] = \beta_n + P_\pi(\mathbf{X}_n)\mathbf{K}_n^{-1}\left(\mathbf{y}_n - \beta_n \mathbf{1}_n\right). \tag{3.16}$$

Then, an a posteriori 95% prediction interval around the mean Bayesian estimator is directly given by:

$$\overline{y}_n^{\mathrm{BQ}} \in \left[\overline{y}_n^{\mathrm{BQ}} - 2\sigma_n^{\mathrm{BQ}}, \overline{y}_n^{\mathrm{BQ}} + 2\sigma_n^{\mathrm{BQ}}\right]. \tag{3.17}$$

Variance-based Bayesian quadrature rule The link between the posterior variance and the squared MMD has been first made by Huszár and Duvenaud (2012) in their Proposition 1: the expected variance in the Bayesian quadrature $\mathrm{Var}(I_\pi(G_n))$ is the MMD between the target distribution π and $\zeta_n = \sum_{i=1}^n \mathbf{w}_{\mathrm{BQ}}^{(i)} \delta(\mathbf{x}^{(i)})$. The proof is reproduced below (as well as in Proposition 6.1 from Kanagawa et al., 2018):

$$\begin{aligned}
\mathrm{Var}(I_\pi(G_n)) &= \mathbb{E}\left[\left(I_\pi(G_n) - I_{\zeta_n}(G_n)\right)^2\right] && (3.18a)\\
&= \mathbb{E}\left[\left(\langle G_n, P_\pi\rangle_{\mathcal{H}(k)} - \left\langle G_n, P_{\zeta_n}\right\rangle_{\mathcal{H}(k)}\right)^2\right] && (3.18b)\\
&= \mathbb{E}\left[\left\langle G_n, P_\pi - P_{\zeta_n}\right\rangle^2_{\mathcal{H}(k)}\right] && (3.18c)\\
&= \|P_\pi - P_{\zeta_n}\|^2_{\mathcal{H}(k)} && (3.18d)\\
&= \mathrm{MMD}(\pi, \zeta_n)^2. && (3.18e)
\end{aligned}$$

Note that the transition from equation Eq. (3.18c) to Eq. (3.18d) relies on the property stating that if G is a standard Gaussian process then $\forall g \in \mathcal{H}(k) : \langle G, g\rangle_{\mathcal{H}(k)} \sim \mathcal{N}(0, \|g\|^2_{\mathcal{H}(k)})$. The method that sequentially builds a quadrature rule by minimizing this variance is called by the authors "sequential Bayesian quadrature". According to the previous proof, this criterion can be seen as an optimally-weighted version of the kernel herding criterion, as stated in the title of the paper from Huszár and Duvenaud (2012). Later, Briol et al. (2015) proved the weak convergence of $I_\pi(G_n)$ towards the target integral. Closer to wind turbine applications, Huchet (2019) and Huchet et al. (2019) introduced the "Adaptive Kriging Damage Assessment" method: a Kriging-based method for mean damage estimation that is very close to sequential Bayesian quadrature. However, This type of method inherits the limits from both KH and BQ since it searches for optimal design points among a candidate set and computes an inverse variance-covariance matrix. These numerical operations both scale hardly in high dimension.

Remark 5. Every quadrature method introduced in this section has been built without any observation of the possibly costly function g. Therefore, they cannot be categorized as active learning approaches. Contrarily, Kanagawa and Hennig (2019) presents a set of methods for BQ with transformations (i.e., adding a positivity constraint on the function g), which are truly active learning methods.

3.4 Numerical experiments

This section presents numerical results computed on two different analytical toy cases, respectively in dimension 2 (toy case #1) and dimension 10 (toy case #2), with easy-to-evaluate functions $g(\cdot)$ and associated input distributions π. Therefore, reference values can easily be computed with great precision. For each toy case, a large reference Monte Carlo sample ($N_{\text{ref}} = 10^8$) is taken. This first benchmark compares the mean estimation of toy cases given by a quasi-Monte Carlo technique (abbreviated by QMC in the next figures) which consists herein using a Sobol' sequence, and kernel herding with the three kernels defined in Table 3.4. Notice that the quasi-Monte Carlo designs are first generated on the unit hypercube and then, transformed using the generalized Nataf transformation to follow the target distribution (Lebrun and Dutfoy, 2009a). Additionally, the performances of kernel herding for both uniform and optimally-weighted Eq. (3.16) estimators are compared.

All the following results and methods (i.e., the kernel-based sampling and BQ methods) have been implemented in a new open source Python package named `otkerneldesign`[5]. This development mostly relies on the open source software `OpenTURNS` ("Open source initiative for the Treatment of Uncertainties, Risks'N Statistics") devoted to uncertainty quantification and statistical learning (Baudin et al., 2017). Finally, note that the numerical experiments for the toy cases are available in the Git repository named `ctbenchmark`[6].

3.4.1 Benchmark results on analytical toy-cases

The toy cases were chosen to cover a large panel of complex probabilistic integration problems, completing the ones from Fekhari et al. (2022). To assess the complexity of numerical integration problems, Owen (2003) introduced the concept of the "effective dimension" of an integrand function (number of the variables that actually impact the integral). The author showed that functions built on sums yield a low effective dimension (unlike functions built on products). In the same vein, Kucherenko et al. (2011) build three classes of integrand sorted by difficulty depending on their effective dimension:

- *class A*: problem with a few dominant variables.
- *class B*: problem without unimportant variables, and important low-order interaction terms.
- *class C*: problems without unimportant variables, and important high-order interaction terms.

The 10-dimensional "GSobol function" (toy case #2) with a set of coefficient $\{a_i = 2\}_{i=1}^{10}$ has an effective dimension equal to 10 and belongs to the hardest class C from Kucherenko et al. (2011). In the case of the two-dimensional Gaussian mixture problem, the complexity is carried by the

[5] https://efekhari27.github.io/otkerneldesign/master/index.html
[6] https://github.com/efekhari27/ctbenchmark

Table 3.5 Analytical toy-cases

Toy-case #1	$dim = 2$	$g_1(\mathbf{x}) = x_1 + x_2$	Gaussian mixture from Fig. 3.8
Toy-case #2	$dim = 10$	$g_2(\mathbf{x}) = \prod_{i=1}^{10} \frac{\lvert 4x_i - 2\rvert + a_i}{1+a_i}, \{a_i = 2\}_{i=1}^{10}$	Gaussian $\mathcal{N}(\mathbf{0.5}, \mathbf{I}_{10})$

mixture of Gaussian distributions with highly nonlinear dependencies. Probabilistic integration results are presented in Fig. 3.10 (toy case #1) and Fig. 3.11 (toy case #2). Kernel herding samples using the energy-distance kernel are in red, while quasi-Monte Carlo samples built from Sobol' sequences are in grey. Convergences of the arithmetic means are plotted on the left and MMDs on the right. The respective BQ estimators of the means are plotted in dashed lines.

Remark 6. Different kernels are used in these numerical experiments. First, the generation kernel is used by the kernel herding algorithm to generate designs (with the heuristic tuning defined in Section 3.3.2). Second, the BQ kernel allows computation of the optimal weights (arbitrarily set up as a Matérn $5/2$ with the heuristic tuning). Third, the evaluation kernel must be common to allow a fair comparison of the computed MMD results (same as the BQ kernel).

Results analysis for toy case #1. Convergence plots are provided in Fig. 3.10. KH consistently converges faster than quasi-Monte Carlo in this case, especially for small sizes in terms of MMD. BQ weights tend to reduce the fluctuations in the mean convergence, which ensures better performance for any size. Overall, applying the weights enhances the convergence rate.

Results analysis for toy case #2. Convergence plots are provided in Fig. 3.11. Although quasi-Monte Carlo is known to suffer the "curse of dimensionality", KH does not outperform it drastically in this example. In fact, KH with uniform weights performs worse than quasi-Monte Carlo while optimally-weighted KH does slightly better. Moreover, the results confirm that $\mathrm{MMD}_{\mathrm{BQ}} < \mathrm{MMD}_{\mathrm{unif}}$ for all our experiments. The application of optimal weights to the quasi-Monte Carlo sample slightly improves the estimation in this case. Note that the prediction interval around the BQ estimator is not plotted for the sake of readability.

In these two toy cases, the MMD is shown to quantify numerical integration convergence well, which illustrates the validity of the inequality given in Eq. (B.6c), similar to the Koksma-Hlawka inequality (as recalled in Eq. (3.6)).

3.4.2 Application to the Teesside wind turbine fatigue estimation

Let us summarize the mean damage estimation strategies studied in this chapter. The diagram represented in Fig. 3.12 describes the different workflows computed. The simplest workflow is represented by the grey horizontal sequence. It directly subsamples a design of experiments from a large and representative dataset (previously referred to as candidate set). This workflow simply estimates the mean damage by computing an arithmetic average of the outputs.

Alternatively, one can respectively fit a joint distribution and sample from it. In our case, this distribution is only known empirically via the candidate set. Since its dependence structure is complex (see Fig. 3.3), a parametric method might fit the distribution poorly (and therefore

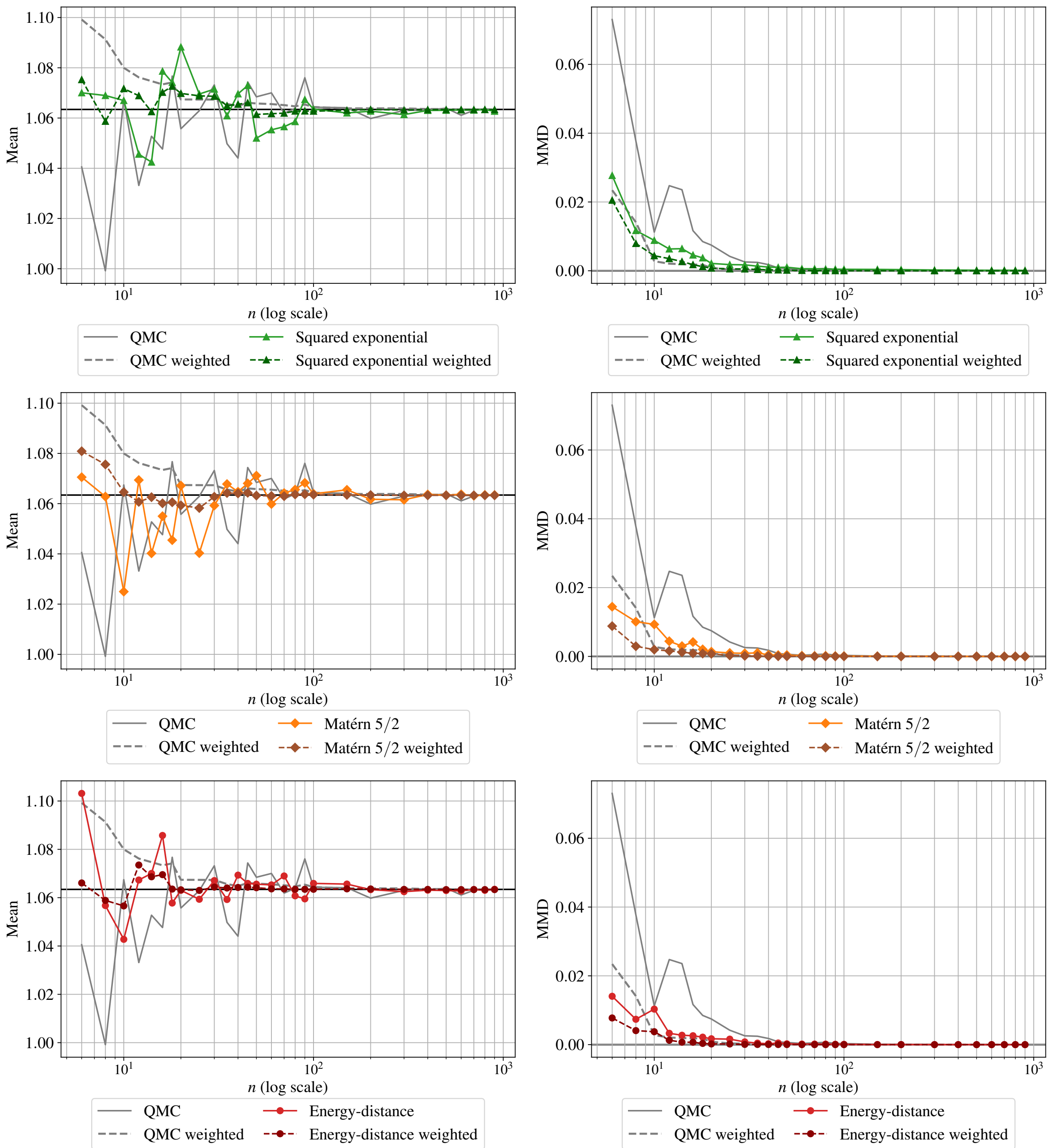


Figure 3.10 Analytical benchmark results on the toy-case #1

lead to a poor estimation of the quantity). Then, a nonparametric fit using the empirical Bernstein copula (introduced in Subsection 3.2.3) coupled with a kernel density estimation on each marginal is applied to the candidate set (with the EBC parameter $m = 100 > m_{\mathrm{MISE}}$ to avoid bias, see Lasserre, 2022, p.117). The sampling on this hybrid joint distribution is realized with a quasi-Monte Carlo method. A Sobol' low-discrepancy sequence generates a uniform sample in the unit hypercube, which can then be transformed according to a target distribution. Remember that quasi-Monte Carlo sampling is also sensitive to the choice of a low-discrepancy sequence, each presenting different properties (e.g., Sobol', Halton, Faure, etc.). Finally, the

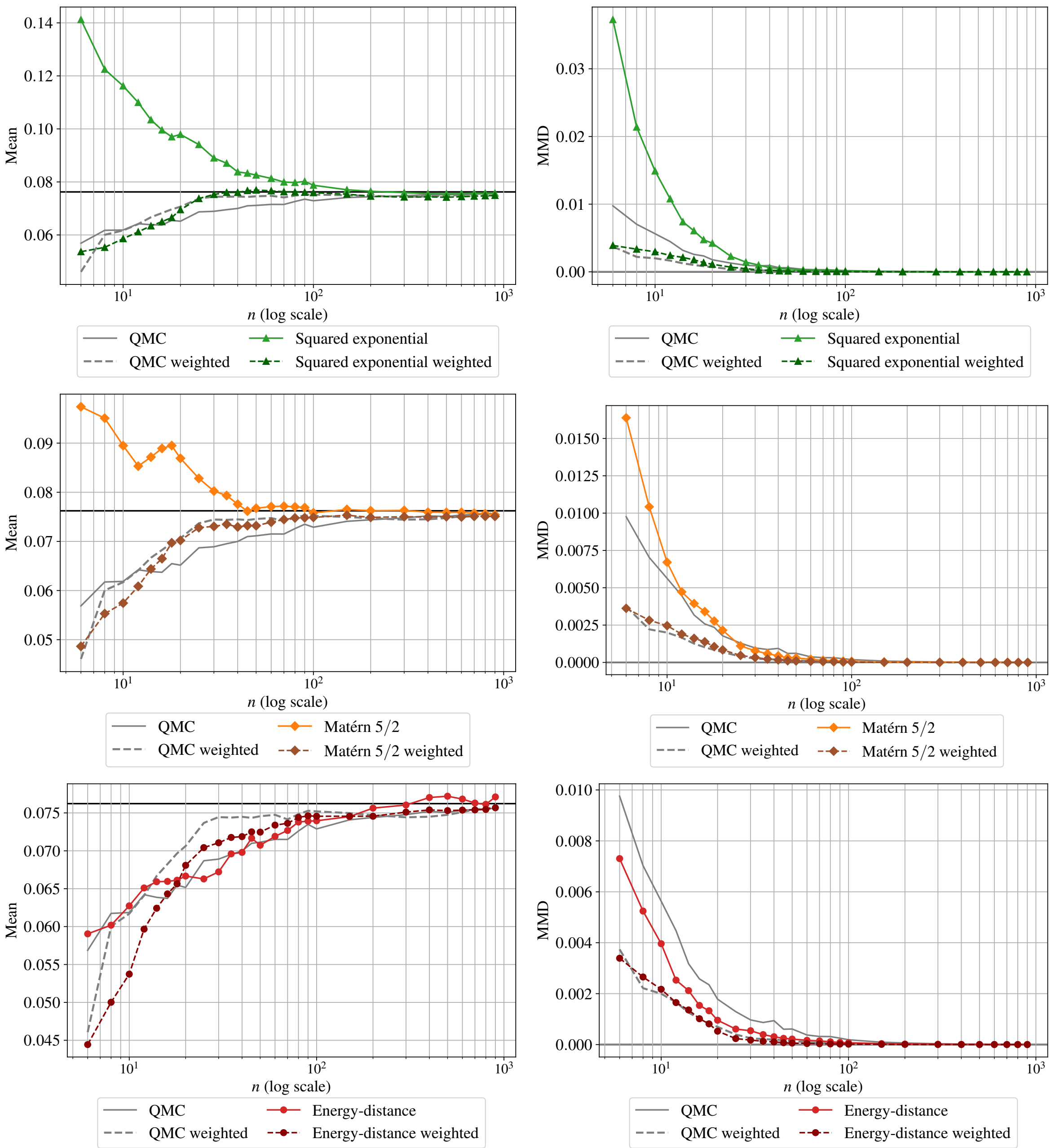


Figure 3.11 Analytical benchmark results on the toy-case #2

estimation by an arithmetic mean can be replaced by an optimally weighted mean. To do so, optimal weight must be computed, using the formulas introduced in Eq. (3.15).

The copulogram in Fig. 3.13 illustrates the intensity of the computed damages, proportionally to the color scale. Note that the numerical values of the damage scale are kept confidential since it models the state of an operating asset. Before analyzing the performance of the KH on this industrial application, let us notice that the copulogram Fig. 3.13 seems to be in line with the global sensitivity analysis presented in Murcia et al. (2018) and Li and Zhang (2020). In particular, the fact that the scatter plot of mean wind speed vs. turbulence wind speed is the main factor explaining the variance of the output $Y = g(\mathbf{X})$. Judging from these references,

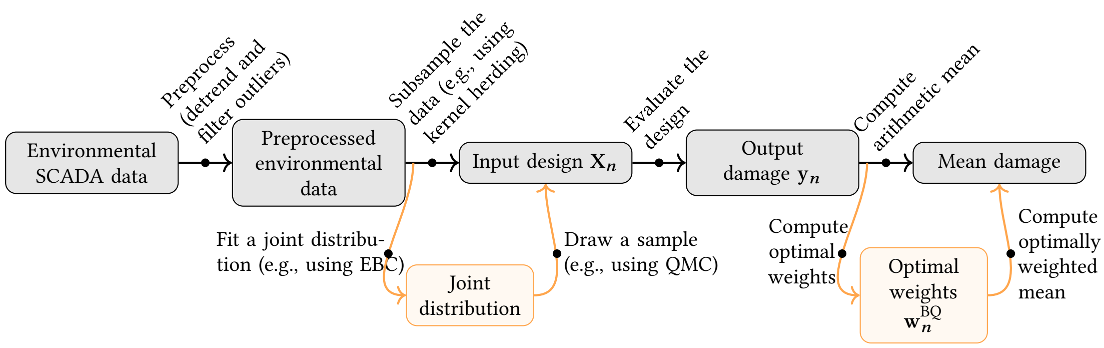


Figure 3.12 Mean damage estimation workflows for the industrial use case. The orange parts represent optional alterations to the workflow: the first one is an alternative to input data subsampling where the underlying distribution is sampled from, the second one improves mean damage calculation by using optimal weights over the output data

the numerical model does not seem to have a highly effective dimension, however, the input dependence structure is challenging and the damage assessment induces strong nonlinearities (see Eq. (3.2)).

The results presented are compared in the following to a large reference Monte Carlo sample (size 2000) with a confidence interval computed by bootstrap (see Fig. 3.14). This reference is represented by a horizontal line intersecting with the most converged Monte Carlo estimation. Once again, the mean damage scale is hidden for confidentiality reasons, but all the plots are represented for the same vertical scale. The performance of the KH is good: it quickly converges towards the confidence interval of the Monte Carlo obtained with the reference sample. In addition, the Bayesian quadrature estimator also offers a posteriori prediction interval, which can reassure the user. The BQ prediction intervals are smaller than the ones obtained by bootstrap on the reference Monte Carlo sample.

To provide more representative results, note that a set of scale parameters is computed with a Kriging procedure to define the kernel used to compute BQ intervals. Since other methods do not generate independent samples, bootstrapping them is not legitimate. Contrarily to the other kernels, we observe that the energy-distance kernel presents a small bias with the MC reference for most of the azimuth angles computed in this experiment. Meanwhile, combining nonparametric fitting with quasi-Monte Carlo sampling also delivers good results as long as the fitting step does not introduce a bias. In our case, any potential bias due to poor fitting would be the result of a poorly tuned empirical Bernstein copula. Fortunately, Nagler et al. (2017) formulated recommendations regarding how to tune empirical Bernstein copulas. We follow these recommendations in the present work.

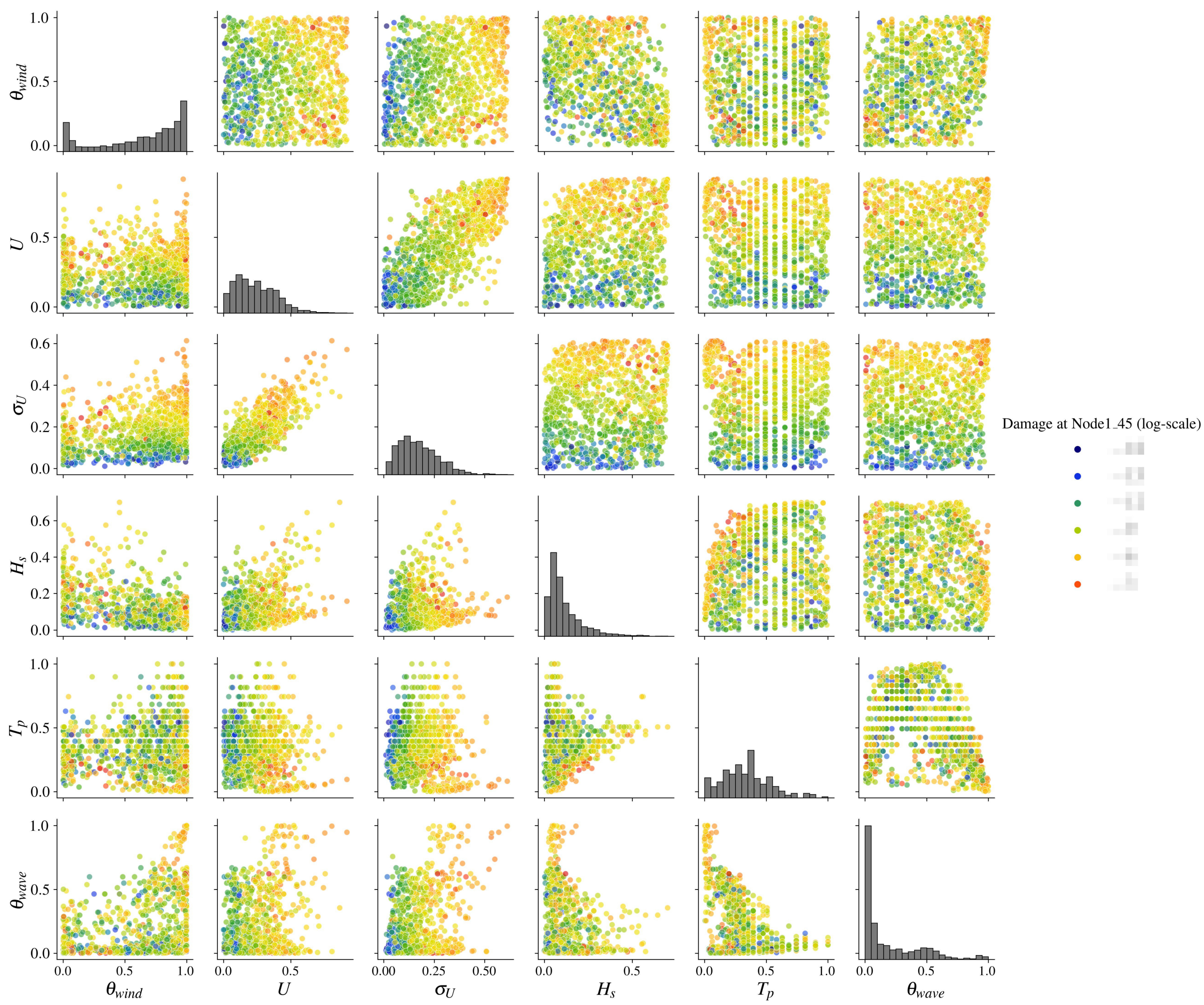


Figure 3.13 Copulogram of the kernel herding design of experiments with corresponding outputs in color (log-scale) on the Teesside case ($n = 10^3$). The color scale ranges from blue for the lowest values to red for the largest. Marginals are represented by histograms (diagonal), the dependence structure with scatter plots in the ranked space (upper triangle). Scatter plots on the bottom triangle are set in the physical space.

3.5 Conclusion

Wind energy assets are subject to highly uncertain environmental conditions. Uncertainty propagation through numerical models is performed to ensure their structural integrity (and energy production). For this case, the method recommended by the standards (regular grid sampling) is intractable for even moderate-fidelity simulators. In practice, such an approach can lead to poor uncertainty propagation, especially when facing simulation budget constraints.

In the present chapter, a real industrial wind turbine fatigue estimation use case is investigated, considering site-specific data. As a perspective, other sites with different environmental conditions could be studied. This use case induces two practical constraints: first, usual active learning methods are hard to set up on such a model (mainly due to the nonlinearity of the variable of interest), and they restrict the use of high-performance computing facilities; second,

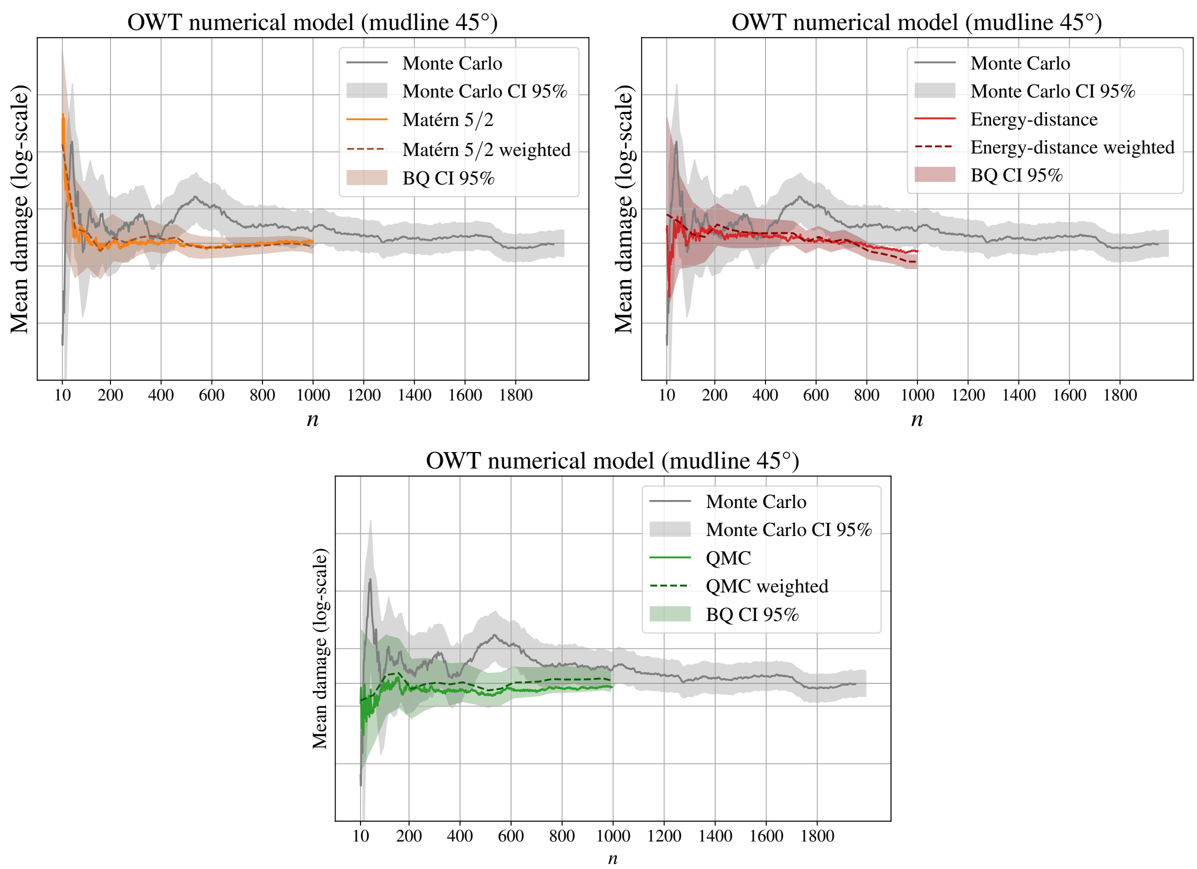


Figure 3.14 Mean estimation convergence (at the mudline, azimuth $\theta = 45$ deg.) on the Teesside case. Monte Carlo confidence intervals are all computed by bootstrap

the input distribution of the environmental conditions presents a complex dependence structure which is hard to infer with common parametric approaches.

In this work, the association of kernel herding sampling with Bayesian quadrature for central tendency estimation is explored theoretically and numerically. This method fits with the practical constraints induced by the industrial use case. To be more specific, the kernel herding method easily subsamples the relevant points directly from a given dataset (here, from the measured environmental data). Moreover, the method is fully compatible with intensive high-performance computer use. Moreover, the present work outlined an upper bound based on the maximum mean discrepancy (MMD) on numerical integration absolute error. Kernel herding and Bayesian quadrature both aim at finding the quadrature rule minimizing the MMD, and therefore the absolute integration error. The numerical experiments confirm that the MMD is an appropriate criterion since it leads to results being better or equivalent to quasi-Monte Carlo sampling. Finally, the proposed numerical benchmark relies on a Python package, called `otkerneldesign`, which implements the methods and allows anyone to reproduce the results.

The limits of the proposed method are reached when the input dimension of the problem increases, requiring a larger candidate set and therefore a larger covariance matrix. Moreover, the numerical experiments show that the method can be sensitive to the choice of the kernel and its tuning (although good practices can be derived). From a methodological viewpoint, further interpretation of the impact of the different kernels could be explored. Besides, extensions of kernel herding sampling for quantile estimation could be investigated, in a similar fashion as the work on randomized quasi-Monte Carlo for quantiles proposed by Kaplan et al. (2019). Kernel herding could also be used to quantize conditional distributions, using the so-called "conditional kernel mean embedding" concept reviewed by Klebanov et al. (2020). Finally, regarding the industrial use case, the next step should be to perform a reliability analysis by considering another group of random variables (related to the wind turbine) or to explore the possibilities offered by reliability-oriented sensitivity analysis in the context of kernel-based indices, as studied in Marrel and Chabridon (2021).

Chapter 4

Kernel-based surrogate model validation

Parts of this chapter are adapted from the following publication:

> 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.

4.1 Introduction

The development of methods to validate and certify the predictivity of supervised learning models is essential to the industry. Estimating the predictivity of these models can either be done by cross-validation or using a suitably selected test sample (as introduced in Section 1.8). Both in a given-data context (i.e., machine learning) or a simulated data context (i.e., computer experiment), guarantees on the validation procedure are increasingly asked. Certain risk-averse industries (e.g., nuclear) impose to establish these guarantees from independent test sets, i.e., datasets that have not been used either to train or to select the learning model (Borovicka et al., 2012; Xu and Goodacre, 2018; Iooss, 2021). Using the prediction residuals on this test set, an independent evaluation of the proposed learning model can be done, enabling the estimation of relevant performance metrics, such as the mean-squared error for regression problems, or the misclassification rate for classification problems.

The present chapter introduces methods to choose a "good" test set, either within a given dataset or within the input space of the model, as recently motivated in Iooss (2021); Joseph and Vakayil (2022). The construction of test sets is studied as an uncertainty propagation of the learning model's error, on which an average error may be estimated using the Bayesian quadrature methods introduced in Chapter 3 for mean estimation.

A first choice concerns the size of the test set. No optimal choice exists, and, when only a finite dataset is available, classical machine learning (ML) handbooks (Hastie et al., 2009; Goodfellow et al., 2016) provide different heuristics on how to split it, e.g., 80%/20% between the training and test samples, or 50%/25%/25% between the training, validation (used for model selection) and test samples. This point is not formally addressed in the following (see Xu and Goodacre, 2018 for a numerical study of this issue). A second issue concerns how the test sample is picked within the input space. The simplest, and most common way to build a test sample is to extract an independent Monte Carlo sample (Hastie et al., 2009). For small test sets, these randomly chosen points may fall too close to the training points or leave large areas of the input space unsampled, and a more constructive method to select points inside the input domain is therefore preferable. Similar concerns motivate the use of space-filling designs when choosing a small set of runs for computationally expensive computer experiments on which a model will be identified (Fang et al., 2006; Pronzato and Müller, 2012).

When the test set must be a subset of an initial dataset, the problem amounts to selecting a certain number of points within a finite collection of points. A review of classical methods for solving this issue is given in Borovicka et al. (2012). For example, the CADEX and DUPLEX algorithms (Kennard and Stone, 1969; Snee, 1977) can sequentially extract points from a database to include them in a test sample, using an inter-point distance criterion.

Several algorithms have also been proposed for the case where points need to be added to an already existing training sample. When the goal is to assess the quality of a model learned using a known training set, one may be tempted to locate the test points the furthest away from the training samples, such that, in some sense, the union of the training and test sets is space-filling. As this chapter shows, test sets built in this manner do enable a good assessment of the quality of models learned with the training set if the observed residuals are appropriately weighted. Moreover, the incremental augmentation of a design can be useful when the assessed model turns out to be of poor quality, or when an additional computational budget is available after a first study (Sheikholeslami and Razavi, 2017; Shang and Apley, 2020). Different empirical strategies have been proposed for incremental space-filling design (Iooss et al., 2010; Crombecq et al., 2011; Li et al., 2017), which basically entail the addition of new points in the zones poorly covered by the current design. Shang and Apley (2020) have recently proposed an improvement of the CADEX algorithm, called the "fully-sequential space-filling" (FSSF) design. Nogales Gómez et al. (2021) also proposed an improved version of such design enforcing boundary avoidance. Although they are developed for different purposes, nested space-filling designs (Qian et al., 2009) and sliced space-filling designs (Qian and Wu, 2009) can also be used to build sequential designs.

This work provides insights into these subjects in two main directions: *(i)* definition of a new predictivity criteria through an optimal weighting of the test points residuals, and *(ii)* use of test sets built by incremental space-filling algorithms, namely FSSF, support points Mak and Joseph (2018) and kernel herding Chen et al. (2010), the latter two algorithms being typically used to provide a representative sample of a desired theoretical or empirical distribution. Besides, this chapter presents a numerical benchmark analysis comparing the behavior of the three algorithms on a selected set of test cases and an industrial case.

This chapter is organized as follows. Section 4.2 defines the predictivity criterion considered and proposes different methods for its estimation. Section 4.3 presents the algorithms used for test-point selection. Our numerical results are presented in Sections 4.4 and 4.5: in Section 4.4 a test set is freely chosen within the entire input space, while in Section 4.5 an existing data set can be split into a training sample and a test set. Finally, Section 4.6 concludes and outlines some perspectives.

4.2 Predictivity assessment criteria for an ML model

In this section, a new criterion to assess the predictive performance of a model is proposed, enhancing a standard model quality metric by suitably weighting the errors observed on the test set. Let us denote by $\mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^d$ the space of the input variables $\mathbf{x} = (x_1, \dots, x_d)$ of the model. Then let $y(\mathbf{x}) \in \mathbb{R}$ (resp. $y(\mathbf{x}') \in \mathbb{R}$) be the observed output at point $\mathbf{x} \in \mathcal{D}_{\mathbf{x}}$ (resp. $\mathbf{x}' \in \mathcal{D}_{\mathbf{x}}$). Considering the training sample denoted by $(\mathbf{X}_m, \mathbf{y}_m)$, with $\mathbf{y}_m = [y(\mathbf{x}^{(1)}), \dots, y(\mathbf{x}^{(m)})]^\top$. The test sample is denoted by $(\mathbf{X}_n, \mathbf{y}_n) = (\mathbf{x}^{(i)}, y(\mathbf{x}^{(i)}))_{1\leq i\leq n}$. Remember that the intersection between these two samples is empty, $\mathbf{X}_m \cap \mathbf{X}_n = \varnothing$.

4.2.1 The predictivity coefficient

Let us denote by $\eta_m(\mathbf{x})$ the prediction at point $\mathbf{x}$ of a model learned using $(\mathbf{X}_m, \mathbf{y}_m)$ (Hastie et al., 2009; Rasmussen and Williams, 2006). A classical measure for assessing the predictive ability of η_m, in order to evaluate its validity, is the *predictivity coefficient*. Considering the probability measure π that weights how comparatively significant it is to accurately predict y over the different regions of $\mathcal{D}_{\mathbf{x}}$. For example, the input could be a random vector with a known distribution: in that case, this distribution would be a reasonable choice for π. The true (i.e., ideal) value of the predictivity is defined as the following normalization of the Integrated Square Error (ISE):

$$Q^2_\pi(\eta_m) = 1 - \frac{\mathrm{ISE}_\pi(\eta_m)}{\mathrm{Var}_\pi(g(\mathbf{X}))}, \tag{4.1}$$

where

$$\mathrm{ISE}_\pi(\eta_m) = \int_{\mathcal{D}_{\mathbf{x}}} [y(\mathbf{x}) - \eta_m(\mathbf{x})]^2 \, \mathrm{d}\pi(\mathbf{x}),$$
$$\mathrm{Var}_\pi(g(\mathbf{X})) = \int_{\mathcal{D}_{\mathbf{x}}} \left[y(\mathbf{x}) - \int_{\mathcal{D}_{\mathbf{x}}} y(\mathbf{x}') d\pi(\mathbf{x}') \right]^2 \mathrm{d}\pi(\mathbf{x}).$$

The ideal predictivity $Q^2_{\mathrm{ideal}}(\pi)$ is usually estimated by its empirical version calculated over the test sample $(\mathbf{X}_n, \mathbf{y}_n)$ (see Da Veiga et al., 2021, p. 32):

$$\widehat{Q}^2_n = 1 - \frac{\sum_{i=1}^n \left[y(\mathbf{x}^{(i)}) - \eta_m(\mathbf{x}^{(i)}) \right]^2}{\sum_{i=1}^n \left[y(\mathbf{x}^{(i)}) - \overline{y}_n \right]^2}, \tag{4.2}$$

where $\overline{y}_n = (1/n) \sum_{i=1}^n y(\mathbf{x}^{(i)})$ denotes the empirical mean of the observations in the test sample. Note that the calculation of $\widehat{Q}^2_n$ only requires access to the predictor $\eta_m(\cdot)$. To compute $\widehat{Q}^2_n$, one does not need to know the training set which was used to build $\eta_m(\cdot)$. This estimator $\widehat{Q}^2_n$ is the *coefficient of determination* (also called "Nash-Sutcliffe criterion" Nash and Sutcliffe, 1970), which is a standard notion in regression studies (Kleijnen and Sargent, 2000; Iooss et al., 2010). It compares the prediction errors obtained with the model η_m with those obtained when prediction equals the empirical mean of the observations. Thus, the closer $\widehat{Q}^2_n$ is to one, the more accurate the surrogate model is (for the test set considered). On the contrary, $\widehat{Q}^2_n$ close to zero (negative values are possible too) indicates poor prediction abilities, as there is little improvement compared to a prediction by the simple empirical mean of the observations. The next section shows how a suitable weighting of the residual on the training sample may be key to improving the estimation of $\widehat{Q}^2_n$.

4.2.2 Weighting the test sample

The simplest way to estimate the $\mathrm{ISE}_\pi(\mathbf{X}_m, \mathbf{y}_m)$ (present on the numerator of the predictivity coefficient) is by computing the arithmetic mean of the squared residuals evaluated on the test set $\mathbf{X}_n$. Writing the equivalent discrete measure $\xi_n = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}^{(i)}}$, with $\delta_\mathbf{x}$ the Dirac measure at $\mathbf{x}$, this estimate can be expressed as:

$$\mathrm{ISE}_{\xi_n}(\eta_m) = \frac{1}{n}\sum_{i=1}^{n}\left[y(\mathbf{x}^{(i)}) - \eta_m(\mathbf{x}^{(i)})\right]^2 .$$

When the points $\mathbf{x}^{(i)}$ of the test set $\mathbf{X}_n$ are distant from the points of the training set $\mathbf{X}_m$, the squared prediction errors $|y(\mathbf{x}^{(i)}) - \eta_m(\mathbf{x}^{(i)})|^2$ tend to represent the worst possible error situations, and $\mathrm{ISE}_{\xi_n}(\eta_m)$ tends to overestimate $\mathrm{ISE}_\pi(\eta_m)$. In this section, a statistical model for the prediction errors is assumed in order to be able to quantify this potential bias when sampling the residual process, enabling its subsequent correction.

Let us assume that the prediction error $\delta_m(\mathbf{x}) = y(\mathbf{x}) - \eta_m(\mathbf{x})$ is a realization of a Gaussian Process (GP) with mean $\widehat{\delta}_m(\mathbf{x})$ and covariance kernel $\sigma^2 K_{|m}$, $\delta_m(\mathbf{x}) \sim \mathrm{GP}(\widehat{\delta}_m(\mathbf{x}), \sigma^2 K_{|m})$

$$\begin{cases} \widehat{\delta}_m(\mathbf{x}) = \mathbf{k}_m^\top(\mathbf{x})\mathbf{K}_m^{-1}(\mathbf{y}_m - \boldsymbol{\eta}_m), \\ \sigma^2 K_{|m}(\mathbf{x}, \mathbf{x}') = \mathbb{E}[\delta_m(\mathbf{x})\delta_m(\mathbf{x}')] = \sigma^2\left[K(\mathbf{x}, \mathbf{x}') - \mathbf{k}_m^\top(\mathbf{x})\mathbf{K}_m^{-1}\mathbf{k}_m(\mathbf{x}')\right] . \end{cases} \tag{4.3}$$

Where $\boldsymbol{\eta}_m = [\eta_m(\mathbf{x}^{(1)}), \ldots, \eta_m(\mathbf{x}^{(m)})]^\top$, $\mathbf{k}_m(\mathbf{x})$ is the column vector $[K(\mathbf{x}, \mathbf{x}^{(1)}), \ldots, K(\mathbf{x}, \mathbf{x}^{(m)})]^\top$, and $\mathbf{K}_m$ is the $m \times m$ covariance matrix whose element (i, j) is given by $\{\mathbf{K}_m\}_{i,j} = K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, with K a positive definite kernel. Note that in the case of a learning model η_m which interpolates the observations $\mathbf{y}_m$, the errors observed at the learning points $\mathbf{X}_m$ equal zero, leading finally to the posterior $\mathrm{GP}(0, \sigma^2 K_{|m})$ for $\delta_m(\mathbf{x})$.

The prediction model error above allows us to study how well $\mathrm{ISE}_\pi(\eta_m)$ is estimated using a test set $\mathbf{X}_n$. The expected squared error when estimating $\mathrm{ISE}_\pi(\eta_m)$ by $\mathrm{ISE}_{\xi_n}(\eta_m)$, is defined as $\overline{\Delta}^2(\xi_n, \pi; \eta_m)$:

$$\begin{aligned} \overline{\Delta}^2(\xi_n, \pi; \eta_m) &= \mathbb{E}\left[\left(\mathrm{ISE}_{\xi_n}(\eta_m) - \mathrm{ISE}_\pi(\eta_m)\right)^2\right] \\ &= \mathbb{E}\left[\left(\int_{\mathcal{D}_\mathbf{x}} \delta_m^2(\mathbf{x})\, \mathrm{d}(\xi_n - \pi)(\mathbf{x})\right)^2\right] \\ &= \mathbb{E}\left[\int_{\mathcal{D}_\mathbf{x}^2} \delta_m^2(\mathbf{x})\delta_m^2(\mathbf{x}')\, \mathrm{d}(\xi_n - \pi)(\mathbf{x})\mathrm{d}(\xi_n - \pi)(\mathbf{x}')\right] . \end{aligned}$$

Tonelli's theorem gives

$$\overline{\Delta}^2(\xi_n, \pi; \eta_m) = \int_{\mathcal{D}_\mathbf{x}^2} \mathbb{E}[\delta_m^2(\mathbf{x})\delta_m^2(\mathbf{x}')]\, \mathrm{d}(\xi_n - \pi)(\mathbf{x})\mathrm{d}(\xi_n - \pi)(\mathbf{x}') .$$

Since $\mathbb{E}\left[U^2V^2\right] = 2\left(\mathbb{E}[UV]\right)^2 + \mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right]$ for any one-dimensional normal-centered random variables U and V. The expression can then be written as:

$$\overline{\Delta}^2(\xi_n, \pi; \eta_m) = \int_{\mathcal{D}_{\mathbf{x}}^2} 2\,\mathbb{E}[\delta_m(\mathbf{x})\delta_m(\mathbf{x}')]^2 + \mathbb{E}[\delta_m^2(\mathbf{x})]\,\mathbb{E}[\delta_m^2(\mathbf{x}')]\,\mathrm{d}(\xi_n - \pi)(\mathbf{x})\mathrm{d}(\xi_n - \pi)(\mathbf{x}') \tag{4.4a}$$

$$\overline{\Delta}^2(\xi_n, \pi; \eta_m) = \int_{\mathcal{D}_{\mathbf{x}}^2} \overline{K}_{|m}(\mathbf{x}, \mathbf{x}')\,\mathrm{d}(\xi_n - \pi)(\mathbf{x})\mathrm{d}(\xi_n - \pi)(\mathbf{x}') \tag{4.4b}$$

$$\overline{\Delta}^2(\xi_n, \pi; \eta_m) = \sigma^2\,\mathrm{MMD}^2_{\overline{K}_{|m}}(\xi_n, \pi). \tag{4.4c}$$

Interestingly, the last expression is equivalent to the maximum mean discrepancy (previously introduced in this manuscript and further defined in Appendix B) between π and ξ_n for a kernel $\overline{K}_{|m}(\cdot,\cdot)$. Note that σ^2 only appears as a multiplying factor in Eq. (4.4b), with the consequence that σ^2 does not impact the choice of a suitable ξ_n. The resulting kernel $\overline{K}_{|m}(\cdot,\cdot)$ is differently defined whether *(i)* the learning model $\eta_m(\mathbf{x})$ interpolates $\mathbf{y}_m$ or not *(ii)*:

$$\begin{cases} (i) \Rightarrow \overline{K}_{|m}(\mathbf{x}, \mathbf{x}') & = 2K^2_{|m}(\mathbf{x}, \mathbf{x}') + K_{|m}(\mathbf{x}, \mathbf{x})K_{|m}(\mathbf{x}', \mathbf{x}') \\ (ii) \Rightarrow \overline{K}_{|m}(\mathbf{x}, \mathbf{x}') & = 2\left[K_{|m}(\mathbf{x}, \mathbf{x}') + 2\,\widehat{\delta}_m(\mathbf{x})\widehat{\delta}_m(\mathbf{x}')\right]K_{|m}(\mathbf{x}, \mathbf{x}') \\ & \quad + \left[\widehat{\delta}^2_m(\mathbf{x}) + K_{|m}(\mathbf{x}, \mathbf{x})\right]\left[\widehat{\delta}^2_m(\mathbf{x}') + K_{|m}(\mathbf{x}', \mathbf{x}')\right]. \end{cases} \tag{4.5}$$

The main idea is to replace ξ_n, uniform on $\mathbf{X}_n$, by a nonuniform measure ζ_n supported on $\mathbf{X}_n$, $\zeta_n = \sum_{i=1}^n w_i \delta_{\mathbf{x}^{(i)}}$ with weights $\mathbf{w}_n = (w_1, \ldots, w_n)^\top$ chosen such that the estimation error $\overline{\Delta}^2(\xi_n, \pi; \eta_m)$, and thus $\mathrm{MMD}^2_{\overline{K}_{|m}}(\zeta_n, \pi)$, is minimized. The squared MMD for the kernel $\overline{K}_{|m}$ between π and the weighted measure ζ_n can be expressed as:

$$\mathrm{MMD}^2_{\overline{K}_{|m}}(\zeta_n, \pi) = \varepsilon_{\overline{K}_{|m},\pi} - 2\,\mathbf{w}_n^\top P_{\overline{K}_{|m},\pi}(\mathbf{X}_n) + \mathbf{w}_n^\top \overline{\mathbf{K}}_{|m}(\mathbf{X}_n)\mathbf{w}_n, \tag{4.6}$$

where $P_{\overline{K}_{|m},\pi}(\mathbf{X}_n)$ is the vector of potentials $\left[\int_{\mathcal{D}_{\mathbf{x}}} \overline{K}_{|m}(\mathbf{x}, \mathbf{x}^{(1)})\mathrm{d}\pi(\mathbf{x}), \ldots, \int_{\mathcal{D}_{\mathbf{x}}} \overline{K}_{|m}(\mathbf{x}, \mathbf{x}^{(n)})\mathrm{d}\pi(\mathbf{x})\right]^\top$, and $\overline{\mathbf{K}}_{|m}(\mathbf{X}_n)$ is the $n \times n$ covariance matrix such that $\{\overline{\mathbf{K}}_{|m}(\mathbf{X}_n)\}_{i,j} = \overline{K}_{|m}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, $\forall i, j = 1, \ldots, n$, and $\varepsilon_{\overline{K}_{|m},\pi} = \int_{\mathcal{D}_{\mathbf{x}}^2} \overline{K}_{|m}(\mathbf{x}, \mathbf{x}')\,\mathrm{d}\pi(\mathbf{x})\mathrm{d}\pi(\mathbf{x}')$. The squared MMD defined in Eq. (4.6) is minimized for the following optimal weights $\mathbf{w}_n^*$:

$$\mathbf{w}_n^* = \overline{\mathbf{K}}_{|m}^{-1}(\mathbf{X}_n)\mathbf{p}_{\overline{K}_{|m},\pi}(\mathbf{X}_n). \tag{4.7}$$

Therefore, an optimally weighted estimator of the predictivity coefficient supported on the test set $\mathbf{X}_n$ is defined as:

$$\widehat{Q}^2_{n*} = 1 - \frac{\sum_{i=1}^n w_i^* \left[y(\mathbf{x}^{(i)}) - \eta_m(\mathbf{x}^{(i)})\right]^2}{\frac{1}{n}\sum_{i=1}^n \left[y(\mathbf{x}^{(i)}) - \overline{y}_n\right]^2}, \tag{4.8}$$

with $\overline{y}_n = (1/n)\sum_{i=1}^{n} y(\mathbf{x}^{(m+i)})$. Notice that the weights w_i^* do not depend on the variance parameter σ^2 of the GP model. Moreover, this approach does not constrain the weights, which works better in our experience than the different constrained versions (e.g., non-negativity, summing to one) studied in Pronzato and Rendas (2023).

Remark 7. The optimal estimator $\widehat{Q}^2_{n*}$ focused on the weighting numerator of the coefficient of determination defined in Eq. (4.2). However, the variance estimator on the denominator can also be optimally weighted. Let us write an alternative version of $\widehat{Q}^2_{n*}$

$$\widehat{Q}'^2_n = 1 - \frac{\sum_{i=1}^{n}\left[y(\mathbf{x}^{(i)}) - \eta_m(\mathbf{x}^{(i)})\right]^2}{\sum_{i=1}^{n}\left[y(\mathbf{x}^{(i)}) - \overline{y}_m\right]^2}, \tag{4.9}$$

which compares the performance on the test set of η_m and $\overline{y}_m = (1/m)\sum_{i=1}^{m} y(\mathbf{x}^i)$. Using similar developments as previously it is possible to also apply a weighting procedure to the denominator of $\widehat{Q}'^2_n$, in order to make it resemble its integral version $V'_\pi(\mathbf{y}_m) = \int_{\mathcal{D}_\mathbf{x}}\left[y(\mathbf{x}) - \overline{y}_m\right]^2 \mathrm{d}\pi(\mathbf{x})$ (see Fekhari et al., 2023b).

4.3 Test-set construction

In the previous section, the test set was assumed as given, and a method was proposed to estimate the $\mathrm{ISE}_\pi(\eta_m)$ (with the learning model η_m, built on $\mathbf{X}_m$) by an optimally weighted sum of the residuals. The objective in this section is to construct a test set of size n, denoted by $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\right\} \subset \mathcal{D}_\mathbf{x}$. To evaluate the predictivity of a model learned on a training set $\mathbf{X}_m$, a strategy is to place the test points the furthest away from the training set, to obtain a space-filling design when gathering leaning and test set. The sampling methods used for building the test set should then be space-filling and incremental. The advantage of using an iterative construction is that it can be stopped as soon as the predictivity estimation is considered sufficiently accurate In case the conclusion is that model predictions are not reliable enough, the full design $\mathbf{X}_{m+n} = \mathbf{X}_m \cup \mathbf{X}_n$ and the associated observations $\mathbf{y}_{m+n}$ can be used to update the model. This updated model can then be tested at additional design points, elements of a new test set to be constructed. This section introduces different space-filling methods, later compared for test set construction.

4.3.1 Fully-Sequential Space-Filling design

The Fully-Sequential Space-Filling forward-reflected (FSSF-fr) algorithm (Shang and Apley, 2020) relies on the CADEX algorithm (Kennard and Stone, 1969) (also called the "coffee-house" method Müller, 2007). It constructs a sequence of nested designs in a bounded set $\mathcal{D}_\mathbf{x}$ by sequentially selecting a new point $\mathbf{x}$ as far away as possible from the $\mathbf{x}^{(i)}$ previously selected. New inserted points are selected within a set of candidates $\mathcal{S}$ which may coincide with $\mathcal{D}_\mathbf{x}$ or be a finite subset of $\mathcal{D}_\mathbf{x}$ (which simplifies the implementation, only this case is considered here).

The improvement of FSSF-fr when compared to CADEX is that new points are selected *at the same time* far from the previous design points as well as far from the boundary of $\mathcal{D}_{\mathbf{x}}$.

The algorithm is as follows:

1. Choose $\mathcal{S}$, a finite set of candidate points in $\mathcal{D}_{\mathbf{x}}$, with size $N \gg n$ in order to allow a fairly dense covering of $\mathcal{D}_{\mathbf{x}}$. When $\mathcal{D}_{\mathbf{x}} = [0,1]^d$, Shang and Apley (2020) recommends taking $\mathcal{S}$ equal to the first $N = 1\,000\,d + 2\,n$ points of a Sobol sequence in $\mathcal{D}_{\mathbf{x}}$.

2. Choose the first point $\mathbf{x}^{(1)}$ randomly in $\mathcal{S}$ and define $\mathbf{X}_1 = \{\mathbf{x}^{(1)}\}$.

3. At iteration i, with $\mathbf{X}_i = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(i)}\}$, select

$$\mathbf{x}^{(i+1)} \in \underset{\mathbf{x}\in\mathcal{S}\setminus\mathbf{X}_i}{\arg\max} \left[\min\left(\min_{j\in\{1,\ldots,i\}} \|\mathbf{x}-\mathbf{x}^{(j)}\|, \sqrt{2}\,d\,\mathrm{dist}(\mathbf{x}, R(\mathbf{x})) \right) \right], \tag{4.10}$$

   where $R(\mathbf{x})$ is the symetric of $\mathbf{x}$ with respect to its nearest boundary of $\mathcal{D}_{\mathbf{x}}$, and set $\mathbf{X}_{i+1} = \mathbf{X}_i \bigcup \mathbf{x}^{(i+1)}$.

4. Stop the algorithm when $\mathbf{X}_n$ has the required size.

The role of the reflected point $R(\mathbf{x})$ is to avoid selecting $\mathbf{x}^{(i+1)}$ too close to the boundary of $\mathcal{D}_{\mathbf{x}}$, which is a major problem with standard coffee-house, especially when $\mathcal{D}_{\mathbf{x}} = [0,1]^d$ with d large. While the standard coffee-house (greedy packing) algorithm simply uses $\mathbf{x}^{(i+1)} \in \arg\max_{\mathbf{x}\in\mathcal{S}\setminus\mathbf{X}_i} \min_{j\in\{1,\ldots,i\}} \|\mathbf{x}-\mathbf{x}^{(j)}\|$. The factor $\sqrt{2}d$ in Eq. (4.10) proposed in Shang and Apley (2020) sets a balance between distance to the design $\mathbf{X}_i$ and distance to the boundary of $\mathcal{D}_{\mathbf{x}}$. Another scaling factor, depending on the target design size n is proposed in Nogales Gómez et al. (2021).

FSSF-fr is entirely based on geometric considerations and implicitly assumes that the selected set of points should cover $\mathcal{D}_{\mathbf{x}}$ evenly.

However, in the context of uncertainty quantification, the distribution π of the model inputs is frequently not uniform. It is then desirable to select a test set representative of π. This can be achieved through the inverse transform of the CDF: FSSF-fr constructs $\mathbf{X}_n$ in the unit hypercube $[0,1]^d$, and an "isoprobabilistic" transform $T : [0,1]^d \to \mathcal{D}_{\mathbf{x}}$ is then applied to the points in $\mathbf{X}_i$, T being such that, if $\boldsymbol{U}$ is a random variable uniform on $[0,1]^d$, then $T(\boldsymbol{U})$ follows the target distribution π. The transformation can be applied to each input separately when π is the product of its marginals but is more complicated in other cases, see (Lemaire et al., 2009, Chap. 4). Note that FSSF-fr operates in the bounded set $[0,1]^d$ even if the support of π is unbounded. The other two algorithms presented in this section are able to directly choose points representative of a given distribution π and do not need to resort to such a transformation.

4.3.2 Support points

Support points (Mak and Joseph, 2018) are such that their associated empirical distribution ξ_n has minimum Maximum-Mean-Discrepancy (MMD) with respect to π for the energy-distance kernel of Székely and Rizzo (Székely and Rizzo, 2013),

$$K_E(\mathbf{x}, \mathbf{x}') = \frac{1}{2}\left(\|\mathbf{x}\| + \|\mathbf{x}'\| - \|\mathbf{x} - \mathbf{x}'\|\right). \quad (4.11)$$

The squared MMD between ξ_n and π for the distance kernel equals

$$\mathrm{MMD}^2_{K_E}(\xi_n, \pi) = \frac{2}{n}\sum_{i=1}^{n} \mathbb{E}\|\mathbf{x}^{(i)} - \zeta\| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| - \mathbb{E}\|\zeta - \zeta'\|, \quad (4.12)$$

where ζ and ζ' are independently distributed with π (see Sejdinovic et al., 2013). A key property of the energy-distance kernel is that it is characteristic (Sriperumbudur et al., 2010), meaning that for any couple of probability distributions π and ξ on $\mathcal{D}_\mathbf{x}$, $\mathrm{MMD}^2_{K_E}(\pi, \xi)$ equals zero if and only if $\pi = \xi$. This MMD then defines a norm in the space of probability distributions. Compared to more heuristic methods for solving quantization problems, support points benefit from the theoretical guarantees of MMD minimization in terms of convergence of ξ_n to π as $n \to \infty$.

As $\mathbb{E}\|\mathbf{x}^{(i)} - \zeta\|$ is not known explicitly, in practice π is replaced by its empirical version π_N for a given large-size sample $(\mathbf{x}'^{(k)})_{k=1\ldots N}$. The support points $\mathbf{X}_n^s$ are then given by

$$\mathbf{X}_n^s \in \underset{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}}{\arg\min} \left(\frac{2}{nN}\sum_{i=1}^{n}\sum_{k=1}^{N} \|\mathbf{x}^{(i)} - \mathbf{x}'^{(k)}\| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| \right). \quad (4.13)$$

The function to be minimized can be written as a difference of functions convex in $\mathbf{X}_n$, which yields a difference-of-convex program. In Mak and Joseph (2018), a majorization-minimization procedure, efficiently combined with resampling, is applied to the construction of large designs (up to $n = 10^4$) in high dimensional spaces (up to $d = 500$). The examples treated clearly show that support points are distributed in a way that matches π more closely than Monte Carlo and quasi-Monte Carlo samples.

The method was used to split a dataset into a training set and a test set in Joseph and Vakayil (2022), where the N points $\mathbf{X}_N$ in Eq. (4.13) are those from the dataset. Then $\mathbf{X}_n^s$ gives the test set and the other $N - n$ points are used for training. There is a serious additional difficulty though, as choosing $\mathbf{X}_n^s$ among the dataset corresponds to a difficult combinatorial optimization problem. A possible solution is to perform the optimization in a continuous domain $\mathcal{D}_\mathbf{x}$ and then choose $\mathbf{X}_n^s$ that corresponds to the closest points in $\mathbf{X}_N$ (for the Euclidean distance) to the continuous solution obtained (Joseph and Vakayil, 2022).

The direct determination of support points through Eq. (4.13) does not allow the construction of a nested sequence of test sets. One possibility would be to solve Eq. (4.13) sequentially, one point at a time, in a continuous domain, and then select the closest point within $\mathbf{X}_N$ as the one to be included in the test set. A different approach can be used, based on the greedy minimization of the MMD Eq. (4.12) for the candidate set $\mathcal{S} = \mathbf{X}_N$: at iteration i, the algorithm chooses

$$\mathbf{x}_{i+1}^{s} \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{S}} \left(\frac{1}{N} \sum_{k=1}^{N} \|\mathbf{x}-\mathbf{x}'^{(k)}\| - \frac{1}{i+1} \sum_{j=1}^{i} \|\mathbf{x}-\mathbf{x}^{(j)}\| \right). \tag{4.14}$$

The method requires the computation of the $N(N-1)/2$ distances $\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|$, $i, j = 1, \dots, N$, $i \neq j$, which hinders its applicability to large-scale problems (a test case with $N = 1\,000$ is presented in Section 4.5). Note that Joseph and Vakayil (2022) only studies the split of a given data set into a learning and test set while this chapter builds support points on the input space $\mathcal{D}_{\mathbf{x}}$

Greedy MMD minimization can be applied to other kernels than the distance kernel Eq. (4.11), see Teymur et al. (2021); Pronzato (2022). In the next section the closely related method of kernel herding is recalled (KH) (Chen et al., 2010), after its presentation in Chapter 3 of the present manuscript.

4.3.3 Kernel herding

As introduced in Chapter 3, Lacoste-Julien et al. (2015) proposed a linearization of the MMD minimization using the Frank-Wolfe algorithm. Let us define a positive definite kernel K on $\mathcal{D}_{\mathbf{x}} \times \mathcal{D}_{\mathbf{x}}$, and consider $\xi_i = (1/i) \sum_{j=1}^{i} \delta_{\mathbf{x}^{(i)}}$ as the empirical measure for $\mathbf{X}_i$. In the sequential and uniformly weighted case, this iteration i of kernel herding is expressed as a difference of potentials:

$$\mathbf{x}_{i+1} \quad \in \quad \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{S}} \left[P_{\xi_i}(\mathbf{x}) - P_{\pi}(\mathbf{x}) \right], \tag{4.15}$$

with $\mathcal{S} \subseteq \mathcal{D}_{\mathbf{x}}$ a given candidate set and $P_{\xi_i}(\mathbf{x}) = (1/i) \sum_{j=1}^{i} K(\mathbf{x}, \mathbf{x}^{(j)})$.

Once the targeted measure substituted by an empirical measure π_N based on a sample $(\mathbf{x}'^{(k)})_{k=1\dots N}$ then complete estimation becomes: $P_{\pi_N}(\mathbf{x}) = (1/N) \sum_{k=1}^{N} K(\mathbf{x}, \mathbf{x}'^{(k)})$, which gives

$$\mathbf{x}_{i+1} \quad \in \quad \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{S}} \left[\frac{1}{i} \sum_{j=1}^{i} K(\mathbf{x}, \mathbf{x}^{(j)}) - \frac{1}{N} \sum_{k=1}^{N} K(\mathbf{x}, \mathbf{x}'^{(k)}) \right].$$

When K is the energy-distance kernel Eq. (4.11) the greedy support points from Eq. (4.14) are recovered with a factor $1/i$ instead of $1/(i+1)$ in the second sum.

The candidate set $\mathcal{S}$ in Eq. (4.15) is arbitrary and can be chosen as in Subsection 4.3.1. A neat advantage of kernel herding over support points is that the potential $P_{\pi}(\mathbf{x})$ is sometimes explicitly available. When $\mathcal{S} = \mathbf{X}_N$, this avoids the need to calculate the $N(N-1)/2$ distances $\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|$ and thus allows application to very large sample sizes. This is the case in particular when $\mathcal{D}_{\mathbf{x}}$ is the cross product of one-dimensional sets $\mathcal{X}_{x_i}$, $\mathcal{D}_{\mathbf{x}} = \mathcal{X}_{x_1} \times \cdots \times \mathcal{X}_{x_d}$, π is the product of its marginals $\pi_{[i]}$ on the $\mathcal{X}_{x_i}$, K is the product of one-dimensional kernels $K_{[i]}$, and the one-dimensional integral in $P_{\pi_{[i]}}(x)$ is known explicitly for each $i \in \{1, \dots, d\}$. Indeed, for $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{D}_{\mathbf{x}}$, $P_{\pi}(\mathbf{x}) = \prod_{i=1}^{d} P_{\pi_{[i]}}(x_i)$ (see Pronzato and Zhigljavsky, 2020). When K

is the product of Matérn kernels with regularity parameter 5/2 and correlation lengths θ_i, $K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{d} K_{5/2,\theta_i}(x_i - x_i')$, with

$$K_{5/2,\theta}(x-x') = \left(1 + \frac{\sqrt{5}}{\theta}|x-x'| + \frac{5}{3\theta^2}(x-x')^2\right) \exp\left(-\frac{\sqrt{5}}{\theta}|x-x'|\right), \tag{4.16}$$

the one-dimensional potentials are given in Appendix B for $\pi_{[i]}$ uniform on $[0,1]$ or $\pi_{[i]}$ the standard normal $\S - N(0,1)$. When no observation is available, which is the common situation at the design stage, the correlation lengths have to be set to heuristic values. The values of the correlation lengths empirically show a significant influence over the design. A reasonable choice for $\mathcal{D}_{\mathbf{x}} = [0,1]^d$ is $\theta_i = n^{-1/d}$ for all i, with n the target number of design points (see Pronzato and Zhigljavsky, 2020).

4.3.4 Numerical illustration

As a first numerical illustration, the FSSF-fr (denoted FSSF in the following), support points and kernel herding algorithms were applied in the situation where a given initial design of size m has to be completed by a series of additional points $\mathbf{x}^{(m+1)}, \ldots, \mathbf{x}^{(m+n)}$. The objective is to obtain a full design $\mathbf{X}_{m+n}$ that is a good quantization of a given distribution π.

Figures 4.1 and 4.2 correspond to π uniform on $[0,1]^2$ and π the standard normal distribution $\mathcal{N}(\mathbf{0},\mathbf{I}_2)$, with $\mathbf{I}_2$ the 2-dimensional identity matrix, respectively. All methods are applied to the same candidate set $\mathcal{S}$.

The initial designs $\mathbf{X}_m$ are chosen in the class of space-filling designs, well suited to initialize sequential learning strategies (Santner et al., 2003). When π is uniform, the initial design is a maximin Latin hypercube design (introduced in Subsection 1.5.2) with $m = 10$ and the candidate set is given by the $N = 2^{12}$ first points $\mathbf{S}_N$ of a Sobol sequence in $[0,1]$. When π is normal, the inverse probability transform method is first applied to $\mathbf{S}_N$ and $\mathbf{X}_m$ (this does not raise any difficulty here as π is the product of its marginals). The candidate points $\mathcal{S}$ are marked in gray on Fig. 4.1 and Fig. 4.2 and the initial design is indicated by the red crosses. The index i of each added test point $\mathbf{x}^{(m+i)}$ is indicated (the font size decreases with i). In such a small dimension ($d = 2$), a visual appreciation gives the impression that the three methods have comparable performance. However, FSSF tends to choose points closer to the boundary of $\mathcal{S}$ than the other two, and the support points seem to sample more freely the holes of $\mathbf{X}_m$ than kernel herding, which seems to be closer to a space-filling continuation of the training set. In the next section, these designs are used for estimating the quality of the predictivity metric.

4.4 Numerical experiments I: construction of a training set and a test set

This section presents numerical results obtained on three different test cases, in dimension 2 (test cases 1 and 2) and 8 (test case 3), for which $y(\mathbf{x}) = f(\mathbf{x})$ with $f(\mathbf{x})$ has an easy to evaluate

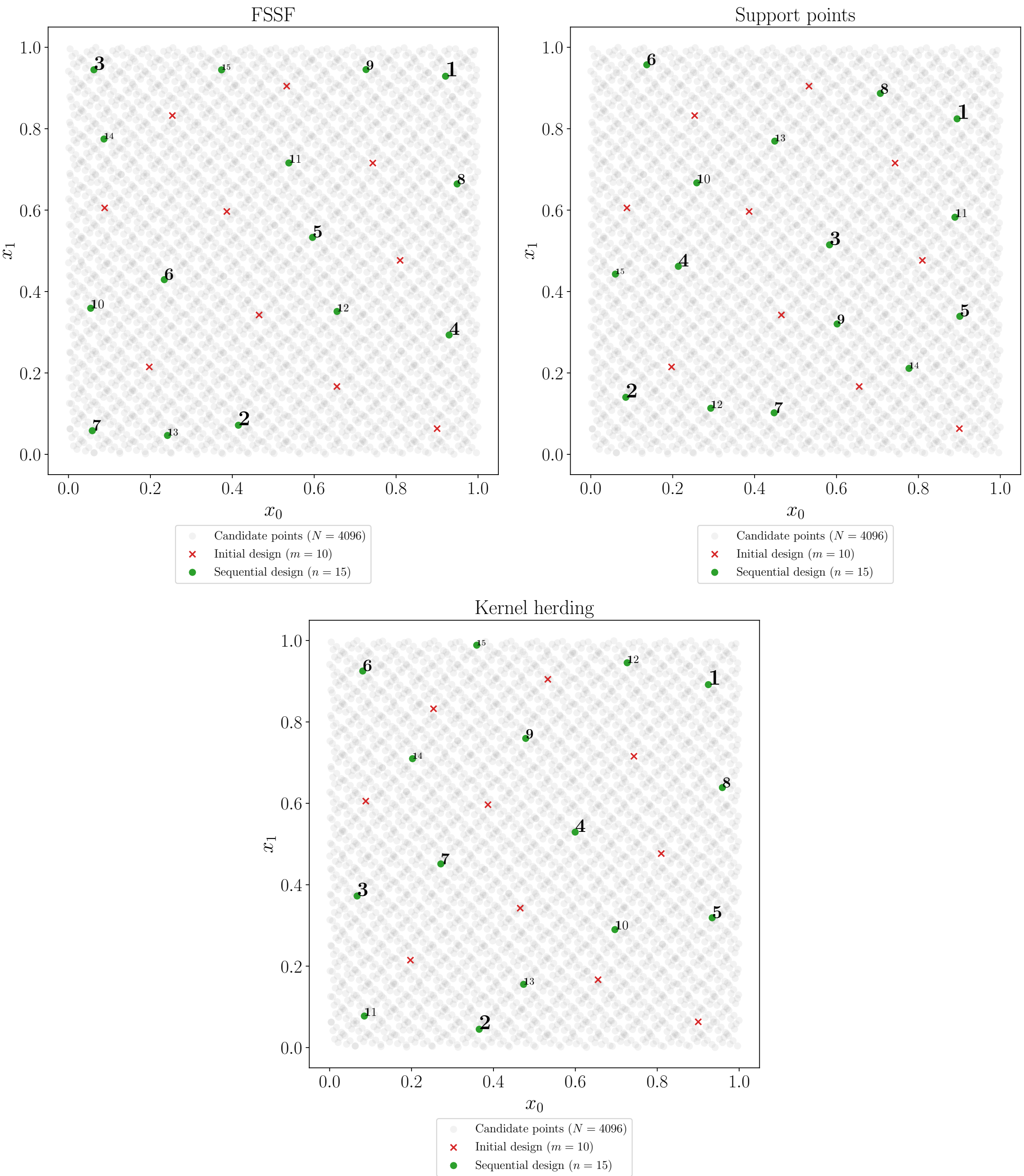


Figure 4.1 Additional points (ordered, green) complementing an initial design (red crosses), π is uniform on $[0, 1]$, the candidate points are in gray.

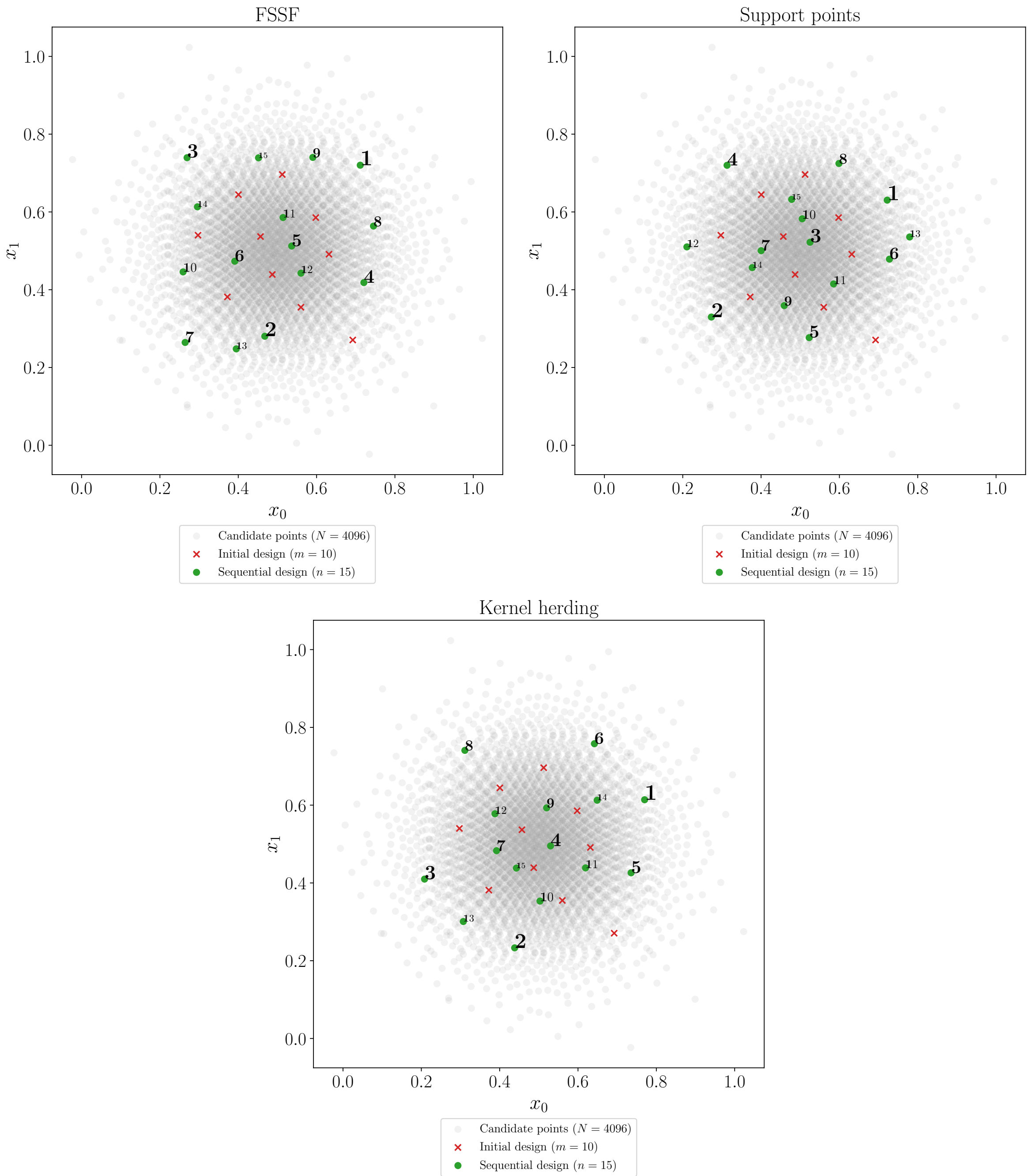


Figure 4.2 Additional points (ordered, green) complementing an initial design (red crosses), π normal, the candidate points are in gray.

analytical expression, see Subsection 4.4.1. This allows a good estimation of Q^2_π (see Eq. (4.1)) by a large Monte Carlo sample (with size $M = 10^6$), which will serve as a reference when assessing the performance of each of the other estimators.

The validation designs are built by FSSF, support points and kernel herding, presented in Subsections 4.3.1, 4.3.2, and 4.3.3, and the performances obtained are compared for each one, considering the uniform and the weighted estimator of Subsection 4.2.2.

4.4.1 Test cases

The training design $\mathbf{X}_m$ and the set $\mathcal{S}$ of potential test set points are as in Subsection 4.3.4. For test cases 1 and 3, π is the uniform measure on $\mathcal{D}_\mathbf{x} = [0,1]^d$, with $d = 2$ and $d = 8$, respectively; $\mathbf{X}_m$ is a maximin Latin hypercube design in $\mathcal{D}_\mathbf{x}$, and $\mathcal{S}$ corresponds to the first N points S_N of Sobol' sequence in $\mathcal{D}_\mathbf{x}$, complemented by the 2^d vertices. In the second test case, $d = 2$, π is the normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, and the sets $\mathbf{X}_m$ and S_N must be transformed as explained in Subsection 4.3.1. There are $N = 2^{14}$ candidate points for test cases 1 and 2 and $N = 2^{15}$ for test case 3 (this value is rather moderate for a problem in dimension 8, but using a larger N yields numerical difficulties for support points; see Subsection 4.3.2).

For each test case, a GP regression model is fitted to the m observations using ordinary Kriging (Rasmussen and Williams, 2006) (a GP model with constant mean), with an anisotropic Matérn kernel with regularity parameter $5/2$, and the correlation lengths θ_i are estimated by maximum likelihood via a truncated Newton algorithm. All calculations were done using the Python package `OpenTURNS`for uncertainty quantification (Baudin et al., 2017). The kernel used for kernel herding is different and corresponds to the tensor product of one-dimensional Matérn kernels Eq. (4.16), so that the potentials $P_\pi(\cdot)$ are known explicitly (see Appendix B); the correlations lengths are set to $\theta = 0.2$ in test cases 1 and 3 ($d = 2$) and to $\theta = 0.7$ in test case 3 ($d = 8$).

Assuming that a model is classified, in terms of the estimated value of its predictivity index Q^2 as "poor fitting" if $Q^2 \in [0.6, 0.8]$, "reasonably good fitting", when $Q^2 \in (0.8, 0.9]$, and "very good fitting" if $Q^2 > 0.9$, for each test case, three different sizes m of the training set are selected such that the corresponding models cover all three possible situations. For all test cases, the impact of the size n of the test set is studied in the range $n \in \{4, \ldots, 50\}$.

Test case 1. This test function is $f_1(\mathbf{x}) = h(2x_1 - 1, 2x_2 - 1)$, $(x_1, x_2) \in \mathcal{D}_\mathbf{x} = [0,1]^2$, with

$$h(u_1, u_2) = \frac{\exp(u_1)}{5} - \frac{u_2}{5} + \frac{u_2^6}{3} + 4u_2^4 - 4u_2^2 + \frac{7u_1^2}{10} + u_1^4 + \frac{3}{4u_1^2 + 4u_2^2 + 1}.$$

Color-coded 3d and contour plots of f_1 for $\mathbf{X} \in \mathcal{D}_\mathbf{x}$ are shown on the left panel of Fig. 4.3, showing that the function is rather smooth, even if its behavior along the boundaries of $\mathcal{D}_\mathbf{x}$, in particular close to the vertices, may present difficulties for some regression methods. The size of the training set for this function is $m \in \{5, 15, 30\}$.

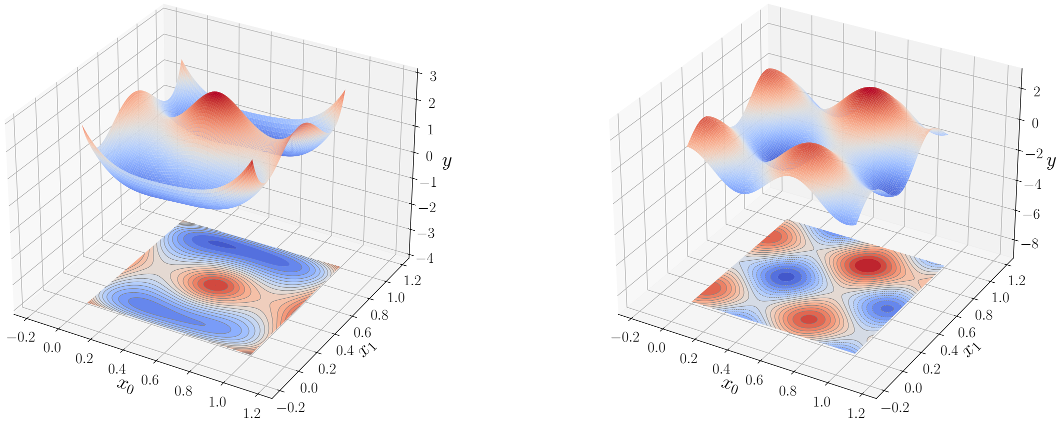


Figure 4.3 Left: $f_1(\mathbf{x})$ (test case 1); right: $f_2(\mathbf{x})$ (test case 2); $\mathbf{x} \in \mathcal{D}_{\mathbf{x}} = [0,1]^2$.

test case 2. The second test function, plotted in the right panel of Fig. 4.3 for $\mathbf{x} \in [0,1]^2$, is

$$f_2(\mathbf{x}) = \cos\left(5 + \frac{3}{2}x_1\right) + \sin\left(5 + \frac{3}{2}x_1\right) + \frac{1}{100}\left(5 + \frac{3}{2}x_1\right)\left(5 + \frac{3}{2}x_2\right).$$

Training set sizes for this test case are $m \in \{8, 15, 30\}$.

test case 3. The third function is the so-called "gSobol" function, defined over $\mathcal{D}_{\mathbf{x}} = [0,1]^8$ by

$$f_3(\mathbf{x}) = \prod_{i=1}^{8} \frac{|4x_i - 2| + a_i}{1 + a_i}, \qquad a_i = i^2 .$$

This parametric function is very versatile as both the dimension of its input space and the coefficients a_i can be freely chosen. The sensitivity to input variables is determined by the a_i: the larger a_i is, the less f is sensitive to x_i. Larger training sets are considered for this test case: $m \in \{15, 30, 100\}$.

4.4.2 Benchmark results and analysis

The numerical results obtained in this section are presented in Figures 4.4, 4.5, and 4.6. Each figure corresponds to one of the test cases and gathers three sub-figures, corresponding to test sets with sizes m yielding poor (left), reasonably good (right) or very good (bottom) fittings.

The baseline value of Q^2_{MC}, calculated with 10^6 Monte Carlo points, is indicated by the black diamonds (the black horizontal lines). Assuming that the error of Q^2_{MC} is much smaller than the errors of all other estimators, this section compares the ability of the methods to approximate Q^2_{MC}. For each sequence of nested test sets ($n \in \{4, \ldots, 50\}$), the observed values of $\widehat{Q}^2_n$ (equation Eq. (4.2)) and Q^2_{n*} (equation Eq. (4.8)), are plotted as the solid and dashed lines, respectively.

The figures also show the value Q^2_{LOO} obtained by Leave-One-Out (LOO) cross-validation, which is indicated at the left of each figure by a red diamond (values smaller than 0.25 are not shown). Note that, contrarily to the other methods considered, for LOO the test set is not disjoint from the training set, and thus the method does not satisfy the conditions set in the Introduction. As the complete model-fitting procedure is repeated for each training sample of size $m-1$, including the maximum-likelihood estimation of the correlation lengths of the Matérn kernel, the closed-form expressions of Dubrule (1983) cannot be used, making the computations rather intensive. The three figures show, and as expected, that the Q^2_{LOO} tends to under-estimate Q^2_{ideal}: by construction of the training set, LOO cross-validation relies on model predictions at points $\mathbf{x}^{(i)}$ far from the other $m-1$ design points used to build the model, and thus tends to systematically overestimate the prediction error at $\mathbf{x}^{(i)}$. The underestimation of Q^2_{ideal} can be particularly severe when m is small, the training set is then necessarily sparse; see Fig. 4.4 where $Q^2_{LOO} < 0.3$ for $m = 5$ and 15.

Let us first concentrate on the non-weighted estimators (solid curves). The two MMD-based constructions, support points (in orange) and kernel herding (in blue), generally produce better validation designs than FSSF (green curves), leading to values of $\widehat{Q}^2_n$ that approach Q^2_{ideal} quicker as n increases. This is particularly noticeable for “good” and “very good” models (central and rightmost panels of all three figures). This supports the idea that test sets should complement the training set $\mathbf{X}_m$ by populating the holes it leaves in $\mathcal{D}_{\mathbf{x}}$ while at the same time being able to mimic the target distribution π, this second objective being more difficult to achieve for FSSF than for the MMD-based constructions.

A comparison of the two MMD-based estimators reveals that support points tend to under-estimate ISE, leading to an over-confident assessment of the model predictivity, while kernel herding displays the expected behavior, with a negative bias that decreases with n. The reason for the positive bias of estimates based on support points designs is not fully understood, but may be linked to the fact that support points tend to place validation points at “mid-range” from the designs (and not at the furthest points like FSSF or kernel herding), see central and rightmost panels in Figure 4.1, and thus residuals at these points are themselves already better representatives of the local average errors.

Let us consider now the impact of the GP-based weighting of the residuals when estimating Q^2 (by Q^2_{n*}), which is related to the relative training-set/validation-set geometry (the manner in which the two designs are entangled in ambient space). The improvement resulting from applying residual weighting is apparent on all panels of the three figures, the dashed curves lying closer to Q^2_{ideal} than their solid counterparts; see in particular kernel herding (blue curve) in Fig. 4.4 and FSSF (green curve) in Fig. 4.5. Unexpectedly, the estimators based on support points seem to be rather insensitive to residual weighting, the dashed and solid orange curves being most of the time close to each other (and in any case, much closer than the green and blue ones). While the reason for this behavior deserves a deeper study, the fact that the support point designs – see Figure 4.1 – sample in a better manner the range of possible training-to-validation

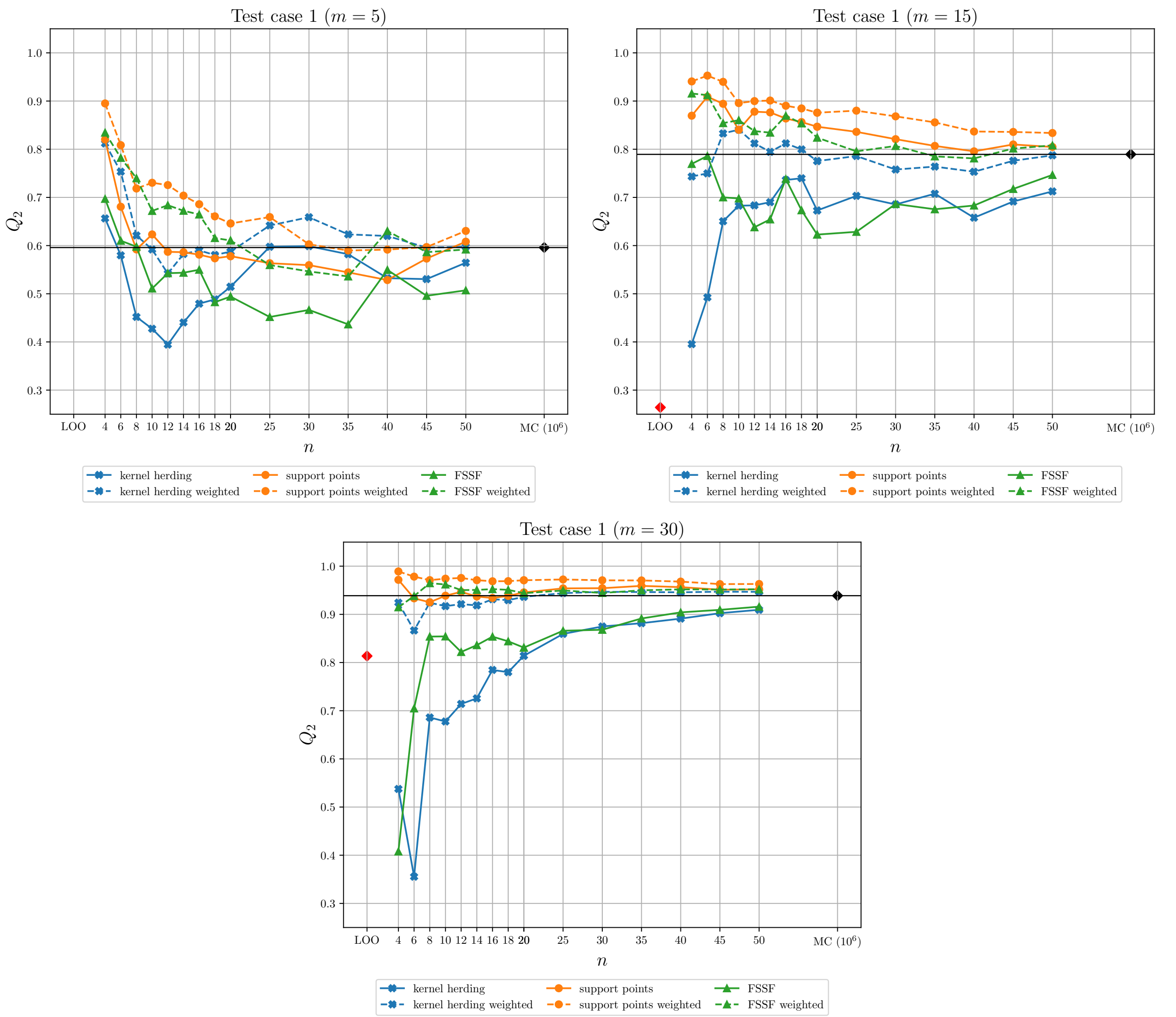


Figure 4.4 test case 1: predictivity assessment of a poor (left), good (right) and very good (bottom) model with kernel herding, support points and FSSF test sets.

distances, being in some sense less space-filling than both FSSF and kernel herding, is again a plausible explanation for this weaker sensitivity to residual weighting.

Consider now a comparison of the behavior across test cases. Setting aside the strikingly singular situation of test case 2, for which kernel herding displays a pathological (bad) behavior for the "very good" model, and all methods present an overall good behavior, the details of the tested function do not seem to play an important role concerning the relative merits of the estimators and validation designs.

Let us finally observe how the methods behave for models of distinct quality (m leading to poor, good or very good models), comparing the three panels in each figure. On the left panels, m is too small for the model η_m to be accurate, and all methods and test-set sizes are able to detect this. For models of practical interest (good and very good), the test sets generated with support points and kernel herding allow a reasonably accurate estimation of Q^2 with a few points. Note, incidentally, that except for test case 2 (where the interplay with a non-uniform measure π complicates the analysis), it is in general easier to estimate the quality of the very

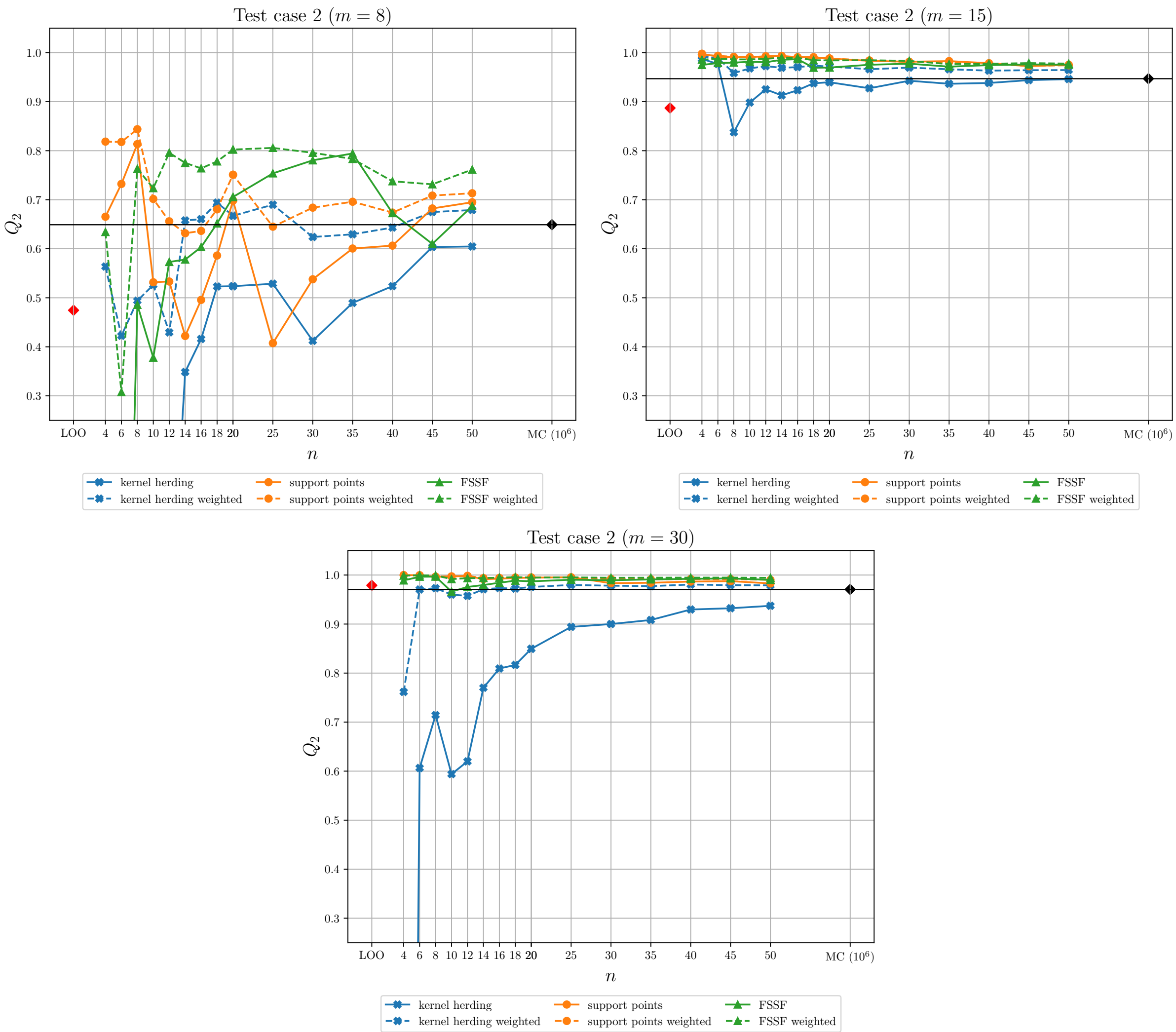


Figure 4.5 test case 2: predictivity assessment of a poor (left), good (right) and very good (bottom) model with kernel herding, support points and FSSF test sets.

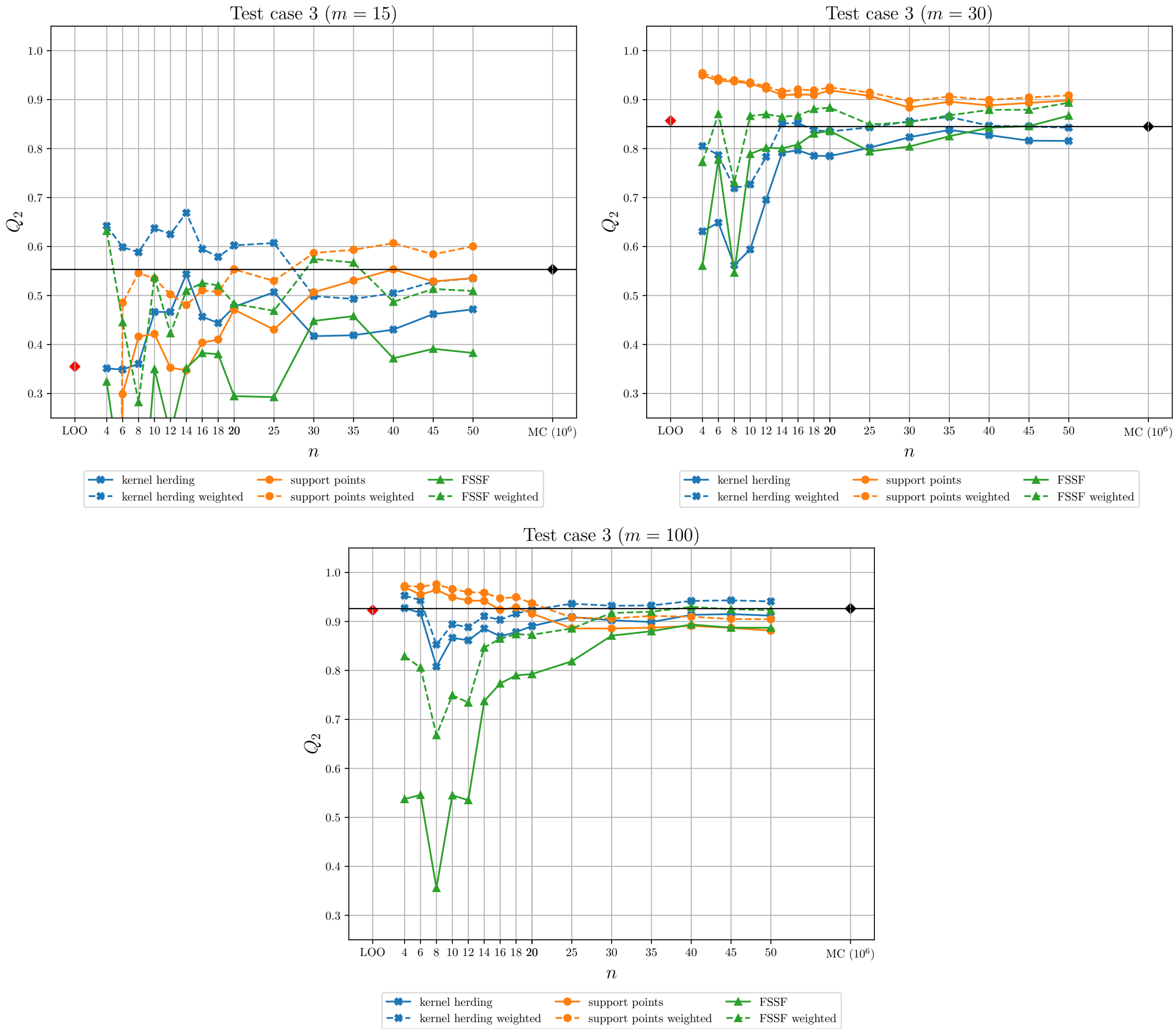


Figure 4.6 test case 3: predictivity assessment of a poor (left), good (right) and very good (bottom) model with kernel herding, support points and FSSF test sets.

good model (right-most panel) than that of the good model (central panel), indicating that the expected complexity (the entropy) of the residual process should be a key factor determining how large the validation set must be. In particular, it may be that larger values of m allow for smaller values of n.

4.5 Numerical experiments II: splitting a dataset into a training set and a test set

This section illustrates the performance of the different designs and estimators considered in this chapter when applied in the context of an industrial application, to split a given dataset of size N into training and test sets, with m and n points respectively, $m + n = N$. In contrast with Joseph and Vakayil (2022), the observations $y(\mathbf{x}^{(i)})$, $i = 1, \ldots, N$, are not used in the splitting mechanism, meaning that it can be performed before the observations are collected and that there cannot be any selection bias related to observations (indeed, the use of observation values in an MMD-based splitting criterion may favor the allocation of the most different observations to different sets, training versus validation).

An ML model is fitted to the training data, and the data collected on the test set are used to assess the predictivity of the model. The influence of the ratio $r_n = n/N = 1 - m/N$ on the quality assessment is investigated. Random Cross-Validation (RCV) is also considered, where n points are chosen at random among the N points of the dataset: for each n, there are $\binom{N}{n}$ possible choices, and $R = 1\,000$ designs were randomly selected among them. A model is fitted on the m complementary points ($m = N - n$), which yields an empirical distribution of Q^2 values for each ratio n/N considered.

4.5.1 Industrial test case CATHARE

The test case corresponds to the computer code CATHARE2 (for "Code Avancé de ThermoHydraulique pour les Accidents de Réacteurs à Eau"), which models the thermal-hydraulic behavior inside nuclear pressurized water reactors (Geffraye et al., 2011). The studied scenario simulates a hypothetical large-break loss of primary coolant accident for which the output of interest is the peak cladding temperature (De Crécy et al., 2008; Iooss et al., 2010). The complexity of this application lies in the large run-time of the computer model (of the order of twenty minutes) and in the high dimension of the input space: the model involves 53 input parameters z_i, corresponding mostly to constants of physical laws, but also coding initial conditions, material properties and geometrical modeling. The z_i were independently sampled according to normal or log-normal distributions. These characteristics make this test case challenging in terms of construction of a surrogate model and validation of its predictivity.

In the following, an existing Monte Carlo sample $\mathbf{Z}_N$ of $N = 1\,000$ points in $\mathbb{R}^{53}$ is used, that corresponds to 53 independent random input configurations (see Iooss et al., 2010 for details). The output of the CATHARE2 code at these N points is also available. To reduce the dimensionality of this dataset, a sensitivity analysis (Da Veiga et al., 2021) screens the inputs that do not impact the output significantly. This dimension-reduction step relies on the Hilbert-Schmidt Independence Criterion (HSIC), which is known as a powerful tool to perform input screening from a single sample of inputs and output values without reference to any specific ML regression model (Da Veiga, 2015). HSIC-based statistical tests and their associated p-values are used to identify (with a 5%-threshold) inputs on which the output is significantly dependent (and therefore, also those of little influence). They were successfully applied to similar datasets from thermal-hydraulic applications in Marrel and Chabridon (2021); Marrel et al. (2022). The screened dataset only includes 10 influential inputs, over which the candidate set $\mathbf{X}_N$ used for the construction of the test-set $\mathbf{X}_n$ (and therefore of the complementary training set $\mathbf{X}_{N-n}$) is defined. The marginal distributions are shown as histograms along the axes of the plots.

To include RCV in the methods to be compared, many (here, $R = 1\,000$) different models η_m must be constructed for each considered design size m. Since Gaussian Process regression proved to be too expensive for this purpose, the comparatively cheaper Partial Least Squares (PLS) method (Wold et al., 2001) is used. For each given training set, the model obtained is a sum of monomials in the 10 input variables. Note that models constructed with different training sets may involve different monomials and have different numbers of monomial terms.

4.5.2 Benchmark results and analysis

Fig. 4.7 compares various ways of extracting an n-point test set from an N-point dataset to estimate model predictivity, for different splitting ratios $n/N \in \{0.1, 0.15, 0.2, \ldots, 0.9\}$.

Consider RCV first. For each value of $r_n = n/N$, the empirical distribution of Q^2_{RCV} obtained from $R = 10^3$ random splittings of $\mathbf{X}_N$ into $\mathbf{X}_m \cup \mathbf{X}_n$ is summarized by a boxplot. Depending on r_n, three behaviors are roughly distinguished. For $0.1 \leq r_n \lesssim 0.3$ the distribution is bi-modal, with the lower mode corresponding to unlucky test-set selections leading to poor performance evaluations. When $0.3 \lesssim n/N \lesssim 0.7$, the distribution looks uni-modal, revealing a more stable performance evaluation. Note that this is (partly) in line with the recommendations discussed in Section 4.1. For $r_n \gtrsim 0.7$, the variance of the distribution increases with r_n: many unlucky training sets lead to poor models. Note that the median of the empirical distribution slowly decreases as r_n increases, which is consistent with the intuition that the model predictivity should decrease when the size of the training set decreases.

For completeness, the red diamond represented on the left of Fig. 4.7 the value of Q^2_{LOO} computed by LOO cross-validation. In principle, being computed using the entire dataset, this value should establish an upper bound on the quality of models computed with smaller training sets. This is indeed the case for small training sets (rightmost values in the figure), for which the predictivity estimated by LOO is above the majority of the predictivity indexes calculated. But at the same time, LOO cross-validation tends to overestimate the errors, which explains the higher predictivity estimated by some other methods when $m = N - n$ is large enough.

Compare now the behavior of the two MMD-based algorithms of Section 4.3, $\widehat{Q}_n^2$ (unweighted) and Q_{n*}^2 (weighted) are plotted using solid and dashed lines, respectively, for both kernel herding (in blue) and support points (in orange). FSSF test sets are not considered, as the application of an iso-probabilistic transformation imposes knowledge of the input distribution, which is not known for this example. Compare first the unweighted versions of the two MMD-based estimators. For small values of the ratio r_n, $0.1 \lesssim r_n \lesssim 0.45$, the relative behavior of support points and kernel herding coincides with what was observed in the previous section, support points (solid orange line) estimating a better performance than kernel herding (solid blue line), which, moreover, is close to the median of the empirical distribution of Q_{RCV}^2. However, for $r_n \geq 0.5$, the dominance is reversed, support points estimating a worse performance than kernel herding.

As r_n increases up to $r_n \lesssim 0.7$ the solid orange and blue curves crossover, and it is now $\widehat{Q}_n^2$ for kernel herding that approximates the RCV empirical median, while the value obtained with support points underestimates the predictivity index. Also, note that for (irrealistic) very large values of r_n both support points and kernel herding estimate lower Q^2 values, which are smaller than the median of the RCV estimates.

Let us now focus on the effect of residual weighting, i.e., in estimators Q_{n*}^2 which use the weights computed by the method of Subsection 4.2.2, shown in dashed lines in Figure 4.7. First, note that while kernel herding weighting leads, as in the previous section, to higher estimates of the predictivity (compare solid and dashed blue lines), this is not the case for support points (solid and dashed orange curves), which, for small split ratios, produces smaller estimates when weighting is introduced. In the large r_n region, the behavior is consistent with what was previously presented, weighting inducing an increase of the estimated predictivity. It is remarkable – and rather surprising – that Q_{n*}^2 for support points (the dashed orange line) does not present the discontinuity of the uncorrected curve.

The sum $\sum_{i=1}^n w_i^*$ of the optimal weights of support points and kernel herding Eq. (4.7) is shown in Fig. 4.8 (orange and blue curves, respectively). The slow increase with n/N of the sum of kernel-herding weights (blue line) is consistent with the increase of the volume of the input region around each validation point when the size of the training set decreases. The behavior of the sum of weights is more difficult to interpret for support points (orange line) but is consistent with the behavior of Q_{n*}^2 on Fig. 4.7. Note that the energy-distance kernel Eq. (4.11) used for support points cannot be used for the weighting method of Subsection 4.2.2 as K_E is not positive definite but only conditionally positive definite. A full understanding of the observed curves would require a deeper analysis of the geometric characteristics of the designs generated by the two MMD methods, in particular of their interleaving with the training designs, which is not compatible with the space constraints of this manuscript.

While a number of unanswered points remain, in particular how deeply the behaviors observed may be affected by the poor predictivity resulting from the chosen PLS modeling methodology, the example presented in this section shows that the construction of test sets via MMD minimization and estimation of the predictivity index using the weighted estimator Q_{n*}^2

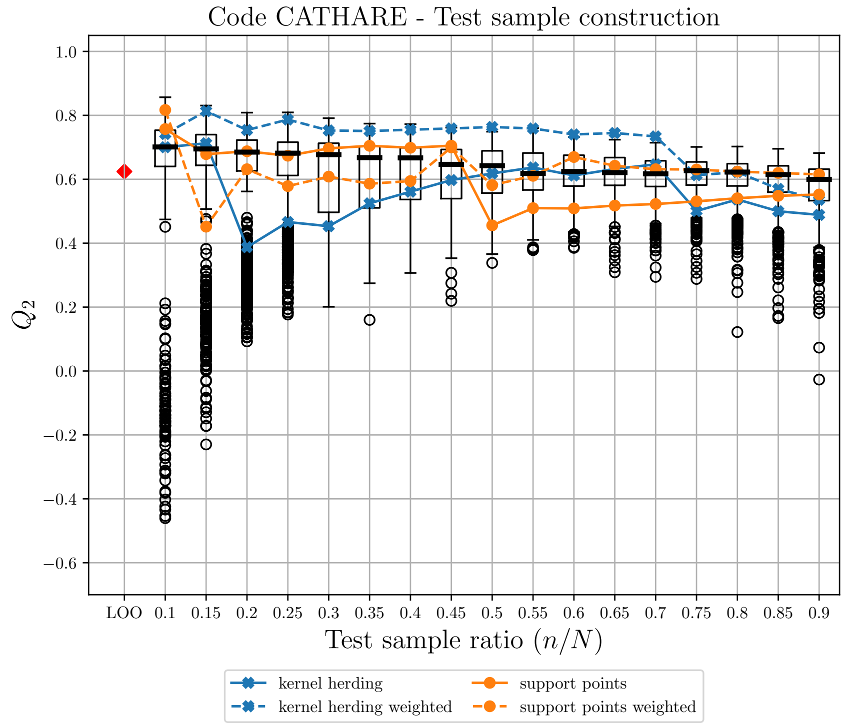


Figure 4.7 test case CATHARE: estimated Q^2. The box plots are for random cross-validation, and the red diamond (left) is for Q^2_{LOO}.

is promising as an efficient alternative to RCV: at a much lower computational cost, it builds performance estimates based on independent data the model developers may not have access to. Moreover, kernel herding proved, in the examples studied in this manuscript, to be a more reliable option for designing the test set, exhibiting a behavior that is consistent with what is expected, and very good estimation quality when the residuals over the design points are appropriately weighted.

4.6 Conclusion

Our study shows that ideas and tools from the design of experiment framework can be transposed to the problem of test-set selection. This chapter explored approaches based on support points, kernel herding and FSSF, considering the incremental construction of a test set (*i*) either as a particular space-filling design problem, where design points should populate the holes left in the design space by the training set, or (*i*) from the point of view of partitioning a given dataset into a training set and a test set.

A numerical benchmark has been performed for a panel of test cases of different dimensions and complexity. Additionally to the usual predictivity coefficient, a new weighted metric (see Pronzato and Rendas, 2023) has been proposed and shown to improve the assessment of the predictivity of a given model for a given test set.

This weighting procedure appears very efficient for interpolators, like Gaussian process regression models, as it corrects the bias when the points in the test set used to predict the

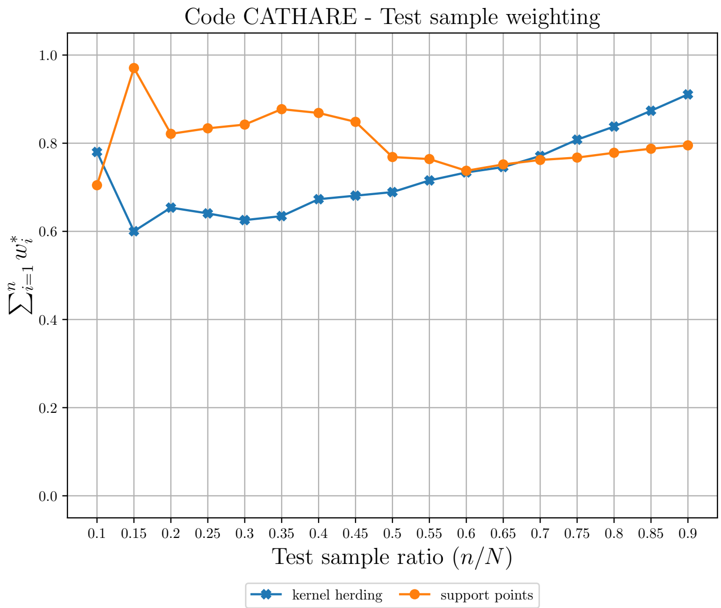


Figure 4.8 test case CATHARE: sum of the weights Eq. (4.7).

errors are far from the training points. For the first three test cases (Section 4.4), pairing one iterative design method with the weight-corrected estimator of the predictivity coefficient Q^2 shows promising results as the estimated Q^2 characteristic is close to the true one even for test-sets of moderate size.

Weighting can also be applied to models that do not interpolate the training data. For the industrial test case of Section 4.5, the true Q^2 value is unknown, but the weight-corrected estimation Q^2_{n*} of Q^2 is close to the value estimated by Leave-One-Out cross-validation and to the median of the empirical distribution of Q^2 values obtained by random k-fold cross-validation. At the same time, estimation by Q^2_{n*} involves a much smaller computational cost than cross-validation methods and uses a dataset fully independent of the one used to construct the model.

To each of the design methods considered to select a test set a downside can be attached. FSSF requires knowledge of the input distribution to be able to apply an iso-probabilistic transformation if necessary; it tends to select many points along the boundary of the candidate set considered. Support points require the computation of the $N(N-1)/2$ distances between all pairs of candidate points, which implies significant memory requirements for large N; the energy-distance kernel on which the method relies cannot be used for the weighting procedure. Finally, the efficient implementation of kernel herding relies on analytical expressions for the potentials P_π, see Appendices A and B, which are available for particular distributions (like the uniform and the normal) and kernels (like Matérn) only. The great freedom in the choice of the kernel K gives a lot of flexibility, but at the same time implies that some non-trivial decisions have to be made; also, the internal parameters of K, such as its correlation lengths, must be specified. Future work should go beyond empirical rules of thumb and study the influence of these choices.

Numerical tests were only computed with independent inputs. Kernel herding and support points are both well suited for probability measures not being equal to the product of their marginals, which is a frequent case in real datasets. Note only incremental constructions were considered, as they allow to stop the validation procedure as soon as the estimation of the model predictivity is deemed sufficiently accurate, but it is also possible to select several points at once, using support points (Mak and Joseph, 2018), or MMD minimization in general (Teymur et al., 2021).

Further developments around this work could be as follows. Firstly, the incremental construction of a test set could be coupled with the definition of an appropriate stopping rule, in order to decide when it is necessary to continue improving the model (possibly by supplementing the initial design with the test set, which seems well suited to this). The $\mathrm{MMD}_{\overline{K}_{|m}}(\zeta_n^*, \pi)$ of Subsection 4.2.2 could play an important role in the derivation of such a rule. Secondly, the approach presented gives equal importance to all the d inputs. However, it seems that inputs with a negligible influence on the output should receive less attention when selecting a test set. A preliminary screening step that identifies the significant inputs would allow the test-set selection algorithm to be applied to these variables only. For example, when a $\mathbf{X}_N \subset \mathbb{R}^d$ dataset is to be partitioned into $\mathbf{X}_m \cup \mathbf{X}_n$, one could use only $d' < d$ components to define the partition, but still use all d components to build the model and estimate its (weighted) Q^2. Note, however, that this would imply a slight violation of the conditions mentioned in the introduction, as it renders the test set dependent on the function observations.

Finally, in some cases, the probability measure π is known up to a normalizing constant. The use of a Stein kernel then makes the potential $P_{K,\pi}$ identically zero (Chen et al., 2018b), which would facilitate the application of kernel herding. Also, more complex problems involve functional inputs, like temporal signals, images, or categorical variables; the application of the methods presented to kernels specifically designed for such situations raises challenging issues.

Part III:

Contributions to rare event estimation

La résignation est un suicide quotidien.

H. Balzac

Chapter 5

Rare event estimation using Bernstein adaptive nonparametric sampling

Parts of this chapter are adapted from the following publication:

❝ 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 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP14).*

☞ This work was rewarded by the "CERRA[1] Student Recognition Award" at the ICASP14 conference.

5.1 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 1.6, the performance of a system is typically modeled by a deterministic scalar function, denoted by $g : \mathcal{D}_{\mathbf{x}} \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}$, and referred to as the *limit-state function* (LSF). A critical threshold on the system's output, denoted by $y_{\mathrm{th}} \in \mathbb{R}$, enables to define the *failure domain* as $\mathcal{F} = \{\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \,|\, g(\mathbf{x}) \leq y_{\mathrm{th}}\}$. Considering a probabilistic framework, the uncertain inputs are modeled by a continuous random vector $\mathbf{X} \in \mathcal{D}_{\mathbf{x}}$ distributed according to a PDF $f_{\mathbf{X}}$. 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* (Rockafellar and Royset, 2015), denoted by p_{f}, representing the probability of the system crossing the threshold y_{th}:

$$p_{\mathrm{f}} = \mathbb{P}(g(\mathbf{X}) \leq y_{\mathrm{th}}) = \int_{\mathcal{D}_{\mathbf{x}}} \mathbb{1}_{\{\mathcal{F}\}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \,\mathrm{d}\mathbf{x}. \tag{5.1}$$

The usual convention presented in Section 1.6 allows to modify the LSF in order to set the threshold to zero. Moreover, as presented previously, the main methods for rare event estimation can be divided into two groups (Morio and Balesdent, 2015): *(i)* first the geometric approaches, such as the first-/second-order reliability method (FORM/SORM) whose aim is to approximate the LSF by a first-/second-order Taylor expansion at the most probable failure point; *(ii)* second the simulation-based techniques such as the crude Monte Carlo method and many other variance reduction techniques. Unfortunately, FORM/SORM methods do not provide much 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 nonadaptive and adaptive versions of the importance sampling (Rubinstein and Kroese, 2008) and splitting techniques (Cérou et al., 2012) such as subset simulation (SS) (Au and Beck, 2001).

In SS, the idea is to write the rare event probability 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 a

[1] CERRA stands for "international Civil Engineering Risk and Reliability Association".

large panel of dedicated algorithms (Papaioannou et al., 2015). However, MCMC algorithms are known to be highly tunable algorithms and to produce non-i.i.d. samples, which consequently, cannot always be used for direct unbiased statistical estimation (e.g., for p_f but also for sensitivity indices).

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 Zhang, 1996; Morio, 2011). The main drawback of cross-entropy adaptive importance sampling (CE-AIS) is its limited flexibility, which makes it perform badly in the case of multimodal failure domains. Meanwhile, nonparametric adaptive importance sampling (NAIS) inherits its drawbacks from the multivariate KDE (i.e., significant performance drop for medium to high dimension). The limitations associated with these two classes of IS have been addressed by several authors (Kurtz and Song, 2013; Papaioannou et al., 2016; Geyer et al., 2019; Uribe et al., 2021).

The present work proposes a new rare event estimation method, “Bernstein adaptive nonparametric conditional sampling” (BANCS), adopting the same adaptive importance sampling structure as CE-AIS or NAIS while using a different mechanism to fit conditional distributions. This algorithm decomposes the fit of the intermediary conditional distributions into two steps: a fit of their marginals by univariate KDE, and a fit of their copula with the *empirical Bernstein copula* (abbreviated EBC and introduced in Section 2.2). Compared to direct multivariate KDE in NAIS, this decomposition should simplify the inference in medium to high dimension. Additionally, unlike SS, the proposed method generates i.i.d. samples of the intermediary conditional distributions. 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). To do so, the present chapter introduces kernel-based reliability-oriented sensitivity indices and their direct estimation as a post-processing of the BANCS algorithm.

In this chapter, Section 5.2 will introduce the BANCS algorithm for rare event estimation. Then, Section 5.3 will apply this method to three toy-cases and analyze the results with respect to NAIS and SS performances. Section 5.4 will present a kernel-based reliability-oriented sensitivity index and illustrate its estimation on samples generated by the BANCS algorithm. Then, the last section present some conclusions and research perspectives.

Remark 8. BANCS was initially proposed in Fekhari et al. (2023a), using the failure probability estimator from the SS algorithm (see Eq. (1.62)). Switching to an adaptive importance sampling estimator (see e.g., Eq. (1.60)) notably improved the method.

5.2 Bernstein adaptive nonparametric conditional sampling (BANCS)

The BANCS algorithm uses the same structure as other adaptive importance sampling methods (e.g., NAIS or CE-AIS introduced in Section 1.6) while employing a different approach to fit the intermediate conditional distributions. The proposed algorithm is described in Algorithm 1 at iteration k, after estimating the intermediary p_0-quantile $\widehat{q}_{[k]}^{p_0}$, a nonparametric model is fitted on the set $\mathbf{A}_{[k+1]}$ of all samples leading to values below $\widehat{q}_{[k]}^{p_0}$ (also called “elite set”). This inference is done by coupling a set of marginals fitted by univariate KDE, with a copula fitted using the EBC. The generation of the next i.i.d. N-sized sample $\mathbf{X}_{[k+1],N}$ from the conditional distributions is straightforward and does not require any MCMC sampling like in SS. Note that the BANCS method does not require any iso-probabilistic transform either.

Algorithm 1 Bernstein adaptive nonparametric conditional sampling (BANCS).

▷ ***Inputs:***

$f_{\mathbf{X}}$: joint PDF of the inputs

$g(\cdot)$: LSF

$y_{\text{th}} \in \mathbb{R}$: threshold defining the failure event

N: number of samples per iteration

$m \in \mathbb{N}$: parameter of the EBC fitting

$p_0 \in]0,1[$: empirical quantile order (rarity parameter)

▷ ***Algorithm:***

Set $k = 0$ and $\widehat{f}_{[0]} = f_{\mathbf{X}}$

Sample $\mathbf{X}_{[0],N} = \left\{\mathbf{x}_{[0]}^{(i)}\right\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \widehat{f}_{[0]}$

Evaluate $Y_{[0],N} = \left\{g(\mathbf{x}_{[0]}^{(i)})\right\}_{i=1}^{N}$

Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $Y_{[0],N}$

while $\widehat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**

  Compute IS weights $\left\{w_{[k]}^{(i)}\right\}_{i=1}^{N} = \left\{\frac{f_{\mathbf{X}}(\mathbf{x}_{[k]}^{(i)})}{\widehat{f}_{[k]}(\mathbf{x}_{[k]}^{(i)})}\right\}_{j=1}^{N}$

  Build a weighted elite set $\mathbf{A}_{[k+1]} = \sum_{l=1}^{k}\sum_{i=1}^{N} w_{[k]}^{(j)}\, \mathbb{1}_{\left\{g(\mathbf{x}_{[l]}^{(i)}) \leq \widehat{q}_{[k]}\right\}}\left(\mathbf{x}_{[k]}^{(i)}\right)$

  Fit marginals of the set $\mathbf{A}_{[k+1]}$ by KDE $\left\{\widehat{F}_j\right\}_{j=1}^{d}$

  Fit the copula of the set $\mathbf{A}_{[k+1]}$ by EBC $B_{\mathbf{m}}(C_n)$

  Build a CDF $\widehat{F}_{[k+1]}(\mathbf{x}) = B_m(C_n)\left(\widehat{F}_1(x_1), \dots, \widehat{F}_d(x_d)\right)$

  Sample $\mathbf{X}_{[k+1],N} = \left\{\mathbf{x}_{[k+1]}^{(i)}\right\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \widehat{f}_{[k+1]}$

  Evaluate $Y_{[k+1],N} = \left\{g(\mathbf{x}_{[k+1]}^{(i)})\right\}_{i=1}^{N}$

  Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $Y_{[k+1],N}$

  Set $k = k + 1$

end while

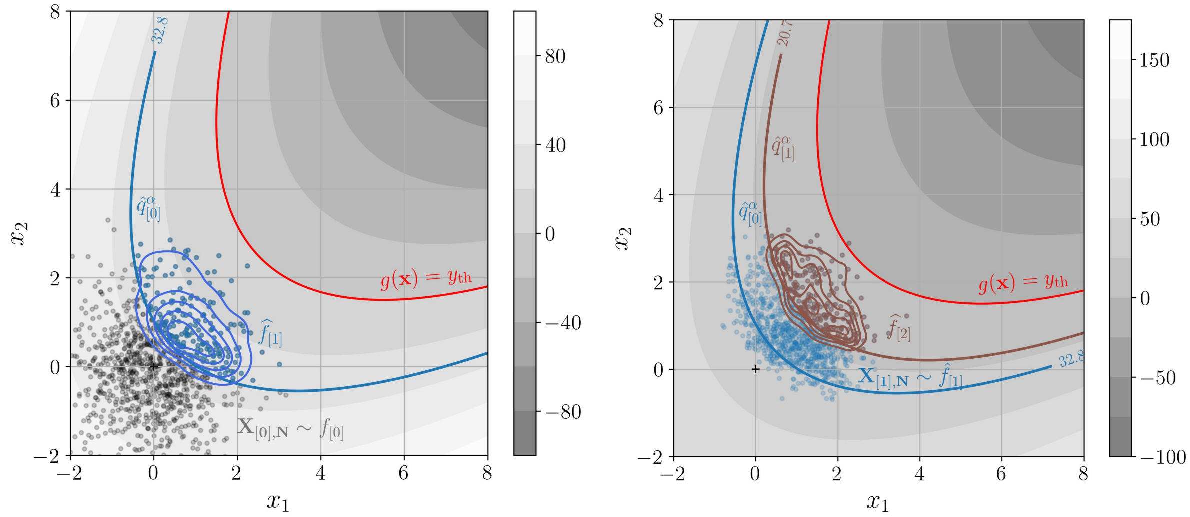
Set $k_\# = k$

Estimate $\widehat{p}_{\text{f}}^{\text{BANCS}} = \frac{1}{N}\sum_{i=1}^{N} w_{[k_\#]}^{(i)}\, \mathbb{1}_{\left\{g(\mathbf{x}_{[k_\#]}^{(i)}) \leq y_{\text{th}}\right\}}\left(\mathbf{x}_{[k_\#]}^{(i)}\right)$

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

▷ ***Outputs:***

$\widehat{p}_{\text{f}}^{\text{BANCS}}$, estimate of p_{f}



(a) Iteration $k = 0$.

(b) Iteration $k = 1$.

Figure 5.1 BANCS algorithm applied to toy-case #1: illustration of conditional sampling and nonparametric fit at the first and second iterations.

Nonparametric inference requires to tune the scaling parameters for KDE and the polynomial order m (considered equal for all the input dimensions) for EBC. In the current implementation of BANCS, the KDE is tuned using Silverman's rule[2] (Silverman, 1981) while the EBC is tuned to minimize an asymptotic mean integrated squared error (AMISE). As discussed in Chapter 2, for a dataset with size n and dimension d, the AMISE tuning for the EBC was defined by Sancetta and Satchell (2004) as:

$$m_{\text{AIMSE}} = 1 + N^{2/(d+4)} \tag{5.2}$$

In our experience, EBC tuning in Eq. (5.2) works well for the BANCS algorithm and will be systematically used in the following. This tuning sets a rather low polynomial order to the EBC, which avoids overfitting issues. 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 estimation should suit the AMISE tuning.

Ultimately, the estimator of the probability from Eq. (5.1) is written as a simple IS estimator on the last conditional distribution with PDF $\widehat{f}_{[k_\#]}$:

$$\widehat{p_{\mathrm{f}}}^{\text{BANCS}} = \frac{1}{N} \sum_{i=1}^{N} \frac{f_{\mathbf{X}}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)}{\widehat{f}_{[k_\#]}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)} \mathbb{1}_{\left\{g(\mathbf{x}^{(i)}_{[k_\#]}) \leq y_{\text{th}}\right\}}\left(\mathbf{x}^{(i)}_{[k_\#]}\right), \quad \left\{\mathbf{x}^{(i)}_{[k_\#]}\right\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \widehat{f}_{[k_\#]}. \tag{5.3}$$

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

$$\widehat{\mathrm{Var}}\left(\widehat{p_{\mathrm{f}}}^{\text{BANCS}}\right) = \frac{1}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{f_{\mathbf{X}}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)}{\widehat{f}_{[k_\#]}\left(\mathbf{x}^{(i)}_{[k_\#]}\right)}\right)^2 \mathbb{1}_{\left\{g(\mathbf{x}^{(i)}_{[k_\#]}) \leq y_{\text{th}}\right\}}\left(\mathbf{x}^{(i)}_{[k_\#]}\right) - \left(\widehat{p_{\mathrm{f}}}^{\text{BANCS}}\right)^2\right] \tag{5.4}$$

[2]In the multivariate case, Silverman's rule is defined as follows: $\eta_{\text{Silv.}} = \left(\frac{1}{N}\frac{4}{d+2}\right)^{1/(d+4)}$.

Fig. 5.1 illustrates the nonparametric fit and conditional sampling in the BANCS method on a two-dimensional reliability problem (later introduced as "toy-case #1", see Subsection 5.3.1). At the iteration $k = 0$, the conditional distribution fitted (with PDF represented by the blue isolines) slightly crosses the quantile border $\widehat{q}_{[0]}^{p_0}$ (in blue). Ideally, the conditional distribution fitted, with PDF denoted by $\widehat{f}_{[1]}$, should be bounded by this border. At the second iteration, the PDF of the conditional distribution denoted by $\widehat{f}_{[2]}$, is represented by the brown isolines. Unlike the inference of $\widehat{f}_{[1]}$, the fit of $\widehat{f}_{[2]}$ is realized on the samples from $\mathbf{X}_{[1],N} \sim \widehat{f}_{[1]}$ and $\mathbf{X}_{[0],N} \sim \widehat{f}_{[0]}$ which crossed the quantile $\widehat{q}_{[1]}^{p_0}$. Indeed, the samples $\mathbf{X}_{[0],N} \sim \widehat{f}_{[0]}$ are weighted according to an IS procedure (see the lines 14 and 15 from Algorithm 1). Including the samples from the previous iterations refines the goodness-of-fit near the quantile border and significantly improves BANCS performances.

5.3 Numerical experiments

In the present section, the performances of BANCS algorithm are compared with the ones from SS and NAIS algorithms. The efficiency of SS depends on the choice and tuning of the MCMC algorithm (Papaioannou et al., 2015). The proposed work uses the `OpenTURNS` implementation of the SS[3] (integrating a component-wise Metropolis-Hastings algorithm), and the `OpenTURNS` implementation of NAIS[4]. An implementation of the BANCS method and the following numerical experiments are available in a public Git repository[5].

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 the usual SS implementation. The following intermediate sample sizes $N \in \{300, 500, 700, 1000, 2000, 5000, 10000\}$ are simulated. Let us recall that the EBC tuning is set to minimize the AMISE, such that $m = 1 + N^{\frac{2}{d+4}}$. Finally, 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 $\widehat{\delta} = \frac{\sigma_{\widehat{p}_f}}{\mu_{\widehat{p}_f}}$. Since the SS estimator only provide an approximated value of its dispersion, an empirical estimation on repetitions is the most neutral way to compare the variability of several methods.

5.3.1 Analytical toy-cases

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

Toy-case #1: Parabolic reliability problem. This reliability problem, is defined by the function $g_1 : \mathbb{R}^2 \to \mathbb{R}$:

$$g_1(\mathbf{x}) = (x_1 - x_2)^2 - 8\,(x_1 + x_2 - 5)\,, \tag{5.5}$$

[3]SS: https://openturns.github.io/openturns/latest/user_manual/_generated/openturns.SubsetSampling.html
[4]NAIS: https://openturns.github.io/openturns/latest/user_manual/_generated/openturns.NAIS.html
[5]BANCS: 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: $p_{f,1} = \mathbb{P}(g_1(\mathbf{X}) \leq 0) = 1.31 \times 10^{-4}$.

Toy-case #2: Four-branch reliability problem. This reliability problem (originally proposed by Waarts (2000)), is defined by the following function $g_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}$:

$$g_2(\mathbf{x}) = \min \begin{pmatrix} 3 + 0.1\,(x_1 - x_2)^2 - \frac{(x_1+x_2)}{\sqrt{2}} \\ 3 + 0.1\,(x_1 - x_2)^2 + \frac{(x_1+x_2)}{\sqrt{2}} \\ (x_1 - x_2) + \frac{7}{\sqrt{2}} \\ (x_2 - x_1) + \frac{7}{\sqrt{2}} \end{pmatrix}, \tag{5.6}$$

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. This reliability problem (inspired by Lemaître et al., 2015), is defined by the following function $g_3 : \mathbb{R}^3 \longrightarrow \mathbb{R}$:

$$g_3(\mathbf{x}) = \sin(x_1) + 7\,\sin(x_2)^2 + \frac{x_3^4\,\sin(x_1)}{10} - 10.5. \tag{5.7}$$

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. This reliability problem (proposed by Yun et al., 2018), is defined by the following function $g_4 : \mathbb{R}^7 \longrightarrow \mathbb{R}$:

$$g_4(\mathbf{x}) = 15.59 \times 10^4 - \frac{x_1 x_3^2}{2x_3^2}\,\frac{x_2^4 - 4x_5 x_6 x_7^2 + x_4(x_6 + 4x_5 + 2x_6 x_7)}{x_4 x_5 (x_4 + x_6 + 2x_6 x_7)}, \tag{5.8}$$

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

Toy-case #5: Medium-dimensional nonlinear oscillator problem. This reliability problem (initially introduced by De Stefano and Der Kiureghian, 1990), is defined by the following function $g_5 : \mathbb{R}^8 \longrightarrow \mathbb{R}$:

$$g_5(\mathbf{x}) = F_s - 3k_s \sqrt{\frac{\pi S_0}{4\zeta_s \omega_s^3}\,\frac{\zeta_a \zeta_s}{\zeta_p \zeta_s (4\zeta_a^2 + \theta^2) + \gamma \zeta_a^2}\,\frac{\omega_p\,(\zeta_p \omega_p^3 + \zeta_s \omega_s^3)}{4\zeta_a \omega_a^4}}, \tag{5.9}$$

where $\mathbf{x} = \left(m_p, m_s, k_p, k_s, \zeta_p, \zeta_s, F_s, S_0\right)$, $\omega_p = \sqrt{k_p/m_p}$, $\omega_s = \sqrt{k_s/m_s}$, $\omega_a = (\omega_p + \omega_s)/2$, $\zeta_a = (\zeta_p + \zeta_s)/2$, $\gamma = m_s/m_p$, $\theta = (\omega_p - \omega_s)/\omega_a$. The distribution of the input random vector $\mathbf{X}$, is defined as a product of marginals following the distributions given in Table 5.1. The reliability problem consists in evaluating: $p_{f,5} = \mathbb{P}(g_5(\mathbf{X}) \leq 0) = 3.78 \times 10^{-7}$. A representation of this high-dimensional function is plotted in Fig. 5.3, with two-dimensional cross-cuts of the LSF passing through FORM's design point P^*. This visualization (inspired by Dubourg, 2011; Bourinet, 2018; Chabridon, 2018) allows to perceive the nonlinearity of the reliability problem near the design point. Note that the color scale associated to the output values is log-transformed to increase the color gradient at the border of the failure domain.

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

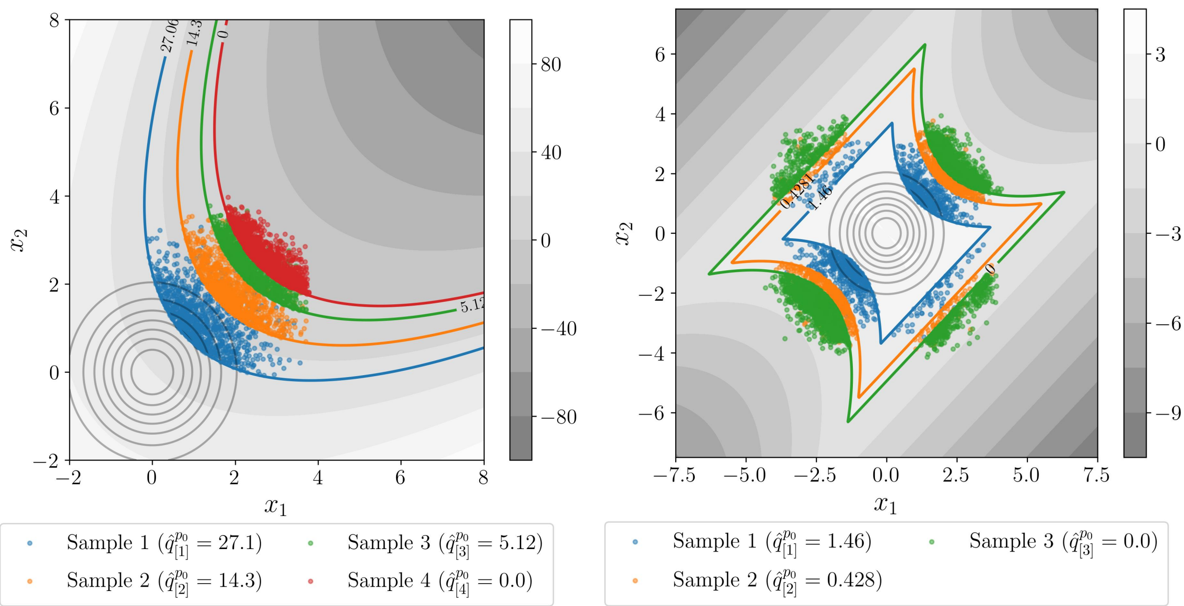
Table 5.1 Input probabilistic model for toy-case #5.

The problems introduced in the present section will be used in a numerical benchmark of BANCS (toy-cases #2, #4, and #4) and to study the ROSA as a post-processing of BANCS (toy-cases #3 and #5).

5.3.2 Benchmark results and analysis

In the present section, the BANCS algorithm is applied to three 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. 5.1, all BANCS iterations from this case are illustrated in Fig. 5.2 (a). This figure shows the intermediate quantiles $\left\{\widehat{q}_{[k]}^{p_0}\right\}_{k=1}^{k_\#}$ which are estimated over conditional samples of size $N = 10^4$. The failure samples corresponding to each elite set are represented in the same color as their p_0-quantile border. Fig. 5.2 (b) provides the same kind of illustration for BANCS applied to test-case #2, a well-known system reliability problem. On these two-dimensional cases, one can visualize the empirical quantiles driving the conditional distributions towards the failure domain(s), and the ability of the nonparametric inference to capture multi-modal patterns.

After this first illustration of BANCS, a more extensive benchmark is proposed. To present a fair assessment of the estimators' dispersion, each experiment is independently repeated 100 times. Then, a boostrap procedure on this set is performed to compute a confidence interval of the mean failure probability. Empirical variances and coefficients of variation (COV) associated with the probability estimators are also computed on the sets 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 as explained in Papaioannou et al., 2015). Fig. 5.4 summarizes the results of this benchmark comparing BANCS with SS and NAIS w.r.t. various sample sizes $N \in \{300, 500, 700, 1000, 2000, 5000, 10000\}$. On the left side, the estimates of failure probabilities repeated 100 times are displayed by their average over the repetitions (full lines) and their Bootsrap confidence intervals. The reference value for each problem is also represented by a horizontal black line. On the right side, an estimate of the coefficient of variation provides a normalized information on the dispersion of the estimators. Note that all the estimates are



(a) Toy-case #1. (b) Toy-case #2.

Figure 5.2 Illustration of BANCS iterations for the two-dimensional reliability problems in test cases #1 and #2 (taking $N = 10^4$ and $p_0 = 0.1$). Only the samples exceeding the intermediary thresholds are represented.

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 $d = 7$), and test-case #5 (nonlinear oscillator problem with $d = 8$). BANCS consistently shows promising results on all three cases. The variance of its estimator is always smaller or equivalent to the SS variance. 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 is therefore removed. BANCS does not encounter the same difficulty as it separately fits the copula and the marginals (which can be easily truncated). On this rare and complex case (notice the nonlinearity in the cross-cuts displayed in Fig. 5.3), BANCS is less accurate than SS for small-sized (i.e., $N \leq 10^3$), but becomes equivalent to SS for larger sample sizes. In general, BANCS shows equivalent or better performances than SS and NAIS (on these first toy cases), while providing i.i.d. sampling and offering a level of flexibility due to the nonparametric inference.

5.4 Reliability-oriented sensitivity analysis

As introduced in Section 1.7, GSA can be viewed as a tool to assess the global impact of the input variability on the output variable of interest. When estimating quantities of interest related to the output distribution's tail (typically, risk measures such as quantiles or failure probabilities), the analysis should be dedicated to the subdomain of interest (e.g., the failure domain). In other

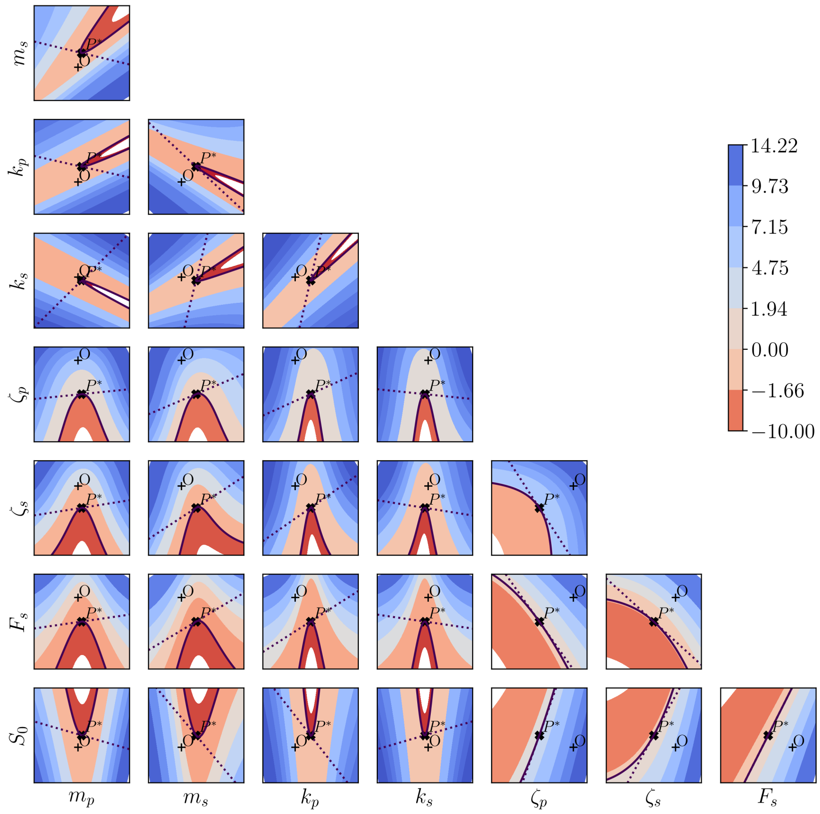
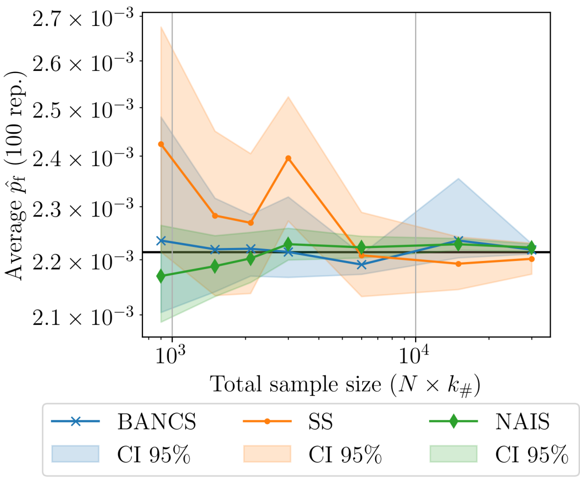
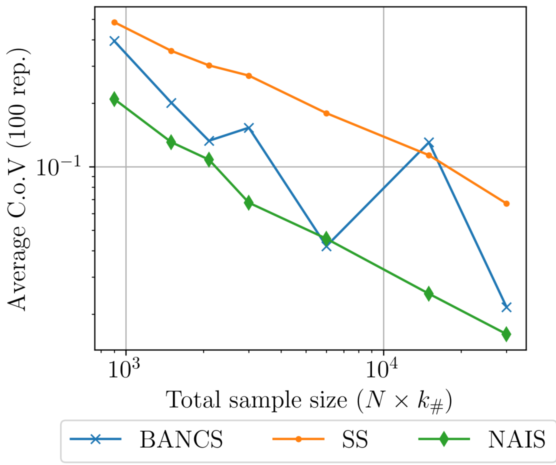


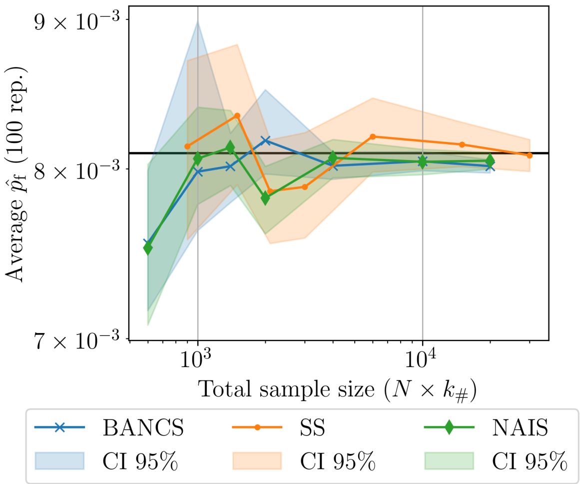
Figure 5.3 Cross-cut visualization of the limit-state function in test-case #5. FORM's most-probable failure point P^* is given by the black cross. The LSF (full line) delimits the safe domain (in blue) and the failure domain (in red). FORM approximation around P^* (represented by the dashed lines).



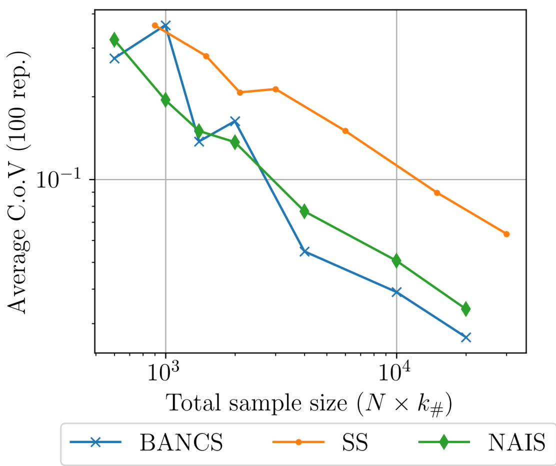
(a) Failure probability (toy-case #2).



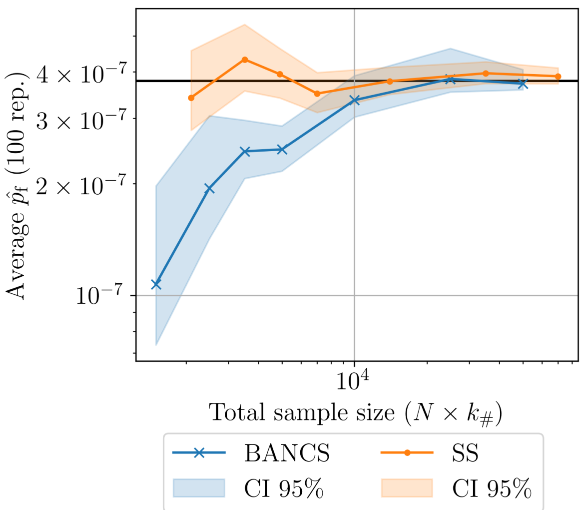
(b) Coefficient of variation (toy-case #2).



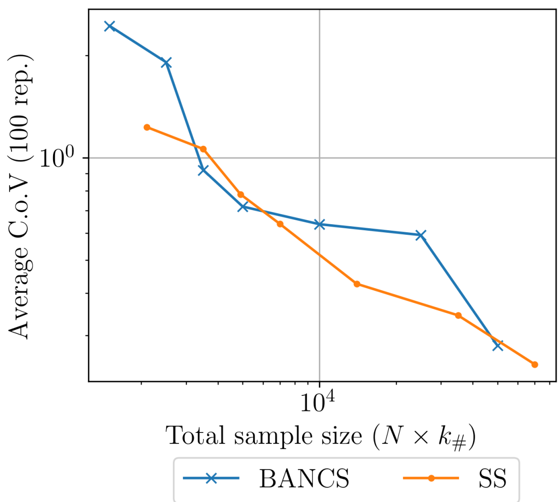
(c) Failure probability (toy-case #4).



(d) Coefficient of variation (toy-case #4).



(e) Failure probability (toy-case #5).



(f) Coefficient of variation (toy-case #5).

Figure 5.4 Reliability analysis benchmark between BANCS, SS and NAIS using 100 repetitions of each experiment (with $p_0 = 0.1$ for every methods). The confidence intervals are obtained by bootstrap on the repetitions. The reference failure probabilities are represented by the horizontal black lines.

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 this issue, the topic of *reliability-oriented sensitivity analysis* (ROSA) is an active research field where several GSA methods were adapted to the analysis of risk measures. Historically, the FORM importance factors represent a first importance measure adapted to ROSA. However, it is only a local measure (obtained as a post-processing of FORM), since it captures the influence near the most-probable failure point. Later on, global sensitivities such as the Sobol' indices were adapted to the indicator function (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 indepencence criterion (HSIC) for reliability. This idea 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 w.r.t. exceeding a threshold on the output;
- *Conditional sensitivity analysis* (CSA), studying the impact of input variables within the restricted domain (i.e., conditionally to this domain).

Finally, several new indices have been proposed in the ROSA context to deal with dependent inputs, such as the Shapley effects by Il Idrissi et al. (2021), further improved by Demange-Chryst et al. (2023). More recently, Ehre et al. (2024) proposed a variance-based approach suiting this context.

All in all, this section aims at illustrating the estimation of HSIC-based 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) is available. These consecutive samples gradually move towards the failure domain and become rarer from the perspective of the initial input distribution. This section first recalls the TSA and CSA adaptation from Raguet and Marrel (2018) and Marrel and Chabridon (2021) of the HSIC indices, whose formulation was introduced for GSA in Section 1.7. Then, the computation of the so-called "target-HSIC" and "conditional-HSIC" indices is presented as a simple post-precessing of the BANCS reliability analysis. This procedure is illustrated on test cases #3 and #5.

5.4.1 Target and conditional HSIC indices

A first approach for TSA is to directly apply any sensitivity measure to the binary variable $\mathbb{1}_{\{g(\mathbf{X}) \leq y_{\mathrm{th}}\}}$. However, this strategy does not distinguish points in the vicinity of the LSF from points that are far from this border. In the present work, a threshold relaxation using the weight transformation proposed in Raguet and Marrel (2018) and further discussed in Marrel and Chabridon (2021) is applied to gather more information. Let us consider the weight function

$w_{\mathcal{F}} : \mathbb{R} \longrightarrow [0,1]$ 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 its respective normalized version, called the target R^2_{HSIC} (T-R^2_{HSIC}) are defined as:

$$T\text{-HSIC}(X_j, Y) = \text{HSIC}(X_j, w_{\mathcal{F}}(Y)) = \text{MMD}^2(\mathbb{P}_{(X_j, w_{\mathcal{F}}(Y))}, \mathbb{P}_{X_j} \otimes \mathbb{P}_{w_{\mathcal{F}}(Y)}), \tag{5.10a}$$

$$T\text{-}R^2_{\text{HSIC}}(X_j, Y) = \frac{T\text{-HSIC}(X_j, w_{\mathcal{F}}(Y))}{\sqrt{T\text{-HSIC}(X_j, X_j)\, T\text{-HSIC}(w_{\mathcal{F}}(Y), w_{\mathcal{F}}(Y))}}. \tag{5.10b}$$

The stronger the dependence, the more X_j is influential 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.

As for the conditional indices, let us define the probability of Y conditionally to the rare event $\{Y \in \mathcal{F}\}$, considering the measurable space $(\Omega, \mathcal{A})$:

$$\mathbb{P}_{Y|\{Y \in \mathcal{F}\}}(A) = \frac{\int_A \mathbb{1}_{g(\mathbf{X}) \leq 0}(y)\, \mathrm{d}\mathbb{P}_Y}{p_{\mathrm{f}}}, \quad \forall A \in \mathcal{A}. \tag{5.11}$$

After applying the same threshold relaxation as earlier, one can write $\forall A \in \mathcal{A}$:

$$\mathbb{P}_Y^{w_{\mathcal{F}}}(A) = \frac{\int_A w_{\mathcal{F}}(y)\, \mathrm{d}\mathbb{P}_Y}{\int_\Omega w_{\mathcal{F}}(y)\, \mathrm{d}\mathbb{P}_Y}, \tag{5.12}$$

and the conditional expectation:

$$\mathbb{E}[Y | Y \in \mathcal{F}] = \mathbb{E}_{Y \sim \mathbb{P}_Y^{w_{\mathcal{F}}}}[Y] = \frac{\int_\Omega y w_{\mathcal{F}}(y)\, \mathrm{d}\mathbb{P}_Y}{\int_\Omega w_{\mathcal{F}}(y)\, \mathrm{d}\mathbb{P}_Y}. \tag{5.13}$$

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

$$C\text{-HSIC}(X_j, Y) = \text{HSIC}_{(X_j, Y) \sim \mathbb{P}^{w_{\mathcal{F}}}_{(X_j, Y)}}(X_j, w_{\mathcal{F}}(Y)), \tag{5.14a}$$

$$C\text{-}R^2_{\text{HSIC}}(X_j, Y) = \frac{C\text{-HSIC}(X_j, w_{\mathcal{F}}(Y))}{\sqrt{C\text{-HSIC}(X_j, X_j)\, C\text{-HSIC}(w_{\mathcal{F}}(Y), w_{\mathcal{F}}(Y))}}. \tag{5.14b}$$

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. In the context of TSA, the asymptotic tests proposed by Gretton et al. (2006) are well suited to the sample sizes encountered in reliability analysis. However, the asymptotic approach no longer holds for CSA, which is replaced by the permutation-based statistical test proposed by De Lozzo and Marrel (2016). 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.

5.4.2 ROSA as a post-processing 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 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 the distribution $\widehat{f}_{[k+1]}$ which is conditional to the threshold $\widehat{q}_k^{p_0}$.

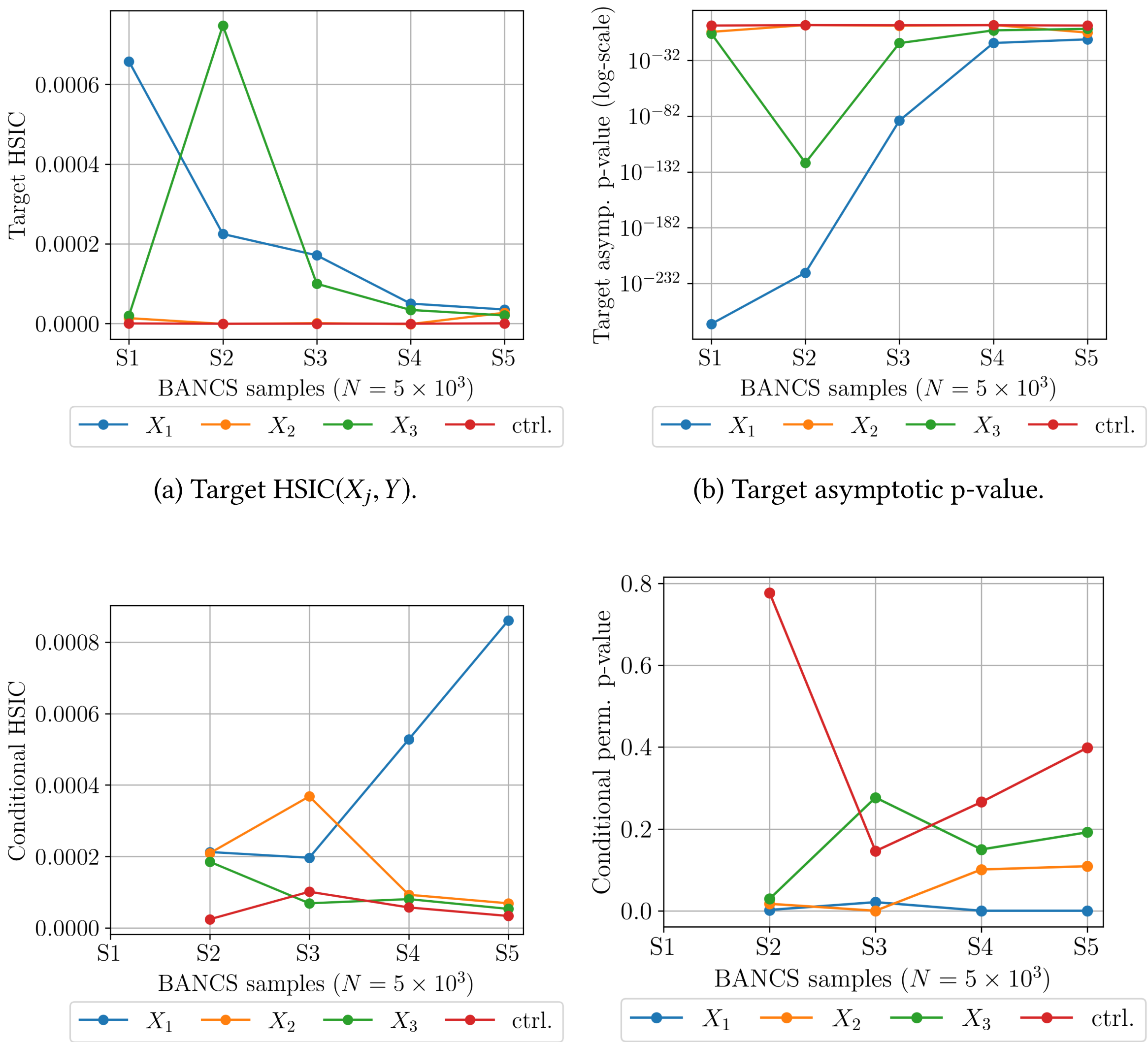
The target and conditional sensitivity analysis are presented via their respective HSIC and the p-value of their independence test. These quantities are plotted as function of the consecutive samples S_k to study the ROSA evolution. Fig. 5.5 represents these results for the test-case #3 while Fig. 5.6 shows the results for test-case #5. To ease the interpretation, a variable without any relation to the output was added. This variable always follows a standard normal distribution and is called the control variable (abbreviated "ctrl." in the figures).

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. The p-value obtained asymptotically for the TSA and by permutations for the CSA, confirm the interpretation of the HSICs. One can conclude that the three variables all have a distinct impact on the failure probability. The variable X_1 has an interesting behavior as the problem becomes rarer, its impact on the TSA decreases while its influence on the CSA increases. It is overall the most influential variable on this ROSA study.

Test-case #5: On the oscillator problem, the variables F_s, ζ_s, and ζ_p have a high target HSIC and low target p-value, meaning that they lead to the failure domain. Interestingly, the T-HSIC and C-HSIC of the resistance variable F_s decrease as the problem becomes rarer. However, F_s remains of prime importance on the CSA. On this medium dimension problem, the variables m_p and k_s have among the lowest T-HSIC.

A similar study was conducted on this test-case by Bourinet (2018) using local ROSA measures, called the score-functions, which rely on derivatives of the failure probability with respect to the moments of the inputs. The results from this work, based on consecutive samples from a SS, are reproduced in Fig. 5.7. Considering a subset probability p_0, the nested failure probabilities are denoted by $\pi_s = \Pi_{k=1}^{s} p_0$, and the inputs first moments by (μ_i, σ_i).

Variables with score-function close to zero should have a small influence on the reliability. Then, one can notice that the conclusions of this local ROSA are quite in line with the target



(a) Target HSIC(X_j, Y).

(b) Target asymptotic p-value.

(c) Conditional HSIC(X_j, Y).

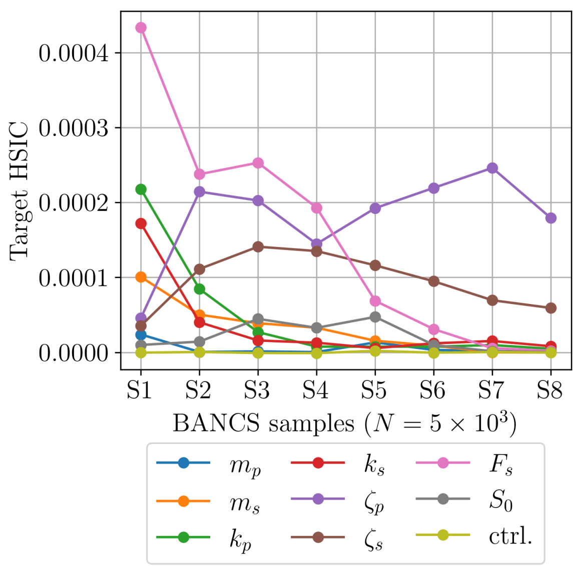
(d) Conditional p-value by perm. (10^3 perm.).

Figure 5.5 Target and conditional HSIC as a post-processing of BANCS reliability analysis of test-case #3 (modified Ishigami). The consecutive samples from BANCS are denoted by $\{S_k\}_{k=1}^{k_\#}$ (each with size $N = 5 \times 10^3$, with $p_0 = 0.25$).

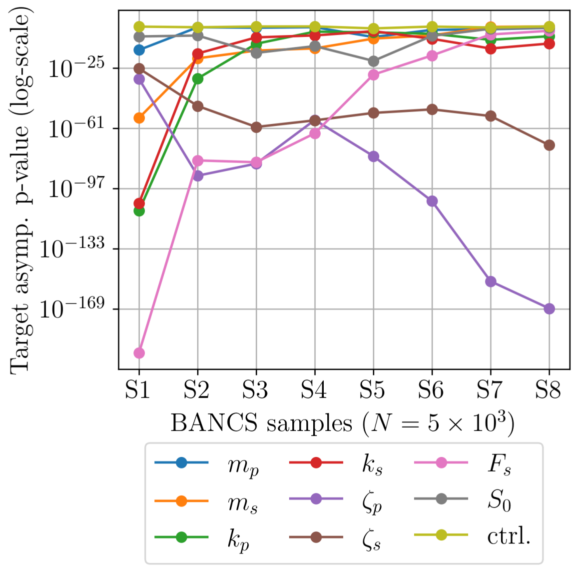
HSIC. The consequence of fixing these variables to their mean values should be further studied to propose solid recommendations for screening in ROSA.

5.5 Conclusion

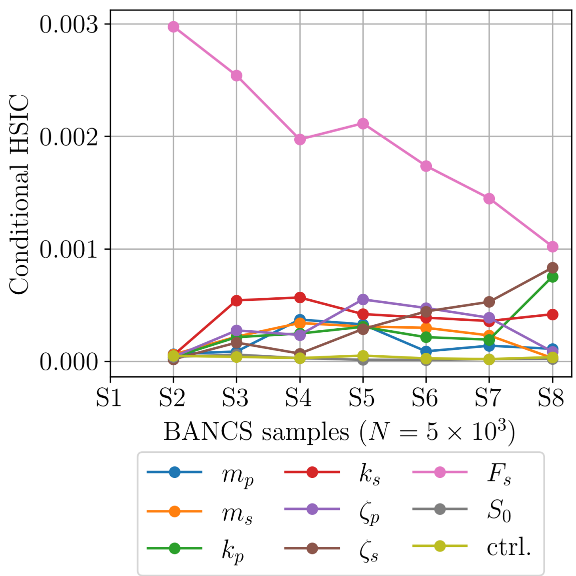
In this chapter a contribution to rare event estimation was presented. BANCS is a new adaptive importance sampling method using nonparametric Bernstein copula to fit the successive conditional distributions. Its performance was compared in a numerical benchmark with other methods: SS and NAIS. The current implementation is at least as efficient as the other methods on the analytical problems studied, however numerous potential improvement remain unexplored. A first perspective lies in the optimization of the EBC polynomial order by applying a validation procedure similar to what is done for surrogate models (among the elite set, a



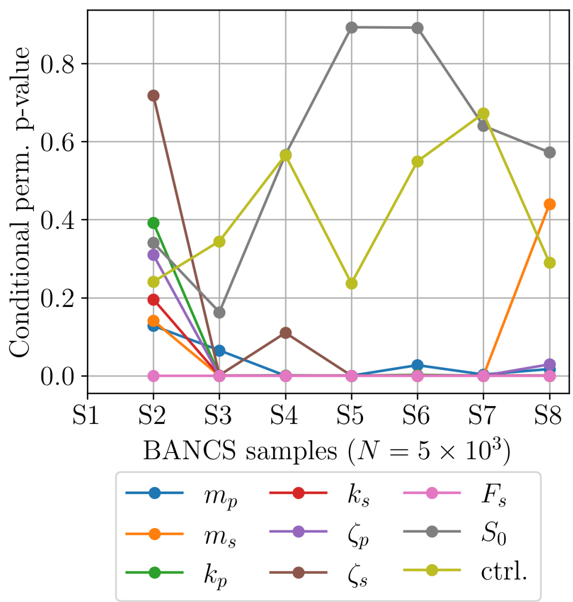
(a) Target HSIC(X_j, Y).



(b) Target asymptotic p-value.



(c) Conditional HSIC(X_j, Y).



(d) Conditional p-value by perm. (10^3 perm.).

Figure 5.6 Target and conditional HSIC as a post-processing of BANCS reliability analysis of test-case #5 (oscillator problem). The consecutive samples from BANCS are denoted by $\{S_k\}_{k=1}^{k_\#}$ (each with size $N = 5 \times 10^3$, with $p_0 = 0.25$).

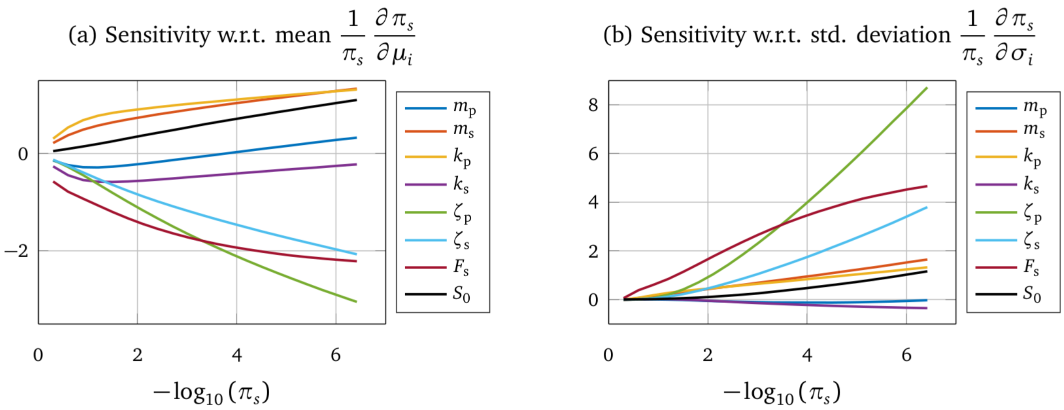


Figure 5.7 Normalized score-functions of $\pi_s = \Pi_{k=1}^{s} p_0$ w.r.t. the inputs mean μ_i and standard deviation σ_i in the standard normal space (source: Bourinet, 2018, p.54). The consecutive probabilities result from a SS (with $N = 10^6$ and $p_0 = 0.5$ per subset).

test set could be selected using the results from Chapter 4). Then, instead of estimating the intermediary quantiles by Monte Carlo, other sampling methods could be tested (e.g., LHS, randomized QMC, see Kaplan et al., 2019, or importance sampling including all the samples). To tackle high-dimentional problems, the inference of copulas by blocks could be interesting (see e.g., Lasserre, 2022).

BANCS also presents multiple advantages, first, it does not require a transformation in the standard space, second, its flexibility allows to capture multimodal problems. A third advantage is that the samples generated at each iteration are i.i.d., offering the possibility to assess global reliability-oriented sensitivity analysis as a post-processing. In this chapter, ROSA was investigated using the HSIC (following the work of Marrel and Chabridon, 2021) but the use of Shapley indices could provide tools to understand problems with dependent inputs (Il Idrissi et al., 2021). This complementary analysis is essential to understand the influence of the inputs on the failure probability. Further studies should be conducted to guarantee the screening of noninfluential variables in the context of rare events.

Chapter 6

Application to wind turbine fatigue reliability and robustness

6.1 Introduction

One of the main goals of this work is to evaluate the reliability of an OWT's monopile foundation with respect to fatigue solicitations. Let us recall the usual approach to assess fatigue damage over the structure's lifetime (see IEC-61400-1, 2019, Appendix H). First, the lifetime duration T is discretized into $N_T \in \mathbb{N}$ 10-minutes intervals $\{t^{(i)}\}_{i=1}^{N_T}$. The 10-minutes duration results from the typical wind energy distribution (see Fig. ??), presenting a "short-term" behavior (for turbulent wind with a return period below 10 minutes) and "long-term" behavior (otherwise). Then the environmental conditions can be considered as a stochastic process $\{\mathbf{X}(t,\omega),\ t \in T,\ \omega \in \Omega\}$.For a realization $\mathbf{X}(t^{(i)}, \omega^{(r)})$ and a given set of parameters z related to the system (defined in Table ??), one can perform a 10-minutes Turbsim-DIEGO simulation and post-process it (see Subsection ??) to obtain a corresponding cumulative damage:

$$d_{\mathrm{c}}^{10\mathrm{min.}}(\mathbf{X}(t^{(i)}, \omega^{(r)})|Z = z). \tag{6.1}$$

To cumulate the damage over the lifetime, let us write the sum of 10-minutes damages, each averaged over $n_{\mathrm{rep}} \in \mathbb{N}$ pseudo-random seed repetitions:

$$\begin{aligned} D(z) &= \frac{1}{n_{\mathrm{rep}}} \sum_{i=1}^{N_T} \sum_{r=1}^{n_{\mathrm{rep}}} d_{\mathrm{c}}^{10\mathrm{min.}}(\mathbf{X}(t^{(i)}, \omega^{(r)})|Z = z) && (6.2\mathrm{a}) \\ &= N_T \frac{1}{N_T\, n_{\mathrm{rep}}} \sum_{i=1}^{N_T} \sum_{r=1}^{n_{\mathrm{rep}}} d_{\mathrm{c}}^{10\mathrm{min.}}(\mathbf{X}(t^{(i)}, \omega^{(r)})|Z = z) && (6.2\mathrm{b}) \\ &\approx N_T\, \mathbb{E}_{\mathbf{X}}\left[d_{\mathrm{c}}^{10\mathrm{min.}}(\mathbf{X}|Z = z)\right], && (6.2\mathrm{c}) \end{aligned}$$

where $\mathbf{X} \sim f_{\mathbf{X}}$ is the random vector of the long-term environmental conditions, defined in Table ??.

In Chapter 3, kernel herding was proposed as a method for given-data subsampling to propagate the uncertain environmental conditions on Teesside's OWT model. This method showed equivalent performances to quasi-Monte Carlo for estimating the lifetime damage in Eq. (6.2c), while being more flexible. In the linear cumulative damage model considered by the community (Miner's rule), a damage value higher than one leads to fatigue failure by convention. The present chapter assesses the probability of such rare event, considering both the environmental uncertainties (aggregated according to Eq. (6.2c)) and the uncertainties related to the system itself (described by the random variable $\mathbf{Z} \in \mathcal{D}_{\mathbf{Z}}$ with PDF $f_{\mathbf{Z}}$). Assuming that the critical damage D_{cr} is a random variable centered around one (equivalent to a "resistance" variable in reliability analysis), this failure probability is written:

$$p_{\mathrm{f}} = \int_{\mathcal{D}_{\mathbf{Z}}} \mathbb{1}_{\{D(\mathbf{z}) \geq D_{\mathrm{cr}}\}}\, f_{\mathbf{Z}}(\mathbf{z})\, \mathrm{d}\mathbf{z}\,. \tag{6.3}$$

Less information is available to define the probabilistic model of the system uncertainties than the environmental uncertainties. Therefore, the robustness of the failure probability to the probabilistic model $\mathbf{Z}$ should be studied. To do so, a perturbation-based approach called the *perturbed-law based sensitivity indices* (PLI) is used in this chapter (Lemaître et al., 2015).

The present chapter is structured as follows: Section 6.2 presents the construction of a surrogate model of $D(\cdot)$, then Subsection 6.3.1 analyses the reliability of the monopile foundation of Teesside's turbine for a nominal distribution of $\mathbf{Z}$, and finally Subsection 6.3.2 proposes a robustness analysis of p_{f} by perturbing the law of $\mathbf{Z}$ and D_{cr}.

6.2 Surrogate modeling for reliability analysis

The prohibitive computational cost of the function $D(\cdot)$ requires fitting a surrogate model. This section presents the specific high-performance computer (HPC) wrapper developed for this application and its use to build a learning set for a Gaussian process regression.

6.2.1 High-performance computer evaluation

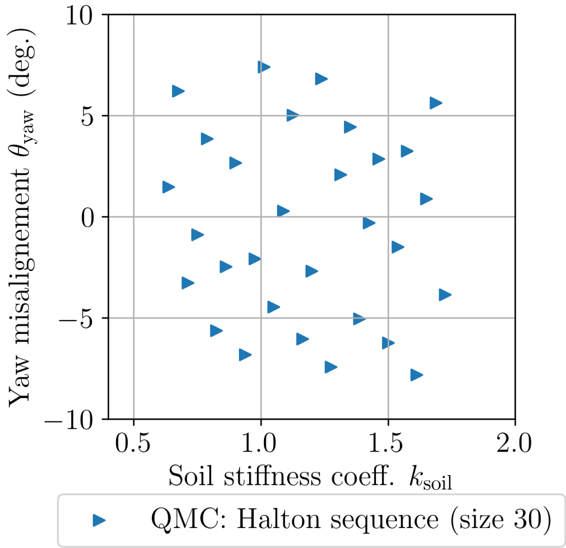
The wrapper of the numerical chain including Turbsim and Diego (illustrated in Fig. **??**) for reliability analysis has a nested double loop structure. The outer loop pilots the realizations of $\mathbf{Z}$ while the inner loop concerns the environmental conditions and their repetitions with n_{rep} different pseudo-random seeds. At this stage, the goal is to approximate $\widehat{D}(k_{\mathrm{soil}}, \theta_{\mathrm{yaw}}) : \mathbb{R}^2 \longrightarrow \mathbb{R}$:

$$\widehat{D}(k_{\mathrm{soil}}, \theta_{\mathrm{yaw}}) = N_T \sum_{i=1}^{n_{\mathrm{X}}} \sum_{r=1}^{n_{\mathrm{rep}}} d_{\mathrm{c}}^{10\mathrm{min.}}(\mathbf{x}^{(i)}, \omega^{(r)} | K_{\mathrm{soil}}, \Theta_{\mathrm{yaw}}), \tag{6.4}$$

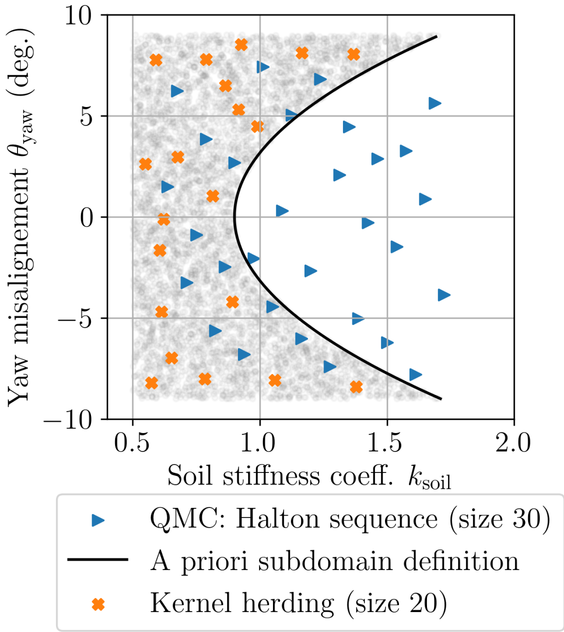
where $\{\mathbf{x}^{(i)}\}_{i=1}^{n_{\mathrm{X}}}$ are defined by kernel herding, and $(K_{\mathrm{soil}}, \Theta_{\mathrm{yaw}})^\top$ are the system variables which are direct inputs of DIEGO (i.e., not part of the post-processing). According to the convergence results obtained in Chapter 3, the kernel herding size is fixed at $n_{\mathrm{X}} = 200$ and the repetitions at $n_{\mathrm{rep}} = 11$, which implies a total of 2200 Turbsim-DIEGO simulations per evaluation of the function $\widehat{D}(\cdot)$. In this setup, the CRONOS HPC from EDF allows us to simultaneously perform those 2200 simulation in parallel. The random variable associated with the S-N curve uncertainty is introduced later as a product factor of Eq. (6.4). Note that the cumulative damage studied is actually the maximum value of $\widehat{D}(\cdot)$ over the discretized azimuth angles (illustrated in Fig. 3.4), at the mudline level.

6.2.2 Design of experiments

To build a learning set, a space-filling design of experiments is created on the joint domain of $(K_{\mathrm{soil}}, \Theta_{\mathrm{yaw}})$. This design was first composed of 30 points generated by a Halton sequence (illustrated in Fig. 6.1.a) which was the most space-filling sequential method in two dimensions. Its evaluation and analysis showed that the highest damage values are the result of high yaw



(a) Halton sequence.



(b) Composite design.

Figure 6.1 Learning set of the mean damage surrogate model. A Halton sequence is first built (in blue) and competed by kernel herding points (in orange) in a subdomain defined a priori (in gray).

misalignment errors and low soil stiffness. Therefore, the design was completed in a second phase by 20 points targeting these areas by applying a kernel herding in a subdomain defined a priori (see the candidate set represented by the gray points in Fig. 6.1.b). Finally, the learning set is the union of the two complementary designs, later referred as the “composite design”, in reference to the heterogeneous composition of composite materials. This composite design, denoted by $\mathbf{Z}_{n_Z}$, has a size of $n_Z = 50$, which represents over 10^5 Turbsim-DIEGO simulations (each requiring around 45 minutes of CPU time). Fig. 6.2 shows the lifetime cumulated damage evaluated on the composite design (with normalized values corresponding to the color scale).

6.2.3 Gaussian process regression

A Gaussian process regression with Matérn 5/2 and constant trend is fitted on the composited design $\mathbf{Z}_{n_Z}$ according to the Kriging equation introduced in Section 1.8. The resulting surrogate model $\widetilde{D} : \mathbb{R}^2 \rightarrow \mathbb{R}$, is represented by the blue three-dimensional surface in Fig. 6.3, and its learning set $\mathbf{Z}_{n_Z}$ by the black crosses. A complementary visualization of this surrogate is proposed for cross-sections with fixed values $\theta_{\text{yaw}} = 0$ on Fig. 6.4.a, and $k_{\text{soil}} = 1$ on Fig. 6.4.b. On these two figures, learning points are plotted in a gray scale and the surrogate model in a blue scale. The darkest the shade, the closest to the cross-section the points are.

To validate this surrogate model, a leave-one-out (LOO) procedure is realized. Fig. 6.5.b represents the LOO squared-residuals at each points of the design (with values corresponding to the color scale). High residual values are mostly due to the strong nonlinearity of the code in

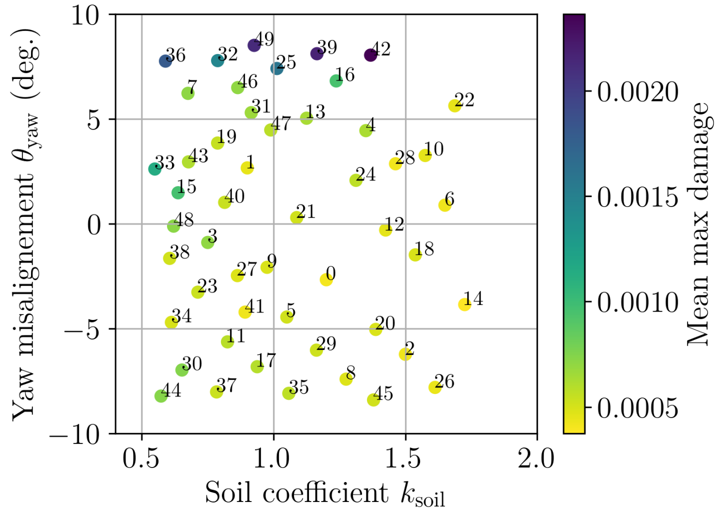


Figure 6.2 Normalized mean damage evaluated on the composite design illustrated in Fig. 6.1.b.

some areas (as revealed by the cross-section in Fig. 6.4.b). Fig. 6.5.a shows the quantile-quantile plot comparing the LOO predictions with the lifetime damage evaluations on the learning set. The general coefficient of predictivity of $\widehat{Q}^2_{LOO} = 0.72$ is considered acceptable in this small data context, especially as the LOO procedure was shown to underestimate the true performance metric in Chapter 4. However, more points could be added to complete the learning set in regions with high nonlinearities to enhance the surrogate presented.

Remark 9.

- Active learning methods for reliability (see Subsection 1.8.3) could be a great option for such costly function. However, the stochasticity of the function would disturb the learning criterion. Since the nonlinearities seem restricted to a small area, the present approach should be more robust.
- Another approach could be to fit a stochastic surrogate (Binois et al., 2019; Baker et al., 2022; Zhu, 2022) on a learning set before averaging on the pseudo-random seed repetitions.

6.3 Reliability and robustness analysis

In the wind energy industry, acceptable risk levels for fatigue are defined by the standards. The target failure probability of the order of magnitude of 10^{-4} over the last year of exploitation is indicated by the IEC-61400-1 (2019) while the recommendations of other standards are reviewed by Wang et al. (2022). Nielsen and Sørensen (2021) discussed the relevance of this risk level, defined from an economic point of view, and offered quantitative guidelines for lifetime extension. In this section, a probabilistic reliability analysis is conducted using the surrogate model of the lifetime cumulative damage. Afterwards, the robustness of this reliability analysis is studied by applying perturbations to the input distributions and computing the perturbed-law based sensitivity indices (PLI) initially proposed by Lemaître et al. (2015).

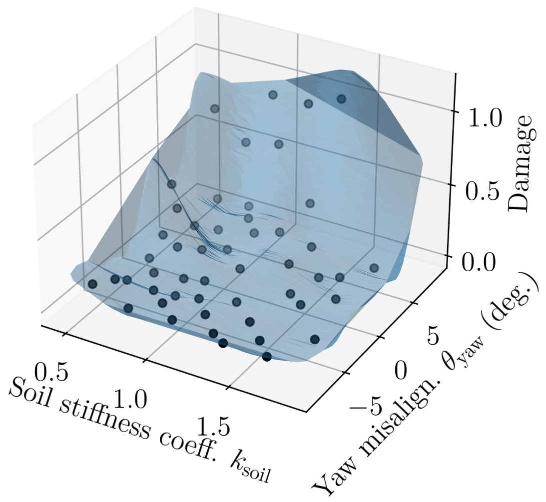
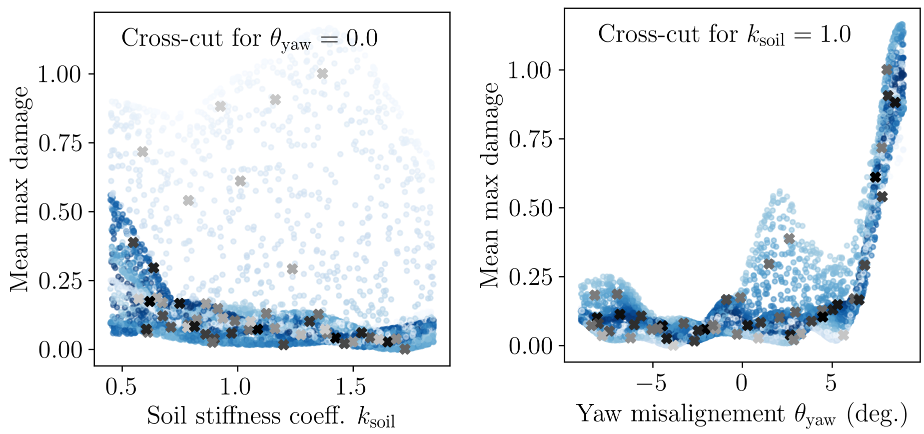


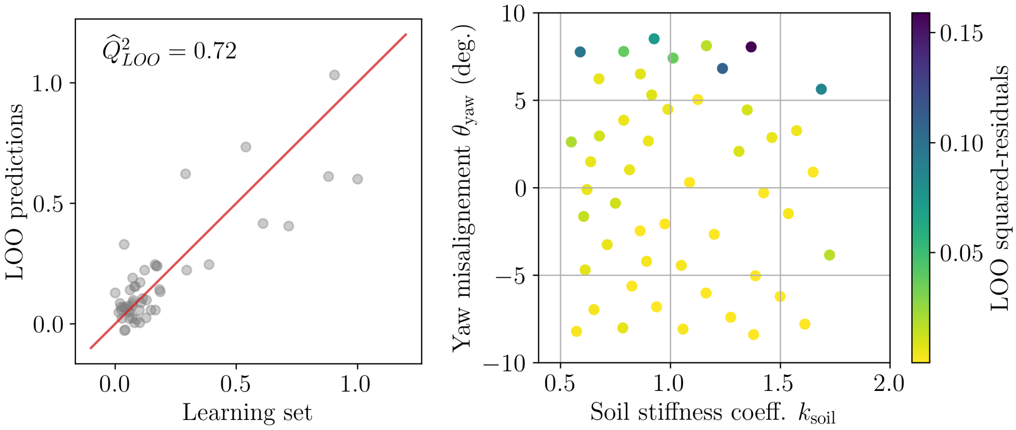
Figure 6.3 Three-dimensional plot of the surrogate model $\widetilde{D}$ (in blue) and learning set (in black).



(a) $\theta_{yaw} = 0$.

(b) $k_{soil} = 1$.

Figure 6.4 Cross-section of the surrogate model $\widetilde{D}$ (in shades of blue) for given values of k_{soil} and θ_{yaw}. The darkest the shade, the closest to the cross-section.



(a) Quantile-quantile plot.

(b) LOO Squared-residuals.

Figure 6.5 Leave-one-out validation results of the surrogate model $\widetilde{D}$.

6.3.1 Nominal reliability analysis

The surrogate model of the lifetime cumulative damage $\widetilde{D}$ is first modified to include the S-N curve uncertainty ε, defined in Subsection **??**:

$$\widetilde{D}'(\mathbf{z}) = \widetilde{D}'(k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon) = \frac{1}{\varepsilon}\widetilde{D}(k_{\text{soil}}, \theta_{\text{yaw}}). \tag{6.5}$$

The probability introduced in Eq. (6.3) then becomes:

$$p_{\text{f}} = \int_{\mathcal{D}_{\mathbf{Z}}} \mathbb{1}_{\{\widetilde{D}'(\mathbf{z}) \geq D_{\text{cr}}\}} f_{\mathbf{Z}}(\mathbf{z})\, \mathrm{d}\mathbf{z}, \tag{6.6}$$

Table 6.1 presents the estimates of this quantity by different methods (FORM, FORM-importance sampling and subset simulation, see Section 1.6) and for two hypotheses regarding the distribution of the resistance variable. D_{cr} either follows a lognormal distribution (which has a short left tail), or a normal distribution (with a heavier left tail). All the methods deliver similar values of p_{f} even if subset simulation requires more time. The adequation of FORM with the simulation methods reveals that the LSF in this case is almost linear. Such a conclusion might differ depending on the OWT model studied (e.g., a floating model could present a different behavior). The probabilities are, as expected, much lower under the hypothesis of lognormal distribution for D_{cr} than for a normal distribution. However, this significant difference questions the robustness of this result to the probabilistic model of D_{cr} and $\mathbf{Z}$.

Table 6.1 Nominal reliability analysis (IS and SS size $N = 5 \times 10^4$, SS $p_0 = 0.1$).

	$D_{\text{cr}} \sim$ **Lognormal**		$D_{\text{cr}} \sim$ **Normal**	
Reliability method	$\widehat{p_{\text{f}}}$	$\widehat{\text{cov}}$	$\widehat{p_{\text{f}}}$	$\widehat{\text{cov}}$
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%

6.3.2 Robustness analysis using the perturbed-law sensitivity indices

The method of Lemaître et al. (2015), later called *perturbed-law based sensitivity indices* (PLI) by Sueur et al. (2017) relies on perturbating densities. The goal is to assess the robustness of a quantity of interest (a failure probability in our case) to these perturbations. An application of the PLI on a thermal-hydraulic numerical code from the nuclear industry was proposed in Iooss et al. (2022).

Assuming a random variable $Z_j \sim f_j \in \mathcal{D}_{Z_j}$ with mean $\mathbb{E}[Z_j] = \mu$ and variance $\text{Var}(Z_j) = \sigma^2$, the strategy is to find the "closest" distribution $f_{j\delta}$ under the constraint of moment perturbation of magnitude δ. The notion of proximity between distributions is quantified by Lemaître et al. (2015) in terms of Kullback–Leibler divergence (KL). For example, a relative mean perturbation

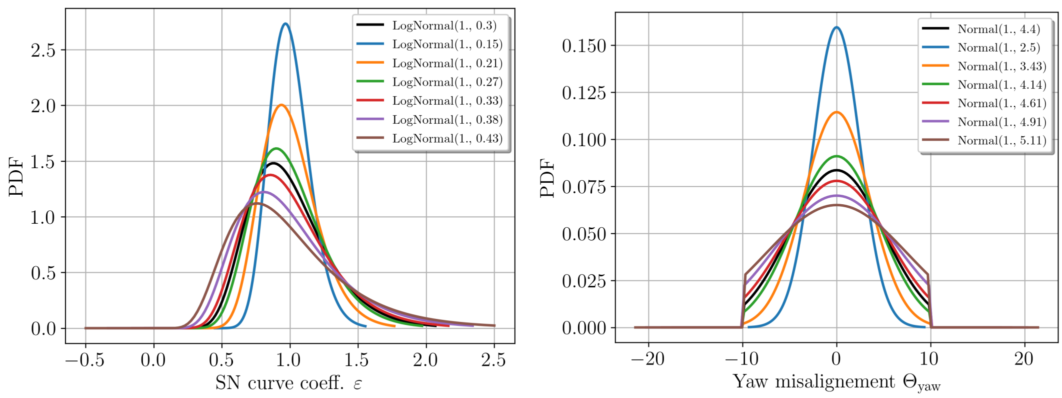


Figure 6.6 Perturbations in terms of standard deviation of a lognormal distribution (left) and a truncated normal distribution (right).

is defined as:

$$f_{j\delta} = \underset{\substack{\pi \in \mathcal{P}, \\ \text{s.t.,} \mathbb{E}_\pi[Z_j] = \mathbb{E}_{f_j}[Z_j](1+\delta)}}{\arg\min} \mathrm{KL}(\pi || f_j), \quad \delta \in \mathbb{R}. \tag{6.7}$$

Note that the perturbed distribution $f_{j\delta}$ might not belong to the parametric family of f_j. This is typically the case for bounded distributions, for example when perturbating the mean of a uniform distribution. To simplify the perturbation problem studied in Lemaître et al. (2015); Gauchy et al. (2022), the perturbations realized in the following will conserve the initial parametric family (which seems reasonable for distributions in the exponential family).

The adapted expression of the PLI used hereafter (Gauchy et al., 2022) reflects the relative impact of a pertubation on the quantity of interest:

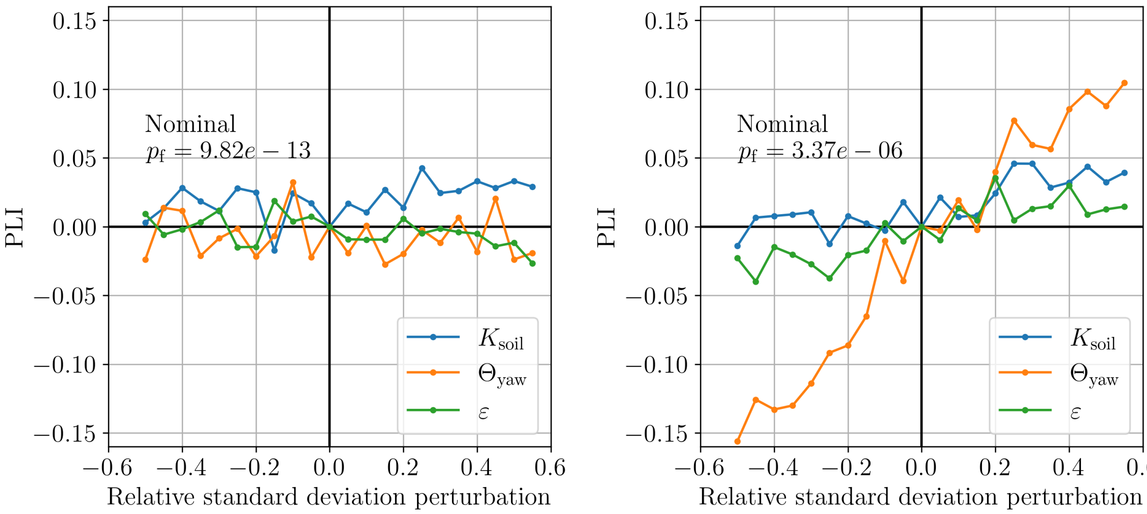
$$\mathrm{PLI}(f_{j\delta}) = \frac{p_{\mathrm{f},j\delta} - p_{\mathrm{f}}}{p_{\mathrm{f}}}, \tag{6.8}$$

where $p_{\mathrm{f},j\delta}$ is the probability obtained when injecting $f_{j\delta}$ in Eq. (6.6).

In our case, each variable in $\mathbf{Z}$ is perturbed one by one in terms of relative standard deviation, such that $\sigma_{j\delta} = \sigma_j(1+\delta)$. The illustration of such perturbations is illustrated in Fig. 6.6, for distributions of K_{soil} on the left, and of Θ_{yaw} on the right. This strategy assumes that the analyst has enough information to determine the mean of the variables Z_j.

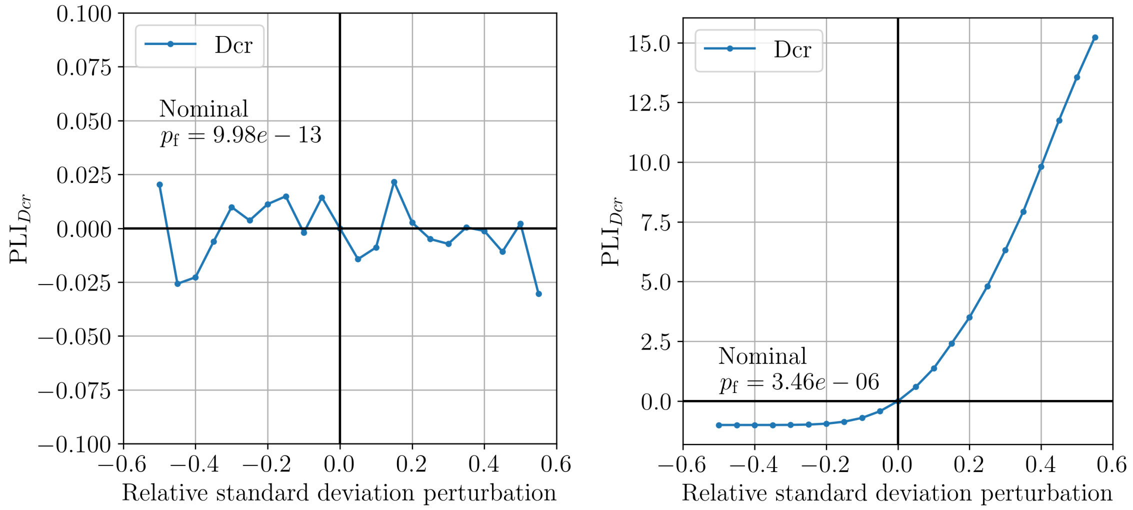
The resulting PLI are presented in Fig. 6.7 for relative perturbations of the standard deviations of $(K_{\mathrm{soil}}, \Theta_{\mathrm{yaw}}, \varepsilon)$. Each failure probability is independently estimated by FORM-IS. In the hypothesis of a normal D_{cr}, the most important variable is Θ_{yaw}, while the fluctuations are quite stable in the hypothesis of a lognormal D_{cr}.

When perturbating the standard deviation of the resistance variable D_{cr}, the same phenomenon is witnessed in Fig. 6.8. The perturbations have nearly no consequences, assuming that $D_{\mathrm{cr}} \sim$ LogNormal, but a lot of influence when $D_{\mathrm{cr}} \sim$ Normal. As a perspective, this study could be completed by a joint perturbation of both standard deviation and mean of the resistance variable.



(a) $D_{cr} \sim$ Lognormal. (b) $D_{cr} \sim$ Normal.

Figure 6.7 Perturbed-law based indices for relative perturbations of the standard deviations of $(K_{soil}, \Theta_{yaw}, \varepsilon)$. The failure probabilities studied are each estimated by FORM-IS method with sample size $N = 5 \times 10^4$.



(a) $D_{cr} \sim$ Lognormal. (b) $D_{cr} \sim$ Normal.

Figure 6.8 Perturbed-law based indices for relative perturbations of the standard deviation of D_{cr}. The failure probabilities studied are each estimated by FORM-IS method with sample size $N = 5 \times 10^4$.

6.4 Conclusion

In this chapter, a fully probabilistic reliability analysis was performed on the monopile foundation of an OWT located in Teesside, UK. Our demonstrative study first required an important computational effort (i.e., over 10^5 Turbsim-DIEGO simulations), made possible with the development of a tailored wrapper deployed a high-performance facility. These simulations served the construction of a surrogate model emulating the lifetime damage function.

The second phase of this chapter addressed the estimation of a failure probability using this surrogate. A general conclusion is that this probability is highly dependent of the probabilistic model describing the critical damage. To challenge the modeling hypothesis of the critical damage and the system variables, a robustness analysis was realized. It consists in studying the impact of perturbating the input distributions on an output quantity. The robustness of the failure probability was evaluated with the formalism of the perturbed-law based indices, introduced by Lemaître et al. (2015). This additional study mostly confirmed the importance of the critical damage definition.

From an industrial point of view, the failure probabilities obtained are lower than, the risk levels targeted in the standards (Wang et al., 2022), which is in favor of lifetime extension. However, let us outline that this demonstrative results did not consider:

- The periods in parked position which could significantly increase the damage (see Velarde et al., 2020). Different studies showed that the aerodynamic damping created by the rotation of the turbine reduces tower vibrations, and therefore fatigue (Liu et al., 2017);
- The rapid transition phases occasioned by emergency stopping;
- The early damage produced during the installation of the monopile foundations by hydraulic impact piling (i.e., hammering).
- The stress-concentration resulting from soldering the structure;

An interesting industrial perspective could be to reproduce a similar study on a floating OWT model and compare the conclusions.

From a methodological aspect, a stochastic surrogate model could be built by considering all the damages before averaging them over the environmental repetitions. Quantile estimation on stochastic models was studied by Browne et al. (2016), however, rare event estimation on stochastic functions remains an open question recently addressed by Pires et al. (2023).

Conclusion and perspectives

Summary of the main contributions and perspectives

Uncertainty quantification applied to offshore wind turbine (OWT) raises numerous methodological questions (Veers et al., 2019). In the present thesis, two main statistical tools were exploited for OWT uncertainty quantification: the kernel-based dissimilarity measure called maximum mean discrepancy (Gretton et al., 2006), and the empirical Bernstein copula (Sancetta and Satchell, 2004) for multivariate nonparametric inference. This work first tackled problems related to uncertainty quantification of offshore environmental conditions (i.e., metocean), including their perturbation induced by the turbine's wake. In a second part, it dealt with uncertainty propagation for various goals such as mean estimation, mean assessment of surrogate models' predictivity, or rare event probability estimation. For reproducibility purposes, most numerical developments related to this work are documented and openly accessible.

Uncertainty quantification of environmental conditions

The first contribution to this topic is a study on the efficiency of a semiparametric inference on the metocean data. In the proposed strategy, the copula is approximated by EBC, while the marginals can either be fitted by parametric or nonparametric methods. The flexibility of the EBC suits the complex dependence structures found in metocean datasets. As perspectives, a more extensive numerical comparison (e.g., with vine copula Vanem, 2016; Lin and Dong, 2019) could be performed, for example using the MMD as a statistic for multivariate testing. Note that the MMD could also serve as a metric to optimize the Bernstein polynomial order. Then, the goodness-of-fit for nonparametric methods might be studied in the same fashion as the validation of machine learning models (by wisely choosing a set of points to test the model). Finally, EBC could also be studied in the context of extreme conditions assessment, which is usually in favor of parametric approaches (Vanem et al., 2024).

After defining a probabilistic model of the ambient metocean conditions, the influence of the turbines' wake was studied. An uncertainty propagation of the ambient conditions through a simplified wake model of a wind farm provided the wake-perturbed metocean distribution at each turbine. Then the MMD was proposed as a dissimilarity measure between these distributions. A matrix of dissimilarities was obtained by computing the MMD between every pair of turbines, later used in a clustering to gather turbines with similar conditions. By creating clusters of turbines solicited with similar environmental conditions, this approach aims at reducing the number loading studies at the farm scale (e.g., fatigue assessment). These contributions are related to the following publications:

- 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).*

- E. Vanem, E. Fekhari, N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2024). "A joint probability distribution model for multivariate wind and wave conditions". In: *Journal of Offshore Mechanics and Arctic Engineering*, In press.

- E. Vanem, Ø. Lande and E. Fekhari, (2024). "A simulation study on the usefulness of the Bernstein copula for statistical modeling of metocean variables". To appear in: *Proceedings of the ASME 2024 43th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2024).*

Reliability and robustness analysis of offshore wind turbines subject to fatigue damage

This work is related to uncertainty propagation on the multi-physics numerical model of an offshore wind turbine DIEGO. A first contribution concerns the estimation of lifetime cumulative fatigue damage as a mean against the environmental conditions (see e.g., Müller and Cheng, 2018). The use of Bayesian quadrature methods such as the kernel herding (Chen et al., 2010; Huszár and Duvenaud, 2012) is an efficient and flexible solution for given-data uncertainty propagation (i.e., directly subsampling from a large dataset without inference). Our numerical experiments showed conclusive results in comparison to Monte Carlo and quasi-Monte Carlo sampling. This development was gathered in a Python package named `otkerneldesign`.

As a perspective, the damage surrogate modeling could be further studied (e.g., following the work of Slot et al., 2020). Recent techniques for stochastic emulators could be tested on this case (Baker et al., 2022; Zhu and Sudret, 2023; Lüthen et al., 2023). However, the method used should also be robust to the highly skewed distribution of the output variable. From a mechanical aspect, other cumulative models for fatigue rule could be considered rather Miner's rule. For example, nonlinear models (Rocher et al., 2020) take into account the order of apparition of the cyclic solicitations.

Using kernel herding to estimate lifetime cumulative damage over the environmental conditions, variables related to the system could be modified. Therefore, a second contribution deals with the reliability analysis on few system variables: soil stiffness, S-N curve parameters, yaw misalignment, and the damage resistance. This study was made possible by the use of high-performance computers and a surrogate model built on a space-filling design over the system variables. Beyond the evaluation of a monopile foundation's failure probability, this thesis studied its robustness to the perturbation of inputs distributions. The perturbed-law based sensitivity indices (Lemaître et al., 2015) were applied to assess the relative influence of perturbating the inputs' moments.

An industrial perspective regarding the reliability analysis is its application to a floating wind turbine case. This setup could introduce more nonlinearitites arising from the stronger impact of the waves on a floating structure. Along with the fatigue reliability, the connection with other failure modes such as extreme loading could be studied (see for example the recent study on ultimate limit states by Wang et al., 2023). These contributions are related to the following publications:

- E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2024). "Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature". In: *Data-Centric Engineering*, In press.

- E. Fekhari, B. Iooss, V. Chabridon and J. Muré (2022). "Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool". In: *Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022).*

Surrogate model predictivity assessment

In the validation process of a surrogate model, cross-validation may be costly to implement or simply not acceptable to guarantee an independent evaluation. The use of space-filling designs of experiments to define a complementary test set to the learning set was proposed. Additionally, a method to compute optimal weights using Bayesian quadrature allows to improve the estimation of coefficients of predictivity.

To extend the incremental construction of test sets, the definition of a MMD-based stopping rule could be developed. Additionally, the computational effort could be reduced by only performing the validation on variables with a significant impact on the output, determined by screening techniques. This contribution is related to the following publication:

- 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.

Adaptive rare event estimation using Bernstein copula

A new method for rare event estimation was proposed in this work. This adaptive importance sampling method, named Bernstein adaptive nonparametric conditional sampling (BANCS), relies on the EBC to infer consecutive conditional distributions. Decomposing this inference between copula and marginals brings more flexibility and showed good performances compared to NAIS (Zhang, 1996) and SS (Au and Beck, 2001). This contribution offers numerous research perspectives. Rather than Monte Carlo sampling on the conditional distributions, other methods could help speed-up the quantile estimation (e.g., randomized QMC (Kaplan et al., 2019)). Instead of fixing a sampling size to each subset, a stopping criteria could arise from the accuracy in the quantile estimation. To improve the inference of the conditional distributions, further work on optimizing the EBC polynomial order should be carried on. Finally, an adaptation of BANCS could be proposed for high-dimensional problems by dividing the d-dimensional copula into independent blocks.

As this method generates i.i.d. samples, they can be used in a post-processing step for ROSA. The same way FORM delivers local sensitivities, a good practice could be to systematically study this ROSA to better understand the reliability. Working on the interpretation of different indices, an interesting objective could be to certify that some inputs do not affect the reliability. This contribution is related to the following publication:

> ❝ 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 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14).*

Summary of the numerical developments

`otkerneldesign` This Python package could be numerically improved by improving matrix management and operation. Using methods as the "hierarchical matrices" (Börm et al., 2003) could considerably ease the manipulation of large matrices and speed-up their operations. Additionally, the increasing use of CPU for parallel computing could be useful. This could for example be realized by the use of the GPU programming interface CUDA (for "compute unified device architecture"). Additionally, different MMD estimators could be offered to the user (Gretton et al., 2006).

`bancs` This Python package would need a proper documentation and a more exhaustive benchmark using the Python package `otbenchmark`[1]. In addition, various methodological perspectives mentioned earlier could be developed.

`copulogram` This visualization tool would need a proper documentation to better illustrate the added value of this plot.

Offshore wind turbine reliability wrapper An important development in this thesis was related to coupling `OpenTURNS` with the offshore numerical model TurbSim-DIEGO on a high-performance computer. This implementation should be adapted to a more challenging floating model.

[1]https://github.com/mbaudin47/otbenchmark/

Bibliography

Abdallah, I., Lataniotis, C., and Sudret, B. (2019). Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators – Application to extreme loads on wind turbines. *Probabilistic Engineering Mechanics*, 55:67 – 77.

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 industrialsystems*. 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.

Andrieu-Renaud, C., Sudret, B., and Lemaire, M. (2004). The PHI2 method: a way to compute time-variant reliability. *Reliability Engineering & System Safety*, 84(1):75–86.

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.

Bai, H., Shi, L., Aoues, Y., Huang, C., and Lemosse, D. (2023). Estimation of probability distribution of long-term fatigue damage on wind turbine tower using residual neural network. *Mechanical Systems and Signal Processing*, 190:110101.

Bai, Z., Radhakrishna-Rao, C., and Zhao, L. (1989). Kernel Estimators of Density Function of Directional Data. In *Multivariate Statistics and Probability*, pages 24–39. Academic Press.

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.

Basu, A., Shioya, H., and Park, C. (2011). *Statistical inference: the minimum distance approach.* CRC press.

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.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications.* John Wiley & Sons.

Bénard, C., Biau, G., Da Veiga, S., and Scornet, E. (2022). SHAFF: Fast and consistent SHApley eFfect 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.

Bitner-Gregersen, E. (2015). Joint met-ocean description for design and operations of marine structures. *Applied Ocean Research*, 51:279–292.

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.

Blondel, F. and Cathelain, M. (2020). An alternative form of the super-Gaussian wind turbine wake model. *Wind Energy Science*, 5(3):1225–1236.

Börm, S., Grasedyck, L., and Hackbusch, W. (2003). Introduction to hierarchical matrices with applications. *Engineering analysis with boundary elements*, 27(5):405–422.

Borovicka, T., Jirina, M. J., Kordik, P., and Jirina, M. (2012). Selecting representative data sets. In Karahoca, A., editor, *Advances in data mining, knowledge discovery and applications*, pages 43–70. INTECH.

Bouezmarni, T., Ghouch, E., and Taamouti, A. (2013). Bernstein estimator for unbounded copula densities. *Statistics & Risk Modeling*, 30(4):343–360.

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.

Briol, F.-X., Oates, C., Girolami, M., and Osborne, M. (2015). Frank-Wolfe Bayesian Quadrature: Probabilistic Integration with Theoretical Guarantees. In *Advances in Neural Information Processing Systems.*

Browne, T., Iooss, B., Le Gratiet, L., Lonchampt, J., and Remy, E. (2016). Stochastic simulators based optimization by Gaussian process metamodels–application to maintenance investments planning issues. *Quality and Reliability Engineering International*, 32(6):2067–2080.

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.

Burton, T., Jenkins, N., Bossanyi, E., Sharpe, D., and Graham, M. (2021). *Wind energy handbook.* John Wiley & Sons.

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., Bourinet, J.-M., Morio, J., and Gayton, N. (2017). Evaluation of failure probability under parameter epistemic uncertainty: application to aerospace system reliability assessment. *Aerospace Science and Technology*, 69:526–537.

Chabridon, V., Balesdent, M., Perrin, G., Morio, J., Bourinet, J.-M., and Gayton, N. (2021). *Global Reliability-oriented Sensitivity Analysis under Distribution Parameter Uncertainty*, pages 237–277. Wiley Online Library.

Chen, T., Wang, X., Yuan, G., and Liu, J. (2018a). Fatigue bending test on grouted connections for monopile offshore wind turbines. *Marine Structures*, 60:52–71.

Chen, W. Y., Mackey, L., Gorham, J., Briol, F.-X., and Oates, C. (2018b). Stein points. In *International Conference on Machine Learning*, pages 844–853. PMLR.

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.

Clenshaw, C. and Curtis, A. (1960). A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.

Cottin, C. and Pfeifer, D. (2014). From Bernstein polynomials to Bernstein copulas. *J. Appl. Funct. Anal*, 9(3-4):277–288.

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.

Crombecq, K., Laermans, E., and Dhaene, T. (2011). Efficient space-filling and non-collapsing sequential design strategies for simulation-based modelling. *European Journal of Operational Research*, 214:683–696.

Csiszár, I. (1963). Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, 8:85–108.

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 Crécy, A., Bazin, P., Glaeser, H., Skorek, T., Joufcla, J., Probst, P., Fujioka, K., Chung, B., Oh, D., Kyncl, M., Pernica, R., Macek, J., Meca, R., Macian, R., D'Auria, F., Petruzzi, A., Batet, L., Perez, M., and Reventos, F. (2008). Uncertainty and sensitivity analysis of the LOFT L2-5 test: Results of the BEMUSE programme. *Nuclear Engineering and Design*, 12:3561–3578.

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.

De Stefano, M. and Der Kiureghian, A. (1990). An efficient algorithm for second-order reliability analysis. Technical report, Report No. UCB/SEMM-90/20. Dept of Civil and Environmental Engineering, University of California, Berkeley.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bulletins de l'Académie Royale de Belgique*, 65(1):274–292.

Demange-Chryst, 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. (2022). *Structural and system reliability*. Cambridge University Press.

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.

Dick, J. and Pillichshammer, F. (2010). *Digital nets and sequences: discrepancy theory and quasi–Monte Carlo integration*. Cambridge University Press.

Dimitrov, N. (2013). *Structural reliability of wind turbine blades: Design methods and evaluation*. PhD thesis, Technical University of Denmark.

Dimitrov, N., Kelly, M., Vignaroli, A., and Berg, J. (2018). From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases. *Wind Energy Science*, 3:767 – 790.

DNV-RP-C203 (2016). DNV-RP-C203: Fatigue design of offshore steel structures. Technical report, Det Norske Veritas.

DNV-ST-0437 (2016). DNV-ST-0437: Loads and site conditions for wind turbines. Technical report, Det Norske Veritas.

Doubrawa, P., Quon, E., Martinez-Tossas, L., Shaler, K., Debnath, M., Hamilton, N., Herges, T., Maniaci, D., Kelley, C., Hsieh, A., et al. (2020). Multimodel validation of single wakes in neutral and stratified atmospheric conditions. *Wind Energy*, 23(11):2027–2055.

Doubrawa, P., Shaler, K., and Jonkman, J. (2023). Difference in load predictions obtained with effective turbulence vs. a dynamic wake meandering modeling approach. *Wind Energy Science Discussions*, 2023:1–28.

Dowling, N. E. (1972). Fatigue Failure Predictions for Complicated Stress-Strain Histories. *Journal of Materials, JMLSA*, 7:71 – 87.

Dubourg, V. (2011). *Adaptive surrogate models for reliability analysis and reliability-based design optimization*. PhD thesis, Université Blaise Pascal.

Dubrule, O. (1983). Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15(6):687–699.

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.

Ehre, M., Papaioannou, I., and Straub, D. (2024). Variance-based reliability sensitivity with dependent inputs using failure samples. *Structural Safety*, 106:102396.

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.

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC.

Féjer, L. (1933). On the infinite sequences arising in the theories of harmonic analysis, of interpolation, and of mechanical quadratures. *Bulletin of the American Mathematical Society*, 39(8):521–534.

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 4^{th} 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., Chabridon, V., and Muré, J. (2022). Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool. In *Proceedings of the 5th International Conference on Renewable Energies Offshore*, pages 837–846.

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.

Fuhg., J., Fau, A., and Nackenhorst, U. (2021). State-of-the-Art and Comparative Review of Adaptive Sampling Methods for Kriging. *Archives of Computational Methods in Engineering*, 28:2689–2747.

Gauchy, C., Stenger, J., Sueur, R., and Iooss, B. (2022). An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91.

Geffraye, G., Antoni, O., Farvacque, M., Kadri, D., Lavialle, G., Rameau, B., and Ruby, A. (2011). CATHARE2 V2.5_2: A single version for various applications. *Nuclear Engineering and Design*, 241:4456–4463.

Genest, C., Nešlehová, J. G., and Rémillard, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *Journal of Multivariate Analysis*, 159:82–110.

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.

González-Barrios, J. and Hoyos-Argüelles, R. (2021). Estimating checkerboard approximations with sample d-copulas. *Communications in Statistics-Simulation and Computation*, 50(12):3992–4027.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT Press.

Graf, P., Stewart, G., Lackner, M., Dykes, K., and Veers, P. (2016). High-throughput computation and the applicability of Monte Carlo integration in fatigue load estimation of floating offshore wind turbines. *Wind Energy*, 19(5):861–872.

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.

Han, Q., Hao, Z., Hu, T., and Chu, F. (2018). Non-parametric models for joint probabilistic distributions of wind speed and direction data. *Renewable Energy*, 126:1032–1042.

Hansen, M. and Henriksen, L. (2013). *Basic DTU Wind Energy controller*. Number 0028 in DTU Wind Energy E. DTU Wind Energy.

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.

Hartigan, J. (1975). Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hawchar, L., El Soueidy, 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.

Heinrich, J. and Weiskopf, D. (2013). State of the art of parallel coordinates. *Eurographics (state of the art reports)*, pages 95–116.

Heredia-Zavoni, E. and Montes-Iturrizaga, R. (2022). Environmental contours using nonparametric copulas. *Ocean Engineering*, 266:112971.

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.

Huchet, Q., Mattrand, C., Beaurepaire, P., Relun, N., and Gayton, N. (2019). AK-DA: An efficient method for the fatigue assessment of wind turbine structures. *Wind Energy*, 22(5):638–652.

Huszár, F. and Duvenaud, D. (2012). Optimally-Weighted Herding is Bayesian Quadrature. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 377 – 386.

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. (2021). Sample selection from a given dataset to validate machine learning models. In *Proceedings of 50th Meeting of the Italian Statistical Society (SIS2021)*, pages 88–93, Pisa, Italy.

Iooss, B., Boussouf, L., Feuillard, V., and Marrel, A. (2010). Numerical studies of the metamodel fitting and validation processes. *International Journal of Advances in Systems and Measurements*, 3:11–21.

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.

Ishihara, T. and Qian, G. W. (2018). A new Gaussian-based analytical wake model for wind turbines considering ambient turbulence intensities and thrust coefficient effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 177:275–292.

Janssen, P., Swanepoel, J., and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, 142(5):1189 – 1197.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461.

Joe, H. (2014). *Dependence modeling with copulas.* CRC press.

Joe, H. and Kurowicka, D. (2011). *Dependence modeling: vine copula handbook.* World Scientific.

Johlas, H., Martínez-Tossas, L., Lackner, M., Schmidt, D., and Churchfield, M. (2020). Large eddy simulations of offshore wind turbine wakes for two floating platform types. In *Journal of Physics: Conference Series*, volume 1452, page 012034.

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.

Joseph, V. R. and Vakayil, A. (2022). SPlit: An optimal method for data splitting. *Technometrics*, 64(2):166–176.

Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards applied mathematics series*, 12:27–30.

Kaimal, J., Wyngaard, J., Izumi, Y., and Coté, O. (1972). Spectral characteristics of surface-layer turbulence. *Quarterly Journal of the Royal Meteorological Society*, 98(417):563–589.

Kanagawa, M. and Hennig, P. (2019). Convergence Guarantees for Adaptive Bayesian Quadrature Methods. In *Advances in Neural Information Processing Systems*, volume 32.

Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. (2018). Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. arXiv:1807.02582.

Kanner, S., Aubault, A., Peiffer, A., and Yu, B. (2018). Maximum dissimilarity-based algorithm for discretization of metocean data into clusters of arbitrary size and dimension. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 51319.

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.

Katsikogiannis, G., Sørum, S., Bachynski, E., and Amdahl, J. (2021). Environmental lumping for efficient fatigue assessment of large-diameter monopile wind turbines. *Marine Structures*, 77:102939.

Kennard, R. and Stone, L. (1969). Computer aided design of experiments. *Technometrics*, 11:137–148.

Kim, T., Natarajan, A., Lovera, A., Julan, E., Peyrard, E., Capaldo, M., Huwart, G., Bozonnet, P., and Guiton, M. (2022). A comprehensive code-to-code comparison study with the modified IEA15MW-UMaine Floating Wind Turbine for H2020 HIPERWIND project. *Journal of Physics: Conference Series*, 2265(4):042006.

Kiriliouk, A., Segers, J., and Tsukahara, H. (2021). *Resampling procedures with empirical beta copulas*, pages 27–53. Springer.

Klebanov, I., Schuster, I., and Sullivan, T. (2020). A rigorous theory of conditional mean embeddings. *SIAM Journal on Mathematics of Data Science*, 2(3):583–606.

Kleijnen, J. and Sargent, R. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120:14–29.

Koutsourelakis, P. (2004). Reliability of structures in high dimensions. Part II. Theoretical validation. *Probabilistic engineering mechanics*, 19(4):419–423.

Kucherenko, S., Feil, B., Shah, N., and Mauntz, W. (2011). The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering & System Safety*, 96:440–449.

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.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Kurtz, N. and Song, J. (2013). Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety*, 42:35–44.

Lacoste-Julien, S., Lindsten, F., and Bach, F. (2015). Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38, pages 544–552.

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.

Lasserre, J.-B. (2023). Chebyshev and equilibrium measure vs bernstein and lebesgue measure.

Lasserre, M. (2022). *Apprentissages dans les réseaux bayésiens à base de copules non-paramétriques*. PhD thesis, Sorbonne Université.

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).

Lebrun, R. and Dutfoy, A. (2009a). A generalization of the Nataf transformation to distributions with elliptical copula. *Probabilistic Engineering Mechanics*, 24(2):172–178.

Lebrun, R. and Dutfoy, A. (2009b). Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics*, 24(4):577–584.

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, W., Lu, L., Xie, X., and Yang, M. (2017). A novel extension algorithm for optimized Latin hypercube sampling. *Journal of Statistical Computation and Simulation*, 87:2549–2559.

Li, X., Mikusiński, P., and Taylor, M. D. (1998). Strong approximation of copulas. *Journal of Mathematical Analysis and Applications*, 225(2):608–623.

Li, X. and Zhang, W. (2020). Long-term fatigue damage assessment for a floating offshore wind turbine under realistic environmental conditions. *Renewable Energy*, 159:570–584.

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.

Lin, Y. and Dong, S. (2019). Wave energy assessment based on trivariate distribution of significant wave height, mean period and direction. *Applied Ocean Research*, 87:47–63.

Liu, X., Lu, C., Li, G., Godbole, A., and Chen, Y. (2017). Effects of aerodynamic damping on the tower load of offshore horizontal axis wind turbines. *Applied Energy*, 204:1101–1114.

Lovera, A., Fekhari, E., Jézéquel, B., Dupoiron, M., Guiton, M., and Ardillon, E. (2023). Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis. In *Journal of Physics: Conference Series*, volume 2505, page 012011. IOP Publishing.

Lüthen, N., Marelli, S., and Sudret, B. (2023). A spectral surrogate model for stochastic simulators computed from trajectory samples. *Computer Methods in Applied Mechanics and Engineering*, 406:115875.

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., and Chabridon, V. (2022). The ICSCREAM methodology: Identification of penalizing configurations in computer experiments using screening and metamodel – Applications in thermal-hydraulics. *Nuclear Science and Engineering*, 196:301–321.

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.

Melchers, R. (1989). Importance sampling in structural systems. *Structural safety*, 6(1):3–10.

Milano, D. (2021). *Numerical prototype for floating offshore wind turbines.* PhD thesis, The University of Edinburgh.

Montes-Iturrizaga, R. and Heredia-Zavoni, E. (2016). 'multivariate environmental contours using c-vine copulas'. *Ocean engineering*, 118:68–82.

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.

Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Advances in applied probability*, 29(2):429–443.

Müller, W. G. (2007). *Collecting Spatial Data.* Springer, 3rd edition.

Murcia, J., Réthoré, P., Dimitrov, N., Natarajan, A., Sørensen, J., Graf, P., and Kim, T. (2018). Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates. *Renewable Energy*, 119:910–922.

Müller, K. and Cheng, P. (2018). Application of a Monte Carlo procedure for probabilistic fatigue design of floating offshore wind turbines. *Wind Energy Science*, 3:149 – 162.

Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods. *Dependence Modeling*, 5:99–120.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I–A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Nelsen, R. (2006). *An introduction to copulas.* Springer.

Nielsen, J. and Sørensen, J. (2021). Risk-based derivation of target reliability levels for life extension of wind turbine structural components. *Wind Energy*, 24(9):939–956.

Nogales Gómez, A., Pronzato, L., and Rendas, M.-J. (2021). Incremental space-filling design based on coverings and spacings: improving upon low discrepancy sequences. *Journal of Statistical Theory and Practice*, 15(4):77.

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.

O'Hagan, A. (1991). Bayes-Hermite quadrature. *Journal of Statistical Planning and Inference*, 29:245–260.

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., Papadimitriou, C., and Straub, D. (2016). Sequential importance sampling for structural reliability analysis. *Structural safety*, 62:66–75.

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.

Park, H. and Jun, C. (2009). A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36:3336–3341.

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.

Pires, A., Moustapha, M., Marelli, S., and Sudret, B. (2023). Surrogate-based reliability analysis for noisy models. In *14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP14)*. Trinity College Dublin.

Plischke, E. and Borgonovo, E. (2019). Copula theory and probabilistic sensitivity analysis: Is there a connection? *European Journal of Operational Research*, 277(3):1046–1059.

Póczos, B., Ghahramani, Z., and Schneider, J. (2012). Copula-Based Kernel Dependency Measures. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, ICML'12, pages 1635–1642.

Powell, M. (1994). *A direct search optimization method that models the objective and constraint functions by linear interpolation.* Springer.

Pronzato, L. (2022). Performance analysis of greedy algorithms for minimising a Maximum Mean Discrepancy. preprint, https://hal.inria.fr/hal-03114891/.

Pronzato, L. and Müller, W. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22:681–701.

Pronzato, L. and Rendas, M.-J. (2023). Validation of Machine Learning Prediction Models. preprint, https://hal.science/hal-03818234.

Pronzato, L. and Zhigljavsky, A. (2020). Bayesian quadrature and energy minimization for space-filling design. *SIAM/ASA Journal on Uncertainty Quantification*, 8:959 – 1011.

Qian, P., Ai, M., and Wu, C. (2009). Construction of Nested Space-Filling Designs. *Annals of Statistics*, 37:3616–3643.

Qian, P. and Wu, C. (2009). Sliced space filling designs. *Biometrika*, 96:945–956.

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.

Rackwitz, R. (2001). Reliability analysis – a review and some perspectives. *Structural safety*, 23(4):365–395.

Raguet, H. and Marrel, A. (2018). Target and conditional sensitivity analysis with emphasis on dependence measures. arXiv preprint arXiv:1801.10047.

Rahman, S. (2016). The f-sensitivity index. *SIAM/ASA journal on uncertainty quantification*, 4(1):130–162.

Raoult, C., Joly, A., Andreevsky, M., and Joly-Laugel, A. (2018). Anemoc-3: amélioration de la base de données d'état de mer anemoc-2 par prise en compte des effets de la marée anemoc-3: improving the anemoc-2 sea state database by adding tide effects. In *16e journées de l'hydrodynamique.*

Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*, volume 1. Springer.

Rocher, B., Schoefs, S., François, M., Salou, A., and Caouissin, A.-L. (2020). A two-scale probabilistic time-dependent fatigue model for offshore steel wind turbines. *International Journal of Fatigue*, 136:105620.

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.

Rose, D. (2015). *Modeling and estimating multivariate dependence structures with the Bernstein copula.* PhD thesis, Ludwig Maximilian University of Munich.

Roustant, O., Barthe, F., and Iooss, B. (2017). Poincaré inequalities on intervals – application to sensitivity analysis. *Electronic Journal of Statistics*, 11(2):3081 – 3119.

Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, 7:387–412.

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.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer.* John Wiley & Sons.

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.

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments.* Springer.

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. (2019). *Surrogate models for uncertainty quantification in the context of imprecise probability modelling.* PhD thesis, ETH Zurich.

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.

Sheikholeslami, R. and Razavi, S. (2017). Progressive Latin hypercube sampling: An efficient approach for robust sampling-based analysis of environmental models. *Environmental Modelling & Software*, 93:109–126.

Silverman, B. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(1):97–99.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231.

Slot, R. M., Sørensen, J. D., Sudret, B., Svenningsen, L., and Thøgersen, M. L. (2020). Surrogate model uncertainty in wind turbine reliability assessment. *Renewable Energy*, 151:1150 – 1162.

Snee, R. (1977). Validation of regression models: Methods and examples. *Technometrics*, 19:415–428.

Sobol', 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., Fukumizu, K., Gretton, A., Schölkopf, B., and Lanckriet, G. (2012). On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550 – 1599.

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.

Sueur, R., Iooss, B., and Delage, T. (2017). Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case. preprint, https://arxiv.org/abs/1707.01296.

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.

Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143:1249 – 1272.

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.

Teixeira, R., Nogal, M., O'Connor, A., Nichols, J., and Dumas, A. (2019a). Stress-cycle fatigue design with Kriging applied to offshore wind turbines. *International Journal of Fatigue*, 125:454–467.

Teixeira, R., O'Connor, N., and Nogal, M. (2019b). Probabilistic sensitivity analysis of offshore wind turbines using a transformed Kullback-Leibler divergence. *Structural Safety*, 81:101860.

Teymur, O., Gorham, J., Riabiz, M., and Oates, C. (2021). Optimal Quantisation of Probability Measures Using Maximum Mean Discrepancy. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pages 1027 – 1035.

Thunnissen, D. P. (2005). *Propagating and mitigating uncertainty in the design of complex multidisciplinary systems.* PhD thesis, California Institute of Technology.

Torre, E., Marelli, S., Embrechts, P., and Sudret, B. (2019). A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. *Probabilistic Engineering Mechanics*, 55:1–16.

Trefethen, L. (2008). Is Gauss quadrature better than Clenshaw–Curtis? *SIAM review*, 50(1):67–87.

Uribe, F., Papaioannou, I., Marzouk, Y., and Straub, D. (2021). Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):818–847.

Van den Bos, L. (2020). *Quadrature Methods for Wind Turbine Load Calculations.* PhD thesis, Delft University of Technology.

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.

Vanem, E. (2016). Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*, 49:180–205.

Vanem, E., Fekhari, E., Dimitrov, N., Kelly, M., Cousin, A., and Guiton, M. (2023). A joint probability distribution model for multivariate wind and wave conditions. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 86847, page V002T02A013. American Society of Mechanical Engineers.

Vanem, E., Lande, O., and Fekhari, E. (2024). A simulation study on the usefulness of the Bernstein copula for statistical modeling of metocean variables. International Conference on Offshore Mechanics and Arctic Engineering (to appear).

Veers, P., Dykes, K., Lantz, E., Barth, S., Bottasso, C., Carlson, O., Clifton, A., Green, J., Green, P., Holttinen, H., et al. (2019). Grand challenges in the science of wind energy. *Science*, 366(6464):eaau2027.

Velarde, J., Kramhøft, C., and Sørensen, J. (2019). Global sensitivity analysis of offshore wind turbine foundation fatigue loads. *Renewable Energy*, 140:177 – 189.

Velarde, J., Kramhøft, C., Sørensen, J., and Zorzi, G. (2020). Fatigue reliability of large monopiles for offshore wind turbines. *International journal of fatigue*, 134:105487.

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, H., Gramstad, O., Schär, S., Marelli, S., and Vanem, E. (2023). Comparison of Probabilistic Structural Reliability Methods for Ultimate Limit State Assessment of Wind Turbines. *arXiv preprint arXiv:2312.04972.*

Wang, L., Kolios, A., Liu, X., Venetsanos, D., and Cai, R. (2022). Reliability of offshore wind turbine support structures: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 161:112250.

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.

Waskom, M. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

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.

Wilkie, D. and Galasso, C. (2021). Gaussian process regression for fatigue reliability analysis of offshore wind turbines. *Structural Safety*, 88:102020.

Wise, A. and Bachynski, E. (2020). Wake meandering effects on floating wind turbines. *Wind Energy*, 23(5):1266–1285.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Xu, Y. and Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2:249–262.

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.

Zhu, X. and Sudret, B. (2023). Stochastic polynomial chaos expansions to emulate stochastic simulators. *International Journal for Uncertainty Quantification*, 13(2).

Zong, H. and Porté-Agel, F. (2020). A momentum-conserving wake superposition method for wind farm power prediction. *Journal of Fluid Mechanics*, 889:A8.

Zwick, D. and Muskulus, M. (2015). The simulation error caused by input loading variability in offshore wind turbine structural analysis. *Wind Energy*, 18:1421 – 1432.

Appendix A

Univariate distribution fitting

This appendix recalls the main methods to infer a univariate distribution considering a n-sized i.i.d. sample $X_n = \left\{x^{(1)}, \ldots, x^{(n)}\right\} \in \mathbb{R}^n$. Inference techniques are split into two main groups, the parametric ones assuming that the underlying distribution belongs to a parametric distribution, and the nonparametric ones otherwise. In general, nonparametric methods require a large amount of data but allow more flexibility. In practice, nontrivial distributions (e.g., multimodal) might be easier to model using nonparametric approaches.

To assess the quality of this inference, a panel of goodness-of-fit methods are proposed (Saporta, 2006), this appendix recalls a few of them.

Main parametric methods

Moments method

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

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

one can solve the system of equations $\left(m(\boldsymbol{\theta}) = \widehat{m}_n;\ \sigma^2(\boldsymbol{\theta}) = \widehat{\sigma}_n^2\right)$ to determine the optimal set of parameters $\boldsymbol{\theta}$ in this situation. Note that some families of distributions might be more suited to this method because of the analytical expression of their moments. However, 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, maximizes a given correspondence metric between the dataset X_n and a parametric distribution with density $f_X(\boldsymbol{\theta})$. This metric is the *likelihood* function, defined as:

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

with the PDF taking the set of parameters $\boldsymbol{\theta}$ written: $f_X(x^{(i)};\boldsymbol{\theta})$. 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 $\widehat{\boldsymbol{\theta}}^*$ of parameters minimizing the following expression:

$$\widehat{\boldsymbol{\theta}}^* = \underset{\boldsymbol{\theta}\in\mathcal{D}_{\boldsymbol{\theta}}}{\arg\min} \left(- \sum_{i=1}^{n} \ln(f_X(x^{(i)};\boldsymbol{\theta})) \right). \tag{A.3}$$

Fig. A.1 illustrates the MLE by considering the fit of two distinct Weibull distributions w.r.t. a small set of observations $X_n = \{1, 2, 3, 4, 6\}$ (represented by the black bars on the x-axis). Note that the quick analytical results from the moment method can be used as a starting point of the MLE optimization.

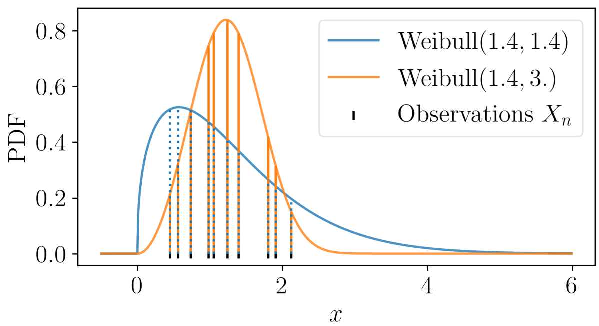


Figure A.1 Adequation of two different Weibull models using their likelihood with a sample of observations (black crosses).

Main nonparametric methods

Empirical CDF and histogram

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

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

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.

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 $k : \mathbb{R} \to \mathbb{R}_+$ and a scaling parameter $h > 0$, also called bandwidth, the kernel density estimator is defined as:

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

Different types of kernels are used for KDE, such as the standard normal, triangular, Epachnikov or uniform. The choice of bandwidth results in a bias-variance trade-off, that has been extensively discussed in the literature (Wand and Jones, 1994). Fig. A.2 illustrates the KDE for three different scaling parameters $h \in \{0.1, 0.2, 0.4\}$ applied on a set of observations (represented by the black bars on the x-axis).

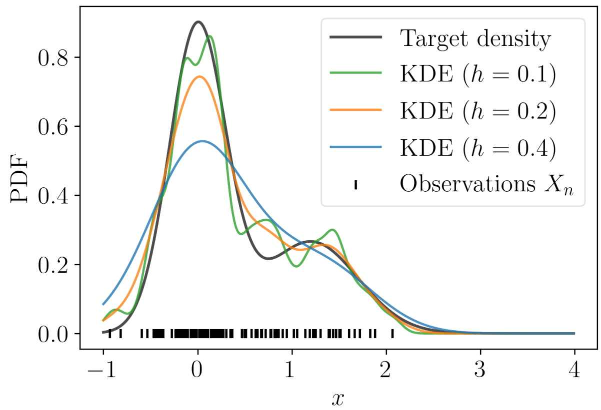


Figure A.2 Fit of a bimodal density by KDE using different tuning parameters.

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 many 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.

The AIC and BIC are expressed as follows:

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

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.

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 $\left\{\alpha^{(i)}\right\}_{i=1}^{n} = \left\{\widehat{F}_X(x^{(i)})\right\}_{i=1}^{n}$. The following Fig. A.3 is a QQ-plot of the KDE model fitted in Fig. A.2. The closer the scatter plot gets to the first bisector line the better the fit is.

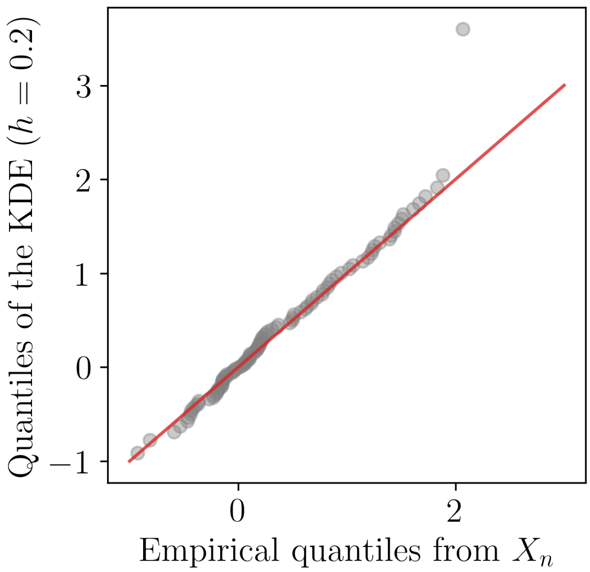


Figure A.3 QQ-plot between the data from Fig. A.2 and a KDE model.

Appendix B

Dissimilarity measures between probability distributions

Beyond the discrepancy measure to the uniform distribution presented in Section 1.5, this Appendix introduces two families of measures commonly used to quantify the dissimilarity between two probability distributions π and ζ.

Csizár f-divergences

The general definition of the "f-divergence" of π from ζ, also called Ciszár f-divergence after the seminal work of Csiszár (1963), is given by:

$$D_f(\pi||\zeta) = \int_\Omega f\left(\frac{\mathrm{d}\pi}{\mathrm{d}\zeta}\right)\mathrm{d}\zeta, \tag{B.1}$$

with $f(\cdot)$ and convex function such that $f(1) = 0$, and where π is absolutely continuous w.r.t. ζ. Let us recall some well-know divergences which are special cases of f-divergences:

- Kullback–Leibler (Kullback and Leibler, 1951): $f(t) = t\ln(t)$;
- Hellinger of order $\alpha \in \mathbb{R}_+\backslash\{1\}$ (Jeffreys, 1946): $f(t) = (1-\sqrt{t})$;
- Total variation: $f(t) = \frac{1}{2}|t-1|$.

The reader may refer to Basu et al. (2011) for further details on f-divergences.

Integral probability metrics

Another family of dissimilarity measures between probability distributions called the "integral probability metrics" (IPM) (Müller, 1997), is defined as:

$$\gamma_{\mathcal{H}}(\pi,\zeta) = \sup_{g\in\mathcal{H}}\left|\int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x})\mathrm{d}\pi(\mathbf{x}) - \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x})\mathrm{d}\zeta(\mathbf{x})\right|, \tag{B.2}$$

where $\mathcal{H}$ is a class of measurable functions on $\mathcal{D}_{\mathbf{X}} \subset \mathbb{R}^d$ that sets the type of IPM. For example, the total variation distance considers all the functions with value in $[-1, 1]$; furthermore, the Wasserstein distance relies on a class of Lipschitz functions; then a kernel-based distance called the "maximum mean discrepancy" (MMD) uses a specific Hilbert space.

Between the wide panel of distances[1], a particular focus is dedicated in this section to the MMD. This distance was successfully use in diverse contexts for the consistency of its estimators, and its closed-form expression (even allow exact computation is some cases) (Sriperumbudur et al., 2012).

Kernel discrepancy

This section first introduces a kernel-based discrepancy called the maximum mean discrepancy, generalizing the concept of discrepancy to non-uniform measures.

Reproducing kernel Hilbert space and kernel mean embedding Let us first assume that the function g belongs 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 $g : \mathcal{D}_{\mathbf{X}} \to \mathbb{R}$. Considering a symmetric and positive definite function $k : \mathcal{D}_{\mathbf{X}} \times \mathcal{D}_{\mathbf{X}} \to \mathbb{R}$, later called a "reproducing kernel" or simply a "kernel", an RKHS verifies the following axioms:

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

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:

$$P_\pi(\mathbf{x}) = \int_{\mathcal{D}_{\mathbf{X}}} k(\mathbf{x}, \mathbf{x}') \mathrm{d}\pi(\mathbf{x}'). \tag{B.3}$$

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:

$$P_{\zeta_n}(\mathbf{x}) = \int_{\mathcal{D}_{\mathbf{X}}} k(\mathbf{x}, \mathbf{x}') \mathrm{d}\zeta_n(\mathbf{x}') = \sum_{i=1}^{n} w_i k(\mathbf{x}, \mathbf{x}^{(i)}). \tag{B.4}$$

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

[1] See the "Taxonomy of principal distances and divergences" proposed by F. Nielsen in: https://franknielsen.github.io/Divergence/Poster-Distances.pdf

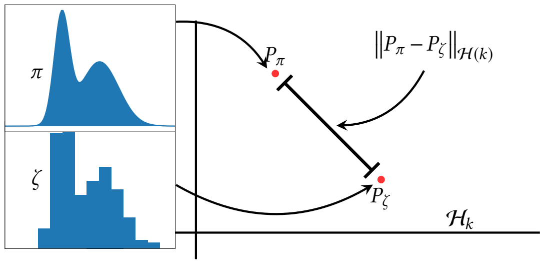


Figure B.1 Kernel mean embedding of a continuous and discrete probability distribution

support is the current design $\mathbf{X}_n$. 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. 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, Sec. 3.4), (Pronzato and Zhigljavsky, 2020, Sec. 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.

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:

$$\varepsilon_\pi = \int_{\mathcal{D}_\mathbf{X}} P_\pi(\mathbf{x}) \mathrm{d}\pi(\mathbf{x}) = \iint_{\mathcal{D}_\mathbf{X}^2} k(\mathbf{x}, \mathbf{x}')\, \mathrm{d}\pi(\mathbf{x}) \mathrm{d}\pi(\mathbf{x}'). \tag{B.5}$$

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. (3.6) (see Briol et al., 2019):

$$\left|\sum_{i=1}^n w_i g(\mathbf{x}^{(i)}) - \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\pi(\mathbf{x})\right| = \left|\left\langle g, P_{\zeta_n}(\mathbf{x})\right\rangle_{\mathcal{H}(k)} - \langle g, P_\pi(\mathbf{x})\rangle_{\mathcal{H}(k)}\right| \tag{B.6a}$$

$$= \left|\left\langle g, \left(P_{\zeta_n}(\mathbf{x}) - P_\pi(\mathbf{x})\right)\right\rangle_{\mathcal{H}(k)}\right| \tag{B.6b}$$

$$\leq \|g\|_{\mathcal{H}(k)} \left\|P_\pi(\mathbf{x}) - P_{\zeta_n}(\mathbf{x})\right\|_{\mathcal{H}(k)}. \tag{B.6c}$$

Maximum mean discrepancy A metric of discrepancy 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:

$$\mathrm{MMD}(\pi, \zeta) = \sup_{\|g\|_{\mathcal{H}(k)} \leq 1} \left| \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\pi(\mathbf{x}) - \int_{\mathcal{D}_\mathbf{X}} g(\mathbf{x}) \mathrm{d}\zeta(\mathbf{x}) \right| \tag{B.7}$$

According to the inequality in Eq. (B.6c), $\mathrm{MMD}(\pi, \zeta) = \left\|P_\pi - P_\zeta\right\|_{\mathcal{H}(k)}$, 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: $\mathrm{MMD}(\pi, \zeta) = 0 \Leftrightarrow \pi = \zeta$. This property makes the MMD a metric on $\mathcal{D}_\mathbf{X}$. The squared MMD has been used for various purposes such as statistical testing (Gretton et al., 2006), numerical integration (Chen et al., 2010), and global sensitivity analysis (Da Veiga, 2015). It can be written as follows:

$$\begin{aligned}
\mathrm{MMD}(\pi,\zeta)^2 &= \left\| P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x}) \right\|^2_{\mathcal{H}(k)} && \text{(B.8a)}\\
&= \left\langle \left(P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})\right), \left(P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})\right) \right\rangle_{\mathcal{H}(k)} && \text{(B.8b)}\\
&= \langle P_\pi(\mathbf{x}), P_\pi(\mathbf{x}) \rangle_{\mathcal{H}(k)} - 2\left\langle P_\pi(\mathbf{x}), P_\zeta(\mathbf{x}) \right\rangle_{\mathcal{H}(k)} + \left\langle P_\zeta(\mathbf{x}), P_\zeta(\mathbf{x}) \right\rangle_{\mathcal{H}(k)} && \text{(B.8c)}\\
&= \iint_{\mathcal{D}^2_{\mathbf{X}}} k(\mathbf{x},\mathbf{x}')\,\mathrm{d}\pi(\mathbf{x})\mathrm{d}\pi(\mathbf{x}') - 2\iint_{\mathcal{D}^2_{\mathbf{X}}} k(\mathbf{x},\mathbf{x}')\,\mathrm{d}\pi(\mathbf{x})\mathrm{d}\zeta(\mathbf{x}') + \iint_{\mathcal{D}^2_{\mathbf{X}}} k(\mathbf{x},\mathbf{x}')\,\mathrm{d}\zeta(\mathbf{x})\mathrm{d}\zeta(\mathbf{x}'). && \text{(B.8d)}
\end{aligned}$$

Appendix C

Rare event estimation algorithm

Subset simulation (SS)

The next algorithm describes the subset simulation methods introduced by Au and Beck (2001).

Algorithm 2 Subset simulation (SS).

▷ ***Inputs:***
$f_{\mathbf{X}}$, joint PDF of the inputs
$g(\cdot)$, LSF
$y_{\text{th}} \in \mathbb{R}$, threshold defining the failure event
N, number of samples per iteration
$m \in \mathbb{N}$, parameter of the EBC fitting
$p_0 \in]0, 1[$, empirical quantile order (rarity parameter)
▷ ***Algorithm:***
Set $k = 0$ and $f_{[0]} = f_{\mathbf{X}}$
Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{[0]}$
Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^{N}$
Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $G_{[0],N}$
while $\widehat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**
  Subsample $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \widehat{q}_{[k]}^{p_0}\}_{j=1}^{n}$
  Fit marginals of the subset $\mathbf{A}_{[k+1],n}$ by KDE $\{\widehat{F}_i\}_{i=1}^{d}$
  Fit the copula of the subset $\mathbf{A}_{[k+1],n}$ by EBC $B_{\mathbf{m}}(C_n)$
  Build a CDF $\widehat{F}_{[k+1]}(\mathbf{x}) = B_{\mathbf{m}}(C_n)(\widehat{F}_1(x_1), \dots, \widehat{F}_d(x_d))$
  Sample $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} \widehat{f}_{[k+1]}$
  Sample by MCMC $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.d}}{\sim} f_{\mathbf{X}|F_{[k+1]}}$ (with $\mathbf{A}_{[k+1],n}$ as initialization points)
  Evaluate $G_{[k+1],N} = \{g(\mathbf{X}_{[k+1]}^{(j)})\}_{j=1}^{N}$
  Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$
  Set $k = k + 1$
end while
Set total iteration number $k_\# = k - 1$
Estimate $\widehat{p_{\text{f}}} = (1 - p_0)^{k_\#} \cdot \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{X}_{[k_\#]}^{(j)}) \geq y_{\text{th}}\}}(\mathbf{X}_{[k_\#]^{(j)}})$
▷ ***Outputs:***
$\widehat{p_{\text{f}}}$, estimate of p_{f}

Appendix D

Uncertainty quantification practice with OpenTURNS

The present Appendix presents minimalistic Python/OpenTURNS examples implementing some of the uncertainty quantification methods presented in Chapter 1.

OpenTURNS 1 (Bivariate distribution). Definition of a probabilistic uncertainty model.

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

OpenTURNS 2 (Numerical integration). Construction of multivariate quadrature rules.

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

OpenTURNS 3 (Design of experiments). Construction of LHS and optimized LHS w.r.t. to a space-filling metric (e.g., L2-centered discrepancy) by simulated annealing algorithm.

```
#!/usr/bin/python3
import openturns as ot
marginals=[ot.Uniform(0.0, 1.0), ot.Uniform(0.0, 1.0)]
distribution=ot.ComposedDistribution(marginals)
# Build a LHS
n=10
LHS_exp=ot.LHSExperiment(distribution, n)
LHS_design=LHS_exp.generate()
# Build an optimized LHS using L2-centered discrepancy
LHS_exp=ot.LHSExperiment(distribution, n)
SF_metric=ot.SpaceFillingC2()
SA_profile=ot.GeometricProfile(10., 0.95, 20000)
LHS_opt=ot.SimulatedAnnealingLHS(LHS_exp, SF_metric, SA_profile)
LHS_opt.generate()
LHS_design=LHS_opt.getResult().getOptimalDesign()
```

OpenTURNS 4 (Rare event estimation). Estimation of rare events with various methods.

```
import openturns as ot
marginals=[ot.Normal(0.0, 1.0), ot.Exponential(1.0)]
distribution=ot.ComposedDistribution(marginals)
# Build a limit-state function and failure event
g=ot.SymbolicFunction(["x1", "x2"], ["(x1 - x2) ^ 2"])
X=ot.RandomVector(distribution)
Y=ot.CompositeRandomVector(g, X)
failure_event=ot.ThresholdEvent(Y, ot.LessOrEqual(), 0.)
# Estimate pf using FORM
starting_p=distribution.getMean()
FORM_algo=ot.FORM(ot.Cobyla(), failure_event, starting_p)
FORM_algo.run()
FORM_results=FORM_algo.getResult()
design_point=FORM_results.getStandardSpaceDesignPoint()
FORM_pf=FORM_results.getEventProbability()
# Estimate pf using Monte Carlo
MC_exp=ot.MonteCarloExperiment()
MC_algo=ot.ProbabilitySimulationAlgorithm(failure_event, MC_exp)
MC_algo.run()
MC_results=MC_algo.getResult()
MC_pf=MC_results.getProbabilityEstimate()
MC_pf_confidence=MC_results.getConfidenceLength(0.95)
# Estimate pf using importance sampling
aux_distribution=ot.Normal(design_point, [1.0, 1.0])
standard_event=ot.StandardEvent(failure_event)
IS_exp=ot.ImportanceSamplingExperiment(aux_distribution)
IS_algo=ot.ProbabilitySimulationAlgorithm(standard_event, IS_exp)
IS_algo.run()
IS_results=IS_algo.getResult()
IS_pf=IS_results.getProbabilityEstimate()
IS_pf_confidence=IS_results.getConfidenceLength(0.95)
# Estimate pf using subset simulation
SS_algo=ot.SubsetSampling(failure_event)
SS_algo.run()
SS_results=SS_algo.getResult()
SS_pf=SS_results.getProbabilityEstimate()
SS_pf_confidence=SS_results.getConfidenceLength(0.95)
```

OpenTURNS 5 (Sobol' indices). Estimation of the Sobol' indices to assess global sensitivity analysis on the Ishigami analytical problem.

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

OpenTURNS 6 (Gaussian process regression). Fit of an ordinary Kriging model fitting.

```
#!/usr/bin/python3
import openturns as ot
g=ot.SymbolicFunction(['x'], ['x * sin(x) + sin(6 * x)'])
x_train=ot.Uniform(0., 12.).getSample(7) # n=7
y_train=g(x_train)
basis=ot.ConstantBasisFactory(1).build() # d=1
cov_model=ot.MaternModel([1.], 1.5)
algo=ot.KrigingAlgorithm(x_train,y_train,cov_model,basis)
algo.run()
Kriging_results=algo.getResult()
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 à 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 :

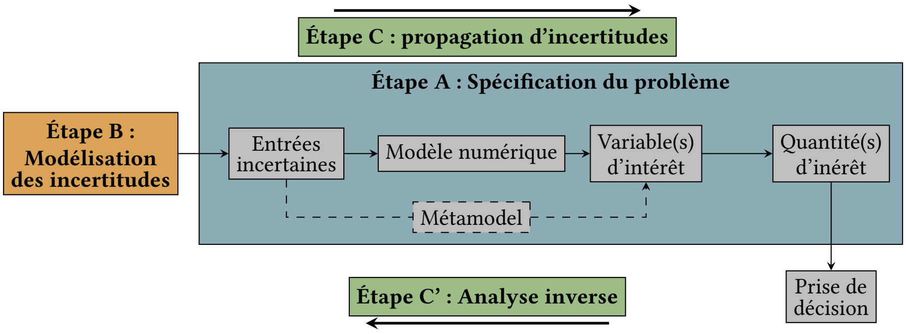


FIGURE E.1 Schéma générique de la quantification des incertitudes (De Rocquigny et al. (2008), adapté par Ajenjo (2023))

- 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 prédiction 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 (2024). "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 Sancetta 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 14th 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 Indepencence 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.	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.
Jour. Pap.	E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2024). "Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature". In : *Data-Centric Engineering*.
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 : *Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)*, 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"[1]. In : *Proceedings of the 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)*, 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 : *Proceedings of the ASME 2023 42th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2023)*, Melbourne, Australia. (Paper)

	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 (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 : *SIAM Conference on Uncertainty Quantification (SIAM UQ22)*, Atlanta, USA. (Talk)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). "Model predictivity assessment : incremental test-set selection and accuracy evaluation". In : *22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)*, 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 : *Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)*, Clermont-Ferrand, France (Poster)
	E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2023). "Rare event estimation using nonparametric Bernstein adaptive sampling". In : *Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)*, 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)

[1]This 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 au Chapitre 5.
- 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.

[2] Documentation :https://efekhari27.github.io/otkerneldesign/master/
[3] Dépôt: https://github.com/efekhari27/bancs
[4] Repository: https://github.com/efekhari27/ctbenchmark
[5] Repository: https://github.com/efekhari27/copulogram