

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 fulfilment of the requirements for the degree of

Doctor of Philosophy
(in Computer Science)

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

Director	Dr. Bertrand IOOSS	Senior Researcher	EDF R&D, Chatou
Co-advisors	Dr. Vincent CHABRIDON	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	ix
List of Tables	xi
List of Acronyms	xiii
Introduction	1
I Introduction to uncertainty quantification and wind energy	11
II Contributions to uncertainty quantification and propagation	13
III Contributions to rare event estimation	15
1 Reliability and robustness analysis of an offshore wind turbine subject to fatigue	17
1.1 Introduction	18
1.2 Surrogate modeling for reliability analysis	19
1.2.1 High-performance computer evaluation	19
1.2.2 Design of experiments	20
1.2.3 Gaussian process regression	20
1.3 Reliability and robustness analysis	23
1.3.1 Nominal reliability analysis	23
1.3.2 Robustness analysis using perturbed-law sensitivity indices	24
1.4 Conclusion	26
Bibliography	29
Appendix A Univariate distribution fitting	35

Appendix B Dissimilarity measures between probability distributions	39
Appendix C Rare event estimation algorithm	43
Appendix D Uncertainty quantification practice with OpenTURNS	45
Appendix E Résumé étendu de la thèse	49

List of Figures

1	General uncertainty quantification framework (De Rocquigny et al., 2008, adapted by Ajenjo, 2023)	3
1.1	Learning set of the mean damage surrogate model. A Halton sequence is first built (in blue) and completed by a set KH-generated points (in orange) in a subdomain defined a priori (in gray).	20
1.2	Normalized mean damage evaluated on the composite design illustrated in Fig. 1.1.b.	21
1.3	Three-dimensional plot of the surrogate model \tilde{D} (in blue) and learning set (in black).	22
1.4	Cross-section of the surrogate model \tilde{D} (in shades of blue) for given values of K_{soil} and Θ_{yaw} . The darker the shade, the closer to the cross-section.	22
1.5	Leave-one-out validation results of the surrogate model \tilde{D}	22
1.6	Perturbations in terms of standard deviation of a lognormal distribution (left) and a truncated normal distribution (right).	25
1.7	Perturbed-law based indices for relative perturbations of the standard deviations of $(K_{\text{soil}}, \Theta_{\text{yaw}}, \varepsilon)$. The failure probabilities studied are each estimated by FORM-IS method with sample size $N = 5 \times 10^4$	25
1.8	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 $N = 5 \times 10^4$	26
A.1	Adequation of two different Weibull models using their likelihood with a sample of observations (black crosses).	36
A.2	Fit of a bimodal density by KDE using different tuning parameters.	37
A.3	QQ-plot between the data from Fig. A.2 and a KDE model.	38
B.1	Kernel mean embedding of a continuous and discrete probability distribution .	41
E.1	Schéma générique de la quantification des incertitudes (De Rocquigny et al. (2008), adapté par Ajenjo (2023))	52

List of Tables

1.1 Nominal reliability analysis (size $N = 5 \times 10^4$ for IS and SS. $p_0 = 0.1$ for SS). . .	24
--	----

List of Acronyms

IPM Integral probability metrics. [39](#)

ROSA Reliability-Oriented Sensitivity Analysis. [8](#)

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 known 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.

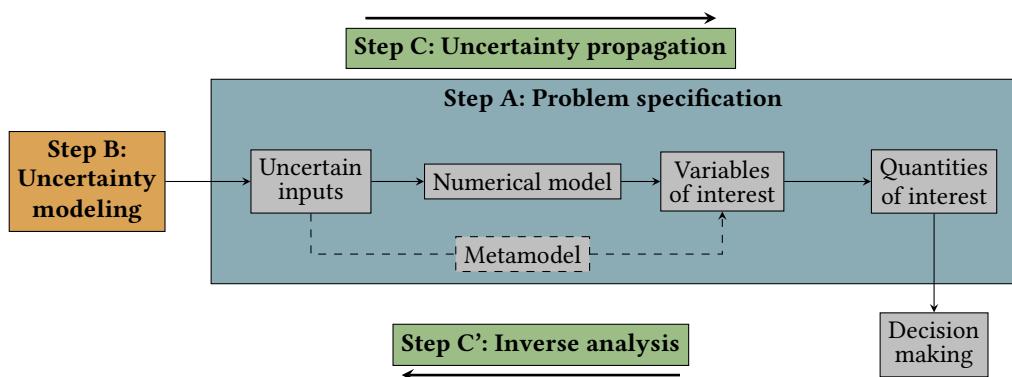


Figure 1 General uncertainty quantification framework (De Rocquigny et al., 2008, adapted by Ajenjo, 2023)

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 HIPERWIND¹ (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

¹HIPERWIND 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 perturbing 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.	<u>E. Fekhari</u> , B. Iooss, J. Muré, L. Pronzato and M. J. Rendas (2023). “Model predictivity assessment: incremental test-set selection and accuracy evaluation”. In: <i>Studies in Theoretical and Applied Statistics</i> , pages 315–347. Springer.
Jour. Pap.	<u>E. Fekhari</u> , V. Chabridon, J. Muré and B. Iooss (2024). “Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature”. In: <i>Data-Centric Engineering</i> , In press. <u>E. Vanem</u> , <u>E. Fekhari</u> , N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2024). “A joint probability distribution for multivariate wind-wave conditions and discussions on uncertainties”. In: <i>Journal of Offshore Mechanics and Arctic Engineering</i> , In press.
Int. Conf. Pap.	<u>E. Fekhari</u> , M. Baudin, V. Chabridon, and Y. Jebroun (2021). “otbenchmark: an open source Python package for benchmarking and validating uncertainty quantification algorithms”. In: <i>Proceedings of the 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021)</i> , Athens, Greece. (Paper & Talk)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). “Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool”. In: <i>Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)</i> , Lisbon, Portugal. (Paper & Talk)
	<u>E. Fekhari</u> , V. Chabridon, J. Muré and B. Iooss (2023). “Bernstein adaptive nonparametric conditional sampling: a new method for rare event probability estimation” ² . In: <i>Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)</i> , Dublin, Ireland. (Paper & Talk)
	<u>A. Lovera</u> , <u>E. Fekhari</u> , B. Jézéquel, M. Dupoirion, M. Guiton and E. Ardillon (2023). “Quantifying and clustering the wake-induced perturbations within a wind farm for load analysis”. In: <i>Journal of Physics: Conference Series (WAKE 2023)</i> , Visby, Sweden. (Paper)
	<u>E. Vanem</u> , Ø. Lande, <u>E. Fekhari</u> (2024). “A joint probability distribution model for multivariate wind and wave conditions”. In: <i>Proceedings of the ASME 2024 43th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2024)</i> , Singapore. (Paper to appear)
Int. Conf. Short Abs.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). “Numerical Studies of Bayesian Quadrature Applied to Offshore Wind Turbine Load Estimation”. In: <i>SIAM Conference on Uncertainty Quantification (SIAM UQ22)</i> , Atlanta, USA. (Talk)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). “Model predictivity assessment: incremental test-set selection and accuracy evaluation”. In: <i>22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)</i> , Trondheim, Norway. (Talk)
Nat. Conf.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). “Kernel-based quadrature applied to offshore wind turbine damage estimation”. In: <i>Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)</i> , Clermont-Ferrand, France. (Poster)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2023). “Rare event estimation using nonparametric Bernstein adaptive sampling”. In: <i>Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)</i> , Le Croisic, France. (Talk)

Invited Lec. Le Printemps de la Recherche 2022, Nantes, France. «Traitement des incertitudes pour la gestion d'actifs éoliens». (Talk)

Journées Scientifiques de l'Eolien 2024, Saint-Malo, France. «Evaluation probabiliste de la fiabilité en fatigue des structures éoliennes en mer». (Talk)

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:

- This Python package generates designs of experiments based on kernel methods such as kernel herding (Chen et al., 2018) and support points (Mak and Joseph, 2018). `otkerneldesign`³ 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.

²Rewarded by the “CERRA Student Recognition Award” (<https://icasp14.com/presenter/awards/>)

³Documentation: <https://efekhari27.github.io/otkerneldesign/master/>

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

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

⁶Repository: <https://github.com/efekhari27/copulogram>

PART I:

INTRODUCTION TO UNCERTAINTY QUANTIFICATION AND WIND ENERGY

Toute pensée émet un coup de dé.

S. MALLARMÉ

PART II:

CONTRIBUTIONS TO UNCERTAINTY QUANTIFICATION AND PROPAGATION

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

VOLTAIRE

PART III:

CONTRIBUTIONS TO RARE EVENT ESTIMATION

La résignation est un suicide quotidien.

H. BALZAC

Chapter **1**

Reliability and robustness analysis of an offshore wind turbine subject to fatigue

1.1	Introduction	18
1.2	Surrogate modeling for reliability analysis	19
1.2.1	High-performance computer evaluation	19
1.2.2	Design of experiments	20
1.2.3	Gaussian process regression	20
1.3	Reliability and robustness analysis	23
1.3.1	Nominal reliability analysis	23
1.3.2	Robustness analysis using perturbed-law sensitivity indices	24
1.4	Conclusion	26

1.1 Introduction

One of the main goals of this work is to evaluate the reliability of an OWT's monopile foundation w.r.t. fatigue solicitations. Let us recall the approach to assess fatigue damage over the structure's lifetime (see [IEC-61400-1, 2019](#), Appendix H). First, the lifetime before decommissioning time $t_d \in \mathbb{R}_+$ forms the interval $T = [0, t_d]$, which can be used to define a probability space $(T, \mathcal{B}(T), \mathcal{U}(T) = \lambda/t_d)$. Then, for all $t \in T$ and assuming the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the random vector of the environmental conditions $\mathbf{X}(t, \cdot)$ is a measurable function on $(\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Therefore, one can define the product probability space $(T \times \Omega, \mathcal{B}(T) \otimes \mathcal{A}, \mathcal{U}(T) \otimes \mathbb{P})$ and the random vector \mathbf{X} , which is a measurable function on $(T \times \Omega, \mathcal{B}(T) \otimes \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For a realization of \mathbf{X} , denoted by $\mathbf{X}(t^{(i)}, \omega^{(r)})$, and a given set of parameters $\mathbf{z} = (k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon)$ related to the system (defined in Table ??), one can perform a 10-minute Turbsim-DIEGO simulation (see Subsection ??) to obtain the corresponding cumulative damage:

$$d_c^{10\text{min}} \left(\mathbf{X}(t^{(i)}, \omega^{(r)}) \mid \mathbf{Z} = \mathbf{z} \right). \quad (1.1)$$

In practice, T is often discretized into $N_T \in \mathbb{N}$ 10-minute intervals $\{t^{(i)}\}_{i=1}^{N_T}$ ([IEC-61400-1, 2019](#), Appendix H). The 10-minute 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, which is defined in Table ??). To cumulate the damage over the lifetime, one can write the sum of 10-minute damages, each averaged over $n_{\text{rep}} \in \mathbb{N}$ pseudo-random seed repetitions:

$$D(\mathbf{z}) = N_T \mathbb{E} [d_c^{10\text{min}}(\mathbf{X} \mid \mathbf{Z} = \mathbf{z})] \quad (1.2a)$$

$$\approx N_T \frac{1}{N_T n_{\text{rep}}} \sum_{i=1}^{N_T} \sum_{r=1}^{n_{\text{rep}}} d_c^{10\text{min}} \left(\mathbf{X}(t^{(i)}, \omega^{(r)}) \mid \mathbf{Z} = \mathbf{z} \right). \quad (1.2b)$$

In Chapter ??, KH 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 QMC for estimating the lifetime damage in Eq. (1.2b), while being more flexible than QMC. In the linear cumulative damage model typically used by the community (Miner's rule), a damage value higher than one leads to ruin of the structure by convention. The present chapter assesses the probability of such a rare event, considering both the environmental uncertainties (aggregated according to Eq. (1.2b)) and the uncertainties related to the system itself (described by the random vector $\mathbf{Z} \in \mathcal{D}_Z$ with joint PDF f_Z). Assuming that the critical damage D_{cr} is a random variable centered around one (equivalent to a “resistance” variable in the well-known “resistance-solicitation” paradigm in reliability analysis [Lemaire et al., 2009](#)), this failure probability is written as:

$$p_f = \int_{\mathcal{D}_Z} \mathbb{I}_{\{D(\mathbf{z}) \geq D_{\text{cr}}\}} f_Z(\mathbf{z}) dz. \quad (1.3)$$

Less information is available to define the probabilistic model of the system uncertainties than the environmental ones. Therefore, the robustness analysis of the failure probability w.r.t. the probabilistic model Z should be studied. To do so, a perturbation-based approach using the *perturbed-law based sensitivity indices* (PLI), originally introduced by (Lemaître et al., 2015), is used in this chapter. However, the reliability and robustness analysis requires a number of evaluations of the Turbsim-DIEGO simulator $D(\cdot)$ that would be prohibitive without the use of surrogate models.

The present chapter is structured as follows: Section 1.2 presents the construction of a surrogate model of $D(\cdot)$, then Subsection 1.3.1 analyses the reliability of the monopile foundation of Teesside's turbine for a nominal distribution of Z , and finally Subsection 1.3.2 proposes a robustness analysis of p_f by perturbing the laws of both Z and D_{cr} .

1.2 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 how it is used to build a learning set for a GP regression model.

1.2.1 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 Z while the inner loop concerns the environmental conditions and their repetitions for n_{rep} different pseudo-random seeds. At this stage, the goal is to approximate $D(\cdot)$ for any value of $\mathbf{z} = (k_{soil}, \theta_{yaw}, \varepsilon = 1)$, since ε is a coefficient applied to the damage as a post-processing (see Subsection ??). Using a KH design to explore the environmental conditions as discussed in Chapter ??, this approximation is written as:

$$D(\mathbf{z}) \approx D^{KH}(\mathbf{z}) = N_T \frac{1}{n_X n_{rep}} \sum_{i=1}^{n_X} \sum_{r=1}^{n_{rep}} d_c^{10\min} \left(\mathbf{x}^{(i)}, \omega^{(r)} | \mathbf{Z} = \mathbf{z} = (k_{soil}, \theta_{yaw}, \varepsilon = 1) \right), \quad (1.4)$$

where a KH design with size $n_X \in \mathbb{N}$ is denoted by $\{\mathbf{x}^{(i)}\}_{i=1}^{n_X}$. According to the convergence results obtained in Chapter ??, the KH size is fixed at $n_X = 200$ and the repetitions at $n_{rep} = 11$, which implies a total of 2200 Turbsim-DIEGO simulations per evaluation of the function $D^{KH}(\cdot)$. In this setup, the CRONOS HPC from EDF R&D 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. (1.4). Note that the cumulative damage studied is actually the maximum value of $D^{KH}(\cdot)$ over the discretized azimuth angles (illustrated in Fig. ??), at the mudline level.

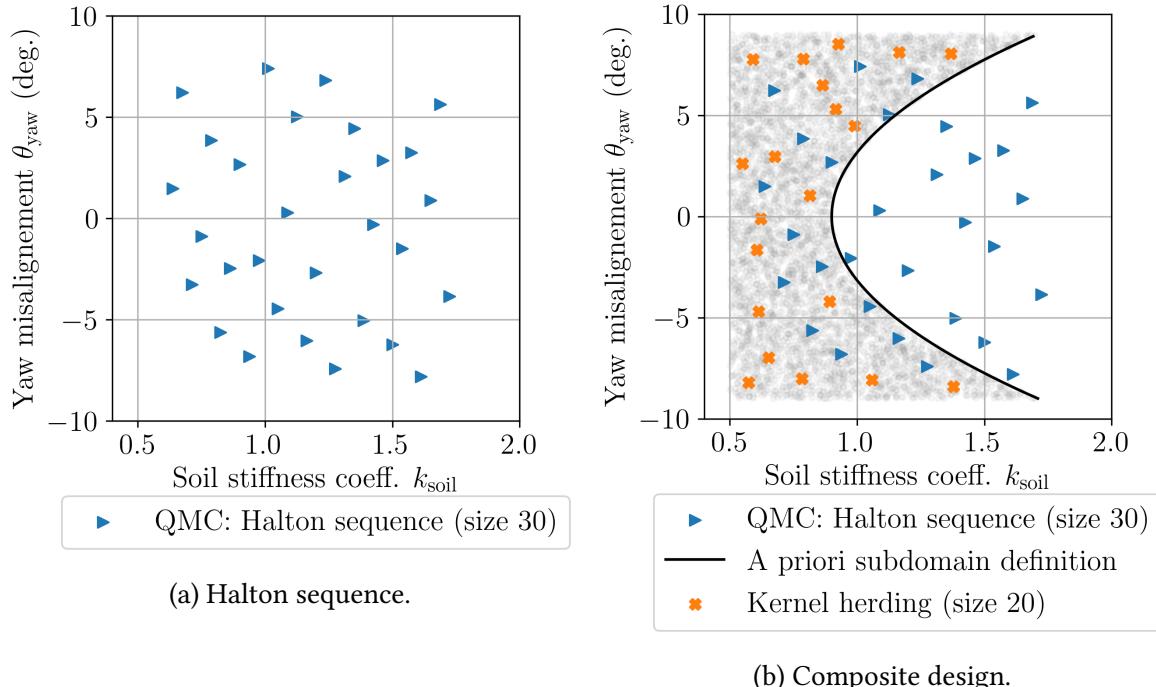


Figure 1.1 Learning set of the mean damage surrogate model. A Halton sequence is first built (in blue) and competed by a set KH-generated points (in orange) in a subdomain defined a priori (in gray).

1.2.2 Design of experiments

To build a learning set, a space-filling design of experiments is created on the joint domain of $(K_{\text{soil}}, \Theta_{\text{yaw}})$. This design was first composed of 30 points generated by a Halton sequence (illustrated in Fig. 1.1.a) which was the most space-filling sequential method in two dimensions (compared to other QMC sequences and KH). Its evaluation and analysis showed that the highest damage values are the result of high absolute values of Θ_{yaw} or low values of K_{soil} . Therefore, the design was completed in a second phase by 20 points targeting these areas by applying kernel herding to a subdomain defined a priori (see the candidate set represented by the gray points in Fig. 1.1.b). Finally, the learning set is the union of the two complementary designs, later referred as the “composite design” (see Fig. 1.1.b). This composite design, denoted by Z_{n_Z} , has a size of $n_Z = 50$ points, which represents over 10^5 Turbsim-DIEGO simulations (each requiring around 45 minutes of CPU time). Fig. 1.2 shows the lifetime cumulated damage evaluated on the composite design (with normalized values corresponding to the color scale).

1.2.3 Gaussian process regression

A GP regression with Matérn 5/2 and constant trend is fitted on the composited design Z_{n_Z} according to the Kriging equations introduced in Section ???. The resulting surrogate model $\tilde{D} : \mathbb{R}^2 \rightarrow \mathbb{R}$, is represented by the blue three-dimensional surface in Fig. 1.3, and its learning set 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. 1.4.a, and $k_{\text{soil}} = 1$ on Fig. 1.4.b. On these two

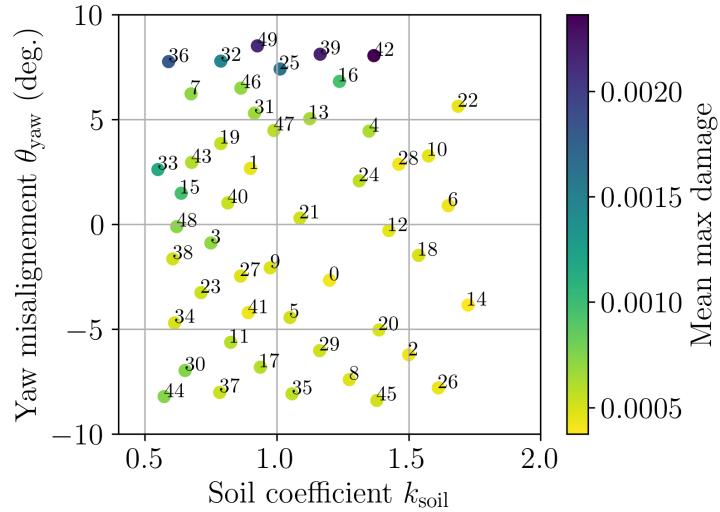


Figure 1.2 Normalized mean damage evaluated on the composite design illustrated in Fig. 1.1.b.

figures, learning points are plotted in a gray scale and the surrogate model in a blue scale. The darker the shade, the closer to the cross-section the points are.

One can first notice that the damage is not symmetric w.r.t. the yaw misalignment θ_{yaw} . When the nacelle perfectly aligns itself with the wind direction, this angle is equal to zero. For this bottom-fixed OWT, introducing a yaw error of the same amplitude has more impact in one direction than the other. In the meantime, the soil stiff k_{soil} has, as expected, a monotonic influence on the damage.

To validate this surrogate model, a leave-one-out (LOO) procedure is realized. Fig. 1.5.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 some areas (as revealed by the cross-section in Fig. 1.4.b). Fig. 1.5.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 $\widehat{Q}_{\text{LOO}}^2 = 0.72$ is considered acceptable in this small data context, especially as the LOO procedure was shown to underestimate the true performance metric in Chapter ???. However, the surrogate could be enhanced by adding points to the learning set in areas with high nonlinearities.

Remark 1. Active learning methods for reliability assessment (see Subsection ??) could be a great option for a such costly computer model. 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.

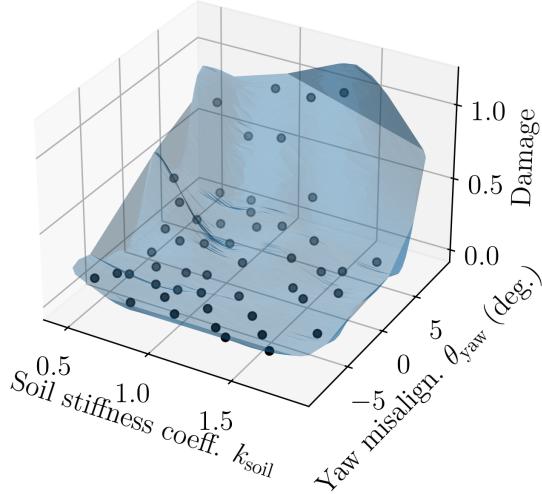


Figure 1.3 Three-dimensional plot of the surrogate model \tilde{D} (in blue) and learning set (in black).

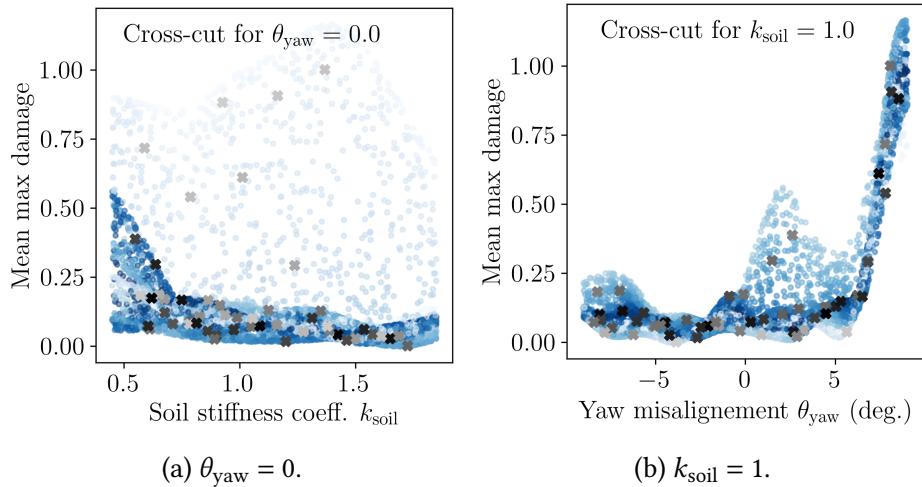


Figure 1.4 Cross-section of the surrogate model \tilde{D} (in shades of blue) for given values of K_{soil} and Θ_{yaw} . The darker the shade, the closer to the cross-section.

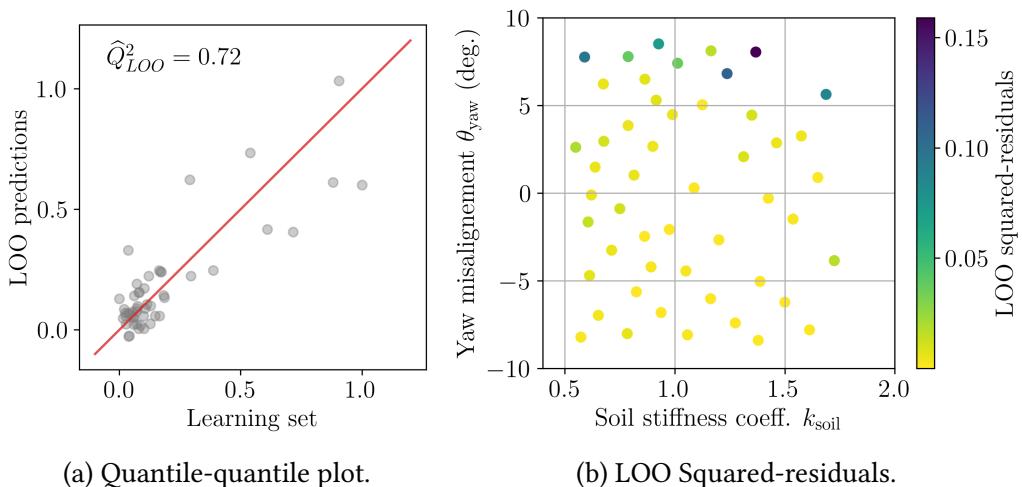


Figure 1.5 Leave-one-out validation results of the surrogate model \tilde{D} .

1.3 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 recommended 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 reliability analysis is conducted using the surrogate model of the lifetime cumulative damage previously constructed in Subsection 1.2.3. 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. \(2015\)](#).

1.3.1 Nominal reliability analysis

The S-N curve uncertainty can be expressed as a simple coefficient ε (see Subsection ??), which is applied to the lifetime cumulative damage as follows:

$$D(k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon) = \frac{1}{\varepsilon} D(k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon = 1). \quad (1.5)$$

Therefore, the surrogate model of the lifetime cumulative damage \tilde{D} is modified without updating its learning such as:

$$\tilde{D}'(\mathbf{z}) = \tilde{D}'(k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon) = \frac{1}{\varepsilon} \tilde{D}(k_{\text{soil}}, \theta_{\text{yaw}}, \varepsilon = 1). \quad (1.6)$$

The probability introduced in Eq. (1.3) then becomes:

$$p'_f = \int_{\mathcal{D}_Z} \mathbb{1}_{\{\tilde{D}'(\mathbf{z}) \geq D_{\text{cr}}\}} f_Z(\mathbf{z}) d\mathbf{z}. \quad (1.7)$$

Table 1.1 presents the estimates of this quantity by various methods (FORM, FORM-IS and SS, see Section ??) and for two different distributions 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 methods deliver similar values of p_f even if FORM-IS converges faster than SS when comparing the coefficients of variation (COV). The adequation of FORM with the simulation methods reveals that the LSF in this case is almost linear. This linearity is mostly due to the nature of the LSF, following a resistance-solicitation paradigm as introduced earlier. 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 w.r.t. the probabilistic model of both D_{cr} and Z .

Table 1.1 Nominal reliability analysis (size $N = 5 \times 10^4$ for IS and SS, $p_0 = 0.1$ for SS).

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

1.3.2 Robustness analysis using 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 perturbing input densities. The goal is to assess the robustness of a quantity of interest (here, a failure probability) w.r.t. these perturbations. This type of index is for example used in studies of thermal-hydraulic numerical models for nuclear safety ([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 (KL) divergence. For example, a relative mean perturbation is defined as:

$$f_{j\delta} = \arg \min_{\substack{\pi \in \mathcal{P}, \\ \text{s.t. } \mathbb{E}_\pi[Z_j] = \mathbb{E}_{f_j}[Z_j](1+\delta)}} \text{KL}(\pi || f_j), \quad (1.8)$$

where $\delta \in \mathbb{R}$ denotes the relative perturbation, and \mathcal{P} is a family of distributions.

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 (e.g., when perturbing the mean of a uniform distribution). Separate approaches exist to define perturbations for the problem studied, such as [Lemaître et al. \(2015\)](#), or [Gauchy et al. \(2022\)](#). However, to ease the computation in the following, the perturbations 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 perturbation on the quantity of interest:

$$\text{PLI}(f_{j\delta}) = \frac{p_{f,j\delta} - p_f}{p_f}, \quad (1.9)$$

where $p_{f,j\delta}$ is the probability obtained when injecting $f_{j\delta}$ in Eq. (1.7). In the following, each variable in 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. 1.6, 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 .

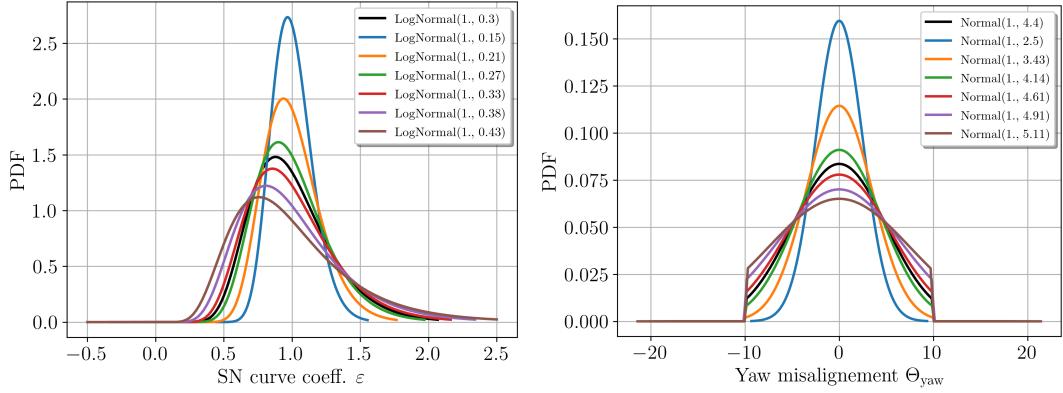


Figure 1.6 Perturbations in terms of standard deviation of a lognormal distribution (left) and a truncated normal distribution (right).

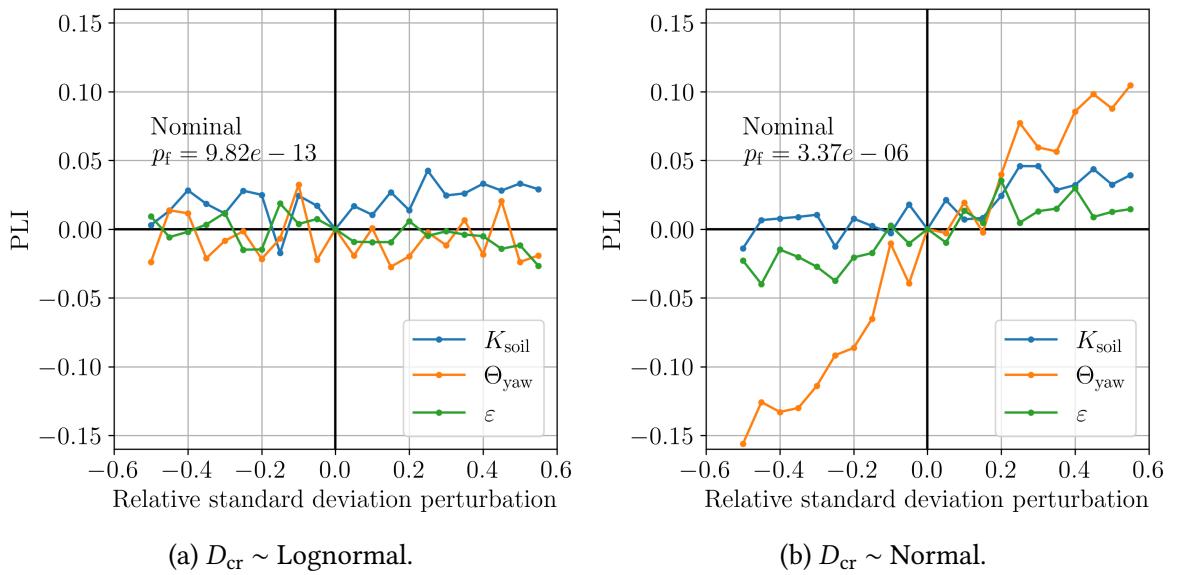


Figure 1.7 Perturbed-law based indices for relative perturbations of the standard deviations of $(K_{soil}, \Theta_{yaw}, \varepsilon)$. The failure probabilities studied are each estimated by FORM-IS method with sample size $N = 5 \times 10^4$.

The resulting PLI are presented in Fig. 1.7 for relative perturbations of the standard deviations of $(K_{soil}, \Theta_{yaw}, \varepsilon)$. 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 perturbing the standard deviation of the resistance variable D_{cr} , the same phenomenon is witnessed in Fig. 1.8. The perturbations have nearly no consequences, assuming that $D_{cr} \sim \text{LogNormal}$, but a lot of influence when $D_{cr} \sim \text{Normal}$. As a perspective, this study could be completed by a joint perturbation of both standard deviation and mean of the resistance variable.

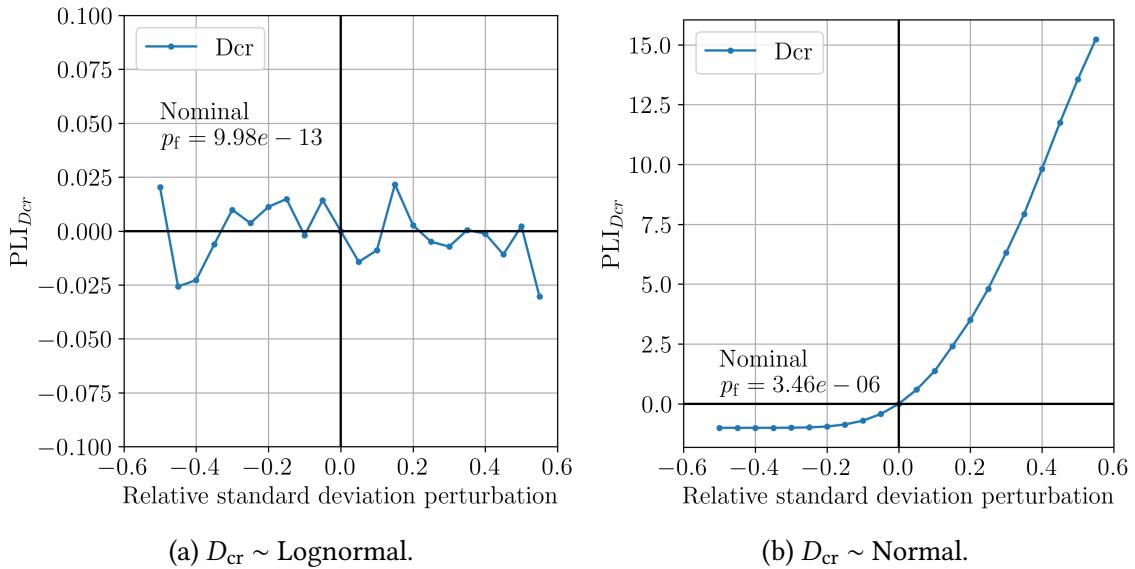


Figure 1.8 Perturbed-law based indices for relative perturbations of the standard deviation of D_{cr} . The failure probabilities studied are each estimated by FORM-IS method with sample size $N = 5 \times 10^4$.

1.4 Conclusion

In this chapter, a fully probabilistic reliability analysis was performed on the monopile foundation of an OWT located in Teesside, UK. This first reliability study required a significant computational effort (i.e., over 10^5 Turbsim-DIEGO simulations), made possible with the development of a tailored wrapper deployed on a HPC facility. These simulations served the construction of a surrogate model emulating the lifetime damage function.

The second section of this chapter addressed the estimation of a failure probability using this surrogate model. A general conclusion is that this probability is highly dependent of the probabilistic model describing the critical damage. To challenge the modeling assumptions of the critical damage and the system variables, a robustness analysis was realized. It consists in studying the impact of perturbing 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 current standards ([Wang et al., 2022](#)), which is in favor of lifetime extension. However, the results of this demonstrative study may be questionable with respect to several assumptions that have been made:

- The periods in parked position which could significantly increase the damage (see [Velarde et al., 2020](#)). Several 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, which might increase fatigue;
- 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 drawn from both studies.

As a final perspective, a stochastic surrogate model could be built by considering all the damages without averaging them over the environmental repetitions. The risk assessment would then be conducted on a stochastic function. In this particular context, quantile estimation was studied by [Browne et al. \(2016\)](#), however, rare event estimation on stochastic functions remains an open question recently discussed by [Pires et al. \(2023\)](#).

Bibliography

- Ajenjo, A. (2023). *Info-gap robustness assessment of reliability evaluations for the safety of critical industrial systems*. PhD thesis, Université Bourgogne Franche-Comté.
- 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.
- 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.
- 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.
- Binois, M., Huang, J., Gramacy, R., and Ludkovski, M. (2019). Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics*, 61(1):7–23.
- 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.
- 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.
- 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.
- 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.

- Chabridon, V. (2018). *Reliability-oriented sensitivity analysis under probabilistic model uncertainty—Application to aerospace systems*. PhD thesis, Université Clermont Auvergne.
- Chen, T., Wang, X., Yuan, G., and Liu, J. (2018). Fatigue bending test on grouted connections for monopile offshore wind turbines. *Marine Structures*, 60:52–71.
- 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.
- 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.
- Csiszár, I. (1963). Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. *Magyer Tud. Akad. Mat. Kutato Int. Kozl.*, 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., 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.
- De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley & Sons.
- DNV-ST-0437 (2016). DNV-ST-0437: Loads and site conditions for wind turbines. Technical report, Det Norske Veritas.
- 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.
- Fang, K., Liu, M.-Q., Qin, H., and Zhou, Y.-D. (2018). *Theory and application of uniform experimental designs*, volume 221. Springer.
- Fekhari, E., Iooss, B., Muré, J., Pronzato, L., and Rendas, J. (2023). 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.
- Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.
- 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.
- 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.
- 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.

- Huchet, Q. (2019). *Kriging based methods for the structural damage assessment of offshore wind turbines*. PhD thesis, Université Blaise Pascal.
- IEC-61400-1 (2019). IEC 61400-1: Wind energy generation systems - Part 1: Design requirements. Technical report, International Electrotechnical Commission (IEC).
- 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.
- 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.
- 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.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.
- Larsen, G. C., Madsen, H., Thomsen, K., and Larsen, T. (2008). Wake meandering: a pragmatic approach. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 11(4):377–395.
- Lataniotis, C. (2019). *Data-driven uncertainty quantification for high-dimensional engineering problems*. PhD thesis, ETH Zürich.
- Le Maître, O. and Knio, O. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media.
- 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.
- 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.
- 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.
- Mak, S. and Joseph, V. (2018). Support points. *The Annals of Statistics*, 46:2562 – 2592.
- 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.
- Milano, D. (2021). *Numerical prototype for floating offshore wind turbines*. PhD thesis, The University of Edinburgh.
- Morio, J. and Balesdent, M. (2015). *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach*. Woodhead Publishing, Elsevier.
- Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Advances in applied probability*, 29(2):429–443.

- 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.
- 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.
- 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.
- Pronzato, L. and Zhigljavsky, A. (2020). Bayesian quadrature and energy minimization for space-filling design. *SIAM/ASA Journal on Uncertainty Quantification*, 8:959 – 1011.
- Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, 7:387–412.
- Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(3):535–562.
- Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.
- 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.
- Shang, B. and Apley, D. (2020). Fully-sequential space-filling design algorithms for computer experiments. *Journal of Quality Technology*, 53:1 – 24.
- 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.
- 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.
- 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.
- 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.
- Wand, M. and Jones, M. (1994). *Kernel smoothing*. CRC press.
- 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.

Zhang, P. (1996). Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(435):1245–1253.

Zhu, X. (2022). *Surrogate Modeling for Stochastic Simulators Using Statistical Approaches*. PhD thesis, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich.

Appendix **A**

Univariate distribution fitting

This appendix recalls the main methods to infer a univariate distribution considering a n -sized i.i.d. sample $X_n = \{x^{(1)}, \dots, x^{(n)}\} \in \mathbb{R}^n$. 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 $f_X(\theta)$, whose first moments (e.g., $m(\theta)$ and $\sigma^2(\theta)$) match the empirical moments of the sample X_n (e.g., \widehat{m}_{X_n} and $\widehat{\sigma}^2$). After computing the empirical moments:

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

one can solve the system of equations $(m(\theta) = \widehat{m}_n; \sigma^2(\theta) = \widehat{\sigma}_n^2)$ to determine the optimal set of parameters θ in this situation. 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 $f_X(\theta)$. This metric is the *likelihood* function, defined as:

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

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

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

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.

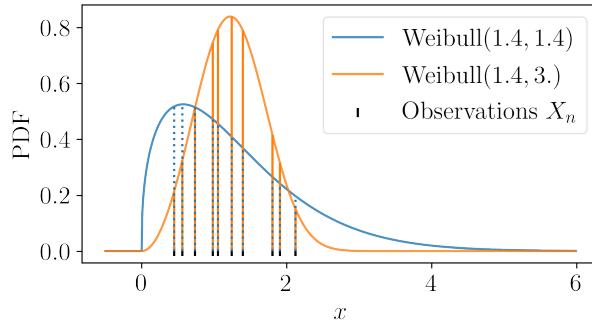


Figure A.1 Adequation of two different Weibull models using their likelihood with a sample of observations (black crosses).

Main nonparametric methods

Empirical CDF and histogram

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

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

A histogram consists of sorting and gathering the observations in a sample X_n into a finite number of categories. These categories are called bins and each regroups the same number of observations (identical binwidth). The number of bins is the only tuning parameter of this method.

Kernel density estimation

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

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

Different types of kernels are used for KDE, such as the standard normal, triangular, Epanchennikov 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).

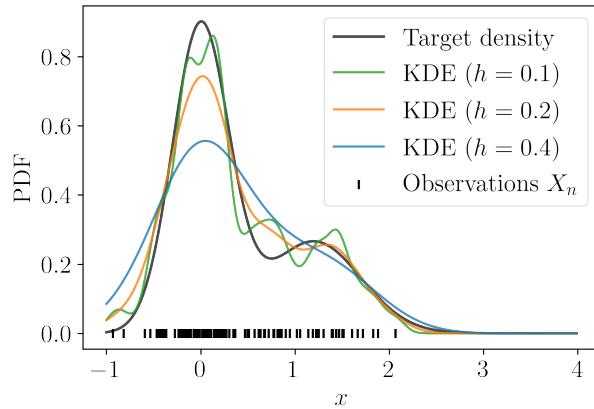


Figure A.2 Fit of a bimodal density by KDE using different tuning parameters.

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:

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

with the likelihood $\mathcal{L}(\boldsymbol{\theta}|X_n)$ 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 $\{\alpha^{(i)}\}_{i=1}^n = \{\widehat{F}_X(x^{(i)})\}_{i=1}^n$. 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.

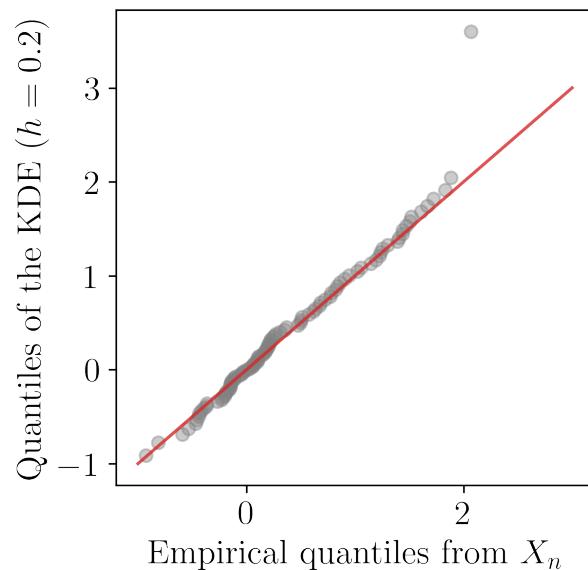


Figure A.3 QQ-plot between the data from Fig. A.2 and a KDE model.

Appendix B

Dissimilarity measures between probability distributions

Beyond the discrepancy measure to the uniform distribution presented in Section ??, 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 Csiszár f -divergence after the seminal work of [Csiszár \(1963\)](#), is given by:

$$D_f(\pi || \zeta) = \int_{\Omega} f\left(\frac{d\pi}{d\zeta}\right) d\zeta, \quad (\text{B.1})$$

with $f(\cdot)$ a convex function such that $f(1) = 0$, and where π is absolutely continuous w.r.t. ζ . Let us recall some well-known divergences which are special cases of f -divergences:

- Kullback–Leibler ([Kullback and Leibler, 1951](#)): $f(t) = t \ln(t)$;
- Hellinger of order $\alpha \in \mathbb{R}_+ \setminus \{1\}$ ([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:

$$\gamma_{\mathcal{H}}(\pi, \zeta) = \sup_{g \in \mathcal{H}} \left| \int_{\mathcal{D}_X} g(x) d\pi(x) - \int_{\mathcal{D}_X} g(x) d\zeta(x) \right|, \quad (\text{B.2})$$

where \mathcal{H} is a class of measurable functions on $\mathcal{D}_X \subset \mathbb{R}^d$ 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 distances¹, a particular focus is dedicated in this section to the MMD. This distance was successfully used in diverse contexts for the consistency of its estimators, and its closed-form expression (even allow exact computation in 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 $g : \mathcal{D}_X \rightarrow \mathbb{R}$. Considering a symmetric and positive definite function $k : \mathcal{D}_X \times \mathcal{D}_X \rightarrow \mathbb{R}$, later called a “reproducing kernel” or simply a “kernel”, an RKHS verifies the following axioms:

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

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

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

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:

$$P_{\zeta_n}(\mathbf{x}) = \int_{\mathcal{D}_X} k(\mathbf{x}, \mathbf{x}') d\zeta_n(\mathbf{x}') = \sum_{i=1}^n w_i k(\mathbf{x}, \mathbf{x}^{(i)}). \quad (\text{B.4})$$

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

¹See the “Taxonomy of principal distances and divergences” proposed by F. Nielsen in: <https://franknielsen.github.io/Divergence/Poster-Distances.pdf>

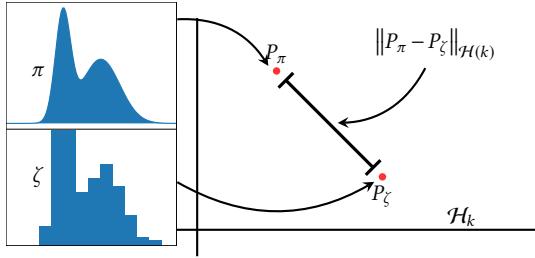


Figure B.1 Kernel mean embedding of a continuous and discrete probability distribution

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. (2023). 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:

$$\varepsilon_\pi = \int_{\mathcal{D}_X} P_\pi(\mathbf{x}) d\pi(\mathbf{x}) = \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\pi(\mathbf{x}'). \quad (\text{B.5})$$

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. (??) (see Briol et al., 2019):

$$\left| \sum_{i=1}^n w_i g(\mathbf{x}^{(i)}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\pi(\mathbf{x}) \right| = \left| \langle g, P_{\zeta_n}(\mathbf{x}) \rangle_{\mathcal{H}(k)} - \langle g, P_\pi(\mathbf{x}) \rangle_{\mathcal{H}(k)} \right| \quad (\text{B.6a})$$

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

$$\leq \|g\|_{\mathcal{H}(k)} \|P_\pi(\mathbf{x}) - P_{\zeta_n}(\mathbf{x})\|_{\mathcal{H}(k)}. \quad (\text{B.6c})$$

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 :

$$\text{MMD}(\pi, \zeta) = \sup_{\|g\|_{\mathcal{H}(k)} \leq 1} \left| \int_{\mathcal{D}_X} g(\mathbf{x}) d\pi(\mathbf{x}) - \int_{\mathcal{D}_X} g(\mathbf{x}) d\zeta(\mathbf{x}) \right| \quad (\text{B.7})$$

According to the inequality in Eq. (B.6c), $\text{MMD}(\pi, \zeta) = \|P_\pi - P_\zeta\|_{\mathcal{H}(k)}$, meaning that the MMD fully relies on the difference of potentials. Moreover, Sriperumbudur et al. (2010) defines a kernel as “characteristic kernel” when the following equivalence is true: $\text{MMD}(\pi, \zeta) = 0 \Leftrightarrow \pi = \zeta$. This property makes the MMD a metric on \mathcal{D}_X . The squared MMD has been used for 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:

$$\text{MMD}(\pi, \zeta)^2 = \|P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})\|_{\mathcal{H}(k)}^2 \quad (\text{B.8a})$$

$$= \langle (P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})), (P_\pi(\mathbf{x}) - P_\zeta(\mathbf{x})) \rangle_{\mathcal{H}(k)} \quad (\text{B.8b})$$

$$= \langle P_\pi(\mathbf{x}), P_\pi(\mathbf{x}) \rangle_{\mathcal{H}(k)} - 2 \langle P_\pi(\mathbf{x}), P_\zeta(\mathbf{x}) \rangle_{\mathcal{H}(k)} + \langle P_\zeta(\mathbf{x}), P_\zeta(\mathbf{x}) \rangle_{\mathcal{H}(k)} \quad (\text{B.8c})$$

$$= \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\pi(\mathbf{x}') - 2 \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\pi(\mathbf{x}) d\zeta(\mathbf{x}') + \iint_{\mathcal{D}_X^2} k(\mathbf{x}, \mathbf{x}') d\zeta(\mathbf{x}) d\zeta(\mathbf{x}'). \quad (\text{B.8d})$$

Appendix **C**

Rare event estimation algorithm

Subset simulation (SS)

The next algorithm describes the subset simulation methods introduced by [Au and Beck \(2001\)](#).

Algorithm 1 Subset simulation (SS).

▷ **Inputs:**

f_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 $f_{[0]} = f_X$

Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^N \stackrel{\text{i.i.d.}}{\sim} f_{[0]}$

Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^N$

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

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

Subsample $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \hat{q}_{[k]}^{p_0}\}_{j=1}^n$

Fit marginals of the subset $\mathbf{A}_{[k+1],n}$ by KDE $\{\hat{F}_i\}_{i=1}^d$

Fit the copula of the subset $\mathbf{A}_{[k+1],n}$ by EBC $B_m(C_n)$

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

Sample $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^N \stackrel{\text{i.i.d.}}{\sim} \hat{f}_{[k+1]}$

Sample by MCMC $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^N \stackrel{\text{i.d.}}{\sim} f_{\mathbf{X}|F_{[k+1]}}$ (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 $\hat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$

Set $k = k + 1$

end while

Set total iteration number $k_{\#} = k - 1$

Estimate $\hat{p}_f = (1 - p_0)^{k_{\#}} \cdot \frac{1}{N} \sum_{j=1}^N \mathbb{1}_{\{g(\mathbf{X}_{[k_{\#}]}^{(j)}) \geq y_{\text{th}}\}} (\mathbf{X}_{[k_{\#}]}^{(j)})$

▷ **Outputs:**

\hat{p}_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 ??.

OpenTURNS 1 (Bivariate distribution). Definition of a probabilistic uncertainty model.

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

OpenTURNS 2 (Numerical integration). Construction of multivariate quadrature rules.

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

OpenTURNS 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.

```

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

OpenTURNS 4 (Rare event estimation). Estimation of rare events with various methods.

```

1  import openturns as ot
2  marginals=[ot.Normal(0.0, 1.0), ot.Exponential(1.0)]
3  distribution=ot.ComposedDistribution(marginals)
4  # Build a limit-state function and failure event
5  g=ot.SymbolicFunction(["x1", "x2"], ["(x1 - x2) ^ 2"])
6  X=ot.RandomVector(distribution)
7  Y=ot.CompositeRandomVector(g, X)
8  failure_event=ot.ThresholdEvent(Y, ot.LessOrEqual(), 0.)
9  # Estimate pf using FORM
10 starting_p=distribution.getMean()
11 FORM_algo=ot.FORM(ot.Cobyla(), failure_event, starting_p)
12 FORM_algo.run()
13 FORM_results=FORM_algo.getResult()
14 design_point=FORM_results.getStandardSpaceDesignPoint()
15 FORM_pf=FORM_results.getEventProbability()
16 # Estimate pf using Monte Carlo
17 MC_exp=ot.MonteCarloExperiment()
18 MC algo=ot.ProbabilitySimulationAlgorithm(failure_event, MC_exp)
19 MC algo.run()
20 MC_results=MC algo.getResult()
21 MC_pf=MC_results.getProbabilityEstimate()
22 MC_pf_confidence=MC_results.getConfidenceLength(0.95)
23 # Estimate pf using importance sampling
24 aux_distribution=ot.Normal(design_point, [1.0, 1.0])
25 standard_event=ot.StandardEvent(failure_event)
26 IS_exp=ot.ImportanceSamplingExperiment(aux_distribution)
27 IS algo=ot.ProbabilitySimulationAlgorithm(standard_event, IS_exp)
28 IS algo.run()
29 IS_results=IS algo.getResult()
30 IS_pf=IS_results.getProbabilityEstimate()
31 IS_pf_confidence=IS_results.getConfidenceLength(0.95)
32 # Estimate pf using subset simulation
33 SS algo=ot.SubsetSampling(failure_event)
34 SS algo.run()
35 SS_results=SS algo.getResult()
36 SS_pf=SS_results.getProbabilityEstimate()
37 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.

```

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

OpenTURNS 6 (Gaussian process regression). Fit of an ordinary Kriging model fitting.

```

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


Appendix E

Résumé étendu de la thèse

Introduction

Contexte industriel

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

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

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

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

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

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

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

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

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

Verrous scientifiques et objectifs de la thèse

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

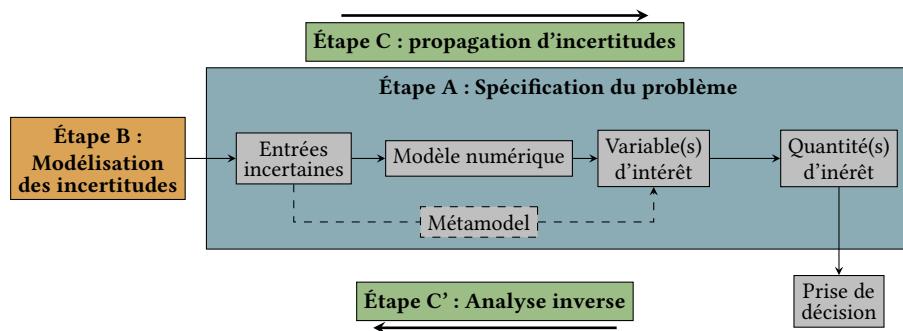


FIGURE E.1 Schéma générique de la quantification des incertitudes (De Rocquigny et al. (2008), adapté par Ajenjo (2023))

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

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

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

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

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

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

Chapitre 1 – Traitement des incertitudes en simulation numérique

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

“E. Fekhari, V. Chabridon, J. Muré and B. Iooss (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 Sanctetta and Satchell (2004). L'algorithme appelé “*Bernstein adaptive nonparametric conditional sampling*” (BANCS) est comparée à la méthode du *subset simulation* pour plusieurs problèmes de fiabilité des structures. Les premiers résultats sont encourageants, mais le contrôle du biais de l'estimateur doit être plus amplement investigué. Ce travail a mené à la publication et au développement informatique suivant :

- E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Bernstein adaptive nonparametric conditional sampling : a new method for rare event probability estimation”. In : *Proceedings of the 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 Independence Criterion*, pour évaluer la dynamique de la sensibilité fiabiliste Marrel and Chabridon (2021).

Conclusion

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

Communications et publications dans revues à comité de lecture

- | | |
|-----------------|---|
| Book Chap. | E. Fekhari, B. Iooss, J. Muré, L. Pronzato and M.J. Rendas (2023). “Model predictivity assessment : incremental test-set selection and accuracy evaluation”. In : <i>Studies in Theoretical and Applied Statistics</i> , pages 315–347. Springer. |
| Jour. Pap. | E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2024). “Given-data probabilistic fatigue assessment for offshore wind turbines using Bayesian quadrature”. In : <i>Data-Centric Engineering</i> . |
| Int. Conf. Pap. | E. Fekhari, B. Iooss, V. Chabridon, J. Muré (2022). “Efficient techniques for fast uncertainty propagation in an offshore wind turbine multi-physics simulation tool”. In : <i>Proceedings of the 5th International Conference on Renewable Energies Offshore (RENEW 2022)</i> , Lisbon, Portugal. (Paper & Talk) |
| | E. Fekhari, V. Chabridon, J. Muré and B. Iooss (2023). “Bernstein adaptive nonparametric conditional sampling : a new method for rare event probability estimation” ¹ . In : <i>Proceedings of the 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 14)</i> , Dublin, Ireland. (Paper & Talk) |
| | E. Vanem, E. Fekhari, N. Dimitrov, M. Kelly, A. Cousin and M. Guiton (2023). “A joint probability distribution model for multivariate wind and wave conditions”. In : <i>Proceedings of the ASME 2023 42th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2023)</i> , Melbourne, Australia. (Paper) |

A. Lovera, E. Fekhari, B. Jézéquel, M. 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.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Numerical Studies of Bayesian Quadrature Applied to Offshore Wind Turbine Load Estimation". In : <i>SIAM Conference on Uncertainty Quantification (SIAM UQ22)</i> , Atlanta, USA. (Talk)
	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Model predictivity assessment : incremental test-set selection and accuracy evaluation". In : <i>22nd Annual Conference of the European Network for Business and Industrial Statistics (ENBIS 2022)</i> , Trondheim, Norway. (Talk)
Nat. Conf.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2022). "Kernel-based quadrature applied to offshore wind turbine damage estimation". In : <i>Proceedings of the Mascot-Num 2022 Annual Conference (MASCOT NUM 2022)</i> , Clermont-Ferrand, France (Poster)
Invited Lec.	<u>E. Fekhari</u> , B. Iooss, V. Chabridon, J. Muré (2023). "Rare event estimation using nonparametric Bernstein adaptive sampling". In : <i>Proceedings of the Mascot-Num 2023 Annual Conference (MASCOT-NUM 2023)</i> , Le Croisic, France (Talk)
	Le Printemps de la Recherche 2022, Nantes, France. "Traitement des incertitudes pour la gestion d'actifs éoliens". (Talk)
	Journées Scientifiques de l'Eolien 2024, Saint-Malo, France. "Evaluation probabiliste de la fiabilité en fatigue des structures éoliennes en mer". (Talk)

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

Développements informatiques open source

`otkerneldesign`²

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

`bancs`³

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

`ctbenchmark`⁴

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

`copulogram`⁵

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

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

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

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

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

