Uncertainty quantification methodology for seismic fragility curves of mechanical structures : Application to a piping system of a nuclear power plant

Clément Gauchy

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2022IPPAX100

Thèse de doctorat

Méthodologie de quantification des incertitudes pour courbes de fragilité sismique de structures mécaniques - Application à un système de tuyauterie d'une centrale nucléaire

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École Polytechnique & au CEA Saclay

École doctorale n°574 École Doctorale de Mathématique Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 9 Novembre 2022, par

CLÉMENT GAUCHY

Composition du Jury :

Cyril Feau Ingénieur chercheur, CEA Saclay	Encadrant
Fabrice Gamboa Professeur, Université Paul Sabatier (IMT)	Rapporteur
Josselin Garnier Professeur, École Polytechnique (CMAP)	Directeur de thèse
Olivier Le Maitre Directeur de recherche, École Polytechnique (CMAP)	Examinateur
Amandine Marrel Ingénieur chercheuse, CEA Cadarache	Examinatrice
Clémentine Prieur Professeure, Université Grenoble Alpes (LJK)	Examinatrice
Bruno Sudret Professeur, ETH Zürich	Président du jury
Alexandros Taflanidis Professeur, University of Notre Dame	Rapporteur
Emmanuel Viallet Ingénieur expert fellow EDF, EDF DT	Invité

Uncertainty quantification methodology for seismic fragility curves of mechanical structures - Application to a piping system of a nuclear power plant

A manuscript submitted by

Clément GAUCHY

in fulfillment of the requirements
for the degree of Doctor of Philosophy

(Applied Mathematics)

Commissariat à l'Énergie Atomique et aux énergies alternatives
Gif sur Yvette, France

&

École Polytechnique
Palaiseau, France

defended publicly on the 9th of November 2022 in front of a defense committee made up of:

Dr. Cyril FEAU	CEA, Gif sur Yvette	Advisor
Pr. Fabrice GAMBOA	Université Toulouse III	Reviewer
Pr. Josselin GARNIER	École Polytechnique, Palaiseau	Thesis director
Pr. Olivier LE MAITRE	École Polytechnique, Palaiseau	Examiner
Dr. Amandine MARREL	CEA, Saint-Paul-lès-Durance	Examiner
Pr. Clémentine PRIEUR	Université Grenoble Alpes	Examiner
Pr. Bruno SUDRET	ETH Zürich	President
Pr. Alexandros TAFLANIDIS	University of Notre Dame, Indiana USA	Reviewer
Mr. Emmanuel VIALLET	EDF DT, Lyon	Invited

This manuscript was typeset using LaTeX compiled with pdflatex. It is based on the *Masters / Doctoral Thesis* template (version 2.5). The source files were edited using VScode. The text font used is *Palatino*. Graphical illustrations were produced with Ti*k*Z, PowerPoint (using IguanaTeX) and Matplotlib (running Python 3.7). The bibliography was compiled using BibLaTeX using Biber backend.

BibTeX entry:

```
@phdthesis{GauchyPhd2022,
author = {Gauchy, C.},
year = 2022,
title = {{Uncertainty quantification methodology for seismic fragility curves of mechanical
structures - Application to a piping system of a nuclear power plant}},
school = {Institut Polytechnique de Paris}
}
```

Pourquoi le sentiment s'est il ancré en moi de bonne heure que, si le voyage seul - le voyage sans idée de retour - ouvre pour nous les portes et peut changer vraiment notre vie, un sortilège plus caché, qui s'apparente au maniement de la baguette de sourcier, se lie à la promenade entre toutes préférée, à l'excursion sans aventure et sans imprévu qui nous ramène en quelques heures à notre point d'attache, à la clôture de la maison familière ?

Julien Gracq, *Les eaux étroites*

Remerciements

C'est une erreur de croire que les scientifiques sont par nature des gens rationnels, ils ne le sont pas plus que n'importe qui ; les scientifiques sont avant tout des gens fascinés par les régularités du monde.

Michel Houellebecq, *anéantir*

Les remerciements étant généralement le passage le plus lu des manuscrits de thèse, je me dois d'y citer toutes les multiples rencontres qui furent déterminantes dans l'élaboration de cet ouvrage. Je me dois tout d'abord de remercier mes deux rapporteurs, Messieurs Gamboa et Taflanidis, pour la lourde tâche qui leur a été incombée. *I address my heartfelt thanks and deepest gratitude to Prof. Taflanidis for reviewing this PhD thesis.* Mes plus profonds remerciements également aux membres extérieurs de mon jury, Mesdames Marrel et Prieur ainsi que Messieurs Le Maitre, Sudret et enfin Monsieur Viallet en qualité de membre invité. Vos remarques et questions me sont précieuses.

Les mathématiques prirent une place prépondérante dans ma vie à partir de mon arrivée au lycée, au contact de brillants professeurs tels que Mme Rebinguet et M. Blanc. Il s'en suivit deux années de retraite monastique que constitue la vie du préparationnaire au Lycée Pierre-de-Fermat. Ainsi, dans l'ancien Couvent des Jacobins, j'ai pu expérimenter pour la première fois la pratique de la recherche mathématique – je m'en rendis compte bien plus tard - à travers les cours de M. Brevet en deuxième année. Sous votre égide, les mathématiques devinrent vivantes. Elles n'étaient plus qu'un simple inventaire à la Prévert de propositions et autres lemmes, mais création humaine, produit brut d'un artisanat de l'esprit. Les devoirs surveillés - conçus entièrement par vos soins - permettaient de voir les joyaux que nous pouvions fabriquer, résultat d'un savant alliage de théorèmes, au-delà même de leur intérêt pratique pour la préparation aux concours.

Mon entrée à Centrale marqua le début d'une brève séparation avec les mathématiques. Néanmoins, l'idée d'une carrière de recherche ne quittait pas mon esprit. Je me dois ainsi de remercier profondément M. Herbin et Mme Lafitte pour m'avoir poussé à faire un Master recherche en plus de ma troisième année à l'école. Cette année passée au Laboratoire de Mathématique d'Orsay, inaugurant son nouveau bâtiment longeant l'Yvette, effaça enfin tous les doutes quant à mon envie de faire la recherche. J'y retrouvais là-bas l'énergie créatrice des mathématiques. Je remercie la bienveillance et la gentillesse de M. Giraud et M. Alquier, qui sûrent nous conseiller, nous guider et nous rassurer sur nos orientations futures. Merci à ceux que j'ai rencontrés là-bas : Rémi, Louis et enfin Martin, avec qui je partage une jeunesse toulousaine et les livres de Julien Gracq. Je te souhaite tout le meilleur pour ta dernière année de thèse !

Lors de ma recherche de stage de fin d'études, j'ai candidaté à une offre de stage à EDF R&D, sur le site de Chatou. Le premier entretien sur l'île des Impressionnistes fut également ma première rencontre avec le domaine de la quantification des incertitudes pour la maîtrise des risques industriels. J'y ai vu tout ce qui me tenait à cœur, que ce soit dans les aspects théoriques que industriels. Cette rencontre avec Bertrand Iooss, Jerôme Stenger et Roman Sueur fût un de ces moments déterminant, de ces aiguillages, qui nous font prendre des directions qui bouleversent nos existences. Jerôme, merci pour ta folie et ta grande disponibilité, j'ai eu sans nul doute le meilleur encadrant de stage possible ! Je te souhaite le meilleur pour ta première année d'enseignement aux Antilles. Bertrand, tes compétences sur les enjeux scientifiques et industriels sont inestimables, et de plus toujours accompagnées d'une profonde humanité. Ton esprit punk amène une fraîcheur

et une énergie qui a la saveur d'une éternelle jeunesse. Ces six mois de stage furent probablement une des meilleures périodes de ma vie, entouré des stagiaires, doctorants et agents EDF dont l'infinie bienveillance procura à cette période un caractère solaire. Merci à Azénor, Justine, Antoine, Mathilde, Louis et Anne pour tous ces moments passés au Bocal. Merci également aux doctorants de l'époque, Paul, Thomas, Sami, Pablo et Alvaro pour tous les échanges qui ont pu avoir lieu et votre bonne humeur qui vous transmettiez au sein du bâtiment S. Les agents EDF ne seront bien sûr pas oubliés dans ces remerciements. Roman pour les passionnantes discussions philosophiques dès 8h du matin en salle café. Elias pour ta profonde gentillesse (et surtout tes conseils d'ordre sartorial pour ma soutenance), je te souhaite une dernière année de thèse brillante et stimulante. Merlin pour ses a priori toujours plein d'optimisme et de joie. Vincent pour nos innombrables discussions sur probablement tous les sujets possibles, pour ta culture et ta sensibilité, pour ton aide précieuse sous des formes multiples - volontaire ou non – et ton invincible détermination. Je pense désormais aux collaborations fructueuses et passionnantes que nous aurons ensemble ! Je remercie également toutes les pauses café - les matins d'été - sur le balcon terrasse du bâtiment S, vue sur la Seine bordée de péniches et de saules pleureurs. Le fantôme d'Auguste Renoir n'était alors pas loin de nous.

Durant ma thèse qui se déroula entre le CEA Saclay et l'Ecole Polytechnique, je fis de multiples rencontres qui m'influencèrent grandement que ce soit sur le plan humain que scientifique. Tout d'abord, mes profonds remerciements à mon encadrant au CEA Cyril et à mon directeur de thèse Josselin. Cyril, je te suis éternellement reconnaissant de t'être lancé dans l'encadrement d'une thèse de mathématiques appliquées. Tu as toujours su te montrer compréhensif et ouvert d'esprit face à mes multiples suggestions et idées, il fallait bien quelqu'un de ta trempe pour que l'on accomplisse ensemble pareille odyssée ! Josselin, merci pour ton écoute, ta perspicacité, et surtout ta bienveillance si précieuse. Tu fus de très bon conseil pour cadrer le sujet d'un point de vue mathématique et débloquer de nombreux verrous scientifiques. Ce sera un plaisir de collaborer ensemble dans le futur. Un immense merci à tous mes collègues du laboratoire EMSI du CEA Saclay : Sandra pour m'avoir dévoilé toutes les arcanes de la table vibrante en pause-café. Pierre-Etienne pour avoir refait le monde ensemble des milliards de fois à s'en donner le tournis. Matthieu pour tes cinquante nuances de filtres de Kalman, ta bonne humeur et les discussions scientifiques et vélocipèdiques. Giuseppe, *grazie per avermi fatto ridere così tanto in questi tre anni*. Bruno pour m'avoir introduit au monde impitoyable des tenseurs d'endommagement anisotropique qui nécessitent d'avoir en offrande tous les tableaux de la pièce 12A. Sylvie pour ton aide très précieuse en début de thèse pour toutes les démarches administratives. Merci Antoine pour avoir été un super stagiaire, je suis ravi que l'on puisse continuer à collaborer ensemble dans le cadre de ta thèse. Et enfin merci à Darius, Vincent, Quentin, Martial, Michel, Reine, Alain et Thierry pour les discussions aux thèmes très variées en pause-café.

Toujours sur un plan scientifique et humain, je tiens à remercier les organisateurs et membres du GdR MASCOT-NUM ainsi que les habitués de l'école d'été ETICS, dont la joie et la bienveillance m'ont été d'une aide si précieuse dans ces multiples périodes d'isolement liées au Covid. Je pense ici aux doctorants : Marouane, Guillaume. Mais également aux autres, Gaël, Claire, Cécile, Amandine, Delphine et Sophie et ceux que j'oublie... Et enfin au plus sage d'entre nous, Sébastien Da Veiga. J'ai beaucoup appris en discutant avec toi. Nous aurons assurément de nombreuses collaborations dans ta nouvelle vie bretonne ! Un grand merci également au comité exécutif qui permet à ce GdR CNRS d'exister : Anthony Nouy, Julien Bect, Guillaume Perrin, Clémentine Prieur.

Merci également aux doctorants du CMAP, j'ai eu la chance d'évoluer dans un cadre fort agréable et riche d'échanges divers et variés avec chacun d'entre vous. Mille mercis à Constantin, Solange pour les discussions entre toulousains, Benjamin, Thomas, Arthur pour avoir été un coburreau affable et représenter fièrement la minorité des EDPistes du CMAP au milieu des dizaines

de milliers de doctorants en statistiques et machine learning. À l'équipe de choc qui assuma durant deux ans la noble mission de diffusion de la connaissance via le séminaire des doctorants du CMAP/CMLS : Clément M. pour les multiples quiproquos provoqués par la non-identifiabilité due à nos prénoms, Louis, Baptiste, Leila et Kevin. Remerciements spéciaux à Baptiste, tu fus un fidèle acolyte durant ces trois années où nous avons fait les 400 coups à notre façon, la semaine au Motel 6 d'Atlanta avec vue sur la 2x6 voies restera mémorable. A bientôt pour de nouvelles aventures CEAesque !

Bien évidemment je souhaite remercier toute ma famille à qui je dois évidemment beaucoup. Merci à mes parents pour m'avoir toujours soutenu dans mes choix et qui se sont assurés que je ne manquais de rien. Merci à ma sœur et à mon beau-frère de me faire souvent rire ! Des remerciements tout particulier à mon grand-père Charles et ma grand-mère Marie-Thérèse à qui je dois beaucoup de mes choix d'études et de carrière.

Je remercie tous mes amis de Centrale qui m'ont fait marrer pendant mes trois années d'études. Je pense en particulier à Ronan, Gabriel, Nox, Julien, Kangoo, Stook, Bebert, Gabriel (encore), Djou et ceux que j'oublie...

En guise de conclusion, merci à toutes les rencontres que j'ai pu faire durant ces vingt-six années et demi sur cette planète dont voici une liste non exhaustive : Tim, urbaniste le jour et DJ la nuit, Brou, et Drou l'homme qui murmure à l'oreille des chaudières.

Quand il faudra fermer le livre
Ce sera sans regretter rien

Louis Aragon

Contents

List of Figures

List of Tables

List of Abbreviations

r.v.	random variable
i.e.	*id est*
w.r.t.	with respect to

Notations

a	Intensity measure
A	Intensity measure (r.v.)
z	Engineering demand parameter
Z	Engineering demand parameter (r.v.)
$\mathbf{x}$	Mechanical parameters
$\mathbf{X}$	Mechanical parameters (r.v.)
$\mathbf{X}^{(j)}$	j-th mechanical parameter (r.v.)
y	Output parameter
η	standard lognormal r.v.
ψ	Seismic fragility curve
ψ_{ref}	Reference seismic fragility curve
Ψ	Epistemic seismic fragility curve
$\Psi^{(1)}$	Epistemic seismic fragility curve estimator
$\Psi^{(2)}$	Posterior predictive seismic fragility curve
$\bar{\Psi}$	Averaged epistemic fragility curve
α	Median parameter for the log-normal model
β	Logarithmic standard deviation for the log-normal model
$\mathcal{M}$	Black box computer model
G	Gaussian process

Abstract

Nuclear power plants are complex engineering systems for which reliability has to be guaranteed during its operational lifetime due to the hugely negative consequences provoked by nuclear accidents to human health or to the environment. The effects of natural hazards such as earthquakes on these facilities are included in the risk analysis but are difficult to estimate because of their randomness. Since the 1980s, a probabilistic seismic risk assessment framework has been developed to evaluate the reliability of structures, systems and components (SSC) of nuclear facilities against seismic risk. This framework is relying on a specific quantity of interest: the *seismic fragility curve*. At the scale of these facilities, these curves represent the conditional probabilities of failure of the SSCs given a scalar value derived from a seismic loading indicating its "strength" and which is called seismic intensity measure. The management of the various sources of uncertainty inherent to the problem to be addressed is often divided into two categories: (i) the so-called *random* uncertainties that arise from the natural variability of physical phenomena that are difficult to measure or control, and (ii) the so-called *epistemic* uncertainties that are associated with the lack of knowledge of the system under study and that can be reduced, in the short term, by means of experimental campaigns for example. In seismic probabilistic risk assessment studies for the nuclear industry, the main source of random uncertainty is the seismic loading and the sources of epistemic uncertainties are attributed to the mechanical parameters of the structure considered. In this framework, this thesis aims at understanding the effect of epistemic uncertainties on a seismic fragility curve by using an uncertainty quantification methodology. However, as numerical mechanical models are often computationally expensive, a metamodeling step, based on Gaussian process regression, is proposed. In practice, the sources of epistemic uncertainties are first modeled using a probabilistic framework. After establishing a Gaussian process metamodel of the numerical mechanical model, they are then propagated through the surrogate model. The propagation of epistemic uncertainties as well as the sensitivity analysis are then carried out on the seismic fragility curve via the metamodel, using a reduced number of calls to the mechanical computer code. This methodology thus allows both propagating and ranking the most influential epistemic sources of uncertainty on the fragility curve itself, at a reduced numerical cost. In addition, several procedures for planning numerical experiments are proposed to lighten the computational load, while ensuring the best possible estimation accuracy on the seismic fragility curve. The methodologies presented in this thesis are finally tested and evaluated on an industrial test case from the nuclear industry, namely a section of piping equipping French pressurized water reactors.

Keywords: Seismic reliability • Uncertainty quantification • Design of experiments • Gaussian process • Nuclear industry

Résumé

Les centrales nucléaires sont des systèmes complexes dont la fiabilité doit être garantie tout au long de leur durée de vie opérationnelle en raison des conséquences négatives des accidents nucléaires sur la santé humaine et l'environnement. Les effets des risques naturels tels que les séismes sur ces installations sont intégrés à l'analyse des risques mais sont difficiles à estimer en raison de leur caractère aléatoire. Depuis les années 1980, un cadre d'évaluation probabiliste des risques sismiques a été développé pour évaluer la fiabilité des structures, systèmes et équipements (SSE) des installations nucléaires face au risque sismique. Ce cadre s'appuie sur une quantité d'intérêt spécifique: la *courbe de fragilité sismique*. A l'échelle de ces installations, ces courbes représentent les probabilités conditionnelles de défaillance des SSE sachant une valeur scalaire dérivée d'un chargement sismique indiquant sa « force » et qui est appelée intensité de mesure sismique. La gestion des diverses sources d'incertitudes inhérentes au problème à traiter est souvent divisée en deux catégories : (i) les incertitudes dites *aléatoires* qui proviennent de la variabilité naturelle des phénomènes physiques difficiles à mesurer ou à contrôler et (ii) les incertitudes dites *épistémiques* que l'on associe au manque de connaissance du système étudié et qui peuvent être réduites, à court terme, par le biais de campagnes expérimentales par exemple. Dans les études d'évaluation probabiliste du risque sismique pour l'industrie nucléaire, la principale source d'incertitude aléatoire est le chargement sismique et les sources d'incertitudes épistémiques sont attribuées aux paramètres mécaniques de la structure considérée. Dans ce cadre, cette thèse vise à appréhender l'effet des incertitudes épistémiques sur une courbe de fragilité sismique en utilisant une méthodologie de quantification des incertitudes. Toutefois, comme les modèles mécaniques numériques sont souvent coûteux en temps de calcul, une étape de métamodélisation, basée sur la régression par processus Gaussien, est proposée. En pratique, les sources d'incertitudes épistémiques sont en premier lieu modélisées en utilisant un cadre probabiliste. Après l'établissement d'un métamodèle par processus Gaussien du modèle mécanique numérique, elles sont ensuite propagées à travers celui-ci. La propagation des incertitudes épistémiques ainsi que l'analyse de sensibilité sont alors menées sur la courbe de fragilité sismique via le métamodèle, à partir d'un nombre réduit d'appels au code de calculs mécaniques. Cette méthodologie permet donc à la fois de propager et de classer les sources d'incertitudes épistémiques les plus influentes sur la courbe de fragilité elle-même, à un coût numérique réduit. En outre, plusieurs procédures de planification d'expériences numériques sont proposées pour alléger la charge de calcul, tout en garantissant la meilleure précision d'estimation possible sur la courbe de fragilité sismique. Les méthodologies présentées dans cette thèse sont finalement testées et évaluées sur un cas test industriel issu de l'industrie nucléaire, à savoir un tronçon de tuyauterie équipant des réacteurs à eau pressurisée français.

Mots-clés: Fiabilité sismique • Quantification des incertitudes • Planification d'expériences • Processus Gaussien • Industrie nucléaire

CHAPTER 1

Introduction

Context

Nuclear power plants (NPP) are some of the most complex engineering systems ever designed by mankind, involving almost every domain of physics and engineering like fluid mechanics, thermohydraulics, neutronics, structural mechanics, fluid structure interactions to say the least. As any complex engineering system, it is not deemed failure-proof, and due to the very long lifetime of nuclear power plants, nuclear safety is a truly complex and challenging task. Moreover, the negative consequences provoked by nuclear accidents - in terms of harm to human health and to the surrounding environment - are so huge that the safety level imposed by national regulatory agencies is one of the highest ever imposed to a complex engineering system.

One of the possible sources of nuclear accidents comes from natural hazard such as an earthquake event in the site of the NPP. The dynamical behavior during a seismic event of the various structural components of the NPP has to be properly assessed to guarantee its operational conditions. The main challenge of seismic risk assessment of NPPs is the inherent randomness of earthquakes in terms of temporal and spectral content of the seismic loading. Since the 1980s, the seismic probabilistic risk assessment (SPRA) framework has been developed in the USA to properly assess the safety of NPP's components under seismic excitations, accounting for the randomness of seismic hazard through a probabilistic framework.

Since the last decades, the NPP safety has increasingly relied on high-fidelity computer models that aim to be a *digital twin* of the system simulated. The use of numerical simulations is motivated by the possibility to simulate physical phenomena that can not be observed or replicated on test specimens due to their complexity (e.g. the dynamical behavior under seismic loading of the whole piping system of a NPP). The computer model in structural dynamics often relies on solving numerically ordinary or partial differential equations using for instance finite element simulations.

Assessing the reliability and safety of mechanical structures involves taking into account various sources of uncertainties. Uncertainties may come from natural hazard such as wind or seismic loading, but also from the physical properties of the structures (e.g. the material properties or the boundary conditions). These uncertainties have to be taken into account to ensure a satisfactory safety level. Such an objective can be achieved thanks to a general *uncertainty quantification framework* presented in the next section.

Uncertainty quantification framework

Uncertainty quantification (UQ) gathers a wide variety of both theoretical and applied tools from probability theory, computational statistics and stochastic simulation. UQ framework is by definition interdisciplinary and can be applied to many engineering branches. UQ can be summarized into a few fundamental steps that can be applied to any engineering branch of interest (see e.g. Sudret, 2007; De Rocquigny et al., 2008; Iooss, 2009). These steps are detailed below:

- **Problem specification:** the first step is naturally the definition of the engineering system studied and the conception of the computer model that will simulate this system. It includes defining the input variables of the computer model and its output quantities that matter for studying the physical phenomena of interest.
- **Uncertainty modeling:** the second step consists in listing all the possible sources of uncertainties affecting the input variables of the computer model. The most classical mathematical model for the representation of the uncertainty is the probabilistic framework.
- **Uncertainty propagation:** the third step is dedicated to the propagation of the uncertainties tainting the input variables to the output variables through the computer model. Therefore, the output variable is tainted by uncertainties as well. Statistical tools can then be applied on the output variables to estimate various quantities of interest (e.g. mean, quantile, probability of exceedance...)
- **Inverse analysis:** this last step can be split into two parts: First, the *calibration* of the computer model w.r.t. available data, which is out of scope of this manuscript. Second, the *sensitivity analysis* step aims at studying how the uncertainty on the output variable can be apportioned to each input variable (or subset of input variables).

In this manuscript, we suppose that the computer model is previously calibrated. The inverse analysis step will be devoted to the sensitivity analysis step.

Problem statement

This thesis deals with the problem of *seismic fragility curve* estimation of a given mechanical structure. This particular quantity of interest is a key quantity of the SPRA framework and is the probability of failure of a mechanical structure conditional to a *seismic intensity measure*, which is commonly a scalar value. A seismic fragility curve is often represented graphically as shown in Figure 1.1. After the pioneering work done in R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984, seismic fragility curves estimation is now usually performed using numerical simulations based on computer models (see e.g. Karim and Yamazaki, 2001; Kim and Shinozuka, 2004; Zentner, 2010). However, the mechanical and material parameters of the structure are affected by uncertainties arising from manufacturing variability. Moreover, the boundary conditions (e.g. fixation of the structure to the support) are very often uncertain due to a lack of knowledge and/or a lack of data. Thus, these uncertainties may affect the seismic fragility curves of the structure.

In this thesis, the emphasis is on taking into account the uncertainties that taint the mechanical parameters of a structure on its seismic fragility curve, using the UQ framework. Therefore, the main problematic under study in this manuscript can be stated as follows.

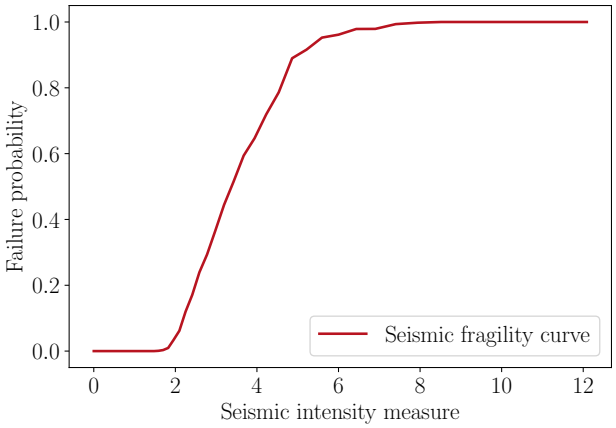


FIGURE 1.1: Illustration of a seismic fragility curve.

How to build an UQ methodology for the seismic fragility curves of mechanical structures ?

This problematic can be decomposed into the following questions:

Q1 - How to propagate the mechanical parameters uncertainties into seismic fragility curves with a reasonable computational time ?

Q2 - How the uncertainty on the seismic fragility curve can be apportioned to the different sources of epistemic uncertainties ?

Q3 - How to planify the numerical simulations to improve the estimation accuracy of the seismic fragility curves and to reduce the number of simulations in the same time ?

The two questions **Q1** and **Q2** are devoted to the adaptation of the general UQ framework to seismic safety studies, and more specifically for the seismic fragility curves which are key quantities of interest in the SPRA framework. The last question **Q3** is a key problematic in goal-oriented UQ. Indeed, it concerns the planification of the computer calls w.r.t. all its input parameters to accurately estimate the quantity of interest (here the seismic fragility curve). Several scientific objectives of this thesis and a brief *résumé* of this manuscript are stated in the next section.

Objectives and outline

The objectives of this thesis are listed below:

O1 Propose a state-of-the-art review of the different existing statistical methods for seismic fragility curves estimation in the SPRA framework;

O2 Develop a methodology to account for epistemic uncertainties in the seismic fragility curves with appropriate statistical tools;

O3 Bring forward sensitivity analysis tools that are seismic fragility curve-oriented, taking into account the estimation uncertainties of the seismic fragility curves;

O4 Propose numerical experimental designs to improve the estimation of seismic fragility curves while managing in the same time the number of calls of the computer model;

O5 Apply the proposed tools to a realistic computer model of a NPP component.

Letting the introduction and conclusion chapters aside, the present manuscript is composed of seven chapters aiming to carry out the five scientific objectives stated above. The contents of each chapter is detailed below.

Chapter 2 proposes a brief review of the principal methods for statistical estimation of seismic fragility curve. The methodology for the estimation of seismic fragility curves using numerical experiments is also recalled, from the generation of synthetic ground motions to the statistical methods of estimation of seismic fragility curves.

Chapter 3 presents a novel estimation methodology of seismic fragility curves using importance sampling in order to improve their estimation accuracy.

Chapter 4 presents an overview about the UQ framework. First, the fundamental notions are described, such as the *black-box* viewpoint, the sources of uncertainties and their probabilistic modeling. Then, an adaptation of the UQ framework to earthquake engineering is proposed, focusing on the seismic fragility curves.

Chapter 5 addresses the core problem of surrogate modeling of black-box computer models. Due to the computational cost of seismic fragility curves estimation, the computer model is replaced by a surrogate using statistical methods. The main advantage is a cheap estimation of the seismic fragility curve, but they are now tainted by a surrogate modeling uncertainty. Surrogate modeling using Gaussian process regression is then developed in this chapter, due to its ability to propose both a prediction and a quantification of the uncertainty on its predictions.

Chapter 6 presents a sensitivity analysis methodology on the seismic fragility curve. The focus is put on the more specific framework of global sensitivity analysis (GSA) that aims at considering the overall probabilistic framework on the input parameters of the computer model. After a brief state-of-the-art review of GSA, two global sensitivity indices are proposed as well as their estimators. A numerical procedure is detailed to account for the uncertainty arising from the surrogate model.

Chapter 7 proposes a sequential planning of experiments procedure based on Gaussian process to improve the estimation accuracy of the seismic fragility curve using a Gaussian process surrogate model. After a brief recall on *stepwise uncertainty reduction* (SUR) techniques, a SUR strategy is then developed for seismic fragility curves estimation.

Chapter 8 presents a nuclear engineering representative test-case issued from an experimental program of the *Commissariat à l'énergie atomique et aux énergies alternatives* (i.e. French national laboratory on nuclear and carbon-free energies). It consists in a mock-up of a piping system of a French pressurized water reactor (PWR). In this chapter, the tools developed in the three previous chapters are applied to discuss the use of the proposed methodology to a real test-case.

The chapters can be organized into two groups: On the one hand, Chapters 2 and 3 are more about the classical methods of seismic fragility curve estimation and a possible improvement using the statistical technique of importance sampling, the question of epistemic uncertainties is not addressed there. On the other hand, chapters 4, 5, 6, 7 and 8 concern the accounting of epistemic uncertainties on the seismic fragility curve and the development of an UQ methodology on the

seismic fragility curves, They are intended to be read in chronological order to be consistent with the various mathematical tools used in these chapters. Table 1.1 provides a brief recap of the contents of each chapter of this manuscript and specify their links to the scientific objectives. For each chapter, a mention "SOTA" or "NEW" indicates if the content of the chapter is more "state-of-the-art" related or an unpublished / novel contribution.

Moreover, I had the opportunity to supervise the six-months research internship of Antoine Van Biesbroeck. The research topic concerned the elicitation of prior distributions concerning Bayesian methods for estimating seismic fragility curves. Since its content differs from the main topic of this thesis, his work has been exposed in Appendix C.

TABLE 1.1: Summary of the content of the thesis

Keywords	**Chapter**	**Content**	**Objectives**
Seismic fragility curve / Ground motion generation / Log normal model	Chapter 2	SOTA	O1
Importance sampling / Planning of experiments / Confidence intervals	Chapter 3	NEW	O4/O5
Uncertainty modeling / Probability theory / Black-box computer models	Chapter 4	SOTA	O2
Surrogate modeling / Gaussian process regression	Chapter 5	SOTA/NEW	O2
Fragility curve-oriented GSA / Kriging-based GSA / Kernel methods	Chapter 6	NEW	O3
Sequential planning of experiments / Gaussian process	Chapter 7	NEW	O4
Piping system of a PWR / Seismic safety study	Chapter 8	NEW	O5

Publications and communications

The research contribution presented in this manuscript has been the subject of the following publications and communications listed below.

Jour. Pap. **Gauchy, C.**, C. Feau, and J. Garnier (2021a). *Importance sampling based active learning for parametric seismic fragility curve estimation*. DOI: 10.48550/ARXIV.2109.04323

Gauchy, C., C. Feau, and J. Garnier (2022c). *Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging*. DOI: 10.48550/ARXIV.2210.06266

Int. Conf. **Gauchy, C.**, C. Feau, and J. Garnier (2022a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *SIAM conference on Uncertainty Quantification 2022*. Atlanta, USA. (Talk)

Nat. Conf. **Gauchy, C.**, C. Feau, and J. Garnier (2020a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *Mascot-num PhD days (MASCOT-NUM 2020)*. Grenoble, France. (Poster)

Gauchy, C., C. Feau, and J. Garnier (2020b). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *5ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2020)*. Saint Pierre d'Oléron, France. (Poster)

Gauchy, C., C. Feau, and J. Garnier (2021d). "Propagation of epistemic uncertainties in seismic risk assessment". In: *15th Mascot-Num annual conference (MASCOT-NUM 2021)*. Online (Poster)

Gauchy, C., C. Feau, and J. Garnier (2021b). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *51ème journées de Statistique de la Société Française de Statistique (JdS 2021)*. Online (Talk)

Gauchy, C., C. Feau, and J. Garnier (2021c). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *6ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2021)*. Erdeven, France (Talk)

Gauchy, C., C. Feau, and J. Garnier (2022d). "Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging". In: *16th Mascot-Num annual conference (MASCOT-NUM 2022)*. Clermont-Ferrand, France (Poster)

Gauchy, C., C. Feau, and J. Garnier (2022b). "Estimation of seismic fragility curves by sequential design of experiments". In: *52ème journées de Statistique de la Société Française de Statistique (JdS 2022)*. Lyon, France (Talk)

CHAPTER 2

Statistical methods for seismic fragility curves estimation: a brief review

Contents

2.1 Context and motivations

Among the natural hazards, earthquakes are known to cause huge structural damages while being extremely difficult to forecast. Since the last 1980s, a probabilistic framework has been developed to evaluate the annual probability of occurrence of severe damage on structures caused by seismic ground motions. The Seismic Probabilistic Risk Assessment (SPRA) (see R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984) is thus a widely used methodology for seismic structural reliability which is now the most common approach to evaluate the seismic risk on structures in nuclear facilities. The key elements of SPRA studies are the determination of the seismic hazard, the seismic fragility evaluations of the components of a structure and the system analysis including the construction of a logical fault tree model of the system. The law of total probability allows us to evaluate the annual probability of failure due to seismic hazard of the system studied.

In this framework, fragility analysis of the structure is a crucial step. This boils down generally to the estimation of a so-called fragility curve, which is the probability of failure of a mechanical structure conditional to seismic Intensity Measure (IM). A seismic IM is a scalar quantity derived from the earthquake accelerogram signal (for instance the maximal absolute soil acceleration during a seismic event at a designated site) in order to measure an earthquake strength. Fragility curve is also a key quantity for Performance Based Earthquake Engineering (PBEE) (see Ghobarah, 2001; Porter, R. Kennedy, and Bachman, 2007) framework. In this context, they are based on a different definition - the probability of exceeding a certain damage state for a given Engineering Demand Parameter (EDP) such as interstory drift ratio for example - and are often referred to as fragility functions (see e.g. Porter, R. Kennedy, and Bachman, 2007 and the proceedings in PEER, 2004), even if this denomination is not reserved for this last definition and also includes the previous one (see e.g. Baker, 2015; Kumar and Gardoni, 2013; Lallemant, Kiremidjian, and Burton, 2015). The concept of fragility curve is not limited to seismic loading and can be used for wind loading (Quilligan, O'Connor, and Pakrashi, 2012) or multi-hazard storm conditions (Bernier and Padgett, 2019).

Estimation of seismic fragility curve is based on three different approaches:

- **Expert judgment:** Expert opinions on the structure studied and the ground motion activity of the specific site studied are used to define a fragility curve for the structure, this was the first method of fragility curve estimation for SPRA studies proposed in R. Kennedy, C. Cornell, et al., 1980 concerning nuclear power plants in the USA. This method does not require computational power, which makes it very appealing in a time when computational resources were scarce and not as powerful as today. However, the major drawback is the inherent subjectivity of the method.

- **Empirical observations:** Fragility curve can be estimated after observational campaigns made by civil engineers in regions that have a recent seismic activity. Several structures are observed to determine their damage level. After this data gathering step, statistical models are used to provide a fragility curve estimate. This method is more adapted for countries with huge seismic activity such as the USA or Japan.

- **Computer simulations:** The rise of computational power in the late 1990s made possible the use of mechanical numerical models (such as Finite Elements) to replace infeasible real experiments by their numerical counterparts. This method is particularly interesting due to its huge flexibility: It is possible to simulate a wide number of structures for seismic ground motions of different sites in the world. However, numerical models of the seismic excitation have to be developed and calibrated on real seismic ground motion signals.

This chapter presents a brief review of the different methods for estimating seismic fragility curves with a focus on the SPRA framework and is organized as follows. Section 2.2 presents the different definitions of the seismic fragility curve and the most common estimation methods that can be found in the literature. Section 2.3 focuses on the particular case of seismic fragility curves estimation using numerical mechanical simulations and finally Section 2.4 illustrates the main concepts of this chapter on a single d.o.f. (degree of freedom) oscillator with nonlinear restoring force.

2.2 Seismic fragility curves

2.2.1 Definition

In this work, the fragility curve of a mechanical structure is defined as the conditional probability of failure for a given seismic intensity measure:

$$\psi(a) = \mathbb{P}(Z \geq C|A = a) \,, \tag{2.1}$$

A is the random variable of the seismic intensity measure, Z is the random variable of the mechanical demand (observed or computed using numerical simulations) of the structure studied and C is a threshold for which the structure studied is considered to be in a non-acceptable behavior in terms of safety. Usually, the seismic intensity measure is a scalar value derived from the seismic ground motion accelerogram. As recalled in Grigoriu and Radu, 2021, seismic intensity measures must follow three properties:

- *Efficiency:* the conditional distribution of Z given A must have small variance;
- *Sufficiency:* the conditional distribution of Z given A must summarize the effect of the seismic hazard;
- *Scale robust:* The conditional distribution of Z given A is unchanged when using scaling seismic signals. Scaling here means adjusting a set of seismic signals to a specific intensity measurement value A by multiplying them by ad hoc constants. In practice such method is nevertheless not recommended (see Section 2.2.3).

2.2.2 SPRA framework for simulation-based approaches

The present work is part of the SPRA framework which ultimately consists in evaluating the mean annual frequency of failure (or undesirable outcome) λ_f defined by (see e.g. R. Kennedy, 1999; A. Cornell, 2004; Baker, 2015; Kumar and Gardoni, 2013):

$$\lambda_f = \int_0^\infty \psi(x) \, |dH(x)| \tag{2.2}$$

where $\psi(x) = \mathbb{P}(\{\text{failure}\}|x)$ is the IM-based fragility curve we want to estimate via simulation-based approaches and $H(x)$ is mean annual frequency of exceedance of a ground motion of level IM $= x$ (i.e. the seismic hazard curve at a designated site). In A. Cornell, 2004, the author recalls the main assumption on which the equation (2.2) is based. This is the so-called sufficiency assumption of the IM with respect to the magnitude (M), source-to-site distance (R) and other parameters thought to dominate the seismic hazard at the site of interest (e.g. $\mathbb{P}(\{\text{failure}\}|\text{M,R,IM}) = \mathbb{P}(\{\text{failure}\}|\text{IM})$). However, he rightly advises remaining cautious in practice and to choose as far as possible seismic signals compatible with the scenario of interest because it is difficult to establish the sufficiency of an IM candidate. Accordingly, this is what we did in this thesis to perform nonlinear dynamic analyses.

They are various methods to estimate the fragility curve of a structure, after a brief review about the major estimation techniques, we briefly present three main methods: the first one is based on expert judgments, the second one is based on a lognormal assumption on the seismic fragility curve and finally the third and last one is based on linear regression in log-space.

2.2.3 Estimation of seismic fragility curves : a brief state-of-the-art

In earthquake engineering, various data sources can be exploited to estimate these curves, namely: expert judgments supported by test data (see e.g. R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984; Park, Hofmayer, and Chokshi, 1998; Zentner, Gündel, and Bonfils, 2017), experimental data (see e.g. Park, Hofmayer, and Chokshi, 1998; Gardoni, Der Kiureghian, and Mosalam, 2002; Choe, Gardoni, and Rosowsky, 2007), post-earthquake damage results called empirical data (see Lallemant, Kiremidjian, and Burton, 2015; Straub and Der Kiureghian, 2008) and analytical results given by more or less refined numerical models. For approaches based on nonlinear dynamic structural analyses, three families of procedures are commonly used to collect data in order to assess seismic fragility curves: Incremental Dynamic Analysis (IDA), Multiple Stripes Analysis (MSA) and Cloud Analysis (CA). In most cases, these approaches are also coupled with a parameterization of the fragility curve since, compared to non-parametric estimations, parametric ones require small sample sizes. The lognormal model historically introduced in the SPRA framework (see R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984) is currently the most used (see e.g. Lallemant, Kiremidjian, and Burton, 2015; Shinozuka et al., 2000; Ellingwood, 2001; Kim and Shinozuka, 2004; Mandal, S. Ghosh, and Pujari, 2016; F. Wang and Feau, 2020; Mai, Konakli, and Sudret, 2017; Trevlopoulos, Feau, and Zentner, 2019), even if its validity is questionable (see e.g. Lallemant, Kiremidjian, and Burton, 2015; Mai, Konakli, and Sudret, 2017; Trevlopoulos, Feau, and Zentner, 2019; Zentner, 2017). IDA is based on scaled accelerograms until the failure threshold of interest. Fragility curves are then considered as empirical cumulative distribution functions. The main disadvantage of IDA is that excessive scaling can lead to signals with unrepresentative frequency content and duration, which can imply biased results in nonlinear structural responses (see e.g. Luco and Bazzurro, 2007; Altieri and Patelli, 2020). Strong evidence against scaling accelerograms is provided in Grigoriu, 2011 considering a more theoretical approach. Although not recommended, this approach is still implemented (see e.g. Mandal, S. Ghosh, and Pujari, 2016; C. Zhao, Yu, and Mo, 2020; Y. Zhao et al., 2021). MSA is based on multiple accelerograms selected or scaled to match specific IMs. Thus, for each IM value, the structural analyses provide a fraction of the ground motions that cause failure. Finally, CA is a kind of generalization of MSA in the sense that it is based on a single accelerogram for each IM value, i.e. no scaled accelerogram is used. Depending on the context, different techniques can be employed to estimate the fragility curves. For example, for parametric estimation, we distinguish: the method of moments (which is mainly used for the IDA-based methodology), the Maximum Likelihood Estimation (MLE) by assuming the independence of the observations (which can be questionable when empirical data are concerned, see (Straub and Der Kiureghian, 2008)), and the minimization of the Sum of Squared Errors (SSE). For non-parametric estimation, kernel smoothing can be used (see e.g. Lallemant, Kiremidjian, and Burton, 2015; Mai, Konakli, and Sudret, 2017) as well as other methodologies (see e.g. Trevlopoulos, Feau, and Zentner, 2019; Altieri and Patelli, 2020). Note that most of these strategies are compared in Baker, 2015; Lallemant, Kiremidjian, and Burton, 2015; Mai, Konakli, and Sudret, 2017 and Lallemant, Kiremidjian, and Burton, 2015 give a clear presentation of the advantages and disadvantages of each of them.

Beyond these methods, techniques based on machine learning can also be used, including: linear regression or generalized linear regression (see Lallemant, Kiremidjian, and Burton, 2015), classification - based techniques in Bernier and Padgett, 2019; Kiani, Camp, and Pezeshk, 2019; Sainct et al., 2020, kriging in Gidaris, Taflanidis, and Mavroeidis, 2015, polynomial chaos expansion in Mai, Spiridonakos, et al., 2016, artificial neural networks in Sainct et al., 2020; Mitropoulou and Papadrakakis, 2011; Z. Wang et al., 2018. Some of them were coupled with adaptive techniques to reduce the number of calculations to be performed (see Sainct et al., 2020; Gidaris, Taflanidis, and Mavroeidis, 2015). The Bayesian framework is also relevant in this context since

it allows either (i) to fit numerical models (metamodels, mathematical expressions based on engineering judgments, etc.) to experimental data to "directly" (see below) estimate the fragility curves (see Gardoni, Der Kiureghian, and Mosalam, 2002) or (ii) to use empirical data or analytical data to fit the parametric models of the fragility curves (see Straub and Der Kiureghian, 2008).

Physics-based approaches are also suitable for estimating fragility curves. Let us quote for example the approach developed as part of the PBEE framework, which makes it possible to estimate non-parametric fragility curves by carrying out a rigorous reliability analysis once the capacity and demand models have been developed (see Gardoni, Der Kiureghian, and Mosalam, 2002). In doing so, it also offers the advantage of being able to carry out sensitivity analyses (see Choe, Gardoni, and Rosowsky, 2007). The capacity model is first built on deterministic engineering principles, thus exhibiting the main variables of the mechanical problem. Then, an additive probabilistic model is constructed by introducing a bias correction term and a noise term to represent the error of the model, which depend on a number of unknown parameters. The demand model is built with similar considerations. This approach was implemented to estimate the EDP-based fragility curves of reinforced concrete columns, having multiple failure modes, for which a large experimental database is available in order to estimate the unknown parameters in the Bayesian framework as in Gardoni, Der Kiureghian, and Mosalam, 2002; Choe, Gardoni, and Rosowsky, 2007. The same methodology was extended in Gardoni, Mosalam, and D. A., 2003 for estimations of IM-based fragility curves of reinforced concrete bridges considering, among others, the results of numerical simulations and, later, effects of corrosion in Choe, Gardoni, Rosowsky, and Haukaas, 2009. Several studies have subsequently been carried out using this methodology.

2.2.4 Estimation based on expert judgment: another definition of the fragility curve

When data are scarce or incomplete, the estimation of seismic fragility curves is only possible based on the judgment of engineers or experts in the domain. The safety factor method (see R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984) was developed in the 1980s in order to propose an economical method for seismic fragility curve estimation. This method supposes that each safety factor follows a lognormal distribution. The *seismic capacity* A_c corresponds to the conditional distribution $A|(Z \geq C)$. A simplified model assumes that A_c is defined by the product of the safety factors, so that it is also lognormally distributed. Denote by $(F_i)_{1\leq i\leq n}$ the safety factors with medians $(\check{F}_i)_{1\leq i\leq n}$ and logarithmic standard deviation $(\beta_i)_{1\leq i\leq n}$, the ground acceleration capacity A_c is then defined by:

$$A_c = \prod_{i=1}^{n} F_i \, . \tag{2.3}$$

Moreover, the seismic capacity randomness is divided into an epistemic and aleatory part:

$$A_c = A_m \varepsilon_R \varepsilon_U \, , \tag{2.4}$$

where A_m is the median capacity, ε_R and ε_U are two lognormal random variables with unit median and log standard deviation respectively β_R and β_U. Consider the case where there are no epistemic uncertainties (i.e. $\beta_U = 0$). The fragility curve for an acceleration level a is in this case defined by:

$$\psi(a) = \mathbb{P}(A_c \leq a) = \Phi\left(\frac{\log\left(a/A_m\right)}{\beta_R}\right) \, , \tag{2.5}$$

where Φ is the cdf of the standard Gaussian distribution $\mathcal{N}(0,1)$. Consider now that $\beta_U > 0$, the fragility curve at ground acceleration a and confidence level Q is defined by:

$$\psi(a;Q) = \Phi\left(\frac{\log\left(a/A_m\right) + \beta_U \Phi^{-1}(Q)}{\beta_R}\right) , \tag{2.6}$$

This confidence level Q provides conservative estimation of the fragility curve w.r.t. the lack of knowledge of the system due to the epistemic uncertainties modeled by the random variable ε_U. Indeed, we can notice that:

$$\mathbb{P}_{\varepsilon_U}(\psi(a;Q) \geq \mathbb{P}_{\varepsilon_R}(A_c \leq a|\varepsilon_U)) = Q ,$$

where for a random variable X, the notation $\mathbb{P}_X$ corresponds to the probability measure defined w.r.t X. In practice, the fragility curve is obtained using tabulated values for the median of the safety factors $\check{F}_i$, their log standard deviation β_i which are divided into a random parts and an epistemic part (the interested reader can consult Table 2 in R. Kennedy and Ravindra, 1984 for a practical example).

Remark that in this framework, the fragility curve defined as $\mathbb{P}(A_c \leq a) = \mathbb{P}(A \leq a|Z \geq C)$ is not the same as the definition given in Equation 2.1. However, it is possible to retrieve the original fragility curve using Bayes theorem and total probability theorem:

$$\begin{aligned} \mathbb{P}(A \leq a|Z \geq C) &= \frac{\mathbb{P}(Z \geq C|A \leq a)\mathbb{P}(A \leq a)}{\mathbb{P}(Z \geq C)} \\ \mathbb{P}(Z \geq C|A \leq a) &= \int_0^a \mathbb{P}(Z \geq C|A = \alpha) d\mathbb{P}_A(\alpha) . \end{aligned} \tag{2.7}$$

The safety factor methodology estimates a fragility curve which is the cumulative distribution function of the random variable $A|(Z \geq C)$, and it can be related to the fragility curve definition in Equation 2.1 thanks to Equation 2.7. In the rest of this manuscript, we will only study the fragility curve defined in Equation 2.1.

2.2.5 Seismic fragility curves estimation using maximum likelihood and the lognormal model

This approach is based on real observations on damaged structures after a seismic excitation, it allows the estimation of fragility curves of real structures submitted to real seismic excitations, without requiring expert knowledge or numerical simulations of the structure. However, the efficiency of this method is limited due to the data scarcity (this method of estimation is only possible in countries with huge seismic activity such as the USA or Japan). Moreover, the inference is also challenging due to the binary nature of the data (the structure is considered damaged or not). However, this method can be easily generalized to the case where the mechanical demand is obtained through numerical simulations.

The most classical approach for fragility curve estimation was derived in Shinozuka et al., 2000, it was used to estimate fragility curves of bridges after the 1995 Kobe earthquake: Consider a dataset $\mathcal{D}_n = (A_i, S_i)_{1 \leq i \leq n}$ where A_i is the intensity measure of the i-th bridge (the intensity measure chosen in Shinozuka et al., 2000 was the PGA) and S_i is a binary random variable equal to 0 if no damage is observed on the i-th bridge and 1 if it is considered damaged. A lognormal assumption is considered for the fragility curve ψ. Defining $\theta = (\alpha, \beta)^T \in (0, +\infty)^2$, the fragility curve takes the following analytical form:

$$\psi_\theta(a) = \Phi\left(\frac{\log\left(\frac{a}{\alpha}\right)}{\beta}\right) , \tag{2.8}$$

we can thus write the likelihood using the lognormal model of the fragility curve:

$$\mathcal{L}(\mathcal{D}_n;\theta) = \prod_{i=1}^{n} \psi_\theta(A_i)^{S_i}(1-\psi_\theta(A_i))^{1-S_i} , \tag{2.9}$$

The estimation of the fragility curve thus boils down to the estimation of the parameter θ. The maximum likelihood estimator (MLE) then writes:

$$\widehat{\theta}_n^{\mathrm{MLE}} = \underset{\theta\in(0,+\infty)^2}{\operatorname{argmax}} \mathcal{L}(\mathcal{D}_n;\theta) . \tag{2.10}$$

The uncertainty of the estimator $\widehat{\theta}_n^{\mathrm{MLE}}$ is classically assessed using the bootstrap method (see Hastie, Tibshirani, and Friedman, 2001). The nonparametric bootstrap method consists in sampling $m \leq n$ points in $\mathcal{D}_n$ with repetition B times to obtain B dataset $(\mathcal{D}^*_{m,b})_{1\leq b\leq B}$ of size m. The bootstrap estimators $(\theta^{\mathrm{MLE},*}_{m,b})_{1\leq b\leq B}$ are obtained by solving:

$$\theta^{\mathrm{MLE},*}_{m,b} = \underset{\theta\in(0,+\infty)^2}{\operatorname{argmax}} \mathcal{L}(\mathcal{D}^*_{m,b};\theta) . \tag{2.11}$$

In Shinozuka et al., 2000, a parametric bootstrap method is proposed. It consists in generating n samples $(S^*_i)_{1\leq i\leq n}$ such that $S^*_i \sim \mathcal{B}(\psi_{\widehat{\theta}_n^{\mathrm{MLE}}}(A_i))$ where $\mathcal{B}(p)$ denotes the Bernouilli distribution with parameter p. This procedure is repeated B times in order to have B datasets of the form $(A_i, S^*_i)_{1\leq i\leq n}$. The bootstrap estimators $(\theta^{\mathrm{MLE},*}_{n,b})_{1\leq b\leq B}$ are computed in the same fashion as in Equation 2.11.

The interested reader will remark that the parametric formulation of the seismic fragility curve define in Equation (2.8) seems arbitrary. It comes actually from the pioneering works of R. Kennedy, C. Cornell, et al., 1980 which motivates the use of such a parametric form. Moreover, this model is directly linked to the probit model (see Finney, 1971). Another well known model for conditional probability estimation is the logit model (or logistic regression) and can also be used for seismic fragility curve estimation (see e.g. Sainct et al., 2020).

2.2.6 Fragility curve estimation using linear regression in log-space

One of the main advantages of estimating fragility curve using numerical simulation is the possibility to define an engineering demand parameter (EDP) which is a positive scalar. The EDP can be an interstory drift, a local strain or stress, a maximal displacement during a time frame... A widely used method to estimate fragility curves is to suppose a lognormal regression model between the IM of the seismic ground motion and the EDP (Ellingwood and Kinali, 2009; Zentner, Gündel, and Bonfils, 2017). Denote by $\mathcal{D}_n = (A_i, Z_i)_{1\leq i\leq n}$ a dataset of n numerical simulations using artificial seismic signals of intensity measure A_i and for each simulation the EDP obtained is Y_i. The EDP Z is thus modeled as a lognormal random variable such that:

$$Z = bA^c\eta , \tag{2.12}$$

where η is a standard lognormal random variable with median 1 and logarithmic standard deviation σ. b, and c are two positive parameters. Using a logarithmic transformation $Y = \log(Z)$ we can write that

$$Y = \log(b) + c\log(A) + \sigma\varepsilon \,, \tag{2.13}$$

where $\varepsilon \sim \mathcal{N}(0,1)$ is a standard Gaussian r.v. Using the definition of a fragility curve one can write:

$$\psi^{\mathrm{LR}}(a) = \Phi\left(\frac{\log(b) + c\log(a) - \log(C)}{\sigma}\right) \tag{2.14}$$

We can use the dataset $\mathcal{D}_n = (A_i, Z_i)_{1\leq i\leq n}$ to estimate the two parameters b and c of the log-linear model. We obtain the following least-squares estimators:

$$\widehat{b}_n, \widehat{c}_n = \underset{b,\, c\, \in\mathbb{R}_+\setminus\{0\}\times\mathbb{R}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \log(b) - c\log(A_i))^2 \,. \tag{2.15}$$

The residual standard deviation σ is estimated using the empirical estimator $\widehat{\sigma}_n$:

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \log(\widehat{b}_n) - \widehat{c}_n\log(A_i))^2 \,, \tag{2.16}$$

the estimator of the fragility curve using linear regression then writes:

$$\widehat{\psi}_n^{\mathrm{LR}}(a) = \Phi\left(\frac{\log(\widehat{b}_n) + \widehat{c}_n\log(a) - \log(C)}{\widehat{\sigma}_n}\right) \,. \tag{2.17}$$

One main advantage of this method is the possibility to estimate the fragility for any failure criterion C with a learning sample $\mathcal{D}_n$. LS estimators derived from Equation 2.15 can be expressed analytically. Of course, the MLE estimator defined in Equation 2.10 can also be used when numerical simulations are carried out by defining $S_i = \mathbb{1}_{Z_i>C}$.

2.3 Seismic fragility curves estimation using numerical simulations

Due to the increase of the computational power in the recent years and due to the lack of real data of mechanical structures damages subjected to earthquakes, it is now more common to replace real experiments (or empirical data) by so-called computer experiments, based on numerical models for both the structure studied and the seismic ground motions (see for examples Karim and Yamazaki, 2001; Kim and Shinozuka, 2004; Zentner, 2010). It allows to study the robustness of structures to earthquakes without relying on costly real experiments.

The use of computer experiments imposes to propose a numerical generator of seismic ground motions. More precisely, the goal is to have a numerical model that generates signals of ground accelerations $t \to a(t)$ for $t \in [0, T]$ which have the same properties as real ground accelerations of an earthquake. After this step, we can proceed the dynamical analysis of the structure studied using synthetic seismic signals generated using the seismic ground motion generator. There are two types of stochastic ground motions models: the source-based model that propagates seismic waves in a ground medium after random occurrences of fault ruptures at a source point and site-based model that are obtained by fitting the model on known earthquakes at a given site. For

more details see the review in Rezaeian and Der Kiureghian, 2008. Basically, seismic ground acceleration signals are usually represented as a filtered and modulated white noise process (see e.g. Rodolfo Sargoni and Hart, 1973; Rezaeian and Der Kiureghian, 2010) with a deterministic modulating function and a power spectral density which depends on the physical characteristics of the earthquake source and the ground materials. In this manuscript, we will use the stochastic ground motion model developed in Rezaeian and Der Kiureghian, 2008; Rezaeian and Der Kiureghian, 2010 which has the advantage of modeling both temporal and spectral nonstationarities of real seismic acceleration signals.

2.3.1 Ground motion generation using Rezaeian's model

For this manuscript, we have chosen to enrich a set of seismic acceleration records selected in a real ground motion database using the model defined in Rezaeian and Der Kiureghian, 2010. This generator is a parametrized modulated and filtered white-noise process. Its main advantage is its capacity to reproduce both temporal and spectral non-stationarities of real seismic acceleration signals, and it has been used in several recent works (see e.g. Mai, Konakli, and Sudret, 2017; Sainct et al., 2020).

Following this model, a seismic ground motion acceleration temporal signal $s(t)$ with $t \in [0, T]$ writes as:

$$s(t) = \frac{q(t, \boldsymbol{\alpha})}{\sigma_f(t)} \int_{-\infty}^{t} h(t - \tau, \boldsymbol{\beta}(\tau)) w(\tau) d\tau \, . \tag{2.18}$$

$q(t, \boldsymbol{\alpha})$ is a deterministic non-negative piecewise modulating function that is defined as:

$$q(t, \boldsymbol{\alpha}) = \begin{cases} \alpha_1 t^2 / T_1^2 & \text{if } 0 \leq t \leq T_1 \\ \alpha_1 & \text{if } T_1 \leq t \leq T_2 \\ \alpha_1 \exp(-\alpha_2 (t - T_2)^{\alpha_3}) & \text{if } T_2 \leq t \leq T \end{cases} \tag{2.19}$$

This modulating function depends on the vector of parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, T_1, T_2) \in \mathbb{R}_+^5$. $w(t)$ is a white-noise process and $h(t, \boldsymbol{\beta}(t))$ is the Impulse Response Function (IRF) of a linear filter that depends on the vector of parameters $\boldsymbol{\beta}$. $\sigma_f(t)^2 = \int_{-\infty}^{t} h(t - \tau, \boldsymbol{\beta}(\tau))^2 d\tau$ is the variance of the stochastic process defined in the integral in Equation 2.18. Remark that $\boldsymbol{\beta}$ depends on time in order to achieve spectral nonstationarity of the ground motion. The IRF writes as follows:

$$h(t - \tau, \boldsymbol{\beta}(\tau)) = \frac{\omega_f(\tau)}{\sqrt{1 - \zeta_f^2}} \exp(-\zeta_f \omega_f(\tau)(t - \tau)) \sin(\omega_f(\tau) \sqrt{1 - \zeta_f^2}(t - \tau)) \mathbb{1}_{t \geq \tau} \, , \tag{2.20}$$

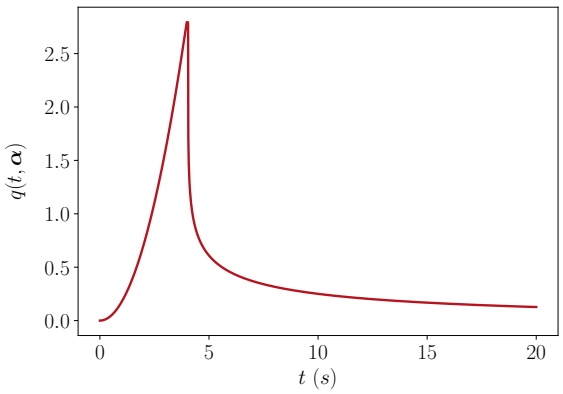
where $\boldsymbol{\beta}(\tau) = (\omega_f(\tau), \zeta_f)$, $\omega_f(\tau)$ is the natural time-dependent pulsation and $\zeta_f \in [0, 1]$ is the constant damping ratio of the linear filter. A linear form $\omega_f(\tau) = \omega_0 + \frac{t}{T}(\omega_n - \omega_0)$ is chosen for the time-dependent pulsation. Therefore, the IRF is parametrized by $\boldsymbol{\lambda} = (\omega_0, \omega_n, \zeta_f) \in \mathbb{R}_+^3$.

The ground motion model parameters $\boldsymbol{\theta}^{\mathrm{GM}} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}) \in \mathbb{R}_+^8$ are identified for each of the N_r real seismic signals. This gives a dataset $(\boldsymbol{\theta}_i^{\mathrm{GM}})_{1 \leq i \leq N_r}$ of N_r seismic ground motion model parameters and synthetic seismic signals can be generated by sampling several realizations of the white-noise w. However, these signals will have features very similar to those of the real seismic signals present in the database. In order to generate more diverse synthetic seismic signals, a second level of uncertainty is added by considering that the model parameters $\boldsymbol{\theta}^{\mathrm{GM}}$ is a random variable with probability density function π. The pdf π is estimated on $(\boldsymbol{\theta}_i^{\mathrm{GM}})_{1 \leq i \leq N_r}$ by Kernel Density

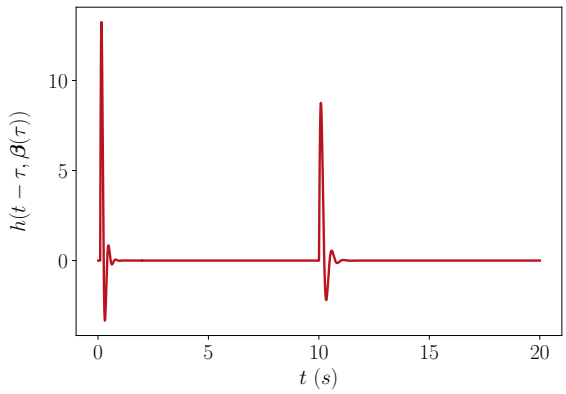
Estimation (KDE) using a Gaussian kernel in $\mathbb{R}^8$, this gives us the pdf π_{KDE}. Finally, the generation of seismic synthetic ground motions requires two steps:

1. Sample $\boldsymbol{\theta}^{\text{GM}} \sim \pi_{\text{KDE}}$;
2. sample a realization of the white noise $w(t)$ and compute the signal $s(t)$ with parameter $\boldsymbol{\theta}^{\text{GM}}$.

It is thus possible with this stochastic ground motion model to generate artificial seismic ground motions that are fitted on a dataset of real earthquakes. This is a site-based method of seismic ground motion generation as explained in Rezaeian and Der Kiureghian, 2008.

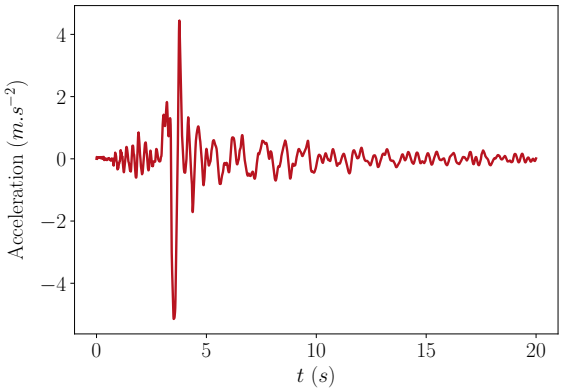


(A) Modulating function of the stochastic ground motion model

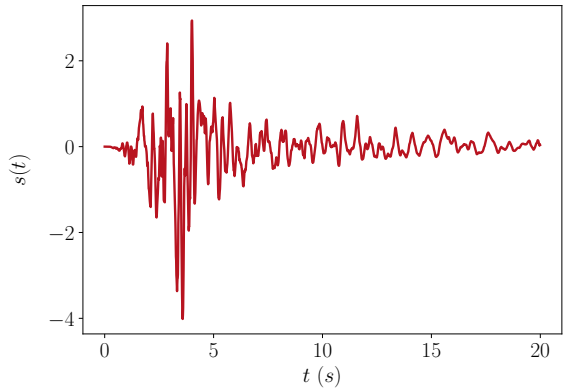


(B) Response of the time-varying filter to unit pulses at two time points $\tau = 10^{-2}\ s.$ and $\tau = 10\ s.$

FIGURE 2.1: Graphical representation of the modulating function and IRF of the linear filter with parameters θ^{GM} fitted on a real earthquake acceleration signal.



(A) Real earthquake acceleration signal.



(B) Simulated earthquake

FIGURE 2.2: Real and simulated acceleration signal of an earthquake using the model of Rezaeian. The parameters of the stochastic ground motion model are fitted on the real earthquake signal.

Figure 2.1 shows the modulating function and IRF of the ground motion model whose parameters have been fitted on an acceleration signal record of the European Strong Motion Database (Ambraseys et al., 2004) in the domain $5.5 < M < 6.5$and $R < 20$km. Figure 2.2 compares the

original acceleration signal record to a realization of the ground motion stochastic process after parameters fitting.

2.3.2 Reference fragility curves obtained by nonparametric estimation

The previous methods for fragility curves estimation rely on a lognormal assumption which can be questionable in practice. Recently, nonparametric methods for estimating fragility curves have been proposed. Mai, Konakli, and Sudret, 2017 compare MLE estimation of fragility curves with binned Monte-Carlo simulations and kernel density estimation of the joint density of (A, Z) and the density of the intensity measure A. Lallemant, Kiremidjian, and Burton, 2015 propose to estimate fragility curves using generalized additive models and Gaussian kernel smoothing. Sainct et al., 2020 perform nonparametric seismic fragility curve estimation using support vector machines. Trevlopoulos, Feau, and Zentner, 2019 propose to perform nonparametric estimation of fragility curves using Monte-Carlo simulation and parametric models averaging based on K-means clustering of the seismic intensity measures data. We will develop the latter nonparametric estimation method in this section.

Let us consider N_c clusters and their associated centroids $(C_i)_{1\leq i\leq N_c}$. The seismic fragility curve value at each centroid is approximated by the empirical probability of failure in each cluster.

$$\widehat{\psi}_n^{\mathrm{MC}}(C_i) = \frac{1}{n_i} \sum_{Z_j \in K_i} \mathbb{1}_{Z_j > C} \ , \tag{2.21}$$

where K_i denotes the ith cluster for $1 \leq i \leq N_c$, and n_i the sample size of the ith cluster. The overall fragility curve is then approximated using linear interpolation between IM values. Confidence intervals for this nonparametric fragility curve can be considered using the asymptotic Gaussian distribution of estimator $\widehat{\psi}_n^{\mathrm{MC}}(C_i)$. Indeed, its asymptotic variance can be approximated by

$$\mathrm{Var}(\widehat{\psi}_n^{\mathrm{MC}}(C_i)) \approx \frac{\widehat{\psi}_n^{\mathrm{MC}}(C_i)(1 - \widehat{\psi}_n^{\mathrm{MC}}(C_i))}{n_i} \tag{2.22}$$

This Monte-Carlo approximation of the seismic fragility curve will be used as a reference fragility curve to compare between different parametric models.

2.4 Application: a single d.o.f. oscillator with nonlinear restoring force

This section aims to propose a practical example of seismic fragility curve estimation by statistical methods using a simplified mechanical model that describes well essential features of nonlinear responses in some real structures. The model is a nonlinear single degree of freedom oscillator. Its equation of motion is:

$$\ddot{z}(t) + 2\xi\omega\dot{z}(t) + f^{\mathrm{NL}}(z(t)) = -s(t) \ , \tag{2.23}$$

where $\dot{z}$ and $\ddot{z}$ are the relative velocity and acceleration of the unit mass of the system submitted to the acceleration signal s with null initial condition in velocity and displacement. ξ is the damping ratio, $\omega = 2\pi f$ the circular frequency of the oscillator, and f^{NL} is the nonlinear restoring force. Figure 2.3 represents the behavior of f^{NL} relatively to the yield displacement z_d and the post-yield stiffness α_y. Figure 2.4 is a schematic representation of the nonlinear oscillator with its parameters. This toy case has the advantage to have a fast computation of the relative displacement z thanks to the finite-difference method.

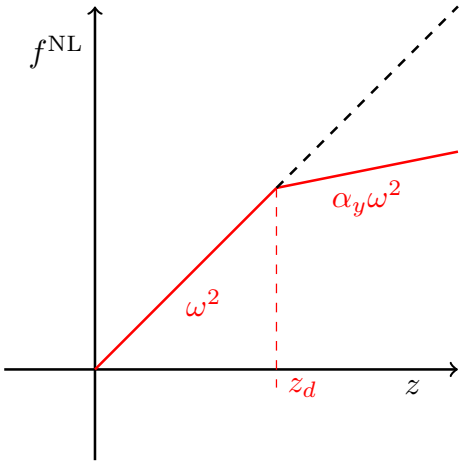


FIGURE 2.3: Schematic representation of the nonlinear restoring force f^{NL} w.r.t. the relative displacement of the oscillator z.

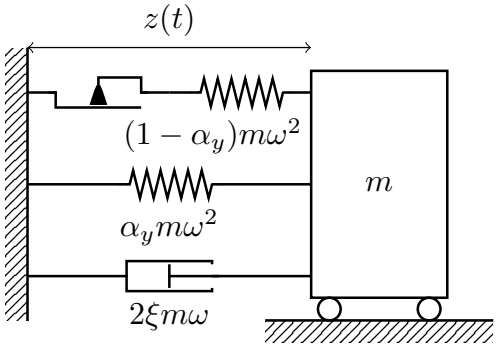


FIGURE 2.4: Nonlinear single degree of freedom oscillator with kinematic hardening.

In this section, $f = 5$ Hz and $\xi = 2\%$. The yield displacement $z_d = 5.10^{-3}\ m$ and the post yield stiffness $\alpha_y = 0.2$. The engineering demand parameter in this example is defined by:

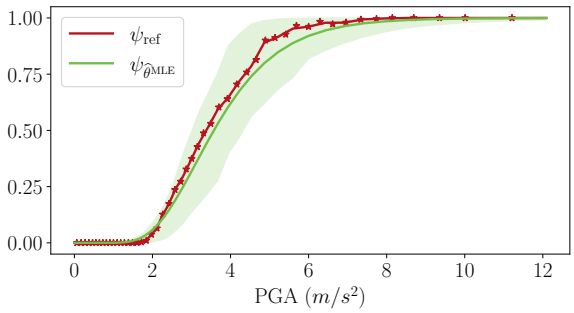
$$Z = \max_{t\in[0,T]} |z(t)| \ . \tag{2.24}$$

It corresponds to the maximal absolute displacement of the nonlinear oscillator during the seismic excitation. The intensity measure used in this example is the Peak Ground Acceleration (PGA) and it is defined as:

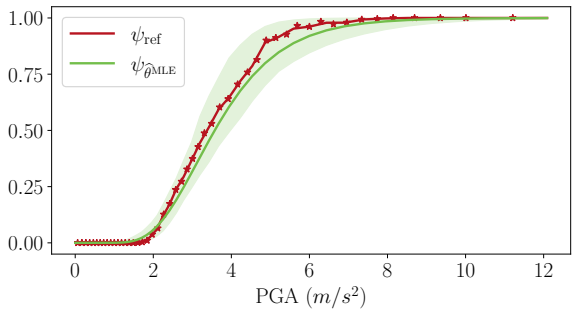
$$A = \max_{t\in[0,T]} |s(t)| \ . \tag{2.25}$$

The PGA is thus the maximal absolute acceleration of the seismic signal. 10^6 synthetic seismic ground motions are generated using the model described in Section 2.3.1 using the data from Sainct et al., 2020. A learning sample of $n = 200$ simulations of the nonlinear oscillator is considered to estimate the fragility curve with a failure threshold $C = 1.6z_d$ using the lognormal model (Section 2.2.5) and the linear regression model (Section 2.2.6). The failure threshold of $C = 1.6z_d$ roughly corresponds to the 90% level quantile of the distribution of Z on the overall dataset of 10^6 synthetic seismic ground motions. Uncertainty on the estimators is assessed using parametric and nonparametric bootstrap described in Section 2.2.5 with 100 bootstrap samples of size 100. Numerical results are shown in Figure 2.5. Remark that the bootstrap confidence interval is narrower

for the parametric bootstrap than for the nonparametric bootstrap. This is expected due to the more restrictive assumption in the parametric bootstrap method, where the model for the fragility curve is supposed well-specified (i.e. the true fragility curve is lognormal).



(A) Estimation of the fragility curve of the nonlinear oscillator with nonparametric bootstrap using the lognormal model.



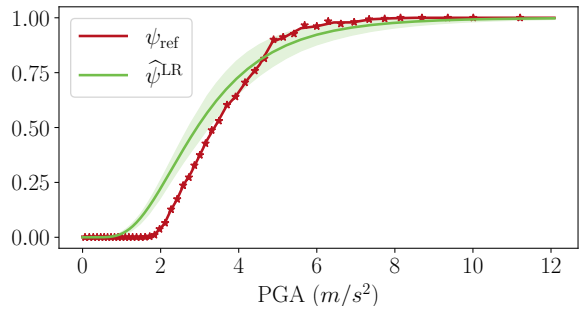
(B) Estimation of the fragility curve of the nonlinear oscillator with parametric bootstrap using the lognormal model.

FIGURE 2.5: Fragility curve estimation using the lognormal model with maximum likelihood estimation. The green shaded area corresponds to the interquantile range of the fragility curve between the 10% and 90% level quantiles for each PGA value, estimated using the two bootstrap methods. The red solid curve corresponds to the nonparametric estimation of the fragility curve using K-means clustering with the 10^6-sized dataset with $K = 50$.

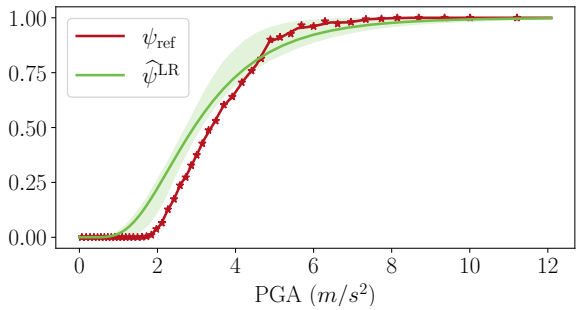
The fragility curve of the nonlinear oscillator is also evaluated using linear regression in log space as described in Section 2.2.6. The same learning set of size $n = 200$ is used, estimation uncertainty is assessed using parametric and nonparametric bootstrap in the same fashion as in Section 2.2.5. Figure 2.6 shows the numerical results of the estimation. Remark that the confidence intervals are narrower than with the lognormal assumption of the fragility curve. However, the bias between the estimator of the fragility curve and the reference fragility curve is larger than with MLE estimation using the lognormal fragility curve model. An interpretation can be proposed thanks to Zentner, Gündel, and Bonfils, 2017: the linear regression can extrapolate the behavior of the structure even when no failure event $(Z \geq C)$ is observed in the training data. This property makes the estimation of the fragility curve with small sample size possible, but with a possible large bias. On the contrary, a sufficient amount of failure event in the training data is needed to propose a good estimation of the fragility curve with MLE estimation, this permits to decrease the bias, but it also imposes to have a bigger sample size to provide a narrow confidence interval.

2.5 Conclusion

A brief review of the different statistical methods to estimate seismic fragility curves has been proposed in this chapter, from the early work in the 1980s by R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984 which propose an estimation based on expert judgment to statistical methods based on numerical simulations. A classical example studied in earthquake engineering consisting of a nonlinear single degree of freedom oscillator with kinematic hardening has been presented in order to illustrate the different statistical methods of fragility curves estimation. For the rest of this manuscript, we will focus mainly on estimation of fragility curves using numerical simulations. Indeed, the next chapter will develop a methodology to improve the estimation of seismic fragility curves using importance sampling, allowing to reduce the estimation uncertainty and the learning sample size in the same time.



(A) Estimation of the fragility curve of the nonlinear oscillator with nonparametric bootstrap.



(B) Estimation of the fragility curve of the nonlinear oscillator with parametric bootstrap.

FIGURE 2.6: Fragility curve estimation using linear regression. The green shaded area corresponds to the interquantile range of the fragility curve between the 10% and 90% level quantiles for each PGA value, estimated using the two bootstrap methods. The red solid curve corresponds to the nonparametric estimation of the fragility curve using K-means clustering with the 10^6-sized dataset with $K = 50$.

CHAPTER 3

Importance sampling-based active learning for parametric seismic fragility curve estimation

Contents

This chapter consists in the submitted paper:

Gauchy, C., C. Feau, and J. Garnier (2021a). *Importance sampling based active learning for parametric seismic fragility curve estimation*. DOI: 10.48550/ARXIV.2109.04323

3.1 Introduction

3.1.1 Contribution

The aim of this work is to address the issue of optimized estimation of intensity measure (IM) based seismic fragility curves within a rigorous mathematical framework when complex numerical models - i.e expensive numerical models - are used to obtain the data. In our case, the complex numerical codes are "black boxes" whose results, for a given IM value, are simply binary, namely: failure or not failure.

To tackle this issue, we propose and implement an active learning methodology based on adaptive importance sampling as in Chu et al., 2011 in a statistical learning context (Hastie, Tibshirani, and Friedman, 2001), called Importance Sampling based Active Learning (IS-AL). In addition, we address the question of Confidence Intervals (CIs) related to the size of the samples used, in order to reflect the estimation uncertainty, because bootstrap method cannot be used.

Adaptive importance sampling was introduced for classical Monte Carlo integral approximation in Kloek and Van Dijk, 1978, and later studied in Oh and J. Berger, 1992. Moreover, adaptive importance sampling is also used in industrial applications and have already been discussed, implemented and tested for probability estimation of rare event (e.g. failure state) in reliability analysis as in Gong and W. Zhou, 2018; Papaioannou, Geyer, and Straub, 2019. By applying it to the parametric estimations of the IM-based fragility curves, we show by asymptotic analysis and numerical simulations that IS-AL allows for (i) a rapid convergence of the estimated fragility curve towards the true (unknown) fragility curve and (ii) a rigorous quantification of the estimation uncertainty. It gives asymptotic CIs and confidence ellipsoids for the quantities of interest as well as statistical tests to determine whether the asymptotic regime has been reached and whether asymptotic CIs and confidence ellipsoids can be used.

The proposed methodology relies on parametric approximations of fragility curves for any IM of interest. Although the validity of parametric models is both questionable and difficult to assess (see e.g. Lallemant, Kiremidjian, and Burton, 2015; Mai, Konakli, and Sudret, 2017; Trevlopoulos, Feau, and Zentner, 2019; Zentner, 2017), some numerical experiments based on the seismic responses of simple mechanical systems - i.e. few degrees of freedom systems - suggest that the choice of an appropriate IM makes it possible to reduce the potential biases between reference fragility curves - that can be obtained by massive Monte Carlo simulations - and their parametric approximations. This point is illustrated in the application section 3.6 of this chapter. Remember, however, that in practice, the selection of an optimal IM is not a trivial matter (see e.g. Hariri-Ardebili and Saouma, 2016a; Ciano, Gioffrè, and Grigoriu, 2020) and Machine Learning techniques can be used for this purpose (e.g. Sainct et al., 2020), knowing that the references A. Cornell, 2004; Luco and A. Cornell, 2007 and Padgett, Nielson, and DesRoches, 2008 give optimality criteria for selection of such IM.

In this work, the methodology is applied to different test cases and compared with more traditional approaches such as MLE often used by practitioners (see e.g. Baker, 2015; Lallemant, Kiremidjian, and Burton, 2015; Straub and Der Kiureghian, 2008; Shinozuka et al., 2000; F. Wang and Feau, 2020; Mai, Konakli, and Sudret, 2017; Zentner, 2010). In order to avoid the scaling of

the accelerograms, the stochastic model of modulated and filtered white-noise process defined in Rezaeian and Der Kiureghian, 2010 (see Chapter 2, Section 2.3.1) is used to enrich a set of real ground motion records selected in a database for a given (M,R) scenario. This stochastic model is chosen because it well encompasses the temporal and spectral non-stationarities of real seismic signals. Furthermore, it has already been used in several works (see e.g. Mai, Konakli, and Sudret, 2017; Sainct et al., 2020; Kwong and Chopra, 2015).

3.1.2 Organization

In section 3.2 the statistical framework is defined for any parametric fragility curve model and any IM. Section 3.3 is dedicated to the presentation of the IS-AL algorithm applied to IM-based seismic fragility curves estimation for the commonly used lognormal model. Section 3.4 summarizes the main theoretical results of this work, which are proved in the appendices. These results concern a criterion for evaluating the convergence of the IS-AL strategy and the definition of asymptotic confidence ellipsoids for the fragility curve parameters. Section 3.5 presents the performance metrics used in this work to compare IS-AL, random sampling and MLE strategies. Finally, in section 3.6, IS-AL performance is assessed on analytical and industrial test cases of increasing complexity.

3.2 Seismic fragility curve estimation: a statistical learning framework

3.2.1 General framework

We consider the following situation. Let $\mathcal{X}$ be a compact set of $\mathbb{R}$, X a $\mathcal{X}$-valued random variable and $S \in \{0,1\}$ a random label. In SPRA studies $X = \log \text{IM}$ - more generally we can define $X = \psi(\text{IM})$ where ψ is an increasing function of IM such as a Box-Cox transform (G. E. Box and Cox, 1964) - and S is the indicator variable of the failure of the structure. The pair (X, S) has the probability distribution P over $\mathcal{X} \times \{0,1\}$:

$$P(dx, ds) = \big[\mu(x)\delta_1(ds) + (1-\mu(x))\delta_0(ds)\big]p(x)dx \,, \tag{3.1}$$

where δ_j is the Dirac distribution at j, p is the marginal probability density function (pdf) of X, and the IM-based fragility curve $\mu(x)$ is the conditional expectation of S (i.e. the conditional probability of failure):

$$\mu(x) = \mathbb{E}[S|X = x] \,. \tag{3.2}$$

The aim of the paper is to estimate the curve $\mu(x)$ from datapoints $(X_i, S_i)_{i=1}^n$ that may be independent and identically distributed with the distribution P or that may be selected by a more appropriate scheme. As mentioned in the introduction, it is a classical assumption to use a parametric form for the fragility curve μ to tackle the need for time consuming mechanical simulations, we thus consider the space of functions $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$, where $x \mapsto f_\theta(x)$ is a function from $\mathbb{R}$ to $[0,1]$ for any θ and $\Theta \subset \mathbb{R}^m$. The goal is to minimize the quadratic risk:

$$g(\theta) = \mathbb{E}[(\mu(X) - f_\theta(X))^2] \,, \tag{3.3}$$

in order to find (provided it exists and is unique):

$$\theta_* = \underset{\theta\in\Theta}{\operatorname{argmin}}\, g(\theta) \,. \tag{3.4}$$

Unfortunately, the observable data are $(X_i, S_i)_{i=1}^n$, we do not observe directly $\mu(X_i)$. But considering that:

$$\begin{aligned}\mathbb{E}[(S - f_\theta(X))^2] &= \mathbb{E}[(S - \mu(X))^2] + \mathbb{E}[(\mu(X) - f_\theta(X))^2] \\ &\quad + 2\mathbb{E}[(\mu(X) - f_\theta(X))(S - \mu(X))] \\ &= \mathbb{E}[\mu(X)(1 - \mu(X))] + \mathbb{E}[(\mu(X) - f_\theta(X))^2] \end{aligned} \tag{3.5}$$

because $\mathbb{E}[S^2|X] = \mathbb{E}[S|X] = \mu(X)$, we can observe that the minimization with respect to θ of $\mathbb{E}[(S - f_\theta(X))^2]$ is equivalent to the minimization of $\mathbb{E}[(\mu(X) - f_\theta(X))^2]$. Hence, we will consider the quadratic risk

$$r(\theta) = \mathbb{E}[(S - f_\theta(X))^2] \, . \tag{3.6}$$

In the context of classical learning, when we observe n datapoints $(X_i, S_i)_{i=1}^n$ drawn independently from the probability distribution $P(dx, ds)$ over $\mathcal{X} \times \{0, 1\}$, the expectation can be approximated by the empirical mean:

$$\widehat{R}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} (S_i - f_\theta(X_i))^2 \, . \tag{3.7}$$

The corresponding passive estimator (the term passive is used to highlight the absence of any particular sampling strategy) is then:

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \, \widehat{R}_n(\theta). \tag{3.8}$$

Conversely to classical learning, active learning aims at selecting the most useful numerical experiments to be carried out in order to form the learning set. In the passive strategy, the datapoints X_i are sampled from the original probability distribution with pdf p drawn from a stochastic ground-motion model. In the same way as in Chu et al., 2011, we propose an active learning strategy, called Importance Sampling based Active Learning (IS-AL). It consists to draw the datapoints X_i from an instrumental probability distribution with pdf q that is chosen in an adaptive way. In our context, it is straightforward to use a rejection method applied to the stochastic ground-motion model in order to generate seismic loads with a desired intensity measure distribution. Let us recall in fact that the label S_i (which gives the failure state of the structure) is, in our case, expensive to obtain because it comes from complex numerical simulations of mechanical structures while the artificial seismic signals are inexpensive to generate.

The main objective of this procedure is to reduce the variance implied by the empirical approximation of the quadratic risk $r(\theta)$. Importance Sampling is a classical variance reduction technique for Monte Carlo estimation used in structural reliability (Papaioannou, Geyer, and Straub, 2019; Zuniga, Murangira, and Perdrizet, 2021). If the $(X_i)_{i=1}^n$ are sampled with the pdf q and $(S_i)_{i=1}^n$ are the labels obtained from n calls to the mechanical model, then the importance sampling estimator of the empirical quadratic risk is:

$$\widehat{R}_n^{\mathrm{IS}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{p(X_i)}{q(X_i)} (S_i - f_\theta(X_i))^2. \tag{3.9}$$

In the rest of the chapter, we will denote by $r(\theta) = \mathbb{E}_{(X,S)\sim P}[\ell_\theta(X, S)]$ with $\theta \mapsto \ell_\theta(x, s)$ a positive loss function for the sake of generalization. For the numerical applications, only the case of the quadratic loss

$$l_\theta(x, s) = (s - f_\theta(x))^2, \tag{3.10}$$

will be considered.

3.2.2 Problem regularization for the parametric lognormal model

For applications to seismic fragility curves estimation, a classical space of functions to approximate μ is $\mathcal{F} = \{\Phi(\frac{\log(\mathrm{IM}/\alpha)}{\beta}), (\alpha, \beta) \in \Theta\}$ where Φ is the cumulative distribution function of the standard Gaussian distribution (R. Kennedy, C. Cornell, et al., 1980), $\theta = (\alpha,\ \beta)^T$, and Θ a compact set of $(0, +\infty)^2$ (therefore $m = 2$). Compactness of Θ is a common assumption in our applications. From an engineer perspective, it is possible to bound α and β. However, in practice, the lower bound for β may be reached by the different estimators. Consequently, inspired by Bayesian inference theory (G. Box and Tiao, 1973), we introduce a regularization term $\Omega(\theta; \beta_{\mathrm{reg}})$ to tackle this issue (we will take $\Omega(\theta; \beta_{\mathrm{reg}}) = \beta_{\mathrm{reg}}/\beta$ below). The squared loss (3.7) is then replaced by:

$$\widehat{R}_{n,\mathrm{reg}}(\theta; \beta_{\mathrm{reg}}) = \frac{1}{n}\sum_{i=1}^{n} \ell_\theta(X_i, S_i) + \frac{\Omega(\theta; \beta_{\mathrm{reg}})}{n} \,. \tag{3.11}$$

The derivation of the importance sampling estimator of the regularized square loss is straightforward:

$$\widehat{R}^{\mathrm{IS}}_{n,\mathrm{reg}}(\theta; \beta_{\mathrm{reg}}) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q(X_i)}\ell_\theta(X_i, S_i) + \frac{\Omega(\theta; \beta_{\mathrm{reg}})}{n} \,. \tag{3.12}$$

This regularization is motivated by the intrinsic difficulty of estimating the standard deviation β of the lognormal model when β is small (Keller et al., 2015). Fragility curves with small β are hard to distinguish due to the convergence towards a degenerate $0-1$ fragility curve.

3.3 Principles of the IS-AL strategy

This section focuses on the choice of an optimal density q (section 3.3.1) as well as on the description of the IS-AL strategy (section 3.3.2).

3.3.1 Choice of an optimal density q

The heuristic used to find a good instrumental probability distribution family is presented in Chu et al., 2011. The first idea would be to minimize the variance of the importance sampling risk estimator (3.12):

$$\mathrm{Var}\big(\widehat{R}^{\mathrm{IS}}_{n,\mathrm{reg}}(\theta)\big) = \frac{1}{n}\Big\{ \iint_{\mathcal{X}\times\{0,1\}} \frac{p(x)}{q(x)}\ell_\theta(x,s)^2 P(dx, ds) - r(\theta)^2 \Big\} \,, \tag{3.13}$$

with respect to q within the set of all pdfs. If we denote by $\tilde{\ell}^2_\theta(x) = \mathbb{E}[\ell_\theta(X,S)^2 | X = x]$ the squared loss averaged on S:

$$\tilde{\ell}^2_\theta(x) = \mu(x)\ell_\theta(x,1)^2 + (1-\mu(x))\ell_\theta(x,0)^2 \,, \tag{3.14}$$

the variance of the importance sampling risk estimator (3.12) can be expressed as

$$\mathrm{Var}\big(\widehat{R}^{\mathrm{IS}}_{n,\mathrm{reg}}(\theta)\big) = \frac{1}{n}\Big\{ \int_{\mathcal{X}} \frac{p(x)^2}{q(x)}\tilde{\ell}^2_\theta(x)dx - r(\theta)^2 \Big\} \,,$$

and we look for

$$q_\theta^* = \underset{q}{\operatorname{argmin}} \int_{\mathcal{X}} \frac{p(x)^2}{q(x)} \tilde{\ell}_\theta^2(x) dx \, . \tag{3.15}$$

Using Jensen's inequality (Robert and Casella, 2004, Theorem 3.12), we can solve the optimization problem and we can find that the optimal sampling pdf is of the form

$$q_\theta^*(x) \propto \tilde{\ell}_\theta(x) p(x) \, ,$$

which depends on μ because $\tilde{\ell}_\theta$ depends on μ [Here and below $\propto$ means equality up to a multiplicative constant]. Hence an approximation step is made by replacing μ by f_θ in (3.14):

$$\tilde{\ell}_\theta^2(x) \approx f_\theta(x) \ell_\theta(x,1)^2 + (1 - f_\theta(x)) \ell_\theta(x,0)^2 \, . \tag{3.16}$$

Hence the instrumental density becomes:

$$q_\theta(x) \propto p(x) \sqrt{f_\theta(x) \ell_\theta(x,1)^2 + (1 - f_\theta(x)) \ell_\theta(x,0)^2} \, . \tag{3.17}$$

Note that the instrumental distribution depends on θ, the parameter we aim to estimate. Moreover, using IS-AL with the instrumental density q_θ directly could increase the variance if the density has light tails. We propose finally a defensive strategy as illustrated in Hesterberg, 1995; Owen and Y. Zhou, 2000. The instrumental density becomes

$$q_{\theta,\varepsilon}(x) = \varepsilon p(x) + (1 - \varepsilon) q_\theta(x), \tag{3.18}$$

with $\varepsilon \in [0,1]$. ε is a mixing parameter, between the original marginal pdf $p(x)$ and the instrumental one $q_\theta(x)$, meaning that one time out of $1/\varepsilon$ the element is drawn from the pdf $p(x)$. This distribution allows to bound the likelihood ratio:

$$\frac{p(x)}{q_{\theta,\varepsilon}(x)} = \frac{1}{\varepsilon + (1 - \varepsilon) \frac{q_\theta(x)}{p(x)}} < \frac{1}{\varepsilon} \, . \tag{3.19}$$

Thus the defensive strategy bounds the variance even if the likelihood ratio $p(x)/q_\theta(x)$ is large.

3.3.2 Description of the IS-AL strategy

3.3.2.1 Algorithm

The procedure for computing the IS-AL estimator $\widehat{\theta}_n^{\mathrm{IA}}$ is described in Algorithm 1. Its main objective is to use an updated instrumental density $q_{\theta,\varepsilon}$ at each step. Note that (i) the algorithm needs to start from a certain parameter value $\widehat{\theta}_0^{\mathrm{IA}}$ and (ii) we choose $\Omega(\theta; \beta_{\mathrm{reg}}) = \beta_{\mathrm{reg}}/\beta$ for the regularization term in equation (3.12).

Additionally, a convergence criterion is presented in section 3.4.2 and an asymptotic confidence ellipsoid for θ_* centered on $\widehat{\theta}_n^{\mathrm{IA}}$ is defined by equation (3.37).

3.3.2.2 Initialization and choice of $\widehat{\theta}_0^{\mathrm{IA}}$

Regarding the initialization, as expected, the closer $\widehat{\theta}_0^{\mathrm{IA}}$ is from the true parameter θ_* the faster IS-AL converges and is in asymptotic normal regime (see Section 3.4). A naive approach is to get a small sample of size n_0 (e.g. $n_0 = 20$) $(X_i, S_i)_{i=1}^{n_0}$ from the original marginal density p of X and

Algorithm 1 Importance Sampling based Active Learning (IS-AL)

1. Choice of $\widehat{\theta}_0^{\mathrm{IA}}$ (section 3.3.2.2) and estimations of β_{reg} and ε (section 3.3.2.3).
2. For $i = 1, \dots, n$:
   (a) Draw X_i from the distribution with pdf $q_{\widehat{\theta}_{i-1}^{\mathrm{IA}},\varepsilon}$.
   (b) Call the mechanical simulation at point X_i to get label S_i
   (c) Compute

$$\widehat{\theta}_i^{\mathrm{IA}} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \widehat{R}_{i,\mathrm{reg}}^{\mathrm{IA}}(\theta;\beta_{\mathrm{reg}}), \tag{3.20}$$

$$\widehat{R}_{i,\mathrm{reg}}^{\mathrm{IA}}(\theta;\beta_{\mathrm{reg}}) = \frac{1}{i}\sum_{j=1}^{i} \frac{p(X_j)}{q_{\widehat{\theta}_{j-1}^{\mathrm{IA}},\varepsilon}(X_j)}\ell_\theta(X_j,S_j) + \frac{\beta_{\mathrm{reg}}}{n\beta}. \tag{3.21}$$

to compute the passive learning estimator $\widehat{\theta}_{n_0}$ (equation (3.8)). This crude estimation can be used as the initial parameter $\widehat{\theta}_0^{\mathrm{IA}}$ to start IS-AL.

A better approach is to consider a metamodel - in the broad sense - of the mechanical simulation. As often used by practitioners, a numerical resolution based on a modal base projection can be implemented to get an estimate of the fragility curve corresponding to the linear behavior of the structure of interest. It is then possible to get a huge amount of datapoints of the reduced model (e.g. an independent and identically distributed sample of $n_{\mathrm{red}} = 10^3$–10^5 pairs $(X_i, S_{\mathrm{red},i})_{i=1}^{n_{\mathrm{red}}}$ where X_i is sampled with the original pdf p and $S_{\mathrm{red},i}$ is the associated label obtained with the reduced model). The initial parameter $\widehat{\theta}_0^{\mathrm{IA}}$ is then chosen to be equal to $\widehat{\theta}_{n_{\mathrm{red}}}$. Statistical metamodels could also be used such as Gaussian Processes in Echard, Gayton, Lemaire, and Relun, 2013 or Support Vector Machines in Sainct et al., 2020.

In our applications reduced models are only used to give us prior knowledge on the fragility curve shape, encapsulated in the initial parameter of the IS-AL procedure. We then initialize IS-AL with a small sample of 20 datapoints with the instrumental density $q_{\widehat{\theta}_0^{\mathrm{IA}},\varepsilon}$ (equation (3.18)). In other words, in Step 2 of Algorithm 1, we do not update $\widehat{\theta}_i^{\mathrm{IA}}$ during the first 20 steps.

3.3.2.3 Estimations of β_{reg} and ε

The regularization parameter, called $\widehat{\beta}_{\mathrm{reg}}^{\mathrm{IA}}$, is determined by minimizing the Leave One Out error on the initialization sample (see previous section):

$$\widehat{\beta}_{\mathrm{reg}}^{\mathrm{IA}} = \underset{\beta}{\operatorname{argmin}}\, \frac{1}{n_0}\sum_{i=1}^{n_0} \ell_{\widehat{\theta}^{\mathrm{IS}}_{n_0,-i}(\beta)}(X_i,S_i)\,, \tag{3.22}$$

where

$$\widehat{\theta}^{\mathrm{IS}}_{n_0,-i}(\beta) = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \frac{1}{n_0-1}\sum_{j=1,\ i\neq j}^{n_0} \ell_\theta(X_j,S_j) + \frac{\Omega(\theta;\beta_{\mathrm{reg}})}{n_0-1}.$$

Regarding the choice of the defensive parameter value ε, it is cumbersome and there is no direct methodology for its estimation. Moreover, its value depends strongly of the problem studied

as shown in Bect, Sueur, et al., 2015. Nevertheless, in section 3.6.2, we propose a benchmark in order to evaluate the "optimal" value of ε for the class of structures considered in this study.

3.4 Theoretical results

This section summarizes the main theoretical results of this work. Section 3.4.1 addresses the issue of the consistency and asymptotic normality for the IS-AL estimator. Then, in section 3.4.2, a convergence criterion is proposed in order to be able to use the asymptotic confidence ellipsoids defined in section 3.4.3. A discussion is finally proposed about the practical use of the convergence criterion in section 3.4.4.

3.4.1 Consistency and asymptotic convergence of the IS-AL estimator

We derive some theoretical properties for the estimator $\widehat{\theta}_n^{\mathrm{IA}}$, consisting in its consistency towards the parameter θ_* defined by (3.4) and its asymptotic normality by adapting several proofs of Delyon and Portier, 2018 about asymptotic optimality of adaptive importance sampling. Detailed proofs of the following results are given in the Appendix. The proofs are given in a more general context of empirical risk minimization, instead of IS-AL specifically. Indeed, we consider that these theoretical results can be used in a broader manner for other kinds of applications.

We first prove in Appendix B.1 the consistency of the IS-AL estimator $\widehat{\theta}_n^{\mathrm{IA}}$ using Algorithm 1:

$$\widehat{\theta}_n^{\mathrm{IA}} \xrightarrow[n\to+\infty]{} \theta_* \text{ in probability .} \tag{3.23}$$

Then, we prove in Apppendix B.2 the convergence of $\sqrt{n}(\widehat{\theta}_n^{\mathrm{IA}} - \theta_*)$ to a Gaussian random variable with mean zero and covariance matrix:

$$G_{\theta_*,\varepsilon} = \ddot{r}(\theta_*)^{-1} V(q_{\theta_*,\varepsilon})(\ddot{r}(\theta_*)^{-1})^T \,, \tag{3.24}$$

where

$$V(q_{\theta_*,\varepsilon}) = \mathbb{E}\left[\frac{p(X)}{q_{\theta_*,\varepsilon}(X)} \ell_{\theta_*}(X,S) \nabla f_{\theta_*}(X) \nabla f_{\theta_*}(X)^T\right] \,, \tag{3.25}$$

and $\ddot{r}(\theta_*)$ is the Hessian of $r(\theta)$ at θ_*.

A straightforward corollary of equation (3.24) is that, if $G_{\theta_*,\varepsilon}$ is nonsingular (which we assume from now on), then for any $\xi \in (0,1)$:

$$\mathbb{P}\left(n(\widehat{\theta}_n^{\mathrm{IA}} - \theta_*)^T G_{\theta_*,\varepsilon}^{-1} (\widehat{\theta}_n^{\mathrm{IA}} - \theta_*) < q^{\xi}_{\chi^2(m)}\right) \xrightarrow[n\to+\infty]{} \xi \,, \tag{3.26}$$

with $q^{\xi}_{\chi^2(m)}$ the ξ-quantile of the $\chi^2(m)$ distribution (remember that $\theta = (\alpha,\beta)^T$ and $m = 2$ for the lognormal model). Remark that the matrix $G_{\theta_*,\varepsilon}$ depends on the unknown parameter θ_*. It is thus possible to use a plug-in estimator:

$$\widehat{G}_n = \widehat{\ddot{r}}_n(\widehat{\theta}_n^{\mathrm{IA}})^{-1} \widehat{V}_n(\widehat{\theta}_n^{\mathrm{IA}}) (\widehat{\ddot{r}}_n(\widehat{\theta}_n^{\mathrm{IA}})^{-1})^T \,, \tag{3.27}$$

with

$$\widehat{\ddot{r}}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q_{\widehat{\theta}_{i-1}^{\mathrm{IA}},\varepsilon}(X_i)} \ddot{\ell}_\theta(X_i,S_i) \,, \tag{3.28}$$

$$\widehat{V}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)^2}{q_{\theta,\varepsilon}(X_i) q_{\widehat{\theta}^{\mathrm{IA}}_{i-1},\varepsilon}(X_i)} \dot{\ell}_\theta(X_i, S_i)\dot{\ell}_\theta(X_i, S_i)^T \,, \tag{3.29}$$

and $\ddot{\ell}_\theta(x, s)$ the Hessian of $\ell_\theta(x, s)$ with respect to θ. We have:

$$\widehat{G}_n^{-1} \to G_{\theta_*,\varepsilon}^{-1} \text{ in probability.} \tag{3.30}$$

The proof is in Appendix B.3. Using asymptotic normality of $\widehat{\theta}^{\mathrm{IA}}_n$, we can show that: $n(\widehat{\theta}^{\mathrm{IA}}_n - \theta_*)^T G^{-1}_{\theta_*,\varepsilon}(\widehat{\theta}^{\mathrm{IA}}_n - \theta_*) \to \chi^2(m)$. Using Slutsky's lemma, we have the following convergence in distribution:

$$n(\widehat{\theta}^{\mathrm{IA}}_n - \theta_*)^T \widehat{G}_n^{-1}(\widehat{\theta}^{\mathrm{IA}}_n - \theta_*) \xrightarrow[n\to+\infty]{} \chi^2(m) \,. \tag{3.31}$$

3.4.2 Convergence criterion using a statistical hypothesis test

The estimation of the generalization error without a validation set is often based on Cross Validation. When IS-AL is used, the data points (X_i, S_i) are no longer independent and identically distributed. We propose to use a convergence criterion that ensures that asymptotic normality is reached. Consider two independent datasets $\mathcal{D}_1 = (X_{i,1}, S_{i,1})_{i=1}^n$ and $\mathcal{D}_2 = (X_{i,2}, S_{i,2})_{i=1}^n$ generated with IS-AL. Let $\widehat{R}^{\mathrm{IA}}_{n,\mathrm{reg},j}$ be the weighted loss for $\mathcal{D}_j$ for $j = 1, 2$ defined as in (3.20). Denote:

$$\widehat{\theta}^{\mathrm{IA}}_{n,j} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \widehat{R}^{\mathrm{IA}}_{n,\mathrm{reg},j}(\theta; \beta_{\mathrm{reg}}), \qquad j = 1, 2 \,.$$

We introduce the following quantity:

$$\Delta\dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg}} = \dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg},1}(\widehat{\theta}^{\mathrm{IA}}_{n,2}; \beta_{\mathrm{reg}}) - \dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg},2}(\widehat{\theta}^{\mathrm{IA}}_{n,1}; \beta_{\mathrm{reg}}),$$

then we have:

$$\sqrt{n}\Delta\dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 8V(q_{\theta_*,\varepsilon}, \dot{\ell}_{\theta_*})) \tag{3.32}$$

as $n \to +\infty$. Denote

$$\widehat{W}_n = \frac{n}{8}(\Delta\dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg}})^T \widehat{V}^{-1}_{n,12}\Delta\dot{\widehat{R}}^{\mathrm{IA}}_{n,\mathrm{reg}}, \tag{3.33}$$

$$\widehat{V}_{n,12} = \frac{1}{2}\big(\widehat{V}_{n,1}(\widehat{\theta}^{\mathrm{IA}}_{n,1}) + \widehat{V}_{n,2}(\widehat{\theta}^{\mathrm{IA}}_{n,2})\big), \tag{3.34}$$

with $\widehat{V}_{n,j}$ the empirical estimator in equation (3.29) for the j-th IS-AL dataset $\mathcal{D}_j$ for $j = 1, 2$.

By equation (3.32) and by Slutsky's lemma, $\widehat{W}_n$ converges weakly to $\chi^2(m)$. It is, therefore, possible to define a convergence criterion inspired by statistical test theory to check the asymptotic normality of $\widehat{\theta}^{\mathrm{IA}}_n$. Our convergence criterion is equivalent to the hypothesis test:

$$(\mathcal{H}_0) \,: \widehat{W}_n \text{ follows } \chi^2(m) \quad \text{against} \quad (\mathcal{H}_1) \,: \widehat{W}_n \text{ does not follow } \chi^2(m) \,. \tag{3.35}$$

For $\xi \in (0, 1)$, we then consider the statistical test which rejects $(\mathcal{H}_0)$ if:

$$\widehat{W}_n > q_{1-\xi}^{\chi^2(m)} \,, \tag{3.36}$$

where $q_{1-\xi}^{\chi^2(m)}$ denotes the $(1-\xi)$-quantile of the $\chi^2(m)$ distribution. Hence, this statistical test is of asymptotic level ξ.

3.4.3 Asymptotic confidence ellipsoid

Thanks to the equation (3.31), it is possible to construct an asymptotic confidence ellipsoid of level $\xi \in (0,1)$ defined by:

$$\mathcal{E}^{\mathrm{IA}}_{n,\xi} = \{\theta : n(\theta - \widehat{\theta}^{\mathrm{IA}}_n)^T \widehat{G}_n^{-1} (\theta - \widehat{\theta}^{\mathrm{IA}}_n) < q^{\chi^2(m)}_{1-\xi}\}\,, \tag{3.37}$$

with:

$$\mathbb{P}(\theta_* \in \mathcal{E}^{\mathrm{IA}}_{n,\xi}) \xrightarrow[n\to+\infty]{} \xi.$$

Because the convergence criterion $\widehat{W}_n$ indicates when the estimator follows the asymptotic Gaussian distribution, it also indicates at which sample size n the value $\mathbb{P}(\theta_* \in \mathcal{E}^{\mathrm{IA}}_{n,\xi})$ is close to its theoretical value ξ.

3.4.4 Discussion about the practical use of the convergence criterion

An apparent drawback of this convergence criterion is that it doubles the computational cost, due to the necessity of having two independent IS-AL estimators $\widehat{\theta}^{\mathrm{IA}}_{n,1}$ and $\widehat{\theta}^{\mathrm{IA}}_{n,2}$ to compute $\widehat{W}_n$. It is, however, possible to use the estimator

$$\widehat{\theta}^{\mathrm{IA}}_{n,12} = \frac{\widehat{\theta}^{\mathrm{IA}}_{n,1} + \widehat{\theta}^{\mathrm{IA}}_{n,2}}{2}, \tag{3.38}$$

which has an asymptotic variance that is half the one of $\widehat{\theta}^{\mathrm{IA}}_{n,1}$ and $\widehat{\theta}^{\mathrm{IA}}_{n,2}$. Indeed, it is straightforward that $\sqrt{n}(\widehat{\theta}^{\mathrm{IA}}_{n,12} - \theta_*)$ converges in distribution to a zero mean Gaussian random variable with covariance matrix $G_{\theta_*,\varepsilon}/2$. It is, therefore, possible to define an asymptotic confidence ellipsoid which exploits all the data points used to build the estimator $\widehat{\theta}^{\mathrm{IA}}_{n,12}$ of θ_*:

$$\mathcal{E}^{\mathrm{IA}}_{n,12,\xi} = \{\theta : 2n(\theta - \widehat{\theta}^{\mathrm{IA}}_{n,12})^T \widehat{G}_{n,12}^{-1} (\theta - \widehat{\theta}^{\mathrm{IA}}_{n,12}) < q^{\chi^2(m)}_{1-\xi}\},$$

with $\widehat{G}_{n,12} = \widehat{\ddot{r}}^{-1}_{n,12} \widehat{V}_{n,12} (\widehat{\ddot{r}}^{-1}_{n,12})^T$, $\widehat{V}_{n,12}$ defined by (3.34), $\widehat{\ddot{r}}_{n,12} = \frac{1}{2}\widehat{\ddot{r}}_{n,1}(\widehat{\theta}^{\mathrm{IA}}_{n,1}) + \frac{1}{2}\widehat{\ddot{r}}_{n,2}(\widehat{\theta}^{\mathrm{IA}}_{n,2})$ and $\widehat{\ddot{r}}_{n,j}$ defined as (3.28) with the dataset $\mathcal{D}_j$, $j = 1, 2$.

3.5 Performance evaluation of the IS-AL strategy compared to the random sampling and MLE strategies

This section explains how to assess the performance of IS-AL with respect to Random Sampling (RS) and MLE strategies. In section 3.5.1, RS and MLE principles are briefly summarized. Performance metrics inspired from Chabridon et al., 2017; Morio and Balesdent, 2015 to check the quality of IS-AL strategy are detailed in section 3.5.2. Finally, the statistical procedure used to assess the quality of the IS-AL asymptotic confidence ellipsoid compared to that of a classical approach such as MLE is given in section 3.5.3.

3.5.1 RS and MLE principles

RS strategy consists in applying the IS-AL algorithm with the proposal probability density q being the marginal probability density p of the intensity measure. This boils down to classical

empirical risk minimization for supervised learning. The RS estimator $\widehat{\theta}_n^{\mathrm{RS}}$ is then defined by:

$$\widehat{\theta}_n^{\mathrm{RS}} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \widehat{R}_{n,\mathrm{reg}}^{\mathrm{RS}}(\theta;\beta_{\mathrm{reg}}), \tag{3.39}$$

$$\widehat{R}_{n,\mathrm{reg}}^{\mathrm{RS}}(\theta;\beta_{\mathrm{reg}}) = \frac{1}{n}\sum_{i=1}^{n}\ell_\theta(X_i,S_i) + \frac{\beta_{\mathrm{reg}}}{n\beta}. \tag{3.40}$$

As mentioned in the introduction, MLE is a classical estimation method in the field of seismic probabilistic risk assessment and fragility curve estimation (see e.g. Straub and Der Kiureghian, 2008; Lallemant, Kiremidjian, and Burton, 2015; Shinozuka et al., 2000; F. Wang and Feau, 2020; Mai, Konakli, and Sudret, 2017; Baker, 2015; Zentner, 2010). It is defined by the estimator $\widehat{\theta}_n^{\mathrm{MLE}}$ that maximizes the likelihood given a dataset $(X_i,S_i)_{i=1}^n$ that is sampled at random from the original marginal density p of X:

$$\widehat{\theta}_n^{\mathrm{MLE}} = \underset{\theta\in\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} S_i\log(f_\theta(X_i)) + (1-S_i)\log(1-f_\theta(X_i)). \tag{3.41}$$

The initializations of the RS and MLE algorithms are based on 20 data points drawn at random from the original distribution p. For the RS algorithm, the regularization parameter, called $\widehat{\beta}_{\mathrm{reg}}^{\mathrm{RS}}$, is computed using Leave One Out cross validation as for the IS-AL algorithm.

3.5.2 Performance metrics for the numerical benchmarks

This section aims to provide performance metrics, inspired from Chabridon et al., 2017; Morio and Balesdent, 2015, to assess IS-AL performances, in comparison with the RS and MLE strategies, on test cases.

3.5.2.1 Performance metrics based on the training errors

For the IS-AL strategy, the training error is called $\widehat{R}_n^{\mathrm{IA}} = \widehat{R}_{n,\mathrm{reg}}^{\mathrm{IA}}(\widehat{\theta}_n^{\mathrm{IA}};\beta_{\mathrm{reg}}^{\mathrm{IA}})$ and is defined by equation (3.21).

For the RS and MLE strategies, the training errors are respectively called $\widehat{R}_n^{\mathrm{RS}} = \widehat{R}_{n,\mathrm{reg}}^{\mathrm{RS}}(\widehat{\theta}_n^{\mathrm{RS}};\beta_{\mathrm{reg}}^{\mathrm{RS}})$ and $\widehat{R}_n^{\mathrm{MLE}} = \widehat{R}_{n,\mathrm{reg}}^{\mathrm{MLE}}(\widehat{\theta}_n^{\mathrm{MLE}};\beta_{\mathrm{reg}}^{\mathrm{RS}})$, and are defined by :

$$\widehat{R}_{n,\mathrm{reg}}^{\bullet}(\theta;\beta_{\mathrm{reg}}^{\mathrm{RS}}) = \frac{1}{n}\sum_{i=1}^{n}\ell_\theta(X_i,S_i) + \frac{\beta_{\mathrm{reg}}^{\mathrm{RS}}}{n\beta}$$

where $\bullet$ is for RS or MLE. Note that for MLE the penalization $\widehat{\beta}_{\mathrm{reg}}^{\mathrm{RS}}$ is only used to define similar training errors as for IS-AL and RS algorithms, in order to compare the same quantity.

Thus, the performance metrics are :

- the *Relative Standard Deviation*

$$\mathrm{RSD}_n^{\bullet} = \frac{\sqrt{\mathbb{V}[\widehat{R}_n^{\bullet}]}}{\mathbb{E}[\widehat{R}_n^{\bullet}]}, \tag{3.42}$$

  where $\bullet$ is for IA, RS and MLE.

- the *Relative Bias*

$$\mathrm{RB}_n^\bullet = \frac{|\mathrm{b} - \mathbb{E}[\widehat{R}_n^\bullet]|}{\mathrm{b}}, \tag{3.43}$$

where $\bullet$ is for IA, RS and MLE, and $\mathrm{b} = \mathbb{E}[\mu(X)(1-\mu(X))]$.

- The *efficiency*

$$\nu_n^\bullet = \frac{\mathbb{V}[\widehat{R}_n^\bullet]}{\mathbb{V}[\widehat{R}_n^{\mathrm{IA}}]}, \tag{3.44}$$

where $\bullet$ is for RS and MLE. A value of $\nu_n^\bullet > 1$ shows that IS-AL has a smaller loss variance than RS or MLE.

The above metrics are empirically calculated using R replications of the three procedures (IS-AL, RS and MLE) and b is estimated using a massive Monte Carlo estimator of μ. In practice, this is only possible when the numerical models are not numerically expensive. This is the case for the first two numerical test cases addressed in Section 3.6, but not for the third industrial case.

3.5.2.2 Performance metrics based on the testing errors

Additionally, we define the testing error by:

$$\widehat{Q}_{n,\mathrm{reg}}^\bullet(\theta;\beta_{\mathrm{reg}}) = \frac{1}{n_t}\sum_{i=1}^{n_t} \ell_\theta(X_i^{(t)}, S_i^{(t)}) + \frac{\beta_{\mathrm{reg}}}{n_t \beta},$$

where $(X_i^{(t)}, S_i^{(t)})_{i=1}^{n_t}$ is a testing set (independently and identically distributed with the original distribution P). The associated performance metrics are calculated by replacing $\widehat{R}_{n,\mathrm{reg}}^\bullet(\theta;\beta_{\mathrm{reg}})$ with $\widehat{Q}_{n,\mathrm{reg}}^\bullet(\theta;\beta_{\mathrm{reg}})$ (also called $\widehat{Q}_n^\bullet$ in the following) as defined in section 3.5.2.1.

3.5.3 Benchmark on the confidence ellipsoids IS-AL and MLE

This section aims to propose a procedure to evaluate the quality of the IS-AL asymptotic confidence ellipsoid, $\mathcal{E}_{n,\xi}^{\mathrm{IA}}$, compared to that of the MLE, $\mathcal{E}_{n,\xi}^{\mathrm{MLE}}$. This procedure is based on the use of R replications of the IS-AL and MLE algorithms, as for the evaluation of the performance metrics. So, we first define the empirical estimator of $\mathbb{P}\left(\theta_* \in \mathcal{E}_{n,\xi}^{\mathrm{IA}}\right)$ (resp. $\mathbb{P}\left(\theta_* \in \mathcal{E}_{n,\xi}^{\mathrm{MLE}}\right)$), namely the Coverage Probability (CP), in order to numerically (i) verify the definitions of the ellipsoids and (ii) evaluate their convergences with respect to the size n of the samples. Then, to quantify the effectiveness of the IS-AL strategy on reducing the variance of the estimate of the fragility curve, compared to that of the MLE, we define and compare their Confidence Ellipsoid Volumes (CEVs).

Section 3.5.3.1 gives the definitions of the confidence ellipsoid and the coverage probability for MLE. As the IS-AL confidence ellipsoid is defined in section 3.4.3, section 3.5.3.2 deals only with the associated CP. Finally section 3.5.3.3 defines the CEVs for both procedures.

3.5.3.1 Confidence ellipsoid and coverage probability for MLE

In order to define the asymptotic confidence ellipsoid for the MLE and to compute the associated CP, we use (i) the asymptotic normality of the MLE estimator (Bachoc, 2013) and (ii) the independence property of the samples that allows the use of the bootstrap method.

We first consider R replications of MLE estimator $\widehat{\theta}_n^{\mathrm{MLE}}$ for different sample size n and build B bootstrap samples of size n for each replication in order to compute a bootstrap covariance:

$$\widehat{V}_{n,r}^{\mathrm{MLE}} = \frac{1}{B}\sum_{b=1}^{B} n(\theta_{b,r}^{\mathrm{MLE},*} - \widehat{\theta}_{n,r}^{\mathrm{MLE}})(\theta_{b,r}^{\mathrm{MLE},*} - \widehat{\theta}_{n,r}^{\mathrm{MLE}})^T \,, \tag{3.45}$$

where $\widehat{\theta}_{n,r}^{\mathrm{MLE}}$ is the MLE estimator for the r-th replication of size n and $\theta_{b,r}^{\mathrm{MLE},*}$ the bootstrap MLE estimator for the b-th bootstrap sample of the r-th replication.

Thus, the bootstrapped confidence ellipsoid for MLE is defined by:

$$\mathcal{E}_{n,\xi}^{\mathrm{MLE},r} = \{\theta : n(\theta - \widehat{\theta}_{n,r}^{\mathrm{MLE}})^T(\widehat{V}_{n,r}^{\mathrm{MLE}})^{-1}(\theta - \widehat{\theta}_{n,r}^{\mathrm{MLE}}) \leq q_{1-\xi}^{\chi^2(m)}\} \,, \tag{3.46}$$

while the bootstrap CP writes:

$$\mathrm{CP}_n^{\mathrm{MLE},r} = \frac{1}{R}\sum_{r=1}^{R} \mathbb{1}_{\theta_* \in \mathcal{E}_{n,\xi}^{\mathrm{MLE},r}} \,. \tag{3.47}$$

3.5.3.2 Coverage probability for IS-AL

The IS-AL asymptotic confidence ellipsoid $\mathcal{E}_{n,\xi}^{\mathrm{IA}}$ is defined in section 3.4.3 by equation (3.37). So, as for MLE, the associated CP is computed by considering R replications of IS-AL, namely:

$$\mathrm{CP}_n^{\mathrm{IA},r} = \frac{1}{R}\sum_{r=1}^{R} \mathbb{1}_{\theta_* \in \mathcal{E}_{n,\xi}^{\mathrm{IA},r}} \,,$$

where $\mathcal{E}_{n,\xi}^{\mathrm{IA},r}$ is the asymptotic confidence ellipsoid of the r-th replication of the IS-AL procedure of size n.

3.5.3.3 Confidence ellipsoid volumes for IS-AL and MLE

A qualitative criterion to measure the sharpness of a confidence ellipsoid is its volume (Golestaneh et al., 2018). So, to evaluate the effectiveness of the IS-AL strategy on the reduction of the variance of the fragility curve estimations, we define the CEVs, for respectively the MLE and IS-AL strategies, as follows:

$$\mathrm{CEV}_n^{\mathrm{MLE},r} = \det\left(\frac{\widehat{V}_{n,r}^{\mathrm{MLE}}}{n}\right) \text{ and } \mathrm{CEV}_n^{\mathrm{IA},r} = \det\left(\frac{\widehat{G}_{n,r}}{n}\right), \tag{3.48}$$

where $\widehat{G}_{n,r}$ is the estimated covariance matrix (3.27) of the r-th replication of IS-AL procedure of size n.

3.6 Numerical results

To evaluate IS-AL efficiency, a numerical benchmark has been performed with three test cases with increasing complexity:
1) a synthetic test case with known fragility curve and probability distribution of the seismic log-intensity measure X,

2) a nonlinear elasto-plastic oscillator with kinematic hardening subjected to synthetic signals generated from the modulated and filtered white-noise ground-motion model (Rezaeian and Der Kiureghian, 2010), as in Sainct et al., 2020,
3) an industrial test case of a nuclear facility's pipeline-system, submitted to the same artificial signals as the two test cases.

For test cases 2 and 3, 97 acceleration records selected from the European Strong Motion Database (Ambraseys et al., 2004) in the domain $5.5 < M < 6.5$ and $R < 20\text{km}$ - where M is the magnitude and R the distance from the epicenter - are considered in order to identify the parameters of the ground-motion model. 10^5 realizations of synthetic signals are then generated to form the unlabeled pool.

The oscillator test case aims to evaluate the effectiveness of the IS-AL strategy before its application to an industrial test case which is numerically much more costly. Moreover, since it well represents the essential features of the nonlinear responses of a large variety of real structures subjected to earthquakes, this test case allows to determine the value of the hyperparameter ε - thanks to a numerical benchmark - because there is no ad hoc procedure to do this.

3.6.1 Synthetic test case

Here we benchmark our methodology while having full knowledge of the true fragility curve. We generate 30,000 datapoints (X_i, S_i) with the fragility curve $\mu(x) = \Phi(\frac{x-\log(\alpha_*)}{\beta_*})$ with $(\alpha_*, \beta_*) = (0.3, 0.4)$. The original marginal distribution of X is here a Gaussian distribution with mean $\log\left(\frac{\alpha_*}{5}\right)$ and variance 1.69. The parameters have been chosen so that the data generated are qualitatively close to the nonlinear oscillator test case presented in section 3.6.2. The unlabeled pool consists of 20,000 datapoints X_i. 10,000 datapoints (X_i, S_i) will be our validation set for testing error estimation, using crude Monte Carlo.

Figure 3.1 shows (i) the target fragility curve μ in dashed red line, (ii) a kernel density estimation of the density p based on the whole dataset in green and (iii) a kernel density density estimation q of the 120 datapoints X_i obtained by IS-AL in red.

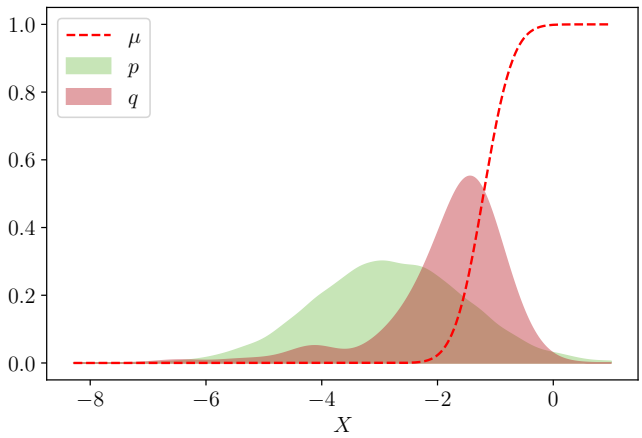
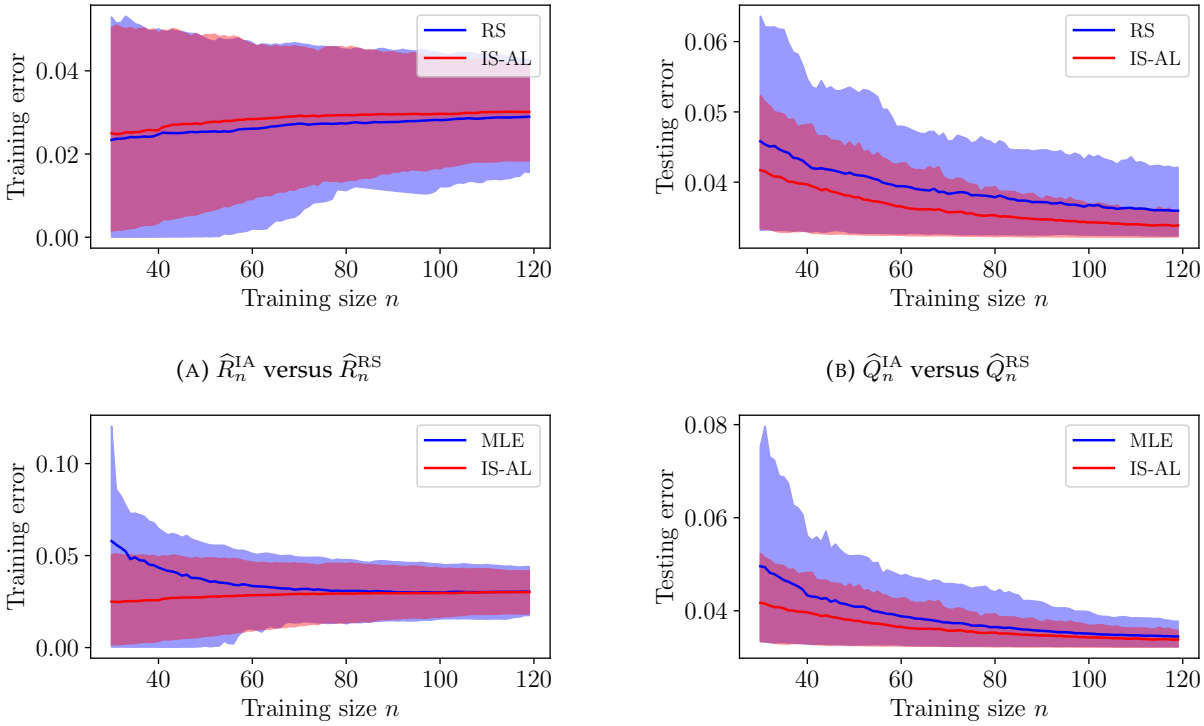


FIGURE 3.1: Synthetic test case with lognormal fragility curve with parameters $(\alpha_*, \beta_*) = (0.3, 0.4)$ and $X \sim \mathcal{N}(\frac{\alpha_*}{5}, 1.69)$. Comparison of the original marginal density p of X with the empirical density q of the $n = 120$ datapoints X_i obtained by IS-AL.

Figure 3.2 shows the training and testing errors for $R = 500$ replications of the IS-AL, RS and MLE algorithms. The algorithms are initialized with 20 datapoints and $n = 100$ datapoints are

extracted from the unlabeled pool with the three procedures. The regularization parameters $\beta_{\text{reg}} \in (10^{-4}, 10^{-1})$ were determined by cross validation with the 20 datapoints used for initialization for each replication of the IS-AL and MLE strategies. We also use a defensive parameter value ε of 10^{-3} (see section 3.6.2 for justification).



(A) $\widehat{R}_n^{\text{IA}}$ versus $\widehat{R}_n^{\text{RS}}$

(B) $\widehat{Q}_n^{\text{IA}}$ versus $\widehat{Q}_n^{\text{RS}}$

(C) $\widehat{R}_n^{\text{IA}}$ versus $\widehat{R}_n^{\text{MLE}}$

(D) $\widehat{Q}_n^{\text{IA}}$ versus $\widehat{Q}_n^{\text{MLE}}$

FIGURE 3.2: Results of the numerical benchmark for the synthetic test case: the thick lines represent the mean loss for $R = 500$ replications, the shaded areas represent the ranges between the quantiles at 90% and 10% of the 500 replications of the IS-AL, RS and MLE procedures. For this case, the bias is known and is equal to $\mathbb{E}[\mu(X)(1 - \mu(X))] \simeq 0.032$.

TABLE 3.1: Performance metrics for the synthetic test case for $n = 120$ (see Section 3.5.2.1)

	Train			Test		
$\bullet$	RS	MLE	IS-AL	RS	MLE	IS-AL
$\text{RSD}_{120}^{\bullet}$ (%)	38	36	34	12	9	8.5
$\nu_{120}^{\bullet}$	1.2	1.2	×	2.3	1.1	×
$\text{RB}_{120}^{\bullet}$ (%)	8.6	4.2	5	13	9	6.7

As depicted by Figure 3.2 and Table 3.1, IS-AL does not seem to reduce the training error. This result is normal because IS-AL selects seisms whose intensity measures maximize $\tilde{\ell}_\theta$, which can be seen as a marginalized training loss variance of the observations. In other words, as illustrated in Figure 3.1 with the density $q(x)$, IS-AL selects "difficult" points - typically values of x for which $\mu(x)$ takes values between 0 and 1 - and therefore the training error can be large because it is not

representative of the generalization error as the testing one. RS, MLE and IS-AL strategies really distinguish themselves on the testing error, which is smaller for IS-AL. Moreover, IS-AL quickly converges to the known bias equal to $\mathbb{E}[\mu(X)(1-\mu(X))] \simeq 0.032$. In comparison with RS and MLE strategies, the variance of IS-AL is smaller after 120 iterations: ν_{120}^{MLE} is smaller than ν_{120}^{RS}, meaning that MLE is competitive with IS-AL in this synthetic case.

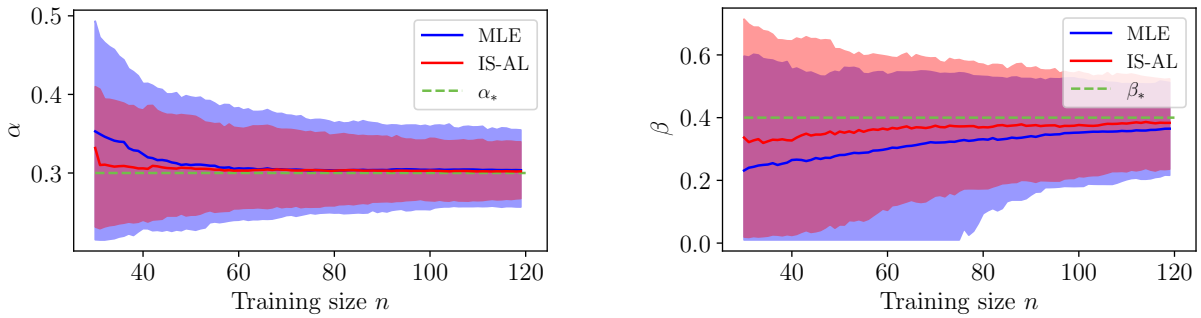


FIGURE 3.3: Results of the numerical benchmark for the synthetic test case: empirical distributions of the parameters α and β are represented by ranges between the empirical 10% and 90% quantiles of 500 replications. The shaded blue and red areas correspond respectively to MLE and IS-AL. The dashed green lines correspond to the true parameters α_* and β_*.

Figure 3.3 compares the distributions of the parameters α and β for several sample sizes using the 500 replications of MLE and IS-AL estimators and they are similar when $n > 100$. Indeed, the statistical model is in this case well specified (i.e. failure events follow a Bernoulli distribution with a lognormal probability of failure) and thus MLE is supposed to perform well as shown in Bachoc, 2013. Note that up to $n = 80$, the MLE strategy can produce degenerate fragility curves because $\beta \simeq 0$. The IS-AL algorithm avoids this pitfall due to the regularization parameter.

3.6.2 A nonlinear oscillator

This test case aims to validate the overall strategy developed in this work on a simple but representative case, because this is not possible for complex structures like the one in section 3.6.3. This section is therefore particularly comprehensive, from the initialization of the IS-AL algorithm to the estimations of the fragility curves, via the choice ε and the numerical verification of the theorems.

3.6.2.1 Presentation of the oscillator

This second test case - illustrated in Figure 3.4 - relates to a single degree of freedom elasto-plastic oscillator which exhibits kinematic hardening. It has been used in previous studies such as Trevlopoulos, Feau, and Zentner, 2019; Sainct et al., 2020. For a unit mass m, its equation of motion is:

$$\ddot{z}(t) + 2\zeta\omega_L\dot{z}(t) + f_{NL}(z(t)) = -s(t)\ ,$$

with $s(t)$ an artificial seismic signal. $\dot{z}(t)$ and $\ddot{z}(t)$ are respectively the velocity and the acceleration of the mass while ζ is the damping ratio and ω_L the pulsation of the oscillator. The nonlinear force f_{NL} is governed by two parameters: the post-yield stiffness, a, and the yield displacement, Y.

With this model, the quantity of interest is the maximum displacement of the mass, $D = \max_{t\in[0,T]} |z(t)|$, where T is the duration of the seismic excitation. The failure state is then defined by the $\{0,1\}$-valued variable $S = \mathbb{1}_{(D>C)}$, where $C = 2Y$ is chosen to be approximately the 90% quantile of the maximal linear displacement of the unlabeled pool of size 10^5.

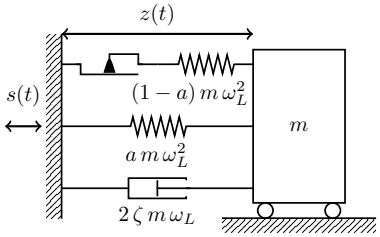


FIGURE 3.4: Elasto-plastic mechanical oscillator with kinematic hardening, with parameters $f_L = 5$ Hz and $\zeta = 2\%$. The yield limit is $Y = 5.10^{-3}$ m and the post-yield stiffness is 20% of the elastic stiffness, hence $a = 0.2$.

In order to check the performances of the IS-AL algorithm, the unlabeled training set consists in 9.10^4 seismic signals and the testing set is composed of 10^4 signals. The benchmark study consists in $R = 500$ replications with $n = 120$ sampled seismic signals using IS-AL (that includes the initial 20 points) and 120 for the RS and MLE strategies.

3.6.2.2 Initialization of the IS-AL procedure

In this test case, for IS-AL initialization, we use the underlying elastic oscillator as a cheap model. The initialization parameter $\widehat{\theta}_0^{\mathrm{IA}}$ is approximated by $\widehat{\theta}_{10^5}^{\mathrm{RS}}$ (equation (3.8)) using the 10^5-sized dataset. In addition, the PGA is first considered as IM. Even if the PGA is not known to be the best indicator, doing so helps to verify the relevance of the methodology in a "less favorable" case. Note that the influence of the IM on the results is discussed in section 3.6.2.8. As shown in Figure 3.5, the parameter $\widehat{\theta}_0^{\mathrm{IA}}$ could be considered "close to" the true parameter θ_*. Thus, 20 datapoints are queried on the nonlinear oscillator with the instrumental density $q_{\widehat{\theta}_0^{\mathrm{IA}},\varepsilon}$ (equation (3.18)) before launching the adaptive strategy.

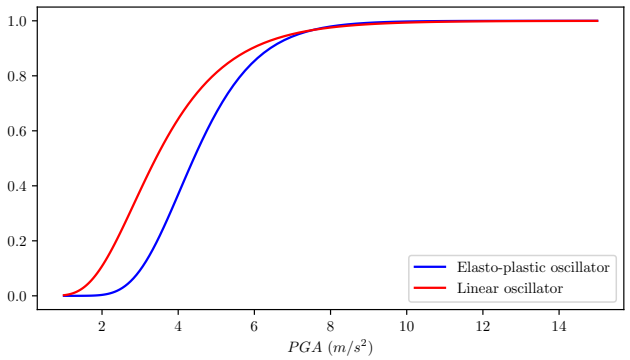


FIGURE 3.5: Lognormal fragility curves of the linear elastic and the nonlinear oscillators obtained by using least squares minimization on the total 10^5 synthetic seismic signals of the dataset.

3.6.2.3 Choice of ε

As mentioned in section 3.3.2.3, there is no direct methodology for the choice of the ε value. One thus benefits from this simple test case to implement a numerical benchmark in order to obtain a reasonable value of ε for the class of structures for which the oscillator represents the global nonlinear behavior under seismic excitation. This benchmark consists in evaluating the IS-AL efficiency with respect to the RS strategy, ν_n^{RS} (equation (3.44)), as a function of ε when IM = PGA and $n = 120$. Results are given in Table 3.2.

TABLE 3.2: Defensive parameter ε influence on ν_{120}^{RS} when IM = PGA.

ε	Train	Test
10^{-1}	1.3	1.2
10^{-2}	2.1	3.9
10^{-3}	2.2	3.3

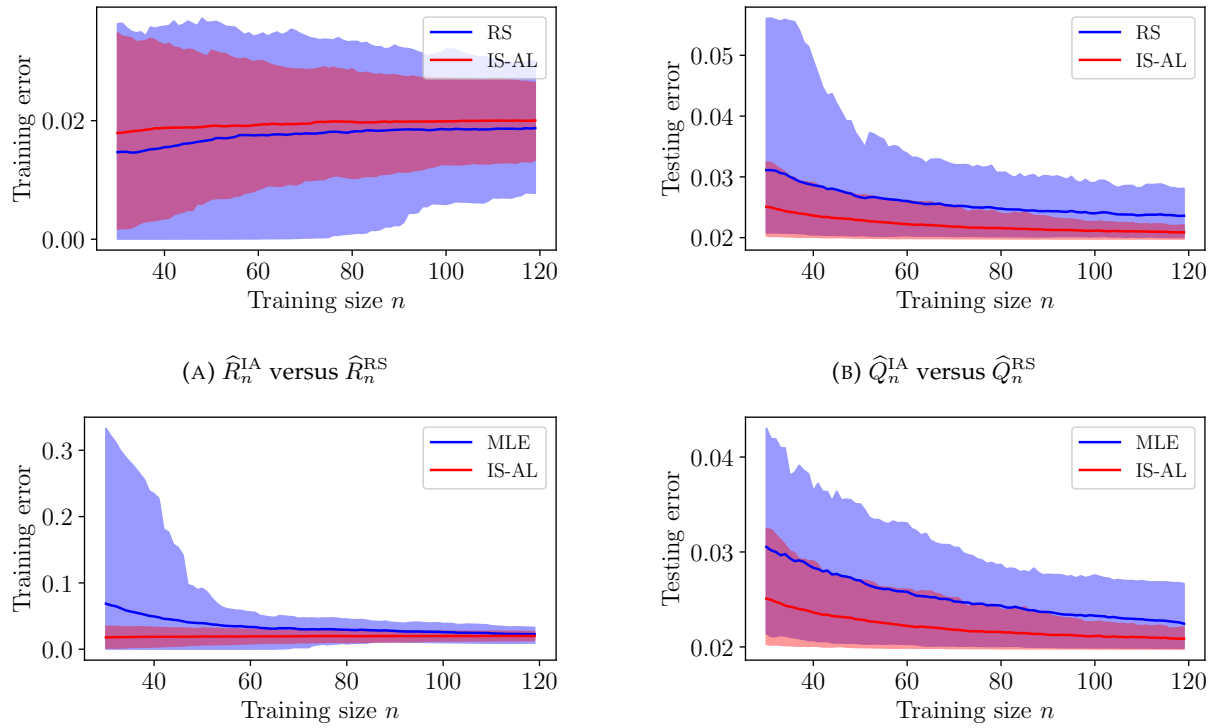
They show that ν_{120}^{RS} does not change between $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-3}$. ν_{120}^{RS} is smaller when $\varepsilon = 10^{-1}$ meaning that this value is too conservative because there are too many elements drawn from the pdf p. Accordingly, all the results will presented hereafter with a defensive parameter $\varepsilon = 10^{-3}$. This implies that the defensive strategy plays essentially no role here, but gives theoretical convergence guarantees.

3.6.2.4 Performance metrics

Figure 3.6 compares the IS-AL, MLE and RS training and testing errors as functions of n. The mean training loss for the 500 replications is higher for IS-AL than for RS. Indeed, the instrumental density is chosen to sample seismic signals that maximize the loss variance, resulting in a high training error. Moreover, the mean testing error of IS-AL is also significantly smaller than for RS and quickly converges to the "minimal" error related to the term $\mathbb{E}[\mu(X)(1-\mu(X))]$ in (3.5). This is shown in Table 3.3 by a significantly smaller value of relative bias $\mathrm{RB}_{120}^{\mathrm{IA}}$ (1%) for the testing error than with RS (12%). With respect to the MLE, one cannot make equivalent remarks insofar as the two errors are "artificial" (see definitions in section 3.5.2.1) and only plotted for illustration purpose. However, Table 3.3 shows that the IS-AL strategy has overall better performance than the other two strategies.

TABLE 3.3: Performance metrics for the elasto-plastic oscillator for $n = 120$ when IM = PGA (see Section 3.5.2.1).

	Train			Test		
$\bullet$	RS	MLE	IS-AL	RS	MLE	IS-AL
$\mathrm{RSD}_{120}^{\bullet}$ (%)	47	65	30	19	14	11
$\nu_{120}^{\bullet}$	2.2	5.9	$\times$	3.3	1.8	$\times$
$\mathrm{RB}_{120}^{\bullet}$ (%)	11.1	7.6	5	12	6.4	1



(A) $\widehat{R}_n^{\mathrm{IA}}$ versus $\widehat{R}_n^{\mathrm{RS}}$

(B) $\widehat{Q}_n^{\mathrm{IA}}$ versus $\widehat{Q}_n^{\mathrm{RS}}$

(C) $\widehat{R}_n^{\mathrm{IA}}$ versus $\widehat{R}_n^{\mathrm{MLE}}$

(D) $\widehat{Q}_n^{\mathrm{IA}}$ versus $\widehat{Q}_n^{\mathrm{MLE}}$

FIGURE 3.6: Results of the numerical benchmark for the elasto-plastic mechanical oscillator with the PGA as IM (same notations as for the synthetic test case). The empirical distributions of the training and testing errors are represented by the range between the empirical 90% and 10% quantiles of the 500 replications.

3.6.2.5 Empirical distributions of the parameters α and β

Figure 3.7 shows the empirical distributions of the parameters α and β for several sample sizes using 500 replications of MLE and IS-AL estimators. Remark in this case that IS-AL performs better than the MLE by reducing the variances of the parameters' estimators. The effects are particularly visible for the parameter β, when the active learning strategy and the regularization play their role in reducing the standard deviations of the estimators without increasing bias. Indeed, MLE performances are downgraded when the model is not well specified (Bachoc, 2013). We remark that parameter estimation is quite unstable for IS-AL for low sample sizes. Indeed, the number of failure events for low sample sizes is often 0, which makes impossible a correct estimation of the fragility curve's parameters.

Figure 3.8 helps to visualize how IS-AL reduces the uncertainty of the fragility curve estimation: IS-AL is designed to sample seismic ground motions in the transition zone between 0 and 1 of the fragility curve, this phenomenon is responsible for the uncertainty reduction.

3.6.2.6 Convergence criterion

Figure 3.9 shows the value of the test statistics $\widehat{W_n}$ (see section 3.4.2) for two independent IS-AL realizations. This result expresses that the IS-AL algorithm achieves asymptotic normality from $n = 100$ because the value of $\widehat{W}_n$ is less than the quantile 90% of the distribution $\chi^2(2)$.

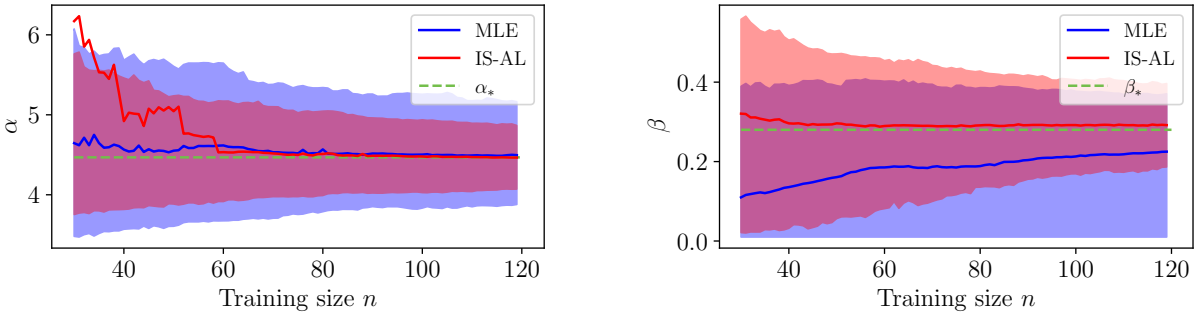


FIGURE 3.7: Results for the elasto-plastic mechanical oscillator with the PGA as IM: the empirical distributions of the parameters α and β are represented by the empirical 90% and 10% quantiles of the 500 replications and correspond to the shaded blue and red areas for MLE and IS-AL, respectively. The dashed green lines correspond to the values α_* and β_*, which have been here approximated by $\widehat{\alpha}_N$, $\widehat{\beta}_N$ for $N = 10^5$.

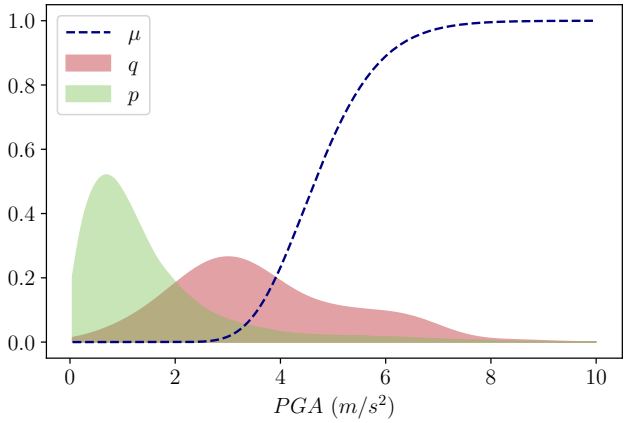


FIGURE 3.8: Comparison of the original marginal density p of PGA with the estimated density q of sampled PGA using IS-AL with $n = 100$ datapoints for the nonlinear oscillator. Fragility curve is approximated by $\mu(\text{PGA}) = \Phi(\frac{\log(\text{PGA}/\widehat{\alpha}_N)}{\widehat{\beta}_N})$ for $N = 10^5$.

3.6.2.7 CPs and CEVs

Figure 3.10 shows the CP values for the nonlinear oscillator for a training size n between 100 and 500 for the fragility curve estimation by MLE or IS-AL. The true parameter θ_* for this case has been approximated by $\widehat{\theta}_N$ for $N = 10^5$. The numerical results show that the CP values are close to the theoretical and nominal value of 90%, which validates the theoretical results of the section 3.4.

Figure 3.11 shows the CEVs for the MLE and IS-AL estimators. For $R = 200$ replications, these results show that for all the values of n considered $\text{CEV}_n^{\text{IA}} < \text{CEV}_n^{\text{MLE}}$. This indicates that MLE and IS-AL succeed in generating confidence ellipsoids that have the required coverage probability but MLE does so by generating ellipsoids that are much larger than the ones generated by IS-AL. We can then conclude that IS-AL is much more efficient.

We emphasize that the convergence criterion $\widehat{W}_n$, illustrated in Figure 3.9, gives us at which sample size the IS-AL reaches asymptotic normality and thus at which sample size asymptotic

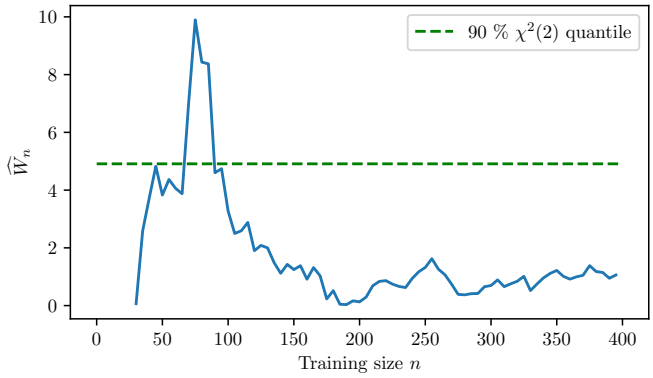


FIGURE 3.9: Values of the test statistic $\widehat{W}_n$ for two independent IS-AL realizations, when IM = PGA.

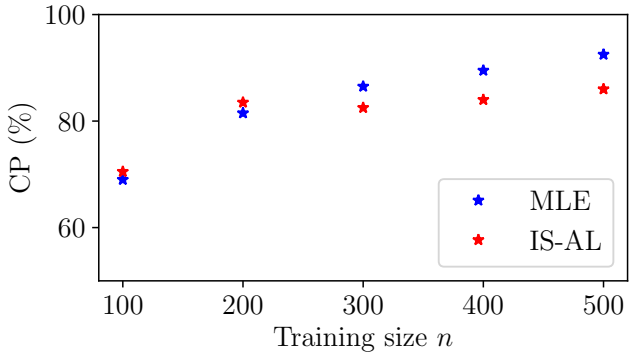


FIGURE 3.10: CP values (for the confidence ellipsoid with level $1-\xi = 0.9$ for θ) as a function of the training size n of IS-AL and MLE when IM = PGA. $R = 100$ IS-AL replications are used to estimate the CP for each training size n. $B = 200$ bootstrap samples are generated for the MLE to build the confidence ellipsoid at level $1-\xi = 0.9$ for n between 200 and 500, $B = 300$ bootstrap samples are generated for $n = 100$ due to numerical instabilities.

confidence ellipsoid can be used. Even though $\text{CP}_{100}^{\text{IA}}$ is less than the theoretical 90%, 70% is considered as acceptable in practice.

3.6.2.8 Empirical distributions of the fragility curves and influence of the IM value

The choice of the seismic IM is crucial for the accuracy of fragility curves estimates, especially when parametric models are concerned. So, empirical distributions of the fragility curves for IS-AL and RS methods are shown in Figure 3.12 when IM = PGA, and in Figure 3.13 when IM is the spectral acceleration (SA) at 5 Hz and 2% damping ratio. The parametric fragility curves estimated with a dataset of 10^4 seismic ground motions, are also shown in order to validate both the model choice and the uncertainty reduction provided by IS-AL.

With the PGA, a bias between the lognormal fragility curve and the k-means nonparametric fragility curve, called μ_{MC}, can be seen in Figure 3.12. This phenomenon could be explained by the small correlation between maximal displacement of the oscillator during the seismic excitation

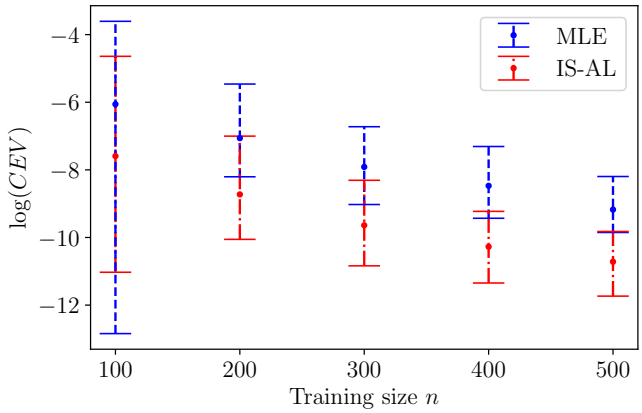


FIGURE 3.11: CEVs (for the confidence ellipsoid with level $1 - \xi = 0.9$ for θ) as a function of the training size n for IS-AL and MLE strategies when IM = PGA. The points are the medians over the $R = 200$ replications while the vertical lines are the ranges between the 10% and 90% quantiles.

and the PGA, which conveys small information about the seismic ground motion (Ciano, Gioffrè, and Grigoriu, 2020).

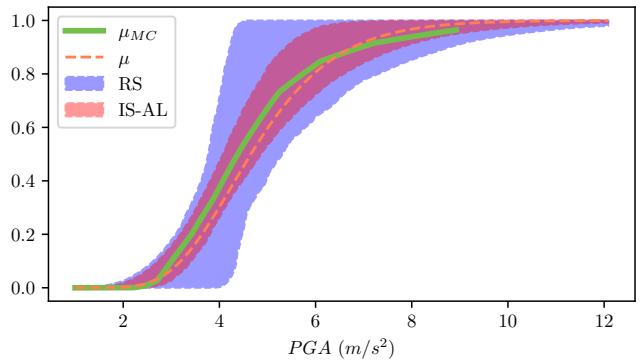


FIGURE 3.12: Empirical distribution of the fragility curves estimated by RS and IS-AL for the nonlinear oscillator for a training size of 120 mechanical computations. The dashed orange line and solid green line are respectively the parametric estimation μ using 10^4 seismic ground motions and the k-means nonparametric estimation of the fragility curve using 10^5 seismic ground motions μ_{MC}. The red and blue shaded areas correspond respectively to the 90% to 10% quantile ranges for the fragility curve dataset computed with IS-AL or RS. Remark that the nonparametric fragility curve is only plotted for $PGA < 10m/s^2$ due to the lack of seismic signals with PGA above that threshold.

The results presented in Figure 3.13 show a reduction of the bias between the nonparametric and the parametric fragility curve. This illustrates that, for the class of structures and for the seismic signal generator considered in this study, the parametric lognormal model has a better fit with the reference SA-based fragility curve than with reference PGA-based fragility curve.

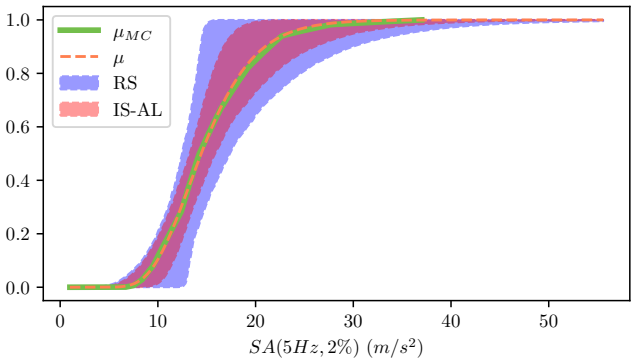


FIGURE 3.13: Empirical distributions of the fragility curves of RS and IS-AL for the nonlinear oscillator when IM is the spectral acceleration at 5 Hz and 2% damping ratio for a training size of 120 mechanical computations. The notations are the same as those in the Figure 3.12. Remark that the bias between the nonparametric fragility curve μ_{MC} and the parametric fragility curve μ is smaller than the one obtained when using the PGA as the intensity measure (compare with Figure 3.12).

3.6.2.9 Confidence interval for parametric fragility curves : towards the engineering practice

After assessing the validity of the asymptotic confidence ellipsoid for IS-AL thanks to the computation of the CP values in section 3.6.2.7, we can use the asymptotic Gaussian distribution to construct the CI of the parametric fragility curve, as in the engineering practice. Thus, using a single run of the IS-AL procedure, we estimate the asymptotic covariance matrix $\widehat{G}_n$ and sample fragility curve parameters from the asymptotic distribution $\mathcal{N}\left(\widehat{\theta}_n^{\mathrm{IA}}, \frac{\widehat{G}_n}{n}\right)$. For the sake of comparison, we also construct the CI on a single replication of the MLE procedure using the bootstrap technique.

Figure 3.14 represents the CIs for the fragility curve at level 90% for IS-AL and MLE strategies, for a single replication of size $n = 200$ of each procedure. Remark that the fragility curves estimated by MLE can be degenerated (i.e. as a unit step function), which implies that the CI for MLE is too conservative. This is consistent with the results of the figures 3.7 and 3.11.

3.6.2.10 Synthesis

In this section, we have shown that the IS-AL-based methodology is (i) efficient to reduce the variance of the fragility curve estimation and (ii) can be applied regardless of the IM of interest. However, in practice, it is more suitable to use an IM as correlated as possible to the response of the structure to minimize potential biases due to the use of a parametric model. In addition, we have shown that, if the computation times allow it, it is possible to know when to stop the IS-AL algorithm, in order to build asymptotic confidence ellipsoids.

3.6.3 Industrial test case: safety water supply pipe of a Pressurized Water Reactor (PWR)

3.6.3.1 Description of the piping system

The following test case corresponds to a piping system which is a simplified part of a secondary line of a French PWR. The numerical model was validated based on seismic tests performed on the shaking table Azalée of the EMSI laboratory of CEA/Saclay. The experimental

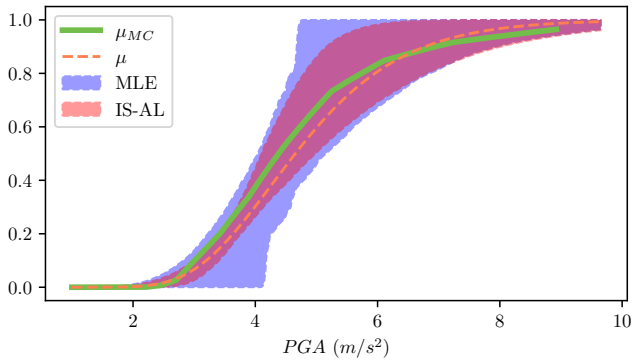
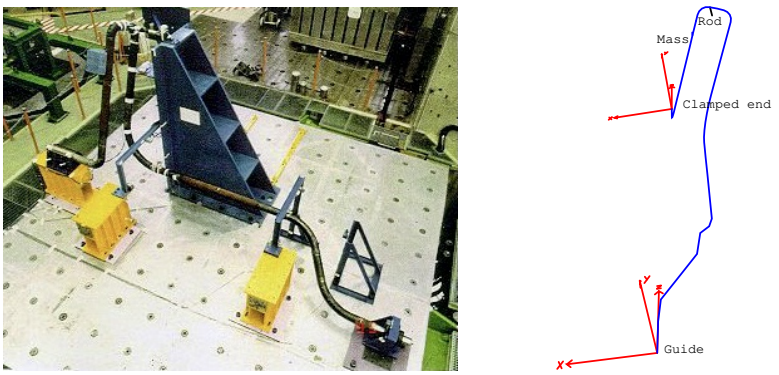


FIGURE 3.14: CIs of the parametric fragility curves of the elasto-plastic oscillator obtained with 500 samples of the parameter asymptotic Gaussian distribution for IS-AL and 500 bootstraped estimators with MLE, both for a training size of 300 mechanical computer simulations. The red and blue shaded areas correspond to the ranges between the 95% and 5% quantiles for respectively IS-AL and MLE. The solid green line corresponds to the Monte Carlo estimation (k-means nonparametric estimation) of the fragility curve based on a dataset of 10^5 of synthetic seismic signals.

program, called ASG program, and the main results are outlined in Touboul, Sollogoub, and Blay, 1999. In Figure 8.2a a view of the mock-up mounted on the shaking table is shown. The Finite Element (FE) model, based on beam elements, is depicted in Figure 8.2b.



(A) (B)

FIGURE 3.15: (a) Overview of the ASG mock-up on the CEA's shaking table and (b) ASG FE model.

The mock-up is a 114.3 mm outside diameter and 8.56 mm thickness pipe with a 0.47 elbow characteristic parameter, in carbon steel TU42C, filled with water without pressure. It contains three elbows and a mass modeling a valve (120 kg) which corresponds to more than 30% of the specimen total mass. As shown in Figure 8.2b, one end of the mock-up is clamped whereas the other is supported by a guide in order to prevent the displacements in the X and Y directions. Additionally, a rod is placed on the top of the specimen in order to limit the mass displacements in the Z direction. In the tests, excitation act in the X direction.

Numerical comparisons are carried out with the homemade FE code CAST3M (Charras and

Kichenin, 2011). Concerning the FE model, the boundary conditions are adjusted in order to obtain shapes and frequencies similar of those of the first two eigenmodes of the mock-up in the X and Y directions, respectively at 5.1 Hz and 6.6 Hz. As measured in the experiments, a critical damping ratio of 1% is considered for these two eigenmodes with a damping Rayleigh assumption. Finally, regarding the nonlinear constitutive law of the material, a bilinear law exhibiting kinematic hardening is used to reproduce the overall nonlinear behavior of the mock-up with satisfactory agreement compared to the results of seismic tests (Touboul, Sollogoub, and Blay, 1999).

In the context of this test case, the yield stress of the bilinear law is equal to $3.5\ 10^8$ Pa, the Young modulus is equal to $1.92\ 10^{11}$ Pa whereas the hardening modulus is equal to $4.3\ 10^8$ Pa. Moreover, since for the synthetic signals considered in this work (the same as those used in the reference Sainct et al., 2020 and in the second test case of this paper) the piping system remains in the linear domain, they are filtered by a fictitious linear single-mode building at 5 Hz and damped at 2%. Finally, we consider excessive out-of-plane rotation of the elbow located near the clamped end of the mock-up as failure criterion, as recommended in Touboul, Blay, et al., 2006. Since the weight of the mass is not completely taken up by the mechanical assembly, the overall behavior of the mock-up exhibits ratcheting.

In the following, the random variable R_e corresponds to the maximum of the out-of-plane rotation of the elbow. The binary variable which indicates the failure state is defined by $S = \mathbf{1}_{R_e > C}$ where C is the admissible rotation in degree. In our case, $C = 4.38°$. This value is the 90%-level quantile from a sample of 2000 mechanical simulations.

3.6.3.2 Performance metrics

For this test case, the numerical benchmark is based on 50 replications of 120 signals sampled using IS-AL (that includes the initial 20 points) with a defensive parameter $\varepsilon = 10^{-3}$ and 120 signals for the RS and MLE strategies.

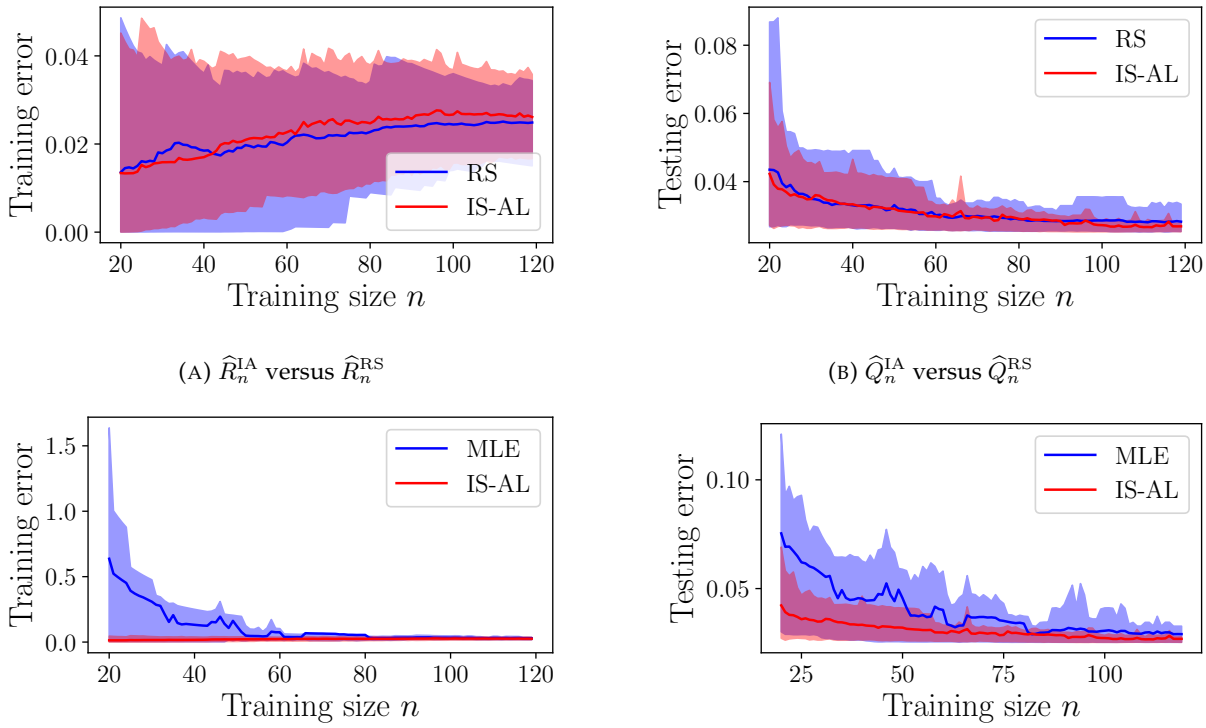
The IS-AL procedure is initialized by considering the linear FE model of the ASG mock-up and a numerical resolution based on a modal base projection. Thus, the initialization parameter $\widehat{\theta}_0^{\mathrm{IA}}$ is approximated by $\widehat{\theta}_{2000}^{\mathrm{RS}}$ using a 2000-sized dataset randomly selected from the 10^5 synthetic seismic ground motions. Then, 20 datapoints are queried with the instrumental density $q_{\widehat{\theta}_0^{\mathrm{IA}},\varepsilon}$ before launching the adaptive strategy. For the training, 100 signals are then chosen in a pool of 1500 CAST3M computations while 500 additional computations are carried out in order to compute the testing error.

Figure 3.16 compares the IS-AL, MLE and RS training and testing errors as functions of n. Remark that the training loss of MLE is greater than the training loss of IS-AL. This numerical artifact is essentially due to the regularization term β_{reg}: if the β parameter estimated by MLE is small, the penalization term $\beta_{\mathrm{reg}}/\beta$ can be very high.

Table 3.4 shows that the IS-AL strategy has overall better performance than the other two strategies.

3.6.3.3 Empirical distributions of the parameters α and β

Figure 3.17 compares the distributions of parameters α and β for several sample sizes between MLE and IS-AL using 50 replications. As with the nonlinear oscillator, the β parameter estimated with IS-AL is less likely to be close to 0 than when it is estimated with MLE. This motivates further the use of active learning to have a better accuracy for fragility curves parameters estimates with the same computational cost as state of the art estimation methods.



(A) $\widehat{R}_n^{\mathrm{IA}}$ versus $\widehat{R}_n^{\mathrm{RS}}$

(B) $\widehat{Q}_n^{\mathrm{IA}}$ versus $\widehat{Q}_n^{\mathrm{RS}}$

(C) $\widehat{R}_n^{\mathrm{IA}}$ versus $\widehat{R}_n^{\mathrm{MLE}}$

(D) $\widehat{Q}_n^{\mathrm{IA}}$ versus $\widehat{Q}_n^{\mathrm{MLE}}$

FIGURE 3.16: Numerical benchmark of the ASG piping system. The empirical distributions of the training and testing errors are computed, the red and blue shaded areas correspond to the area between the empirical quantiles of levels 10% and 90% of the 50 replications for respectively IS-AL, RS and MLE.

TABLE 3.4: Performance metrics for the ASG piping system for $n = 120$ when IM = SA (see Section 3.5.2.1)

	Train			Test		
$\bullet$	RS	MLE	IS-AL	RS	MLE	IS-AL
$\mathrm{RSD}_{120}^{\bullet}$ (%)	40.5	46	34	24.1	28	12
$\nu_{120}^{\bullet}$	0.93	2.4	×	1.3	5.8	×
$\mathrm{RB}_{120}^{\bullet}$ (%)	7.2	8.6	5.5	18	8.4	0.3

3.6.3.4 Fragility curve estimations

Figure 3.18 illustrates the uncertainty reduction provided by IS-AL on the fragility curve estimate. Motivated by the results obtained for the nonlinear oscillator, the fragility curve of the piping system is here expressed as a function of the pseudo-spectral acceleration of the initial set of synthetic signals (i.e not filtered signals), calculated at 5 Hz and 1% damping ratio.

Figure 3.19 represents the confidence interval on the fragility curve for IS-AL and MLE using a single replication of 200 CAST3M computations (20 computations used for initialization and 180 computations for IS-AL), obtained with the same methodology as that presented in section 3.6.2.9. Remark that the Monte Carlo estimation of the fragility curve (i) belongs to the confidence interval

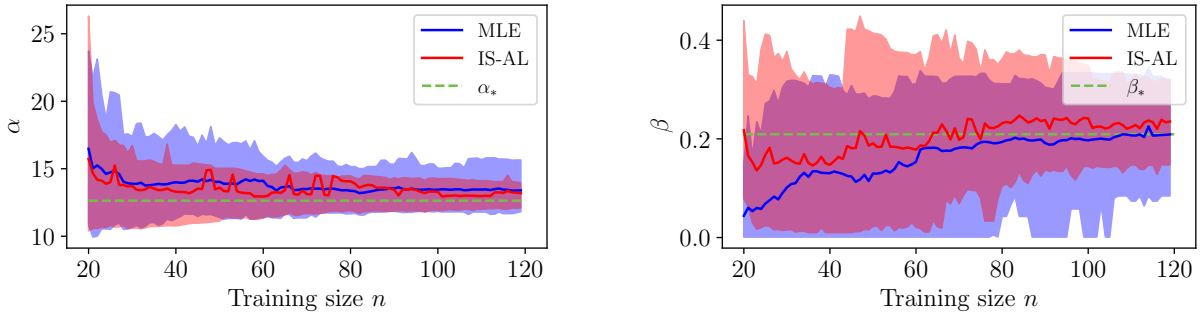


FIGURE 3.17: Results for the ASG piping system: the empirical distributions of parameters α and β are represented by the empirical 90% and 10% quantiles of 50 replications and correspond to the shaded blue and red areas respectively for MLE and IS-AL. The dashed green lines correspond to the values α_* and β_*, which have been here approximated by $\widehat{\alpha}_N$, $\widehat{\beta}_N$ for $N = 2000$.

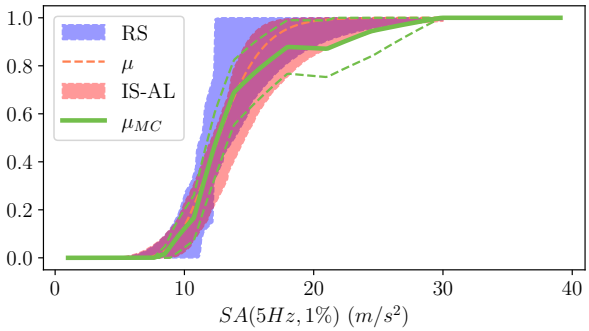


FIGURE 3.18: Empirical distributions of the fragility curves of RS and IS-AL for the ASG piping system. The red and blue areas correspond respectively to the ranges between the 10% and 90% quantiles of the fragility curve dataset generated with IS-AL and RS with $n = 120$ training datapoints (that includes the initialization points). The dashed orange line corresponds to a parametric fragility curve estimation using least squares minimization with a dataset of 2000 seismic ground motions and FE simulations of the piping system. The solid green line corresponds to a Monte Carlo estimation (k-means nonparametric estimation) of the fragility curve using the same 2000-sized dataset. The dashed green curves correspond to the 95% confidence interval of the Monte Carlo fragility curve estimation.

of IS-AL for seisms with relatively small spectral acceleration (ii) is not accurate for high spectral accelerations due to the lack of seismic signals of such intensities (see confidence intervals of the Monte Carlo fragility curve estimations).

As for the nonlinear oscillator, the figures 3.18 and 3.19 suggest that for the RS and MLE strategies, even with $n = 120$ points, it is possible to obtain samples for which a β estimate is close to 0, which IS-AL avoids.

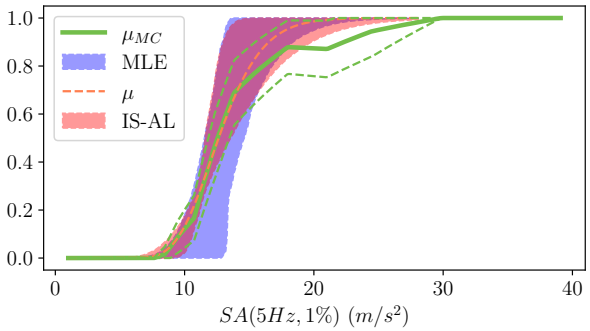


FIGURE 3.19: Parametric fragility curve of the ASG piping system confidence interval obtained with 500 samples of the parameter asymptotic Gaussian distribution for IS-AL and 500 bootstraped estimators with MLE. The red and blue shaded areas correspond to the ranges between the 95% and 5% quantiles for respectively IS-AL and MLE. The solid green line corresponds to a Monte Carlo estimation (k-means non-parametric estimation) of the fragility curve using our 2000 sized dataset of CAST3M computations. The dashed green curves correspond to the 95% confidence interval of the Monte Carlo fragility curve estimation.

3.7 Conclusion

In this chapter, for problems involving expensive numerical simulations, we have introduced an original methodology to improve the accuracy of parametric IM-based fragility curves estimations without increasing the sample size, thanks to an active learning strategy based on importance sampling. Defensive strategy has been implemented to control the likelihood ratio and the possible increase of the training loss variance in the early steps. We use a penalized least square loss to avoid an identifiability issue of the standard deviation of the lognormal model. We define a convergence criterion that indicates asymptotic normality of the estimator and provides asymptotic confidence intervals and ellipsoids. This methodology concerns problems for which aleatory uncertainties (i.e. due to the excitation) have a predominant contribution in the variability of the structural response. Taking into account epistemic uncertainties will be the subject of the next chapters. We have illustrated the performance of the proposed active learning procedure with numerical examples dealing with a synthetic case up to the FE mechanical simulation of a piping system of a French PWR. In comparison with the engineering practice based on the joint use of the MLE and the boostrap techniques, the proposed methodology is more efficient. For the same number of calculations, the IS-AL procedure reduces the variance of the parametric estimation of the fragility curve and gives theoretical guarantees on the convergence of the estimations. Finally, note that the performances of the IS-AL algorithm increase, compared to the classical methods, with the failure thresholds. In practice, even it is difficult to estimate in advance what the effective gain will be, it will be in favor of the proposed algorithm for well-designed structures facing the seismic risk.

CHAPTER 4

Uncertainty studies for black-box computer models

Les ordinateurs, plus on s'en sert moins, moins ça a de chances de mal marcher.

Devise Shadok

Contents

4.1 Introduction

Numerical simulation consists in representing a real experiment using a computer code. Computer models are now essential for simulating and designing complex systems such as mechanical structures in industrial facilities. Computer simulation is now considered as a third branch for studying phenomena, after theory and real experiments. Its main advantage is to replace costly or infeasible real experiments by numerical simulations.

In order to assess that the system studied is always in operational conditions, it is necessary to precisely estimate the uncertainties tainting specific quantities of interest of the system. In some industries, it is even mandatory for regulatory reasons (such as in the nuclear industry). Uncertainty Quantification (UQ) aims at developing specific methodologies to address this issue using a probabilistic framework (in most cases). It can be the estimation of the probability of a rare

event, optimizing quantities of interests under constraints, or assessing which design parameter uncertainty influence the most the uncertainty of the system's quantity of interest (e.g. sensitivity analysis).

UQ concerns more globally the assessment of uncertainties sources on a model representing an observed phenomenon. UQ framework can be used for several types of decision-making processes such that:

- Design of a system to optimize its exploitation under costs and risks constraints;
- Estimation of a regulatory criterion;
- Understand more precisely the behavior of a phenomenon for R&D purposes.

In the scope of this thesis, the phenomenon studied is the dynamical response of a mechanical structure to seismic ground motions.

4.2 Uncertainty quantification

4.2.1 A general framework

Uncertainty quantification of a numerical model comes after a modeling phase of a physical system. Although in this manuscript the models considered will be related to structural mechanics, all kind of physical models can be considered in this framework (e.g. thermohydraulics, neutronics, fluid mechanics, etc).

Figure 4.1 describes the general framework of uncertainty quantification of computer models in engineering studies (see e.g. Sudret, 2007; De Rocquigny et al., 2008; Iooss, 2009). The UQ framework is divided in several (some of them optional) steps:

- **Step A:** Specification of the problem studied and its main objectives. Definition of the input and output variables, the quantity of interest, and the definition of which input parameters are uncertain;
- **Step B:** Uncertainty modeling of the input parameters. It consists in describing all the sources of uncertainties affecting the parameters and choosing a framework to describe them (e.g. probabilistic framework);
- **Step B':** The validation and verification (VVUQ) phase of the computer model w.r.t. available data. This step consists mainly of solving ill-posed inverse problems in order to calibrate the computer model. It is beyond the scope of this manuscript;
- **Step C:** The uncertainty propagation step consists in evaluating the uncertainty of the model output coming from input parameters uncertainty. The model output uncertainty is thus summarized by a quantity of interest $\phi(Y)$ which is in most cases the mean of the computer model output $\mathbb{E}_{\mathbf{X}}[\mathcal{M}(\mathbf{X})]$ or a failure probability $\mathbb{E}_{\mathbf{X}}[\mathbb{1}_{(\mathcal{M}(\mathbf{X})>C)}]$;
- **Step C':** The sensitivity analysis step aims at apportion the output uncertainty to each input parameter uncertainty. It permits to rank the input parameters by influence on the model output variability and thus help to make information based design choices. This could guide the assessment of which parts or elements of the system studied deserve priority in subsequent modeling and data gathering steps.

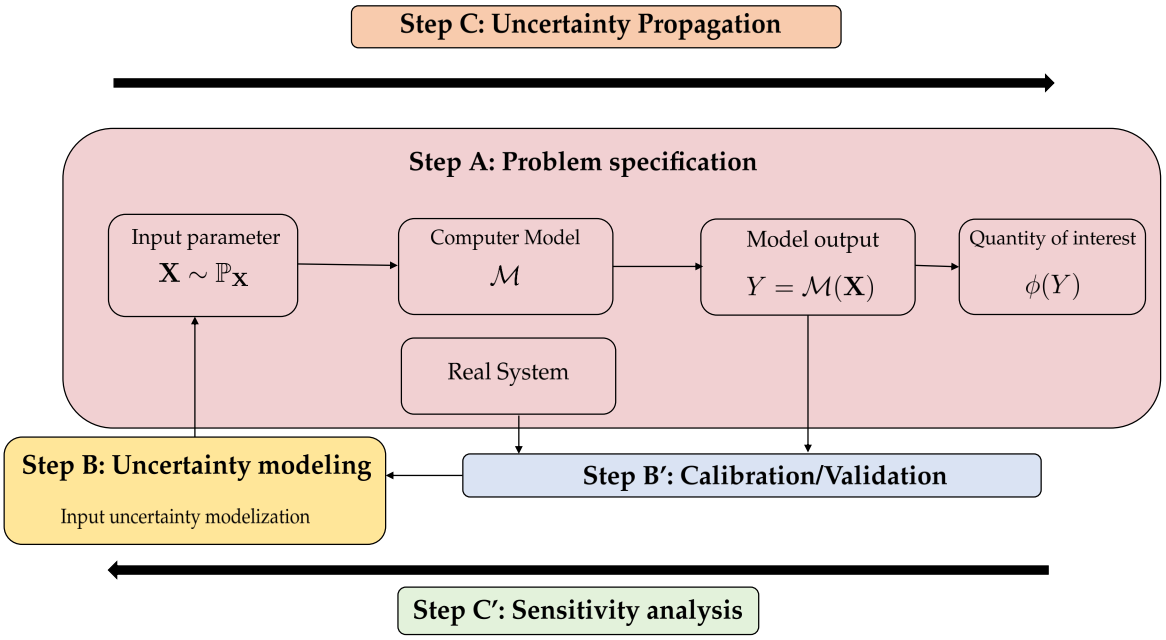


FIGURE 4.1: General framework for uncertainty quantification of computer models

The methodology for solving these different steps is now well-known. For instance, the Bayesian calibration of computer codes using Gaussian process (M. C. Kennedy and O'Hagan, 2001), the substitution of costly computer codes by surrogate models based on Gaussian process (Sacks et al., 1989) and variance-based sensitivity indices for sensitivity analysis purposes (Sobol', 1993; Sobol', 2001). However, research in UQ is still popular, and it is possible to identify some recent trends:

- Theoretical and applied aspects about input uncertainty modeling has been recently studied in Lemaitre, 2014; Meynaoui, 2019; Stenger, 2020; Gauchy et al., 2021;
- Global sensitivity analysis (GSA) methodologies for dependent input parameters has been developed in Chastaing, 2013; Broto, 2020. GSA for goal-oriented purposes was also recently developed in Browne, 2017; Chabridon, 2018;
- The development of surrogate models of computer code using multifidelity (e.g. where several computer models are available) has been made in Le Gratiet, 2013; Stroh, 2018, or with adaptive approximation of the computer model in Haberstich, 2020.

The next sections will describe more precisely the different steps of the UQ framework.

4.2.2 Black-box computer model

Numerical simulations are usually carried out using computer models. Due to the high complexity of the physical phenomena studied and in order to have a flexible framework, such computer codes will be seen throughout this manuscript as a black-box function. This means that a computer model (or computer code) will be formally seen as a scalar valued function:

$$\begin{aligned} \mathcal{M}\colon \mathcal{X} \subset \mathbb{R}^p &\to \mathcal{Y} \subset \mathbb{R} \\ \mathbf{x} &\mapsto y = \mathcal{M}(\mathbf{x}) \end{aligned} \tag{4.1}$$

$\mathbf{x} = (x_1, \ldots, x_p)$ is a set of p input parameters of the computer model. They are generally physical quantities depending on the type of physics studied for the application (example: Young modulus for mechanics). In most general cases, the output y of the computer model is considered to be a scalar. The computer model complexity is summarized into the mapping $\mathcal{M}(.)$. The computer model can be in most simple cases an analytical formula or complex multiphysics computational model based for instance on Finite Elements (FE) in mechanics or Computational Fluid Dynamics (CFD) in thermohydraulics. This complexity justifies the consideration of the computer model as a black-box function that is only known pointwise. It is important to also notice that the computer model can be very expensive to evaluate as well as in terms of computational time (computer model based on FE or CFD can take several days for a single run based on a fixed value of input parameters $\mathbf{x}$) as in terms of financial budget (in some cases the computer model is outsourced in a specialized company and each single run results in fees to this company). The computer model $\mathcal{M}$ can be deterministic (e.g. for a fixed $\mathbf{x}$ the value of $y = \mathcal{M}(\mathbf{x})$ is fixed) or stochastic (two runs of the computer model at fixed input parameters $\mathbf{x}$ gives different values for the output y).

Nowadays, it is not enough to perform simulation of a system only on a fixed set of input parameters $\mathbf{x}$ and uncertainties coming from the input parameters can be very impactful on the behavior of the real phenomenon simulated by the computer model. The potential impact of these uncertainties can lead to undesirable consequences (e.g. financial consequences) up to dramatic consequences (e.g. direct threats to human life, huge environmental negative impact). Thus, providing a rational inventory of the different sources of uncertainties and a clear mathematical framework to model them is a key step in UQ methodology (see Figure 4.1).

4.2.3 Sources of uncertainties

For an efficient use of the computer model, the uncertainty tainting the input parameters of the computer model has to be identified and quantified for several reasons. It can be due to safety reasons like in aerospatial or nuclear industry where the need of high safety margins are imposed by regulators. But as pointed out in Soize, 2005; Sudret, 2007 it is also related to the concept itself of simulation of a physical system: the system simulated by the computer model with a given set of input parameters may satisfy the imposed design criteria but in practice, the real system is a man-made physical system in which uncertainty arises in all cases. Thus, the very distinction between simulation and reality imposes the need to quantify the sources of uncertainties.

From an engineering point of view, after the characterization step of the different uncertainties tainting the system studied. It is important to classify which uncertainties are reducible in a near-term with a reasonable budget to help engineers to make information based design choices on these sources of uncertainties. This leads to introduce two main types of uncertainties:

- **Aleatory uncertainty** refers to natural variability of a physical phenomenon that remains irreducible. For instance, seismic ground motions are considered aleatory due to the difficulty to predict them;
- **Epistemic uncertainty** refers to the lack of knowledge of the analyst and may be reducible by gathering more information (by data acquisition or expert judgement). Indeed, the word "epistemic" comes from the Greek word $\epsilon\pi\iota\sigma\tau\eta\mu\eta$, meaning "knowledge". In the nuclear or

aerospatial industry, the uncertainty on the mechanical parameters of a structure is considered epistemic because the R&D budget in these industries allows making more measurements and complex uncertainty models of these variables than in general cases.

Following the lines of Der Kiureghian and Ditlevsen, 2009, the distinction between aleatory and epistemic uncertainties is subjective: it is a pragmatic classification between manageable uncertainties for which engineers can allocate resources to reduce them to those for which it is impossible.

After identifying and classifying the sources of uncertainties, the analyst have to model these uncertainties with a mathematical framework such as the probabilistic framework. Some approaches consider non-probabilistic paradigm to model uncertainty such as random set (Tonon and B. A., 1998) or Dempster-Shafer theory (Shafer, 1976), but this is beyond the scope of this manuscript.

4.2.4 Probabilistic modeling of uncertainty

This section gives some insights about measure theory which is the mathematical core of the probabilistic modeling that will be use extensively in this manuscript. Let $(\Omega, \mathcal{A}, \mathbb{P})$ by a probability space. For $\mathcal{X} \subset \mathbb{R}^p$ we denote $\mathcal{B}(\mathcal{X})$ the Borel σ-algebra on $\mathcal{X}$. A random vector $\mathbf{X} = (X_1, ..., X_p)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a measurable function such that:

$$\begin{aligned} \mathbf{X} \colon \Omega &\to \mathcal{X} \\ \omega &\mapsto \mathbf{X}(\omega) \end{aligned} \tag{4.2}$$

The probability measure $\mathbb{P}_{\mathbf{X}}$ of $\mathbf{X}$ can be defined by two manners:

- The joint cumulative distribution function (CDF) $F_{\mathbf{X}} \colon \mathcal{X} \mapsto [0, 1]$ such that: $F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_p \leq x_p)$;
- The joint probability density function (PDF) $f_{\mathbf{X}} \colon \mathcal{X} \mapsto \mathbb{R}^+$ only in cases where $\mathbf{X}$ is absolutely continuous w.r.t. the Lebesgue measure on $\mathcal{X}$. It is easily obtained from the CDF such that: $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^p F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \ldots \partial x_p}$.

We now precise a fundamental theorem of probability theory that is the theoretical cornerstone of UQ framework.

Theorem 1 (Transport theorem)**.** *Let $\mathbf{X}$ be a p-dimensional random vector with $f_{\mathbf{X}}$ its joint pdf. Assume that $\varphi \colon \mathbb{R}^p \mapsto \mathbb{R}$ is a measurable function. Then $\mathbb{E}[\varphi(\mathbf{X})]$ is given by:*

$$\mathbb{E}[\varphi(\mathbf{X})] = \int_{\mathbb{R}^p} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{4.3}$$

if the integral is absolutely convergent.

The transport theorem is the way engineers performs the uncertainty propagation step of UQ framework. Indeed, given $Y = \mathcal{M}(\mathbf{X})$ the computer model output, the quantity of interest $\phi(Y)$ is most of the time a generalized moment that can be expressed as:

$$\begin{aligned}\phi(Y) &= \mathbb{E}_Y[\varphi(Y)] \\ &= \mathbb{E}_{\mathbf{X}}[\varphi(\mathcal{M}(\mathbf{X}))] \\ &= \int_{\mathbb{R}^p} \varphi(\mathcal{M}(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \,.\end{aligned} \tag{4.4}$$

However, the p-dimensional integral in 4.4 is most of the time not analytic and the engineer has to rely on numerical methods of integration. The most classical one is the Monte-Carlo method, which is based on the law of large numbers theorem.

Theorem 2 (Strong law of large numbers)**.** *Let $\mathbf{X}$ be a p-dimensional random vector with $f_{\mathbf{X}}$ its joint pdf. For a sample $(\mathbf{X}_i)_{1\leq i\leq N}$ of independent realizations of $\mathbf{X}$, if for $\varphi\colon \mathbb{R}^p \mapsto \mathbb{R}$ a measurable function we have $\mathbb{E}[|\varphi(\mathbf{X}_1)|] < +\infty$ then :*

$$\frac{1}{N}\sum_{i=1}^{N} \varphi(\mathbf{X}_i) \xrightarrow[N\to+\infty]{\text{a.s.}} \int_{\mathbb{R}^p} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{4.5}$$

In order to compute the quantity of interest $\phi(Y)$, a sample $(\mathbf{X}_i)_{1\leq i\leq N}$ of the input parameters of the computer code is generated using its known pdf function. Then the computer model output $Y_i = \mathcal{M}(\mathbf{X}_i)$ is computed for each input parameter vector in the N-sized sample. Finally, the quantity of interest is approximated thanks to the law of large numbers:

$$\phi(Y) \approx \frac{1}{N}\sum_{i=1}^{N} \varphi(Y_i) \tag{4.6}$$

4.3 Framework of uncertainty quantification for earthquake engineering

The aim of this section is to show how to extend the general framework presented above into seismic probabilistic risk assessment. Indeed, it is possible to see seismic fragility curve as a goal oriented quantity of interest of a stochastic computer model.

4.3.1 The specific case of stochastic computer models

In most of UQ literature, the computer model is always considered deterministic w.r.t. the input parameters vector $\mathbf{X}$. However, in some engineering studies an uncontrolled uncertainty can appear. For instance, seismic probabilistic risk assessment (SPRA) has to account for the uncontrolled seismic hazard coming from a ground motion loading that is considered as the realization of a stochastic process. In this case, we will consider that the computer model is a function of two input parameters:

- The seismic intensity measure A which quantifies the strength of the seismic loading;
- The vector of parameters of the structure studied denoted $\mathbf{X} \in \mathbb{R}^p$.

Note that the vector $\mathbf{X}$ is a random vector of controllable input parameters of the computer model with probability distribution $\mathbb{P}_{\mathbf{X}}$, whereas A depends on the seismic loading and hence cannot be considered controllable. In SPRA we will thus have to consider the case of a stochastic computer model which can be formalized as follows:

$$\begin{aligned} \mathcal{M}\colon \mathbb{R}^{p+1} &\to \mathbb{R} \\ (a,\mathbf{x}) &\mapsto y(a,\mathbf{x}) = \log(z(a,\mathbf{x})) = g(a,\mathbf{x}) + \varepsilon \,, \end{aligned} \tag{4.7}$$

where $z(a,\mathbf{x})$ is the mechanical demand of the simulated structure, the log mechanical demand $y(a,\mathbf{x})$ is usually considered due to the presence of multiplicative noise on the mechanical demand for most of the structures. g is the deterministic part of the computer model and ε the stochastic part which depends on an uncontrollable parameter ω and of $(a,\mathbf{x})$ in terms of interaction with ω:

$$\varepsilon = \varepsilon(\omega, a, \mathbf{x}) \,, \tag{4.8}$$

where $\mathbb{E}[\varepsilon|A,\mathbf{X}] = 0$. ω can be seen as the noise coming from the stochastic ground motion, and which is difficult to represent with a reasonable number of scalar parameters. We emphasize that with this kind of models, several mechanical simulations with the same value of $(a,\mathbf{x})$ and a different seismic ground motion will give different outputs.

4.3.2 Uncertainty quantification for seismic fragility curves

Seismic fragility curve is a peculiar quantity of interest arising from seismic probabilistic risk assessment (SPRA) studies (R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984). It is the probability of exceedance of a threshold C of a mechanical demand Z conditional to a seismic intensity measure A. Following the lines of Der Kiureghian and Ditlevsen, 2009, we classify the sources of uncertainties in SPRA for the nuclear industry between aleatory and epistemic uncertainties. In the frame of this manuscript, the seismic ground motion uncertainty will be considered as aleatory uncertainty while the mechanical parameters of the structure will be considered as epistemic uncertainty. The main goal of the remainder of this manuscript is to study the impact of the epistemic uncertainties on the seismic fragility curves. Indeed, measuring the impact of the epistemic uncertainties on the seismic fragility curves will help mechanical engineers to propose information based design choices and goal oriented risk assessment while in the same time using the existing tools of SPRA. We thus propose a new definition of fragility curve that will be the conditional probability of exceedance of a threshold to a seismic intensity measure and a mechanical parameter vector:

$$\Psi(a,\mathbf{x}) = \mathbb{P}_{\varepsilon}(z(A,\mathbf{X}) > C | A = a, \mathbf{X} = \mathbf{x}) \,. \tag{4.9}$$

This new definition allows us to consider the random function $a \to \Psi(a,\mathbf{X})$ as the goal-oriented functional output for UQ studies in a SPRA context. From this functional output we can derive functional quantities of interest such as the mean fragility curve:

$$\bar{\Psi}(a) = \mathbb{E}_{\mathbf{X}}[\Psi(a,\mathbf{X})] \,. \tag{4.10}$$

One can also be interested in more conservative statistical quantities that will be useful for risk analysis. We introduce the seismic fragility quantiles curves $a \to q_\gamma(a)$ of level $\gamma \in [0,1]$ such that:

$$q_\gamma(a) = \inf_{q\in\mathbb{R}} \{\mathbb{P}_{\mathbf{X}}(\Psi(a,\mathbf{X}) \leq q) \geq \gamma\} \,. \tag{4.11}$$

The effect of epistemic uncertainties on the fragility curve is assessed by uncertainty quantification using a Monte-Carlo sample $(\mathbf{X}_i)_{1\leq i\leq N}$ w.r.t. the probability measure $\mathbb{P}_{\mathbf{X}}$. Then, the quantity of interest is approximated thanks to the law of large numbers:

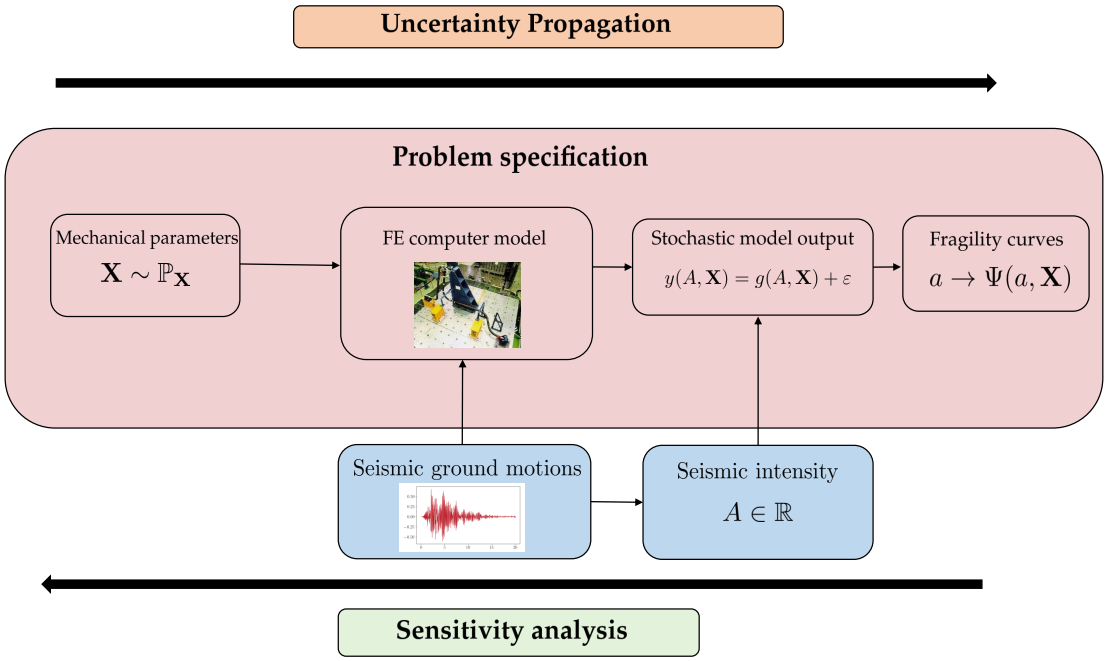


FIGURE 4.2: Uncertainty quantification framework adapted to seismic probabilistic risk assessment

$$\begin{aligned} \bar{\Psi}(a) &\approx \frac{1}{N}\sum_{i=1}^{N} \Psi(a, \mathbf{X}_i) \ , \\ q_\gamma(a) &\approx \inf_{q\in\mathbb{R}} \{ \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{(\Psi(a,\mathbf{X}_i)\leq q)} \geq \gamma \} \ . \end{aligned} \tag{4.12}$$

Figure 4.2 shows the adaptation of the general UQ framework to seismic probabilistic risk assessment in the same spirit as in Figure 4.1. Denote the two sources of randomness in this framework: the aleatory uncertainty coming from the seismic ground motions and the epistemic uncertainty coming from the mechanical parameters of the structure. This distinction between sources of uncertainties is once again subjective and context dependent: in the case of nuclear safety, mechanical parameters uncertainty is epistemic due to the very high level of safety imposed by regulatory agencies and thus uncertainty reduction on the mechanical parameters of the structure is considered feasible in a near term. This is for instance generally not the case for seismic risk assessment in civil engineering where mass production of structures and financial constraints make infeasible the reduction of uncertainty on their mechanical parameters, which is the main definition of epistemic uncertainty.

4.4 Conclusion

This chapter sets the main framework of uncertainty studies in engineering using black-box computer models. Probability theory allows us to formalize mathematically the concept of uncertainty and how to define valuable quantities of interest that can be insightful in the context of

uncertainty studies. This framework can be extended to seismic probabilistic risk assessment of structures for nuclear safety, by introducing an aleatory source of uncertainty: the seismic ground motions. A goal-oriented statistic coming from the aleatory uncertainty is then defined: the seismic fragility curve. The mechanical parameters of the structure are considered tainted by epistemic uncertainties, represented by a probability distribution. This uncertainty is then propagated into the seismic fragility curve. However, the Monte-Carlo sample size N and the number of simulations necessary to obtain an estimation of $a \rightarrow \Psi(a, \mathbf{X}_i)$ make the direct use of the stochastic computer model $\mathcal{M}$ too costly for practical applications to mechanical structures. We thus will have to build a surrogate for $\mathcal{M}$ using statistical learning methods to make the uncertainty propagation step and the sensitivity analysis step numerically tractable. In the next chapter, we will propose to build a surrogate based on Gaussian process regression in order to perform uncertainty propagation. The main advantage of this surrogate is to propose both a prediction and an uncertainty on the mechanical computer model output, which is very appealing in a context of seismic safety studies.

CHAPTER 5

Surrogate modeling of black-box computer models using Gaussian process

Au-delà de l'outil, et à travers lui, c'est la vieille nature que nous retrouvons, celle du jardinier, du navigateur ou du poète.

Saint-Exupéry, *Terre des hommes*

Contents

5.1 Introduction

The increasing complexity of computer simulation and the need of a huge number of simulations for Uncertainty Quantification (UQ) and optimization motivate the elaboration of meta-models built using only a limited number of numerical simulations. These meta-models are used as a surrogate (or emulator) of the numerical computer code. A surrogate can be simply defined as being a substitute to a real computer code. Its main advantage is to perform fast evaluation of the output quantities of the computer model, allowing the engineers to save money and computation time while keeping the same operational objectives. However, one main drawback is the approximation error between the real computer model and the surrogate model, coming from the very definition of the surrogate. If this approximation error is not properly measured and controled, it is not possible to be confident in the prediction by a surrogate of statistical quantities of interest for an UQ study. This need of confidence measurement of a surrogate is even more significant in the context of the nuclear industry, where the safety level imposed by national regulatory agencies is very high. Since the late 1990s, Gaussian process (GP) regression imposed itself as a canonical methodology for surrogate modeling of computer models. A main property of Gaussian process that makes it stand out of the other surrogates methods is its probabilistic nature which allows to provide a quantification of the uncertainty of the predictions of the surrogate itself. Moreover, the uncertainty on the GP predictions can be quantified analytically, this makes GP regression virtuous in terms of energy consumption for its computation.

GP regression is also known as kriging in the geostatiscal litterature of the 1960s (Matheron, 1962). The term kriging comes from the South-African engineer Danie G. Krige who aims to estimate the spatial distribution of ore concentration in a mine field using a small number of drilling (Krige, 1951). The kriging predictor is just a weighted linear combination of the observed ore concentration at the drilling places. In the field of computer experiments, the Gaussian hypothesis is added due to the high dimension of the input parameters of the computer model (Iooss, 2009, Section 3.3). GP regression has been then popularized in the field of computer experiments (Sacks et al., 1989) and machine learning (Rasmussen and C. Williams, 2005).

This chapter aims at presenting the main methodology of surrogate modeling using Gaussian process with a practical implementation for seismic fragility curves estimation. In Section 5.3, the so-called kriging equations used to obtain the predictive distribution of the Gaussian process and the Bayesian viewpoint of Gaussian process regression are presented. Then in Section 5.4 aspects about the role of the covariance function of the Gaussian process on its regularity and how to infer a covariance function from observed data are detailed. In Section 5.5 we propose a methodology to estimate seismic fragility curve using a Gaussian process surrogate. Finally, Section 5.6 proposes an application on a nonlinear single d.o.f. oscillator.

5.2 General considerations about Gaussian process regression

This section introduces the main concepts of GP regression. At first, general definitions will be introduced. Then the kriging equations that are the core of GP regression technique will be derived for the simple and the universal case. The practical use of a kriging surrogate is presented. Finally,

the kriging equations are written in the case of noisy observations, which leads to a Bayesian viewpoint of GP regression.

Definition 1. *Let $(\Omega, \mathcal{G}, \mathbb{P}_G)$ be a probability space and $\mathcal{X}$ a subset of $\mathbb{R}^p$. A Gaussian process on $\mathcal{X}$ is a collection of real-valued random variables $G = \{G(\mathbf{x}) \colon \Omega \to \mathbb{R} | \mathbf{x} \in \mathcal{X}\}$ such that for any subset $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\} \subset \mathcal{X}$, the vector $(G(\mathbf{x}_1), \ldots, G(\mathbf{x}_d))^T$ is a Gaussian vector.*

The distribution of the Gaussian process is uniquely determined by its mean function defined as:

$$m : \mathbf{x} \to \mathbb{E}_G[G(\mathbf{x})] \, , \tag{5.1}$$

and its covariance function Σ defined as:

$$\begin{aligned} \Sigma \colon \mathcal{X} \times \mathcal{X} &\to \mathbb{R} \\ (\mathbf{x}, \mathbf{x}') &\mapsto \Sigma(\mathbf{x}, \mathbf{x}') = \mathbb{E}_G[(G(\mathbf{x}) - m(\mathbf{x}))(G(\mathbf{x}') - m(\mathbf{x}'))] \, . \end{aligned} \tag{5.2}$$

Generally a parametric model of the following form is considered for the mean function:

$$m(\mathbf{x}) = \sum_{j=1}^{r} \eta_j u_j(\mathbf{x}) = \eta^T u(\mathbf{x}) \, , \tag{5.3}$$

where $\eta = (\eta_1, \ldots, \eta_r)^T \in \mathbb{R}^r$ a vector of parameters and $u = (u_1, \ldots, u_r)$ is a set of basis functions from $\mathcal{X}$ to $\mathbb{R}$. The GP G with mean function m and covariance function Σ will be simply denoted by:

$$G \sim \mathcal{GP}(m, \Sigma) \, .$$

5.3 Kriging equations

This section will present the core of GP regression and its application to approximation of functions. Let us consider a function $g \colon \mathcal{X} \mapsto \mathbb{R}$, this function can represent a computer model as in Sacks et al., 1989 or a spatially distributed natural variable (see e.g. Krige, 1951; Stein, 1999; Chilès and Delfiner, 2012). For $\omega \in \Omega$, we denote by $G(.;\omega)$ a realization of the GP G. In order to perform GP regression, we will make the assumption that g is a realization of the GP G. This boils down more formally to say:

$$\exists \omega \in \Omega \, , \forall \mathbf{x} \in \mathcal{X} \, , \; g(\mathbf{x}) = G(\mathbf{x}; \omega) \, . \tag{5.4}$$

As pointed out in Chilès and Delfiner, 2012, Section 1.2, this hypothesis has to be considered as a modeling choice. This does not mean that this choice is arbitrary - it can be suggested by prior knowledge on the function g - but that this choice cannot be falsified in K. Popper sense, i.e. it cannot by refuted by objective observations. Moreover, GP regression can also be viewed in a Bayesian way and will be discussed in Section 5.3.4.

We define the learning set $\mathcal{D}_n = (\mathbf{x}_i, g(\mathbf{x}_i))_{1 \leq i \leq n}$. The goal of kriging is to predict for any $\mathbf{x} \in \mathcal{X}$ the value of g by only knowing its evaluations $g^n = (g(\mathbf{x}_1), \ldots, g(\mathbf{x}_n))^T$ in the learning set. We define also the observation vector by $G^n = (G(\mathbf{x}_1), \ldots, G(\mathbf{x}_n))^T$. The main advantage of GP regression is that the conditional distribution of $G(\mathbf{x})$ knowing $G^n = g^n$, denoted by $(G(\mathbf{x})|G^n = g^n)$ is Gaussian with an analytic expression for the mean $m_n(\mathbf{x})$ and variance $s_n^2(\mathbf{x})$ obtained using the so-called kriging equations. We can derive these equations in two manners given some assumptions:

- The case where the mean function m of the GP is known is called simple kriging.
- The case where the mean function has to be estimated using the learning set $\mathcal{D}_n$ is called universal kriging.

Through all this section, the covariance function of the GP is supposed known. The case of unknown covariance functions will be treated in Section 5.4.

5.3.1 Simple kriging

We will start by the case of simple kriging. The mean function m is supposed to be known. The following theorem gives the expression of the kriging equations in this case:

Theorem 3 (Simple kriging equation)**.** *Under the assumption that the mean function m is known, the conditional distribution of $G(\mathbf{x})$ given $(G^n = g^n)$ is:*

$$(G(\mathbf{x})|G^n = g^n) \sim \mathcal{N}\left(m_n(\mathbf{x}), s_n^2(\mathbf{x})\right) \ , \tag{5.5}$$

where the kriging mean $m_n(\mathbf{x})$, and the kriging covariance $s_n^2(\mathbf{x}, \mathbf{x}')$ are given by:

$$\begin{aligned} m_n(\mathbf{x}) &= m(\mathbf{x}) + c_n(\mathbf{x})^T \Sigma_n^{-1}(g^n - m^n) \ , \\ s_n^2(\mathbf{x}, \mathbf{x}') &= \Sigma(\mathbf{x}, \mathbf{x}') - c_n(\mathbf{x})^T \Sigma_n^{-1} c_n(\mathbf{x}') \ . \end{aligned} \tag{5.6}$$

$m^n = (m(\mathbf{x}_1), \ldots, m(\mathbf{x}_n))^T$ is the mean vector at observations points, $\Sigma_n = (\Sigma(\mathbf{x}_i, \mathbf{x}_j))_{1 \leq i,j \leq n}$ is the covariance matrix of the Gaussian vector G^n and $c_n(\mathbf{x}) = (\Sigma(\mathbf{x}_1, \mathbf{x}), \ldots, \Sigma(\mathbf{x}_n, \mathbf{x}))^T$. This conditional GP is referred as the kriging surrogate.

The kriging variance $s_n^2(\mathbf{x}, \mathbf{x})$ will be denoted $s_n^2(\mathbf{x})$ for the sake of notation simplicity. Some comments about the simple kriging equations can be made:

- The kriging variance expression depends on the input observation points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and the new point $\mathbf{x}$ where we want to predict the value of g. This means that the kriging variance is easy and fast to compute because it does not depend on the value of $g(\mathbf{x})$. This property is particularly useful for performing design of experiments (see Echard, Gayton, and Lemaire, 2011).
- Remark that if $\mathbf{x} = \mathbf{x}_i$ for $1 \leq i \leq n$, the kriging variance $s_n^2(\mathbf{x}_i) = 0$ and the prediction $m_n(\mathbf{x}_i) = g(\mathbf{x}_i)$. This means that the kriging surrogate interpolates the known observations. This is of peculiar interest for application to deterministic computer models.
- The expressions of the kriging mean and variance are explicit and analytic (i.e. it only involves matrix multiplications and inversions). In most industrial applications, the sample size n is not large, thus their numerical evaluation is fast and economical in computational resources (e.g. the computational time or the energy consumption).

Figure 5.1 gives an example of kriging prediction with the univariate function $g(x) = 2\pi x \sin(2\pi x)$. A very useful graphical tool to illustrate the prediction uncertainty of the conditional GP is the prediction interval of level δ. It is defined by:

$$\left[m_n(\mathbf{x}) - \Phi^{-1}(1 - \delta/2)s_n(\mathbf{x}), m_n(\mathbf{x}) + \Phi^{-1}(1 - \delta/2)s_n(\mathbf{x})\right] \ , \tag{5.7}$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. The green shaded area in Figure 5.1 corresponds to the predictive interval of level 95%. As expected, the

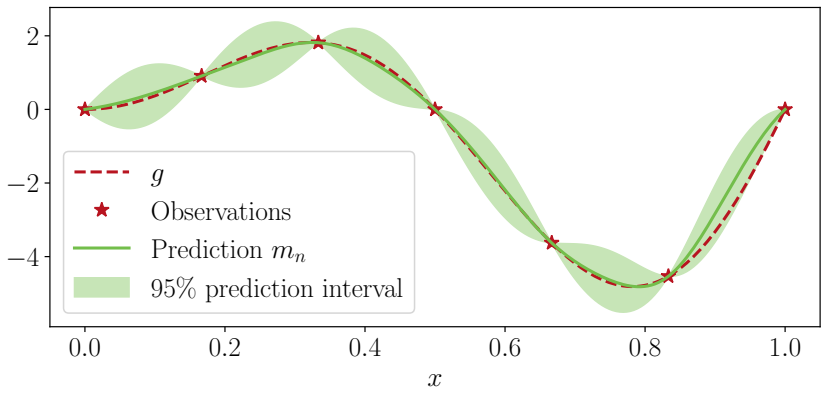


FIGURE 5.1: Illustration of kriging prediction with a simple univariate example with a zero mean Gaussian process and a Gaussian covariance function

prediction interval length is larger in regions where no observation of g has been made and is null at points where there is an observation.

It is also possible to sample a realization $G_n(.;\omega)$ from the distribution of the conditional GP $(G|G^n = g^n)$. Given a set of evaluation points $(t_q)_{1\leq q\leq Q}$, we can derive the Q-dimensional vector mean μ_n and the $Q \times Q$ covariance matrix K_n of the Gaussian vector $(G_n(t_1), \ldots, G_n(t_Q))^T$ thanks to the simple kriging equations:

$$\begin{aligned} \mu_n &= \mu + C_n^T \Sigma_n^{-1}(g^n - m^n) \,, \\ K_n &= K - C_n^T \Sigma_n^{-1} C_n \,. \end{aligned} \tag{5.8}$$

$\mu = (m(t_1), \ldots, m(t_q))$ is the mean vector at evaluation points, $K = (\Sigma(t_i, t_j))_{1\leq i,j\leq Q}$ the covariance matrix of the GP G at the evaluation points, $C_n = (\Sigma(\mathbf{x}_i, t_j))$, $\Sigma_n = (\Sigma(\mathbf{x}_i, \mathbf{x}_j))_{1\leq i,j\leq n}$ and $m^n = (m(\mathbf{x}_1), \ldots, m(\mathbf{x}_n))^T$. Figure 5.2 represents several realizations of the kriging surrogate $(G|G^n = g^n)$ using a regular grid of 200 evaluation points in $[0, 1]$. Numerical implementation requires a Cholesky decomposition of the covariance matrix K_n. Note that the complexity of computing the conditional covariance matrix and sampling the conditional GP is in $O(\max(n^3, Q^3))$. This means that when n (respectively Q) is large (approximately more than 10^3), the computation of the conditional covariance matrix (respectively sampling realizations of the conditional GP) becomes numerically intractable. Fortunately in our applications, the learning size n is usually small. However, the number of evaluation points Q can be very high and requires more advanced methodologies (see e.g. Chilès and Delfiner, 2012, Chapter 7; Le Gratiet, 2013, Section 6.4.3).

There is also another way to obtain the kriging equations 5.6. Indeed, the kriging mean is the Best Linear Unbiased Predictor (BLUP) from the observations (see e.g. Chilès and Delfiner, 2012, Section 3.3.4; Stein, 1999, Section 1.5). The BLUP is obtained by solving the following optimization problem:

$$\lambda_n(\mathbf{x}) = \underset{\lambda(\mathbf{x})\in\mathbb{R}^n}{\operatorname{argmin}} \, \mathbb{E}_G[\left(\lambda(\mathbf{x})^T G^n - G(\mathbf{x})\right)^2] \,. \tag{5.9}$$

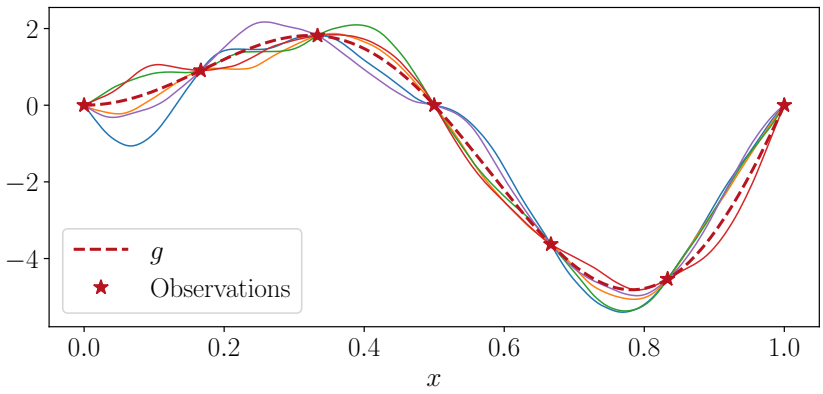


FIGURE 5.2: Realizations of the kriging surrogate using a zero mean Gaussian process and a Gaussian covariance

The BLUP has the same mean and variance function as the kriging surrogate:

$$\begin{aligned} m_n(\mathbf{x}) &= \lambda_n(\mathbf{x})^T g^n \,, \\ s_n^2(\mathbf{x}) &= \mathbb{E}_G[\big(\lambda_n(\mathbf{x})^T G^n - G(\mathbf{x})\big)^2] \,. \end{aligned} \tag{5.10}$$

The BLUP formulation is useful to derive the universal kriging equations, presented in the next subsection.

5.3.2 Universal kriging

Now we consider the case where the mean vector m is unknown with the parametric form $m(\mathbf{x}) = \eta^T u(\mathbf{x})$ defined in Equation 5.3. The BLUP formulation of the kriging surrogate is used to get the kriging equations in this context. We are looking for a predictor of the form:

$$m_n(\mathbf{x}) = \lambda_n(\mathbf{x})^T g^n \,, \tag{5.11}$$

such that $\lambda_n(\mathbf{x})$ is the solution of the problem:

$$\begin{aligned} \lambda_n(\mathbf{x}) = &\underset{\lambda(\mathbf{x})\in\mathbb{R}^n}{\operatorname{argmin}} \mathbb{E}_G[\big(\lambda(\mathbf{x})^T G^n - G(\mathbf{x})\big)^2] \,, \\ &\text{s.t. } U^T\lambda(\mathbf{x}) = u(\mathbf{x}) \,. \end{aligned} \tag{5.12}$$

$U = (u_j(\mathbf{x}_i))_{i,j}$ is the matrix of the basis functions evaluated at observation points. The equality constraint in Equation 5.12 comes from the unbiasedness property of the BLUP. This optimization problem admits a closed form solution obtained using Lagrange multipliers methods (see e.g. Le Gratiet, 2013, Section 1.5.1 for practical details of the computation). We then obtain:

$$\lambda_n(\mathbf{x}) = \Sigma_n^{-1} c_n(\mathbf{x}) + \Sigma_n^{-1} U (U^T \Sigma_n^{-1} U)^{-1} (u(\mathbf{x}) - U\Sigma_n^{-1} c_n(\mathbf{x})) \,, \tag{5.13}$$

the kriging variance is obtained using the following equality under the constraint $U^T\lambda(\mathbf{x}) = u(\mathbf{x})$:

$$\mathbb{E}_G[\big(\lambda(\mathbf{x})^T G^n - G(\mathbf{x})^2\big)] = \Sigma(\mathbf{x},\mathbf{x}) - 2\lambda(\mathbf{x})^T c_n(\mathbf{x}) + \lambda(\mathbf{x})^T \Sigma_n \lambda(\mathbf{x}) \,, \tag{5.14}$$

then by plugging the value $\lambda_n(\mathbf{x})$ in Equation 5.11 (resp. Equation 5.14) we obtain the universal kriging mean (resp. variance):

$$\begin{aligned} m_n(\mathbf{x}) &= u(\mathbf{x})^T\widehat{\eta}_n + c_n(\mathbf{x})^T\Sigma_n^{-1}(g^n - U\widehat{\eta}_n), \\ s_n(\mathbf{x})^2 &= \Sigma(\mathbf{x},\mathbf{x}) - c_n(\mathbf{x})^T\Sigma_n^{-1}c_n(\mathbf{x}) \\ &\quad + (u(\mathbf{x})^T - c_n(\mathbf{x})^T\Sigma_n^{-1})(U^T\Sigma_n^{-1}U)^{-1}(u(\mathbf{x})^T - c_n(\mathbf{x})^T\Sigma_n^{-1})^T \ , \end{aligned} \tag{5.15}$$

where $\widehat{\eta}_n = (U^T\Sigma_n^{-1}U)^{-1}U^T\Sigma_n^{-1}g^n$. The kriging equations in the universal case are similar to the ones in the simple case, the value of η being replaced by its estimation $\widehat{\eta}_n$. Remark also the additional term in the kriging variance, coming from the uncertainty in the estimation of η.

5.3.3 Kriging with noisy observations

In some situations, Gaussian process regression can also be used in the more general context of nonparametric regression, where a noise is added between the function to approximate g and observed data y. Recall the classical equation of nonparametric regression:

$$y(\mathbf{x}) = g(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{5.16}$$

where $g\colon \mathbb{R}^p \mapsto \mathbb{R}$ is coined the regression function. $\varepsilon(\mathbf{x})$ is a stochastic process independent of g such that $\mathbb{E}[\varepsilon(\mathbf{x})] = 0$. This will represent the noise that taints our observations $y(\mathbf{x})$. Throughout this manuscript, we will suppose that $\varepsilon(\mathbf{x}) \sim \mathcal{N}(0, \sigma_\varepsilon(\mathbf{x})^2)$ and $\mathrm{Cov}(\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}')) = \mathbb{1}_{\mathbf{x}=\mathbf{x}'}\sigma_\varepsilon(\mathbf{x})^2$. The Gaussian assumption on ε allows to define the Gaussian process Y such that:

$$Y(\mathbf{x}) = G(\mathbf{x}) + \varepsilon(\mathbf{x}) \ . \tag{5.17}$$

Indeed, ε and G being two Gaussian processes, their sum is also a Gaussian process. The mean functions of Y and G are the same and the covariance function of Y writes:

$$\mathrm{Cov}(Y(\mathbf{x}), Y(\mathbf{x}')) = \Sigma(\mathbf{x},\mathbf{x}') + \sigma_\varepsilon(\mathbf{x})^2\delta(\mathbf{x},\mathbf{x}') \ , \tag{5.18}$$

where

$$\delta(\mathbf{x},\mathbf{x}') = \begin{cases} 0 & \text{if } \mathbf{x} \neq \mathbf{x}', \\ 1 & \text{if } \mathbf{x} = \mathbf{x}'. \end{cases} \tag{5.19}$$

We can then derive the kriging equations for the GP $(G(\mathbf{x})|Y^n = y^n)$:

$$\begin{aligned} m_n(\mathbf{x}) &= m(\mathbf{x}) + c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}(y^n - m^n) \ , \\ s_n^2(\mathbf{x},\mathbf{x}') &= \Sigma(\mathbf{x},\mathbf{x}') - c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}c_n(\mathbf{x}') \ , \end{aligned} \tag{5.20}$$

where $\Delta_n = \mathrm{Diag}(\sigma_\varepsilon(\mathbf{x}_i)^2)_{1\leq i\leq n}$ the diagonal matrix composed of the observations noise variances at the observation points. Remark that the only difference with the simple kriging equation defined in Equation 5.6 is the replacement of the covariance matrix Σ_n by $\Sigma_n + \Delta_n$. This results in the loss of the interpolation property of the kriging mean function. Indeed, the noise tainted the observations results in a non-zero variance $s(\mathbf{x}_i)^2$ at observation points $\mathbf{x}_i$. Moreover, Gaussian process regression in the presence of noisy observations can be addressed in a Bayesian framework. This is the topic of the following subsection.

5.3.4 Bayesian viewpoint of kriging

Gaussian process regression became very popular in the Machine Learning community thanks to the book of Rasmussen and C. Williams, 2005. In this book, GP regression is seen as a Bayesian nonparametric regression framework. Indeed, a lot of classical procedures can be reinterpreted in a Bayesian way, such that penalized maximum likelihood, which can be seen as a maximum a posteriori estimator using a well-chosen prior (e.g. the Laplace prior for lasso regression). As shown in R.B. Gramacy, 2020, Section 5.3.2, since we can sample realizations of a GP without observing data, the GP before observations can be seen as a prior distribution. Moreover, the kriging equations raise the conditional distribution $(G|\mathcal{D}_n)$, which can be interpreted as a posterior distribution. Again referring to R.B. Gramacy, 2020, this interpretation raises three questions:

- On which space is the prior distribution ?
- What is the likelihood pairing the posterior and the prior distribution and how to apply Bayes theorem in this framework ?
- What brings the Bayesian viewpoint to our comprehension of the GP regression framework ?

Given the statistical model defined in Equation 5.17, and the three quantities $(\mathbf{x}_i, G(\mathbf{x}_i), Y(\mathbf{x}_i))$, the $(G(\mathbf{x}_i))$-values can be seen as a latent (i.e. the $G(\mathbf{x}_i)$-values are never observed) Gaussian random field between the observation points $\mathbf{x}_i$ and the observed outputs $Y(\mathbf{x}_i)$. The GP structure on G boils down to put a prior on the latent $G(\mathbf{x}_i)$-values:

$$G^n \sim \mathcal{N}_n(0, \Sigma_n) \, , \tag{5.21}$$

where $\mathcal{N}_n$ defines the n-dimensional multivariate Gaussian distribution. Because no data is observed at this point, we can call this distribution a prior. The likelihood comes from the noise of the observations. The vector of observations $Y^n = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))^T$ indeed follow a multivariate Gaussian distribution due to the noise structure:

$$Y^n | G^n \sim \mathcal{N}_n(G^n, \Delta_n) \, , \tag{5.22}$$

where $\Delta_n = \text{Diag}(\sigma_\varepsilon(\mathbf{x}_i)^2)_{1 \leq i \leq n}$. The prior and the likelihood can now be combined using Bayes theorem to obtain the posterior distribution:

$$\mathbb{P}(G^n | Y^n) = \frac{\mathbb{P}(Y^n | G^n)\mathbb{P}(G^n)}{\mathbb{P}(Y^n)} \, . \tag{5.23}$$

The denominator $\mathbb{P}(Y^n)$ is often called marginal likelihood or evidence. It can be obtained by integrating over the latent probability distribution of the GP G:

$$\mathbb{P}(Y^n) = \int \mathbb{P}(Y^n | G^n) d\mathbb{P}(G^n) \, . \tag{5.24}$$

The Bayesian viewpoint sheds a new light on GP regression: The GP G put a latent structure on the regression function g, which can be seen as a prior distribution on g. This raises the question of the choice of model for G, which boils down to the choice of the covariance function $\Sigma(.,.)$. The effect of $\Sigma(.,.)$ on the GP is thus of keen interest, both on the theoretical and applied point of view. Answering this question is the goal of the following section.

5.4 Covariance functions and model selection

This section aims at studying the choice of covariance function Σ to model a good prior GP on the regression function g. First, we will present theoretical results concerning the link between the covariance function of the GP and peculiar Hilbert spaces of functions called Reproducing Kernel Hilbert Spaces (RKHS). These results will give us insights about the effect of the covariance function on the regularity of the realizations of the GP. Moreover, we will present a more practical approach for choosing the covariance function. We will restrict to parametric families of covariance functions, hence the choice of the covariance function boils down to a choice of hyperparameters. We will introduce a Maximum Likelihood Estimator (MLE) of the hyperparameters of the covariance function, and then we will present a fully Bayesian approach for hyperparameters tuning.

5.4.1 Covariance functions and Gaussian process regularity

In this subsection, the main properties of covariance functions will be detailed, theoretical results about Reproducing Kernel Hilbert Spaces (RKHS) will be recalled in order to give insights about the regularity of the GP realizations. In this part, the covariance function will be denoted $k(.,.)$.

First, we will recall the necessary and sufficient conditions for a function to be a covariance function:

Definition 2. *Let $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$. The mapping k is commonly called a kernel. Two properties on k are necessary and sufficient for k to be the covariance function of a GP:*

- *k is symmetric, i.e. $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$ for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^p$.*
- *k is positive semi-definite:*

$$\forall m \in \mathbb{N},\ \forall \mathbf{x}_1, \dots, \mathbf{x}_m \in \mathcal{X}, \forall a_1, \dots, a_m \in \mathbb{R},\ \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \ .$$

Given a covariance kernel k, in which functional space the GP realizations with covariance k lies ? In order to answer this question, we need first to provide a spectral decomposition of k. This decomposition comes from the eigenvalues and eigenfunctions decomposition of k. The following theorem is an extension of the Mercer's theorem (Mercer, 1909) with a probability measure μ with support on $\mathcal{X}$ and a continuous positive kernel satisfying the property $\sup_{\mathbf{x} \in \mathcal{X}} k(\mathbf{x}, \mathbf{x}) < +\infty$ with $\mathcal{X}$ a non-empty set of $\mathbb{R}^p$.

Theorem 4 (Mercer's theorem)**.** *Let us consider a continuous symmetric positive semi-definite kernel $k(\mathbf{x}, \mathbf{x}')$, $\mathbf{x}, \mathbf{x}' \in \mathcal{X} \subset \mathbb{R}^p$ - such that $\sup_{\mathbf{x} \in \mathcal{X}} k(\mathbf{x}, \mathbf{x}) < +\infty$ and $\mathcal{X}$ is a nonempty open set - and a probability measure μ on $\mathcal{X}$ such that $\mu(U) > 0$ for any nonempty open subset U of $\mathcal{X}$. The kernel $k(\mathbf{x}, \mathbf{x}')$ can be decomposed as follows:*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{q \geq 0} \lambda_q \phi_q(\mathbf{x}) \phi_q(\mathbf{x}') \ ,$$

where $\phi_q \in L^2_\mu(\mathcal{X})$ are the eigenfunctions of the trace class operator

$$(T_k f)(\mathbf{x}) = \int k(\mathbf{x}, \mathbf{u}) f(\mathbf{u}) d\mu(\mathbf{u}) \ ,$$

and $(\lambda_q)_{q\geq 0}$ is the nonnegative sequence of eigenvalues sorted in decreasing order. Furthermore, $(\phi_q)_{q\geq 0}$ is an orthonormal basis of $L^2_\mu(\mathcal{X})$ and ϕ_q is continuous for all q such that $\lambda_q \neq 0$.

The regularity of the Gaussian process is closely related to the rate of convergence to 0 of the eigenvalues sequence $(\lambda_q)_{q\geq 0}$. We can derive from Theorem 4 some interesting insights about the decomposition that will be of peculiar interest:

- By definition the eigenfunction ϕ_q satisfies the following equality:

$$\lambda_q \phi_q(\mathbf{x}) = \int k(\mathbf{x},\mathbf{u})\phi_q(\mathbf{u})d\mu(\mathbf{u}) \ .$$

- The orthonormal property of the eigenfunctions $(\phi_q)_{q\geq 0}$ implies that:

$$\int \phi_q(\mathbf{x})\phi_p(\mathbf{x})d\mu(\mathbf{x}) = \mathbb{1}_{p=q} \ .$$

- For covariance kernels such that $\forall \mathbf{x} \in \mathcal{X},\ k(\mathbf{x},\mathbf{x}) = \sigma^2$, we have:

$$\sigma^2 = \sum_{q\geq 0} \lambda_q \phi_q(\mathbf{x})^2 = \sum_{q\geq 0} \lambda_q \ .$$

Thanks to the Mercers's decomposition theorem, we can now introduce a special Hilbert space of functions called Reproducing Kernel Hilbert Space (RKHS). This type of Hilbert space is often used for kernel methods in machine learning. However, it can be introduced from several perspectives as explained in Aubin-Frankowski, 2021, Section 1.2: One approach focuses closely on functional analysis (see e.g. Berlinet and Thomas-Agnan, 2011; Saitoh and Sawano, 2016) where the input set $\mathcal{X}$ is a subset of $\mathbb{R}^p$, whereas another approach focuses more on defining kernel functions on complex input spaces such that graphs or probability measures and extending classical linear procedures to nonlinear settings (see Schölkopf, B. Smola, et al., 2002). In this section, we will present the functional analysis definition of the RKHS and its links with the GP sample paths regularity. The RKHS can be defined by its Mercer representation (Steinwart and Christmann, 2008, Theorem 4.51; Kanagawa et al., 2018):

Theorem 5 (Mercer representation). *Let $\mathcal{X} \subset \mathbb{R}^p$ be a nonempty open set, $k\colon \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$ be a continuous symmetric positive semi-definite kernel, μ be a probability measure on $\mathcal{X}$ and $(\phi_q, \lambda_q)_{q\geq 0}$ be defined as in Theorem 4. Then the RKHS $\mathcal{H}_k$ of k is given by:*

$$\mathcal{H}_k = \left\{ g := \sum_{q\geq 0} \alpha_q \sqrt{\lambda_q}\phi_q \ :\ \|g\|^2_{\mathcal{H}_k} := \sum_{q\geq 0} \alpha_q^2 < +\infty \right\} \ , \tag{5.25}$$

with the associated scalar product for $g, h \in \mathcal{H}_k$ defined by

$$\langle g, h\rangle_{\mathcal{H}_k} := \sum_{q\geq 0} \alpha_q \beta_q \ ,$$

for $g = \sum_{q\geq 0} \alpha_q \sqrt{\lambda_q}\phi_q \ , h = \sum_{q\geq 0} \beta_q \sqrt{\lambda_q}\phi_q$. It boils down to say that $(\sqrt{\lambda_q}\phi_q)_{q\geq 0}$ forms an orthonormal basis of $\mathcal{H}_k$.

A first useful result is that the kriging mean function m_n belongs to the RKHS $\mathcal{H}_k$ when the covariance function is k (see Kanagawa et al., 2018, Remark 3.8). However, the GP sample paths

do not belong to the RKHS $\mathcal{H}_k$. In order to provide the GP sample paths space, we have to define the powers of RKHSs and kernels (Kanagawa et al., 2018, Section 4.3):

Definition 3 (Powers of RKHS and kernels). *Under the same assumptions of Theorem 5, let $0 < \theta \leq 1$ be a constant and assume that $\sum_{q\geq 0} \lambda_q^\theta \phi_q(\mathbf{x})^2 < +\infty$ holds for all $\mathbf{x} \in \mathcal{X}$, where $(\lambda_q, \phi_q)_{q\geq 0}$ is the eigensystem coming from the Mercer's decomposition. Then the power of RKHS $\mathcal{H}_k$ is defined as:*

$$\mathcal{H}_k^\theta = \left\{ g := \sum_{q\geq 0} \alpha_q \lambda_q^{\theta/2} \phi_q \ : \ \|g\|_{\mathcal{H}_k}^2 := \sum_{q\geq 0} \alpha_q^2 < +\infty \right\} , \tag{5.26}$$

with the scalar product for $g, h \in \mathcal{H}_k^\theta$ is given by

$$\langle g, h \rangle_{\mathcal{H}_k^\theta} = \sum_{q\geq 0} \alpha_q \beta_q ,$$

for $g = \sum_{q\geq 0} \alpha_q \lambda_q^{\theta/2} \phi_q$, $h = \sum_{q\geq 0} \beta_q \lambda_q^{\theta/2} \phi_q$. The θ-th power of the kernel k is defined by

$$k^\theta(\mathbf{x}, \mathbf{x}') := \sum_{q\geq 0} \lambda_q^\theta \phi_q(\mathbf{x}) \phi_q(\mathbf{x}') \ , \ \mathbf{x}, \mathbf{x}' \in \mathcal{X} . \tag{5.27}$$

The power of a RKHS can be viewed as an intermediate function space between $L_\mu^2(\mathcal{X})$ and $\mathcal{H}_k$ with the constant $0 < \theta \leq 1$ determining the proximity between $\mathcal{H}_k^\theta$ and $\mathcal{H}_k$ (Steinwart and Scovel, 2012, Theorem 4.6). If $\theta = 1$, we have $\mathcal{H}_k^\theta = \mathcal{H}_k$, and $\mathcal{H}_k^\theta$ approaches $L_\mu^2(\mathcal{X})$ as $\theta \to 0^+$. The spaces $\{\mathcal{H}_k^\theta / 0 < \theta \leq 1\}$ are in fact nested, i.e. for all $0 < \theta < \theta' \leq 1$:

$$\mathcal{H}_k = \mathcal{H}_k^1 \subset \mathcal{H}_k^{\theta'} \subset \mathcal{H}_k^\theta \subset L_\mu^2(\mathcal{X}) ,$$

$\mathcal{H}_k^\theta$ gets larger as θ decreases, which means that $\mathcal{H}_k^\theta$ contains less smooth functions than $\mathcal{H}_k$.

The power of RKHS are in fact very useful to characterize the function space of the GP-sample path as shown in Kanagawa et al., 2018, Theorem 4.12:

Theorem 6. *Under the same assumptions as Theorem 5, let $0 < \theta \leq 1$ be a constant and assume that $\sum_{q\geq 0} \lambda_q^\theta \phi_q(\mathbf{x})^2 < +\infty$ holds for all $\mathbf{x} \in \mathcal{X}$, where $(\lambda_q, \phi_q)_{q\geq 0}$ is the eigensystem coming from the Mercer's decomposition. Consider $G \sim \mathcal{GP}(0, k)$. Then the following statements are equivalent:*

1. $\sum_{q\geq 0} \lambda_q^{1-\theta} < +\infty$.
2. *There exists a version $\tilde{G}$ of G such that $\tilde{G} \in \mathcal{H}_k^\theta$ with probability 1.*

Theorem 6 characterizes the GP-sample path function spaces as the powers of the RKHS $\mathcal{H}_k^\theta$ with θ satisfying the condition $\sum_{q\geq 0} \lambda_q^{1-\theta} < \infty$. The GP-sample paths function space thus depends of the rate of convergence to 0 of the eigenvalues $(\lambda_q)_{q\geq 0}$, the faster the rate of convergence the smoother is the GP-sample path space. This theorem demonstrates also the insight that the GP-sample path function space is always less smooth that the RKHS generated by covariance kernel k.

5.4.2 Karhunen-Loeve decomposition of a Gaussian process and Nyström method

On a more practical viewpoint, we can derive from the Mercer's decomposition a representation of realization of a Gaussian process and thus a way to simulate numerically Gaussian process

realizations. This method will be used extensively for kriging-based sensitivity analysis presented in Chapter 6.

Theorem 7 (Karhunen-Loeve decomposition)**.** *Let us consider a zero-mean Gaussian process G with a covariance function k and admitting the following Mercer's decomposition:*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{q \geq 0} \lambda_q \phi_q(\mathbf{x}) \phi_q(\mathbf{x}') \ .$$

Then, G can be represented in the following form:

$$G(\mathbf{x}) = \sum_{q \geq 0} \sqrt{\lambda_q} \phi_q(\mathbf{x}) \varepsilon_q \ ,$$

where $(\varepsilon_q)_{q \geq 0}$ are independent and identically distributed real valued random variables such that $\varepsilon_1 \sim \mathcal{N}(0,1)$.

Theorem 7 is particularly interesting for our applications. Indeed, if we know the eigensystem $(\lambda_q, \phi_q)_{q \geq 0}$ of the covariance function, it is possible to generate realization of the Gaussian vector $(G(\mathbf{x}_1), \ldots, G(\mathbf{x}_Q))^T$ with a computational cost in $O(Q)$, on the contrary of direct sampling with Cholesky decomposition method which is in $O(Q^3)$. However, we can only generate approximations of the GP realizations because the series of the Karhunen-Loeve decomposition has to be truncated for numerical implementation. In practice, the eigensystem $(\lambda_q, \phi_q)_{q \geq 0}$ is also unknown, the Nyström method (Rasmussen and C. Williams, 2005, Section 4.2.3) aims at numerically approximating the eigensystem and thus the Karhunen-Loeve decomposition. According to Mercer's decomposition, we have to solve the eigenproblem $\forall q \in \mathbb{N}$:

$$\lambda_q \phi_q(\mathbf{x}) = \int_{\mathcal{X}} k(\mathbf{x}, \mathbf{u}) \phi_q(\mathbf{u}) d\mu(\mathbf{u}) \ . \tag{5.28}$$

We can approximate the right-hand side of the equation by Monte-Carlo:

$$\int_{\mathcal{X}} k(\mathbf{x}, \mathbf{u}) \phi_q(\mathbf{u}) d\mu(\mathbf{u}) \approx \frac{1}{N} \sum_{i=1}^{N} k(\mathbf{x}, \mathbf{u}_i) \phi_q(\mathbf{u}_i) \ , \tag{5.29}$$

where the $\mathbf{u}_i$-values are drawn from the probability distribution μ. We then consider the eigenfunction ϕ_q at point $(\mathbf{u}_i)_{1 \leq i \leq N}$ in order to raise the following eigenproblem:

$$\lambda_q \Phi_q = \frac{K_N}{N} \Phi_q \ , \tag{5.30}$$

where $\Phi_q^T = (\phi_q(\mathbf{u}_1), \ldots, \phi_q(\mathbf{u}_N))$, $(K_N)_{i,j} = k(\mathbf{u}_i, \mathbf{u}_j)$. The qth eigenfunction is then approximated by:

$$\phi_q(\mathbf{x}) = \frac{1}{N \lambda_q} \mathbf{k}(\mathbf{x})^T \Phi_q \ , \tag{5.31}$$

where $\mathbf{k}(\mathbf{x})^T = (k(\mathbf{x}, \mathbf{u}_1), \ldots, k(\mathbf{x}, \mathbf{u}_N))$. Given a point $\mathbf{x}$, we can sample $G(\mathbf{x})$ by considering the truncated Karhunen-Loeve decomposition of G:

$$G(\mathbf{x}) \approx \frac{1}{N} \sum_{q \leq T} \frac{\mathbf{k}(\mathbf{x})^T \Phi_q}{\sqrt{\lambda_q}} \varepsilon_q \ . \tag{5.32}$$

The only remaining difficulty is the choice of the truncation T, corresponding to the number of eigenfunctions to consider in the Karhunen-Loeve decomposition of G. Denote by G_{KL} the truncated Karhunen-Loeve decomposition:

$$G_{\mathrm{KL}}(\mathbf{x}) = \sum_{q=0}^{T} \sqrt{\lambda_q}\phi_q(\mathbf{x})\varepsilon_q \ , \tag{5.33}$$

We will now discuss the effect of the truncation T on the integrated mean square error (IMSE):

$$\mathrm{IMSE}(G, G_{\mathrm{KL}}) = \int_{\mathcal{X}} \mathbb{E}\left[\left(G(\mathbf{x}) - G_{\mathrm{KL}}(\mathbf{x})\right)^2\right] d\mu(\mathbf{x}) \ . \tag{5.34}$$

It is possible to control $\mathrm{IMSE}(G, G_{\mathrm{KL}})$ using the eigenvalues $(\lambda_q)_{q\geq 0}$ using the following lemma:

Lemma 1. *Under the assumptions of Theorem 4, we have the following equation:*

$$\mathit{IMSE}(G, G_{\mathrm{KL}}) = \sum_{q\geq T+1} \lambda_q \ .$$

Proof. Denote by $G_k(\mathbf{x}) = \sum_{q=0}^{k} \sqrt{\lambda_q}\phi_q(\mathbf{x})$, thus

$$\begin{aligned} \mathrm{IMSE}(G_m, G_k) &= \int_{\mathcal{X}} \mathbb{E}\left[\left(\sum_{q=k+1}^{m} \sqrt{\lambda_q}\varepsilon_q\phi_q(\mathbf{x})\right)^2\right] d\mu(\mathbf{x}) \\ &= \sum_{k+1\leq q_1,q_2\leq m} \sqrt{\lambda_{q_1}\lambda_{q_2}}\mathbb{E}[\varepsilon_{q_1}\varepsilon_{q_2}] \int_{\mathcal{X}} \phi_{q_1}(\mathbf{x})\phi_{q_2}(\mathbf{x}) d\mu(\mathbf{x}) \ , \end{aligned} \tag{5.35}$$

The random variables $(\varepsilon_q)_{q\geq 0}$ being standard i.i.d. Gaussian variables, we have $\mathbb{E}[\varepsilon_{q_1}\varepsilon_{q_2}] = \mathbb{1}_{q_1=q_2}$. The eigenfunctions $(\phi_q)_{q\geq 0}$ being an orthonormal basis of $L^2(\mu)$, we have $\int_{\mathcal{X}} \phi_{q_1}(\mathbf{x})\phi_{q_2}(\mathbf{x})d\mu(\mathbf{x}) = \mathbb{1}_{q_1=q_2}$ as well. We then obtain the following expression:

$$\mathrm{IMSE}(G_m, G_k) = \sum_{q=k+1}^{m} \lambda_q \ .$$

Since $\sum_{q\geq 0} \lambda_q < \infty$, $(G_k)_{k\leq 1}$ is a Cauchy sequence. Due to the completeness of $L^2(\Omega\times\mathcal{X}, d\mathbb{P}\otimes d\mu)$, we have

$$\lim_{m\to\infty} \mathrm{IMSE}(G_m, G_k) = \mathrm{IMSE}(G, G_k) = \sum_{q\geq k+1} \lambda_q$$

□

For covariance kernels such that $\forall \mathbf{x} \in \mathcal{X}, k(\mathbf{x}, \mathbf{x}) = \sigma^2 = \sum_{q\geq 0} \lambda_q$, it is thus possible to choose T in order to control the IMSE of the truncated Karhunen-Loeve decomposition of a GP. However, it matters also to verify how close is the covariance function of the truncated Gaussian process G_{KL} to the original one. It boils down to study the Mean Covariance Difference (MCD):

$$\mathrm{MCD}(G, G_{\mathrm{KL}}) = \int_{\mathcal{X}\times\mathcal{X}} \mathbb{E}[|G(\mathbf{x})G(\mathbf{u}) - G_{\mathrm{KL}}(\mathbf{x})G_{\mathrm{KL}}(\mathbf{u})|] d\mu(\mathbf{x})d\mu(\mathbf{u}) \ . \tag{5.36}$$

We can bound the MCD using the IMSE thanks to the following lemma:

Lemma 2. *Under the assumptions of Theorem 4, considering a GP with a covariance kernel such that* $\forall \mathbf{x} \in \mathcal{X}, k(\mathbf{x}, \mathbf{x}) = \sigma^2$. *We have the following upper bound:*

$$MCD(G, G_{\mathrm{KL}}) \leq 2\sigma\sqrt{IMSE(G, G_{\mathrm{KL}})}$$

Proof. Using the following

$$\begin{aligned}
\mathrm{MCD}(G, G_{\mathrm{KL}}) &= \int_{\mathcal{X}\times\mathcal{X}} \mathbb{E}[|G(\mathbf{x})G(\mathbf{u}) - G(\mathbf{x})G_{KL}(\mathbf{u}) + G(\mathbf{x})G_{KL}(\mathbf{u}) - G_{\mathrm{KL}}(\mathbf{x})G_{\mathrm{KL}}(\mathbf{u})|]d\mu(\mathbf{x})d\mu(\mathbf{u}) \\
&\leq \int_{\mathcal{X}\times\mathcal{X}} \mathbb{E}[|G(\mathbf{x})||G(\mathbf{u}) - G_{KL}(\mathbf{u})|]d\mu(\mathbf{x})d\mu(\mathbf{u}) \\
&+ \int_{\mathcal{X}\times\mathcal{X}} \mathbb{E}[|G(\mathbf{u})||G(\mathbf{x}) - G_{KL}(\mathbf{x})|]d\mu(\mathbf{x})d\mu(\mathbf{u})
\end{aligned} \tag{5.37}$$

We can use the Cauchy-Schwarz inequality to obtain

$$\int_{\mathcal{X}\times\mathcal{X}} \mathbb{E}[|G(\mathbf{x})||G(\mathbf{u}) - G_{KL}(\mathbf{u})|]d\mu(\mathbf{x})d\mu(\mathbf{u}) \leq \left(\int_{\mathcal{X}} \mathbb{E}[G(\mathbf{x})^2]d\mu(\mathbf{x})\right)^{1/2} \mathrm{IMSE}(G, G_{KL})^{1/2} ,$$

from the last expression we can derive the upper bound on $\mathrm{MCD}(G, G_{\mathrm{KL}})$. □

It is thus possible to control the MCD between the GP and its truncated Karhunen-Loeve decomposition using the IMSE.

5.4.3 Classical parametric models of covariance function

In practice, the covariance function is chosen among a parametric family of functions which are known to be admissible covariance functions. Parametrized covariance functions have the typical form:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 r_\rho(\mathbf{x}, \mathbf{x}') ,$$

σ^2 controls the variance of the GP, r_ρ is a correlation function parametrized by the vector $\rho = (\rho_1, \ldots, \rho_p)^T \in \mathbb{R}^p$ with $\rho_j > 0$ for all j. The parameter ρ_j is called the characteristic length in the j-th direction. Indeed, this parameter controls for $\mathbf{x}, \mathbf{x}'$ the level of correlation between $G(\mathbf{x})$ and $G(\mathbf{x}')$. Figure 5.3 gives an example of realizations of a GP with different values of characteristic lengths. A common way to define a covariance function in high-dimensional space is through tensorized 1-dimensional covariance function. Moreover, all the covariance functions considered in this manuscript will be stationary, i.e. they only depend on $\mathbf{h} = \mathbf{x} - \mathbf{x}'$. Assume that $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ such that $\mathbf{x} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)})$, the covariance functions considered in this manuscript have the following form:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{j=1}^{p} w_{\rho_j}(h^{(j)}) , \tag{5.38}$$

where w_{ρ_j} is a 1-dimensional correlation function. Examples of stationary correlation functions from Rasmussen and C. Williams, 2005 are given in Table 5.1. In this manuscript, we will put a particular focus on the Matérn covariance model. This covariance model includes a parameter controlling the degree of differentiability of the GP and includes the exponential model for a specific parameter value and Gaussian model as a limiting case. The function Γ is the Euler gamma function and K_ν is the modified Bessel function of the second kind of order $\nu \in (0, +\infty)$

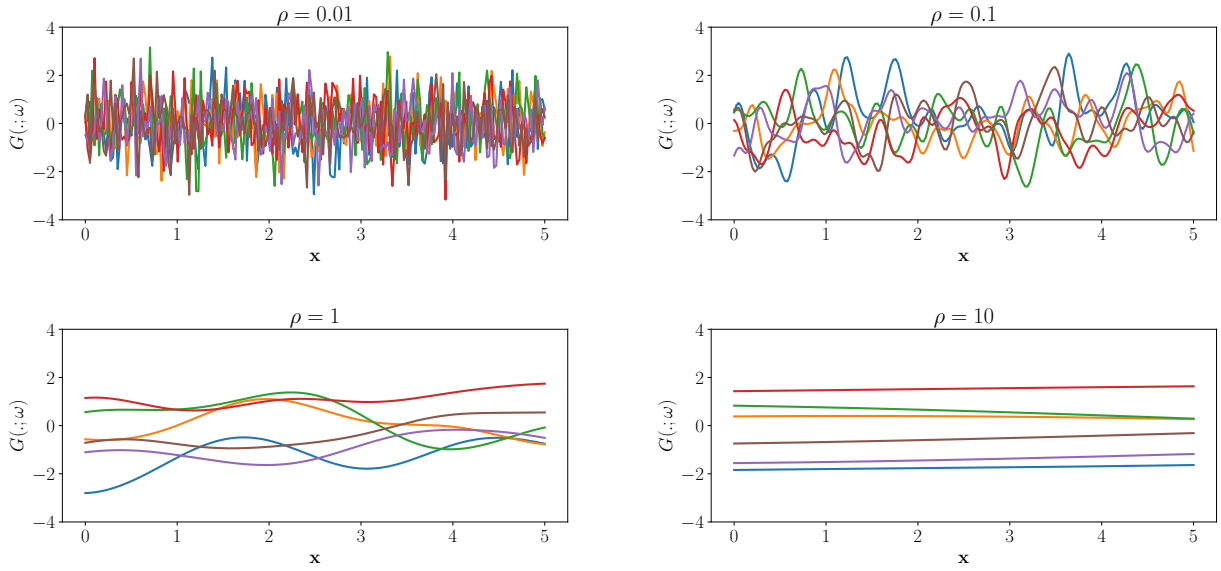


FIGURE 5.3: Realizations of a zero-mean Gaussian process with covariance function $\exp\left(\frac{-(\mathbf{x}-\mathbf{x}')^2}{2\rho^2}\right)$ for different values of ρ.

(Abramowitz and Stegun, 1965). Remark that Matérn correlation functions have simpler expressions when $\nu = p + 1/2\ , p \in \mathbb{N}$ as they can be written as a product of an exponential and a polynomial function of order p. When $\nu = 1/2$ the Matérn correlation function corresponds to an exponential correlation and when $\nu \to +\infty$, the Matérn correlation converges to the Gaussian correlation function.

TABLE 5.1: Examples of one-dimensional correlation functions

Name	*Formula*
Matérn $\nu > 0$	$w_\rho(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\lvert h\rvert}{\rho}\right)^\nu K_\nu \left(\frac{\sqrt{2\nu}\lvert h\rvert}{\rho}\right)$
Gaussian (Matérn $\nu \mapsto +\infty$)	$w_\rho(h) = \exp\left(-\frac{h^2}{2\rho^2}\right)$
Matérn $\nu = 5/2$	$w_\rho(h) = \left(1 + \frac{\sqrt{5}\lvert h\rvert}{\rho} + \frac{5h^2}{3\rho^2}\right) \exp\left(\frac{-\sqrt{5}\lvert h\rvert}{\rho}\right)$
Matérn $\nu = 3/2$	$w_\rho(h) = \left(1 + \frac{\sqrt{3}\lvert h\rvert}{\rho}\right) \exp\left(\frac{-\sqrt{3}\lvert h\rvert}{\rho}\right)$
Exponential (Matérn $\nu = 1/2$)	$w_\rho(h) = \exp\left(-\frac{\lvert h\rvert}{\rho}\right)$

Another interesting property is the representation of the Matérn covariance through a stochastic partial differential equation (SPDE) as highlighted in Lindgren, Rue, and Lindström, 2011. Indeed, a Gaussian process $y(\mathbf{x})\ , \mathbf{x} \in \mathbb{R}^p$ with the Matérn covariance is the only stationary solution of the following linear fractional SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2} y(\mathbf{x}) = \mathcal{W}(\mathbf{x})\ , \alpha = \nu + p/2\ , \mathbf{x} \in \mathbb{R}^p \tag{5.39}$$

with the Laplace operator $\Delta y = \sum_{j=1}^{p} \partial^2 y / \partial^2 \mathbf{x}^{(j)}$, a standard Gaussian white noise process $\mathcal{W}$ with unit variance. Depending on ν and p, the operator $(\kappa^2 - \Delta)^{\alpha/2}$ is fractionary and must be defined appropriately. The characteristic length ρ is empirically linked to ν and κ through the relation $\rho \approx \sqrt{8\nu}/\kappa^2$. Actually this representation is very useful in practice in low dimensional cases. Indeed, it is possible to derive a weak solution of Equation 5.39 and compute a finite element approximation of y. This allows sampling of approximations of the GP y while remaining computationally tractable.

Moreover, the parameter ν controls the regularity of the process through the parameter α. Indeed, the realizations of a GP with Matérn covariance functions of parameter ν are $\lceil \nu \rceil - 1$ mean square differentiable (Stein, 1999, Section 2.7) where $\lceil \nu \rceil$ is the smallest integer greater than or equal to ν. In practice, a choice of $\nu = 3/2$ or $\nu = 5/2$ is commonly used for GP regression, as it represents a good trade off between the regularity of the realizations and the proximity to $L^2_\mu(\mathcal{X})$ (i.e. in terms of set inclusion) of the RKHS power space generated by the GP realizations. Figure 5.4 shows GP realizations with Matérn covariance model with different values of ν. When $\nu = 1/2$ the realizations are not differentiable while for $\nu \to +\infty$, corresponding to the Gaussian covariance model, the realizations are $\mathcal{C}^\infty$-differentiable.

In practical applications of GP regression, the different hyperparameters of the covariance model are not known. To address this problem several approaches are possible: Two methods are proposed in this manuscript. First, a plug-in approach consisting into replacing the unknown hyperparameters by a maximum likelihood estimator (MLE) using the training data. Second, a full Bayesian approach is presented, by considering prior distributions on the covariance model hyperparameters. Other methods involving cross-validation techniques are proposed in the literature (see Bachoc, 2013).

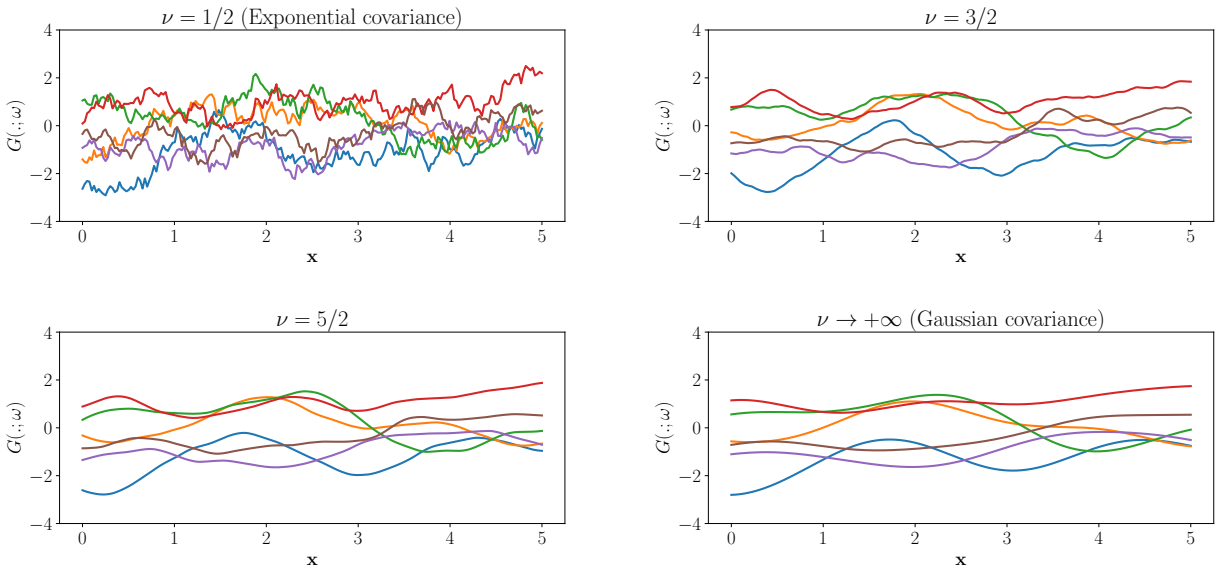


FIGURE 5.4: Realizations of a zero-mean and unit variance Gaussian process with Matérn covariance function with different values of ν. The characteristic length is also unit.

5.4.4 Maximum likelihood estimation of the GP hyperparameters

In this section, we will develop maximum likelihood estimates of GP parameters. Assume that we want to perform simple kriging with a GP $G \sim \mathcal{GP}(0, \Sigma_\theta)$ where $(\Sigma_\theta)_{\theta\in\Theta}$ is a family of admissible covariance functions and θ is unknown and has to be estimated. Given a set of training data $G^n = (G(\mathbf{x}_1), \dots, G(\mathbf{x}_n))^T$, the likelihood considered is the joint density of the random vector G^n (it is also coined marginal likelihood in the GP regression literature). Denote by $\Sigma_n(\theta)$ the covariance matrix of G^n as a function of θ. The log likelihood thus writes:

$$\ell_{\text{simple}}(\theta; G^n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det\left(\Sigma_n(\theta)\right) - \frac{1}{2}(G^n)^T\Sigma_n(\theta)^{-1}G^n \,. \tag{5.40}$$

As pointed out in Rasmussen and C. Williams, 2005, Section 5.4, the three terms in the likelihood can be interpreted: The only term involving the observations $-(G^n)^T\Sigma_n(\theta)^{-1}G^n/2$ corresponds to a data fitting term; $-\log\det\left(\Sigma_n(\theta)\right)/2$ corresponds to a complexity penalty as it only involves the covariance matrix and $n\log(2\pi)/2$ is a normalization constant. The maximum likelihood estimator (MLE) $\widehat{\theta}_n$ thus writes:

$$\widehat{\theta}_n^{\text{MLE}} = \operatorname*{argmax}_{\theta\in\Theta} \ell_{\text{simple}}(\theta; G^n) \,. \tag{5.41}$$

Consider now the case of universal kriging $G \sim \mathcal{GP}(m, \Sigma_\theta)$ where $m(\mathbf{x}) = \eta^T u(\mathbf{x})$. Given the training data G^n, we have to estimate the parameters (η, θ) of both the mean and covariance functions. Consider that $\Sigma_\theta = \sigma^2 W_\rho$ and $\theta = (\sigma, \rho)$. Denote by $W_n(\rho)$ the correlation matrix of the observations and $U = (u_j(\mathbf{x}_i))_{i,j}$. The log likelihood in this case writes:

$$\ell_{\text{universal}}(\theta, \eta; G^n) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2}\log\det\left(W_n(\rho)\right) - \frac{1}{2}(G^n - U\eta)^T W_n(\rho)^{-1}(G^n - U\eta) \,. \tag{5.42}$$

Conditionally to σ and ρ, the MLE of η corresponds to the generalized least squares estimate:

$$\widehat{\eta}_n = (U^T W_n(\rho)^{-1}U)^{-1}UW_n(\rho)^{-1}G^n \,. \tag{5.43}$$

We can plug in the value of $\widehat{\eta}_n$ in the log likelihood (5.42)

$$\widehat{\sigma}_n^2 = \frac{(G^n - U\widehat{\eta}_n)^T W_n(\rho)^{-1}(G^n - U\widehat{\eta}_n)}{n} \,. \tag{5.44}$$

The hyperparameters of the correlation model ρ are found by maximizing the marginal log likelihood:

$$\ell_{\text{marg}}(\rho; G^n) = -\frac{n}{2}\log(2\pi) - n\log(\widehat{\sigma}_n) - \log\det\left(W_n(\rho)\right) \,. \tag{5.45}$$

Consider the case of simple kriging with heteroskedastic noise $Y \sim \mathcal{GP}(0, \Sigma_\theta + \varphi_\beta^2\delta)$ where $\delta(\mathbf{x}, \mathbf{x}') = \mathbb{1}_{\mathbf{x}=\mathbf{x}'}$ and $\mathbf{x} \to \varphi_\beta(\mathbf{x})$ is the heteroskedastic standard deviation parametrized by β. In this case, it is not possible to derive a closed form for (θ, β). Denote $\Delta_n(\beta) = (\text{Diag}(\varphi_\beta(\mathbf{x}_i)^2))_{1\le i\le n}$. The log likelihood writes:

$$\ell_{\text{noisy}}(\theta, \beta; G^n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det\left(\Sigma_n(\theta) + \Delta_n(\beta)\right) - \frac{1}{2}(G^n)^T(\Sigma_n(\theta) + \Delta_n(\beta))^{-1}G^n \,. \tag{5.46}$$

The values of θ and β are obtained by maximizing the log likelihood ℓ_{noisy}. Note that the case of homoskedastic noise corresponds to $\phi_\beta(\mathbf{x}) = \beta$ for $\beta \in (0, +\infty)$.

In practice, numerical optimization of the log likelihood has to be carefully carried out to avoid

numerical errors and instabilities. As pointed out in Rasmussen and C. Williams, 2005, Section 2.2, the log likelihood should be computed using Cholesky decomposition (see Section A.1) to perform numerically stable and fast matrix inversion of the covariance matrix. A simple practical implementation is proposed in Algorithm 2 in the case of simple kriging. To avoid ill-conditioned covariance matrix, a common practice is to add a small Gaussian noise to the data, boiling down to add $\tau^2 I_n$ to the covariance matrix. This is called *nugget regularization*.

Algorithm 2 Numerical computation of the log likelihood

inputs: $\Sigma_n(\theta)$ (covariance matrix), G^n (observations)

1. $L := \text{Cholesky}(\Sigma_n(\theta))$
2. $\alpha = L^T \backslash (L \backslash G^n)$
3. $\ell(\theta; G^n) = -\alpha^T G^n/2 - \sum_i L_{ii} - n \log 2\pi/2$
4. **return:** $\ell(\theta; G^n)$ (log likelihood)

5.4.5 Bayesian estimation of the GP hyperparameters

As pointed out in Section 5.3.4, the GP regression framework can be seen in a Bayesian way. Thus, it is natural to consider a Bayesian framework for estimating the different hyperparameters of the GP. In this part, the so-called "fully Bayesian" approach will be presented, this consists into considering all the parameters of the mean function (the vector η) and the hyperparameters of the covariance function (the variance σ^2 and the characteristic lengths ρ) as random variables. The prior distribution of the different parameters are built in a hierarchical way (see e.g. Santner, B. Williams, and Notz, 2003, Section 4.1.4). This is a common methodology in Bayesian modeling. To simplify the notations, all the probability measures involved in this part are absolutely continuous w.r.t. the Lebesgue measure with a probability density function denoted by π.

The first major assumption is that the mean and variance parameters, η and σ^2 respectively, are independent of the correlation parameters ρ. Hence,

$$\pi(\eta, \sigma^2, \rho) = \pi(\eta, \sigma^2)\pi(\rho) \ .$$

The prior on η and σ^2 is decomposed into

$$\pi(\eta, \sigma^2) = \pi(\eta|\sigma^2)\pi(\sigma^2) \ .$$

From the Bayes rule we can derive the following equation

$$\pi(\eta, \sigma^2, \rho|G^n) \propto \pi(G^n|\eta, \sigma^2, \rho)\pi(\eta|\sigma^2)\pi(\sigma^2)\pi(\rho) \ , \tag{5.47}$$

remark that $\pi(G^n|\eta, \sigma^2, \rho) = \exp\left(\ell_{\text{universal}}(\theta, \eta; G^n)\right)$, it is the likelihood of the vector G^n. The posterior distribution $\pi(\rho|G^n)$ is obtained by marginalizing over $\pi(\eta, \sigma^2)$:

$$\pi(\rho|G^n) = \int \pi(G^n|\eta, \sigma^2, \rho)\pi(\rho)\pi(\eta|\sigma^2)\pi(\sigma^2) d\eta d\sigma^2 \ . \tag{5.48}$$

GP regression is mostly about prediction of the regression function g at an input $\mathbf{x}$. It is natural to derive the posterior prediction distribution $\pi(G(\mathbf{x})|G^n = g^n)$:

$$\pi(G(\mathbf{x})|G^n = g^n) = \int \pi(G(\mathbf{x})|\eta, \sigma^2, \rho, G^n = g^n)\pi(\eta, \sigma^2, \rho|G^n = g^n)d\eta d\sigma^2 d\rho \ . \tag{5.49}$$

Remark that $\pi(G(\mathbf{x})|\eta, \sigma^2, \rho, G^n = g^n)$ is the conditional distribution of the GP coming from the kriging equations (see Equation 5.6). Equation 5.49 illustrates the major drawback of the full Bayesian approach: The posterior predictive distribution in this case is no longer analytically tractable as in the more traditional kriging approach and its computation relies on into approximating a potentially high-dimensional integral. Thus, one has in the general case to resort to approximation by Monte Carlo Markov Chains (MCMC) methods or variational methods such as Laplace approximation (see Robert and Casella, 2004). After sampling from the posterior distribution a sample $(\eta_i, \sigma_i^2, \rho_i)_{1\leq i\leq N}$, the posterior predictive distribution is approximated by Monte-Carlo:

$$\pi(G(\mathbf{x})|G^n = g^n) \approx \frac{1}{N}\sum_{i=1}^{N} \pi(G(\mathbf{x})|\eta_i, \sigma_i^2, \rho_i, G^n = g^n) \ . \tag{5.50}$$

However, because the computation of the likelihood requires the inversion of the correlation matrix, the computational cost of MCMC is typically prohibitive. In GP regression, it is common to estimate ρ by its maximum a posteriori (MAP) estimator:

$$\widehat{\rho}_n^{\text{MAP}} = \underset{\rho}{\text{argmax}}\, \ell(\rho; G^n) + \log \pi(\rho) \ , \tag{5.51}$$

where $\ell(\rho; G^n)$ is the log likelihood. The posterior mode estimator can be seen as a penalized maximum likelihood estimator, with the penalization term $\log \pi(\rho)$ coming from the prior modeling on the parameters. In order to have tractable notations, the correlation matrix will be denoted by W in this section.

MCMC or variational approximations can be computationally cumbersome. Some specific prior distributions of the parameters provide closed form expression for the posterior predictive distribution, alleviating this computational burden. (Those priors are coined as *conjugate priors* in the Bayesian literature).

Conjugate priors

We will consider the following prior distributions, that are known to provide an analytically tractable posterior distribution. Note that in this part, the characteristic length parameter ρ is supposed known.

$$\eta|\sigma^2 \sim \mathcal{N}(\eta_0, \sigma^2 V_0) \tag{5.52}$$

and

$$\sigma^2 \sim \mathcal{IG}(\alpha, \gamma) \tag{5.53}$$

where $\mathcal{IG}$ denotes the inverse gamma probability distribution with density function:

$$f(x) = \frac{\gamma^\alpha e^{-\gamma/x}}{\Gamma(\alpha)x^{\alpha+1}}\mathbb{1}_{x>0} \ .$$

These priors are commonly use in Bayesian GP regression (see e.g. Santner, B. Williams, and Notz, 2003, Section 4.1; Le Gratiet, 2013, Section 1.2.2). The likelihood of the universal kriging

framework defined in Equation 5.42 combined with the prior defined in Equation 5.52 raises the following posterior distribution:

$$\eta|G^n, \sigma^2 \sim \mathcal{N}(Av, A) \,, \tag{5.54}$$

where $A = (U^T W^{-1} U + V_0^{-1})/\sigma^2$ and $v = (U^T W^{-1} G^n + V_0^{-1}\eta_0)/\sigma^2$. The following equation

$$\pi(\sigma^2|G^n) = \pi(G^n|\eta, \sigma^2)\pi(\eta|\sigma^2)\pi(\sigma^2)/\pi(\eta|G^n, \sigma^2)\pi(G^n) \tag{5.55}$$

raises the following posterior distribution for parameter σ^2:

$$\sigma^2|G^n \sim \mathcal{IG}(v_\sigma, Q_\sigma) \,, \tag{5.56}$$

where

$$Q_\sigma \propto 2\gamma + (\eta - \widehat{\eta})(V_0 + (U^T W^{-1} U)^{-1})^{-1}(\eta_0 - \widehat{\eta}) + \bar{Q}_\sigma$$

with $v_\sigma = n/2+\alpha$, $\widehat{\eta} = (U^T W^{-1} U)^{-1} U W^{-1} G^n$ and $\bar{Q}_\sigma = (G^n)^T (W^{-1} - W^{-1}U(U^T W^{-1} U)^{-1} U^T W^{-1}) G^n$.

After obtaining the posterior distribution of the parameters, it is of major interest to derive the posterior predictive distribution $(G(\mathbf{x})|G^n = g^n)$ because the main objective of GP regression is prediction. As pointed out in Equation 5.49, the posterior predictive distribution is obtained through marginalization of the predictive distribution coming from the kriging equations by the posterior distribution of the parameters.

First, let us marginalize by the posterior of η:

$$\pi(G(\mathbf{x})|G^n = g^n, \sigma^2) = \int \pi(G(\mathbf{x})|G^n = g^n, \eta, \sigma^2)\pi(\eta|G^n = g^n, \sigma^2) d\eta \,, \tag{5.57}$$

we obtain through computation the following analytic form for the distribution of $(G(\mathbf{x})|G^n = g^n, \sigma^2)$:

$$(G(\mathbf{x})|G^n = g^n, \sigma^2) \sim \mathcal{N}(m_\eta(\mathbf{x}), s_\eta^2(\mathbf{x})) \,, \tag{5.58}$$

where

$$m_\eta(\mathbf{x}) = u(\mathbf{x})^T Av + c(\mathbf{x})^T \Sigma^{-1}(g^n - UAv) \,, \tag{5.59}$$

$$s_\eta^2(\mathbf{x}) = \sigma^2 \left(1 - \begin{pmatrix} u(\mathbf{x})^T & c(\mathbf{x})^T \end{pmatrix} \begin{pmatrix} -V_0^{-1} & U^T \\ U & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} u(\mathbf{x}) \\ c(\mathbf{x}) \end{pmatrix} \right) . \tag{5.60}$$

Now, we can use the final predictive distribution $(G(\mathbf{x})|G^n = g^n)$ by integrating on the posterior distribution of σ^2. The computation raises the following posterior predictive distribution

$$(G(\mathbf{x})|G^n = g^n) \sim \mathcal{T}(v_\sigma, m_\eta(\mathbf{x}), Q_{\sigma,\eta}(\mathbf{x})) \,, \tag{5.61}$$

where $\mathcal{T}$ is the Student-t distribution, where m_η is defined in Equation 5.59,

$$Q_{\sigma,\eta}(\mathbf{x}) = \frac{Q_\sigma}{v_\sigma} \left(1 + \begin{pmatrix} u(\mathbf{x})^T & c(\mathbf{x})^T \end{pmatrix} \begin{pmatrix} -V_0^{-1} & U^T \\ U & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} u(\mathbf{x}) \\ c(\mathbf{x}) \end{pmatrix} \right) , \tag{5.62}$$

and Q_σ and v_σ are introduced in Equation 5.56.

Now, we will discuss special types of priors which are based on the likelihood. These priors are often coined as *non informative* priors, due to the fact they can be a default choice for the statistician in cases when there are no information about the parameters values before observing the data.

Non-informative priors

When there is no knowledge about the parameters before observing the data, one can build prior distributions based on the likelihood. These priors are denoted as *reference, objective, non-informative* priors in the literature. The most classical objective prior is the Jeffreys prior (Jeffreys, 1961) which is based on Fisher information matrix (Fisher, 1956). Note $\mathbf{X}$ a random variable following the probability distribution $\mathbb{P}_\theta$ with a pdf function $p(\theta; \bullet)$. The likelihood of the parameter $\theta = (\theta_i)_{1 \leq i \leq p}$ with respect to $\mathbf{X}$ is then $p(\theta; \mathbf{X})$. The Fisher information matrix is given by

$$\mathcal{I}(\theta) = -\left(\mathbb{E}_{\mathbf{X} \sim \mathbb{P}_\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(p(\theta; \mathbf{X})) \right] \right)_{1 \leq i,j \leq p} . \tag{5.63}$$

The Jeffreys prior distribution's pdf π_J is then defined by

$$\pi_J(\theta) \propto \sqrt{\det \mathcal{I}(\theta)} . \tag{5.64}$$

Its main interesting property is that it is invariant under reparametrization. Priors based on the likelihood have been extensively studied (see e.g. J. M. Bernardo, 1979; J. O. Berger, J. M. Bernardo, and Sun, 2009). Such kinds of priors for GP regression have been studied in Muré, 2018. The reference prior $\pi_R(\eta, \sigma^2, \rho)$ for the universal GP regression model with a covariance function of the form 5.38 with $\rho \in (0, +\infty)^p$ was developed in Paulo, 2005 and is given by:

$$\pi_R(\eta, \sigma^2, \rho) \propto \frac{\pi_R(\rho)}{\sigma^2} , \tag{5.65}$$

with $\pi_R(\rho) \propto \sqrt{\det \mathcal{I}(\rho)}$, where $\mathcal{I}(\rho)$ is the Fisher information matrix of the observation G^n and is written as follows:

$$\mathcal{I}(\rho) = (\mathcal{I}(\rho)_{i,j})_{1 \leq i,j \leq p} = \begin{cases} n - r & \text{if } i = j = 0 \\ \mathrm{Tr}(V_j) & \text{if } i = 0,\ j \neq 0 \\ \mathrm{Tr}(V_i) & \text{if } j = 0,\ i \neq 0 \\ \mathrm{Tr}(V_i V_j) & \text{otherwise} \end{cases} \tag{5.66}$$

where $V_i = \dot{W}_i Q$ for $1 \leq i \leq p$, and $\dot{W}_i$ is the partial derivative of the correlation matrix W with respect to the i-th characteristic length parameter. The matrix Q is defined as:

$$Q = W^{-1}(I_n - U(U^T W^{-1} U)^{-1} U^T W^{-1}) .$$

In the case of simple kriging (i.e. $r = 0$) the reference prior has the same expression with $Q = I_n$. As pointed out in Gu, 2019, the reference prior π_R have several interesting properties when $\rho_i \to 0$ or $\rho_i \to +\infty$. The first case correspond to $W \approx I_n$ and the second to $W \approx \mathbf{1}_n \mathbf{1}_n^T$ (where $\mathbf{1}_n$ is a vector composed of ones), these cases are degenerate and provoke undesirable behaviors for prediction. However, the use of the reference prior is computationally challenging, mainly coming from the computation of the matrices V_i. To avoid this problem, a new class of prior coined jointly robust priors have been introduced in Gu, 2019, they have the same tail rate decay as the reference prior defined in Equation 5.65 while being analytical and thus computationally much simpler. The

jointly robust prior writes:

$$\pi_{JR}(\eta, \sigma^2, \rho) \propto \frac{\pi_{JR}(\rho)}{\sigma^2} \,, \tag{5.67}$$

where $\pi_{JR}(\rho)$ has the following form:

$$\pi_{JR}(\rho_1, \dots, \rho_p) = C_0 \left(\sum_{i=1}^{p} \frac{C_i}{\rho_i} \right)^a \exp\left(-b \sum_{i=1}^{p} \frac{C_i}{\rho_i} \right) \,, \tag{5.68}$$

where $C_0 = \frac{(p-1)! b^{a+p} \prod_{i=1}^{p} C_i}{\Gamma(a+p)}$ is the normalizing constant. The variables $a,\ b,\ C_i$ are to be chosen by the statistician. Again following Gu, 2019, the parameters C_i can be seen as scale parameters and can be set to $C_i = n^{-1/p} |\mathbf{x}_{\max}^{(i)} - \mathbf{x}_{\min}^{(i)}|$ where $\mathbf{x}_{\max}^{(i)}$ and $\mathbf{x}_{\min}^{(i)}$ are the maximum and minimum input values at the i-th coordinate and the factor $n^{-1/p}$ is the average distance between points in a dataset of n samples in the hypercube $[0,1]^p$. More details about choosing the prior parameters are to be found in Gu, Palomo, and J. Berger, 2018.

The jointly robust prior is also proposed in the case of noisy homoskedastic observations of the form $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Denote by $\zeta = \sigma_\varepsilon^2 / \sigma^2$, the jointly robust prior writes:

$$\pi_{JR}(\eta, \sigma^2, \rho, \zeta) \propto \frac{\pi_{JR}(\rho, \zeta)}{\sigma^2} \,, \tag{5.69}$$

where $\pi_{JR}(\rho, \zeta)$ has the following form:

$$\pi_{JR}(\rho_1, \dots, \rho_p, \zeta) = C_0 \left(\sum_{i=1}^{p} \frac{C_i}{\rho_i} + \zeta \right)^a \exp\left(-b \left(\sum_{i=1}^{p} \frac{C_i}{\rho_i} + \zeta \right) \right) \,. \tag{5.70}$$

The different parameters can be estimated using the MAP estimator defined in Equation 5.51.

5.5 Gaussian process regression for seismic fragility curves estimation

In this section, we will perform GP regression for seismic probabilistic risk assessment (SPRA) purposes, and more precisely by estimating seismic fragility curves using Gaussian processes. After building the GP metamodel of the mechanical computer model, it is possible to carry out an UQ study as detailed in Section 4.3.2. The sources of uncertainties are divided into two parts. The first part is aleatory uncertainties which are related to the stochastic ground motions. A synthetic generator of ground motions based on a filtered modulated white noise (Rezaeian and Der Kiureghian, 2010) is then used (see Chapter 2, Section 2.3.1). Since seisms are complex stochastic signals, it is common in SPRA studies to sum up the seismic hazard coming from the seisms by a so-called intensity measure (IM). This is often a scalar value obtained from the seismic signal such as the peak ground acceleration (PGA) or the pseudo spectral acceleration (PSA) (Grigoriu and Radu, 2021). The choice of the IM is of great importance for fragility curves estimation and the IM has to verify several properties to be representative of the seismic hazard (Grigoriu and Radu, 2021). We denote by a the scalar value corresponding to the IM. The model of the structure can also include epistemic uncertainties about the material and mechanical properties, these parameters will be denoted by the vector $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$. Furthermore, we will denote by z the engineering demand parameter (EDP) of interest. A very common statistical model between the EDP and the combination of structural and seismic uncertainty is the log-normal model:

$$\log(z(a, \mathbf{x})) = g(a, \mathbf{x}) + \varepsilon \,, \tag{5.71}$$

where $\mathbf{x}$ is the vector of the mechanical properties of the structure, a is the intensity measure, $g(a, \mathbf{x})$ is the regression function, and $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is a centered Gaussian noise. Note that this log-normal assumption for the EDP distribution is not necessary for the proposed methodology, any functional transformation of z - such as Box-Cox transformation (see G. E. Box and Cox, 1964) - is possible as long as it is normally distributed after this transformation. For the sake of notation simplicity, we will denote $y(a, \mathbf{x}) = \log(z(a, \mathbf{x}))$.

As mentioned in Chapter 2, seismic fragility curve estimation is usually performed by numerical simulations. The computational burden is thus significant, as it requires a consequent number of numerical simulations to provide an accurate estimation. As also introduced in Chapter 2, the seismic fragility curve can be expressed as

$$\Psi(a, \mathbf{x}) = \mathbb{P}(z(A, \mathbf{X}) > C | A = a, \mathbf{X} = \mathbf{x}) \, , \tag{5.72}$$

where $z(a, \mathbf{x})$ is a scalar mechanical demand parameter obtained through numerical simulation of the structure with mechanical parameters $\mathbf{x}$ subjected to a seismic ground motion of intensity a. C is a threshold of acceptable behavior of the structure regarding the mechanical demand parameter studied.

5.5.1 Gaussian process surrogate with homoskedastic noise

In order to build our surrogate model, we suppose that the regression function g is a realization of a Gaussian process G. We thus define the random observation by:

$$Y(a, \mathbf{x}) = G(a, \mathbf{x}) + \varepsilon \, . \tag{5.73}$$

Remark in Equation 5.73 that thanks to the homoskedastic Gaussian noise assumption on the nugget ε, the random observations $Y(a, \mathbf{x})$ is also a Gaussian process. We make the assumption that G is a zero mean Gaussian process with an anisotropic stationary Matérn $5/2$ covariance function parametrized by its intensity σ and its lengthscales $(\rho_i)_{1 \leq i \leq p+1}$. Given an experimental design made of n simulations of the mechanical computer model, we obtain the dataset $\mathcal{D}_n = ((a_i, \mathbf{x}_i), y(a_i, \mathbf{x}_i))_{1 \leq i \leq n}$. By the maximum likelihood method, we can provide estimates for the unknown covariance function hyperparameters $\sigma, (\rho_i)_{1 \leq i \leq p+1}$ and also the Gaussian noise variance σ_ε (see Marrel, Iooss, Van Dorpe, et al., 2008 for a practical implementation of the method). The dataset $\mathcal{D}_n$ can then be used to derive the conditional distribution of the Gaussian process Y for any $(a, \mathbf{x})$:

$$(Y(a, \mathbf{x}) | \mathcal{D}_n) \sim \mathcal{N}\left(m_n(a, \mathbf{x}), \sigma_n(a, \mathbf{x})^2\right) \, , \tag{5.74}$$

where $m_n(a, \mathbf{x})$ and $\sigma_n(a, \mathbf{x})^2 = s_n(a, \mathbf{x})^2 + \sigma_\varepsilon^2$ are obtained from the kriging equations (see Section 5.3). In the same fashion, we can derive the conditional distribution of the Gaussian process G on the regression function for any $(a, \mathbf{x})$:

$$(G(a, \mathbf{x}) | \mathcal{D}_n) \sim \mathcal{N}\left(m_n(a, \mathbf{x}), s_n(a, \mathbf{x})^2\right) \, . \tag{5.75}$$

The fragility curve is then obtained by replacing the computer model output y by a Gaussian process Y_n which follows the distribution of the Gaussian process Y conditioned to $\mathcal{D}_n$ detailed in Equation (5.74). Hence, for any vector $(a, \mathbf{x})$ we derive the estimator of the fragility curve $\Psi^{(1)}$:

$$\Psi^{(1)}(a, \mathbf{x}) = \mathbb{P}(Y_n(a, \mathbf{x}) > \log(C)) \, . \tag{5.76}$$

We can then use the distribution of Y_n to estimate the fragility curve:

$$\Psi^{(1)}(a, \mathbf{x}) = \Phi\left(\frac{m_n(a, \mathbf{x}) - \log(C)}{\sigma_n(a, \mathbf{x})}\right) , \tag{5.77}$$

where Φ is the cdf of the standard Gaussian distribution. Moreover, the Gaussian process surrogate allows us to propagate the surrogate model uncertainty into the fragility curve, thanks to the conditional distribution of the regression function $(G(a, \mathbf{x})|\mathcal{D}_n)$. We define G_n the Gaussian process with same distribution as the Gaussian process $(G|\mathcal{D}_n)$, then the fragility curve tainted by the uncertainty of the Gaussian process surrogate writes:

$$\begin{aligned} \Psi^{(2)}(a, \mathbf{x}) &= \Phi\left(\frac{G_n(a, \mathbf{x}) - \log(C)}{\sigma_\varepsilon}\right) , \\ G_n(a, \mathbf{x}) &\sim \mathcal{N}\left(m_n(a, \mathbf{x}), s_n(a, \mathbf{x})^2\right) . \end{aligned} \tag{5.78}$$

Remark that $\Psi^{(1)}$ is the mean of $\Psi^{(2)}$ with respect to the distribution of $(G|\mathcal{D}_n)$. In order to estimate the distribution of $\Psi^{(2)}$, we simulate K realizations $(G_{n,k}(a, \mathbf{x}))_{1\leq k\leq K}$ from $(G(a, \mathbf{x})|\mathcal{D}_n)$ to obtain a sample $(\Psi^{(2)}_k)_{1\leq k\leq K}$ of $\Psi^{(2)}$:

$$\Psi^{(2)}_k(a, \mathbf{x}) = \Phi\left(\frac{G_{n,k}(a, \mathbf{x}) - \log(C)}{\sigma_\varepsilon}\right) . \tag{5.79}$$

However, some mechanical structures have nonlinear behavior that can influence the local variability of the log-EDP $y(a, \mathbf{x})$. Thus, a varying nugget with respect to $(a, \mathbf{x})$ could be necessary to retrieve the distribution of $y(a, \mathbf{x})$. This comes with a cost in terms of dataset size, due to the increase in the number of parameters to estimate. We will then propose a simple heteroskedastic Gaussian process surrogate and study its impact on the seismic fragility curve estimation.

5.5.2 Gaussian process surrogate with heteroskedastic noise

In this section, the log-EDP $y(a, \mathbf{x})$ is now supposed to follow the statistical model:

$$\log(z(a, \mathbf{x})) = g(a, \mathbf{x}) + \varepsilon(a, \mathbf{x}) , \tag{5.80}$$

where $\varepsilon(a, \mathbf{x}) \sim \mathcal{N}(0, \sigma_\varepsilon(a, \mathbf{x})^2)$. There are two ways of estimating $\sigma_\varepsilon(a, \mathbf{x})$ described in Kyprioti and Taflanidis, 2021: The first one, called stochastic kriging (SK), is to consider several replications at the same value of the input parameters $(a, \mathbf{x})$ and to provide an empirical estimation of the heteroskedastic standard deviation $\sigma_\varepsilon(a, \mathbf{x})$, the other one is to propose a parametric model of the noise standard deviation $\sigma_\varepsilon(a, \mathbf{x}) = \varphi(a, \mathbf{x}; \beta)$, and to calibrate the parameters vector β using the dataset $\mathcal{D}_n = ((a_i, \mathbf{x}_i), y(a_i, \mathbf{x}_i))_{1\leq i\leq n}$. We decided to implement the second method with a parametric model for several reasons: SK imposes to be intrusive with respect to the stochastic ground motion generator in order to make several replications at precise seismic intensity a, we prefer to consider a framework that is independent of the generator of seismic ground motions, due to the high number and diversity of stochastic generators proposed in the literature. Moreover, SK also imposes to control the design of experiments in order to be able to make replications, but in many applications, like in Marrel, Iooss, and Chabridon, 2021, due to budget or time constraints engineers only have access to a Monte-Carlo dataset $\mathcal{D}_n = ((a_i, \mathbf{x}_i), y(a_i, \mathbf{x}_i))_{1\leq i\leq n}$, this makes the use of SK difficult.

The key aspect of the parametric model of the heteroskedastic noise is the choice of the family of parametric functions $\varphi(a, \mathbf{x}; \beta)$. A sparse representation is preferable in order to limit the dimension of the parameters vector β. Prior knowledge about earthquake engineering helps to reduce the dimension of the input parameters $(a, \mathbf{x})$. Indeed, it is common in earthquake engineering that variability in the EDP is mainly caused by mechanical nonlinearities, which themselves depend on the intensity of the seismic ground motion. This leads to the simplification $\varphi(a, \mathbf{x}; \beta) = \varphi(a; \beta)$. Thus, φ depends on only one variable, reducing drastically the dimension of β. The calibration of β is performed using maximum likelihood estimation as in the homoskedastic case, β is considered as a hyperparameter of the Gaussian process. After calibration of the hyperparameters, we can obtain the conditional distribution of the heteroskedastic Gaussian process for every $(a, \mathbf{x})$.

$$(Y(a, \mathbf{x})|\mathcal{D}_n) \sim \mathcal{N}\left(\check{m}_n(a, \mathbf{x}), \check{\sigma}_n(a, \mathbf{x})^2\right) , \tag{5.81}$$

we can also derive the conditional distribution of the Gaussian process G on the regression function:

$$(G(a, \mathbf{x})|\mathcal{D}_n) \sim \mathcal{N}\left(\check{m}_n(a, \mathbf{x}), \check{s}_n(a, \mathbf{x})^2\right) , \tag{5.82}$$

where $\check{\sigma}_n(a, \mathbf{x})^2 = \check{s}_n(a, \mathbf{x})^2 + \varphi(a; \check{\beta}_n)^2$, $\check{\beta}_n$ is the vector of parameters of the parametrized heteroskedastic standard deviation obtained by maximum likelihood. In the same fashion as for the homoskedastic Gaussian process we can estimate the fragility curve using the conditional distribution:

$$\check{\Psi}^{(1)}(a, \mathbf{x}) = \Phi\left(\frac{\check{m}_n(a, \mathbf{x}) - \log(C)}{\check{\sigma}_n(a, \mathbf{x})}\right) , \tag{5.83}$$

the uncertainty on the Gaussian process $(G|\mathcal{D}_n)$ can be propagated in the fragility curve in the same fashion as for the homoskedastic Gaussian process:

$$\begin{aligned} \check{\Psi}^{(2)}(a, \mathbf{x}) &= \Phi\left(\frac{H_n(a, \mathbf{x}) - \log(C)}{\varphi(a; \check{\beta}_n)}\right) , \\ H_n(a, \mathbf{x}) &\sim \mathcal{N}\left(\check{m}_n(a, \mathbf{x}), \check{s}_n(a, \mathbf{x})^2\right) . \end{aligned} \tag{5.84}$$

The distribution of $\check{\Psi}^{(2)}$ is empirically estimated by generating K realizations $(H_{n,k}(a, \mathbf{x}))_{1 \leq k \leq K}$ from the distribution $\mathcal{N}\left(\check{m}_n(a, \mathbf{x}), \check{s}_n(a, \mathbf{x})^2\right)$ in order to estimate a sample of $\check{\Psi}^{(2)}$:

$$\check{\Psi}^{(2)}_k(a, \mathbf{x}) = \Phi\left(\frac{H_{n,k}(a, \mathbf{x}) - \log(C)}{\varphi(a; \check{\beta}_n)}\right) . \tag{5.85}$$

5.5.3 Uncertainty propagation on seismic fragility curves using Gaussian process surrogates

The Gaussian process surrogates allow us to propagate the uncertainties on $\mathbf{X}$ such that $\mathbf{X} \sim \mathbb{P}_{\mathbf{X}}$ into the fragility curves by considering the random functions $a \to \Psi(a, \mathbf{X})$. As previously introduced in Chapter 4, Section 4.3.2, we can derive from these random fragility curves several statistical quantities of interest such that the mean fragility curve:

$$\bar{\Psi}(a) = \mathbb{E}_{\mathbf{X}}[\Psi(a, \mathbf{X})] . \tag{5.86}$$

Moreover, the mechanical engineer may be interested in more conservative statistical quantities that will be useful for risk analysis. We recall the definition of the seismic fragility quantiles curve

$a \to q_\gamma(a)$ of level $\gamma \in [0,1]$ defined as:

$$q_\gamma(a) = \inf_{q \in \mathbb{R}} \{\mathbb{P}_{\mathbf{X}}(\Psi(a, \mathbf{X}) \leq q) \geq \gamma\} \,. \tag{5.87}$$

The estimation of these quantities of interest can be carried out using a Monte-Carlo sample $(\mathbf{X}_j)_{1 \leq j \leq m}$ and the law of large numbers as detailed in Equation 4.12. For the fragility quantile curve, the seismic fragility curve estimator $\Psi^{(1)}$ can be used to propose the following plug-in estimator:

$$q_\gamma^{(1)}(a) = \inf_{q \in \mathbb{R}} \{\frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{(\Psi^{(1)}(a, \mathbf{X}_j) \leq q)} \geq \gamma\} \,. \tag{5.88}$$

Furthermore, the posterior predictive distribution of the GP surrogates can be used to obtain the posterior distribution of the seismic fragility quantile curve using $\Psi^{(2)}$. Using a sample of $(\Psi_k^{(2)})_{1 \leq k \leq K}$ of $\Psi^{(2)}$, we can estimate a γ_G-level quantile w.r.t. the posterior distribution of the GP surrogate.

$$q_{\gamma_G}^{(2)}(a, \mathbf{X}) = \inf_{q \in R} \left\{ \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{(\Psi_k^{(2)}(a, \mathbf{X}) \leq q)} \geq \gamma_G \right\} . \tag{5.89}$$

A *bi-level* seismic fragility quantile curve is then proposed by taking the $\gamma_{\mathbf{X}}$-level quantile of $q_{\gamma_G}^{(2)}(a, \mathbf{X})$ w.r.t. the probability distribution of $\mathbf{X}$.

$$q_{\gamma_G, \gamma_{\mathbf{X}}}^{(2)}(a) = \inf_{q \in R} \left\{ \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{(q_{\gamma_G}^{(2)}(a, \mathbf{X}_j) \leq q)} \geq \gamma_{\mathbf{X}} \right\} . \tag{5.90}$$

The denomination bi-level meaning that it encompasses both the uncertainty on $\mathbf{X}$ and on the GP surrogate modeling. The procedure of estimation of the bi-level seismic fragility quantile curve is detailed in Algorithm 3. The same procedure can be applied using the heteroskedastic GP surrogate.

Algorithm 3 Uncertainty propagation on seismic fragility curves with Gaussian process

Requirements:

1. a regular grid $(a_t)_{1 \leq t \leq T}$
2. a Monte-Carlo sample $(\mathbf{X}_j)_{1 \leq j \leq m}$ with $\mathbf{X}_j \sim \mathbb{P}_{\mathbf{X}}$
3. a learning sample $\mathcal{D}_n = ((a_i, \mathbf{x}_i), y(a_i, \mathbf{x}_i))_{1 \leq i \leq n}$.

Procedure:

1. For each a_t with $1 \leq t \leq T$
   (a) For each $\mathbf{X}_j$ with $1 \leq j \leq m$
      i. Compute with the kriging equations $m_n(a_t, \mathbf{X}_j)$, $\sigma_n(a_t, \mathbf{X}_j)$ and $s_n(a_t, \mathbf{X}_j)$
      ii. Compute $\Psi^{(1)}(a_t, \mathbf{X}_j)$ using Equation (5.77)
      iii. For $1 \leq k \leq K$, sample $G_{n,k}(a_t, \mathbf{X}_j) \sim \mathcal{N}(m_n(a_t, \mathbf{X}_j), s_n(a_t, \mathbf{X}_j)^2)$ and compute $\Psi_k^{(2)}(a_t, \mathbf{X}_j)$ using Equation (5.79)
2. Estimate the seismic fragility quantile curve at point a_t using the dataset $((\Psi^{(1)}(a_t, \mathbf{X}_j))_{1 \leq j \leq m}$ by Equation (5.88)
3. Estimate the bi-level seismic fragility quantile curve with surrogate uncertainty at point a_t using the dataset $(\Psi_k^{(2)}(a_t, \mathbf{X}_j))_{\substack{1 \leq k \leq K, \\ 1 \leq j \leq m}}$ by using Equation (5.89) and (5.90)

5.6 Uncertainty propagation on a single d.o.f. oscillator with nonlinear restoring force

5.6.1 Presentation of the application

In this section, we apply the methodology of Section 5.5 to the single d.o.f. oscillator with nonlinear restoring force introduced in Section 2.4. The source of epistemic uncertainties in this application are the mechanical parameters of the oscillator. The variables are the mass m, the spring stiffness k, the damping ratio ξ, the yield displacement z_d and the post-yield stiffness α_y. We thus have the parameters vector $\mathbf{x} = (k, m, \xi, z_d, \alpha_y)^T$. We denote by $\mathbf{X}$ the random vector following the probability measure $\mathbb{P}_{\mathbf{X}}$ of the mechanical parameters of which probabilistic model is described in Table 5.2.

TABLE 5.2: Probabilistic model of $\mathbf{X}$ for the nonlinear oscillator.

Variable	Name	Mean
m (kg)	Mass of the system	300
k (N/m)	Stiffness	$2.7\ 10^5$
ξ (1)	Damping ratio	0.015
z_d (m)	Yield displacement	$5\ 10^{-3}$
α_y (1)	Post-yield stiffness	$2\ 10^{-4}$

The variables are supposed mutually independent with uniform distribution parametrized for each variable by its mean and its coefficient of variation (equal to 30% for each input variable). The choice of the uniform distribution comes from the maximum entropy principle (see e.g. E. T. Jaynes, 1957). The choice of the parametrization of the mechanical system is also subjective: a parametrization with physically measurable parameters (such that the mass m) seems preferable in the context of the study of a real mechanical system. However, the mean of the parameters k and m have been chosen such that the natural frequency of the oscillator is 5 Hz. Indeed, this value is commonly obtained on real mechanical structures. The seismic intensity measure chosen for this study is the peak ground acceleration (PGA) previously used in Chapter 2, Section 2.4. Synthetic seismic signals are generated using the methodology of Section 2.3.1. In this study, the engineering demand parameter (EDP) is the maximal absolute displacement of the nonlinear oscillator during the seismic excitation, similarly to Section 2.4.

5.6.2 Building Gaussian process metamodels and estimation of seismic fragility curves

In this part, we estimate the fragility curve of the nonlinear oscillator using Gaussian process surrogates, one with homoskedastic noise assumption and another with an heteroskedastic noise assumption. We perform simple kriging with a zero-mean Gaussian process with an anisotropic tensorized Matérn $5/2$ covariance function. This choice was motivated by the need of keeping a few hyperparameters to keep the surrogate model as simple as possible. The hyperparameters

to be optimized are the variance σ and the lengthscales $(\rho_i)_{0\leq i\leq p}$ ($p = 5$ in this study) where index 0 denotes the intensity measure variable. The input of the surrogate $(a, \mathbf{x})$ is logarithmically transformed, centered and standardized before being taken as input variable by the GP for fitting. All the computations are done in `Python` 3.7 using the library `scikit-learn`.

For the homoskedastic GP surrogate, computation of the MLE and MAP estimators of the hyperparameters with the jointly robust prior described in Section 5.4.5 is carried out. In order to compare the estimates, MAP estimator and MLE of the hyperparameters have been computed on 500 replications of learning sample of size $n = 200$. Numerical results for the hyperparameters σ, σ_ε, ρ_0, ρ_1 are shown in Figure 5.5. Remark that the tail behavior of the jointly robust prior for $\rho_i \to 0$ and $\rho_i \to +\infty$ indeed reduces the uncertainty on the lengthscale parameters and leads to fewer degenerates estimates than the MLE.

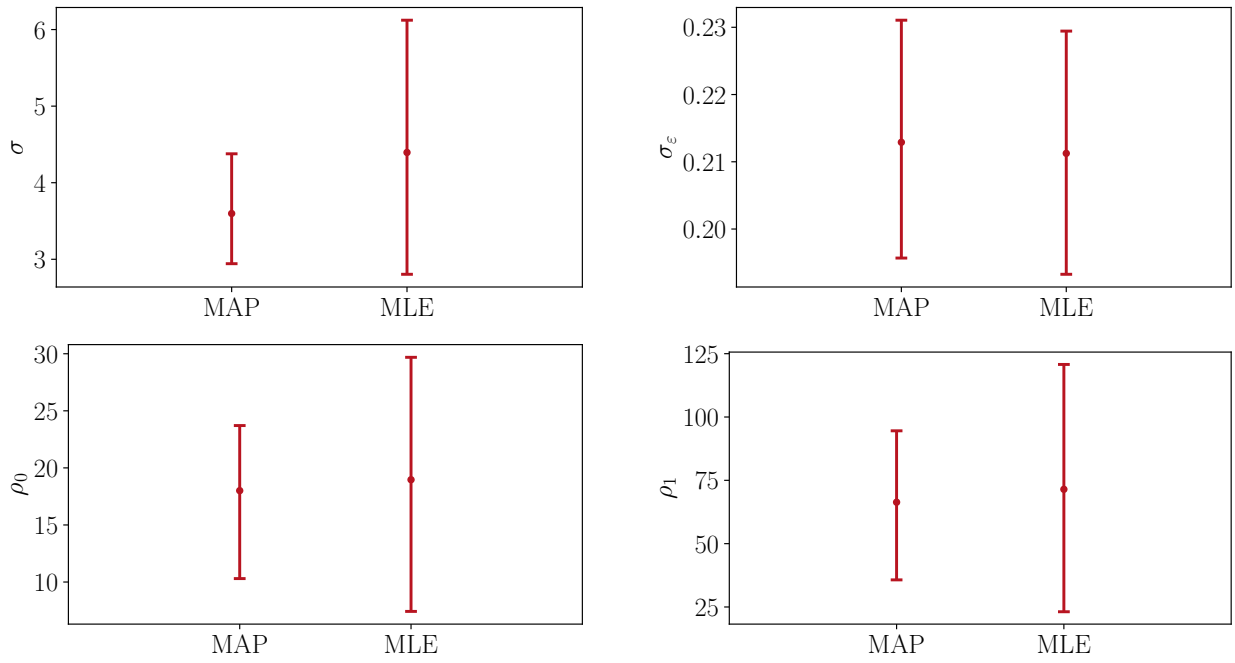


FIGURE 5.5: Comparison of the MAP estimator and the MLE of the hyperparameters. The red solid lines represent the interquantile range between the 10% and 90% quantiles.

A full Bayesian approach is also carried out on a learning sample of $n = 100$ observations. The parameters of the jointly robust prior are chosen according to Gu, Palomo, and J. Berger, 2018. The MCMC algorithm is implemented in `Stan` (see Stan Development Team, 2018) using the No-U-turn sampler (Hoffman and Gelman, 2011). 1000 posterior samples are then obtained with 1000 burn-in samples. Figure 5.6 represents the estimated marginal posterior distributions for different hyperparameters based on the MCMC samples. We can see that the marginal posterior distribution of σ_ε is close to Gaussian and concentrated around its mode. On the contrary, the parameters ρ_1 - corresponding to the stiffness k - is poorly identified, with a heavy tail towards large lengthscale values. The parameter ρ_0 is better identified with a smaller right tail. This behavior is coming from the very strong influence of the seismic signal - through the intensity measure variable a - to the log displacement of the oscillator relatively to the mechanical parameters of the structure. The influence of the mechanical parameters to the log displacement is way more difficult to infer.

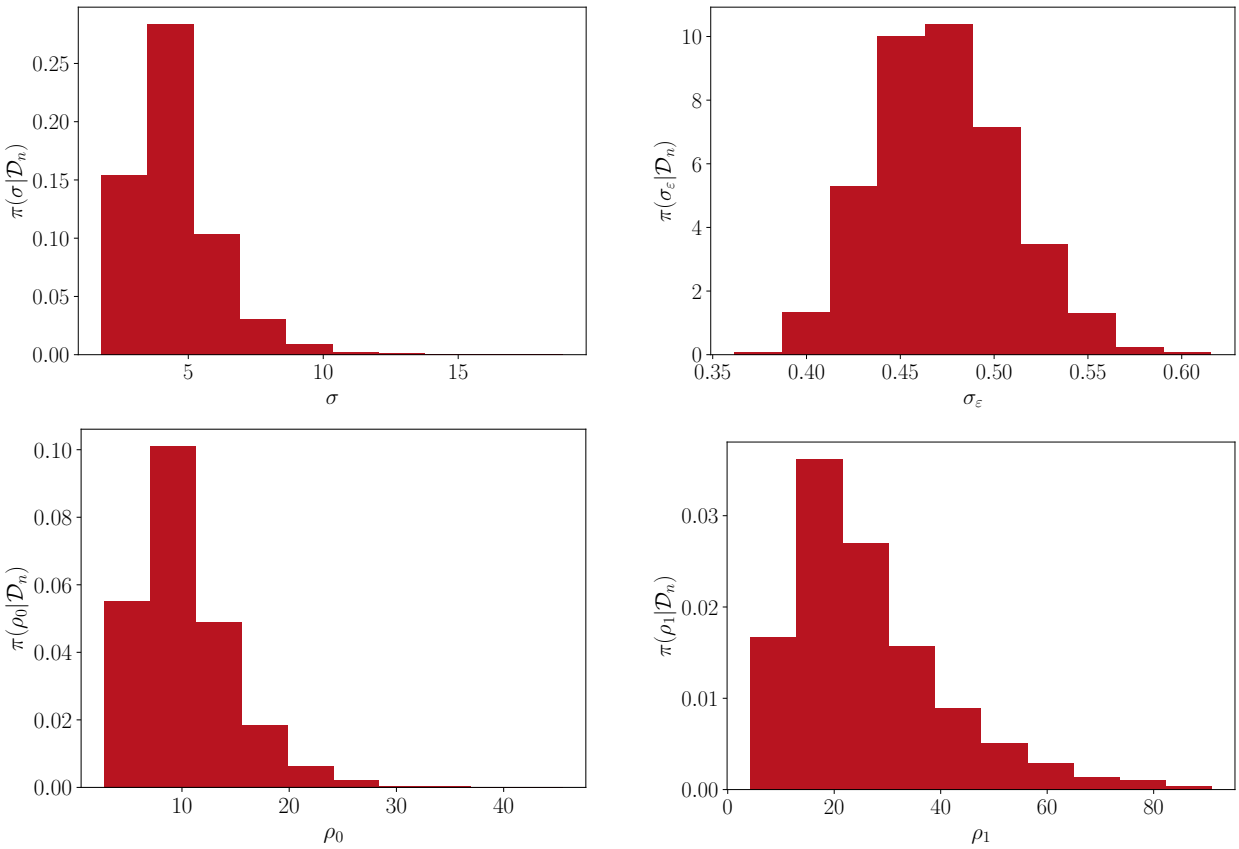


FIGURE 5.6: Histograms of the marginal posterior distribution for parameters σ, σ_ε, ρ_0, ρ_1 using the MCMC posterior samples.

Figure 5.7 represents scatter plots of parameters σ and σ_ε and of σ and ρ_0. Remark that the parameters σ and σ_ε are not correlated on the contrary to σ and ρ_0 where a positive correlation is observed.

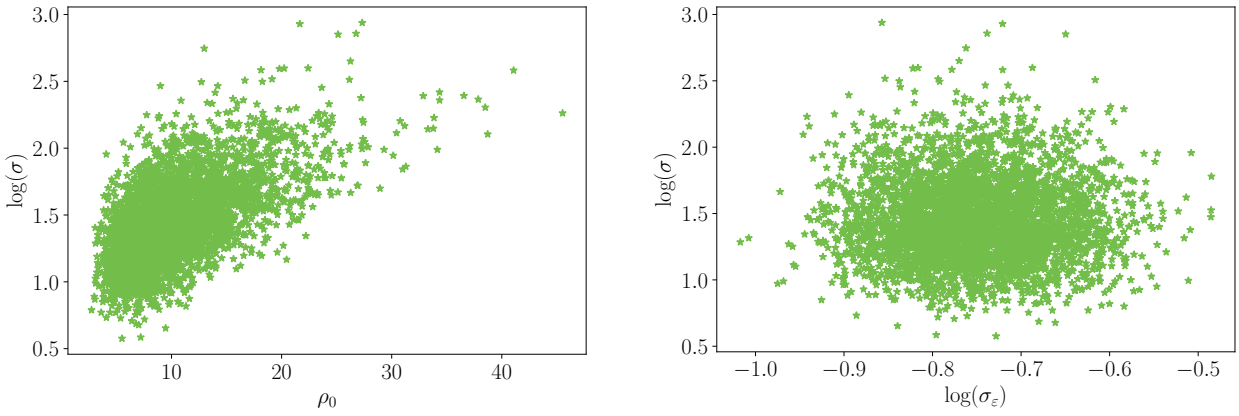


FIGURE 5.7: Scatter plots for different parameters with 1000 posterior samples.

For the rest of the Section, we will perform fragility curve estimation using the homoskedastic GP with the MAP estimator. This is motivated by its computational simplicity and the learning sample size in our applications (e.g. several hundreds). The heteroskedastic GP hyperparameters

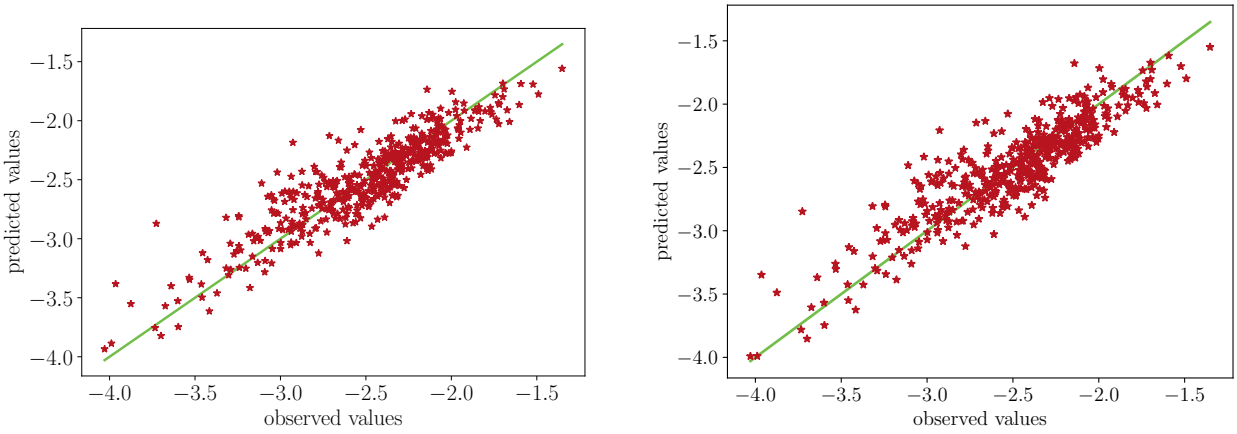
will be estimated by maximum likelihood.

Now we will consider the GP surrogate based on a heteroskedastic noise assumption. For the parametric form of the standard deviation for the heteroskedastic Gaussian process, we consider the following ramp function:

$$\varphi(a;\beta) = \max(\beta_0 + \beta_1 a, \beta_2) \,, \tag{5.91}$$

where $\beta = (\beta_0, \beta_1, \beta_2)$. This parametric formulation depends only on one variable and the small dimension of β allows for its calibration with common sample sizes encountered in our applications ($10 < n < 1000$). This parametric model for the heteroskedastic noise is motivated by the model proposed in Kyprioti and Taflanidis, 2021.

The predictivity quality of our two surrogates is assessed as follows: Figure 5.8 shows the predicted versus observed values of the log-EDP $y(a, \mathbf{x})$ using a learning dataset of $n = 500$ observations, the green solid line corresponds to the identity, the closer the data are from this line the better is the prediction quality of the surrogate. The plots are quite similar for this application, with the heteroskedastic GP predicting the high log-EDP values slightly better than the homoskedastic GP.



(A) Heteroskedastic Gaussian process

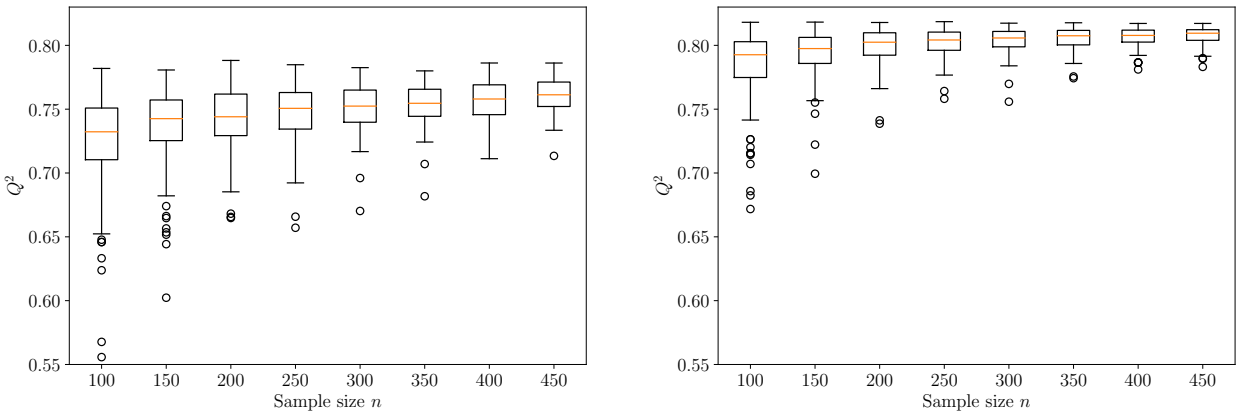
(B) Homoskedastic Gaussian process

FIGURE 5.8: Predicted values versus observed values for the heteroskedastic and homoskedastic Gaussian process surrogate with a dataset size $n = 500$.

Next, we measure quantitatively the predictivity quality of our surrogates using the predictivity coefficient Q^2 defined as:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_t} (y(a_i^t, \mathbf{x}_i^t) - m_n(a_i^t, \mathbf{x}_i^t))^2}{\sum_{i=1}^{n_t} (y(a_i^t, \mathbf{x}_i^t) - \bar{y})^2} \,, \tag{5.92}$$

where $(\mathbf{x}_i^t, y(a_i^t, \mathbf{x}_i^t))_{1 \le i \le n_t}$ is a test dataset, and $\bar{y} = \frac{1}{n_t} \sum_{i=1}^{n_t} y(a_i^t, \mathbf{x}_i^t)$. The predictivity coefficient Q^2 is computed on a test dataset of 1000 mechanical simulations, we consider learning sample size n varying between 100 and 450. The learning datasets are resampled using bootstrap in a total dataset of also 1000 mechanical simulations and 100 replications are made for each n. Figure 5.9 shows boxplots of the predictivity coefficient Q^2 for the heteroskedastic and homoskedastic Gaussian processes. Note that the homoskedastic GP surrogate seems to perform better than the heteroskedastic GP. Indeed, the mean Q^2 over the 100 replications for the homoskedastic GP surrogate is greater than the mean Q^2 of the heteroskedastic GP surrogate.



(A) Heteroskedastic Gaussian process

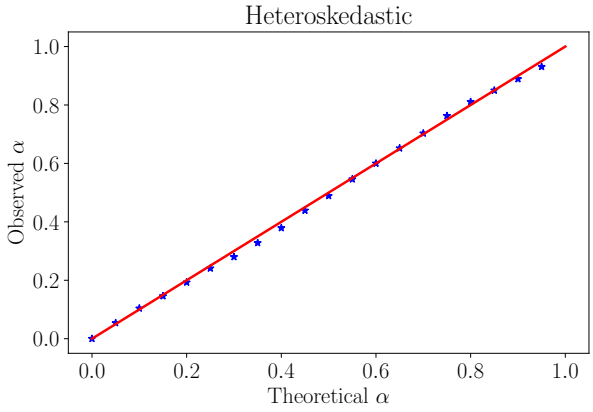
(B) Homoskedastic Gaussian process

FIGURE 5.9: Boxplots of the predictivity coefficients Q^2 estimated with 100 replications for varying size of the learning sample from $n = 100$ to $n = 450$.

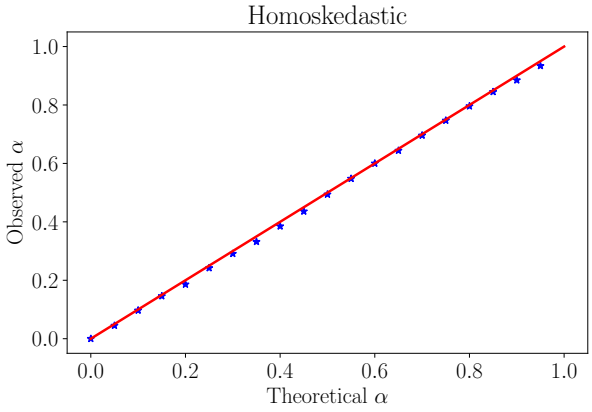
Moreover, we also provide a graphical tool proposed in Marrel, Iooss, and Chabridon, 2021 which consists in evaluating the proportion of data that lies in the α-theoretical confidence interval obtained with heteroskedastic and homoskedastic Gaussian process surrogates. Several values $\alpha \in [0, 1]$ of the prediction interval level are chosen and the theoretical level of the prediction interval is compared to the empirical proportion of the data that belongs actually to this prediction interval. By definition, the more the points are close to the identity line, the better is the quality of the kriging surrogates. Figure 5.10 gives the results for heteroskedastic and homoskedastic Gaussian process surrogate for $n = 100,\ 250,\ 500$ learning dataset sizes. The empirical coverage probabilities are computed on a test sample of size 1000. Remark that the empirical coverages of the homoskedastic and heteroskedastic GP surrogate are almost similar. Hence, the two surrogates may provide almost the same conditional distribution of the log-displacement.

The quantities of interest are estimated empirically using a Monte-Carlo sampling $(\mathbf{X})_{1\leq j\leq m}$ of size $m = 1000$. Numerical results for several training sizes n and failure threshold C are shown in Figures 5.11 and 5.12, the red area corresponds to the area of the 10% and 90% level fragility quantiles estimated using $(\Psi^{(1)}(., \mathbf{X}_j))_{1\leq j\leq m}$ and $(\check{\Psi}^{(1)}(., \mathbf{X}_j))_{1\leq j\leq m}$ for respectively the homoskedastic and heteroskedastic Gaussian processes. The Gaussian process surrogate uncertainty is assessed by sampling $K = 1000$ realizations of G_n and H_n for each value $\mathbf{X}_j,\ 1 \leq j \leq m$, the bi-level fragility quantile curves with $\gamma_{\mathbf{X}} = \gamma_G = 0.1$ and $\gamma_{\mathbf{X}} = \gamma_G = 0.9$ are shown in dashed green and they are estimated empirically from the dataset $(\Psi_k^{(2)}(., \mathbf{X}_j))_{\substack{1\leq k\leq K,\\ 1\leq j\leq m}}$ and $(\check{\Psi}_k^{(2)}(., \mathbf{X}_j))_{\substack{1\leq k\leq K,\\ 1\leq j\leq m}}$ respectively for the homoskedastic and heteroskedastic Gaussian processes. The blue curve corresponds to a nonparametric fragility curve estimation of the mean fragility curve using K-means clustering and binned Monte-Carlo (see Chapter 2, Section 2.3.2) on a large dataset of 20000 mechanical simulations. The dashed blue curves corresponds to the 10% and 90% quantiles of the nonparametric estimator using its asymptotic normal distribution.

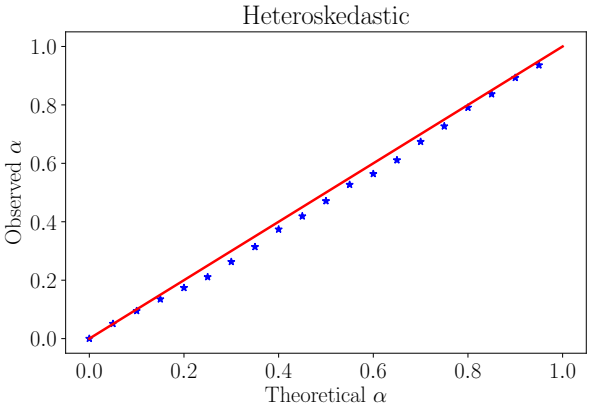
Remark that the interquantile range accounting for metamodeling uncertainty (in green) of the heteroskedastic Gaussian process is wider than for the homoskedastic Gaussian process. This heuristic may validate for this particular test case the use of a homoskedastic Gaussian process rather than a heteroskedastic one.



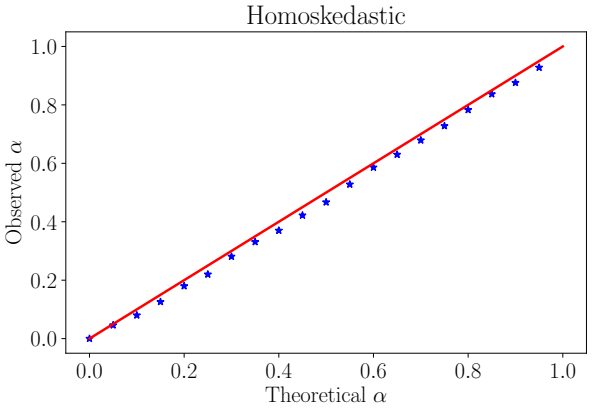
(A) $n = 100$



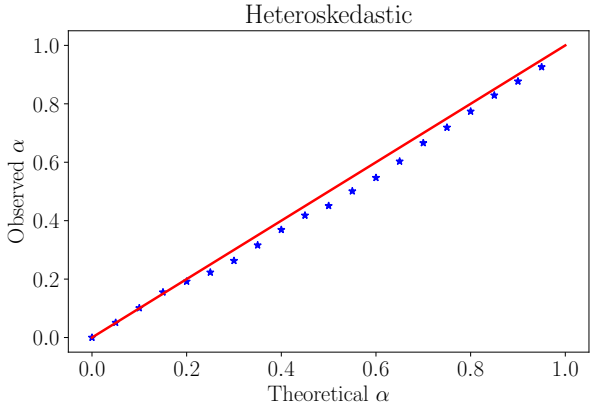
(B) $n = 100$



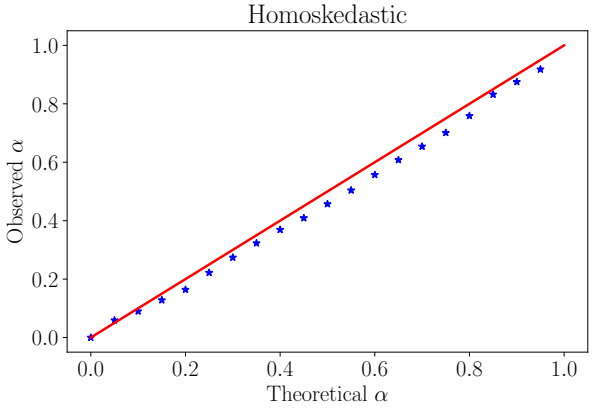
(C) $n = 250$



(D) $n = 250$

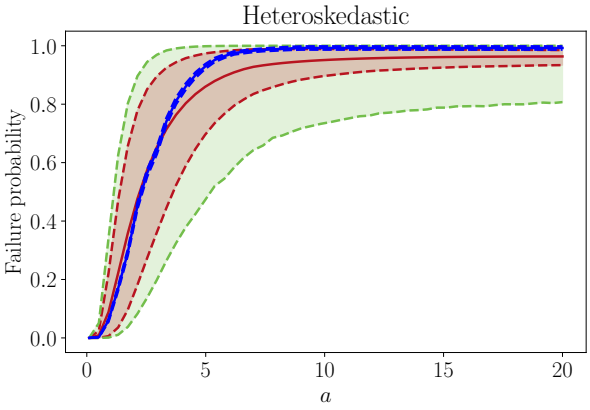


(E) $n = 500$

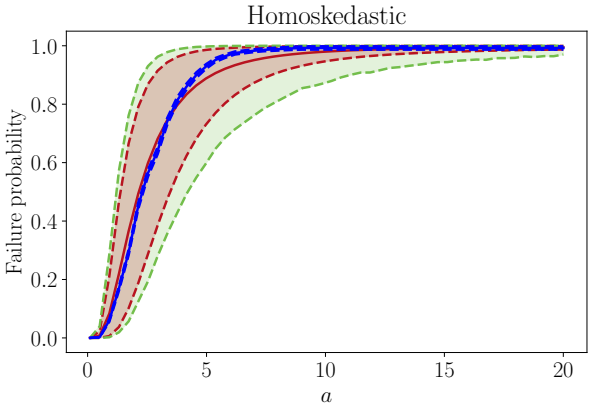


(F) $n = 500$

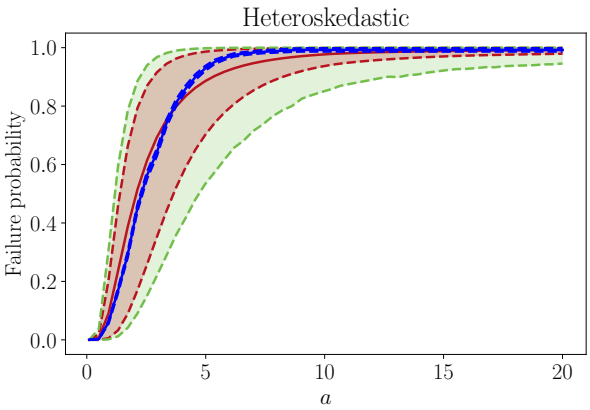
FIGURE 5.10: Observed proportion of the data that lies in the α-theoretical confidence intervals with respect to their theoretical proportion for both heteroskedastic and homoskedastic Gaussian processes with sample size of the learning dataset.



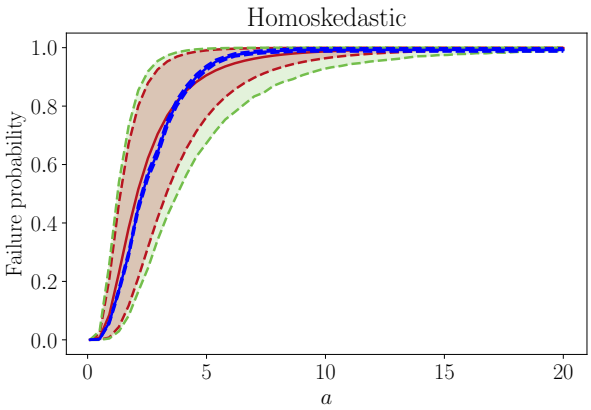
(A) $C = 1.3m_{z_d}$, $n = 100$



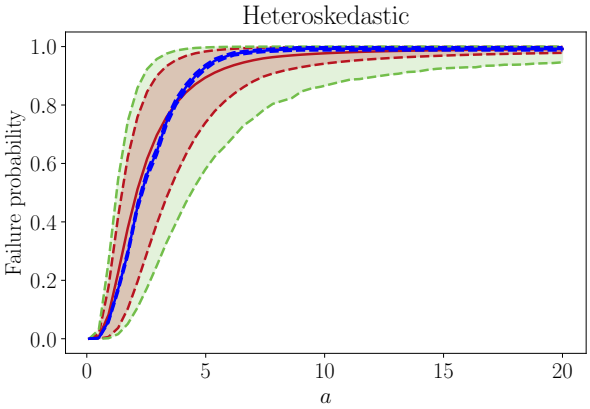
(B) $C = 1.3m_{z_d}$, $n = 100$



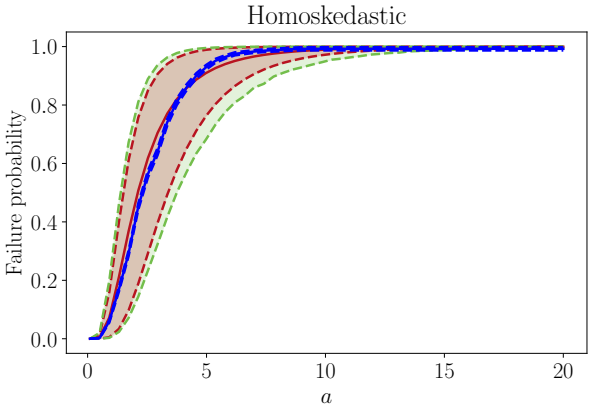
(C) $C = 1.3m_{z_d}$, $n = 250$



(D) $C = 1.3m_{z_d}$, $n = 250$

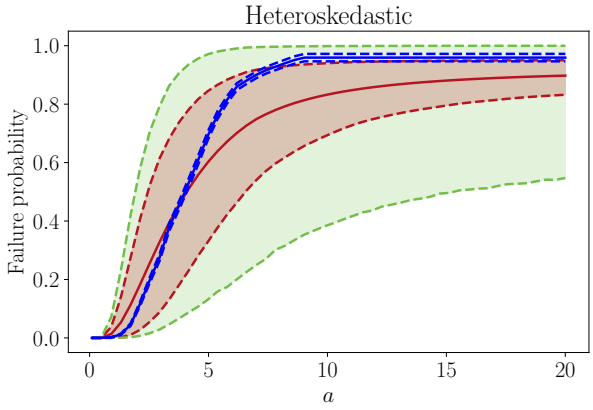


(E) $C = 1.3m_{z_d}$, $n = 500$

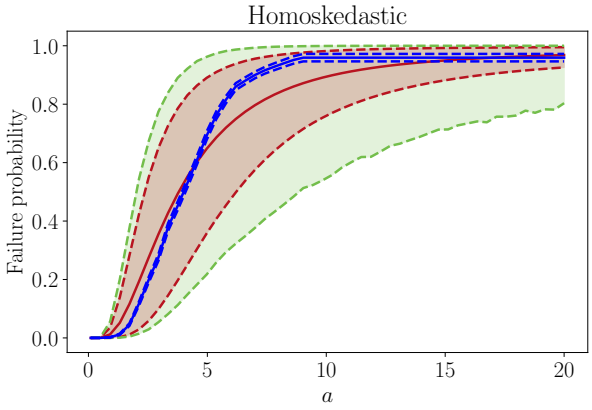


(F) $C = 1.3m_{z_d}$, $n = 500$

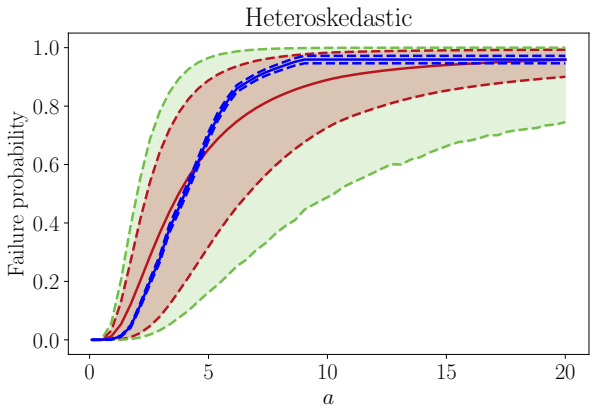
FIGURE 5.11: Uncertainty propagation of the epistemic uncertainties on the seismic fragility curves with failure threshold of $C = 1.3m_{z_d}$ where m_{z_d} is the mean of the random parameter z_d. The red area corresponds to the area of the 10% and 90% level estimated fragility quantiles and the green area to the area of the 10% and 90% level estimated bi-level quantile. The blue curve corresponds to a nonparametric fragility curve estimation of the mean fragility curve using K-means clustering and binned Monte-Carlo.



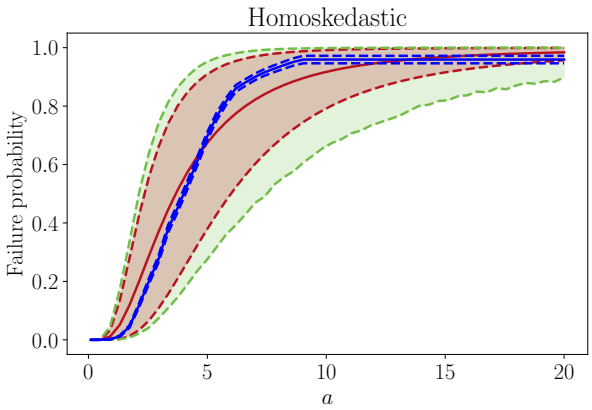
(A) $C = 2.1m_{z_d}$, $n = 100$



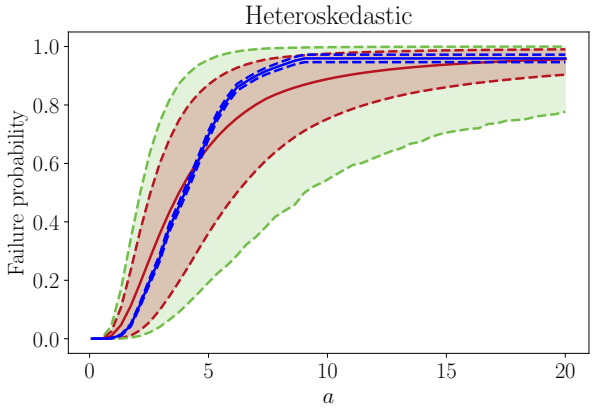
(B) $C = 2.1m_{z_d}$, $n = 100$



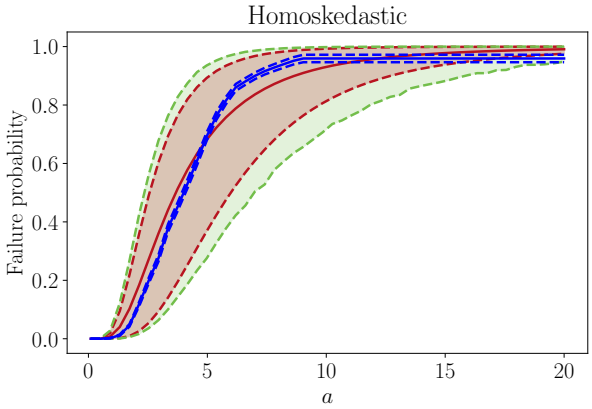
(C) $C = 2.1m_{z_d}$, $n = 250$



(D) $C = 2.1m_{z_d}$, $n = 250$



(E) $C = 2.1m_{z_d}$, $n = 500$



(F) $C = 2.1m_{z_d}$, $n = 500$

FIGURE 5.12: Uncertainty propagation of the epistemic uncertainties on the seismic fragility curves with failure threshold of $C = 2.1m_{z_d}$ where m_{z_d} is the mean of the random parameter z_d. The red area corresponds to the area of the 10% and 90% level estimated fragility quantiles and the green area to the area of the 10% and 90% level estimated bi-level quantile. The blue curve corresponds to a nonparametric fragility curve estimation of the mean fragility curve using K-means clustering and binned Monte-Carlo.

5.7 Synthesis about Gaussian process regression

This chapter provided a review about the metamodeling of black box computer models using Gaussian processes. This section sums up the main aspects that have been covered in this chapter.

Kriging. As shown in the beginning of the chapter, Gaussian process regression was first introduced for geostatistics purposes and coined kriging. This leads to the kriging equations for simple kriging (i.e. the mean function is known) or universal kriging (i.e. the mean is unknown and parametrized by a vector η such that the mean function is linear in η). The case of noisy observations is also covered, as it is of primary importance for our applications to earthquake engineering.

The choice of the covariance function. The covariance function has a key role for studying the behavior of the realizations of a Gaussian process for theoretical purposes. The spectral decomposition of the covariance - thanks to the Mercer theorem - is also very useful in practice for approximating Gaussian process realizations through the Karhunen-Loeve decomposition. Practical aspects for choosing the covariance hyperparameters in a data-driven manner is also covered.

Gaussian process regression seen as a Bayesian nonparametric regression framework. The stochastic nature of the Gaussian process makes very natural a Bayesian interpretation of the kriging framework. Indeed, the Gaussian process may be seen as a latent regression function. This interpretation - made by the machine learning community in the 2000s - leads to a more modern viewpoint of the Gaussian process regression. Moreover, this interpretation naturally leads to the full Bayesian approach, where the parameters and hyperparameters of the Gaussian process are seen as random variables following prior distributions. The Bayesian framework is then used for both inference of the various parameters and for prediction.

5.8 Conclusion

This chapter provides a review about Gaussian process regression and its use for Uncertainty Quantification on seismic fragility curves of structures by numerical simulations. This framework is illustrated on a single d.o.f. oscillator with nonlinear restoring force with its mechanical parameters considered as random variables modeling epistemic uncertainties. Due to the computational cost of mechanical computer models, the epistemic uncertainties propagation on the seismic fragility curve has to be performed using surrogates.

Two surrogate models have been proposed: one with a homoskedastic noise assumption and the other one with a parametrized heteroskedastic noise. The metamodels are validated using different metrics based on their predictivities and their coverages for a test-case consisting in a single d.o.f. oscillator with a nonlinear restoring force.

After assessing the uncertainties on the seismic fragility curves provoked by the epistemic uncertainties, it is of particular interest for the mechanical engineer to apportion the uncertainties on the seismic fragility curves to the different sources of epistemic uncertainties. Thus, in the next chapter, sensitivity analysis on seismic fragility curves using Gaussian process surrogates is discussed.

CHAPTER 6

Sensitivity analysis on seismic fragility curves using Gaussian process surrogates

Toutes les grandes sensibilités sont un peu prémonitoires

Romain Gary, *Les cerfs-volants*

Contents

6.1 Introduction

While the modeling and the propagation of uncertainties can be considered as *forward* UQ, the study of *backward* UQ may be of major interest as well. It boils down to study how the uncertainty on the model output can be apportioned to the different sources of uncertainties in the model inputs (see Saltelli et al., 2004, Section 2.2). Sensitivity analysis (SA) is a backward (or inverse) UQ study, as presented in Chapter 4, which aims at determining which input parameters (often

coined factors in the SA literature) drive the model output behavior. The different objectives of SA can be found in Saltelli et al., 2004, Section 2.4 and are listed below:

- **Factor prioritization**: SA may aim to rank the input parameters in terms of influence on the model output uncertainty. This objective aims to help make informed decisions about which input parameter uncertainty have to be reduced in priority to maximize the reduction of uncertainty on the model output.
- **Factor fixing**: This setting - also coined *screening* - aims at identifying the input parameter (or a subset of input parameters) that has little to no influence on the model output uncertainty. This objective aims at reducing the dimensionality of the input parameters without loosing information on the model behavior.
- **Variance cutting**: This aims to identify a minimal number of input parameters to fix in order to reach a target variance on the model output.
- **Factor mapping**: This aims to identify a subset of input parameters responsible for producing model output values in a given region of interest.

SA techniques are numerous and can be separated in two main families of methods: First, SA can be performed locally in the vicinity of a set of input parameters values, these methods are naturally coined *local* SA techniques. They mainly rely on gradient evaluations or Taylor expansions of the model. On the other hand, SA can be performed by taking into account the uncertainty range of the input parameters. This is coined global sensitivity analysis (GSA). The most famous GSA method is based on the variance-based importance measure called Sobol' indices (Sobol', 1993). They are based on the Hoeffding-Sobol' decomposition suggested by Hoeffding, 1948, which is only valid when the input parameters are mutually independent. For an extension of the Hoeffding-Sobol' to the non-independent case, the interested reader is referred to Chastaing, 2013. In this manuscript, only the case of mutually independent input parameters are considered.

In this chapter, we define global sensitivity indices on seismic fragility curves. First, the aggregated Sobol' indices introduced in Iooss and Le Gratiet, 2019 - a natural extension of the Sobol' indices to a functional output - is applied for GSA on seismic fragility curves. We also focus on recently studied global sensitivity indices based on kernel methods which seem adapted to functional output like seismic fragility curves. These indices are based on an embedding of the fragility curve into a Reproducing Kernel Hilbert Space (RKHS). Moreover, all the global sensitivity indices presented in this chapter are estimated using a Gaussian process surrogate of the mechanical computer model. Thus, the surrogate model uncertainty is propagated into the global sensitivity indices estimate in the same fashion as Le Gratiet, 2013, Chapter 6.

This chapter is organized as follows. First, we introduce the Sobol' indices and the aggregated Sobol' indices in Section 6.2. Then we present in Section 6.3 the global sensitivity indices based on the RKHS, by introducing the Maximum Mean Discrepancy (MMD) distance. Finally, we develop in Section 6.4 a method to propagate the posterior distribution of the Gaussian process surrogate into the global sensitivity indices. For that matter, we recall an important result for sampling a Gaussian process on a large sample. Section 6.5 illustrates the presented sensitivity analysis methodology on a single d.o.f. nonlinear oscillator.

6.2 Variance based global sensitivity indices

We will present in this section the method of Sobol' for global sensitivity analysis (Sobol', 1993) in the case where the model output is a scalar value. The extension of the Sobol' indices to

functional output, in the same fashion as Iooss and Le Gratiet, 2019, will also be presented with some theoretical results.

6.2.1 The method of Sobol'

Let us consider the input parameters space $\mathcal{X} \subset \mathbb{R}^p$ such that $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a measurable product space of the form:

$$(\mathcal{X}, \mathcal{B}(\mathcal{X})) = (\mathcal{X}_1 \times \ldots \times \mathcal{X}_p, \mathcal{B}(\mathcal{X}_1) \otimes \ldots \otimes \mathcal{B}(\mathcal{X}_p)) \, ,$$

where $\mathcal{B}(U)$ denote the Borelian σ-algebra of $U \subset \mathcal{X}$ and $\mathcal{X}_i \subset \mathbb{R}$ is a nonempty open set for $1 \leq i \leq p$. Furthermore, we consider a probability measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ of the form

$$\mu = \mu_1 \otimes \ldots \otimes \mu_p \, .$$

The Hoeffding-Sobol' decomposition introduced in Hoeffding, 1948 states that any function $g(\mathbf{x}) \in L^2_\mu(\mathcal{X})$ can be decomposed into terms of increasing dimensionality such that:

$$g(\mathbf{x}) = g_0 + \sum_{1 \leq i \leq p} g_i(\mathbf{x}^{(i)}) + \sum_{1 \leq i < j \leq p} g_{i,j}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \ldots + g_{1,2,\ldots,p}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}) = \sum_{u \in \mathcal{P}} g_u(\mathbf{x}^{(u)}) \, , \quad (6.1)$$

where $\mathcal{P}$ is the collection of all subsets of $\{1, \ldots, p\}$ and $\mathbf{x}^{(u)}$ is a group of variables such that $\mathbf{x}^{(u)} = (\mathbf{x}^{(i)})_{i \in u}$. The decomposition is also unique if we impose for all $u = (u_1, \ldots, u_k)_{1 \leq k \leq p}$, $1 \leq u_i \leq p$:

$$\int g_u(\mathbf{x}^{(u)}) d\mu_{u_i}(\mathbf{x}^{(u_i)}) = 0 \; \forall i = 1, \ldots, k \, . \quad (6.2)$$

This property has for consequence that the terms $g_u(\mathbf{x}^{(u)})$ are orthogonal in $L^2_\mu(\mathcal{X})$, i.e. for every $g_u(\mathbf{x}^{(u)})$ and $g_v(\mathbf{x}^{(v)})$ such that $u, v \in \mathcal{P}$ and $u \neq v$, we have:

$$\int g_u(\mathbf{x}^{(u)}) g_v(\mathbf{x}^{(v)}) d\mu(\mathbf{x}) = 0 \quad (6.3)$$

Another consequence is that g_0 is the mean of $g(\mathbf{x})$ w.r.t. the measure μ:

$$g_0 = \int g(\mathbf{x}) d\mu(\mathbf{x}) \, . \quad (6.4)$$

As shown in Sobol', 1993; Antoniadis, 1984, if $\mathbf{X} \in \mathbb{R}^p$ is a random variable following the probability distribution μ, the variance of $Y = g(\mathbf{X})$ can be decomposed thanks to the Hoeffding decomposition by:

$$\mathrm{Var}(Y) = V = \sum_{u \subseteq \{1,\ldots,p\}} \mathrm{Var}(g_u(\mathbf{X}^{(u)})) = \sum_{u \subseteq \{1,\ldots,p\}} V_u \, , \quad (6.5)$$

where

$$g_u(\mathbf{X}^{(u)}) = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathbb{E}[Y | \mathbf{X}^{(v)}] \, ,$$

where $|u| = \mathrm{Card}(u)$. Finally, the Sobol' sensitivity indices are given by

$$S_u = \frac{V_u}{V} \, ,$$

where $u \in \mathcal{P}$. These indices represent the part of variance of $Y = g(\mathbf{X})$ due to $\mathbf{X}^{(u)}$ and not explained by $\mathbf{X}^{(v)}$ where $v \subset u$. We can rewrite the Hoeffding decomposition in terms of Sobol' indices as follows.

$$1 = \sum_{i=1}^{p} S_i + \sum_{1 \leq i < j \leq p} S_{ij} + \ldots + S_{1,\ldots,p} = \sum_{u \in \mathcal{P}} S_u \,. \tag{6.6}$$

In particular, the $S_i = \mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(i)}])/V$ are called *first-order* Sobol' indices and measure the part of variance explained by the random variable $\mathbf{X}^{(i)}$. Furthermore, S_{ij} for $i \neq j$ are coined *second-order* Sobol' indices and measure the part of variance of Y explained by the interactions between variables $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ and not explained by the individual effects of $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$.

Finally, the *total-order* Sobol' indices T_u measure the part of variance explained by all the interactions between the subset of variables $\mathbf{X}^{(u)}$ and all the other variables, it is defined as follows

$$T_u = \sum_{v \in \mathcal{P}, u \cap v \neq \emptyset} S_v = 1 - \frac{\mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(-u)}])}{V} \,, \tag{6.7}$$

where $\mathbf{X}^{(-u)}$ is the complementary vector of variables with indices not in u.

Thus, Sobol' indices allow to quantitatively assess the impact of each input variable on the model output variance. One can use Sobol' sensitivity indices for screening purposes: a low total-order Sobol' index means that the input (or subset of inputs) variable is noninfluential and that freezing its value has no effect on the output variance. When the value of the first-order and total-order Sobol' indices are close, this means that there are no interactions with the other inputs, whereas a large difference indicates interactions.

6.2.2 Monte-Carlo estimation of Sobol' indices

In this section, we are interested in proposing estimators of the Sobol' sensitivity indices of the following form:

$$S_i = \frac{V_i}{V} = \frac{\mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(i)}])}{V} \,. \tag{6.8}$$

A naive estimator of $\mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(i)}])$ would require a double Monte-Carlo loop. It is however possible to design better Monte-Carlo estimators to compute first-order and total-order Sobol' indices if the inputs are independent (see e.g. Da Veiga et al., 2021, Section 3.2.1):

Lemma 3 (Pick-freeze formulation of Sobol' indices (Sobol', 1993))**.** *Assume $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are two independent copies of the input parameters vector. For any subset $u \subseteq \mathcal{P}$, define $\tilde{\mathbf{X}}_u$ a vector composed of $\mathbf{X}$ and $\tilde{\mathbf{X}}$ such that $\tilde{\mathbf{X}}_u^{(u)} = \mathbf{X}^{(u)}$ and $\tilde{\mathbf{X}}_u^{(-u)} = \tilde{\mathbf{X}}^{(-u)}$. Denote $Y = g(\mathbf{X})$ and $Y_u = g(\tilde{\mathbf{X}}_u)$, we have*

$$\mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(u)}]) = \mathrm{Cov}(Y, Y_u) \,.$$

Proof. First, remark that from the equality $Y \overset{\mathcal{L}}{=} Y_u$ we have the following

$$\begin{aligned} \mathrm{Cov}(Y, Y_u) &= \mathbb{E}[YY_u] - \mathbb{E}[Y]\mathbb{E}[Y_u] \\ &= \mathbb{E}[YY_u] - \mathbb{E}[Y]^2 \,. \end{aligned}$$

The following equalities hold since $\mathbf{X}^{(-u)}$ and $\tilde{\mathbf{X}}_u^{(-u)}$ are independent

$$\begin{aligned}\mathbb{E}[YY_u] &= \mathbb{E}[\mathbb{E}[YY_u|\mathbf{X}^{(u)}]] \\ &= \mathbb{E}\left[\mathbb{E}[Y|\mathbf{X}^{(u)}]\mathbb{E}[Y_u|\mathbf{X}^{(u)}]\right] \\ &= \mathbb{E}[\mathbb{E}[Y|\mathbf{X}^{(u)}]^2] \,.\end{aligned}$$

Finally, the conditional expectation property $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|\mathbf{X}^{(u)}]]$ leads to

$$\begin{aligned}\mathrm{Cov}(Y, Y_u) &= \mathbb{E}[\mathbb{E}[Y|\mathbf{X}^{(u)}]^2] - \mathbb{E}[\mathbb{E}[Y|\mathbf{X}^{(u)}]]^2 \\ &= \mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(u)}]) \,.\end{aligned}$$

□

Using the pick-freeze formulation, it is possible to estimate the first-order Sobol' indices by considering two random vectors $(\mathbf{X}_j)_{1\leq j\leq m}$ and $(\tilde{\mathbf{X}}_{i,j})_{1\leq j\leq m}$ and by estimating $\mathrm{Var}(\mathbb{E}[Y|\mathbf{X}^{(i)}])$ using the empirical estimator of $\mathrm{Cov}(Y, Y_i)$. Following this idea, Sobol' suggests the following estimator for the first-order Sobol' index (see Sobol', 1993):

$$\widehat{S}_i = \frac{\frac{1}{m}\sum_{j=1}^m g(\mathbf{X}_j)g(\tilde{\mathbf{X}}_{i,j}) - \frac{1}{m}\sum_{j=1}^m g(\mathbf{X}_j)\frac{1}{m}\sum_{j=1}^m g(\tilde{\mathbf{X}}_{i,j})}{\frac{1}{m}\sum_{j=1}^m g(\mathbf{X}_j)^2 - \left(\frac{1}{m}\sum_{j=1}^m g(\mathbf{X}_j)\right)^2} \,. \tag{6.9}$$

This Sobol' index estimator is improved in Janon et al., 2014 with the following estimator

$$\widehat{S}_i = \frac{\frac{1}{m}\sum_{j=1}^m g(\mathbf{X}_j)g(\tilde{\mathbf{X}}_{i,j}) - \left(\frac{1}{2m}\sum_{j=1}^m g(\tilde{\mathbf{X}}_{i,j}) + g(\mathbf{X}_j)\right)^2}{\frac{1}{2m}\sum_{j=1}^m (g(\mathbf{X}_j)^2 + g(\tilde{\mathbf{X}}_{i,j})^2) - \left(\frac{1}{2m}\sum_{j=1}^m g(\tilde{\mathbf{X}}_{i,j}) + g(\mathbf{X}_j)\right)^2} \,. \tag{6.10}$$

Indeed, the estimator in Equation 6.10 is asymptotically efficient for the first-order Sobol' indices. However, they are sometimes not accurate for small values of V_i/V. Note also that first-order Sobol' indices estimator based on rank statistics has been recently introduced (see Gamboa et al., 2022). They can be useful in a *given-data* context (e.g. no further simulations of the computer model is feasible).

6.2.3 Aggregated Sobol' indices for functional output

In this part, we present an adaptation of the Sobol' sensitivity indices to a functional output. It was first introduced in Le Gratiet et al., 2017; Iooss and Le Gratiet, 2019 in the context of POD (Probability of Detection) curves used in non-destructive testing studies. Using the notation $\mathbf{X} =$

$(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(p)})$, we first define the following quantities (with $\Psi_{\mathbf{X}}(a) = \Psi(a, \mathbf{X})$):

$$\begin{aligned}
\bar{\Psi}(a) &= \mathbb{E}_{\mathbf{X}}\left[\Psi(a, \mathbf{X})\right] , \\
\Psi_{\mathbf{X}^{(i)}}(a) &= \mathbb{P}(z(A, \mathbf{X}) > C | A = a, \mathbf{X}^{(i)}) , \\
\Psi_{\mathbf{X}^{(-i)}}(a) &= \mathbb{P}(z(A, \mathbf{X}) > C | A = a, \mathbf{X}^{(-i)}) , \\
D &= \mathbb{E}_{\mathbf{X}}\left[\|\bar{\Psi} - \Psi_{\mathbf{X}}\|^2_{L^2_h(\mathcal{A})}\right] \\
&= \mathbb{E}_{\mathbf{X}}\left[\int_{\mathcal{A}} (\bar{\Psi}(a) - \Psi(a, \mathbf{X}))^2 dh(a)\right] ,
\end{aligned} \tag{6.11}$$

where h is a probability measure on the compact bounded set $\mathcal{A} \subset \mathbb{R}$ with p.d.f. p_A. The aggregated Sobol' indices of first and total-order for the seismic fragility curves then write:

$$\begin{aligned}
S^{FC}_{i,m} &= \frac{\mathbb{E}_{\mathbf{X}}\left[\|\bar{\Psi} - \Psi_{\mathbf{X}^{(i)}}\|^2_{L^2_h(\mathcal{A})}\right]}{D} , \\
T^{FC}_{i,m} &= 1 - \frac{\mathbb{E}_{\mathbf{X}}\left[\|\bar{\Psi} - \Psi_{\mathbf{X}^{(-i)}}\|^2_{L^2_h(\mathcal{A})}\right]}{D} ,
\end{aligned} \tag{6.12}$$

In the next lemma, we demonstrate that these indices follow the same Sobol'-Hoeffding decomposition as in Equation 6.6.

Lemma 4. *Suppose that* $\forall a \in \mathcal{A}$, *the functions* $\mathbf{x} \mapsto \Psi(a, \mathbf{x})$ *are in* $L^2_\mu(\mathcal{X})$, *then we have*

$$\sum_{u \subseteq \{1,\dots,p\}} S^{FC}_u = 1 ,$$

where $S^{FC}_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathbb{E}\left[\frac{\|\Psi_{\mathbf{X}^{(v)}} - \bar{\Psi}\|^2_{L^2_h(\mathcal{A})}}{D}\right]$

Proof. For each $a \in \mathcal{A}$ fixed, if $\Psi(a, \cdot)$ is in $L^2_\mu(\mathcal{X})$, we can apply the variance decomposition stated in Equation 6.5:

$$\mathrm{Var}(\Psi(a, \mathbf{X})) = \sum_{u \subseteq \{1,\dots,p\}} V_u(a) , \tag{6.13}$$

where

$$V_u(a) = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathrm{Var}(\Psi_{\mathbf{X}^{(v)}}(a)) .$$

Denoting that, by definition $\mathrm{Var}(\Psi_{\mathbf{X}^{(v)}}(a)) = \mathbb{E}[(\Psi_{\mathbf{X}^{(v)}}(a) - \bar{\Psi}(a))^2]$ and by using the linearity of the expectation, we can integrate against the measure h on both sides of Equation 6.13. The results of Lemma 4 is then obtained by division by $D = \int_{\mathcal{A}} \mathrm{Var}(\Psi(a, \mathbf{X})) dh(a)$ on both sides of the equation. □

In consequence, the aggregated Sobol' indices can be interpreted in the same fashion as Sobol' indices on a scalar model output. In particular, the first-order aggregated Sobol' index S_i^{FC} corresponds to the averaged part of variance on the intensity measure of $\Psi(., \mathbf{X})$ due to input $\mathbf{X}^{(i)}$. The total-order aggregated Sobol' measure the average part of variance due to all the interactions with input $\mathbf{X}^{(i)}$. For more details about extending Sobol' indices for nonscalar output variables, the interested reader is referred to Da Veiga et al., 2021, Section 3.3.

Now, estimators of the aggregated Sobol' indices are proposed using also the pick-freeze methodology.

Lemma 5. *Using the same notation as in Lemma 3, we have*

$$\mathbb{E}_{\mathbf{X}}\left[\|\bar{\Psi} - \Psi_{\mathbf{X}^{(u)}}\|^2_{L^2_h(\mathcal{A})}\right] = \int_{\mathcal{A}} \mathrm{Cov}\left(\Psi_{\mathbf{X}}(a), \Psi_{\mathbf{X}^{(u)}}(a)\right) dh(a) \,.$$

Proof. It is possible to intervert the expectation and the integral

$$\mathbb{E}_{\mathbf{X}}\left[\|\bar{\Psi} - \Psi_{\mathbf{X}^{(u)}}\|^2_{L^2_h(\mathcal{A})}\right] = \int_{\mathcal{A}} \mathrm{Var}(\Psi_{\mathbf{X}^{(u)}}(a)) dh(a) \,,$$

it is possible to use for all $a \in \mathcal{A}$ Lemma 3 on the scalar random variable $\Psi_{\mathbf{X}^{(u)}}(a)$. □

Now, we only need to approximate the integral on $\mathcal{A}$ against the measure h to propose an estimator of the first and total-order aggregated Sobol' indices. It is possible to perform a Monte-Carlo sampling, however this is computationally cumbersome, as we need to perform one pick-freeze estimation per sample of the seismic intensity measure. A more economical way to approximate the integral is by relying on numerical quadrature algorithms (e.g. Riemann sum, trapezoidal rule, Simpson's rule, etc.). These methods are particularly suitable to our case since $\mathcal{A}$ is one-dimensional and bounded. Consider $(a_t)_{1\leq t\leq T}$ a regular grid of intensity measures of size T. One can propose the following pick-freeze estimator of the first and total-order aggregated Sobol' indices

$$\begin{aligned}
\widehat{S}^{FC}_{i,m} &= \frac{\displaystyle\sum_{t=1}^{T} \left\langle \Psi(a_t, \mathbf{X})\Psi(a_t, \tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t) - \langle \Psi(a_t, \mathbf{X})\rangle_m \left\langle \Psi(a_t, \tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t)}{\displaystyle\sum_{t=1}^{T} \left\langle \Psi(a_t, \mathbf{X})^2\right\rangle_m p_A(a_t) - \langle \Psi(a_t, \mathbf{X})\rangle^2_m p_A(a_t)} \,, \\
\widehat{T}^{FC}_{i,m} &= 1 - \frac{\displaystyle\sum_{t=1}^{T} \left\langle \Psi(a_t, \mathbf{X})\Psi(a_t, \tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t) - \langle \Psi(a_t, \mathbf{X})\rangle_m \left\langle \Psi(a_t, \tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t)}{\displaystyle\sum_{t=1}^{T} \left\langle \Psi(a_t, \mathbf{X})^2\right\rangle_m p_A(a_t) - \langle \Psi(a_t, \mathbf{X})\rangle^2_m p_A(a_t)} \,,
\end{aligned} \tag{6.14}$$

where we denote for any function f:

$$\left\langle f(\mathbf{X}, \tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_{-i})\right\rangle_m = \frac{1}{m}\sum_{j=1}^{m} f(\mathbf{X}_j, \tilde{\mathbf{X}}_{i,j}, \tilde{\mathbf{X}}_{-i,j}) \,. \tag{6.15}$$

6.3 Kernel-based sensitivity indices

Kernel-based methods in machine learning and statistics gain in popularity due to their ability to simplify difficult nonlinear problems into linear problems by embedding the data points into a Reproducing Kernel Hilbert Space (RKHS) (Schölkopf, B. Smola, et al., 2002). The main applications involve independence testing (see e.g. Gretton, Fukumizu, et al., 2007; Fukumizu, Gretton, et al., 2007) and dimensionality reduction (see e.g. Schölkopf, A. Smola, and K.-R. Müller, 1998; Fukumizu, Bach, and Jordan, 2004; Fukumizu, Bach, and Jordan, 2009). A first use of kernel methods for GSA purposes was proposed in Da Veiga, 2015 where the Hilbert Schmidt Independence Criterion (HSIC) is used to propose global sensitivity indices. Global sensitivity indices based on the Maximum Mean Discrepancy (MMD) has been proposed in Da Veiga, 2021; Barr and Rabitz, 2022 and defined with the rationale of Borgonovo, Hazen, and Plischke, 2016. These MMD-based sensitivity indices have also the interesting property of being Sobol' indices on the kernel embedding of the output variable as shown in Da Veiga, 2021. In this section, after a brief remainder on RKHS theory, we will define MMD-based sensitivity indices on the seismic fragility curves. Their main advantages is their resemblance to the Sobol' indices, allowing the same estimation methods to be used.

6.3.1 Embedding of probability distributions in a RKHS

Let us introduce a more general definition of a RKHS $\mathcal{H}$ (see Chapter 5, Section 5.4) through the notion of reproducing kernel

Definition 4. *Let $\mathcal{X}$ be a set and $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\mathbb{R}$ forming an Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if*

- *$\mathcal{H}$ contains all functions of the form*

$$\forall \mathbf{x} \in \mathcal{X},\ k_{\mathbf{x}} : \mathbf{t} \to k(\mathbf{x}, \mathbf{t}) .$$

- *For every $\mathbf{x} \in \mathcal{X}$ and $f \in \mathcal{H}$ the reproducing property holds:*

$$f(\mathbf{x}) = \langle f, k_{\mathbf{x}} \rangle_{\mathcal{H}} .$$

If such a kernel k exists, then $\mathcal{H}$ is called a reproducing kernel Hilbert space (RKHS).

One of the foundations of the RKHS theory is the Moore-Aronszajn theorem (Aronszajn, 1950) which links symmetric semi positive definite kernels to RKHS.

Theorem 8 (Moore-Aronszajn theorem)**.** *Let $\mathcal{X}$ be a set and suppose that the function $k : \mathcal{X} \times \mathcal{X}$ is symmetric positive definite. Then there is a unique Hilbert space of functions on $\mathcal{X}$ for which k is a reproducing kernel*

Consider now a RKHS $\mathcal{H}$ of functions $\mathcal{X} \to \mathbb{R}$ with reproducing kernel k with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. The *kernel mean embedding* $m_\mu \in \mathcal{H}$ of the probability distribution μ is defined as

$$m_\mu = \mathbb{E}_{\mathbf{X} \sim \mu}[k(\mathbf{X}, .)] = \int_{\mathcal{X}} k(\mathbf{x}, .) d\mu(\mathbf{x}) , \tag{6.16}$$

if $\mathbb{E}_{\mathbf{X} \sim \mu}[k(\mathbf{X}, \mathbf{X})] < \infty$. The *maximum mean discrepancy* (MMD) distance between two probability measures μ_1 and μ_2 on $\mathcal{X}$ is defined by the distance in the RKHS between their kernel mean embeddings:

$$\mathrm{MMD}(\mu_1, \mu_2) = \| m_{\mu_1} - m_{\mu_2} \|_{\mathcal{H}} . \tag{6.17}$$

The MMD corresponds to a distance between probability measure if the kernel k is characteristic (i.e. the mean embedding map $\mu \to m_\mu$ is injective). The MMD also has a practical interest as its computation relies only on expectations (see Gretton, Borgwardt, et al., 2012).

$$\mathrm{MMD}(\mu_1, \mu_2)^2 = \mathbb{E}[k(U, U')] + \mathbb{E}[k(V, V')] - 2\mathbb{E}[k(U, V)] \,, \tag{6.18}$$

where $U,\ U' \sim \mu_1$ and $V,\ V' \sim \mu_2$ with U, U', V, V' mutually independent. This means that on the contrary of most of the dissimilarity measures between probability measures, the MMD is evaluated using only expectations of the kernel. This allows for instance for Monte-Carlo approximations of the MMD.

6.3.2 Maximum mean discrepancy based global sensitivity indices

A general framework for designing global sensitivity indices using a dissimilarity measure was proposed in Da Veiga, 2015. Given a dissimilarity $d(\cdot, \cdot)$ between probability measures, denote by $\mathbb{P}_Y$ the marginal probability distribution of Y and by $\mathbb{P}_{Y|\mathbf{X}^{(u)}}$ the marginal probability distribution of Y given the input parameters $\mathbf{X}^{(u)}$. The sensitivity measure S_u^d of the set of input u writes

$$S_u^d = \mathbb{E}_{\mathbf{X}^{(u)}}\left[d\left(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(u)}}\right)\right] \,. \tag{6.19}$$

Indeed, the MMD can be used as a dissimilarity d and thus defining the following unnormalized (i.e. the sum of all the sensitivity indices is not equal to 1) sensitivity index. In fact, the MMD^2 dissimilarity will be used:

$$S_u^{\mathrm{MMD\ ,unorm}} = \mathbb{E}_{\mathbf{X}^{(u)}}\left[\mathrm{MMD}\left(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(u)}}\right)^2\right] \,. \tag{6.20}$$

This sensitivity index is well-defined under the assumption that for all $u \in \mathcal{P}$ and $\mathbb{P}_{\mathbf{X}^{(u)}}$-almost all $\mathbf{x}^{(u)} \in \mathcal{X}_u$, $\mathbb{E}_{U \sim \mathbb{P}_{Y|\mathbf{X}^{(u)}=\mathbf{x}^{(u)}}}[k(U, U)] < \infty$. This sensitivity index remarkably admits an Hoeffding-Sobol' (or FANOVA) decomposition and thus can be thought as a kernelized version of the Sobol' indices (see e.g. Da Veiga et al., 2021, Section 6.2.2).

Lemma 6 (Hoeffding-Sobol' decomposition in a RKHS (Da Veiga et al., 2021))**.** *Assume that* $\mathbf{X}$ *has independent components. Assume further that Mercer's theorem holds w.r.t. the probability distribution of* Y*. Denote* $\mathrm{MMD}^2_{\mathrm{tot}} = \mathbb{E}[k(Y, Y)] - \mathbb{E}[k(Y, \tilde{Y})]$*, where* $\tilde{Y}$ *is an independent copy of* Y*. Then the total MMD can be decomposed as*

$$\mathrm{MMD}^2_{\mathrm{tot}} = \sum_{u \in \mathcal{P}} \mathrm{MMD}^2_u \,, \tag{6.21}$$

where

$$\mathrm{MMD}^2_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} \mathbb{E}_{\mathbf{X}^{(v)}}\left[\mathrm{MMD}^2\left(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(v)}}\right)\right] \,. \tag{6.22}$$

Proof. First, recall the Mercer's decomposition of a kernel k w.r.t. the probability measure $\mathbb{P}_Y$.

$$k(y, y') = \sum_{q=0}^{+\infty} \lambda_q \phi_q(y) \phi_q(y')$$

Consider now the random variable W such that $W = \sum_{q=0}^{+\infty} Y^{[q]}$ where $Y^{[q]} = \sqrt{\lambda_q}\phi_q(Y)$. Note that, since the functions ϕ_q are orthogonal in $L^2_{\mathbb{P}_Y}(\mathcal{Y})$ and using the absolute convergence of the

series, we have

$$\begin{aligned}
\mathrm{Var}(W) &= \sum_{q=0}^{+\infty} \lambda_q \,\mathrm{Var}(\phi_q(Y)) \\
&= \sum_{q=0}^{+\infty} \lambda_q \mathbb{E}[\phi_q(Y)\phi_q(Y)] - \sum_{q=0}^{+\infty} \lambda_q \mathbb{E}[\phi_q(Y)\phi_q(\tilde{Y})] \\
&= \mathbb{E}\left[\sum_{q=0}^{+\infty} \lambda_q \phi_q(Y)\phi_q(Y)\right] - \mathbb{E}\left[\sum_{q=0}^{+\infty} \lambda_q \phi_q(Y)\phi_q(\tilde{Y})\right] \\
&= \mathbb{E}[k(Y,Y)] - \mathbb{E}[k(Y,Y')] \,.
\end{aligned} \tag{6.23}$$

Remark also that, for $u \in \mathcal{P}$

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}^{(u)}}\left[\mathrm{MMD}^2\left(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(u)}}\right)\right] &= \mathbb{E}_{\mathbf{X}^{(u)}}[\|m_{\mathbb{P}_{Y|\mathbf{X}^{(u)}}}\|^2_{\mathcal{H}} - \|m_{\mathbb{P}_Y}\|^2_{\mathcal{H}}] \\
&= \mathbb{E}_{\mathbf{X}^{(u)}}[\mathbb{E}_{U,U'\sim\mathbb{P}_{Y|\mathbf{X}^{(u)}}}[k(U,U')]] - \mathbb{E}_{U,U'\sim\mathbb{P}_Y}[k(U,U')] \\
&= \mathbb{E}_{\mathbf{X}^{(u)}}\left[\mathbb{E}_{U,U'\sim\mathbb{P}_{Y|\mathbf{X}^{(u)}}}\left[\sum_{q=0}^{+\infty} \lambda_q \phi_q(U)\phi_q(U')\right]\right] \\
&\quad - \mathbb{E}_{U,U'\sim\mathbb{P}_Y}\left[\sum_{q=0}^{+\infty} \lambda_q \phi_q(U)\phi_q(U')\right] \,.
\end{aligned} \tag{6.24}$$

Since the convergence of the series is absolute, we can switch the expectations and the summations.

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}^{(u)}}\left[\mathrm{MMD}^2\left(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(u)}}\right)\right] &= \sum_{q=0}^{+\infty} \lambda_q \mathbb{E}_{\mathbf{X}^{(u)}}[\mathbb{E}_{U,U'\sim\mathbb{P}_{Y|\mathbf{X}^{(u)}}}[\phi_q(U)\phi_q(U')]] - \lambda_q \mathbb{E}_{U,U'\sim\mathbb{P}_Y}[\phi_q(U)\phi_q(U')] \\
&= \sum_{q=0}^{+\infty} \lambda_q \,\mathrm{Var}(\mathbb{E}[\phi_q(Y)|\mathbf{X}^{(u)}]) \,.
\end{aligned} \tag{6.25}$$

Moreover, using the classical Sobol'-Hoeffding decomposition for each $Y^{[q]} = \sqrt{\lambda_q}\phi_q(Y)$ raises

$$\begin{aligned}
\mathrm{Var}(W) &= \sum_{q=0}^{+\infty} \lambda_q \,\mathrm{Var}(\phi_q(Y)) \\
&= \sum_{q=0}^{+\infty} \sum_{u\in\mathcal{P}} \sum_{v\subseteq u} (-1)^{|u|-|v|} \,\mathrm{Var}(\mathbb{E}[Y^{[q]}|\mathbf{X}^{(v)}]) \\
&= \sum_{u\in\mathcal{P}} \sum_{v\subseteq u} (-1)^{|u|-|v|} \sum_{q=0}^{+\infty} \lambda_q \,\mathrm{Var}(\mathbb{E}[\phi_q(Y)|\mathbf{X}^{(v)}]) \,.
\end{aligned} \tag{6.26}$$

The lemma follows by considering the two possible expansions of $\mathrm{Var}(W)$. □

Lemma 6 shows that the MMD-based sensitivity index generalizes the Sobol' one in the sense that it measures not only the variance of the conditional expectation of the output variable Y, but

on the RKHS embedding of the output through the feature map $y \mapsto \sum_{q\geq 0} \sqrt{\lambda_q}\phi_q(y)\phi_q(.)$. It is now possible to define normalized MMD-based sensitivity indices

$$\beta_u^k = \frac{\mathrm{MMD}_u^2}{\mathrm{MMD}_{\mathrm{tot}}^2} , \tag{6.27}$$

β_u^k is thus the normalized MMD-based sensitivity index associated to the subset of input u. Total MMD-based sensitivity indices are also defined in the same fashion as for the total Sobol' indices:

$$\beta_{-u}^k = \sum_{v\in\mathcal{P}, u\cap v\neq\emptyset} \beta_v^k = 1 - \frac{\mathbb{E}_{\mathbf{X}^{(-u)}}[\mathrm{MMD}^2(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X}^{(-u)}})]}{\mathrm{MMD}_{\mathrm{tot}}^2} . \tag{6.28}$$

Notice that β_{-u}^k is not equal to β_v^k with $v = -u = \bar{u}$. We have from Lemma 6 the normalization property $\sum_{u\in\mathcal{P}} \beta_u^k = 1$.

In the same spirit as the Sobol' indices, it is possible to define pick-freeze estimator of the MMD-based sensitivity index of the first order

$$\widehat{\beta}_{i,m}^k = \frac{\langle k(Y,\tilde{Y}_i) - k(Y,\tilde{Y})\rangle_m}{\langle k(Y,Y) - k(Y,\tilde{Y})\rangle_m} , \tag{6.29}$$

where $\tilde{Y}_i = g(\tilde{\mathbf{X}}_i)$ where $\tilde{\mathbf{X}}_i$ is defined the same way as in Lemma 3. The MMD-based sensitivity index of the total order writes

$$\widehat{\beta}_{-i,m}^k = 1 - \frac{\langle k(Y,\tilde{Y}_{-i}) - k(Y,\tilde{Y})\rangle_m}{\langle k(Y,Y) - k(Y,\tilde{Y})\rangle_m} . \tag{6.30}$$

In order to define β^k indices on the seismic fragility curves, define $\mathcal{F} = \{a \to \Psi(a,\mathbf{x}), \mathbf{x}\in\mathcal{X}\}$ the space of all possible fragility curves given the variability of the mechanical parameters $\mathbf{x}$. We thus have to define a positive definite kernel on $\mathcal{F}\times\mathcal{F}$, $(\Psi_1,\Psi_2) \to k_{\mathcal{F}}(\Psi_1,\Psi_2)$ for $\Psi_1,\Psi_2\in\mathcal{F}$. According to Ferraty and Vieu, 2006, let $\Delta(\cdot,\cdot)$ be a semi-metric (i.e. it satisfies the axioms of the metric but not necessarily the triangle inequality) defined on the functional space $\mathcal{F}\times\mathcal{F}$, a kernel associated to $\mathcal{F}$ can be defined as $k_{\mathcal{F}}(\Psi_1,\Psi_2) = k(\Delta(\Psi_1,\Psi_2))$ where k is acting on $\mathbb{R}$. For the sake of notations simplicity, the kernel acting on the functional space $\mathcal{F}$ will be denoted by k. For our application, we will choose the so-called Gaussian kernel with squared L^2 norm:

$$k(\Psi_1,\Psi_2) = \exp\left(-\frac{\|\Psi_1-\Psi_2\|^2_{L^2_h(\mathcal{A})}}{2\ell^2}\right) , \tag{6.31}$$

where ℓ is a hyperparameter of the kernel. We will choose ℓ as the variance of the seismic fragility curve in L^2 norm:

$$2\ell^2 = \mathbb{E}\left[\|\Psi_{\mathbf{X}} - \Psi_{\tilde{\mathbf{X}}}\|^2_{L^2_h(\mathcal{A})}\right] ,$$

where $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are two independent random vectors with same probability distribution.

6.4 Kriging-based global sensitivity indices estimation

6.4.1 Posterior distribution of the global sensitivity indices

In this section, we will present an approach inspired from Le Gratiet, 2013 to perform global sensitivity indices estimation on seismic fragility curves using a kriging surrogate model. Indeed, a first natural approach is to plug the fragility curve estimator $\Psi^{(1)}$ into the pick-freeze estimator of the aggregated Sobol' indices (or MMD-based sensitivity indices). Given $(a_t)_{1\leq t\leq T}$ a regular grid of intensity measure of size T, we propose the following *plug-in* estimator of the Sobol' and MMD-based sensitivity indices:

$$\widehat{S}^{FC}_{i,m,n} = \frac{\sum_{t=1}^{T} \left\langle \Psi^{(1)}(a_t,\mathbf{X})\Psi^{(1)}(a_t,\tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t) - \left\langle \Psi^{(1)}(a_t,\mathbf{X})\right\rangle_m \left\langle \Psi^{(1)}(a_t,\tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t)}{\sum_{t=1}^{T} \left\langle \Psi^{(1)}(a_t,\mathbf{X})^2\right\rangle_m p_A(a_t) - \left\langle \Psi^{(1)}(a_t,\mathbf{X})\right\rangle_m^2 p_A(a_t)},$$

$$\widehat{T}^{FC}_{i,m,n} = 1 - \frac{\sum_{t=1}^{T} \left\langle \Psi^{(1)}(a_t,\mathbf{X})\Psi^{(1)}(a_t,\tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t) - \left\langle \Psi^{(1)}(a_t,\mathbf{X})\right\rangle_m \left\langle \Psi^{(1)}(a_t,\tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t)}{\sum_{t=1}^{T} \left\langle \Psi^{(1)}(a_t,\mathbf{X})^2\right\rangle_m p_A(a_t) - \left\langle \Psi^{(1)}(a_t,\mathbf{X})\right\rangle_m^2 p_A(a_t)}, \tag{6.32}$$

$$\widehat{\beta}^{k}_{i,m} = \frac{\left\langle k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}}_i)\right) - k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}{\left\langle k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\mathbf{X})\right) - k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}, \tag{6.33}$$

$$\widehat{\beta}^{k}_{-i,m} = 1 - \frac{\left\langle k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}}_{-i})\right) - k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}{\left\langle k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\mathbf{X})\right) - k\left(\Psi^{(1)}(\cdot,\mathbf{X}),\Psi^{(1)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m},$$

where $\Psi^{(1)}$ is the seismic fragility curve estimator using a GP surrogate defined in Chapter 5, Section 5.5.

However, given the Bayesian interpretation of the GP regression framework (see Chapter 5, Section 5.3.4), it is of particular interest to study the posterior distribution of the sensitivity indices by considering the posterior distribution of the seismic fragility curve $\Psi^{(2)}$ (see Chapter 5, Section 5.5), itself obtained using the posterior distribution of the kriging surrogate. In practice, the posterior distribution of the sensitivity indices is obtained using Q realizations $(\Psi^{(2)}_q)_{1\leq q\leq Q}$ of the seismic fragility curve, this raises the following estimators:

$$\tilde{S}^{FC}_{i,m,n,q} = \frac{\sum_{t=1}^{T} \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\Psi_q^{(2)}(a_t,\tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t) - \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\right\rangle_m \left\langle \Psi_q^{(2)}(a_t,\tilde{\mathbf{X}}_i)\right\rangle_m p_A(a_t)}{\sum_{t=1}^{T} \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})^2\right\rangle_m p_A(a_t) - \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\right\rangle_m^2 p_A(a_t)},$$

$$\tilde{T}^{FC}_{i,m,n,q} = 1 - \frac{\sum_{t=1}^{T} \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\Psi_q^{(2)}(a_t,\tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t) - \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\right\rangle_m \left\langle \Psi_q^{(2)}(a_t,\tilde{\mathbf{X}}_{-i})\right\rangle_m p_A(a_t)}{\sum_{t=1}^{T} \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})^2\right\rangle_m p_A(a_t) - \left\langle \Psi_q^{(2)}(a_t,\mathbf{X})\right\rangle_m^2 p_A(a_t)}. \tag{6.34}$$

$$\tilde{\beta}^k_{i,q} = \frac{\left\langle k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}}_i)\right) - k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}{\left\langle k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\mathbf{X})\right) - k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}, \tag{6.35}$$

$$\tilde{\beta}^k_{-i,q} = 1 - \frac{\left\langle k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}}_{-i})\right) - k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}{\left\langle k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\mathbf{X})\right) - k\left(\Psi_q^{(2)}(\cdot,\mathbf{X}),\Psi_q^{(2)}(\cdot,\tilde{\mathbf{X}})\right)\right\rangle_m}.$$

Moreover, in order to take into account the uncertainty of the Monte-Carlo estimation of the Sobol' indices, we draw, for $b = 1,\dots,B$, the random variables $(u_b(j))_{1\leq j\leq m}$ with equiprobability and with replacement in $\{1,\dots,m\}$ and replace the pick-freeze Monte-Carlo sampling dataset $(\mathbf{X}_j,\tilde{\mathbf{X}}_{i,j},\tilde{\mathbf{X}}_{-i,j})_{1\leq j\leq m}$ by $(\mathbf{X}_{u_b(j)},\tilde{\mathbf{X}}_{i,u_b(j)},\tilde{\mathbf{X}}_{-i,u_b(j)})_{1\leq j\leq m}$. We thus obtain a sample of size $Q\times B$ of aggregated Sobol' indices $(\tilde{S}^{FC}_{i,m,n,q,b})_{\substack{1\leq q\leq Q,\\ 1\leq b\leq B}}$. This sample allows us to quantify the uncertainty of S_i^{FC} coming from the kriging metamodel uncertainty and the pick-freeze Monte-Carlo uncertainty. The posterior distribution estimation of S_i^{FC} is presented in Algorithm 4. The same procedure can be applied to the total order aggregated Sobol' index and the MMD-based sensitivity indices. We can also estimate the part of variance of S_i^{FC} coming from the Monte-Carlo approximation and the part related to the kriging metamodel uncertainty. The part of variance related to the metamodeling writes:

$$\widehat{\sigma}^2_{G_n}(\tilde{S}^{FC}_{i,m,n}) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{Q-1}\sum_{q=1}^{Q}\left(\tilde{S}^{FC}_{i,m,n,q,b} - \left\langle \tilde{S}^{FC}_{i,m,n,b}\right\rangle_Q\right)^2, \tag{6.36}$$

where $\left\langle \tilde{S}^{FC}_{i,m,n,b}\right\rangle_Q = \frac{1}{Q}\sum_{q=1}^{Q}\tilde{S}^{FC}_{i,m,n,q,b}$. Furthermore, it is also possible to evaluate the part of the variance due to Monte-Carlo approximation of the aggregated Sobol' indices:

$$\widehat{\sigma}^2_{\mathrm{MC}_m}(\tilde{S}^{FC}_{i,m,n}) = \frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{S}^{FC}_{i,m,n,q,b} - \left\langle \tilde{S}^{FC}_{i,m,n,q}\right\rangle_B\right)^2, \tag{6.37}$$

where $\left\langle \tilde{S}^{FC}_{i,m,n,q} \right\rangle_B = \frac{1}{B}\sum_{b=1}^{B} \tilde{S}^{FC}_{i,m,n,q,b}$. Again, this computation is feasible for the MMD-based sensitivity indices as well. Following Le Gratiet, 2013, we can use these two variances as a rationale for choosing the number of Monte Carlo samples m and the number of mechanical simulations of the structure n. Indeed, when $\widehat{\sigma}^2_{\mathrm{MC}_m}(\tilde{S}^{FC}_{i,m,n}) \approx \widehat{\sigma}^2_{G_n}(\tilde{S}^{FC}_{i,m,n})$ the Monte Carlo and the kriging metamodel errors have contributions of the same order into the estimation error of the aggregated Sobol' indices. Remark that these variances are defined for each input parameter and each order of the aggregated Sobol' indices. A compromise has to be made for choosing which order and input parameter the engineer must consider.

Algorithm 4 Estimation of the posterior distribution of S_i^{FC}

1. Compute the posterior distribution $(G|\mathcal{D}_n)$ using Equation 5.20
2. Generate using Monte-Carlo the samples $(\mathbf{X}_j, \tilde{\mathbf{X}}_{i,j})_{1\leq j\leq m}$ with respect to the probability measure $\mathbb{P}_{\mathbf{X}}$
3. Set Q the number of realizations of the GP posterior distribution and B the number of bootstrap samples
4. For $1 \leq q \leq Q$
   (a) Sample $G_{n,q}(\mathbf{D})$ where $\mathbf{D} = (a_t, \mathbf{X}_j)_{1\leq t\leq T, 1\leq j\leq m} \cup (a_t, \tilde{\mathbf{X}}_{i,j})_{1\leq t\leq T, 1\leq j\leq m}$.
   (b) For $1 \leq b \leq B$
      i. Sample with replacement in $\{1, \ldots, m\}$ the bootstrap indices $(u_b(j))_{1\leq j\leq m}$ and then define the bootstrap sample $\mathbf{D}^b = (a_t, \mathbf{X}_{u_b(j)})_{1\leq t\leq T, 1\leq j\leq m} \cup (a_t, \tilde{\mathbf{X}}_{i,u_b(j)})_{1\leq t\leq T, 1\leq j\leq m}$
      ii. Compute $\tilde{S}^{FC}_{i,m,n,q}$ using $G_{n,q}(\mathbf{D}^b)$

6.4.2 Sampling realizations of Gaussian process posterior distribution on large samples of evaluation points

The previous section provides a methodology to estimate the posterior distribution of global sensitivity indices using the posterior distribution of a kriging surrogate. Nevertheless, the size of the sample needed is $2mT$ for the Sobol' indices and $3mT$ for the MMD-based indices. In our applications, the sample size of the evaluation set where to sample realizations of the GP surrogate is thus close to 10^6, which makes the use of the Cholesky method computationally intractable. We will thus use the technique of kriging conditioning coupled with Nyström approximation of a GP (see Chapter 5, Section 5.4.2). We have for that matter the following lemma (see Chilès and Delfiner, 2012, Section 7.3, Le Gratiet, 2013, Section 6.4.3) adapted to the heteroskedastic noisy case.

Lemma 7. *(Sampling Gaussian process posterior distribution by kriging conditioning) Consider a learning sample $\mathcal{D}_n = (\mathbf{x}_i, y(\mathbf{x}_i))_{1\leq i\leq n}$ and the following GP regression model:*

$$Y(\mathbf{x}) = G(\mathbf{x}) + \varepsilon(\mathbf{x}) \ ,$$

where $\varepsilon(\mathbf{x}) \sim \mathcal{N}(0, \sigma_\varepsilon(\mathbf{x})^2)$, G is a zero-mean GP with covariance function Σ. Denote the following Gaussian process:

$$\tilde{G}_n(\mathbf{x}) = m_n(\mathbf{x}) - \tilde{m}_n(\mathbf{x}) + G(\mathbf{x}) \ , \tag{6.38}$$

where $m_n(\mathbf{x})$ is the predictive mean of G_n (5.20), $\tilde{m}_n$ is the predictive mean of the GP G_n where the output observations vector y^n is replaced by a virtual output observations vector $\tilde{y}^n$ sampled from the unconditionned GP Y:

$$\tilde{m}_n(\mathbf{x}) = c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}\tilde{y}^n \ , \tag{6.39}$$

where Δ_n is the diagonal matrix built from the $(\sigma_\varepsilon(\mathbf{x}_i)^2)_{1\leq i\leq n}$. Then, we have the following equality in distribution between the two Gaussian processes

$$\tilde{G}_n \overset{\mathcal{L}}{=} G_n \ ,$$

where $G_n \sim (G|\mathcal{D}_n)$.

Proof. The proof consists in proving that the Gaussian process $\tilde{G}_n$ such that

$$\tilde{G}_n(\mathbf{x}) = m_n(\mathbf{x}) - \tilde{m}_n(\mathbf{x}) + G(\mathbf{x})$$

has the same mean and covariance function as the Gaussian process $G_n \sim (G|\mathcal{D}_n)$. Remark that

$$\begin{aligned} \mathbb{E}[\tilde{G}_n(\mathbf{x})] &= m_n(\mathbf{x}) - c(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}\mathbb{E}[\tilde{y}^n] + \mathbb{E}[G(\mathbf{x})] \\ &= m_n(\mathbf{x}) \ , \end{aligned} \tag{6.40}$$

since $\mathbb{E}[\tilde{y}^n] = 0_{\mathbb{R}^n}$ and $\mathbb{E}[G(\mathbf{x})] = 0$. The mean of $\tilde{G}_n(\mathbf{x})$ is equal to the one of $G_n(\mathbf{x})$ for all $\mathbf{x}$. Now, we have to check that the covariance function of $\tilde{G}_n$ is equal to the covariance function of $G(\mathbf{x})$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. By bilinearity of the covariance we have

$$\mathrm{Cov}(\tilde{G}_n(\mathbf{x}), \tilde{G}_n(\mathbf{x}')) = \mathrm{Cov}(\tilde{m}_n(\mathbf{x}), \tilde{m}_n(\mathbf{x}')) - 2\,\mathrm{Cov}(G(\mathbf{x}), \tilde{m}_n(\mathbf{x}')) + \mathrm{Cov}(G(\mathbf{x}), G(\mathbf{x}')) \ .$$

First, we have the equality

$$\mathrm{Cov}(G(\mathbf{x}), G(\mathbf{x}')) = \Sigma(\mathbf{x}, \mathbf{x}') \ .$$

Second, we have

$$\begin{aligned}\mathrm{Cov}(\tilde{m}_n(\mathbf{x}), \tilde{m}_n(\mathbf{x}')) &= c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}\,\mathrm{Cov}(\tilde{y}^n, \tilde{y}^n)(\Sigma_n + \Delta_n)^{-1}c_n(\mathbf{x}') \\ &= c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}c_n(\mathbf{x}')\ ,\end{aligned} \tag{6.41}$$

since $\mathrm{Cov}(\tilde{y}^n, \tilde{y}^n) = \Sigma_n + \Delta_n$. Third, we have

$$\begin{aligned}\mathrm{Cov}(G(\mathbf{x}), \tilde{m}_n(\mathbf{x}')) &= c(\mathbf{x}')^T(\Sigma_n + \Delta_n)^{-1}\,\mathrm{Cov}(G(\mathbf{x}), \tilde{y}^n) \\ &= c_n(\mathbf{x}')^T(\Sigma_n + \Delta_n)^{-1}c_n(\mathbf{x})\ ,\end{aligned} \tag{6.42}$$

since, for $1 \leq i \leq n$, $\mathrm{Cov}(G(\mathbf{x}), G(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)) = \Sigma(\mathbf{x}, \mathbf{x}_i)$. Finally, we obtain

$$\mathrm{Cov}(\tilde{G}_n(\mathbf{x}), \tilde{G}_n(\mathbf{x}')) = \Sigma(\mathbf{x}, \mathbf{x}') - c_n(\mathbf{x})^T(\Sigma_n + \Delta_n)^{-1}c_n(\mathbf{x}')\ ,$$

which is exactly the covariance obtained from the kriging equations (5.20). $\tilde{G}_n$ and G_n have thus the same probability distribution. □

The main trick of Lemma 7 is to sample a GP $\tilde{G}_n$ conditioned by the observed data in $\mathcal{D}_n$ by sampling an unconditioned GP G with the same prior covariance function. Coupled with the Nyström approximation of G (see Chapter 5, Section 5.4.2), sampling GP realizations from an evaluation sample composed of Q datapoints have a computational complexity of $O(Q)$ compared to $O(Q^3)$ with the Cholesky method. Indeed, a main advantage of the Nyström approximation is the possibility to add new points in the evaluation set without having to recompute the decomposition of the eigensystem. Thus, it becomes computationally tractable to estimate the posterior distribution of the Sobol' indices and the MMD-based indices as presented in Section 6.4.1.

6.5 Sensitivity analysis on a single d.o.f. oscillator with nonlinear restoring force

In this section, we apply the methodology of sensitivity described in this Chapter to the single d.o.f. nonlinear oscillator presented in Chapter 5, Section 5.6. The probability distribution on the several mechanical parameters is described in Table 5.2. A homoskedastic Gaussian process is fitted using a learning sample of $n = 250$ mechanical simulations. A Monte-Carlo sample of size $m = 2.10^4$ from the probability distribution of the mechanical parameters is used to perform the pick-freeze estimation of the aggregated Sobol' indices. $Q = 200$ realizations of the Gaussian process posterior distribution and $B = 150$ bootstrap redraws have been carried out to assess both the metamodeling and the Monte-Carlo uncertainty on the Sobol' indices estimate. The eigensystem of the prior GP is approximated with Nyström method using a dataset of size 100, we keep 88 eigenvalues in order to retrieve more than 90% of the GP variance. We estimate the aggregated Sobol indices for the seismic fragility curve with failure threshold of $C = 1.3m_{z_d}$, corresponding roughly to a 80%-level quantile evaluated on the whole simulation dataset of 10^6 mechanical simulations. The measure h on A for the aggregated Sobol indices is chosen as the uniform distribution on peak ground acceleration (PGA) superior to $1\ m/s^2$ and inferior to $10\ m/s^2$ in order to focus the transition area between small and high probability of failure.

Figure 6.1 illustrates the results of the aggregated Sobol' indices. Remark that the stiffness k and the mass m have almost the same interquantile ranges and are the most influential variables on the seismic fragility curves. Indeed, the mechanical oscillator is driven by the natural pulsation

$\omega_0 = \sqrt{k/m}$, since the coefficients of variation of k and m are the same, the first-order Sobol' indices of k and m for the scalar output ω_0 is almost 0.5. Hence, the variables k and m have the same influence on the natural pulsation ω_0. We can then expect the same values of sensitivity indices for the input variables k and m.

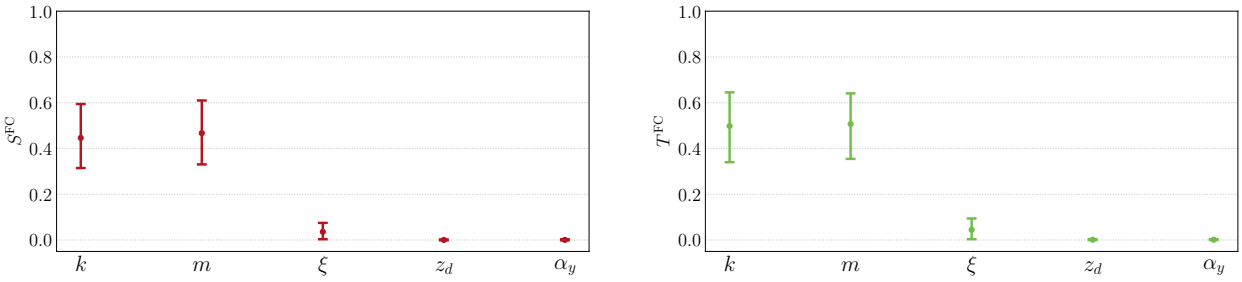


FIGURE 6.1: Aggregated Sobol' first and total order indices for a failure threshold $C = 1.3\, m_{z_d}$. The solid lines represent the interquantile ranges between the 10% and 90% quantiles for both the Monte-Carlo and the homoskedastic Gaussian process surrogate uncertainty.

The same parameters for Monte-Carlo pick-freeze estimation, the bootstrap procedures and the GP surrogate uncertainty are kept for the estimation of the MMD-based sensitivity indices. Figure 6.2 illustrates the results of the estimation of the MMD-based sensitivity indices. First, remark that the ranking of input remains the same as for the one obtained for the aggregated Sobol' indices. Moreover, we can remark that the total-order MMD-based sensitivity indices take larger values than the first order indices whereas the aggregated Sobol' indices of first and total order have very close values. Thus, the aggregated Sobol' indices do not detect interactions between the input parameters on the contrary of the MMD-based sensitivity indices. However, the MMD-based sensitivity indices suffer from a lack of interpretability and the choice of the kernel (or the choice of the lengthscale ℓ in the case of the Gaussian kernel) is still an open question for sensitivity analysis purposes (Barr and Rabitz, 2022). The output space being the space of fragility curves, it is unsure if Mercer's theorem holds in this case and thus Lemma 6 could possibly not be used in this context. However, the MMD-based sensitivity indices still have a practical interest and some properties still hold without the Mercer's theorem as shown in Barr and Rabitz, 2022.

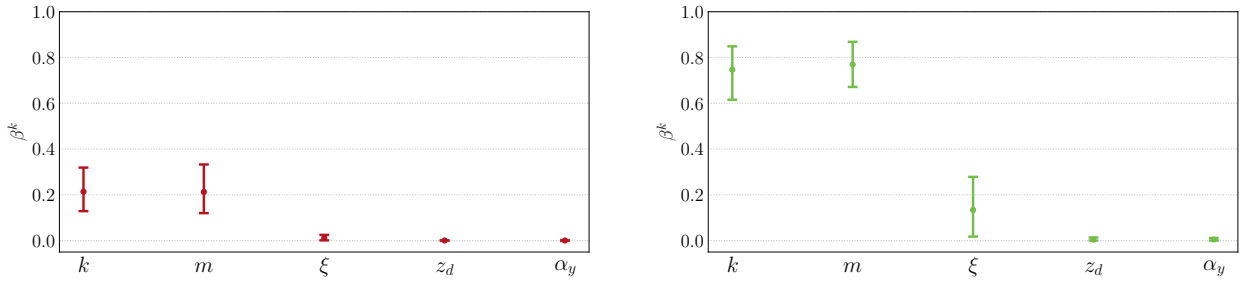


FIGURE 6.2: First and total order MMD-based sensitivity indices for a failure threshold $C = 1.3\, m_{z_d}$. The solid lines represent the interquantile range between the 10% and 90% quantiles for both the Monte-Carlo and the homoskedastic Gaussian process surrogate uncertainty.

6.6 Synthesis and conclusion

In this chapter, the main principles of sensitivity analysis have been recalled. We propose a sensitivity analysis study adapted to seismic fragility curves, by extending the classical sensitivity analysis framework based on scalar output to functional output. After reviewing the main properties of the variance-based sensitivity indices - the Sobol' indices - we choose for our sensitivity analysis study the aggregated Sobol' indices, introduced in Iooss and Le Gratiet, 2019. Since the Sobol'-Hoeffding decomposition still holds for the aggregated Sobol' indices, this is a natural extension of the Sobol' indices to functional output. Moreover, the use of kernel methods for GSA purposes is adapted to our goal of sensitivity analysis of seismic fragility curves. We use for our study a global sensitivity index based on the maximum mean discrepancy (MMD) (see e.g. Da Veiga, 2021; Barr and Rabitz, 2022). Its main advantage is its ability to deal with complex type of data such as functional ones while staying easy to estimate thanks to the pick-freeze Monte-Carlo principle. Finally, we take into account the Gaussian process surrogate uncertainty in the estimation of the sensitivity indices, the posterior distribution of the Gaussian process surrogate being propagated into the sensitivity indices estimate. For that matter one needs to sample Gaussian process realizations on very large evaluation sample. Since this is not tractable with classical Cholesky decomposition, one has to rely on more advanced methods based on kriging conditioning and Nyström approximation. The sensitivity analysis framework developed in this chapter is then applied to a single d.o.f. oscillator with nonlinear restoring force with its mechanical parameters considered tainted by epistemic uncertainties.

CHAPTER 7

Bayesian sequential design of experiments for seismic fragility curves estimation

J'avais déjà décidé de poursuivre mes recherches, car il me semblait, dans mon innocence, que l'incertitude était pire que tout, même si la vérité était terrible.

H.P. Lovecraft, *L'appel de Cthulhu*

Contents

7.1 Introduction

Since Chapter 4, we consider that the learning sample $\mathcal{D}_n = (a_i, \mathbf{x}_i, y_i)_{1 \leq i \leq n}$ is given and used to estimate the seismic fragility curves. However - when it is possible - the choice of the design of experiments (DoE) can improve greatly the accuracy of the estimation of the quantity of interest (QoI) while controlling the learning sample size n, as first shown in Chapter 3. In UQ studies, it is usually possible to control the input parameters of the computer model and thus various DoE algorithms and methods have been developed in this domain. DoE algorithms are conceived for two separate objectives :

1. One may seek to minimize the number of simulations n while keeping a target estimation accuracy;
2. One can fix a simulation budget n and build a DoE to maximize the estimation accuracy of the QoI for this simulation budget.

DoE can be *model-free*, these kinds of DoE are commonly coined *space-filling* design, since they only rely on geometrical properties of the design in the input space. A review about space-filling design is to be found in Pronzato and W. G. Müller, 2012. DoE can also be *model-oriented* in the case of parametric estimation: they rely on optimality criteria based on functionals of the Fisher information matrix of the parameters estimator (see e.g. Pronzato and Pazman, 2013, Chapter 5). Moreover, DoE can also be seen with a Bayesian eye. These Bayesian DoE procedures boil down to optimizing an expected utility function depending on the design (see Chaloner and Verdinelli, 1995).

Introduced by Villemonteix, Vazquez, and Walter, 2009; Vazquez and Bect, 2009, the stepwise uncertainty reduction (SUR) strategies are sequential design of experiments procedures based on the Bayesian nature of Gaussian processes. These algorithms aim at maximizing the estimation accuracy of the QoI for a given simulation budget n, this objective seems particularly adapted in our case. Indeed, the main advantage of SUR strategies are their flexibility to any QoI one can consider. For instance, SUR procedures have been developed for conservative estimation of excursion sets of the input space $\mathcal{X}$ in Azzimonti et al., 2016, quantile estimation in Labopin-Richard and Picheny, 2016 and probability of failure estimation in Bect, Ginsbourger, et al., 2012. SUR strategies in a multi-fidelity setting has been studied in Stroh, 2018. SUR strategies can be also used for optimization of expensive black-box functions as shown in Villemonteix, Vazquez, and Walter, 2009. A theoretical work about the convergence of SUR strategies is also proposed in Bect, Bachoc, and Ginsbourger, 2019.

We propose in this chapter a SUR strategy adapted to estimations of seismic fragility curves with a Gaussian process surrogate. The proposed SUR strategy is then applied to estimation of seismic fragility curve on a single d.o.f. oscillator with nonlinear restoring force. This chapter is organized as follows: First, we present the Bayesian decision-theoretic framework of the SUR strategy in Section 7.2, adapted to stochastic computer models. Second, we develop a SUR strategy for the estimation of seismic fragility curves in Section 7.3. Finally, the proposed DoE procedure is illustrated on a single d.o.f. oscillator in Section 7.4. The SUR strategy developed in this chapter will also be applied to an industrial test case in Chapter 8.

7.2 Bayesian decision-theoretic framework

7.2.1 Problem statement and Bayes risk

Let us consider the following nonparametric regression model for $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$:

$$y(\mathbf{x}) = g(\mathbf{x}) + \varepsilon \,, \tag{7.1}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon(\mathbf{x})^2)$ is a heteroskedastic Gaussian noise. This statistical model corresponds to the stochastic mechanical computer model usually considered in this manuscript. Our objective in this section is to estimate the function $w(\mathbf{x}; g)$ defined as:

$$w(\mathbf{x}; g) = \mathbb{E}_\varepsilon[f(y(\mathbf{x}))] = \mathbb{E}_\varepsilon[f(g(\mathbf{x}) + \varepsilon)] \,, \tag{7.2}$$

where f is a measurable and integrable function. Define $\mathcal{I}_n = (\mathbf{X}_i, y(\mathbf{X}_i))_{1\leq i\leq n}$, we will consider the following class $\mathcal{S}_n$ of admissible design strategies with a simulation budget of size n

$$\mathcal{S}_n = \{\mathbb{X}_n \in \mathcal{X}^n / \, \forall k \in \{1,\dots,n\}, \mathbf{X}_k \text{ is } \sigma(\mathcal{I}_{k-1}) - \text{measurable}\} \, , \tag{7.3}$$

where $\sigma(\mathcal{I}_0)$ is the trivial σ-algebra $\{\emptyset, \Omega\}$ (X_1 being deterministic). This class of designs corresponds to sequential non-randomized designs, the input $\mathbf{X}_1$ is determined prior to any observations of the output y and at any step k, the input $\mathbf{X}_k$ is chosen by using the information given in previous observations $\mathcal{I}_{k-1}$. Given a design $\mathbb{X}_n = (\mathbf{x}_i)_{1\leq i\leq n}$ with corresponding observed values $(y(\mathbf{x}_i))_{1\leq i\leq n}$, one can build an estimator $\widehat{w}_n$ of w. According to the Bayesian principles, one has to define a loss function ℓ between $\widehat{w}_n$ and w to measure the quality of the experimental design $\mathbb{X}_n$. The approximation error $e(\mathbb{X}_n; g)$ is thus defined by

$$e(\mathbb{X}_n; g) = \ell(w, \widehat{w}_n) \, .$$

An example of common loss function is the squared L^2 norm w.r.t. a probability measure μ on $\mathcal{X}$

$$\ell(w, \widehat{w}_n) = \|w - \widehat{w}_n\|^2_{L^2_\mu(\mathcal{X})} \, .$$

Again according to Bayesian principles, we will consider now a Gaussian process prior distribution on the regression function g. Consider the Gaussian process G defined on the probability space $(\Omega, \mathcal{B}, \mathbb{P}_0)$. The statistical model defined in Equation 7.1 now becomes

$$Y(\mathbf{x}) = G(\mathbf{x}) + \varepsilon \, . \tag{7.4}$$

The *Bayes risk* of the estimator $\widehat{w}_n$ is defined by:

$$r_B = \min_{\mathbb{X}_n \in \mathcal{S}_n} \mathbb{E}_{\mathbb{P}_0}[e(\mathbb{X}_n, G)] \, . \tag{7.5}$$

The optimal design $\mathbb{X}_n^*$ is the one minimizing the Bayes risk:

$$\mathbb{X}_n^* = \operatorname*{argmin}_{\mathbb{X}_n \in \mathcal{S}_n} \mathbb{E}_{\mathbb{P}_0}[e(\mathbb{X}_n, G)] \, . \tag{7.6}$$

From a Bayesian viewpoint, the Gaussian process G represents our initial uncertain knowledge about the regression function g. The optimal design can be seen as the one minimizing the loss averaged on our prior knowledge on the function g.

7.2.2 j-step lookahead strategies and SUR criterion

Denote by $\mathcal{F}_n$ the σ-algebra generated by $(\mathbf{x}_i, Y(\mathbf{x}_i))_{1\leq i\leq n}$ ($\mathcal{F}_0$ being the trivial σ-algebra). According to Bect, Ginsbourger, et al., 2012, the optimal design $\mathbb{X}_n^*$ can be determined using backward recursion: Denote the terminal risk

$$R_n = \mathbb{E}_{\mathbb{P}_0}[e(\mathbb{X}_n, G)|\mathcal{F}_n] \, ,$$

and define by backward induction for $0 \leq k \leq n-1$

$$R_k = \min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}[R_{k+1}|\mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] \, . \tag{7.7}$$

Remark that $\mathbb{E}_{\mathbb{P}_0}[R_{k+1}|\mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k]$ is an expectation w.r.t. the random variable $Y_k(\mathbf{x})$ following the posterior distribution $(Y(\mathbf{x})|\mathcal{F}_k)$. Then, we have $R_0 = r_B$ and the experimental design $\mathbb{X}_n^*$

defined by

$$\mathbf{x}_{k+1}^{*} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbb{P}_0}[R_{k+1}|\mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] \,, \tag{7.8}$$

is optimal in the sense of Equation 7.6. It is critical to remark that the state space $(\mathcal{X} \times \mathbb{R})^k$ of the optimization problem at step k is of dimension $k(p+1)$. As a consequence, a direct attempt to obtain the optimal design using Equation 7.8 will be computationally untractable for a simulation budget n more than a few dozen. A possible solution is to relax the optimization problem defined in 7.8 in order to obtain computationally tractable experimental design. Using Equation 7.8, the optimal design can be rewritten as follows

$$\mathbf{x}_{k+1}^{*} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}}\, E_k[\min_{\mathbf{X}_{k+2}} E_{k+1} \ldots \min_{\mathbf{X}_n} E_{n-1} R_n | \mathbf{X}_{k+1} = \mathbf{x}] \,, \tag{7.9}$$

where $E_k = \mathbb{E}_{\mathbb{P}_0}[.|\mathcal{F}_k]$. A general way to relax the optimization problem is to truncate the expansion after j terms. The resulting strategy

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}}\, E_k[\min_{\mathbf{X}_{k+2}} E_{k+1} \ldots \min_{\mathbf{X}_{\min(n,k+j)}} E_{\min(n,k+j)-1} R_n | \mathbf{X}_{k+1} = \mathbf{x}] \,, \tag{7.10}$$

is coined a *j-step lookahead strategy*. Note also that the optimal and the j-lookahead strategies boil down to define a *sampling criterion* $J_k(\mathbf{x})$ which is $\mathcal{F}_k$-measurable such that

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}}\, J_k(\mathbf{x}) \,. \tag{7.11}$$

SUR strategies consist in defining J_k as a 1-step lookahead strategy. The SUR criterion J_k then boils down to

$$J_k(\mathbf{x}) = \mathbb{E}_{\mathbb{P}_0}[e(\mathbb{X}_{k+1}, G)|\mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] \,. \tag{7.12}$$

In order to better interpret the SUR criterion, recall the definition of the approximation error $e(\mathbb{X}_{k+1}, G)$ in terms of loss function:

$$J_k(\mathbf{x}) = \mathbb{E}_{\mathbb{P}_0}[\ell(w(.\,; G), \widehat{w}_{k+1})|\mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] \,. \tag{7.13}$$

The SUR criterion is thus the loss between the QoI w and its estimator $\widehat{w}_{k+1}$ at step $k+1$ computed using a virtual observation sampled from the distribution $(Y(\mathbf{x})|\mathcal{F}_k)$, averaged over the posterior distribution of the observations $(Y(\mathbf{x})|\mathcal{F}_k)$. The SUR design $\mathbb{X}_n^{\mathrm{SUR}} = (\mathbf{x}_i^{\mathrm{SUR}})_{1\leq i\leq n}$ is then defined for $1 \leq k \leq n-1$ by

$$\mathbf{x}_{k+1}^{\mathrm{SUR}} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}}\, J_k(\mathbf{x}) \,. \tag{7.14}$$

7.3 SUR strategy for seismic fragility curve estimation

After introducing the main theoretical tools to construct SUR designs in Section 7.2, the goal of this section is to propose a SUR strategy for the estimation of seismic fragility curve.

7.3.1 Definition of the SUR criterion for seismic fragility curve estimation

We have detailed in Chapter 5, Section 5.5 how to estimate a seismic fragility curve using a Gaussian process surrogate. Consider the following statistical model

$$y(a, \mathbf{x}) = g(a, \mathbf{x}) + \varepsilon \,, \tag{7.15}$$

with $(a, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^p$ where a is the seismic intensity measure and $\mathbf{x}$ is the vector of mechanical parameters of the structure considered, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon(a,\mathbf{x})^2)$ is a Gaussian noise. $y(a,\mathbf{x}) = \log(z(a,\mathbf{x}))$ is the log engineering demand parameter of the structure. The seismic fragility curve for a failure threshold C then writes:

$$\Psi(a, \mathbf{x}; g) = \Phi\left(\frac{g(a,\mathbf{x}) - \log(C)}{\sigma_\varepsilon(a,\mathbf{x})}\right) , \tag{7.16}$$

where Φ is the c.d.f. of the standard Gaussian distribution. Again referring to Chapter 5, Section 5.5, we model the uncertainty on the regression function g by a Gaussian process prior G. Define $\mathcal{F}_n$ the σ-algebra of the observations $(A_i, \mathbf{X}_i, y(A_i, \mathbf{X}_i))_{1\leq i\leq n}$. Thanks to the kriging equations, we have the following posterior distribution $(G(a,\mathbf{x})|\mathcal{F}_n) \sim \mathcal{N}(m_n(a,\mathbf{x}), s_n(a,\mathbf{x})^2)$ for the GP surrogate. Then, we can propose the *posterior mean* as a Bayesian estimator of the seismic fragility curve

$$\begin{aligned}\widehat{\Psi}_n(a,\mathbf{x}) &= \mathbb{E}_{\mathbb{P}_0}[\Psi(a,\mathbf{x};G)|\mathcal{F}_n] \\ &= \Phi\left(\frac{m_n(a,\mathbf{x}) - \log(C)}{\sigma_n(a,\mathbf{x})}\right) ,\end{aligned} \tag{7.17}$$

where $\sigma_n(a,\mathbf{x})^2 = s_n(a,\mathbf{x})^2 + \sigma_\varepsilon(a,\mathbf{x})^2$. A natural loss function in this setup is the squared L^2 loss for a probability measure η such that $\eta = h \otimes \mathbb{P}_{\mathbf{X}}$:

$$\ell(\Psi, \widehat{\Psi}_n) = \|\Psi - \widehat{\Psi}_n\|^2_{L_\eta(\mathcal{A}\times\mathcal{X})} , \tag{7.18}$$

where h is a probability measure on the seismic intensity measure (not necessary the probability measure of observed data) and $\mathbb{P}_{\mathbf{X}}$ the probability on the mechanical parameters of the structure. Following the principles of SUR strategy detailed in Section 7.2, the SUR criterion in this case writes

$$\begin{aligned}J_k(a,\mathbf{x}) &= \mathbb{E}_{\mathbb{P}_0}[\|\Psi - \widehat{\Psi}_{k+1}\|^2_{L_\eta(\mathcal{A}\times\mathcal{X})}|A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] \\ &= \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} \left(\Psi(\alpha,\mathbf{u};G) - \widehat{\Psi}_{k+1}(\alpha,\mathbf{u})\right)^2 dh(\alpha)d\mathbb{P}_{\mathbf{X}}(\mathbf{u}) \middle| A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k\right] .\end{aligned} \tag{7.19}$$

The SUR strategy $(a_i^{\mathrm{SUR}}, \mathbf{x}_i^{\mathrm{SUR}})_{1\leq i\leq n}$ for seismic fragility curves estimation then verifies the following optimization problem

$$a_{k+1}^{\mathrm{SUR}}, \mathbf{x}_{k+1}^{\mathrm{SUR}} = \underset{a,\mathbf{x}\in\mathcal{A}\times\mathcal{X}}{\operatorname{argmin}}\ J_k(a,\mathbf{x}) , \tag{7.20}$$

7.3.2 Practical implementation of the SUR strategy

We will derive in this section the quadrature methods used to compute numerically the SUR criterion defined in Equation 7.19. First, we will rewrite the SUR criterion thanks to the following Lemma

Lemma 8. *The following equality holds*

$$J_k(a,\mathbf{x}) = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}[\Psi(\alpha,\mathbf{u};G)^2|\mathcal{F}_k] - \mathbb{E}_{\mathbb{P}_0}[\widehat{\Psi}_{k+1}(\alpha,\mathbf{u})^2|A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k]dh(\alpha)d\mathbb{P}_{\mathbf{X}}(\mathbf{u})$$

Proof. Given that we have $\sigma(A_{k+1}, \mathbf{X}_{k+1}, \mathcal{F}_k) \subset \mathcal{F}_{k+1}$ and using the property of the conditional expectation we have:

$$J_k(a, \mathbf{x}) = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[\mathbb{E}_{\mathbb{P}_0}\left[\left(\Psi(\alpha, \mathbf{u}; G) - \widehat{\Psi}_{k+1}(\alpha, \mathbf{u})\right)^2 \middle| \mathcal{F}_{k+1}\right] \middle| A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k\right] dh(\alpha) d\mathbb{P}_{\mathbf{x}}(\mathbf{u}) .$$

Since $\widehat{\Psi}_{k+1}$ is the posterior mean w.r.t. the posterior distribution $(G|\mathcal{F}_{k+1})$, we have

$$J_k(a, \mathbf{x}) = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[\mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha, \mathbf{u}; G)^2 - \widehat{\Psi}_{k+1}(\alpha, \mathbf{u})^2 \middle| \mathcal{F}_{k+1}\right] \middle| A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k\right] dh(\alpha) d\mathbb{P}_{\mathbf{x}}(\mathbf{u}) .$$

By splitting the difference, we obtain the result of the Lemma. □

Lemma 8 has a practical interest: it allows to compute the SUR sampling criterion only using a simple Monte-Carlo loop. We will now derive the explicit expression of the two conditional expectations inside the integrand. First, we have

$$\mathbb{E}_{\mathbb{P}_0}[\Psi(\alpha, \mathbf{u}; G)^2 | \mathcal{F}_k] = \mathbb{E}_{Z\sim\mathcal{N}(m_k(\alpha,\mathbf{u}), s_k(\alpha,\mathbf{u})^2)}\left[\Phi\left(\frac{Z - \log(C)}{\sigma_\varepsilon(\alpha, \mathbf{u})}\right)^2\right] . \tag{7.21}$$

Second, we have

$$\mathbb{E}_{\mathbb{P}_0}[\widehat{\Psi}_{k+1}(\alpha, \mathbf{u})^2 | A_{k+1} = a, \mathbf{X}_{k+1} = \mathbf{x}, \mathcal{F}_k] = \mathbb{E}_{Z\sim\mathcal{N}(m_k(a,\mathbf{x}), \sigma_k(a,\mathbf{x})^2)}\left[\Phi\left(\frac{m_{k+1}(\alpha, \mathbf{u}; Z) - \log(C)}{\sigma_{k+1}(\alpha, \mathbf{u}; a, \mathbf{x})}\right)^2\right] , \tag{7.22}$$

where $m_{k+1}(\alpha, \mathbf{u}; Z)$ and $\sigma_{k+1}(\alpha, \mathbf{u}; a, \mathbf{x})$ denote respectively the posterior mean and standard deviation at point $(\alpha, \mathbf{u})$ of the GP conditional on the observations at step k and on the virtual observation Z at point $(a, \mathbf{x})$. For fast computation of the two expectations, we use Gauss-Hermite quadrature (Steen, Byrne, and Gelbard, 1969): Given $(u_q)_{1\leq q\leq Q}$ the roots of the Hermite polynomial $H_Q(x)$ such that

$$H_Q(x) = (-1)^Q e^{x^2} \frac{d^Q}{dx^Q} e^{-x^2} ,$$

with their associate weights $(\omega_q)_{1\leq q\leq Q}$ given by

$$\omega_q = \frac{2^{Q-1} Q! \sqrt{\pi}}{Q^2 |H_{Q-1}(u_q)|^2} .$$

Then, we can provide the following approximation

$$\mathbb{E}_{Z\sim\mathcal{N}(\mu,\sigma^2)}[f(Z)] \approx \frac{1}{\sqrt{\pi}} \sum_{q=1}^{Q} \omega_q f(z_q) , \tag{7.23}$$

where f is a measurable function and $z_q = \mu + \sqrt{2}\sigma u_q$. The integral against the measure $\eta = h \otimes \mathbb{P}_{\mathbf{X}}$ is approximated using a Monte-Carlo sample $(\alpha_i, \mathbf{u}_i)_{1\leq i\leq N}$.

The last difficulty to tackle in order to put in practice the proposed SUR strategy is the resolution of the optimization problem defined in Equation 7.14. The heuristic method proposed in Bect, Ginsbourger, et al., 2012 is to replace the continuous optimization space $\mathcal{A} \times \mathcal{X}$ by a discrete

one composed of M candidate points $(a'_j, \mathbf{x}'_j)_{1\leq j\leq M}$. The optimization problem then simplifies to

$$\begin{aligned} j^* &= \underset{1\leq j\leq M}{\operatorname{argmin}}\, J_k(a'_j, \mathbf{x}'_j) \\ a^{\mathrm{SUR}}_{k+1}, \mathbf{x}^{\mathrm{SUR}}_{k+1} &= a'_{j^*}, \mathbf{x}'_{j^*} \ . \end{aligned} \tag{7.24}$$

Algorithm 5 sums the main results of this section and describes the practical implementation of a SUR strategy for seismic fragility curves estimation. Remark that the result of Lemma 8 has a computational interest. Indeed, the first expectation can be computed outside the loop on the candidate points, allowing for a better computation time.

Algorithm 5 SUR strategy for seismic fragility curve estimation

Requirements

1. A set of observations $(a_i, \mathbf{x}_i, y(a_i, \mathbf{x}_i))_{1\leq i\leq k}$
2. A sample of candidate points $(a'_i, \mathbf{x}'_i)_{1\leq i\leq M}$
3. A Monte-Carlo sample $(\alpha_i, \mathbf{u}_i)_{1\leq i\leq N}$
4. Gauss-Hermite quadrature points $(u_q)_{1\leq q\leq Q}$ and their associated weights $(\omega_q)_{1\leq q\leq Q}$

Procedure:

1. Compute the kriging mean prediction $m_k(\alpha_i, \mathbf{u}_i)$ and variance $s_k(\alpha_i, \mathbf{u}_i)^2$ (see Equation 5.20) for each datapoint in the Monte-Carlo sample
2. Compute $z_{k,q}(\alpha_i, \mathbf{u}_i) = m_k(\alpha_i, \mathbf{u}_i) + \sqrt{2}s_k(\alpha_i, \mathbf{u}_i)u_q$ for $1 \leq q \leq Q$
3. Compute
$$J_{k,1} = \frac{1}{N\sqrt{\pi}} \sum_{i=1}^{N}\sum_{q=1}^{Q} \Phi\left(\frac{z_{k,q}(\alpha_i, \mathbf{u}_i) - \log(C)}{\sigma_\varepsilon(\alpha_i, \mathbf{u}_i)}\right)^2$$
4. For each candidate point $(a'_j, \mathbf{x}'_j)$
   (a) Compute $\tilde{z}_{k,q}(a'_j, \mathbf{x}'_j) = m_k(a'_j, \mathbf{x}'_j) + \sqrt{2}\sigma_k(a'_j, \mathbf{x}'_j)u_q$
   (b) Compute
$$J_{k,2}(a'_j, \mathbf{x}'_j) = \frac{1}{N\sqrt{\pi}} \sum_{i=1}^{N}\sum_{q=1}^{Q} \Phi\left(\frac{m_{k+1}(\alpha_i, \mathbf{u}_i; \tilde{z}_{k,q}(a'_j, \mathbf{x}'_j)) - \log(C)}{\sigma_{k+1}(\alpha_i, \mathbf{u}_i; a'_j, \mathbf{x}'_j)}\right)^2$$
   (c) Compute the SUR criterion $J_k(a'_j, \mathbf{x}'_j) = J_{k,1} - J_{k,2}(a'_j, \mathbf{x}'_j)$
5. Find $j^* = \operatorname{argmin}_j J_k(a'_j, \mathbf{x}'_j)$ and set $a^{\mathrm{SUR}}_{k+1}, \mathbf{x}^{\mathrm{SUR}}_{k+1} = a'_{j^*}, \mathbf{x}'_{j^*}$

7.4 Estimation of seismic fragility curve of a single d.o.f. oscillator using a SUR strategy

In this section, we apply the SUR strategy developed in this chapter to the single d.o.f. nonlinear oscillator presented in Chapter 5, Section 5.6 and Chapter 6, Section 6.5. The probability distribution of the different mechanical parameters is described in Table 5.2. We propose a homoskedastic GP surrogate with a zero mean function and a Matérn $5/2$ tensorized covariance function. In this case, we will only fit the surrogate on the subset of variables (PGA, k, m). Indeed, the proposed SUR strategy requires many candidate points and this number increases exponentially with the dimension. We thus select the most influential variables for building the GP surrogate. Remark also that in SUR strategies, the hyperparameters estimation as well as their uncertainties are not taken into account. By consequence, we take a homoskedastic noise in order to reduce the number of hyperparameters.

The failure threshold considered in this section is $C = 2.1 m_{z_d}$, this corresponds roughly to the 90% quantile of the output variable for the whole dataset of 10^5 mechanical simulations. The performance of the proposed SUR strategy is evaluated by comparing the bias b_n of the estimator $\widehat{\Psi}_n$ to a reference fragility curve Ψ_{ref} estimated by kernel ridge regression with Gaussian kernel on a dataset of 10^4 mechanical simulations.

$$b_n = \int_{\mathcal{A}\times\mathcal{X}} (\widehat{\Psi}_n(\alpha, \mathbf{u}) - \Psi_{\text{ref}}(\alpha, \mathbf{u}))^2 dh(\alpha) d\mathbb{P}_{\mathbf{x}}(\mathbf{u}) \, . \tag{7.25}$$

Moreover, the performance is also assessed with the loss ℓ defined in Equation 7.18. For this application, we choose for the measure h the uniform distribution on the segment $[a_0, a_1]$ where $a_0 = 1 \ m/s^2$ and $a_1 = 20 \ m/s^2$. This choice allows targeting evenly the transition between low and high probabilities of failures. The SUR procedure starts by choosing 10 synthetic seismic ground motions in the whole dataset in a stratified manner, by dividing the full dataset into 10 sets of equal size using the $(j/10)_{1\leq j\leq 9}$-level quantiles of the PGA. The mechanical parameters $\mathbf{x}$ are sampled at random w.r.t. the probability distribution $\mathbb{P}_{\mathbf{X}}$. They are the first points of the design set and are used to optimize the Gaussian process hyperparameters (including σ_ε). After that, a set of candidate points $(a'_j, \mathbf{X}'_j)_{1\leq j\leq M}$ ($M = 1000$) are chosen at random in the whole dataset of 10^5 synthetic seismic ground motions. The candidate points set is built by using the same partition into 10 subdomains as for initialization. We then compute the SUR sampling criterion as described in Algorithm 5. The SUR sampling criterion is estimated with a Monte-Carlo sample of size $N = 5000$ and $Q = 12$ for the Gauss-Hermite quadrature. The GP hyperparameters are optimized every 10 iterations.

We compare 100 replications of the SUR strategy with 100 replications of Monte-Carlo sampling design (sampled in the whole dataset of 10^5 mechanical simulations). Figure 7.1 compares the log bias b_n and the log loss ℓ as a function of the training size. The bias b_n is estimated with a Monte-Carlo sample of size 5000, the loss $\ell(\Psi, \widehat{\Psi}_n)$ is estimated with the same Monte-Carlo sample and with 4000 realizations of the GP surrogate. The SUR outperforms Monte-Carlo design significantly in terms of loss both in mean values and interquantile ranges. The bias is also better for the SUR strategy, however it belongs in the interquantile range of the Monte-Carlo design. Figure 7.2 represents uncertainty propagation of the mechanical parameters on the seismic fragility curve in the same fashion as in Chapter 5, Section 5.5 for a Monte-Carlo sample and a SUR design for a training size $n = 80$. Remark that the GP surrogate uncertainty is significantly smaller (in terms of interquantile range on the seismic fragility curves) with a SUR-based design than a Monte-Carlo

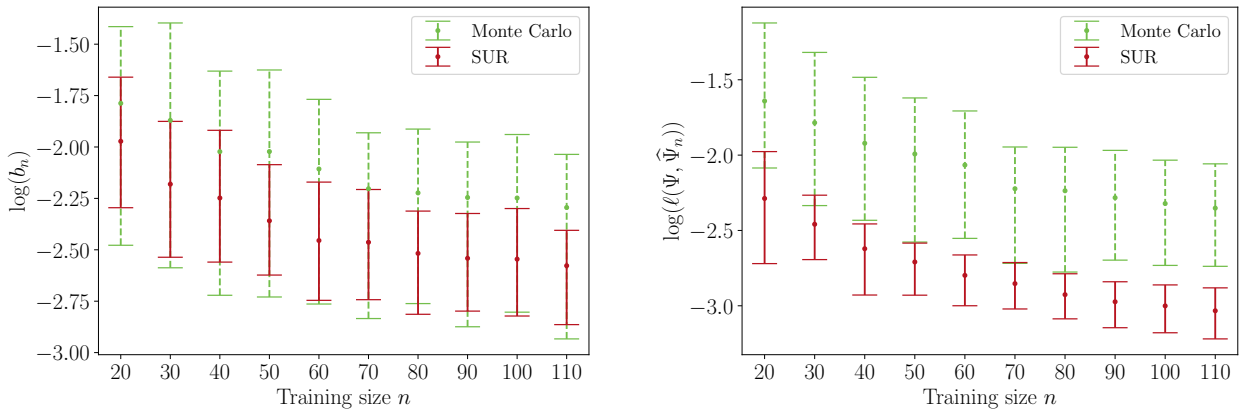


FIGURE 7.1: Comparison of the bias b_n and the loss $\ell(\Psi, \widehat{\Psi}_n)$ between 100 Monte-Carlo designs and 100 runs of the proposed SUR strategy for a failure threshold $C = 2.1m_{z_d}$. The green error bars correspond to the interquantile range between the 10% and 90% quantiles of the observed metric on the 100 replications of the Monte-Carlo design. The red error bars are the same quantities for the replications of the SUR strategy.

design.

However, one can expect less performance of the SUR strategy for a smaller value of the failure threshold C. Indeed, for a high value of C, one can guess that failure states of the mechanical structure considered happen for rare values of the input parameters $(a, \mathbf{x})$ and thus the SUR strategy will be more performant in this setting. We thus perform the same experimental campaign with a smaller failure threshold $C = 1.3m_{z_d}$, which corresponds roughly to a 80%-level quantile of the output variable on the whole dataset. Figure 7.3 illustrates the bias and the loss in this setting. One can notice that even if the SUR strategy raises better performance than Monte-Carlo design, the difference is less striking than for the higher failure threshold $C = 2.1m_{z_d}$. The use of the SUR strategy seems thus more adapted for the estimation of seismic fragility curves with high failure thresholds. Nevertheless, this validates the use of a SUR strategy for seismic fragility curves estimation.

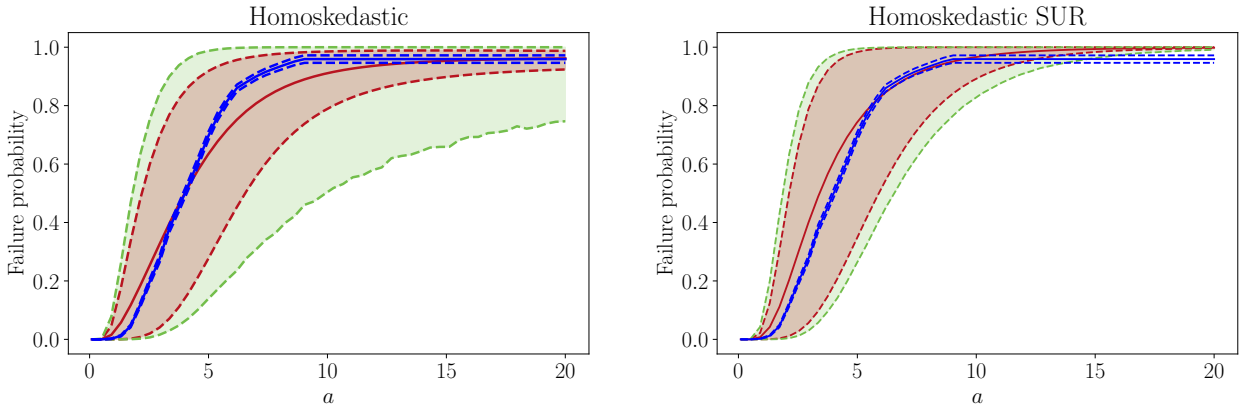


FIGURE 7.2: Comparison of a Monte-Carlo design and a SUR design for a training size $n = 80$ in terms on uncertainty propagation on the seismic fragility curves in the same fashion as in Figure 5.11.

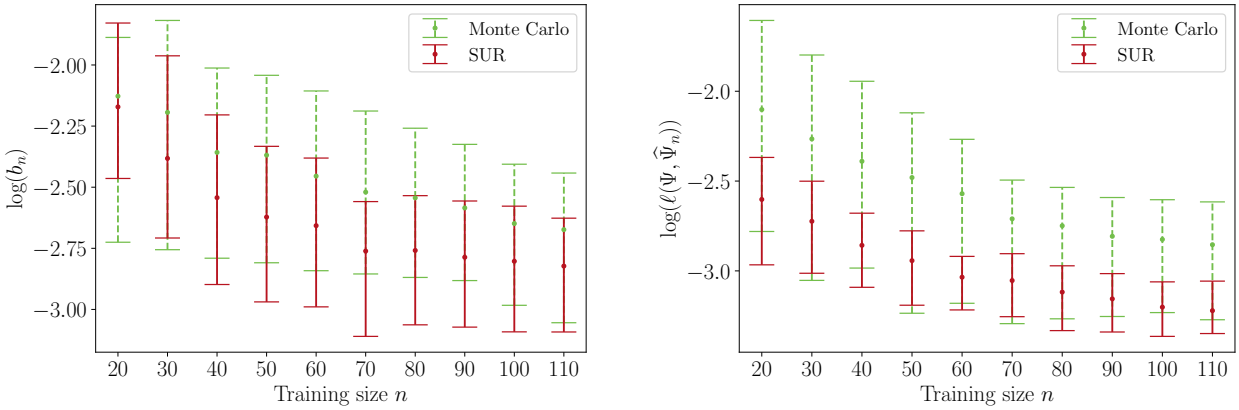


FIGURE 7.3: Comparison of the bias b_n and the loss $\ell(\Psi, \widehat{\Psi}_n)$ between 100 Monte-Carlo designs and 100 runs of the proposed SUR strategy for a failure threshold $C = 1.3m_{z_d}$. The green error bars correspond to the interquantile range between the 10% and 90% quantiles of the observed metric on the 100 replications of the Monte-Carlo design. The red error bars are the same quantities for the replications of the SUR strategy.

7.5 Discussion and conclusion

In this chapter, we review the main methodology of SUR strategies. These DoE procedures are Bayesian sequential design of experiments procedures based on the Bayesian viewpoint of GP regression. After defining the SUR sampling criterion in the general setting of noisy computer models, we derive a SUR sampling criterion for seismic fragility curve in the presence of epistemic uncertainties on the mechanical parameters. The methodology developed in this chapter is then illustrated on a single d.o.f. oscillator with nonlinear restoring force. However, there is still room for improvements on this subject. First, the SUR strategy does not take into account the uncertainty on the hyperparameters of the GP surrogate. Even if a full Bayesian approach as described in Chapter 5, Section 5.4.5 could fit in the SUR strategy framework, the computation time of the SUR sampling criterion might become untractable (as pointed out in Stroh, 2018, Chapter 2). Second, the brute-force optimization based on a set of candidate points described in Equation 7.24 suffers from the curse of dimensionality, which makes SUR strategy computationally cumbersome in a high-dimensional setting. Despite these limitations, the numerical experiments carried out on the nonlinear single d.o.f. oscillator show significant performance gains of the SUR strategy for seismic fragility curves estimation and motivate further research on this topic.

CHAPTER 8

Application to a piping system mock-up of a pressurized water reactor

Contents

8.1 Introduction and motivations

Seismic design of nuclear piping is a complicated problem, since the length of piping systems in nuclear power plants can be up to 100 km and that they play a key role in nuclear safety (e.g. core cooling). Their designs have to satisfy a compromise between flexibility for supporting thermal expansion and rigidity to support high level seismic loading. Thus, laboratory experiments have to be carried out to assess the seismic robustness of piping systems. These experiments are based on mock-ups which represent the earthquake-sensitive parts of piping systems. The mock-ups are then put on shaking table, in order to analyze their behaviors under seismic loading. Since it is too costly to reproduce real experiments on shaking tables, computer models of the mock-ups based on Finite Elements (FE) are built to simulate their temporal behaviors under seismic excitation and calibrated using real experiments. However, in such experimental campaigns the

uncertainties, for instance on the boundary conditions and the material physical parameters of the mock-ups, are generally not assessed and can be considered as epistemic uncertainties.

The objective of this chapter is to propose an UQ study of a computer model of a piping mock-up tainted by epistemic and random uncertainties. The epistemic uncertainties concern the mechanical parameters and the boundary conditions of the model whereas the random uncertainty comes from the inherent randomness of the seismic loading. The epistemic uncertainties are modeled by a random vector and the output variable of the computer model corresponds to an engineering demand parameter of the mock-up. The UQ study is goal-oriented, as it is performed on the seismic fragility curve of the piping structure. The aim of this UQ study is to better understand how the epistemic uncertainties - that are *reducible* in the sense of Chapter 4, Section 4.2.3 - affects the seismic fragility curve of the mock-up, in order to have a further insight of the mock-up behavior under seismic excitations.

This chapter is organized as follows. Section 8.2 describes the piping mock-up considered and the sources of uncertainties. Section 8.3 defines and motivates the probabilistic model put on the input parameters of the piping mock-up computer model. Then, Section 8.4, 8.5, 8.6 present the results of the three methodologies developed in the three previous chapters (i.e., surrogate modeling based on Gaussian process regression and uncertainty propagation, global sensitivity analysis on the seismic fragility curves, Bayesian sequential design of experiments for seismic fragility curves estimation). The most important results and remarks are gathered in Section 8.7.

8.2 Description of the ASG piping mock-up

The ASG (*Alimentation de Secours Général*) piping system is the French acronym for the emergency feedwater system (EFWS) in the pressurized water reactor (PWR) nomenclature. According to Jacquemain, 2015, Chapter 2, the EFWS is used to maintain the level of water in the secondary line of the steam generator of the PWR, it helps to cool the reactor cooling system (RCS) in the event the main feedwater system (MFWS) is not available. A schematic diagram of the different auxiliary systems of a French PWR is given in Figure 8.1.

As motivated in Section 8.1, the behavior of a piping system such as the ASG system during seismic excitations has to be studied thoroughly for nuclear safety purposes. Thus, the test case developed in this chapter is based on a simplified and scale mock-up of a section of the ASG piping system and is part of the ASG experimental program detailed in Touboul, Sollogoub, and Blay, 1999. The mock-up is a 114.3 mm outside diameter and 8.56 mm thickness pipe with a 0.47 elbow characteristic parameter, in carbon steel TU42C, filled with water without pressure. It contains three elbows and a mass modeling a valve (120 kg) which corresponds to more than 30% of the specimen total mass. As shown in Figure 8.2b, one end of the mock-up is clamped whereas the other is supported by a guide in order to prevent the displacements in the X and Y directions. Additionally, a rod is placed on the top of the specimen in order to limit the mass displacements in the Z direction.

The experimental campaign was performed on the shaking table AZALÉE of the CEA EMSI Laboratory and consisted of temporal dynamic loading in the X direction. A view of the mock-up mounted on the AZALÉE shaking table is shown in Figure 8.2a. A FE model, based on beam elements, is implemented with the homemade FE code CAST3M (Charras and Kichenin, 2011). The FE model is depicted in Figure 8.2b. This computer model is calibrated on the dynamical behavior

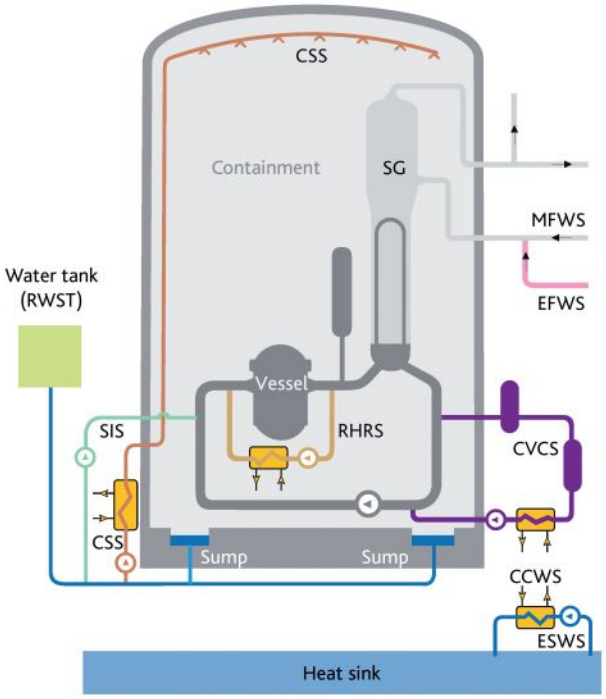
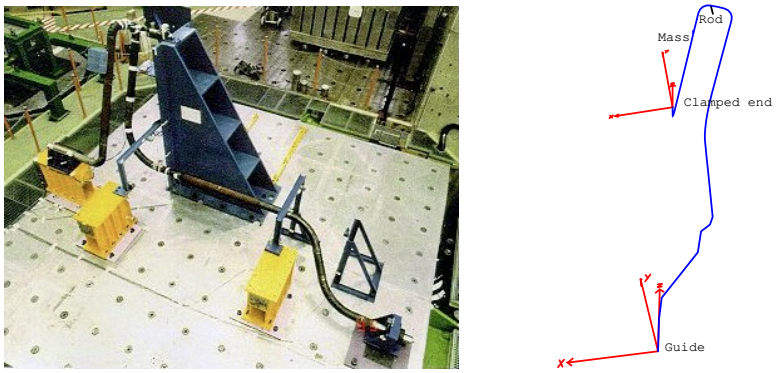


FIGURE 8.1: Schematic diagram of the main auxiliary systems in the reactor building of a French PWR (see Jacquemain, 2015, Chapter 2), the piping system studied in this chapter is the emergency feedwater system (EFWS) shown in pink color. This system aims to maintain the level of water in the main feedwater system (MFWS) when it is not available in order to cool the reactor core through the steam generator (SG). Thus, it is one of the engineered safety systems of the PWR.



(A) (B)

FIGURE 8.2: (a) Overview of the ASG mock-up on AZALÉE shaking table and (b) ASG FE model.

of the mock-up during seismic loading of various intensities.

Numerical computation shows that the maximum stress of the mock-up was found to be in the elbow closest to the clamped end of the mock-up. This stress is associated to an out-of-plane moment. Consequently, the output variable of the computer model studied in this chapter will be

the out-of-plane rotation R (in degree) of this specific elbow, this will be the engineering demand parameter (EDP) used for seismic fragility curve estimation. Furthermore, physical variables describing the material properties of the mock-up and mechanical parameters at the boundary conditions are considered as input variables of the mechanical computer model, they are listed in Table 8.1.

TABLE 8.1: Input variables of the numerical model of the ASG mock-up

Variable number	*Variable*
1	E: Young modulus
2	Sy: Elasticity limit
3	H: Hardening module
4	b: Modal damping ratio
5	RPY151: Rotation stiffness for the P151 guide in Y direction
6	RPX29: Rotation stiffness for the P29 clamped end in X direction
7	RPY29: Rotation stiffness for the P29 clamped end in Y direction
8	TPX29: Translation stiffness for the P29 clamped end in X direction
9	TPY29: Translation stiffness for the P29 clamped end in Y direction
10	TPZ29: Translation stiffness for the P29 clamped end in Z direction

The variables E, Sy, H, b are physical parameters proper to the carbon steel TU42C, influencing the dynamical behavior of the mock-up. The other variables concern the boundary conditions: They describe the translation and rotation stiffness on both the clamped end and the guide of the mock-up. Indeed, they can be considered uncertain and their impacts on the dynamical behavior of the mock-up are generally not assessed. These variables are considered tainted by epistemic uncertainties. Indeed, the uncertainties on these variables may be reducible to ensure the best possible knowledge of the structure, whereas the uncertainty on the seismic loading is aleatory.

The dynamical behavior of the piping mock-up is simulated using the synthetic seismic ground motion generator described in Chapter 2, Section 2.3.1. The ground motions are previously filtered by a linear single d.o.f. oscillator with a natural frequency of 5 Hz and a damping ratio of 2%. This linear oscillator represents roughly the reactor building that support the ASG piping system. This fictitious building is considered deterministic here since we only want to assess the reliability of the piping system. In practice, uncertainties should also be taken into account on the mechanical characteristics of the building.

The computer model of the ASG mock-up is composed of a linear FE model when the maximal stress in the mock-up pipe elbow is less than the elasticity limit Sy and a nonlinear FE model when

the maximal stress is greater. A run of the linear FE model has a computation of a dozen of seconds whereas a run of the nonlinear FE model has a computation time of approximately 10 minutes.

8.3 Probabilistic model for the epistemic uncertainties of the ASG piping mock-up

The input parameters describing the material and the boundary conditions of the mock-up are considered uncertain. The vector of the 10 parameters are denoted by $\mathbf{X}$, and we assume that $\mathbf{X}$ follows a probability distribution described in Table 8.2, the variables being independent, it is sufficient to describe the probability distribution of $\mathbf{X}$ by the marginal probability distribution of each of the input variables.

The mean value of each parameter is calibrated in the following manner: The mock-up is part of a bigger piping system with a known first eigenmode obtained through numerical simulations, so we choose the mean value for the boundary condition's parameters so that the first eigenmode of the mock-up matches the first eigenmode of the mock-up when coupled to the entire piping system. Therefore, computational experiments based on simulations with calibrated mock-up boundary conditions are more representative of the mock-up in its real environment. Of course, these parameters are uncertain, the probability distribution chosen for $\mathbf{X}$ is the uniform distribution, since it follows the principle of maximum entropy (see E. T. Jaynes, 1957). The uniform distribution is parametrized by its mean and the coefficient of variation (c.o.v.). A c.o.v. of 15 % is considered as a good compromise between uncertainty assessment of the input parameters and the physical properties of the system studied. Note that the numerical values used in this chapter are still purely hypothetical and does not intend to represent the real system.

8.4 Step #1: Surrogate model building and uncertainty propagation

In this section, we estimate the fragility curve of the piping system mock-up using a Gaussian process surrogate. The seismic intensity measure chosen in this study is the pseudo-spectral acceleration at 5 Hz and 1 % damping ratio. The spectral acceleration (PSA) at pulsation ω and damping ratio ξ for a seismic acceleration signal $t \to s(t)$ is defined by:

$$a = \max_{t \in [0,T]} \omega^2 |\ell(t)| \,, \tag{8.1}$$

where ℓ is the displacement of a single d.o.f. linear oscillator with natural pulsation ω and damping ratio ξ. This intensity measure is usually chosen in industrial applications due to the high correlation between the engineering demand parameter and the pseudo-spectral acceleration for various types of structure. The statistical regression model considered is thus the following

$$y(a, \mathbf{x}) = g(a, \mathbf{x}) + \varepsilon(a, \mathbf{x}) \,, \tag{8.2}$$

where $y(a, \mathbf{x}) = \log(R(a, \mathbf{x}))$ is the log out-of-plane rotation angle at the mock-up pipe elbow.

8.4.1 Validation of the surrogate models

Due to the high dimension of the input space (10 mechanical input variables and 1 seismic intensity measure), we reduce the dimension of the input space with an HSIC based statistical hypothesis test using the ICSCREAM methodology developed in Marrel, Iooss, and Chabridon, 2021: a Gaussian kernel is used for each input variable and for the output variable (i.e. the log

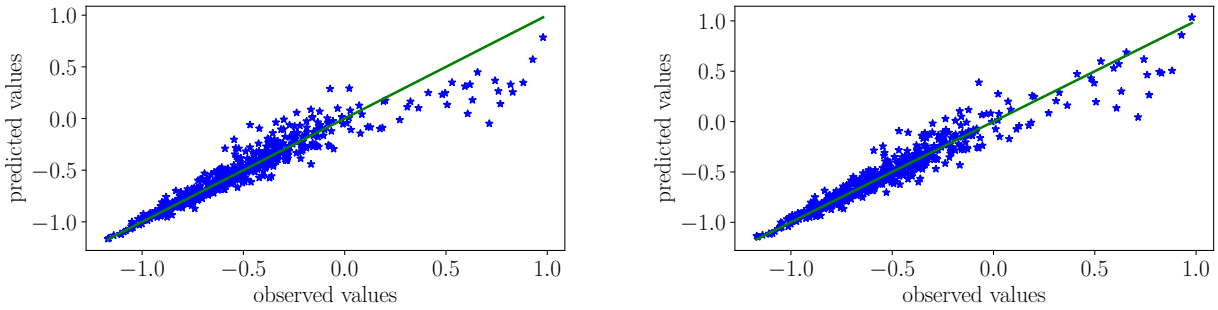
TABLE 8.2: Input variables probabilistic model

Variable	*Distribution*	*Mean*	*c.o.v.*
E (Pa)	Uniform	$1.9236\ 10^{11}$	15%
Sy (MPa)	Uniform	300	15%
H (1)	Uniform	$4.27\ 10^{8}$	15%
b (1)	Uniform	0.01	15%
RPY151 (Nm/rad)	Uniform	$1.1\ 10^{5}$	15%
RPX29 (Nm/rad)	Uniform	$1.1\ 10^{5}$	15%
RPY29 (Nm/rad)	Uniform	$3.3\ 10^{5}$	15%
TPX29 (N/m)	Uniform	$1.0\ 10^{6}$	15%
TYP29 (N/m)	Uniform	$2.0\ 10^{5}$	15%
TPZ29 (N/m)	Uniform	$1.0\ 10^{6}$	15%

rotation of the pipe elbow). 2000 mechanical simulations using the cheaper linear finite element model for the piping system are carried out to perform the screening of the mechanical input variables. 6 input variables are selected (the variables number 1,2,3,8,9,10 in Table 8.1). This screening procedure has a physical interpretation: the variables not detected by the HSIC based statistical hypothesis test are all rotation stiffness on the clamped end for the three axis. The GP surrogates proposed in this chapter are the same as in Chapter 5, Section 5.5. They are zero-mean Gaussian processes with anisotropic tensorized Matérn 5/2 covariance function with homoskedastic Gaussian noise on one hand and a parametrized heteroskedastic Gaussian noise on the other hand.

First, the predictivity of our surrogates is measured qualitatively. Figure 8.3 shows the predicted versus observed values of the log-EDP $y(a, \mathbf{x})$ using a learning dataset of $n = 500$ observations, the green solid line corresponds to the identity, the closer the data are from this line the better is the prediction quality of the surrogate. We can notice that the heteroskedastic Gaussian process underestimates the high values of the log-EDP, the homoskedastic surrogate has also this behavior, but the data are closer to the identity line for high values of the log-EDP.

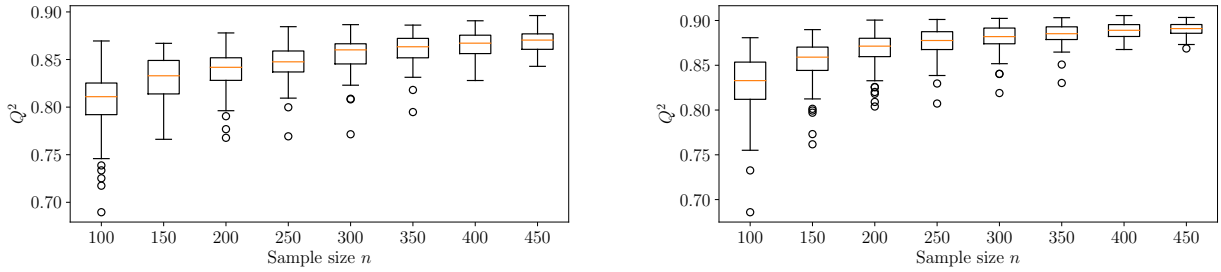
Second, we measure quantitatively the predictivity quality of our surrogates using the predictivity coefficient Q^2 in the same fashion as in Chapter 5, Section 5.5. The predictivity coefficient Q^2 is computed on a test dataset of 1000 nonlinear mechanical simulations with a learning sample size n varying between 100 and 450. The learning datasets are resampled using bootstrap in a total dataset of also 1000 nonlinear mechanical simulations. 100 replications are made for each n. Figure 8.4 shows boxplots of the predictivity coefficient Q^2 for the heteroskedastic and homoskedastic Gaussian processes. We can notice that the homoskedastic Gaussian process surrogate performs slightly better than the heteroskedastic Gaussian process surrogate for both the



(A) Heteroskedastic Gaussian process (B) Homoskedastic Gaussian process

FIGURE 8.3: Predicted values versus observed values for the heteroskedastic and homoskedastic Gaussian process surrogates with a learning dataset size $n = 500$.

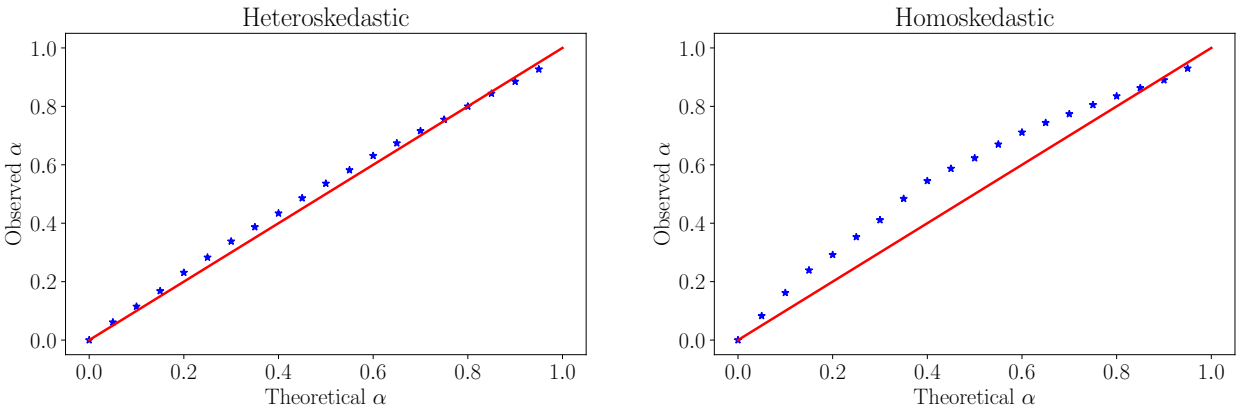
mean and dispersion of the Q^2 predictivity coefficient between replications for each n. This might be the result of the increase of the number of hyperparameters to calibrate for the heteroskedastic Gaussian process.



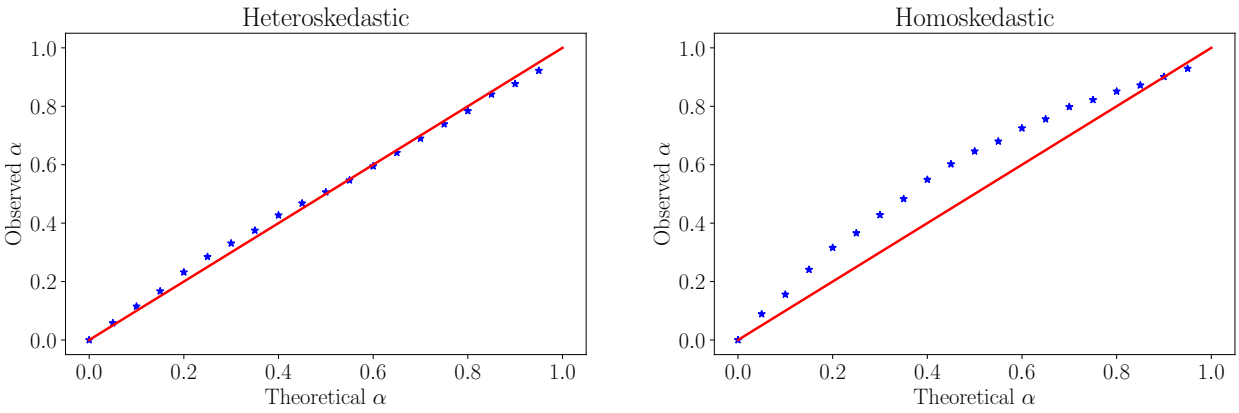
(A) Heteroskedastic Gaussian process (B) Homoskedastic Gaussian process

FIGURE 8.4: Boxplots of the predictivity coefficients Q^2 estimated with 100 replications for learning sample sizes varying from $n = 100$ to $n = 450$.

Moreover, in the same fashion as in Chapter 5, Section 5.5, we study qualitatively the coverage probabilities of our two GP surrogates using $\alpha - \alpha$ plots. The empirical coverage probabilities are estimated on a test sample of size 1000. Figure 5.10 gives the results for heteroskedastic and homoskedastic Gaussian process surrogates for learning sample of size 200 and 500. Note that for both dataset sizes the empirical coverage probabilities with the heteroskedastic surrogate are closer to the identity line than for the homoskedastic surrogate. This can be explained by the flexibility of the variance provided by the heteroskedasticity which allows better adaptation to the distribution of the data than with a fixed value for the variance. Finally, the observations made in Figures 8.4 and 8.5 indicate that the homoskedastic surrogate performs better in terms of predictivity than the heteroskedastic one, but it is less efficient in terms of uncertainty quantification and coverage. According to Ockham's razor principle, more complex models are not always the best choice for the practitioner and are highly dependent on the intended application. The homoskedastic model is the best model in order to approximate the regression function g, but the heteroskedastic one is better than the homoskedastic one to approximate the overall distribution of the data which consists in, due to the Gaussian assumption of the statistical model defined in Equation 5.80, approximating both the variance and the regression function of the data.



(A) $n = 200$ (B) $n = 200$



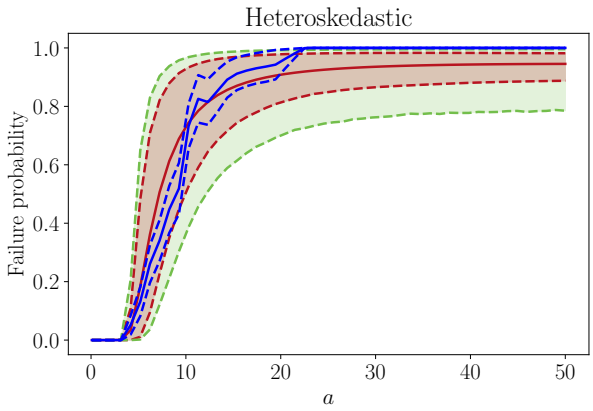
(C) $n = 500$ (D) $n = 500$

FIGURE 8.5: Observed proportion of the data that lies in the α-theoretical confidence intervals with respect to their theoretical proportion for both heteroskedastic and homoskedastic Gaussian processes with sample size of the learning dataset $n = 200$ and $n = 500$.

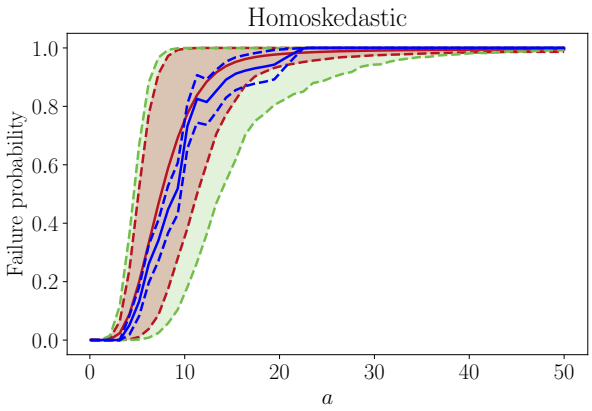
8.4.2 Results and discussion

The GP surrogates allow us to propagate the uncertainty on $\mathbf{X}$ into the seismic fragility curves by considering the random functions $a \to \Psi(a, \mathbf{X})$ as described in Chapter 4, Section 4.3.2. The mean seismic fragility curve and the seismic fragility quantile curves are estimated empirically using a Monte-Carlo sampling $(\mathbf{X}_j)_{1 \leq j \leq m}$ of size $m = 1000$. Numerical results for several training sizes n and failure elbow out-of-plane rotation angles C are shown in Figures 8.6 and 8.7. The failure angles $C = 0.5°$ and $C = 1°$ correspond respectively to a 75 % and 90 % level empirical quantile of the 2000 nonlinear CAST3M simulations. The red area corresponds to the area of the 10% and 90% level fragility quantiles estimated using $(\Psi^{(1)}(., \mathbf{X}_j))_{1 \leq j \leq m}$ and $(\check{\Psi}^{(1)}(., \mathbf{X}_j))_{1 \leq j \leq m}$ for respectively the homoskedastic and heteroskedastic Gaussian process. The Gaussian process surrogate uncertainty is assessed by sampling $Q = 1000$ realizations of the GP posterior distribution for each value $\mathbf{X}_j$, $1 \leq j \leq m$, the bi-level fragility quantile curves with $\gamma_{\mathbf{X}} = \gamma_G = 0.1$ and $\gamma_{\mathbf{X}} = \gamma_G = 0.9$ are shown in dashed green and are estimated empirically from the dataset $(\Psi_q^{(2)}(., \mathbf{X}_j))_{\substack{1 \leq q \leq Q, \\ 1 \leq j \leq m}}$ and $(\check{\Psi}_q^{(2)}(., \mathbf{X}_j))_{\substack{1 \leq q \leq Q, \\ 1 \leq j \leq m}}$ respectively for the homoskedastic and heteroskedastic GP.

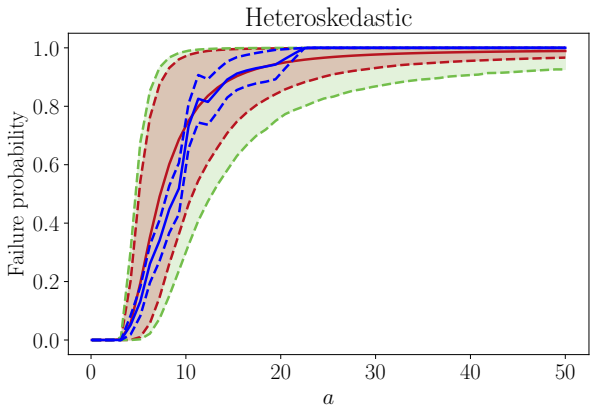
Note that the interquantile range is larger for the homoskedastic Gaussian process than the



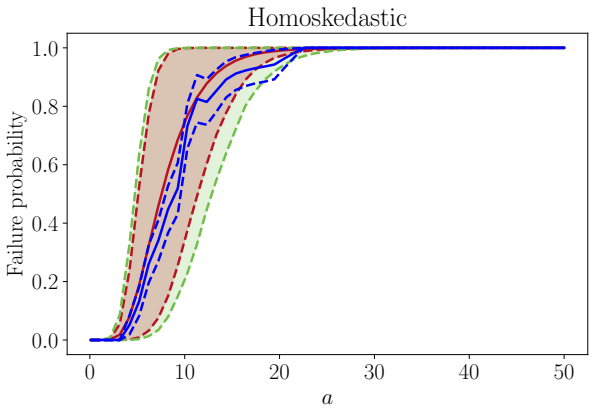
(A) $C = 0.5°,\ n = 200$



(B) $C = 0.5°,\ n = 200$



(C) $C = 0.5°,\ n = 500$



(D) $C = 0.5°,\ n = 500$

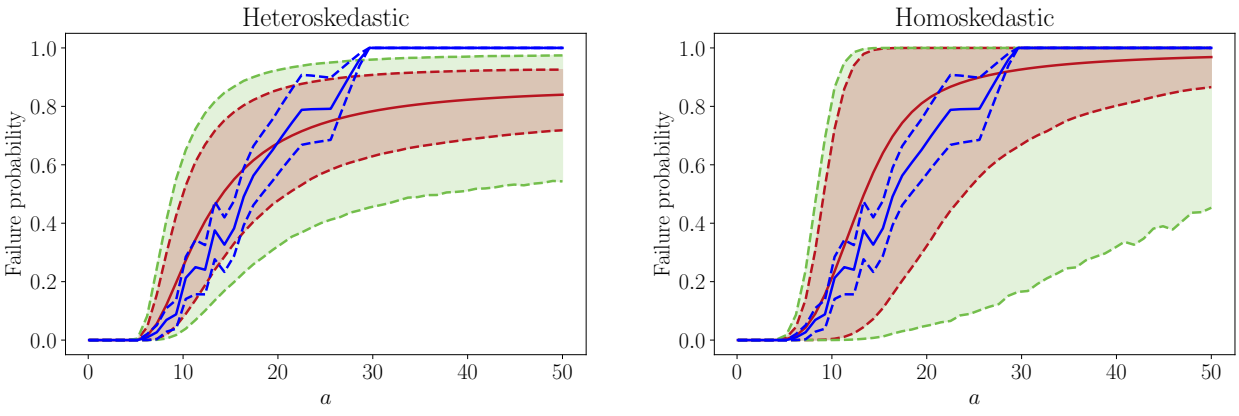
FIGURE 8.6: Uncertainty propagation on the seismic fragility curves with a failure elbow out-of-plane rotation angle $C = 0.5°$. The blue solid curve is the nonparametric Monte-Carlo estimation of the mean fragility curve and the dashed blue curves are the 90 % and 10 % asymptotic quantiles of the nonparametric fragility curve estimator (see Chapter 2, Section 2.3.2)

heteroskedastic Gaussian process for small training datasets ($n = 200$) and especially for a high failure out-of-plane rotation angle of the pipe mock-up elbow ($C = 1°$). This tends to demonstrate that the heteroskedastic surrogate fits better the conditional distribution of the log elbow out-of-plane rotation angle.

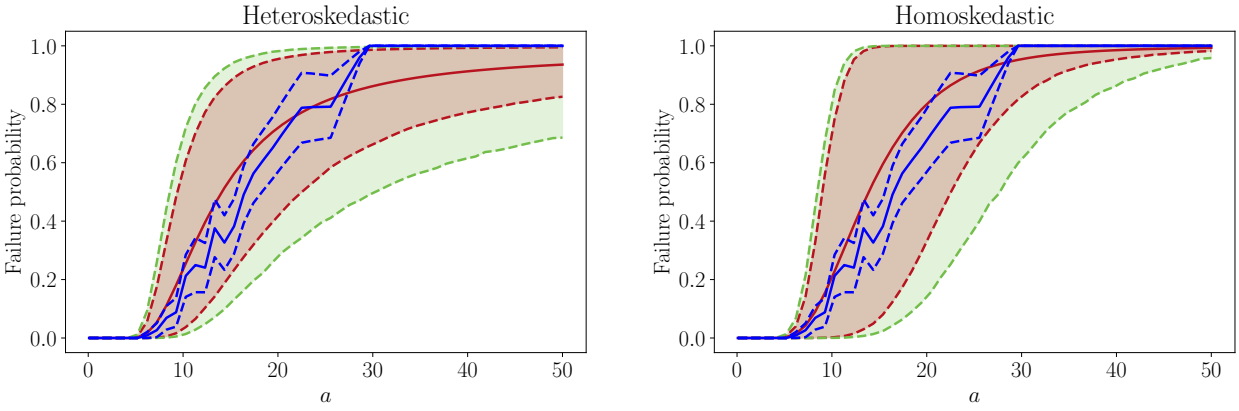
8.5 Step #2: Global sensitivity analysis on the seismic fragility curve

8.5.1 Global sensitivity indices estimation settings

The estimation of the aggregated Sobol indices and the MMD-based sensitivity indices of the seismic fragility curves of the pipe mock-up is performed using the methodology described in Chapter 6, Section 6.4. A training dataset of $n = 500$ simulations and a Monte-Carlo design of size $m = 20000$ have been sampled in order to perform the pick-freeze estimation of the aggregated Sobol indices. $Q = 200$ realizations of the GP surrogate and $B = 150$ bootstrap samples have been carried out to assess the uncertainty of the aggregated Sobol indices both in terms of metamodeling and Monte-Carlo uncertainty. For the failure elbow out-of-plane rotation angle $C = 1°$ we



(A) $C = 1°,\ n = 200$ (B) $C = 1°,\ n = 200$



(C) $C = 1°,\ n = 500$ (D) $C = 1°,\ n = 500$

FIGURE 8.7: Uncertainty propagation on the seismic fragility curves with a failure elbow out-of-plane rotation angle $C = 1°$. The blue solid curve is the nonparametric Monte-Carlo estimation of the mean fragility curve and the dashed blue curves are the 90 % and 10 % asymptotic quantiles of the nonparametric fragility curve estimator (see Chapter 2, Section 2.3.2)

compute the L^2 distance between fragility curves on the interval $a \in [5, 25]$ in order to focus on the transition area between small and high probabilities of failure.

For the MMD-based sensitivity indices, we use the same parameters $n,\ Q,\ B$ as for the estimation of the aggregated Sobol indices. However, we choose $m = 15000$ for the Monte-Carlo design used for the pick-freeze estimation. The kernel used is the Gaussian kernel defined in Equation 6.31.

8.5.2 Results and discussion

The figures 8.8 and 8.9 provide the results for the estimation of both first-order and total-order aggregated Sobol indices for $C = 1°$ using the homoskedastic and heteroskedastic Gaussian process surrogates. Remark that the parameters E, TPX29 and TPY29 are mostly influential on the seismic fragility curve. Indeed, the modal properties of the piping system essentially drive its dynamic behavior and hence its robustness under seismic loading. One of the way to explain why TPY29 is the most influential mechanical parameter of the piping system is the coupling of the eigenmodes between the X direction (i.e. the direction of the seismic load) and the Y direction (i.e.

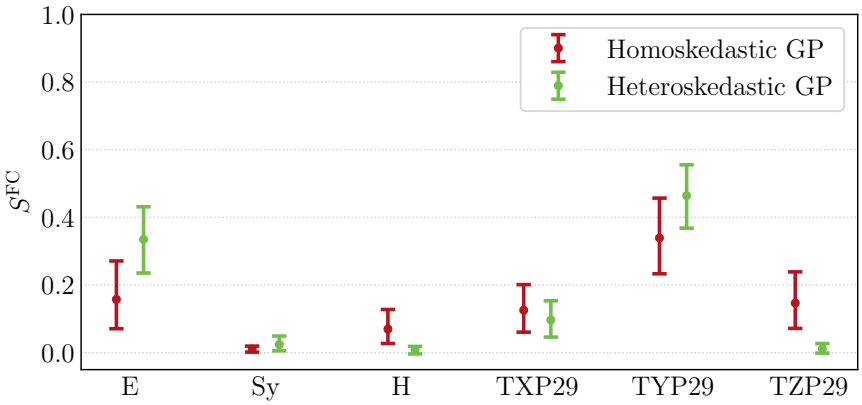


FIGURE 8.8: First-order aggregated Sobol indices for a failure rotation angle $C = 1°$ estimated with the heteroskedastic and homoskedastic GP surrogates

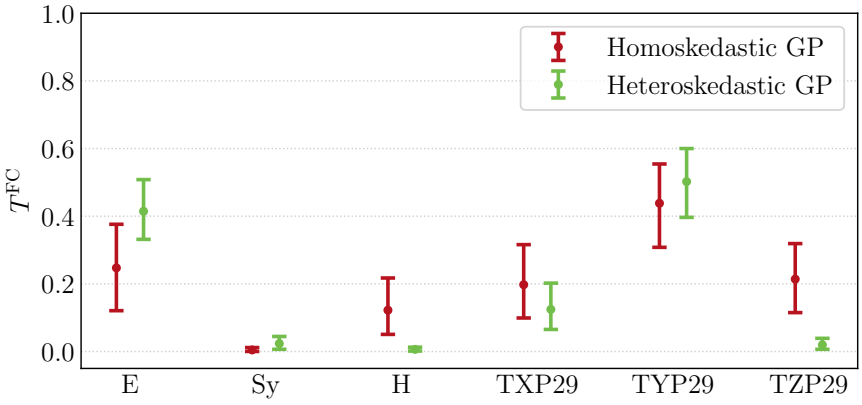


FIGURE 8.9: Total-order aggregated Sobol indices for a failure rotation angle $C = 1°$ estimated with the heteroskedastic and homoskedastic GP surrogates

the direction of the permanent loading due to the piping system's weight). Note that the variable TYP29 corresponds to the stiffness of the clamped end in the Y direction. The influence of variable TYP29 is more clearly detected by the heteroskedastic Gaussian process surrogate, however the two metamodels raise the same ranking of mechanical parameters in terms of aggregated Sobol indices.

The results of the estimation of the MMD-based sensitivity indices are shown in Figure 8.10 and 8.11. Note that the ranking of inputs is the same as for the one obtained with aggregated Sobol indices. However we can remark that the total-order MMD-based indices have higher values than the first-order MMD-based indices whereas the aggregated Sobol indices of first and total order have very close values. This means that the aggregated Sobol indices fail to detect interactions between input parameters. On the contrary, because the MMD-based indices take into account the overall probability distribution of the fragility curves conditional to the input parameters, it is not surprising to detect more clearly interactions between inputs. Note also that the difference between the heteroskedastic and homoskedastic GP surrogates is also more striking than for the aggregated Sobol indices. Similarly to the aggregated Sobol indices, the influence of TYP29 seems more clearly detected by the heteroskedastic GP surrogate than the homoskedastic one, while keeping the same ranking of influence for each mechanical parameter.

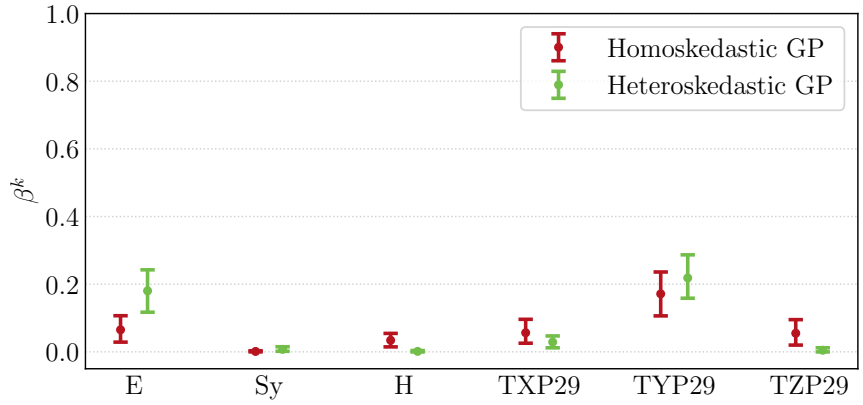


FIGURE 8.10: First-order MMD-based indices for a failure rotation angle $C = 1°$ estimated with the heteroskedastic and homoskedastic GP surrogates

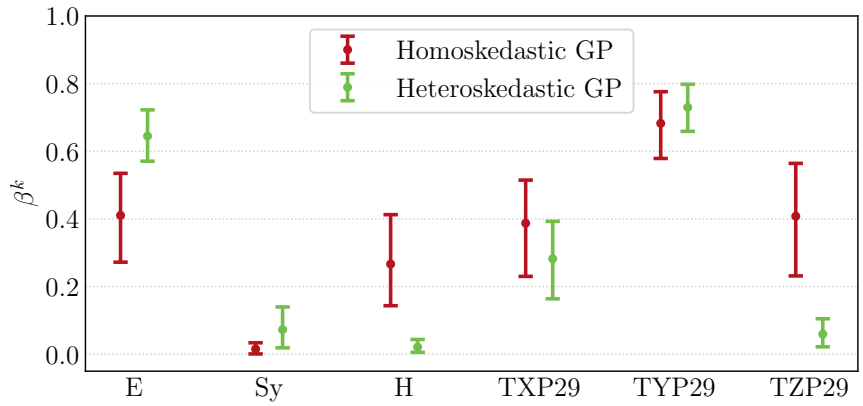


FIGURE 8.11: Total-order MMD-based indices for a failure rotation angle $C = 1°$ estimated with the heteroskedastic and homoskedastic GP surrogates

8.6 Step #3: Bayesian sequential design of experiments for seismic fragility curve estimation

In this third step, the sequential Bayesian design of experiments procedure developed in Chapter 7 is applied to seismic fragility curve estimation of the piping system mock-up.

8.6.1 SUR procedure settings

The SUR procedure in this section is based on a zero-mean Gaussian process with an anisotropic tensorized Matérn 5/2 covariance function and a homoskedastic Gaussian noise assumption. For the same reasons explained in Chapter 7, Section 7.4, the input variables for surrogate fitting are reduced to the most influential variables (PSA, E, TYP29). The failure rotation angle considered is $C = 1°$. This corresponds roughly to the 90% quantile of the output variable for the whole dataset of 2000 CAST3M mechanical simulations. Similarly to Chapter 7, Section 7.4, the performance of the proposed SUR strategy is evaluated by comparing the bias b_n of the estimator $\widehat{\Psi}_n$ to a reference fragility curve Ψ_{ref} estimated by kernel ridge regression with Gaussian kernel on the whole dataset of 2000 CAST3M mechanical simulations. The performance is also assessed with the loss ℓ defined in Equation 7.18. For this application, we choose for the measure h the uniform distribution on the segment $[a_0, a_1]$ where $a_0 = 1\ m/s^2$ and $a_1 = 25\ m/s^2$. This choice allows to target evenly the transition between low and high probabilities of failures. Due to the computational cost of the CAST3M mechanical simulations, we propose a meta-estimation framework as in Bect, Ginsbourger, et al., 2012. The SUR procedure starts by choosing 10 CAST3M simulations in a stratified manner on the pseudo-spectral acceleration, by dividing the full dataset into 10 sets of equal size using the $(j/10)_{1\leq j\leq 9}$-level quantiles of the PSA. They are the first points of the design set and are used to optimize the Gaussian process hyperparameters (including σ_ε). After that, a set of candidate points $(a'_j, \mathbf{X}'_j)_{1\leq j\leq M}$ ($M = 1000$) are chosen at random in the dataset of 2000 CAST3M mechanical simulations. The candidate points set is built by using the same partition into 10 subdomains as for initialization. We then compute the SUR sampling criterion as described in Algorithm 5. The SUR sampling criterion is estimated with a Monte-Carlo sample of size $N = 5000$ and $Q = 12$ for the Gauss-Hermite quadrature. Due to a difficulty to estimate σ_ε for small training sizes, a non-informative prior $\pi(\sigma_\varepsilon) \propto \sigma_\varepsilon^{-2}$ is assumed on the parameter σ_ε. The GP hyperparameters are then obtained with a MAP estimator. The GP hyperparameters are optimized every 10 iterations.

8.6.2 Results and discussion

100 replications of the SUR strategy is compared with 100 replications of Monte-Carlo sampling design (sampled in the whole dataset of 2000 CAST3M simulations). Figure 8.12 compares the log bias b_n and the log loss ℓ as a function of the training size. The bias b_n is estimated with a Monte-Carlo sample of size 5000, the loss $\ell(\Psi, \widehat{\Psi}_n)$ is estimated with the same Monte-Carlo sample and with 4000 realizations of the GP surrogate. Still the SUR outperforms Monte-Carlo design in this case, however one can see that the performance metrics have more variability in this case. In terms of bias, the SUR strategy beats Monte-Carlo design for a training size greater than 100 CAST3M simulations, the posterior variance of the SUR strategy is quite similar to Monte-Carlo design until a training size of 90 CAST3M simulations. The huge variability in the loss and the bias for the SUR strategy compared to the single d.o.f. oscillator test-case developed in Chapter 7, Section 7.4 may come from the smaller size of the datasets (10^5 simulations for the oscillator compared to 2000 CAST3M simulations).

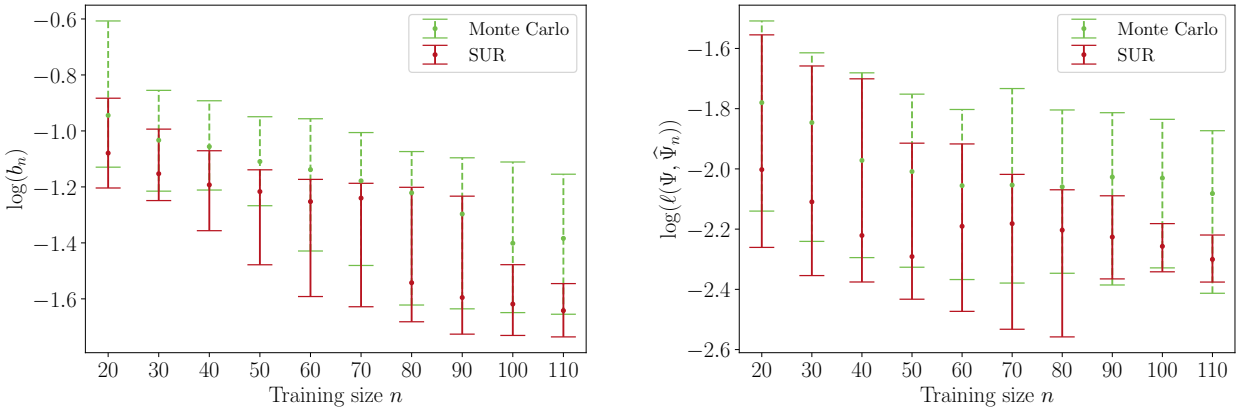


FIGURE 8.12: Comparison of the bias b_n and the loss $\ell(\Psi, \widehat{\Psi}_n)$ between 100 Monte-Carlo design and 100 runs of the proposed SUR strategy for a failure threshold $C = 1°$. The green error bars correspond to the interquantile range between the 10% and 90% quantiles of the observed metric on the 100 replications of the Monte-Carlo design. The red error bars are the same quantities for the replications of the SUR strategy.

Figure 8.13 shows the results of the propagation of uncertainties on the mechanical parameters of the piping structure to the seismic fragility curves with a Monte-Carlo design and a SUR design of size 100. Similarly to the results of Chapter 7, Section 7.4. The GP surrogate uncertainty is significantly smaller (in terms of interquantile range on the seismic fragility curves) with a SUR-based design than a Monte-Carlo design. This further validates the use of such an experimental design strategy on industrial applications.

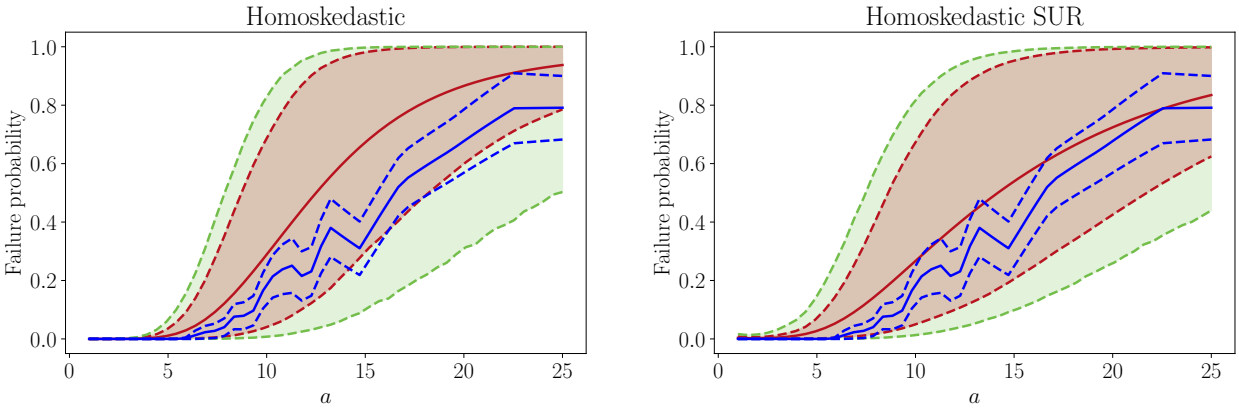


FIGURE 8.13: Comparison of a Monte-Carlo design and a SUR design for a training size $n = 100$ in terms on uncertainty propagation on the seismic fragility curves for the ASG piping system.

8.7 Conclusion

In this chapter, the methodologies developed in the three previous chapters have been illustrated on a mechanical FE computer model of a piping mock-up issued from structural seismic safety for French PWR research experimental program.

The first and second steps present an UQ study of the piping mock-up using a CAST3M finite element computer model: the first step consists in GP surrogates building and is used for uncertainty propagation on the seismic fragility curves of the mock-up. The second step is a sensitivity analysis of the mechanical parameters of the structure tainted by epistemic uncertainties. The third step is an auxiliary step for UQ studies and consists in providing a sequential design of experiments algorithm based on the SUR principles. The design of experiments procedure is based on the Bayesian nature of the GP surrogate and is targeted for seismic fragility curve estimation.

Consequently, the UQ methodology developed in this manuscript has been applied to a finite element computer model of piping mock-up structure, validating the use of these three steps for computer models of real structures. Nonetheless, a nonparametric estimation of the Gaussian noise for the surrogate model building step is possible and motivates further improvements of the methodology.

CHAPTER 9

Conclusion and perspectives

Main contributions of the thesis

The objective of this thesis was to propose efficient methodologies to estimate the seismic fragility curves of mechanical structures while taking into account, if possible, the epistemic uncertainties inherent in their mechanical parameters.

First, the main methods for estimating seismic fragility curves have been reviewed. An estimation procedure based on adaptive importance sampling is then proposed to improve the precision of the estimation of the seismic fragility curve while reducing the number of computer model calls. In a second part, an UQ methodology has been developed to take into account the uncertainties coming from a lack of knowledge on the mechanical structure. Gaussian process surrogates have been proposed to replace the costly mechanical computer model and to perform uncertainty propagation and global sensitivity analysis. Finally, a sequential planning of experiments algorithm using the Gaussian process surrogate have been proposed to improve the estimation of seismic fragility curves.

Parametric seismic fragility curve estimation using importance sampling

The first contribution of this thesis has consisted in the development of an algorithm based on adaptive importance sampling to improve the estimation of parametric lognormal seismic fragility curves. It relies on a statistical learning viewpoint by considering the empirical loss as a Monte-Carlo approximation of the true loss function. Importance sampling is then used as a classical variance reduction technique of the training loss. Since the optimal sampling density depends on the parameters to estimate, the algorithm proposed is of adaptive nature. Theoretical results guaranteeing the consistency and the asymptotic normality of the proposed fragility curve parameters estimator have been developed. The performance of this strategy in terms of loss and coverage probability has been illustrated on several test-cases of increasing complexity. Finally, this work has been subject to the following publication and communications:

- **Gauchy, C.**, C. Feau, and J. Garnier (2020a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *Mascot-num PhD days (MASCOT-NUM 2020)*. Grenoble, France.
- **Gauchy, C.**, C. Feau, and J. Garnier (2020b). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *5ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2020)*. Saint Pierre d'Oléron, France.

- **Gauchy, C.**, C. Feau, and J. Garnier (2021a). *Importance sampling based active learning for parametric seismic fragility curve estimation*. DOI: 10.48550/ARXIV.2109.04323

Gaussian process surrogates and uncertainty propagation

The second contribution of this manuscript has focused on building a surrogate model of the mechanical computer model. It consists in replacing a costly computer model by a predictor based on statistical methods and a learning sample of observed computer model output and input parameters. Gaussian process regression (or kriging) supposes that the computer model is a realization of a Gaussian stochastic process, the main advantage is its ability to propose both a predictor and a quantification of the uncertainty of its predictions that are given analytically. The main results about kriging have been reviewed, as well as the crucial choice of the prior covariance function. Then, the seismic fragility curve estimator has been obtained using the posterior predictive distribution of the Gaussian process surrogate. Moreover, the Bayesian viewpoint on the Gaussian process regression methodology helps us to propagate the uncertainty of the Gaussian process surrogate into the seismic fragility curve estimator. Finally, an uncertainty propagation methodology has been defined to study how the epistemic uncertainties on the mechanical structure affect the seismic fragility curves. Numerical experiments on a single d.o.f. oscillator with nonlinear restoring force have been proposed to illustrate the developed methodology.

Global sensitivity analysis on seismic fragility curves using kriging

The third technical contribution of this thesis deals with the adaptation of the classical global sensitivity analysis techniques on scalar model output to functional output such as seismic fragility curves. The idea was to study how the uncertainty of the seismic fragility curve is apportioned to the uncertainties tainting the mechanical parameters of the structure. To do so, the classical Sobol indices have been extended to the case of functional output thanks to the aggregated Sobol indices, their main advantage is the conservation of the theoretical guarantees of the Sobol indices such as the Sobol-Hoeffding decomposition and the pick - freeze estimation methodology. Then, new kind of global sensitivity indices based on kernel methods have been proposed, which are based on the maximum mean discrepancy (MMD) distance. A kernel function between seismic fragility curves is then defined to compute the MMD-based sensitivity indices. The two global sensitivity indices proposed can be estimated using the pick-freeze formulation. Furthermore, the Gaussian process posterior distribution is propagated into the global sensitivity indices estimates using the method of kriging by conditioning. Therefore, the uncertainty of the metamodel is taken into account in the sensitivity analysis step. The methodology proposed has been illustrated on a single d.o.f. nonlinear oscillator.

The second and the third contributions have been subject to the following communications:

- **Gauchy, C.**, C. Feau, and J. Garnier (2021d). "Propagation of epistemic uncertainties in seismic risk assessment". In: *15th Mascot-Num annual conference (MASCOT-NUM 2021)*. Online
- **Gauchy, C.**, C. Feau, and J. Garnier (2021b). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *51ème journées de Statistique de la Société Française de Statistique (JdS 2021)*. Online
- **Gauchy, C.**, C. Feau, and J. Garnier (2021c). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *6ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2021)*. Erdeven, France

- **Gauchy, C.**, C. Feau, and J. Garnier (2022d). "Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging". In: *16th Mascot-Num annual conference (MASCOT-NUM 2022)*. Clermont-Ferrand, France

A scientific article on this subject is currently under writing process at the time these lines are written and will be soon submitted to a peer-reviewed journal.

Bayesian sequential planning of experiments using Gaussian process

The fourth contribution of this thesis is the elaboration of a sequential planning of experiments procedure based on *stepwise uncertainty reduction* technique. This type of algorithm takes into account the Bayesian nature of Gaussian process to define sampling criterion that aims at maximizing the estimation accuracy of a given quantity of interest. After a brief review on the main principles of SUR strategy, a sequential planning of experiments procedure have been proposed to improve the estimation of seismic fragility curves using Gaussian processes. Numerical experiments have been carried out to study the performance of the proposed SUR algorithm on a test-case composed of a single d.o.f. oscillator with nonlinear restoring force. The contribution is also linked to the following communication.

- **Gauchy, C.**, C. Feau, and J. Garnier (2022b). "Estimation of seismic fragility curves by sequential design of experiments". In: *52ème journées de Statistique de la Société Française de Statistique (JdS 2022)*. Lyon, France

Application to a realistic piping system of a nuclear power plant

The last contribution of this thesis is devoted to the application of the developed methodologies to a realistic mechanical computer model issued from the nuclear industry. This case is based on the dynamical simulation of a piping system mock-up under seismic loading. After a modeling of the possible sources of epistemic uncertainties tainting the piping mock-up, Gaussian process surrogate models with a homoskedastic noise on the one hand and heteroskedastic noise on the other hand are fitted using a learning sample of several computer model simulations. Then, the uncertainties are propagated on the seismic fragility curve. Several statistical quantities have been studied to account for the surrogate models performances. In addition, a sensitivity analysis step has been carried out using the surrogates to apportion the uncertainty on the seismic fragility curve to the uncertainty of the different mechanical parameters. Finally, the Bayesian planning of experiments algorithm has been applied to this test-case to study its performances on a complex computer model used in mechanical simulations for seismic safety studies.

Perspectives

The work presented in this thesis could be extended in several directions set out below.

About surrogate modeling. A first research track could consist in elaborating a Gaussian process surrogate model fitting both the conditional mean and variance of the mechanical computer model output. Indeed, the results presented in this manuscript demonstrate the difficulty of choosing a parametric model for the heteroskedastic noise and therefore motivate a nonparametric estimation of the noise. This can be done using joint Gaussian process modeling as proposed in Marrel, Iooss, Da Veiga, et al., 2012: the first step consists in fitting a GP surrogate model on the computer model output with a homoskedastic Gaussian noise, then a second step aims at fitting

a second GP surrogate model on the squared residuals of the first predictor in order to learn the heteroskedastic variance, the third step uses the GP variance predictor to fit the mean component of the computer model with a third GP surrogate, this time with a heteroskedastic noise assumption. Finally, a fourth GP is fitted on the squared residuals. Another heteroskedastic Gaussian process methodology using a latent-variable viewpoint is proposed in Binois, R. Gramacy, and Ludkovski, 2018.

About sensitivity analysis. The topic of sensitivity analysis using kernel methods is promising and motivates further researches. Indeed, the choice of the kernel for the MMD-based sensitivity analysis is crucial and still an open question as pointed out by Barr and Rabitz, 2022. The problem of the choice of the kernel can be relaxed by choosing a parametric family of kernels like the Gaussian family with a lengthscale parameter. However, on the contrary to kernel methods for statistical tests (see Gretton, Sejdinovic, et al., 2012) no optimality criterion for kernel choosing in GSA context is available in the literature. Another way to solve this problem could be to design a specific kernel for seismic fragility curves.

About SUR strategies. A motivating research track for SUR strategies could be the development of new methodologies for tackling the curse of dimensionality. Indeed, the optimization of the SUR sampling criterion using candidate points set suffers from the curse of dimensionality. Metamodeling the SUR sampling criterion could be an idea but one has to be sure to not downgrade the performances of the SUR strategy. Another possibility could be to carefully design the set of candidate points using space-filling design to explore all the possible interactions between the input variables when the input space is high-dimensional.

About Bayesian estimation of seismic fragility curve. The internship of Antoine Van Biesbroeck (see Appendix C) dealing about the Bayesian estimation of seismic fragility curves using objective priors led to promising results. Antoine will start a PhD on this subject in Autumn 2022 to study Bayesian methods for seismic fragility curves estimation more thoroughly.

APPENDIX A

Mathematical background

A.1 Cholesky decomposition

The Cholesky decomposition of a symmetric definite positive matrix A consists in decomposing A into a product of lower triangular matrix and its transpose

$$LL^T = A\,, \tag{A.1}$$

where L is called the Cholesky factor. The Cholesky decomposition is very useful to solve linear systems of the form $A\mathbf{x} = \mathbf{b}$. First solve the triangular system $L\mathbf{y} = \mathbf{b}$ and then the system $L^T\mathbf{x} = \mathbf{y}$. The solution is written $\mathbf{x} = L^T\backslash(L\backslash\mathbf{b})$ where the notation $A\backslash\mathbf{b}$ is the vector $\mathbf{x}$ solving $A\mathbf{x} = \mathbf{b}$.

A.2 Sampling of a Gaussian vector

Consider a Gaussian vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^T$ with mean vector $\mu \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$. A useful algorithm for generating a sample $\mathbf{y} = (y_1, \ldots, y_p)^T$ from the probability distribution of $\mathbf{Y}$ is relying on the Cholesky decomposition of Σ:

- Consider the vector $\mathbf{v} = (v_1, \ldots, v_p)^T$ such that $v_i \sim \mathcal{N}(0, 1)$
- Write $\Sigma = LL^T$ where L the Cholesky factor
- Compute $\mathbf{y} = \mu + L\mathbf{v}$

APPENDIX B

Proofs of Chapter 3

B.1 Proof of Equation (3.23)

Throughout the appendix, the IS-AL estimator $\widehat{\theta}_n^{\mathrm{IA}}$ is denoted by $\tilde{\theta}_n$ and the loss $\widehat{R}^{\mathrm{IA}}$ by $\tilde{R}$. The proof for the consistency is based on Theorem 2 of Delyon and Portier, 2018. We precise the needed assumptions in a very general way, with a parametric family $\mathcal{F} = \{f_\theta,\ \theta \in \Theta\}$, loss function ℓ_θ and instrumental density q_θ. We will then check that the needed assumptions are satisfied by IS-AL. Set $L(x,s) = \sup_{\theta\in\Theta} \ell_\theta(x,s)$. Assume that Θ is a compact set, $\theta_* = \operatorname{argmin}_{\theta\in\Theta} r(\theta)$ exists and is unique and that:

$$\iint L(x,s)P(dx,ds) < +\infty,$$

$$\sup_{\theta\in\Theta} \iint \frac{L(x,s)^2 p(x)}{q_\theta(x)} P(dx,ds) < +\infty,$$

$$\forall \theta \neq \theta_*, \iint \ell_\theta(x,s)P(dx,ds) > \iint \ell_{\theta_*}(x,s)P(dx,ds).$$

and for any $(x,s) \in \mathcal{X} \times \{0,1\}$, $\theta \in \Theta \mapsto \ell_\theta(x,s)$ is continuous. Thus, we can apply Theorem 2 of Delyon and Portier, 2018 in order to prove the consistency of $\tilde{\theta}_n = \operatorname{argmin}_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q_{\tilde{\theta}_{i-1}}(X_i)} \ell_\theta(X_i,S_i)$. More precisely, these assumptions are verified for IS-AL. Indeed, the regularized squared loss is bounded for the variables $\theta,\ x, s$ when $\theta = (\alpha,\beta)^T$ is in a compact set of $(0,+\infty)^2$. Moreover, the likelihood ratio $\frac{p(x)}{q_{\theta,\varepsilon}(x)}$ with the defensive instrumental density is bounded for $x \in \mathcal{X}$. Concerning the regularization, we have:

$$\left|\tilde{R}_n(\theta) - r(\theta)\right| < \left|\frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q_{\widehat{\theta}_{i-1}^{\mathrm{IA}},\varepsilon}(X_i)} \ell_\theta(X_i,S_i) - r(\theta)\right| + \frac{\Omega(\theta;\beta_{\mathrm{reg}})}{n}\,. \tag{B.1}$$

Thus the condition (17) of Theorem 2 in Delyon and Portier, 2018 is still valid.

B.2 Proof of the asymptotic normality of $\widehat{\theta}_n^{\mathrm{IA}}$

In the same way as in the proof of the consistency of $\tilde{\theta}_n$, we provide a general proof of asymptotic normality. Assume that $\theta \mapsto \ell_\theta$ is three times differentiable at θ_* for all x, s and that the matrix $\ddot{r}(\theta_*)$ exists and is nonsingular. Assume that the third-order derivatives of $\theta \mapsto \ell_\theta(x,s)$ are

dominated in a neighborhood of θ_* by a function that is integrable with respect to P. Assume also that the following conditions are satisfied:

1. The hypotheses needed for the consistency of $\tilde{\theta}_n$ are satisfied,
2. $\exists \eta > 0$ such that $\sup_{\theta\in\Theta} \iint ||\frac{p(x)\dot{\ell}_{\theta_*}(x,s)}{q_\theta(x)}||^{2+\eta} P(dx, ds) < +\infty$,
3. $\sup_{\theta\in\Theta} \iint \frac{p(x)||\ddot{\ell}_{\theta_*}(x,s)\ddot{\ell}_{\theta_*}(x,s)^T||}{q_\theta(x)} P(dx, ds) < +\infty$,
4. there exists a neighborhood $\mathcal{B}$ of θ_* such that $\forall (x, s) \in \mathcal{X} \times \{0, 1\}$, $\sup_{\theta\in\mathcal{B}} \frac{p(x)\| \dddot{\ell}_\theta(x,s)\|}{q_\theta(x)} < +\infty$.

The asymptotic normality of an estimator built such as $\tilde{\theta}_n$ is based on the following arguments highlighted in Theorem 5.41 of Vaart, 1998:

- **(P1)** The random function $\sqrt{n}\Psi_n(\theta_*)$, with $\Psi_n(\theta) = \dot{\hat{R}}_n(\theta) - \dot{r}(\theta)$, converges in law to a centered Gaussian distribution with covariance V_{θ_*}.
- **(P2)** The random function $\dot{\Psi}_n(\theta_*)$ converges in probability to $\mathbb{E}[\ddot{\ell}_{\theta_*}(X, S)]$
- **(P3)** The random function $\ddot{\Psi}_n(\theta_n)$ is bounded in probability for θ_n a deterministic sequence in a neighborhood of θ_*.

Of course, we need all the quantities above to be properly defined, hence we have to restrict ourselves to a loss function $\theta \mapsto \ell_\theta$ that is smooth enough, such as the quadratic loss. We use Theorem 1 of Delyon and Portier, 2018 to prove proposition **(P1)**. Theorem 2.18 in Hall et al., 2014 ensures that **(P2)** and **(P3)** are verified by the assumptions 2), 3) and 4) so that $\dot{\Psi}_n(\theta_*)$ converges toward the matrix $\ddot{r}(\theta_*)$. The sequence $\sqrt{n}(\tilde{\theta}_n - \theta_*)$ is asymptotically normal with mean zero and covariance matrix $\ddot{r}(\theta_*)^{-1}V_{\theta_*}(\ddot{r}(\theta_*)^{-1})^T$. For IS-AL, the functions $\dot{\ell}_\theta$, $\ddot{\ell}_\theta$, $\dddot{\ell}_\theta$ are continuous for variables θ, x on a compact set and thus are bounded for variable θ, in the same way as for the consistency, the likelihood ratio for the defensive instrumental density is bounded for $x \in \mathcal{X}$. Concerning the regularization, the third derivative $\dddot{\Omega}(\theta; \beta_{\text{reg}})$ is continuous on Θ which is compact, hence bounded. Naturally, we have **(P3)** verified. Because $\frac{\dot{\Omega}(\theta;\beta_{\text{reg}})}{n}$ converges in probability to 0, **(P2)** is also verified. Using Slutsky's lemma, **(P1)** is verified.

B.3 Proof of Equation (3.30)

First of all, we precise the needed assumptions for a general proof. Set $L_{1,k,l}(x, s) = sup_{\theta\in\Theta}\ddot{\ell}_\theta(x, s)_{k,l}$ and $L_{2,k,l}(x, s) = sup_{\theta\in\Theta}\frac{p(x)}{q_\theta(x)}(\dot{\ell}_\theta(x, s)\dot{\ell}_\theta(x, s)^T)_{k,l}$ $\forall k, l = 1, ..., m$ and assume that:

1. $\inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \frac{p(x)}{q_\theta(x)}\ddot{\ell}_\theta(x, s)_{k,l} > -\infty$ $\forall k, l = 1, \ldots, m$.
2. $\inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \left(\frac{p(x)}{q_\theta(x)}\right)^2 \dot{\ell}_\theta(x, s)\dot{\ell}_\theta(x, s)^T_{k,l} > -\infty$ $\forall k, l = 1, \ldots, m$.
3. $\iint L_{i,k,l}(x, s)P(dx, ds) < +\infty$, $\forall i \in \{1, 2\}$, $\forall k, l = 1, ..., m$.
4. $\sup_{\theta\in\Theta} \iint \frac{L_i(x,s)^2 p(x)}{q_\theta(x)} P(dx, ds) < +\infty$, $i \in \{1, 2\}$.

The result comes from the uniform convergence of $\widehat{G}_n(\theta) = \widehat{\ddot{r}}_n(\theta)^{-1}\widehat{V}_n(\theta)(\widehat{\ddot{r}}_n(\theta)^{-1})^T$ to G_θ for θ in a neighborhood of θ_*. It boils down to prove uniform convergence of $\widehat{\ddot{r}}_n(\theta)$ and $\widehat{V}_n(\theta)$. The proof is in the same spirit as in Section B.1. We proceed coordinate by coordinate defining $H_i(\theta)_{k,l} = \frac{p(X_i)}{q_\theta(X_i)}\ddot{\ell}_\theta(X_i,S_i)_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}}\frac{p(x)}{q_\theta(x)}\ddot{\ell}_\theta(x,s)_{k,l}$ to prove uniform convergence of $\widehat{\ddot{r}}_n(\theta)$ and $H_i(\theta)_{k,l} = \left(\frac{p(X_i)}{q_\theta(X_i)}\right)^2 \dot{\ell}_\theta(X_i,S_i)\dot{\ell}_\theta(X_i,S_i)^T_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}}\left(\frac{p(x)}{q_\theta(x)}\right)^2 \dot{\ell}_\theta(x,s)\dot{\ell}_\theta(x,s)^T_{k,l}$ for $\widehat{V}_n(\theta)$. Assumptions 3) and 4) ensure the uniform convergence using the proof technique of Theorem 1 of Delyon and Portier, 2018.

B.4 Proof of Equation (3.32)

The proof relies on the Taylor expansions of $\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2})$ and $\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})$ around the parameter value θ_*:

$$\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) = \dot{\tilde{R}}_{n,1}(\theta_*) + \ddot{\tilde{R}}_{n,1}(\theta_*)(\tilde{\theta}_{n,2}-\theta_*) + o\left(\|\tilde{\theta}_{n,2}-\theta_*\|\right).$$
$$\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1}) = \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{\tilde{R}}_{n,2}(\theta_*)(\tilde{\theta}_{n,1}-\theta_*) + o\left(\|\tilde{\theta}_{n,1}-\theta_*\|\right).$$

Using the asymptotic normality of $\tilde{\theta}_n$, we can apply Appendix B.3 in Delyon and Portier, 2018 to prove the convergence of $\ddot{\tilde{R}}_{n,1}(\theta_*)$ and $\ddot{\tilde{R}}_{n,2}(\theta_*)$ to $\ddot{r}(\theta_*)$ in the same spirit as for the proof in Section B.1. We proceed coordinate by coordinate, defining $H_i(\theta_*)_{k,l} = \frac{p(X_i)}{q_{\tilde{\theta}_{i-1}(X_i)}}\ddot{\ell}_{\theta_*}(X_i,S_i)_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}}\frac{p(x)}{q_\theta(x)}\ddot{\ell}_{\theta_*}(x,s)_{k,l}$. Remark that $H_i(\theta_*)_{k,l} \geq 0$, hence we can apply Appendix B.3 in Delyon and Portier, 2018 to obtain the desired convergence. Moreover, the Taylor expansions of $\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,1})$ and $\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,2})$ write:

$$0 = \dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,1}) = \dot{\tilde{R}}_{n,1}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,1}-\theta_*) + o\left(\|\tilde{\theta}_{n,1}-\theta_*\|\right),$$
$$0 = \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,2}) = \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,2}-\theta_*) + o\left(\|\tilde{\theta}_{n,2}-\theta_*\|\right).$$

Finally, the Taylor expansion of $\sqrt{n}(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1}))$ writes:

$$\begin{aligned}
&\sqrt{n}\big(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})\big) \\
&= \sqrt{n}\big(\dot{\tilde{R}}_{n,1}(\theta_*) - \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,2}-\theta_*) - \ddot{r}(\theta_*)(\tilde{\theta}_{n,1}-\theta_*)\big) + o_P(1) \\
&= 2\sqrt{n}(\ddot{\tilde{R}}_{n,1}(\theta_*) - \ddot{\tilde{R}}_{n,2}(\theta_*)) + o_P(1),
\end{aligned} \tag{B.2}$$

because $\sqrt{n}\|\tilde{\theta}_{n,j}-\theta_*\| = O_P(1)$ for $j = 1, 2$. The right-hand side of equation (B.2) weakly converges towards the centered Gaussian distribution with covariance matrix $8V(q_{\theta_*}, \dot{\ell}_{\theta_*})$.

APPENDIX C

Bayesian estimation of seismic fragility curves using objective priors

The content of this appendix is devoted to the work performed by Antoine Van Biesbroeck during his 6th months internship that I supervised during my PhD in CEA Saclay, EMSI Laboratory.

This work deals with Bayesian method for the estimation of seismic fragility curve. Indeed, due to the computational cost of mechanical simulations, the sample size is hence limited. Bayesian methods have the advantage to propose to incorporate prior knowledge on the seismic fragility curve. It is done by assuming a prior probability distribution on the seismic fragility curve parameters. From this prior distribution a posterior distribution is obtained using the Bayes theorem after data observations. However, the elicitation of the prior distribution on the parameters is far from trivial and can influence the posterior distribution for small sample sizes. Therefore, the internship's objective was to defined an *objective* prior distribution - i.e. a prior distribution that maximizes the information brought by the observed data - for the usually considered log normal parametrical model of seismic fragility curve. After a theoretical work on the objective prior distribution definition, the methodology developed was illustrated on a simple mechanical test-case.

C.1 Introduction

In Seismic Probabilistic Risk Assessment (SPRA), fragility curves express the failure probability of a structure as a function of a seismic intensity measure (IM) criterion. They are useful decision making tools widely used for structures subject to earthquakes, such as nuclear power plants (R. Kennedy, C. Cornell, et al., 1980). The estimation of such a fragility curve is still today a complex and hugely studied issue because, in practice, the estimate can be numerically expensive for complex structures modeled by sophisticated numerical mechanical models. Therefore, since data are obtained through time-consuming mechanical simulations, this precludes the use of certain data-consuming methodologies (Gidaris, Taflanidis, and Mavroeidis, 2015). A possible point of view is to see these curves as the evaluations of the complementary cumulative distribution function of a damage measure (DM) conditionally to the intensity measure (IM) $P_f(a) = \mathbb{P}(DM > C|IM = a)$, where C is the failure criterion. In general, a log-normal model is

considered leading to the following expression of the fragility curve

$$P_f(a) = \mathbb{P}(DM > C|IM = a) = \Phi\left(\frac{\ln a - \ln \alpha}{\beta}\right) \tag{C.1}$$

whose study is reduced to the approximation of two parameters α and β, Φ being the cumulative density function of a standard normal distribution. The assumption of log-normal correlation between DM and IM is convenient but remains questionable (Karamlou and Bochini, 2015), and several non-parametric methodologies have also been studied during the last decade (Mai, Konakli, and Sudret, 2017; Sainct et al., 2020; Z. Wang et al., 2018). In this work we will stay under the parametric log-normal model, and focus on a new approach under this framework. Numerous ways already exist to estimate these parameters and a review of a range of the most common ones is proposed by Zentner, Gündel, and Bonfils, 2017. The simplest approaches are based on maximum likelihood estimation (Shinozuka et al., 2000; Gehl, Douglas, and Seyedi, 2015; Baker, 2015), but they suffer from irregularities on low data. The Bayesian approach has the purpose of reaching a better accuracy and regularity, regarding the smallness of the sample size. It consists into attributing a prior distribution to the parameters that need to be estimated. These priors are for now chosen subjectively (Damblin et al., 2014; Keller et al., 2015; Jalayer, De Risi, and Manfredi, 2015), making questionable their relevance, and jeopardizing agreement concerning their validity.

To handle this last issue, our idea is to construct with an objective criterion a good and reliable prior for the parameters. This approach is the one of Nalisnick and Smyth, 2017 in a more general context but they limit their study to a parametric family of priors. Relying on the reference priors theory particularly developed by J. M. Bernardo, 1979; J.. Bernardo and Smith, 1994; J. Bernardo, 2005; J. O. Berger, J. M. Bernardo, and Sun, 2009, we propose a complete approach to build an objective prior suitable for seismic fragility curves, with the aim of estimating them.

In the next section, the lognormal model used for seismic fragility curve estimation is described. Section C.2 presents definitions and choices of reference priors. These objective priors will then be used for the estimation of fragility curves in Section C.4. Section C.5 is finally dedicated to the conclusion.

C.2 Parametric log-normal model for seismic fragility curves

In the context of estimating fragility curves, both the question of the seismic intensity measure and the choice of the parametric model arise. First, numerous criteria are used as a metric for seismic intensity measure Hariri-Ardebili and Saouma, 2016b, and there is no consensus in the literature on the most relevant beyond them. In fact, the choice depends of the mechanical problem and must be done in practice after a prior look to some specific criteria (efficiency, sufficiency, and scale robustness) Grigoriu and Radu, 2021. The Peak Ground Acceleration (PGA), which is the maximal acceleration measured in absolute value, is mostly used because of the facility of its calculation and the commonly admitted relation with the induced seismic forces Solomos, Pinto, and Dimova, 2008. Studied structures are generally sensitive to ground motions on some of their eigenmodes, which makes the consideration of spectral acceleration indicators relevant Feau, 2019. Note that in some works several IMs are taken into consideration at once Lagaros and Fragiadakis, 2007; Sainct et al., 2020.

Second, about choosing a one-dimensional correlation model between inputs (i.e. input excitation reduced to some relevant parameters or IMs) and outputs (damage measure), linear models, for their simplicity and the knowledge we have about them, are the most employed. For seismic fragility curves estimation, the assumed linear link is between the logarithm of the damage measure DM with any of the presented above IMs' logarithm. This model is what we call log-normal

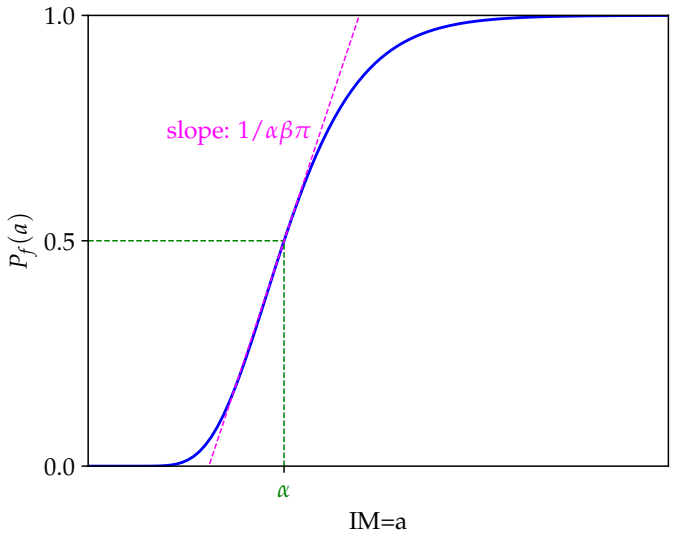


FIGURE C.1: Seismic fragility curve example as a log-normal curve. See equation (C.1) for P_f's expression.

and is expressed below

$$\log DM = c \log IM + d + \sigma\mathcal{N}(0,1).$$

To support this choice, fig. C.2 shows $\log DM$ as a function of $\log IM$. One can see a hazardous chaos scattered around a straight line, here plotted from linear regression calculation. The choice of this model stays imperfect and therefore is questionable. However, we must admit the complexity of such a prediction with a limited number of data. So, in this context, the use of a one-dimensional parametric model is justified because its "imprecision" is counterbalanced by a more precise estimate of its parameters.

Finally, this leads to the analytical expression of fragility curves stated in eq. (C.1). Parameters α and β, computed from C, c, d and σ are the ones we want to estimate. An example of such a fragility curve under the log-normal assumption is provided in fig. C.1.

Problem formalization Our model may be seen as a statistical model with independent observations $((a_1, z_1), \ldots, (a_k, z_k))$, k being the data-set size. For an earthquake i, a_i is its observed IM and z_i is the observation of a failure: $z_i = \mathbb{1}_{DM_i > C}$. The couples (a_i, z_i) are independent realizations of the random variable $(IM, \mathbb{1}_{DM>C})$, defined on a probability space $(\Omega, \mathscr{T}, \mathbb{P})$. We express the k-tuple of them by A and Z, with distribution respectively denoted $\mathbb{P}_A$ and $\mathbb{P}_Z$. Thus, $(A, Z) \in \mathcal{A} \times \mathcal{Z}$ a.s. with $\mathcal{Z} = \{0,1\}^k$ and $\mathcal{A} \subset \mathbb{R}_+^k$. T is the random variable denoting the researched parameters (a realization of T is noted $\theta = (\alpha, \beta)$). The distribution of T is noted π and called the prior, defined on a set $\Theta \subset \mathbb{R}_+^2$. As fragility curves should depend only on the structure's response and not on the earthquake characteristics, A and T can be supposed being independent. We use probability kernels notation to denote conditional distributions: $\mathbb{P}_{X_1,X_2} = \mathbb{P}_{X_1} \circ \mathbb{P}_{X_2|X_1}$ where $\mu \circ \nu(B_1, B_2) = \int_{B_1} \nu(B_2, x) d\mu(x)$. In our problem, given the log-normal relation (C.1) and the independence of the observations, the conditional distribution of Z with respect to (A, T) can be written:

$$\mathbb{P}_{Z|A,T}(B, a, \theta) = \sum_{z \in B} \prod_{i=1}^{k} \Phi\left(\frac{\log \frac{a_i}{\alpha}}{\beta}\right)^{z_i} \left(1 - \Phi\left(\frac{\log \frac{a_i}{\alpha}}{\beta}\right)\right)^{1-z_i}. \tag{C.2}$$

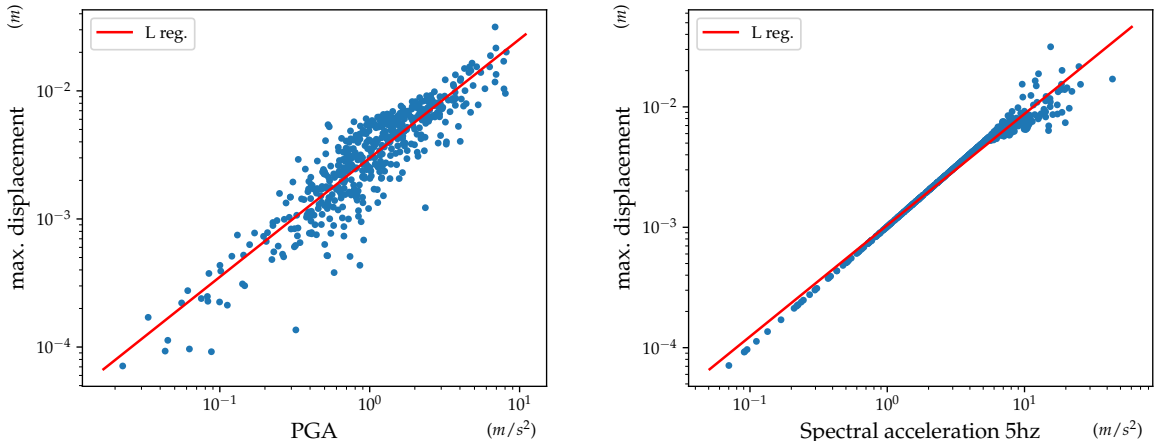


FIGURE C.2: Maximal displacement of the oscillator presented in section C.4.3, as a function of some IMs in log scale. Red lines are linear regressions of the data.

where $B \subset \mathcal{Z}$. $\mathbb{P}_A$ is approximated by kernels from the empirical distribution we observed Wasserman, 2004. The method is the following: we assume $\mathbb{P}_A$ is absolutely continuous with respect to the Lebesgue measure and we approximate its density f_A as

$$f_A(a) \simeq \frac{1}{M} \sum_{i=1}^{M} \frac{1}{h} K\left(\frac{a_i - a}{h}\right) \tag{C.3}$$

where $(a_i)_{i=1}^M$ denote the data and $K(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ the Gaussian kernel. h and σ are generally chosen as $\sigma^2 = \mathrm{Var}(A)$ (approximated numerically) and $h = M^{-1/5}$.

C.3 Information theory and reference priors

This section is dedicated to explore the prior choice in the Bayesian context. We discuss different objective prior definitions. In subsection C.3.1 is introduced the notion of mutual information. Its definition and its usefulness in Bayesian problems is not new J. M. Bernardo, 1979; Muré, 2018, however, its computation under the framework of seismic fragility curves has not been studied yet in the literature.

Then, we have decided to clearly express two reference priors definitions: non asymptotic and asymptotic, respectively detailed in C.3.2 and C.3.3. Non-asymptotic reference priors are generally not studied, as they are used for an introduction for asymptotic ones, making our consideration new. Our point of view is that a dependency of the reference prior with respect to the number of observations can be seen as relevant information rather than an issue. Particularly, our work bring a rigorous mathematical framework and proper demonstrations of some main theorems that did not exist yet.

The link between asymptotic and non-asymptotic reference prior is clearly expressed and discussed in C.3.3.

C.3.1 Average expected information

This section is dedicated to explore the choice of prior in the Bayesian context. Hereafter is discussed the construction of an objective prior. Shannon's information theory provides relevant elements for our problem, the information entropy is a common example of them, and allows to call a distribution informative or non-informative E. Jaynes, 1982. Former approaches as objective prior choices are to look for non-informative ones (i.e. with high entropy), as any information would be considered subjective. In a context of finite number of events, an uniform prior meets this criterion. Our approach goes further, a non-informative prior is a relevant choice as its impact will be little in the posterior computation Muré, 2018. This Bayesian point of view is detailed below, after the next useful definition.

Definition 1 (Kullback-Leibler divergence)**.** Let μ and ν be two probability measures absolutely continuous with respect to a measure λ. We denote p and q their respective Nikodym derivatives with respect to λ. The Kullback-Leibler divergence of μ with respect to ν is

$$KL(\mu||\nu) = \int p(x) \log \frac{p(x)}{q(x)} d\lambda(x).$$

This quantity is often seen as a "distance" between the distributions μ and ν (yet it is not symmetrical). In fact, looking at Shannon's information theory, we can see that this is the information that μ would provide if it were chosen instead of ν (lower weight is given to areas where μ and ν's are close). In the Bayesian framework K-L divergence can be used to get an idea about the information provided by the posterior distribution with respect to the prior. Therefore, our idea consists in maximizing with respect to π the quantity $KL(\mathbb{P}_{T|A,Z}(\cdot, a, z)||\pi)$ with our notations. We can notice that it is a function of a and z, this is why we consider its expected value, hence the following definition.

Definition 2 (Average expected information)**.** Define a statistical model $\mathcal{M} = \{X, \mathbb{P}_{X|\theta}, \theta \in \Theta\}$, the average expected information for a prior π in the model $\mathcal{M}$ is the quantity (if it exists)

$$I(\pi|\mathcal{M}) = \int KL(\mathbb{P}_{T|X}(\cdot, x)||\pi) d\mathbb{P}_X(x)$$

where $\mathbb{P}_{T|X}$ and $\mathbb{P}_X$ are calculated considering $T \sim \pi$ and $\mathbb{P}_{X|T}(\cdot, \theta) = \mathbb{P}_{X|\theta} \, \forall \theta \in \Theta$.

Our problem finally is

$$\text{Find } \underset{\pi \in \mathcal{P}}{\operatorname{argmax}} I(\pi|\mathcal{M}) = \int_{\mathcal{Z} \times \mathcal{A}} KL(\mathbb{P}_{T|A,Z}(\cdot, a, z)||\pi) d\mathbb{P}_{Z,A}(z, a) \tag{C.4}$$

for some class $\mathcal{P}$ of measures over Θ under which the above formula has a sense. In fact, in our work we will consider proper priors, i.e. $\mathcal{P}$ will denote a class of probability measures.

The letter $\mathcal{M}$ reminds us how this quantity depends on the model. While it is not expected to change all along this work, it is noticeable that I's expression and its – if it exists – maximizer depends of the number observations we suppose to have (even if no empirical observation is taken into account for its calculation). It is this last remark which explains the dissociation between the two following subsections in which a finite number of observations is first considered, before examining what happens under the assumptions of an "asymptotic model".

In what follows, the notation $\mathcal{M}^k$ is adopted to express that the number of observations is k. For infinite observations assumption, we will write $\mathcal{M}$. Such an asymptotic model is not properly defined and we do not try to express it. This notation only reminds that the point of view is asymptotic as as if k were growing to infinity. Further details are developed in subsection C.3.3.

Contextual calculation of I In order to derive I in the context of seismic fragility curves, a proper computation of the $\mathbb{P}_{T|A,Z}$ and $\mathbb{P}_{A,Z}$ distributions is needed, as regard to the available quantities described in section C.2. This is done within the two following lemmas, demonstrated in Appendix C.6.

Lemma 1. *Let us denote $\mathscr{Z}$ and $\mathscr{A}$ the sigma-algebras on $\mathcal{Z}$ and $\mathcal{A}$. The joint distribution of (Z, A) can be expressed as follows:*

$$\mathbb{P}_{A,Z}(B_1 \times B_2) = \int_{B_1} \int_{\Theta} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\pi(\theta) d\mathbb{P}_A(a)$$

for any $B_1 \in \mathscr{A}$, $B_2 \in \mathscr{Z}$.

Lemma 2. *Suppose that there exists a measure c defined on $\mathcal{Z}$ such that for any $(a, \theta) \in \mathcal{A} \times \Theta$, $\mathbb{P}_{Z|A,T}(\cdot, a, \theta)$ is absolutely continuous with respect to c with Nikodym derivative equal to $z \mapsto p(z|a, \theta)$. Then the posterior distribution, which is the conditional law of T knowing (A, Z), is defined by the following kernel:*

$$\mathbb{P}_{T|A,Z}(F, a, z) = \frac{\int_F p(z|a, \theta) d\pi(\theta)}{\int_{\Theta} p(z|a, \theta) d\pi(\theta)}$$

for any $F \in \mathscr{B}(\Theta)$, $a \in \mathcal{A}$, $z \in \mathcal{Z}$.

Let us remind that equation (C.2) ensure the satisfactions of lemma 2's hypothesis, c being the discrete measure on $\mathcal{Z}$ and $p(z|a, \theta)$ being the likelihood of our model:

$$p(z|a, \theta) = \prod_{i=1}^{k} \Phi\left(\frac{\log \frac{a_i}{\alpha}}{\beta}\right)^{z_i} \left(1 - \Phi\left(\frac{\log \frac{a_i}{\alpha}}{\beta}\right)\right)^{1-z_i} \tag{C.5}$$

Thus, the quantity $I(\pi|\mathcal{M}^k)$ we want to maximize can be written as we do below.

$$\begin{aligned} I(\pi|\mathcal{M}^k) &= \int_{\mathcal{A}\times\mathcal{Z}} \int_{\Theta} \frac{p(z|a, \theta)}{\int_{\Theta} p(z|a, \theta) d\pi(\theta)} \log \frac{p(z|a, \theta)}{\int_{\Theta} p(z|a, \theta) d\pi(\theta)} d\pi(\theta) d\mathbb{P}_{A,Z}(a, z) \\ &= \int_{\mathcal{A}\times\mathcal{Z}} \int_{\Theta} p(z|a, \theta) \log \frac{p(z|a, \theta)}{\int_{\Theta} p(z|a, \theta) d\pi(\theta)} d\pi(\theta) dc(z) d\mathbb{P}_A(a). \end{aligned} \tag{C.6}$$

C.3.2 Reference priors under a finite number of observations

This maximization problem of the average expected information has already been studied by J.. Bernardo and Smith, 1994, §5.4. Indeed, they developed a heuristic from which they derived an implicit solution, that led them to a first definition of reference priors. In this subsection we propose a definition for a reference prior and express its solution into a clear mathematical framework.

Note that our problem is to maximize with respect to π the function $I(\pi|\mathcal{M}^k)$ defined in equation (C.6). As this function can be written as

$$I(\pi|\mathcal{M}^k) = \int_{\mathcal{T}} \int_{\Theta} g(t, \theta) \log \frac{g(t, \theta)}{\int_{\Theta} g(t, \theta) d\pi(\theta)} d\pi(\theta) d\mu(t) \tag{C.7}$$

where $t = (z, a)$, $\mathcal{T} = \mathcal{Z} \times \mathcal{A}$, $g(t, \theta) = p(z|a, \theta)$ and $\mu = c \otimes \mathbb{P}_A$. We will consider in the following this more general expression with g continuous, non negative and such that its integral over $\mathcal{T}$ is equal to 1. This formulation includes the generally studied problem of maximizing with respect

to p $\tilde{I}(p|\mathcal{M}^k) = \int_{\mathcal{Z}} \int_{\Theta} p(z|\theta)p(\theta) \log \frac{p(z|\theta)}{\int_{\Theta} p(z|\theta)p(\theta)d\theta} d\theta dz$. One can find below our rigorous definition of a reference prior.

Definition 3 (Reference prior)**.** We call $\pi^* \in \mathcal{P}$ a reference prior over the class $\mathcal{P}$ of priors if $\pi^* \in \operatorname{argmax}_{\pi \in \mathcal{P}} I(\pi|\mathcal{M}^k)$.

Maximizing this function leads to a result that depends on k Muré, 2018. As stated before this is generally seen as an issue which makes this maximizer not really considered. Indeed, the works we cited are looking for a reference prior adapted to a model, which they consider intrinsically not dependent on the number of observations. However, as expressed before, the low values of available sample sizes is a central characteristic of our problem. As a consequence, we thought that a proper study of this reference prior could be relevant.

A solution of this problem has been conjectured by J.. Bernardo and Smith, 1994. Under a clear mathematical framework and some regularity assumptions, we bring in C.6 demonstrations of the following results.

Theorem 1. *Assume $\Theta \subset \mathbb{R}^d$ is compact and $g(t,\theta) \in [l_1, l_2] \subset \mathbb{R}_+^* \ \forall t, \theta$. Consider a finite measure ν on Θ and $\mathcal{P}$ the class of the probability measure on Θ that are absolutely continuous with respect to ν and admitting a continuous and positive Nikodym derivative. Then there exists an unique $\pi^* = \operatorname{argmax}_{\pi \in \mathcal{P}} I(\pi|\mathcal{M}^k)$. It is such that $p^* = \frac{d\pi^*}{d\nu}$ verifies :*

$$p^* \propto f \qquad \text{where } f(\theta) = \exp\left(\int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta)p^*(\theta)}{\int_{\Theta} g(t,\tilde{\theta})p^*(\tilde{\theta})d\nu(\tilde{\theta})} d\mu(t) \right).$$

Theorem 2. *Under the assumption of theorem 1, consider $\mathcal{P}' = \mathcal{P} \cap \{\pi, \int_{\Theta} g_i d\pi = c_i \ \forall i \in [\![1, n]\!]\}$. Then there exists an unique $\pi^* = \operatorname{argmax}_{\pi \in \mathcal{P}'} I(\pi|\mathcal{M})$. It is such that $p^* = \frac{d\pi^*}{d\nu}$ verifies*

$$p^* \propto f \times \exp\left(\sum_{i=1}^{n} \lambda_i g_i \right)$$

for some $\lambda_i \in \mathbb{R}$.

Knowing the expression of a reference prior allows us to calculate the reference prior under some additional constraints, following theorem 2. The idea behind that statement is to impose moments constraints after an expert judgement on the parameters, choosing for example $g_i = \mathrm{id}^i$. The required calculation of the resulting λ_i is discussed in C.3.4.

The above theorems give an implicit expression of the optimal priors which may seem useless at a first look. However, reminding what the function g corresponds to, allows us to recognize in the definition of f the posterior distribution density with respect to ν, deduced from the result of lemma 2:

$$\frac{d\mathbb{P}_{T|A,Z}}{d\nu}(\theta, a, z) = p^*(\theta|a, z) = \frac{p(z|a,\theta)p^*(\theta)}{\int_{\Theta} p(z|a,\theta)p^*(\theta)d\nu(\theta)}$$

considering that T is distributed with respect to $\pi^* = p^*\nu$.

This posterior distribution is clearly unknown, yet it can be approximated as follows:

$$\tilde{p}^*(\theta|a, z) = \frac{p(z|a,\theta)h(\theta)}{\int_{\Theta} p(z|a,\theta)h(\theta)d\nu(\theta)} \tag{C.8}$$

for a chosen density h. This basic formula suggested in J.. Bernardo and Smith, 1994 is based on the fact that under an infinite number of observations and some regularity conditions, the

resulting prior should be asymptotically independent of the posterior approximation. This will be discussed in next subsection.

C.3.3 Asymptotic reference priors

We stated that J.. Bernardo and Smith, 1994 pointed out an independence between the prior and the posterior approximation when the number of observations goes to infinity. The study of what we call asymptotic reference priors came out of this remark. The key idea is that such a prior, maximizer of the asymptotic average expected information, will depend only on the model, and not on any uncertainty brought by the observations, satisfying the Bayesian point of view.

This statement could be seen as a consequence of the Bernstein-von Mises' theorem Vaart, 1998. Indeed, according to the theorem, under appropriate assumptions, the posterior density $p(\sqrt{k}(\theta - \hat{\theta}_k)|t)$ converges to a Gaussian density – whose variance is the inverted square-root of the Fisher information $\mathcal{I}(\theta)^{-1/2}$ which is defined later (see definition 6) – where $\hat{\theta}_k$ is an asymptotically sufficient and consistent estimator, such as the Maximum Likelihood Estimator (MLE) under general assumptions. This convergence does not depend on the prior and so any approximation from (C.8) would lead to what follows. Thus, we asymptotically have

$$\log p(\theta|t) \simeq \log \frac{k}{2\pi} - \frac{k}{2}\|\mathcal{I}(\theta)^{1/2}(\hat{\theta}_k - \theta)\|^2 + \frac{1}{2}\log\det\mathcal{I}(\theta)$$

and consequently to the convergence of $k\|\mathcal{I}(\theta)^{1/2}(\hat{\theta}_k - \theta)\|^2$ to a χ^2_2 distribution (whose expected value is $\mathbb{E}\chi^2_2 = 2$), we can approximate

$$\int_{\mathcal{T}} p(t|\theta)\log p(\theta|t)d\mu(t) \simeq \log\frac{k}{2\pi e} + \frac{1}{2}\log\det\mathcal{I}(\theta)$$

combined with theorem 1's reference prior expression, we asymptotically get

$$p^*(\theta) \underset{\sim}{\propto} \exp\frac{1}{2}\log\det\mathcal{I}(\theta).$$

This way, this heuristic developed more rigorously by M. Ghosh, 2011 not only states the intuited asymptotic independence of the reference priors with respect to the construction choices made discussed in (C.8), but also provides an idea of the limit which is verified in what follows.

Mostly studied and theorized in J. Bernardo, 2005; J. O. Berger, J. M. Bernardo, and Sun, 2009, the asymptotic reference priors now benefit of clear frameworks and properties. Muré, 2018, chapter I.2 provides a more general definition of these reference priors that is adapted to our problem. With a few exceptions, we only consider proper priors in this work.

A possible definition for the asymptotic reference prior is the limit of some normalization of what we derive in previous subsection J. M. Bernardo, 1979; J.. Bernardo and Smith, 1994. Yet, we will consider to define it being the maximizer of some non-informative entropy, as we did to define the non-asymptotic reference prior above, and as it is done in recent literature. These definitions actually are equivalent under some regularity assumptions, as we will later see.

The first approach would be to define the asymptotic reference prior as the one which maximizes the limit of the non-informative entropy $I(\pi|\mathcal{M})$. Unfortunately this limit generally is infinite, a more subtle definition is then needed and expressed as follows, we could show that it is equivalent to the heuristic under the right constraints.

Definition 4 (Asymptotic reference prior)**.** Let $\mathcal{M}^k$ denote the model under k observations of the data. We suppose that $\Theta \subset \mathbb{R}^d$ and that there exists an increasing sequence $(\Theta_i)_i$ of compact sets

such that $\bigcup_i \Theta_i = \Theta$. We call a probability distribution $\pi^* \in \mathcal{P}$ an asymptotic reference prior over a class $\mathcal{P}$ if for any other probability distribution $\pi \in \mathcal{P}$ and any i,

$$\liminf_{k\to\infty} I(\pi_i^*|\mathcal{M}^k) - I(\pi_i|\mathcal{M}^k) \geq 0$$

where π_i and π_i^* denote the normalized restrictions of the priors to Θ_i.

Note that contrarily to the previous subsection, Θ is not supposed to be compact anymore. In fact, this asymptotic study is conjectured to lead to a non-implicit solution (contrarily to last section's results) as regard to the posterior convergence we discussed earlier. This will logically lead to solutions that are equal on any compact intersections of Θ, up to a multiplicative constant. That allows us to relax this hypothesis. From now on, we will still consider a measure ν on Θ, not necessarily finite, and focus on priors absolutely continuous with respect to ν.

Next proposition adapted from Muré, 2018 states the uniqueness of a prior satisfying the former definition.

Proposition 1 (Asymptotic reference prior uniqueness)**.** *Assume that for any $\theta \in \Theta$, there exists a weakly consistent estimator of θ and let $\mathcal{P}$ be a convex class of priors. Then if an asymptotic reference prior over $\mathcal{P}$ exists it is unique.*

As we now express clearly the dependency of our variables with respect to the number of observation k, we will denote t_k our data, and $(\mathcal{T}^k, \mathscr{T}^k, \mu_k)$ their living measured space. Next definition set the correct class to which the approximated prior for posterior approximation we evoked in subsection C.3.2 should belong to ensure convergence to the asymptotic reference prior.

Definition 5 (Asymptotic consistency)**.** We call a positive, continuous and proper prior Θ asymptotically consistent if its density h is such that for any $\theta \in \Theta$, $\varepsilon > 0$, denoting $h(\theta|t_k) = \int_\Theta g(t_k, \theta)h(\theta)d\nu(\theta)$, the quantity

$$\int_{|\tau-\theta|\leq\varepsilon} h(\tau|t_k)d\nu(\tau)$$

is a random variable distributed from $g(\cdot, \theta)\mu$ which converges in probability to 1. The class of these density functions is denoted by $\mathcal{Q}$.

The following theorem due to J. O. Berger, J. M. Bernardo, and Sun, 2009 completes our heuristic as it states the convergence of the approximated non-asymptotic reference prior to the asymptotic one.

Theorem 3 (Explicit asymptotic reference prior)**.** *Let $\mathcal{P}$ the class of the probability measure on Θ that are absolutely continuous with respect to ν and admitting a continuous and positive Nikodym derivative. Let h be in $\mathcal{Q}$ and θ_0 be an interior point of Θ. We define the following functions*

$$f_k(\theta) = \exp\left(\int_{\mathcal{T}^k} g(t_k, \theta)\log h(\theta|t_k)d\mu_k(t_k)\right)$$
$$f(\theta) = \lim_{k\to\infty} \frac{f_k(\theta)}{f_k(\theta_0)}.$$

Assume the f_k are continuous and bounded on every compact for k large enough. Then $\pi^* = p^*\nu$ with $p^* \propto f$ is the asymptotic reference prior over class $\mathcal{P}$.

The last theorem expresses the consistency of the non-asymptotic solution approximation we described in subsection C.3.2. With the possibility of simulating some large numbers of observations, it allows the implementation of an algorithm to calculate the asymptotic reference prior

J. O. Berger, J. M. Bernardo, and Sun, 2009. In what follows, we derive an other explicit expression of the asymptotic reference prior, which does not depend of an asymptotical number of observations. This is more suitable to our problem, given the generation complexity of such data. Indeed, Clarke and Barron, 1994 have shown Clarke and Barron, 1994 that Jefferey's prior is an asymptotic reference prior, under the right conditions. One can find below a definition of it.

Definition 6 (Jeffreys prior)**.** Suppose g is twice differentiable with respect to θ and the following integrals converge. The Fisher information matrix $\mathcal{I}^k(\theta) \in \mathbb{R}^{d\times d}$ is defined by its coordinates:

$$\mathcal{I}^k(\theta)_{i,j} = -\mathbb{E}_{t_k\sim g(\cdot,\theta)} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log g(t_k,\theta) = -\int_{\mathcal{T}^k} g(t_k,\theta) \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log g(t_k,\theta) d\mu_k(t_k).$$

We call Jeffreys prior a prior J on Θ absolutely continuous with respect to the Lebesgue distribution, whose density is proportional to $\sqrt{\det \mathcal{I}^k(\theta)}$. Last results presented below propose explicit solutions for asymptotic reference priors under or without constraints.

We remind the relation $\mathcal{I}^k(\theta) = k\mathcal{I}^1(\theta)$ ($\mathcal{I}^1$ will be simply noted $\mathcal{I}$ in the following), ensuring the independence of Jeffreys prior with respect to k.

Theorem 4. *If it is proper, the Jeffreys prior is the asymptotic reference prior over the class of all probability distributions on* Θ*.*

Proposition 2. *Let* $\mathcal{P}$ *be the class defined in theorem 3, we define* $\mathcal{P}_0 = \mathcal{P} \cap \{\pi,\ \int_\Theta g_i d\pi = c_i\ \forall i\}$ *for some continuous* g_i*. Then the asymptotic reference prior* π_0^* *over the class* $\mathcal{P}_0$ *is such that* $p_0^* = \frac{d\pi_0^*}{d\nu}$ *verifies*

$$p_0^* \propto p^* \exp\left(\sum_{i=1}^n \lambda_i g_i\right)$$

for some $\lambda_i \in \mathbb{R}$*, where* p^* *denotes the Nikodym derivative of the asymptotic reference prior over class* $\mathcal{P}$*, derived in theorem 3.*

Proposition 2 is, as far we know, new. It can be demonstrated by adapting theorem 3's proof in J. O. Berger, J. M. Bernardo, and Sun, 2009. Next corollary came from theorem 4 and proposition2.

Corollary 1. *If the Jeffreys prior is proper and continuous, then the asymptotic reference prior over class* $\mathcal{P}_0$ *defined in proposition 2 with* ν *being the Lebesgue measure admits for Nikodym derivative:*

$$p_0^*(\theta) \propto \sqrt{\det \mathcal{I}(\theta)} \exp\left(\sum_{i=1}^n \lambda_i g_i(\theta)\right)$$

for some $\lambda_i \in \mathbb{R}$*.*

C.3.4 Lagrange multipliers derivation

In theorem 2 and proposition 2 we derived under-constraints solutions whose expressions involve the unconstrained solutions and some Lagrange multipliers. The latter are such that the constraints hold but this condition does not necessarily provide a convenient method to compute them. In this section are presented some derivations adapted for some particular and often encountered cases.

Mean constraint The following method is due to Lemaitre, 2014 Lemaitre, 2014, §3.3.1, who derive a solution from a different but similar problem that the one considered there. Suppose the constraint is a mean constraint: π should be such that $\int_\Theta \theta_i d\pi(\theta) = m_i \, \forall i$. Then results stated before ensure the solution admits a density of the form

$$p_0^*(\theta) = K p^*(\theta) e^{\langle \lambda, \theta \rangle}$$

for some $\lambda \in \mathbb{R}^d$ and $K \in \mathbb{R}$, with p^* denoting the unconstrained solution. K depends on λ as it is the normalization constant, its calculation leads to the following:

$$p_0^*(\theta) = p^*(\theta) e^{\langle \lambda, \theta \rangle - \psi(\lambda)} \qquad \psi(\lambda) = \log \int_\Theta e^{\langle \lambda, \theta \rangle} p^*(\theta) d\nu(\theta)$$

ψ can be seen as the logarithm of the moment generating function M_T of the variable $T \sim p^* \nu$. Knowing its expression allows us then the computation of λ as it is shown below. From

$$\nabla M_T(\lambda) = \int_\Theta \theta e^{\langle \lambda, \theta \rangle} p^*(\theta) d\nu(\theta) = \int_\Theta \theta p_0^*(\theta) e^{\psi(\lambda)} d\nu(\theta) = M_T(\lambda) m$$

thus $m = \nabla M_T(\lambda) / M_T(\lambda)$. This equation should enable the calculation of λ.

Mean and variance constraints What is done above can be adapted to the constraints $\int_\Theta \theta d\pi(\theta) = m$ and $\int_\theta (\theta - m)(\theta - m)^T d\pi(\theta) = \Sigma$. Note that this second constraint can be replaced equivalently by $\int_\Theta \theta\theta^T d\pi(\theta) = S$, $S = \Sigma + mm^T$. Therefore we have

$$p_0^*(\theta) = K p^*(\theta) e^{\lambda^T \theta + \mathrm{Tr}(\Lambda^T \theta \theta^T)} = K p^*(\theta) e^{\lambda^T \theta + \theta^T \Lambda^T \theta}$$

for some $K \in \mathbb{R}$, $\lambda \in \mathbb{R}^d$ and $\Lambda \in \mathbb{R}^{d \times d}$. Therefore, denoting $N_T(\lambda, \Lambda)$ the function $\int_\Theta p^*(\theta) e^{\lambda^T \theta + \theta^T \Lambda^T \theta} d\nu(\theta)$ gives us, similarly than before:

$$\nabla_\lambda N_T(\lambda, \Lambda) / N_T(\lambda, \Lambda) = m \qquad \nabla_\Lambda N_T(\lambda, \Lambda) / N_T(\lambda, \Lambda) = S.$$

Note that the previous methods can easily be generalized. Their only need is to know the function $N_T(\lambda) = \int_\Theta p^* \exp \sum_i \lambda_i g_i d\nu$, and be able to solve the equations $\nabla_{\lambda_i} N_T(\lambda) / N_T(\lambda) = c_i$.

C.4 Fragility curves estimation with reference priors

C.4.1 Posterior simulation method

Once a prior for the parameters is chosen, different Bayesian methods exist to estimate fragility curves. As an example, Damblin et al., 2014 introduce loss functions which serve to express the right parameters as their minimizers. A simpler method often conducted in literature (e.g. Jalayer, De Risi, and Manfredi, 2015) is to simulate a range of parameters under the computed posterior distribution. These can give an idea of a desired confidence interval, derived from some simulations' quantiles. This is how we process in this section.

The posterior density is simply calculated – up to a constant – as the product between the prior density and the likelihood. The expression of a commonly known density is generally not recognized (e.g. with Jeffreys prior), we perform for this reason the simulations through an Metropolis-Hasting (MH) algorithm.

The resulting estimations that we made from this method and the consistency of our MH's simulation are discussed later on.

C.4.2 Jeffreys prior calculation

Theorem 4 and corollary 1 expressed the usefulness of the Jeffreys prior in the context of reference prior construction. Also, according to the heuristic we discussed in C.3.3, a good posterior approximation could be derived from the Fisher information matrix. Consequently, this section is devoted to its calculation.

We compute $\mathcal{I}(\theta)$ in our model. Here, $\theta = (\alpha, \beta) \in \mathbb{R}_+^2$ and

$$\mathcal{I}(\theta)_{i,j} = -\int_{\mathcal{Z}\times\mathcal{A}} p(z|a,\theta) \frac{\partial^2}{\partial\theta_i\theta_j} \log p(z|a,\theta) dc(z) d\mathbb{P}_A(a).$$

for $i, j \in [\![1, 2]\!]$. With $c = \delta_0 + \delta_1$ and $p(z|a,\theta)$ as expressed in (C.5) – with $k = 1$ –, i.e.

$$\log p(z|a,\theta) = z \log \Phi\left(\frac{\log a - \log \alpha}{\beta}\right) + (1-z) \log\left(1 - \Phi\left(\frac{\log a - \log \alpha}{\beta}\right)\right).$$

We denote $\gamma = \beta^{-1} \log \frac{a}{\alpha}$, the first order partial derivatives with respect to θ of $\log p(z|a,\theta)$ are the followings:

$$\begin{aligned}
\frac{\partial}{\partial\alpha} \log p(z|a,\theta) &= -\frac{1}{\alpha\beta} z \frac{\Phi'(\gamma)}{\Phi(\gamma)} + \frac{1}{\alpha\beta}(1-z)\frac{\Phi'(\gamma)}{1-\Phi(\gamma)} \\
\frac{\partial}{\partial\beta} \log p(z|a,\theta) &= -\frac{\log\frac{a}{\alpha}}{\beta^2} z \frac{\Phi'(\gamma)}{\Phi(\gamma)} + \frac{\log\frac{a}{\alpha}}{\beta^2}(1-z)\frac{\Phi'(\gamma)}{1-\Phi(\gamma)}
\end{aligned}$$

which lead to the second order partial derivatives computed below:

$$\begin{aligned}
\frac{\partial^2}{\partial\alpha\partial\beta} \log p(z|a,\theta) = -\frac{1}{\beta}\frac{\partial}{\partial\alpha} p(z|a,\theta) &+ \frac{\log\frac{a}{\alpha}}{\alpha\beta^3} z \frac{\Phi''(\gamma)\Phi(\gamma) - \Phi'(\gamma)^2}{\Phi(\gamma)^2} \\
&- \frac{\log\frac{a}{\alpha}}{\alpha\beta^3}(1-z)\frac{\Phi''(\gamma)(1-\Phi(\gamma)) + \Phi'(\gamma)^2}{(1-\Phi(\gamma))^2}
\end{aligned} \tag{C.9}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\alpha^2} \log p(z|a,\theta) = -\frac{1}{\alpha}\frac{\partial}{\partial\alpha} \log p(z|a,\theta) &+ \frac{1}{\alpha^2\beta^2} z \frac{\Phi''(\gamma)\Phi(\gamma) - \Phi'(\gamma)^2}{\Phi(\gamma)^2} \\
&- \frac{1}{\alpha^2\beta^2}(1-z)\frac{\Phi''(\gamma)(1-\Phi(\gamma)) + \Phi'(\gamma)^2}{(1-\Phi(\gamma))^2}
\end{aligned} \tag{C.10}$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2} \log p(z|a,\theta) = -\frac{2}{\beta}\frac{\partial}{\partial\beta} \log p(z|a,\theta) &+ \frac{\log^2\frac{a}{\alpha}}{\beta^4} z \frac{\Phi''(\gamma)\Phi(\gamma) - \Phi'(\gamma)^2}{\Phi(\gamma)^2} \\
&- \frac{\log^2\frac{a}{\alpha}}{\beta^4}(1-z)\frac{\Phi''(\gamma)(1-\Phi(\gamma)) + \Phi'(\gamma)^2}{(1-\Phi(\gamma))^2}
\end{aligned} \tag{C.11}$$

From these, the Fisher information matrix is computed numerically, from approximations of the integrals over $\mathcal{A}$. See appendix C.7 for more details.

As we express in the following, Jeffreys prior needs to be iterated numerous times through our computations (mainly due to the use of MCMC methods). However, as the complexity of its calculation is high due to the integrals to be calculated, we decided to perform a calculation first, based on a fine-mesh grid of $\mathbb{R}_+^2$, giving us an interpolated approximation of Jeffreys prior more suitable for our numerical applications. Fig. C.3 proposed a plot of it. To be precise, 500×500 prior values have been computed for $\alpha \in [10^{-5}, 10]$ and $\beta \in [10^{-3}, 2]$. A linear interpolation has been processed from these.

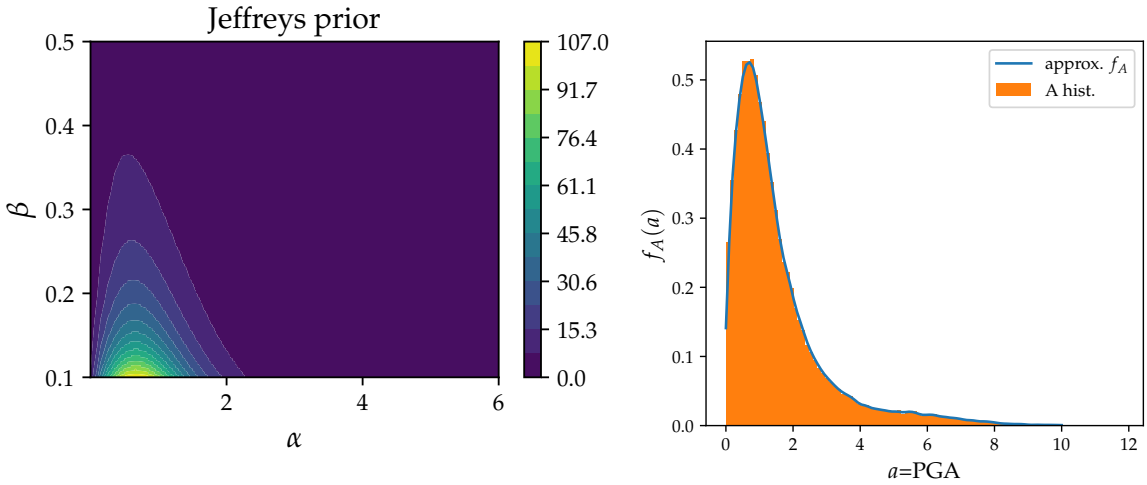


FIGURE C.3: Jeffreys prior plot, calculated from PGA (at left). Approximation of the PGA's probability density function (at right) for the dataset used in section C.4.3.

Jeffreys prior is known to be improper in numerous common cases. As it is clear that it diverges around $(0.65, 0)$ in our study, this does not necessarily implies that it is not proper. However, we have not demonstrated the contrary either. In this work, the prior is numerically truncated, thus an approximation of Jeffreys prior and posterior – yet definitely proper – are actually computed for the experiments presented in section C.4.4.

One can clearly see the influence of the IM's distribution (also plotted in figure C.3) on Jeffreys prior. That is explainable as follows: giving a higher prior weights to fragility curves whose median happens to be around high probability IM values will enhance any learning from the observations. This naturally makes the information brought by the posterior likelihood with respect to the prior higher.

From this remark we deduce how important is the studied IM's distribution to construct an objective prior. Some works, in which these data are uniformly generated or artificially augmented, and so not realistic, are therefore questionable.

C.4.3 Choices and data-set

The results of previous section give us a substantial range of different possible objective priors according to the desired criteria.

First, the dependency of the prior on the observations set's size is an interesting point of view. Moreover, as regard to what we discussed in subsection C.3.3, a good approximation of the datasize-dependent reference prior could be computed in accordance with equation (C.8) using Jeffreys prior for the density h. In fact, many Bayesian studies prefer a focus onto the asymptotic considerations, due to (i) the demonstrated convergence, (ii) the perception of asymptotic model as more objective and (iii) the higher convenience of its calculation. However, we think that it

could be interesting to take into account the information provided by the number of observations, often limited in our field of application. Unfortunately, because of the additional calculations required, we do not go further on this trail and leave its exploration for possible future works.

Secondly, different results have been presented on the subject of objective prior with "under constraints" criteria (e.g. theorem 2 and proposition 2). As it is not the case in our work since we stay on a theoretical case, these results could be useful in a study dealing for example with expert judgements about the parameters. However, one should keep in mind that such a consideration could compromise the 100% objective nature of the prior we have built so far.

Thus, only Jeffreys prior has been derived and used as a prior in this report, considering it as the more general and objective possible prior.

About seismic intensity measures, we focused on the peak ground acceleration (PGA) discussed in section C.2 and expressed hereafter. This choice is mostly common and general in earthquake engineering and one can easily notice that our methodology can be directly adapted to any different choice. Indeed, it only affects numerical calculations. We think that a validation of our work on the PGA makes its validation on others really likely.

Presentation of the problem of interest For this work and our numerical simulations we run our computations from the same seismic and mechanical models that Sainct et al., 2020 used.

A ground motion s is a signal simulated here as a filtered white noise with time dependent parameters following the work done in Rezaeian and Der Kiureghian, 2010:

$$s(t) = q(t, \boldsymbol{\rho}) \left[\frac{1}{\sigma_f(t)} \int_{-\infty}^{t} h[t-\tau, \boldsymbol{\lambda}(\tau)] w(\tau) d\tau \right],$$

where $\sigma_f(t)^2 = \int_{-\infty}^{t} h[t-\tau, \boldsymbol{\lambda}(\tau)] d\tau$ is the standard deviation and q is a time-dependant modulating function

$$\begin{aligned} q(t, \boldsymbol{\rho}) &= 0 && \text{if} \quad t \leq T_0, \\ &= \rho_1 \left(\tfrac{t-T_0}{T_1-T_0} \right)^2 && \text{if} \quad T_0 \leq t \leq T_1, \\ &= \rho_1 && \text{if} \quad T_1 \leq t \leq T_2, \\ &= \rho_1 \exp\left[-\rho_2 (t-T_2)^{\rho_3}\right] && \text{if} \quad t \geq T_2. \end{aligned}$$

For more details about the parameters and the impulse response function h, one can refer to Sainct et al., 2020, §2. From such a signal, several IMs are calculated:

- the $PGA = \max_{t \in [0,T]} |s(t)|$;
- the Peak Ground Velocity $PGV = \max_{t \in [0,T]} \left| \int_0^t s(\tau) d\tau \right|$;
- the Peak Ground Displacement $PGD = \max_{t \in [0,T]} \left| \int_0^t \int_0^\tau s(u) du d\tau \right|$;
- the total energy $E = \int_0^T s^2(\tau) d\tau$;
- the spectral acceleration $SA = 4\pi^2 f_L^2 \max_{t \in [0,T]} |\tilde{y}(t)|$. $\tilde{y}$ being calculated from a linear system modeling structure, as expressed in equation (C.12).

As shown in figure C.4, the simulated structure is a nonlinear single degree of freedom oscillator with kinematic hardening.

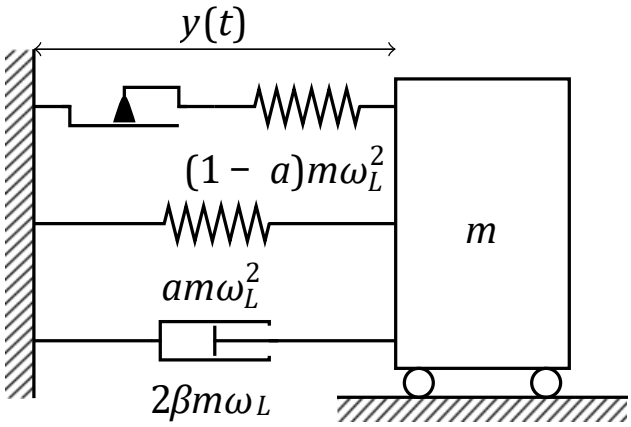


FIGURE C.4: Scheme of the nonlinear single degree of freedom oscillator used for the numerical simulations. m is the mass, $\omega_L = 2\pi f_L$ the circular frequency, a the post-yield stiffness defining kinematic hardening, β the damping ratio and $y(t)$ is the relative displacement of the mass with respect to the ground.

Still following Sainct et al., 2020, its relative displacement y is ruled by the following equation

$$y''(t) + 2\beta 2\pi f_L y'(t) + f^{nl}(y(t)) = -s(t)$$

under a ground motion s. f^{nl} being the nonlinear resisting force, β the damping ratio, and f_L the frequency (in this study $f_L = 5Hz$). In addition, we consider the associated linear system – supposed to be known – whose relative displacement $\tilde{y}$ follows the equation

$$\tilde{y}''(t) + 2\beta 2\pi f_L \tilde{y}'(t) + 4\pi^2 f_L^2 \tilde{y}(t) = -s(t), \tag{C.12}$$

From these computations is derived the displacement measure $DM = \max_{t\in[0,T]} |y(t)|$. The failure is considered when $DM > C$ with $C = 8 \cdot 10^{-3} m$. A total of 10^5 signals and responses have been generated this way. 10% of them resulted into the failure of the structure.

C.4.4 Numerical results

In this section we present the results of the numerical simulations with seismic fragility curves estimation purpose, in accordance with the method described in section C.4.1.

As expressed in the last section, for posterior calculation, we illustrate the Jeffreys prior choice method, which is computed considering the PGA as seismic intensity measure. Then, we compare our estimations with the maximum likelihood estimation (MLE) method. Our resulting curves are compared with a reference log-normal curve, whose parameters are the maximum likelihood estimator based on the total available data set (10^5 observations). To analyze and compare the methods we have done our simulations under two different sample sizes, 51 and 101, since the objective is to assess the quality of the estimates for a “small" number of data.

500 MLE simulations have been performed for the 2 samples and are presented in figure C.5. It is easy to see the "irregularities" that this method involves, especially when dealing with small size samples. These are characterized by some estimations of β as 0, resulting into vertical confidence intervals.

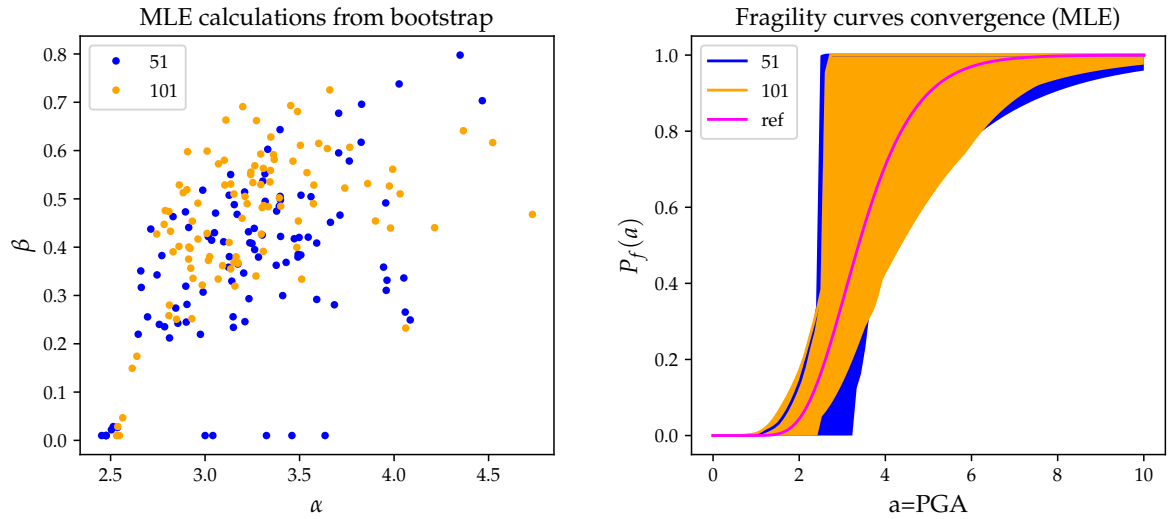


FIGURE C.5: Scatter plots and resulting 95% confidence intervals of 500 MLE simulations, for a sample size of 51 observations first, and 101 then.

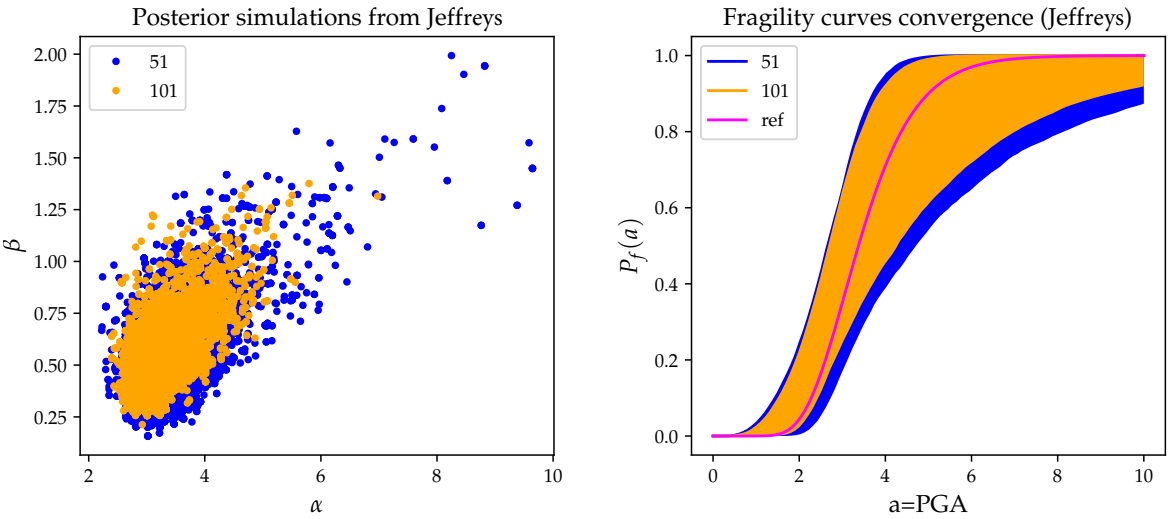


FIGURE C.6: Scatter plots and resulting 95% confidence intervals of 10000 posterior simulations from Jeffreys prior, for a sample size of 51 observations first, and 101 then.

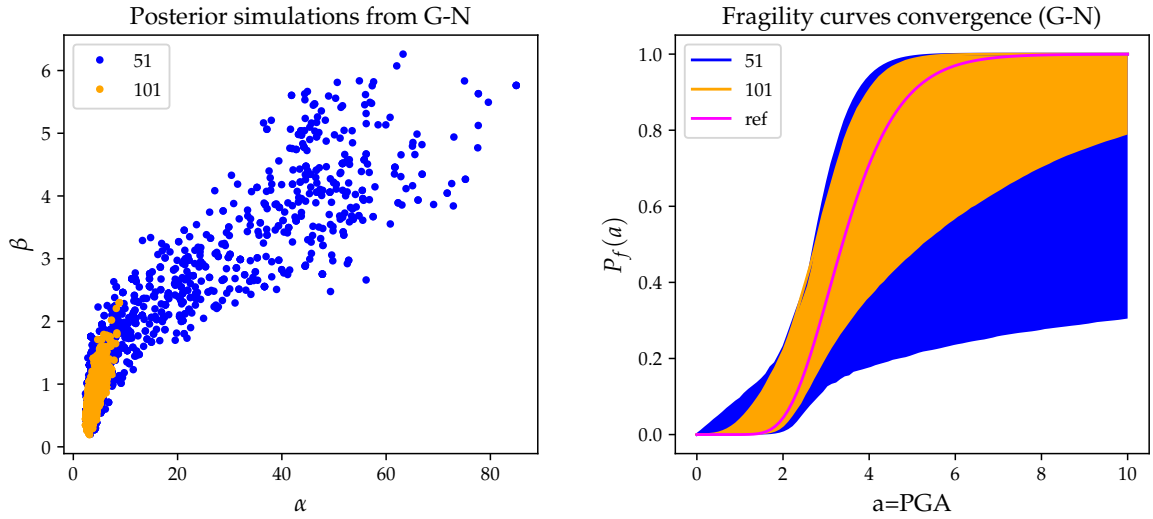


FIGURE C.7: Scatter plots and resulting 95% confidence intervals of 10000 posterior simulations from Gamma-Normal prior, for a sample size of 51 observations first, and 101 then.

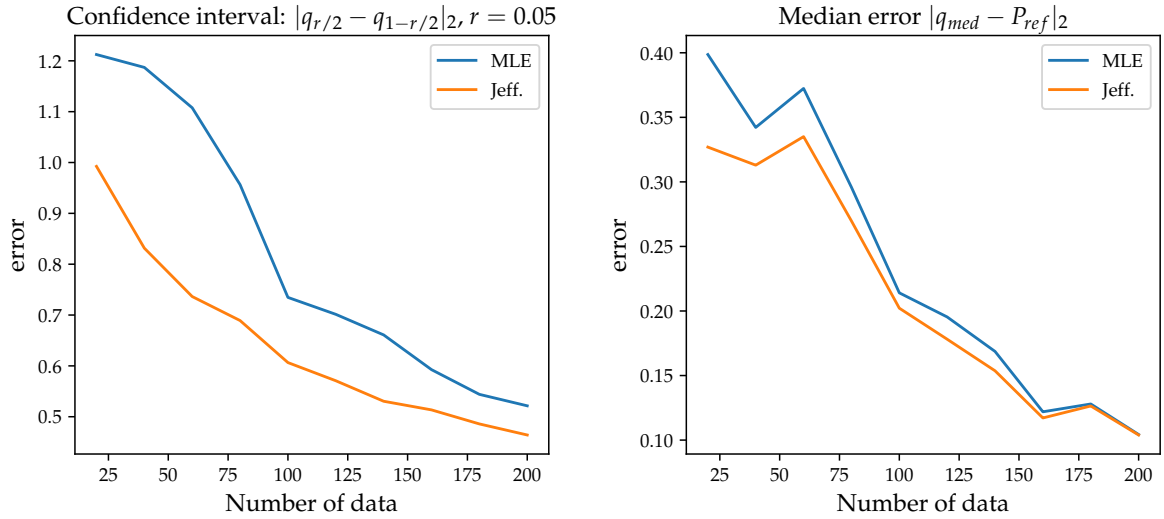


FIGURE C.8: Evolutions of the average, over 10 replications, of the "size" of the 95% confidence interval and the mean square error between the medians of the estimates and the reference fragility curve, as a function of the sample size.

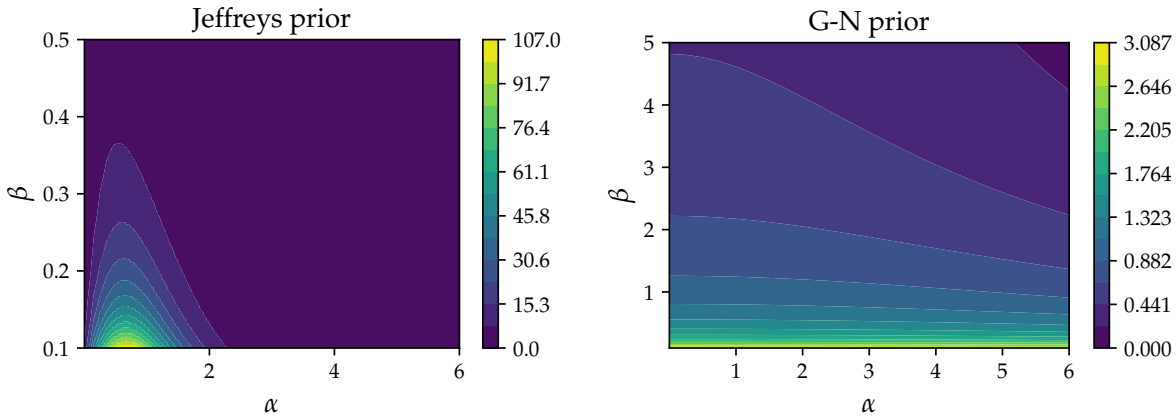


FIGURE C.9: Jeffreys prior (at left) and Gamma-Normal prior (at right).

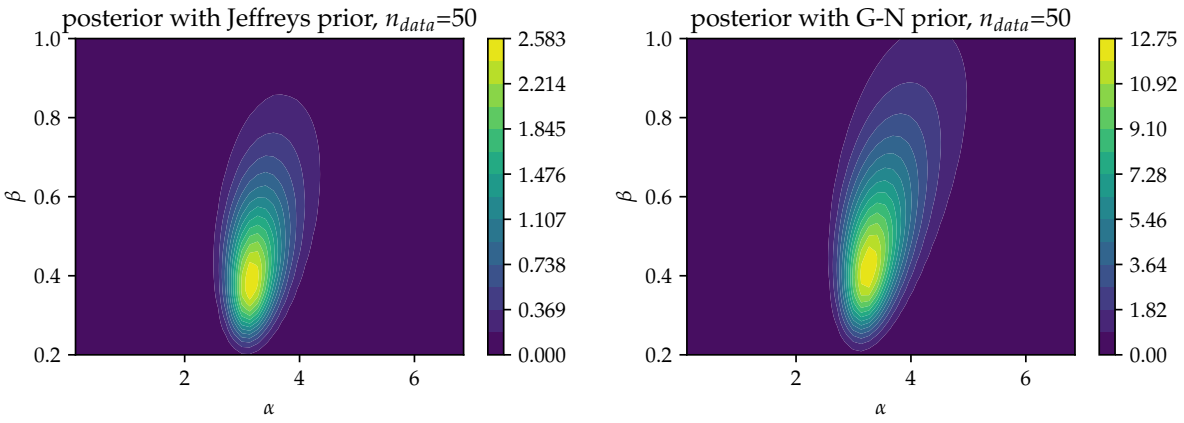


FIGURE C.10: Posterior plots from a number of 50 observations. Priors are Jeffreys (at left) and a Gamma-Normal $\mathcal{GN}(0, 0.1, 0.1, 0.1)$ distribution (at right).

These irregularities do not exist in Bayesian simulations. Plots of figure C.6 come from 10000 simulations of the posterior with Jeffreys prior. The scatter plot does not contain any of such limit points discussed above (i.e. $\beta = 0$), which is see-able on the shape of the confidence intervals that more closely match the shape of the reference curve.

To have a deep insight into these results, one can refer to figure C.8 where the evolutions, with respect to the sample size, of the "size" of the confidence interval and the mean square error between the medians of the estimates and the reference fragility curve are compared. These results correspond to an average based on 10 replications.

We can immediately see the advantage of the Bayesian method over that of the MLE when it comes to small samples. The fact that both curves are getting closer validates our convergence intuitions: with a rising number of observations, the weight of the likelihood become stronger, compared to the prior, in the posterior expression.

In the framework of Bayesian estimation of fragility curves, a Gamma-Normal prior is often chosen, generally fitted with parameters making it poorly informative. In fact, such a subjective choice would be more questionable if it provided too much information. This Gamma-Normal also has the advantage of making calculations and simulations convenient, as in log-normal model the resulting posterior also follows a Gamma-Normal distribution whose parameters are known. For these reasons, we decided to perform, for comparison, a Bayesian estimate from such a Gamma-Normal prior. We have done the choice of the distribution $\mathcal{GN}(0, 0.1, 0.1, 0.1)$ in accordance with Damblin et al., 2014, which is plotted in figure C.9, next to Jeffreys prior plot.

In figure C.10 both posteriors derived from Jeffreys (at left) and Gamma-Normal (at right) priors are plotted. On can notice that the low information feature of Gamma-Normal prior implies that higher value of (α, β)'s areas have a substantial weight in the posterior than in Jeffreys one. This is particularly true on a little sample size as one can observe in our simulations results plotted in figure C.7. Confidence intervals are larger, especially in case $N_{data} = 51$, because numerous large couples (α, β) are simulated. This phenomenon is comparable to overfitting, because it gives a neutral probability to all observable events.

C.4.5 MCMC consistency

As stated before, MH algorithm has been used for posterior simulations purpose. Because we do not have a totally clear view on the prior's features we have chosen to implement a variant that can adapt to the different probability density functions. Our explicit choice is an adaptive MH algorithm with Gaussian transition distribution Haario, Saksman, and Tamminen, 2001 which is summarized in algorithm 6.

Our simulations have been realized by keeping the 10000 last values of a total of 40000 iterations, and one can see in figures C.13 and C.14 the average convergences of the parameters through the MCMC iterations. The expected values all seem to converge relatively quickly (after 10000 iterations) questioning the burn-in of 30000 we did. However, under 51 observations, Gamma-Normal posterior seems to struggle more, and does not seem stable before the 10000 last iterations, hence our choice. This difficulty to converge is probably due to the slow decrease of that posterior we already discussed in previous subsection. The acceptance rates we got, plotted in figure C.15 are acceptable as being asymptotically around 0.4.

One can find the histograms of our resulting simulated parameters in figures C.11 and C.12, to get an idea of the density function of the simulations.

In addition, we computed a 90% confidence zone from the posterior densities. After the simulations, it was checked and confirmed that the same proportion of the simulated variables felt inside this zone.

Algorithm 6 Adaptive MH

Require: π: target distribution

t_0: Markov chain initialization

$N_{iter}, N_{step}, \sigma_p, b, s$: calibration parameters

for $k = 1 \dots N_{iter}$ **do**
  if $k < N_{step}$ **then**
    $t^* = t_{k-1} + \mathcal{N}(0, \sigma_p I)$
  else
    $t^* = t_{k-1} + s(1-b)\mathcal{N}(0, \hat{\Sigma}_{k-1}) + b\mathcal{N}(0, \sigma_p I)$
  end if
  $\alpha_k = \pi(t^*)/\pi(t_{k-1})$
  $t_k = t^*$ with probability α_k, $t_k = t_{k-1}$ with probability $1 - \alpha_k$
  $m_k = \frac{1}{k}\sum_{i=1}^{k} t_i$
  $\hat{\Sigma}_k = \sum_{i=1}^{k}(t_i - m_k)(t_i - m_k)^T$
end for

Ensure: t, α: Markov chain and acceptance rates sequences

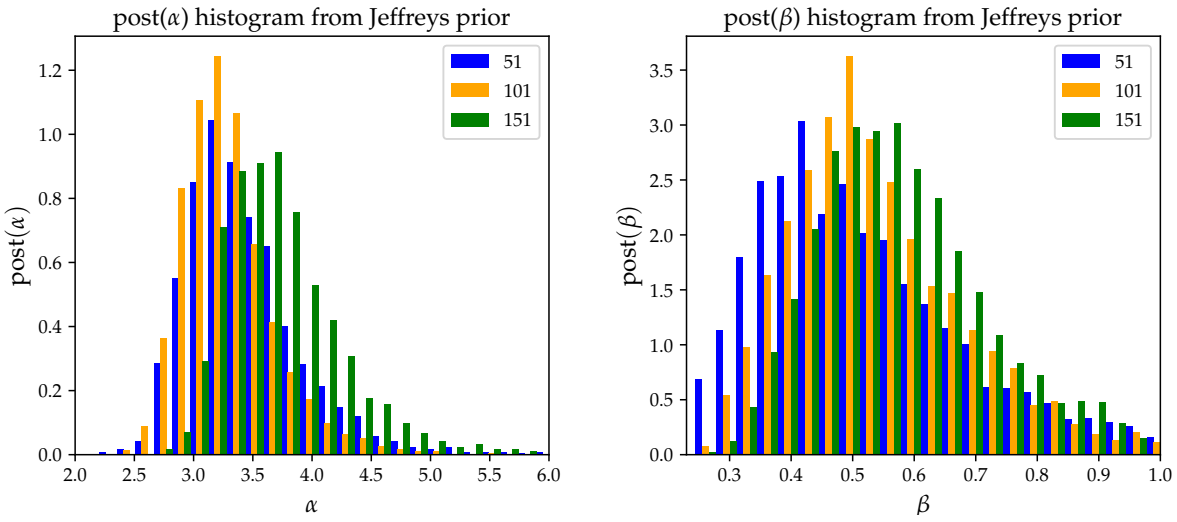


FIGURE C.11: Histograms of resulting simulated α and β through MCMC simulation of Jeffreys posterior. Results of 3 simulations with 3 different sample sizes are plotted.

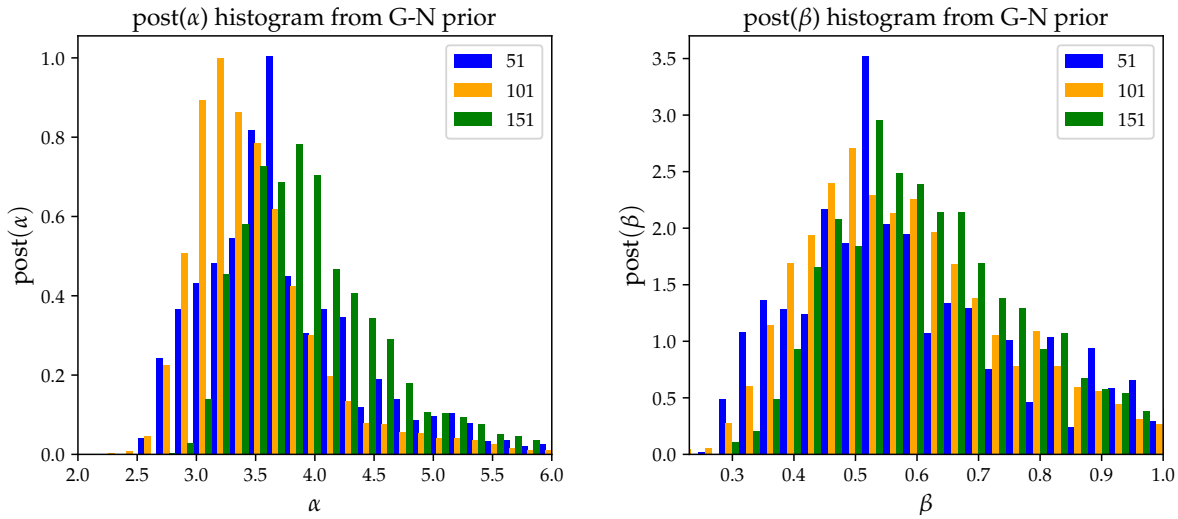


FIGURE C.12: Histograms of resulting simulated α and β through MCMC simulation of G-N posterior. Results of 3 simulations with 3 different sample sizes are plotted.

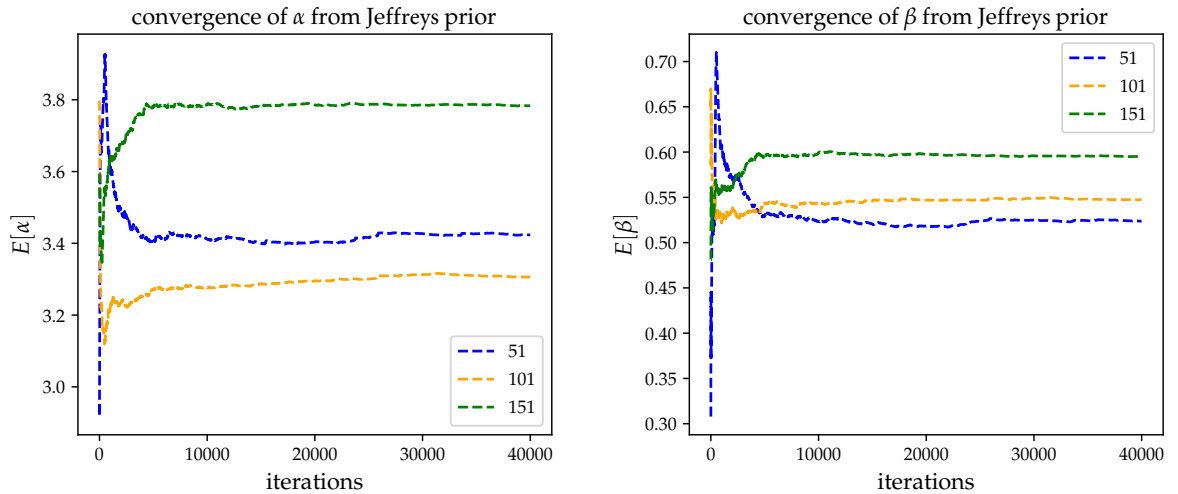


FIGURE C.13: Cumulative average values of α (at left) and β (at right) through MH iterations of Jeffreys posterior. Results of 3 simulations with 3 different sample sizes are plotted.

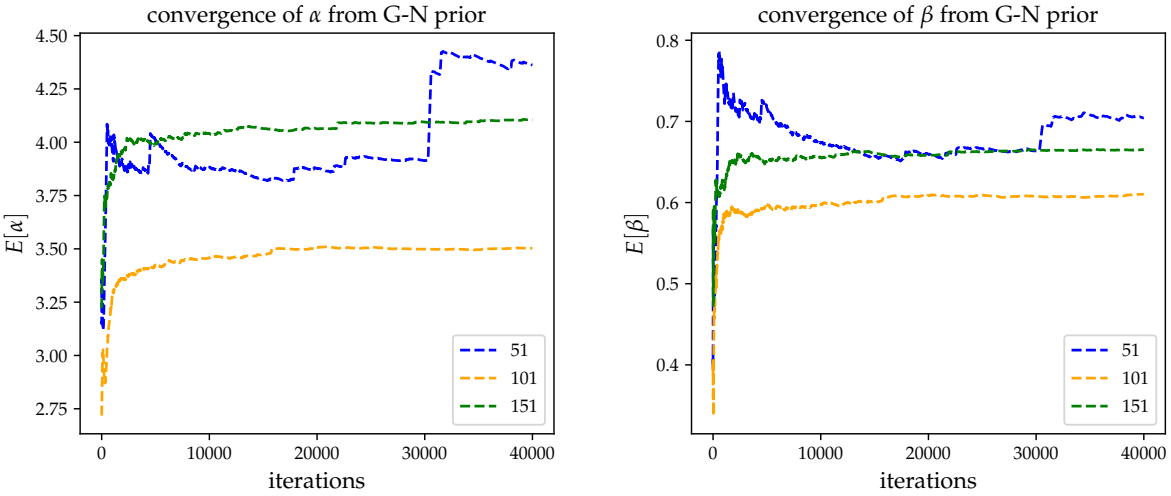


FIGURE C.14: Cumulative average values of α (at left) and β (at right) through MH iterations of G-N posterior. Results of 3 simulations with 3 different sample sizes are plotted.

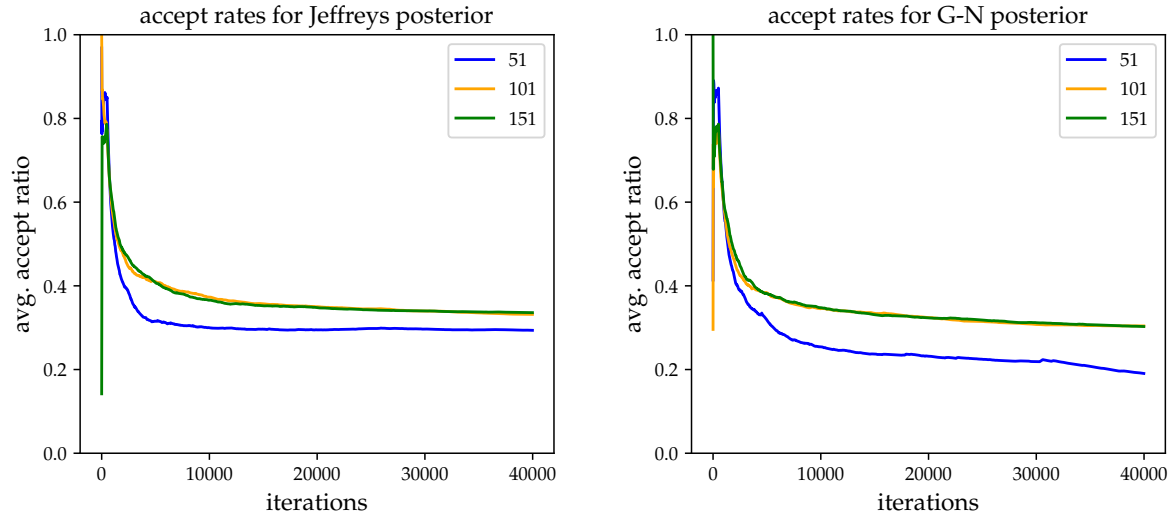


FIGURE C.15: Average acceptance rates through the MH iterations for Jeffreys and Gamma-Normal posteriors. Results of 3 simulations with 3 different sample sizes are plotted.

C.5 Conclusion & prospects

In this work, we have implemented a method for Bayesian estimation of seismic fragility curves, which oversteps the general prior choice issue, whose subjectivity can hardly be entirely justified. We have studied a substantially theoretical and robust framework giving few trails about different objective prior constructions. While we focused in this work on Jeffreys prior, which is a theoretically known choice but never implemented yet for fragility curves estimation, other objective priors are proposed in this work and could be computed in the future. Indeed, expressing the dependency of the expected information with respect to the sample size is an interesting point of view, as well as prior construction under mean and/or variance constraints. Moreover, for a given seismic scenario, we have shown that such an objective prior is correlated with the distribution of seismic intensity measures, such as PGA, stating the importance of a realistic distributions of such IMs in the data-set.

Our numerical simulations highlighted the limit of bootstrapping for MLE and the irregularities it implies, especially on small data-sets. In addition, they demonstrated the superiority of the Bayesian method to remove these irregularities.

Finally, while the construction of a Bayesian estimation method which does not suffer from a validation jeopardized by a prior subjectively chosen is proper progress on the literature, we have been able to notice a better accuracy of it than the one achieved by the most commonly considered prior in other studies of Bayesian estimation of fragility curves.

We conclude by saying that this work provides a methodology that could be generalized in the context of fragility curves estimation. It only requires the calculation of the Jeffreys prior from the distribution of the IM of interest by following our equations. Therefore, the estimations only result from classical MCMC methods, which are easily reproducible. In the future, a methodology based on Gamma-Normal parameters' fitting could be implemented, choosing the objective priors among the class of Gamma-Normal distributions. This would lead to a lower computation cost.

C.6 Proofs of the theoretical results

C.6.1 Lemma 1

We remind that all the terms in the integrals we manipulate in the demonstration of this lemma or of the following are positive, and so Fubini-Tonelli's theorem stands and the integration orders do not matter.

Note that as A and T are supposed independent, their joint distribution is $\mathbb{P}_{A,T} = \mathbb{P}_A \otimes \pi$ and therefore for any $B_1 \in \mathscr{A}$, $B_2 \in \mathscr{Z}$

$$\int_{B_1} \int_{\Theta} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\pi(\theta) d\mathbb{P}_A(a) = \int_{B_1 \times \Theta} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\mathbb{P}_{A,T}(a, \theta)$$

which can be recognized being equal to the composition $\mathbb{P}_{A,T} \circ \mathbb{P}_{Z|A,T}(B_2 \times (B_1 \times \Theta))$ which is by definition equal to $\mathbb{P}_{A,Z,T}(B_1 \times B_2 \times \Theta) = \mathbb{P}_{A,Z}(B_1 \times B_2)$ as the event $\{T \in \Theta\}$ is the universe. We therefore have demonstrated the following equality for every elements B_1 and B_2 of the tributes of $\mathcal{A}$ and $\mathcal{Z}$:

$$\mathbb{P}_{A,Z}(B_1 \times B_2) = \int_{B_1} \int_{\Theta} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\pi(\theta) d\mathbb{P}_A(a)$$

C.6.2 Lemma 2

To demonstrate the result, let us first derive the conditional law of (A, Z) knowing T. Note that according to Jirina's theorem Stroock, 1993, all the requested conditional law exist. We consider the probability kernel ν defined on $(\mathscr{A} \otimes \mathscr{Z}) \times \Theta$ by $\nu(B_1 \times B_2, \theta) = \int_{B_1} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\mathbb{P}_A(a)$ for any $B_1 \in \mathscr{A}$, $B_2 \in \mathscr{Z}$, $\theta \in \Theta$. We calculate below $\pi \circ \nu(B_1 \times B_2 \times F)$, for $B_1 \in \mathscr{A}$, $B_2 \in \mathscr{Z}$, $F \in \mathscr{B}(\Theta)$.

$$\pi \circ \nu(B_1 \times B_2 \times F) = \int_F \nu(B_1 \times B_2, \theta) d\pi(\theta) = \int_F \int_{B_2} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\mathbb{P}_A(a) d\pi(\theta)$$

As T and A are independents, their joint distribution is $\mathbb{P}_{A,T} = \mathbb{P}_A \otimes \pi$, therefore

$$\pi \circ \nu(B_1 \times B_2 \times F) = \int_{B_1 \times F} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\mathbb{P}_{A,T}(a, \theta) = \mathbb{P}_{A,Z,T}(B_1 \times B_2 \times F)$$

we conclude that $\pi \circ \nu = \mathbb{P}_{A,Z,T}$ which implies the equality $\mathbb{P}_{A,Z|T} = \nu$.

To compute the conditional distribution of T knowing (A, Z), we start from the distribution equality $\mathbb{P}_{A,Z} \circ \mathbb{P}_{T|A,Z} = \mathbb{P}_{A,Z,T} = \pi \circ \mathbb{P}_{A,Z|T}$, which can be re-written for any $B_1 \in \mathscr{A}$, $B_2 \in \mathscr{Z}$, $F \in \mathscr{B}(\Theta)$ as:

$$\int_{B_1 \times B_2} \mathbb{P}_{T|A,Z}(F, a, z) d\mathbb{P}_{A,Z}(a, z) = \int_F \mathbb{P}_{A,Z|T}(B_1 \times B_2, \theta) d\pi(\theta) \tag{C.13}$$

Lemma 1 allows us to express $\mathbb{P}_{A,Z}$ as a function of $\mathbb{P}_{Z|A,T}$, and our hypothesis about that last kernel implies that $\mathbb{P}_{A,Z}$ is absolutely continuous with respect to the measure $c \otimes \mathbb{P}_A$:

$$\mathbb{P}_{A,Z}(B_1 \times B_2) = \int_{B_1} \int_\Theta \int_{B_2} p(z|a, \theta) dc(z) d\pi(\theta) d\mathbb{P}_A(a)$$

This allows us to express the left hand side of equation (C.13) as follows:

$$\int_{B_1 \times B_2} \mathbb{P}_{T|A,Z}(F, a, z) d\mathbb{P}_{A,Z}(a, z) = \int_{B_1} \int_{B_2} \mathbb{P}_{T|A,Z}(F, a, z) \int_\Theta p(z|a, \theta) d\pi(\theta) dc(z) d\mathbb{P}_A(a) \tag{C.14}$$

Furthermore, in the right hand side of equation (C.13), the conditional kernel $\mathbb{P}_{A,Z|T}$ can be replaced with the expression of ν defined above, which leads to:

$$\begin{aligned} \int_F \mathbb{P}_{A,Z|T}(B_1 \times B_2, \theta) d\pi(\theta) &= \int_F \int_{B_1} \mathbb{P}_{Z|A,T}(B_2, a, \theta) d\mathbb{P}_A(a) d\pi(\theta) \\ &= \int_{B_1} \int_{B_2} \int_F p(z|a, \theta) d\pi(\theta) dc(z) d\mathbb{P}_A(a) \end{aligned} \tag{C.15}$$

As equation (C.13) expresses that (C.14) and (C.15) are equal and as it stands for any $(B_1, B_2) \in \mathscr{A} \times \mathscr{Z}$, we can conclude

$$\mathbb{P}_{T|A,Z}(F, a, z) \int_\Theta p(z|a, \theta) d\pi(\theta) = \int_F p(z|a, \theta) d\pi(\theta)$$

for any $F \in \mathscr{B}(\Theta)$, $a \in \mathcal{A}$, $z \in \mathcal{Z}$. Hence the result after dividing by $\int_\Theta p(z|a, \theta) d\pi(\theta)$.

C.6.3 Theorem 1

Let $\pi \in \mathcal{P}$, we denote p its Nikodym derivative with respect to ν. $I(\pi, \mathcal{M})$ can be written as follows.

$$I(\pi, \mathcal{M}) = \hat{I}(p) = \int_{\mathcal{T}} \int_{\Theta} g(t,\theta) p(\theta) \log \frac{g(t,\theta)}{\int_{\Theta} g(t,\theta) p(\theta) d\nu(\theta)} d\nu(\theta) d\mu(t)$$

Therefore maximizing $I(\cdot|\mathcal{M})$ over $\mathcal{P}$ is equivalent to maximize $\hat{I}$ over $\mathcal{C}(\Theta, \mathbb{R}_+^*) \cap \{p, \int_{\Theta} p d\nu = 1\}$. In addition, $\hat{I}$ can be expressed as $\hat{I} = \hat{I}_1 + \hat{I}_2$ where

$$\begin{aligned}
\hat{I}_1(p) &= \int_{\mathcal{T}} \int_{\Theta} g(t,\theta) p(\theta) \log g(t,\theta) d\nu(\theta) d\mu(t) \\
\hat{I}_2(p) &= \int_{\mathcal{T}} \psi \left(\int_{\Theta} g(t,\theta) p(\theta) d\nu(\theta) \right) d\mu(t)
\end{aligned}$$

with $\psi(x) = -x \log(x)$ is a strictly concave function. Therefore let us remind that $\mathcal{C}(\Theta, \mathbb{R}_+^*)$ is convex and consider $p \neq q \in \mathcal{C}(\Theta, \mathbb{R}_+^*)$, $\lambda \in]0,1[$, we have:

$$\begin{aligned}
\hat{I}_2(\lambda p + (1-\lambda) q) &= \int_{\mathcal{T}} \psi \left(\lambda \int_{\Theta} g(t,\theta) p(\theta) d\nu(\theta) + (1-\lambda) \int_{\Theta} g(t,\theta) q(\theta) d\nu(\theta) \right) d\mu(t) \\
&> \int_{\mathcal{T}} \left[\lambda \psi \left(\int_{\Theta} g(t,\theta) p(\theta) d\nu(\theta) \right) + (1-\lambda) \psi \left(\int_{\Theta} g(t,\theta) q(\theta) d\nu(\theta) \right) \right] d\mu(t) \\
&> \lambda \hat{I}_2(p) + (1-\lambda) \hat{I}_2(q)
\end{aligned}$$

from which we deduce that $\hat{I}$ is strictly concave.

Note that $\mathcal{C}(\Theta, \mathbb{R})$ provided with norm $\|\cdot\|_{\infty}$ is a Banach set whose $U = \mathcal{C}(\Theta, \mathbb{R}_+^*) \cap \{p, \int_{\Theta} p(\theta) d\theta \in]1-\xi, 1+\xi[\} \, \xi \in]0,1[$ is an open and convex subset as Θ is supposed compact. Next part of the proof will be dedicated to show that $\hat{I}$ is differentiable.

As a consequence of the hypothesis $g(t,\theta) \in [l_1, l_2] \subset \mathbb{R}_+^*$, $|\log g(t,\theta)|$ is bounded. Then for any $p \in U$ we have:

$$\begin{aligned}
|\hat{I}_1(p)| &\leq \int_{\mathcal{T}} \int_{\Theta} g(t,\theta) d\nu(\theta) d\mu(t) \sup_{t,\theta} |\log g(t,\theta)| \|p\|_{\infty} \\
&\leq \nu(\Theta) \sup_{t,\theta} |\log g(t,\theta)| \|p\|_{\infty}
\end{aligned} \tag{C.16}$$

this implies that $\hat{I}_1$ is continuous and then differentiable as it is linear. Its differentiate is $d\hat{I}_1(p) = \hat{I}_1 \, \forall p$.

We consider $0 < \tilde{l}_1 < (1-\xi) l_1 \nu(\Theta)$ and $\tilde{l}_2 > (1+\xi) l_2 \nu(\Theta)$ and write $\hat{I}_2(p) = \Phi_1(p \times G(p))$, with $G = C - \varphi \circ \Phi_2$ where these functions are defined below.

$$\Phi_2 \begin{cases} U \longrightarrow & \overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[) \\ p \longmapsto & \left[(t,\theta) \mapsto \int_{\Theta} g(t,\tilde{\theta}) p(\tilde{\theta}) d\nu(\tilde{\theta}) \right] \end{cases} \qquad \varphi : q \mapsto \log \circ q$$

$$\Phi_1 \begin{cases} \overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\hat{l}_1, \hat{l}_2[) \longrightarrow & \mathbb{R} \\ r \longmapsto & \int_{\mathcal{T}} \int_{\Theta} r(t,\theta) g(t,\theta) d\nu(\theta) d\mu(t) \end{cases} \qquad C = (t,\theta) \mapsto \log g(t,\theta)$$

where $0 < \hat{l}_1 < \min(\log \tilde{l}_1, l_1)$ and $\hat{l}_2 > \max(\log \tilde{l}_2, l_2)$. $\overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[)$ denotes the set of interior points of $\mathcal{C}(\mathcal{T} \times \Theta,]\tilde{\alpha}, \tilde{\beta}[)$. We justify hereafter the correct definition of Φ_2. Let p be in $\mathcal{C}(\Theta, \mathbb{R}_+^*)$. As g is continuous and uniformly bounded, p is bounded and $\nu(\Theta) < \infty$ it is clear that $\Phi_2(p)$ is continuous, then we have

$$(1-\xi)l_1\nu(\Theta) < \Phi_2(p) < (1+\xi)l_2\nu(\Theta)$$

this way $\Phi_2(p) \in \mathcal{C}(\mathcal{T} \times \Theta,](1-\xi)l_1\nu(\Theta), (1+\xi)l_2\nu(\Theta)[)$ which is included in $\overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[)$.

Now we will show that all these functions are differentiable. Φ_2 is linear, and for any $p \in U$, $\|\Phi_2(p)\|_\infty \leq \|p\|_\infty l_1\nu(\Theta)$ thus it is continuous and differentiable. Φ_1 also is linear and for any $r \in \overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\hat{l}_1, \hat{l}_2[)$, $\|\Phi_1(r)\|_\infty \leq \|r\|_\infty \nu(\Theta)$, hence the continuity and the differentiability of Φ. As it is clear that the application $(q, p) \mapsto qp$ is a continuous and differentiable bi-linear application, last task we have to do is to demonstrate the differentiability of φ.

We consider $p \in \overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[)$. Let $\varepsilon > 0$, there exists $\eta > 0$ such that for any $x \in]-\eta, \eta[$,

$$|\log(1+x) - x| \leq \varepsilon |x| \inf_{\mathcal{T}\times\Theta} |p|$$

Therefore if we consider δ such that $\mathcal{B}(p, \delta) \subset \overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[)$ and $\delta / \inf_{\mathcal{T}\times\Theta} |p| < \eta$ it comes for any t, θ and $h \in \mathcal{B}(p, \delta)$,

$$\left|\varphi(p+h) - \varphi(p) - \frac{h}{p}\right|(t,\theta) = \left|\log\left(1 + \frac{h(t,\theta)}{p(t,\theta)}\right) - \frac{h(t,\theta)}{p(t,\theta)}\right| \leq \varepsilon |h(x)|$$

$$\text{Thus } \left\|\varphi(p+h) - \varphi(p) - \frac{h}{p}\right\|_\infty \leq \varepsilon \|h\|_\infty$$

which shows that φ is differentiable with $d\varphi(p)h = h/p$.

Finally, $\hat{I}_2$ is differentiable. And as $\hat{I}_1$ also is as demonstrated at (C.16), $\hat{I}$ is. Moreover, all the linear functions we have studied are $\mathcal{C}^1$, and as $p \mapsto d\varphi(p)$ is continuous on $\overset{\circ}{\mathcal{C}}(\mathcal{T} \times \Theta,]\tilde{l}_1, \tilde{l}_2[)$, $\hat{I}$ actually is $\mathcal{C}^1$ over U. We compute below the differentiate of $\hat{I}$:

$$\begin{aligned} d\hat{I}(p)h &= \Phi_1(hG(p) - pd\varphi(\Phi_2(p))\Phi_2(h)) \\ &= \int_{\mathcal{T}}\int_{\Theta} g(t,\theta)\left(h(\theta)\log\frac{g(t,\theta)}{\int_\Theta g(t,\tilde{\theta})p(\tilde{\theta})d\nu(\tilde{\theta})} - p(\theta)\frac{\int_\Theta g(t,\tilde{\theta})h(\tilde{\theta})d\nu(\tilde{\theta})}{\int_\Theta g(t,\tilde{\theta})p(\tilde{\theta})d\nu(\tilde{\theta})}\right)d\nu(\theta)d\mu(t) \end{aligned} \quad \text{(C.17)}$$

The maximal point we are looking for is the argmax of $\hat{I}$ on U under the constraint $p \in F = g^{-1}(\{0\})$, $g(p) = \int_\Theta p d\nu - 1$. g is clearly continuous and then $\mathcal{C}^1$ on U with $dg(p) = h \mapsto \int_\Theta h d\nu$ clearly surjective for any $p \in F$. F is convex and $\hat{I}$ is concave on F and bounded above, this way, it admits a maximizer p^* over F. According to the Lagrange multipliers theorem Cartan, 2007, there exists $\lambda \in \mathbb{R}$ such that p^* is a critical point of the function $\mathcal{L} = \hat{I} - \lambda g$ i.e.

$$d\mathcal{L}(p^*) = d\hat{I}(p^*) - \lambda dg(p^*) = 0$$

This gives for any $h \in \mathcal{C}(\Theta, \mathbb{R})$ that the expression derived in (C.17) is equal to $\lambda \int_\Theta h(\theta) d\nu(\theta)$. This gives for any $\theta \in \Theta$,

$$\lambda = \int_{\mathcal{T}} g(t,\theta) \left(\log \frac{g(t,\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} - \frac{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} \right) d\mu(t)$$

expressing the exponential of last equation and multiplying by $p^*(\theta)$ leads to what follows.

$$\begin{aligned} p^*(\theta) e^{\lambda+1} &= \exp\left[\int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} d\mu(t) + \log p^*(\theta) \right] \\ p^*(\theta) e^{\lambda+1} &= \exp\left[\int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} d\mu(t) + \int_{\mathcal{T}} g(t,\theta) \log p^*(\theta) d\mu(t) \right] \end{aligned} \tag{C.18}$$

by using the fact that $\int_{\mathcal{T}} g(t,\theta) d\mu(t) = 1$. Final equation (C.19) states the result.

$$p^*(\theta) e^{\lambda+1} = \exp\left[\int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta) p^*(\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} d\mu(t) \right]. \tag{C.19}$$

C.6.4 Theorem 2

This proof follows the proof of theorem 1 that one can find above. We are looking for a maximizer $p^* \in U$ of $\hat{I}$ that satisfies the equality constraints $\int_\Theta g_i(\theta) p^*(\theta) d\nu(\theta) = c_i \, \forall i$ and $\int_\Theta p d\nu = 1$. These constraints are linear and continuous so the concavity of $\hat{I}$ ensure the existence of p^* and we still are under the assumptions of the Lagrange multipliers theorem which states that there exist $\lambda, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$d\hat{I}(p^*)h - \lambda \int_\Theta h d\nu + \sum_{i=1}^n \lambda_i \int_\Theta g_i h d\nu = 0$$

for any $h \in \mathcal{C}(\Theta, \mathbb{R})$. This gives the following, for any $\theta \in \Theta$,

$$\lambda - \sum_{i=1}^n \lambda_i g_i(\theta) = \int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} d\mu(t) - 1$$

and so, similarly as in (C.18), applying exponential function and multiplying by p^* leads to the result:

$$p^*(\theta) e^{\lambda+1} e^{-\sum_{i=1}^n \lambda_i g_i(\theta)} = \exp\left[\int_{\mathcal{T}} g(t,\theta) \log \frac{g(t,\theta) p^*(\theta)}{\int_\Theta g(t,\tilde{\theta}) p^*(\tilde{\theta}) d\nu(\tilde{\theta})} d\mu(t) \right].$$

C.7 Numerical calculation of the Fisher information matrix

In this section we develop the calculation begun in section C.4.2 and explain how we implement it numerically. The equations of the second order partial derivatives of $p(z|a,\theta)$ derived in equations (C.9), (C.10) and (C.11) need to be integrated over $\mathcal{Z}$ and $\mathcal{A}$. Integrating over the discrete variable z first only replace z by $\Phi(\gamma)$ and $(1-z)$ by $1-\Phi(\gamma)$ in the equations.

For the integration with respect to a, we proceed using Simpson's interpolation method,to approximate the integrals (remind f_A's approximation is expressed in (C.3))

$$A_{11} = \int_{\mathcal{A}} \log \frac{a}{\alpha} \frac{\Phi'(\gamma)^2}{\Phi(\gamma)} f_A(a)da \qquad A_{12} = \int_{\mathcal{A}} \log \frac{a}{\alpha} \frac{\Phi'(\gamma)^2}{1-\Phi(\gamma)} f_A(a)da$$
$$A_{21} = \int_{\mathcal{A}} \log^2 \frac{a}{\alpha} \frac{\Phi'(\gamma)^2}{\Phi(\gamma)} f_A(a)da \qquad A_{22} = \int_{\mathcal{A}} \log^2 \frac{a}{\alpha} \frac{\Phi'(\gamma)^2}{1-\Phi(\gamma)} f_A(a)da$$
$$A_{31} = \int_{\mathcal{A}} \frac{\Phi'(\gamma)^2}{\Phi(\gamma)} f_A(a)da \qquad A_{32} = \int_{\mathcal{A}} \frac{\Phi'(\gamma)^2}{1-\Phi(\gamma)} f_A(a)da$$

reminding $\gamma = \beta^{-1} \log \frac{a}{\alpha}$. Therefore, $I(\theta)$ is the following matrix

$$\mathcal{I}(\theta) = \begin{pmatrix} \frac{1}{\alpha^2\beta^2}(A_{31}+A_{32}) & \frac{1}{\alpha\beta^3}(A_{11}+A_{12}) \\ \frac{1}{\alpha\beta^3}(A_{11}+A_{12}) & \frac{1}{\beta^4}(A_{21}+A_{22}) \end{pmatrix}.$$

APPENDIX D

Résumé étendu en français

Contexte

Les *centrales nucléaires* font partie des systèmes d'ingénierie les plus complexes jamais conçus par l'homme, impliquant presque tous les domaines de la physique et de l'ingénierie tels que la mécanique des fluides, la thermohydraulique, la neutronique, la mécanique des structures, les interactions fluide-structure, pour ne citer que quelques exemples. Comme tout système d'ingénierie complexe, il n'est pas à l'abri de défaillances et, en raison de la très longue durée de vie des centrales nucléaires, la sûreté nucléaire est une tâche complexe et difficile. En outre, les conséquences négatives provoquées par les accidents nucléaires - en termes d'atteinte à la santé humaine et à l'environnement - sont si importantes que le niveau de sûreté imposé par les organismes de réglementation nationaux est l'un des plus élevé jamais imposé à un système d'ingénierie complexe.

L'une des sources possibles d'accidents nucléaires pourrait provenir d'un risque naturel tel qu'un tremblement de terre sur le site de la centrale. Le comportement dynamique des différents composants structurels de la centrale lors d'un événement sismique doit être correctement évalué afin de garantir ses conditions opérationnelles. Le principal défi de l'évaluation du risque sismique des centrales nucléaires est le caractère aléatoire inhérent aux séismes en termes de contenu temporel et spectral. Depuis les années 1980, le cadre de l'évaluation probabiliste du risque sismique (SPRA en anglais) a été développé aux États-Unis pour évaluer correctement la sûreté des composants des centrales nucléaires sous excitation sismique, en tenant compte du caractère aléatoire du risque sismique dans un cadre probabiliste.

Depuis les dernières décennies, la sûreté des centrales nucléaires s'appuie de plus en plus sur des modèles numériques haute-fidélité qui visent à être un *jumeau numérique* du système simulé. L'utilisation de simulations numériques est motivée par la possibilité de simuler des phénomènes physiques qui ne peuvent pas être observés ou reproduits par des essais réels en raison de leur complexité (par exemple, le comportement dynamique sous chargement sismique de l'ensemble du système de tuyauterie d'une centrale nucléaire). Le modèle numérique en dynamique des structures repose souvent sur la résolution numérique d'équations différentielles ordinaires ou partielles en utilisant par exemple la simulation par éléments finis.

L'évaluation de la fiabilité et de la sûreté des structures mécaniques implique la prise en compte de diverses sources d'incertitudes. Les incertitudes peuvent provenir de risques naturels tels que le vent ou les chargements sismiques, mais aussi des propriétés physiques des structures (par exemple, les propriétés des matériaux ou les conditions aux limites). Ces incertitudes doivent

être prises en compte pour garantir un niveau de sûreté satisfaisant. Un tel objectif peut être atteint grâce à un *cadre général de quantification des incertitudes* présenté dans la section suivante.

Quantification des incertitudes

La quantification des incertitudes (UQ en anglais) regroupe une grande variété d'outils théoriques et appliqués issus de la théorie des probabilités, des statistiques computationelles et de la simulation stochastique. Le cadre de l'UQ est par définition interdisciplinaire et peut être appliqué à de nombreuses branches de l'ingénierie. La quantification des incertitudes peut être résumé en quelques étapes fondamentales qui peuvent être appliquées à n'importe quelle domaine de l'ingénierie (voir par exemple Sudret, 2007; De Rocquigny et al., 2008; Iooss, 2009). Ces étapes sont détaillées ci-dessous :

- **Spécification du problème:** la première étape est naturellement la définition du système d'ingénierie étudié et la conception du modèle numérique qui simulera ce système. Elle comprend la définition des variables d'entrée du modèle informatique et de ses quantités de sortie qui sont importantes pour l'étude des phénomènes physiques d'intérêt.
- **Modélisation des incertitudes:** la deuxième étape consiste à répertorier toutes les sources possibles d'incertitudes affectant les variables d'entrée du modèle numérique. Le modèle mathématique le plus classique pour la représentation des incertitudes est le cadre probabiliste.
- **Propagation de l'incertitude:** la troisième étape est consacrée à la propagation des incertitudes qui affectent les variables d'entrée aux variables de sortie à travers le modèle numérique. Par conséquent, les variables de sortie sont également entachées d'incertitudes. Des outils statistiques peuvent alors être appliqués aux variables de sortie pour estimer diverses quantités d'intérêt (par exemple, moyenne, quantile, probabilité de dépassement...).
- **Analyse inverse:** cette dernière étape peut être divisée en deux parties: premièrement, la *calibration* du modèle numérique par rapport aux données disponibles, ce qui n'est pas l'objet de ce manuscrit. Deuxièmement, l'étape *d'analyse de sensibilité*, visant à étudier comment l'incertitude sur la variable de sortie peut être attribuée à chaque variable d'entrée (ou sous-ensemble de variables d'entrée).

Dans ce manuscrit, nous supposons que le modèle numérique est préalablement calibré. L'étape d'analyse inverse sera consacrée à l'étape d'analyse de sensibilité.

Énoncé du problème

Cette thèse traite du problème de l'estimation de la *courbe de fragilité sismique* d'une structure mécanique donnée. Cette quantité d'intérêt particulière est clé pour les études probabilistes de sûreté sismique. Elle correspond à la probabilité de défaillance d'une structure mécanique conditionnellement à une *mesure d'intensité sismique*, qui est généralement une valeur scalaire. Une courbe de fragilité sismique est souvent représentée graphiquement comme illustré à la figure 1.1. Après les travaux fondateurs effectués par R. Kennedy, C. Cornell, et al., 1980; R. Kennedy and Ravindra, 1984, l'estimation des courbes de fragilité sismique est maintenant généralement réalisée à l'aide de simulations numériques basées sur des modèles numériques de simulation (voir par exemple Karim and Yamazaki, 2001; Kim and Shinozuka, 2004; Zentner, 2010). Cependant, les

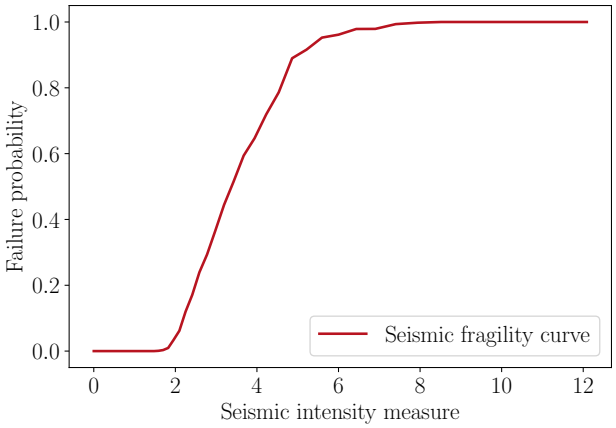


FIGURE D.1: Illustration d'une courbe de fragilité sismique.

paramètres mécaniques et matériaux de la structure sont affectés par des incertitudes provenant de la variabilité intrinsèque à la conception des structures. De plus, les conditions aux limites (par exemple, la fixation de la structure au support) sont très souvent incertaines en raison d'un manque de connaissances et/ou de données. Ainsi, ces incertitudes peuvent affecter les courbes de fragilité sismique de la structure.

Dans cette thèse, l'accent est mis sur la prise en compte des incertitudes qui entachent les paramètres mécaniques d'une structure sur sa courbe de fragilité sismique, en utilisant le cadre de quantification des incertitudes. Par conséquent, la principale problématique étudiée dans ce manuscrit peut être énoncée comme suit.

Comment construire une méthodologie de quantification des incertitudes pour les courbes de fragilité sismique des structures mécaniques ?

Cette problématique peut être décomposée en les questions suivantes :

Q1 - Comment propager les incertitudes des paramètres mécaniques dans les courbes de fragilité sismique avec un temps de calcul raisonnable ?

Q2 - Comment répartir l'incertitude sur la courbe de fragilité sismique entre les différentes sources d'incertitudes épistémiques ?

Q3 - Comment planifier les simulations numériques pour améliorer la précision d'estimation des courbes de fragilité sismique et réduire le nombre de simulations dans le même temps ?

Les deux questions **Q1** et **Q2** sont consacrées à l'adaptation du cadre général de quantification des incertitudes aux études de sûreté sismique, et plus particulièrement aux courbes de fragilité sismique qui sont des quantités clés pour les études probabilistes de sûreté sismique. La dernière question **Q3** est une problématique importante pour la quantification des incertitudes dite "orientée-but". En effet, elle concerne la planification d'expérience du modèle de simulation numérique pour estimer avec précision la quantité d'intérêt (ici la courbe de fragilité sismique). Plusieurs objectifs scientifiques de cette thèse ainsi qu'un bref résumé de ce manuscrit sont énoncés dans la section suivante.

Objectifs et plan

Les objectifs de cette thèse sont énumérés ci-dessous :

O1 Proposer une revue de l'état de l'art des différentes méthodes statistiques existantes pour l'estimation des courbes de fragilité sismique dans le cadre des études probabilistes de sûreté sismique;

O2 Développer une méthodologie pour prendre en compte les incertitudes épistémiques dans les courbes de fragilité sismique avec des outils statistiques appropriés;

O3 Proposer des outils d'analyse de sensibilité adaptées aux courbes de fragilité sismique, en tenant compte des incertitudes d'estimation de ces courbes;

O4 Proposer des plans d'expériences pour améliorer l'estimation des courbes de fragilité sismique tout en limitant le nombres d'appels au code de simulation numérique;

O5 Appliquer les outils proposés à un modèle numérique réaliste d'un équipement de centrale nucléaire.

En laissant de côté les chapitres d'introduction et de conclusion, le présent manuscrit est composé de sept chapitres visant à réaliser les cinq objectifs scientifiques énoncés ci-dessus. Le contenu de chaque chapitre est détaillé ci-dessous.

Chapitre 2 propose une brève revue des principales méthodes d'estimation statistique de la courbe de fragilité sismique. La méthodologie d'estimation des courbes de fragilité sismique à l'aide d'expériences numériques est également rappelée, depuis la génération de séismes artificiels jusqu'aux méthodes statistiques d'estimation des courbes de fragilité sismique.

Chapitre 3 présente une nouvelle méthodologie d'estimation des courbes de fragilité sismique utilisant l'échantillonnage préférentiel afin d'améliorer leur précision d'estimation.

Chapter 4 présente un aperçu du cadre de quantification des incertitudes. Tout d'abord, les notions fondamentales sont décrites, telles que le point de vue *boîte noire*, les différentes sources d'incertitudes et leur modélisation probabiliste. Ensuite, une adaptation du cadre de quantification des incertitudes au génie parasismique est proposée, tel qu'il soit adapté aux courbes de fragilité sismique.

Chapitre 5 aborde le problème central de la construction de métamodèles de codes de calcul de type "boîte noire". En raison du coût de calcul de l'estimation des courbes de fragilité sismique, le modèle numérique est remplacé par un modèle de substitution (dit métamodèle) utilisant des méthodes statistiques. Le principal avantage est une estimation à moindre coût de la courbe de fragilité sismique, mais elle est désormais entachée d'une incertitude de modèle. La métamodèlisation par processus Gaussien est alors développée dans ce chapitre, en raison de sa capacité à proposer à la fois une prédiction et une quantification de l'incertitude sur ses prédictions.

Chapitre 6 présente une méthodologie d'analyse de sensibilité sur la courbe de fragilité sismique. L'accent est mis sur le cadre plus spécifique de l'analyse de sensibilité globale (GSA en anglais) qui vise à considérer le cadre probabiliste global sur les paramètres d'entrée du code de calcul. Après une brève revue de l'état de l'art de l'analyse de sensibilité globale, deux indices de

sensibilité globale sont proposés ainsi que leurs estimateurs. Une procédure numérique est détaillée pour prendre en compte l'incertitude provenant du métamodèle par processus Gaussien.

Chapitre 7 propose une procédure de planification séquentielle d'expériences basée sur un processus Gaussien pour améliorer la précision de l'estimation de la courbe de fragilité sismique. Après un bref rappel des techniques de réduction d'incertitude par étapes (SUR en anglais), une stratégie SUR est ensuite développée pour l'estimation des courbes de fragilité sismique.

Chapitre 8 présente un cas-test représentatif de l'ingénierie nucléaire issu d'un programme expérimental du *Commissariat à l'énergie atomique et aux énergies alternatives*. Il consiste en une maquette d'un système de tuyauterie d'un réacteur français à eau pressurisée (REP). Dans ce chapitre, les outils développés dans les trois chapitres précédents sont appliqués pour discuter de l'utilisation de la méthodologie proposée sur un cas-test réel.

Les chapitres peuvent être organisés en deux groupes: d'une part, les chapitres 2 et 3 concernent les méthodes classiques d'estimation des courbes de fragilité sismique et une amélioration possible par la technique statistique de l'échantillonnage préférentiel, la question des incertitudes épistémiques n'y est pas abordée. En revanche, les chapitres 4, 5, 6, 7 et 8 concernent la prise en compte des incertitudes épistémiques sur la courbe de fragilité sismique et le développement d'une méthodologie de quantification des incertitudes sur les courbes de fragilité sismique, ils sont destinés à être lus dans l'ordre chronologique pour être cohérent avec les différents outils mathématiques utilisés dans ces chapitres.

Par ailleurs, j'ai eu l'occasion de superviser le stage de fin d'études de six mois d'Antoine Van Biesbroeck. Le sujet du stage concernait l'élicitation de distributions *a priori* pour l'estimation bayésienne des courbes de fragilité sismique. Puisque son contenu du stage diffère du sujet principal de cette thèse, son rapport de stage a été placé en Annexe C.

Publications et communications

Les contributions présentées dans ce manuscrit ont fait l'objet de publications et communications qui sont énumérées ci-dessous.

Jour. Pap. **Gauchy, C.**, C. Feau, and J. Garnier (2021a). *Importance sampling based active learning for parametric seismic fragility curve estimation*. DOI: 10.48550/ARXIV.2109.04323

Gauchy, C., C. Feau, and J. Garnier (2022c). *Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging*. DOI: 10.48550/ARXIV.2210.06266

Int. Conf. **Gauchy, C.**, C. Feau, and J. Garnier (2022a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *SIAM conference on Uncertainty Quantification 2022*. Atlanta, USA. (Présentation)

Nat. Conf. **Gauchy, C.**, C. Feau, and J. Garnier (2020a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *Mascot-num PhD days (MASCOT-NUM 2020)*. Grenoble, France. (Poster)

Gauchy, C., C. Feau, and J. Garnier (2020b). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *5ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2020)*. Saint Pierre d'Oléron, France. (Poster)

Gauchy, C., C. Feau, and J. Garnier (2021d). "Propagation of epistemic uncertainties in seismic risk assessment". In: *15th Mascot-Num annual conference (MASCOT-NUM 2021)*. Online (Poster)

Gauchy, C., C. Feau, and J. Garnier (2021b). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *51ème journées de Statistique de la Société Française de Statistique (JdS 2021)*. Online (Présentation)

Gauchy, C., C. Feau, and J. Garnier (2021c). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *6ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2021)*. Erdeven, France (Présentation)

Gauchy, C., C. Feau, and J. Garnier (2022d). "Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging". In: *16th Mascot-Num annual conference (MASCOT-NUM 2022)*. Clermont-Ferrand, France (Poster)

Gauchy, C., C. Feau, and J. Garnier (2022b). "Estimation of seismic fragility curves by sequential design of experiments". In: *52ème journées de Statistique de la Société Française de Statistique (JdS 2022)*. Lyon, France (Présentation)

Bibliography

Abramowitz, M. and I.A. Stegun (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Applied mathematics series. Dover Publications. ISBN: 9780486612720.

Altieri, D. and E. Patelli (2020). "An efficient approach for computing analytical non-parametric fragility curves". In: *Structural Safety* 85, p. 101956. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2020.101956.

Ambraseys, N. N et al. (2004). "Dissemination of European strong-motion data, Volume 2". In: *Proceedings of the 13th Thirteenth World Conference on Earthquake Engineering, Vancouver, Canada. Paper*. 32. Citeseer.

Antoniadis, A. (1984). "Analysis of variance on function spaces". In: *Series Statistics* 15.1, pp. 59–71. DOI: 10.1080/02331888408801747.

Aronszajn, N. (1950). "Theory of Reproducing Kernels". In: *Transactions of the American Mathematical Society* 68.3, pp. 337–404.

Aubin-Frankowski, P.-C. (2021). "Estimation and control under constraints through Kernel methods". PhD thesis. Université Paris sciences et lettres.

Azzimonti, D. et al. (2016). "Quantifying Uncertainties on Excursion Sets Under a Gaussian Random Field Prior". In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1. Publisher: Society for Industrial and Applied Mathematics, pp. 850–874. DOI: 10.1137/141000749.

Bachoc, F. (2013). "Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification". In: *Comput. Stat. Data Anal.* 66, pp. 55–69.

Baker, J. W. (2015). "Efficient Analytical Fragility Function Fitting Using Dynamic Structural Analysis". In: *Earthquake Spectra* 31.1, pp. 579–599. DOI: 10.1193/021113EQS025M.

Barr, J. and H. Rabitz (2022). "A Generalized Kernel Method for Global Sensitivity Analysis". In: *SIAM/ASA Journal on Uncertainty Quantification* 10.1, pp. 27–54. DOI: 10.1137/20M1354829.

Bect, J., F. Bachoc, and D. Ginsbourger (2019). "A supermartingale approach to Gaussian process based sequential design of experiments". In: *Bernoulli* 25.4A. Publisher: Bernoulli Society for Mathematical Statistics and Probability, pp. 2883–2919.

Bect, J., D. Ginsbourger, et al. (2012). "Sequential design of computer experiments for the estimation of a probability of failure". en. In: *Statistics and Computing* 22.3, pp. 773–793. ISSN: 0960-3174, 1573-1375. DOI: 10.1007/s11222-011-9241-4.

Bect, J., R. Sueur, et al. (2015). "Echantillonnage préférentiel et méta-modèles : méthodes bayésiennes optimale et defensive". In: *47èmes Journées de Statistique de la SFdS (JdS 2015)*.

Berger, J. O., J. M. Bernardo, and D. Sun (2009). "The Formal Definition of Reference Priors". In: *The Annals of statistics* 37.2, pp. 905–938. DOI: 10.1214/07-AOS587.

Berlinet, A. and C. Thomas-Agnan (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US. ISBN: 978-1-4419-9096-9.

Bernardo, J. (2005). "Reference Analysis". In: *Handbook of Statistics* 25, pp. 17–90. DOI: 10.1016/S0169-7161(05)25002-2.

Bernardo, J. M. (1979). "Reference Posterior Distributions for Bayesian Inference". In: *Journal of the Royal Statistical Society. Series B* 41.2, pp. 113–147. DOI: 10.1111/j.2517-6161.1979.tb01066.x.

Bernardo, J.. and A. Smith (1994). *Bayesian Theory*. 1st ed. Wiley Series in Probability and Statistics. Wiley. ISBN: 047149464X.

Bernier, C. and J. E. Padgett (2019). "Fragility and risk assessment of aboveground storage tanks subjected to concurrent surge, wave, and wind loads". In: *Reliability Engineering & System Safety* 191, p. 106571. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2019.106571.

Binois, M., R. Gramacy, and M. Ludkovski (2018). "Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments". In: *Journal of Computational and Graphical Statistics* 27.4, pp. 808–821. DOI: 10.1080/10618600.2018.1458625.

Borgonovo, E., G. B. Hazen, and E. Plischke (2016). "A Common Rationale for Global Sensitivity Measures and Their Estimation". In: *Risk Analysis* 36.10, pp. 1871–1895. ISSN: 1539-6924. DOI: 10.1111/risa.12555.

Box, G. E. and D. R. Cox (1964). "An analysis of transformations". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 26.2, pp. 211–243.

Box, G.E.P. and G.t. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Broto, B. (July 2020). "Sensitivity analysis with dependent random variables : Estimation of the Shapley effects for unknown input distribution and linear Gaussian models". PhD thesis. université Paris-Saclay.

Browne, T. (2017). "Regression models and sensitivity analysis for stochastic simulators : applications to non-destructive examination". PhD thesis. Sorbonne Paris Cité.

Cartan, H. (2007). *Cours de calcul différentiel*. 2nd ed. Sciences et techniques. Hermann. ISBN: 9782705667023.

Chabridon, V. (2018). "Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems". PhD thesis. Université Clermont Auvergne.

Chabridon, V. et al. (2017). "Evaluation of failure probability under parameter epistemic uncertainty: application to aerospace system reliability assessment". In: *Aerospace Science and Technology* 69, pp. 526–537.

Chaloner, K. and I. Verdinelli (1995). "Bayesian Experimental Design: A Review". In: *Statistical Science* 10.3, pp. 273–304. DOI: 10.1214/ss/1177009939.

Charras, T. and J. Kichenin (2011). *Développer dans CAST3M*. URL: http://www-cast3m.cea.fr/.

Chastaing, G. (2013). "Indices de Sobol généralisés par variables dépendantes". PhD thesis. Grenoble.

Chilès, J.P. and P. Delfiner (2012). *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. Wiley. ISBN: 978-1-118-13617-1.

Choe, D., P. Gardoni, and D. Rosowsky (2007). "Closed-Form Fragility Estimates, Parameter Sensitivity, and Bayesian Updating for RC Columns". In: *Journal of Engineering Mechanics* 133.7, pp. 833–843. DOI: 10.1061/(ASCE)0733-9399(2007)133:7(833).

Choe, D., P. Gardoni, D. Rosowsky, and T. Haukaas (2009). "Seismic fragility estimates for reinforced concrete bridges subject to corrosion". In: *Structural Safety* 31.4, pp. 275–283. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2008.10.001.

Chu, W. et al. (2011). "Unbiased Online Active Learning in Data Streams". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '11. San Diego, California, USA: Association for Computing Machinery, pp. 195–203. ISBN: 9781450308137.

Ciano, M., M. Gioffrè, and M. Grigoriu (2020). "The role of intensity measures on the accuracy of seismic fragilities". In: *Probabilistic Engineering Mechanics* 60, p. 103041.

Clarke, B. and A. Barron (1994). "Jeffreys' prior is asymptotically least favorable under entropy risk". In: *Journal of Statistical Planning and Inference* 41.1, pp. 37–60. DOI: 10.1016/0378-3758(94)90153-8.

Cornell, A. (2004). "Hazard, ground motions and probabilistic assessments for PBSD". In: *Proceedings of the International Workshop on Performance-Based Seismic Design - Concepts and Implementation*. BLED, SLOVENIA: PEER Center, pp. 39–52.

Da Veiga, S. (2015). "Global sensitivity analysis with dependence measures". In: *Journal of Statistical Computation and Simulation* 85.7, pp. 1283–1305. ISSN: 0094-9655. DOI: 10.1080/00949655.2014.945932.

— (2021). "Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis". In: DOI: 10.48550/ARXIV.2101.05487.

Da Veiga, S. et al. (2021). *Basics and Trends in Sensitivity Analysis*. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611976694.

Damblin, G. et al. (2014). "Approche décisionnelle bayésienne pour estimer une courbe de fragilité". In: *Journal de la Société Française de Statistique* 155.3, pp. 78–103. HAL: hal-01545648.

De Rocquigny, E. et al. (2008). *Uncertainty in industrial practice: A guide to quantitative uncertainty management*. Wiley.

Delyon, B. and F. Portier (2018). "Asymptotic Optimality of Adaptive Importance Sampling". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Montréal, Canada: Curran Associates Inc., pp. 3138–3148.

Der Kiureghian, A. and O. Ditlevsen (2009). "Aleatory or epistemic? Does it matter?" In: *Structural Safety*. Risk Acceptance and Risk Communication 31.2, pp. 105–112. ISSN: 0167-4730. DOI: 10.1016/j.strusafe.2008.06.020.

Echard, B., N. Gayton, and M. Lemaire (2011). "AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation". In: *Structural Safety* 33.2, pp. 145–154. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2011.01.002.

Echard, B., N. Gayton, M. Lemaire, and N. Relun (2013). "A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models". In: *Reliability Engineering & System Safety* 111, pp. 232–240. ISSN: 0951-8320.

Ellingwood, B. R. (2001). "Earthquake risk assessment of building structures". In: *Reliability Engineering & System Safety* 74.3, pp. 251–262. ISSN: 0951-8320.

Ellingwood, B. R. and K. Kinali (2009). "Quantifying and communicating uncertainty in seismic risk assessment". In: *Structural Safety*. Risk Acceptance and Risk Communication 31.2, pp. 179–187. ISSN: 0167-4730. DOI: 10.1016/j.strusafe.2008.06.001.

Feau, C. (2019). "Estimation de la fragilité des structures et équipements sous séisme". In: *Mécanique déterministe ou incertitudes : Où en est-on avec $F = M\gamma$? - Ça passe ou ça casse ?* Aristote. Palaiseau, France.

Ferraty, F. and P. Vieu (2006). *Nonparametric functional data analysis: theory and practice*. Vol. 76. Springer.

Finney, D.J. (1971). "Probit Analysis". In: *Journal of Pharmaceutical Sciences* 60.9, pp. 1432–1432.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd.

Fukumizu, K., F. R. Bach, and M. I. Jordan (2004). "Dimensionality Reduction for Supervised Learning with Reproducing Kernel Hilbert Spaces". In: *Journal of Machine Learning Research* 5.Jan, pp. 73–99. ISSN: ISSN 1533-7928.

— (2009). "Kernel dimension reduction in regression". In: *The Annals of Statistics* 37.4, pp. 1871–1905. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/08-AOS637.

Fukumizu, K., A. Gretton, et al. (2007). "Kernel Measures of Conditional Dependence". In: *Advances in Neural Information Processing Systems*. Vol. 20.

Gamboa, F. et al. (2022). "Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics". In: *Bernoulli* 28.4, pp. 2345–2374. DOI: 10.3150/21-BEJ1421.

Gardoni, P., A. Der Kiureghian, and K. M. Mosalam (2002). "Probabilistic Capacity Models and Fragility Estimates for Reinforced Concrete Columns based on Experimental Observations". In: *Journal of Engineering Mechanics* 128.10, pp. 1024–1038. DOI: 10.1061/(ASCE)0733-9399(2002)128:10(1024).

Gardoni, P., K. M. Mosalam, and Der Kiureghian A. (2003). "Probabilistic seismic demand models and fragility estimates for rc bridges". In: *Journal of Earthquake Engineering* 7.sup001, pp. 79–106. DOI: 10.1080/13632460309350474.

Gauchy, C. et al. (2021). "An Information Geometry Approach to Robustness Analysis for the Uncertainty Quantification of Computer Codes". In: *Technometrics* 64.1, pp. 80–91. DOI: 10.1080/00401706.2021.1905072.

Gauchy, C., C. Feau, and J. Garnier (2020a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *Mascot-num PhD days (MASCOT-NUM 2020)*. Grenoble, France.

— (2020b). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *5ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2020)*. Saint Pierre d'Oléron, France.

— (2021a). *Importance sampling based active learning for parametric seismic fragility curve estimation*. DOI: 10.48550/ARXIV.2109.04323.

— (2021b). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *51ème journées de Statistique de la Société Française de Statistique (JdS 2021)*. Online.

— (2021c). "Propagation of epistemic uncertainties and global sensitivity analysis in seismic risk assessment". In: *6ème École Thématique sur les Incertitudes en Calcul Scientifique (ETICS 2021)*. Erdeven, France.

— (2021d). "Propagation of epistemic uncertainties in seismic risk assessment". In: *15th Mascot-Num annual conference (MASCOT-NUM 2021)*. Online.

— (2022a). "Adaptive Importance Sampling for Seismic Fragility Curves Estimation". In: *SIAM conference on Uncertainty Quantification 2022*. Atlanta, USA.

— (2022b). "Estimation of seismic fragility curves by sequential design of experiments". In: *52ème journées de Statistique de la Société Française de Statistique (JdS 2022)*. Lyon, France.

— (2022c). *Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging*. DOI: 10.48550/ARXIV.2210.06266.

— (2022d). "Uncertainty quantification and global sensitivity analysis of seismic fragility curves using kriging". In: *16th Mascot-Num annual conference (MASCOT-NUM 2022)*. Clermont-Ferrand, France.

Gehl, P., J. Douglas, and D. Seyedi (2015). "Influence of the Number of Dynamic Analyses on the Accuracy of Structural Response Estimates". In: *Earthquake Spectra* 31.1, pp. 97–113. DOI: https://doi.org/10.1193/102912EQS320M.

Ghobarah, A. (2001). "Performance-based design in earthquake engineering: state of development". In: *Engineering Structures* 23.8, pp. 878–884.

Ghosh, M. (2011). "Objective Priors: An Introduction for Frequentists". In: *Statistical Science* 26.2, pp. 187–202. DOI: 10.1214/10-STS338.

Gidaris, I., A. A. Taflanidis, and G. P. Mavroeidis (2015). "Kriging metamodeling in seismic risk assessment based on stochastic ground motion models". In: *Earthquake Engineering & Structural Dynamics* 44.14, pp. 2377–2399.

Golestaneh, F. et al. (2018). "Ellipsoidal Prediction Regions for Multivariate Uncertainty Characterization". In: *IEEE Transactions on Power Systems* 33.4, pp. 4519–4530. DOI: 10.1109/TPWRS.2018.2791975.

Gong, C. and W. Zhou (2018). "Importance sampling-based system reliability analysis of corroding pipelines considering multiple failure modes". In: *Reliability Engineering & System Safety* 169, pp. 199–208.

Gramacy, R.B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Taylor & Francis Limited. ISBN: 978-1-03-224255-2.

Gretton, A., K. M. Borgwardt, et al. (2012). "A Kernel Two-Sample Test". In: *Journal of Machine Learning Research* 13.25, pp. 723–773.

Gretton, A., K. Fukumizu, et al. (2007). "A Kernel Statistical Test of Independence". In: *Advances in Neural Information Processing Systems*. Vol. 20.

Gretton, A., D. Sejdinovic, et al. (2012). "Optimal kernel choice for large-scale two-sample tests". In: *Advances in neural information processing systems* 25.

Grigoriu, M. (2011). "To Scale or Not to Scale Seismic Ground-Acceleration Records". In: *Journal of Engineering Mechanics* 137.4, pp. 284–293. DOI: 10.1061/(ASCE)EM.1943-7889.0000226.

Grigoriu, M. and A. Radu (2021). "Are seismic fragility curves fragile?" In: *Probabilistic Engineering Mechanics* 63, p. 103115. ISSN: 02668920. DOI: 10.1016/j.probengmech.2020.103115.

Gu, M. (2019). "Jointly Robust Prior for Gaussian Stochastic Process in Emulation, Calibration and Variable Selection". In: *Bayesian Analysis* 14.3, pp. 857–885. DOI: 10.1214/18-BA1133.

Gu, M., J. Palomo, and J. Berger (2018). "RobustGaSP: Robust Gaussian Stochastic Process Emulation in R". In: *The R Journal* 11. DOI: 10.32614/RJ-2019-011.

Haario, H., E. Saksman, and J. Tamminen (2001). "An Adaptive Metropolis Algorithm". In: *Bernoulli* 7.2, pp. 223–242. DOI: 10.2307/3318737.

Haberstich, C. (Dec. 2020). "Adaptive approximation of high-dimensional functions with tree tensor networks for Uncertainty Quantification". PhD thesis. Ecole centrale de Nantes.

Hall, P. et al. (2014). *Martingale Limit Theory and Its Application*. Communication and Behavior. Elsevier Science.

Hariri-Ardebili, M. and V. Saouma (2016a). "Probabilistic seismic demand model and optimal intensity measure for concrete dams". In: *Structural Safety* 59, pp. 67–85. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2015.12.001.

— (2016b). "Probabilistic seismic demand model and optimal intensity measure for concrete dams". In: *Structural Safety* 59, pp. 67–85. DOI: 10.1016/j.strusafe.2015.12.001.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

Hesterberg, T. (1995). "Weighted Average Importance Sampling and Defensive Mixture Distributions". In: *Technometrics* 37.2, pp. 185–194.

Hoeffding, W. (1948). "A Class of Statistics with Asymptotically Normal Distribution". In: *The Annals of Mathematical Statistics* 19.3. Publisher: Institute of Mathematical Statistics, pp. 293–325. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177730196.

Hoffman, M. D. and A. Gelman (2011). *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. DOI: 10.48550/ARXIV.1111.4246.

Iooss, B. (Jan. 2009). "Contributions au traitement des incertitudes en modélisation numérique : propagation d'ondes en milieu aléatoire et analyse statistique d'expériences simulées". HDR (French accreditation to supervise research). Université Paul Sabatier - Toulouse III.

Iooss, B. and L. Le Gratiet (2019). "Uncertainty and sensitivity analysis of functional risk curves based on Gaussian processes". In: *Reliability Engineering & System Safety* 187, pp. 58–66. ISSN: 09518320. DOI: 10.1016/j.ress.2017.11.022.

Jacquemain, D. (2015). *Nuclear Power Reactor Core Melt Accidents. Current State of Knowledge*. EDP Sciences. ISBN: 9782759818358.

Jalayer, F., R. De Risi, and G. Manfredi (2015). "Bayesian Cloud Analysis: efficient structural fragility assessment using linear regression". In: *Bulletin of Earthquake Engineering* 13, pp. 1183–1203. DOI: 10.1007/s10518-014-9692-z.

Janon, A. et al. (2014). *Asymptotic normality and efficiency of two Sobol index estimators*. DOI: 10.1051/ps/2013040.

Jaynes, E. (1982). "On the rationale of maximum-entropy methods". In: *Proceedings of the IEEE* 70.9, pp. 939–952. DOI: 10.1109/PROC.1982.12425.

Jaynes, E. T. (1957). "Information theory and statistical mechanics". In: *Physical review* 106.4, p. 620.

Jeffreys, H. (1961). *Theory of Probability*. London: Oxford University Press.

Kanagawa, M. et al. (2018). *Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences*. DOI: 10.48550/ARXIV.1807.02582.

Karamlou, A. and P. Bochini (2015). "Computation of bridge seismic fragility by large-scale simulation for probabilistic resilience analysis". In: *Earthquake Engineering & Structural Dynamics* 44.12, pp. 1959–1978. DOI: 10.1002/eqe.2567.

Karim, K. R. and F. Yamazaki (2001). "Effect of earthquake ground motions on fragility curves of highway bridge piers based on numerical simulation". In: *Earthquake Engineering & Structural Dynamics* 30.12, pp. 1839–1856. ISSN: 1096-9845. DOI: 10.1002/eqe.97.

Keller, M. et al. (2015). "Nonparametric Estimation of the Probability of Detection of Flaws in an Industrial Component, from Destructive and Nondestructive Testing Data, Using Approximate Bayesian Computation". In: *Risk Analysis* 35 (09), pp. 1595–1610.

Kennedy, M. C. and A. O'Hagan (2001). "Bayesian calibration of computer models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 425–464.

Kennedy, R. (1999). "Risk based seismic design criteria". In: *Nuclear Engineering and Design* 192.2, pp. 117–135. ISSN: 0029-5493. DOI: https://doi.org/10.1016/S0029-5493(99)00102-8.

Kennedy, R., C.A. Cornell, et al. (1980). "Probabilistic seismic safety study of an existing nuclear power plant". In: *Nuclear Engineering and Design* 59.2, pp. 315–338. ISSN: 0029-5493.

Kennedy, R. and M.K. Ravindra (1984). "Seismic fragilities for nuclear power plant risk studies". In: *Nuclear Engineering and Design* 79.1, pp. 47–68.

Kiani, J., C. Camp, and S. Pezeshk (2019). "On the application of machine learning techniques to derive seismic fragility curves". In: *Computers & Structures* 218, pp. 108–122. ISSN: 0045-7949. DOI: https://doi.org/10.1016/j.compstruc.2019.03.004.

Kim, S.-H. and M. Shinozuka (2004). "Development of fragility curves of bridges retrofitted by column jacketing". In: *Probabilistic Engineering Mechanics*. Fourth International Conference on Computational Stochastic Mechanics 19.1, pp. 105–112. ISSN: 0266-8920. DOI: 10.1016/j.probengmech.2003.11.009.

Kloek, T. and H. Van Dijk (1978). "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo". In: *Econometrica* 46, pp. 1–19.

Krige, D. G. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". In: *Journal of the Southern African Institute of Mining and Metallurgy* 52.6. Publisher: Southern African Institute of Mining and Metallurgy, pp. 119–139. DOI: 10.10520/AJA0038223X_4792.

Kumar, R. and P. Gardoni (2013). "Second-order Logarithmic formulation for hazard curves and closed-form approximation to annual failure probability". In: *Structural Safety* 45, pp. 18–23. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2013.07.007.

Kwong, N. and A. Chopra (2015). "Evaluation of the Exact Conditional Spectrum and Generalized Conditional Intensity Measure methods for ground motion selection". In: *Earthquake Engineering & Structural Dynamics* 45, pp. 757–777.

Kyprioti, A. P. and A. A. Taflanidis (2021). "Kriging metamodeling for seismic response distribution estimation". In: *Earthquake Engineering & Structural Dynamics* 50.13, pp. 3550–3576. DOI: https://doi.org/10.1002/eqe.3522.

Labopin-Richard, T. and V. Picheny (2016). "Sequential design of experiments for estimating percentiles of black-box functions". In: *Statistica Sinica*. DOI: 10.5705/ss.202016.0160.

Lagaros, N. and M. Fragiadakis (2007). "Fragility Assessment of Steel Frames Using Neural Networks". In: *Earthquake Spectra* 23.4, pp. 735–752. DOI: 10.1193/1.2798241.

Lallemant, D: A. Kiremidjian, and H. Burton (2015). "Statistical procedures for developing earthquake damage fragility curves". In: *Earthquake Engineering & Structural Dynamics* 44.9, pp. 1373–1389. ISSN: 1096-9845. DOI: 10.1002/eqe.2522.

Le Gratiet, L. (Jan. 2013). "Multi-fidelity Gaussian process regression for computer experiments". PhD thesis. Paris 7.

Le Gratiet, L. et al. (2017). "Model Assisted Probability of Detection Curves: New Statistical Tools and Progressive Methodology". In: *Journal of Nondestructive Evaluation* 36.1. ISSN: 0195-9298. DOI: 10.1007/s10921-016-0387-z.

Lemaitre, P. (2014). "Analyse de sensibilité en fiabilité des structures". PhD thesis. Université de Bordeaux.

Lindgren, F., H. Rue, and J. Lindström (2011). "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4, pp. 423–498.

Luco, N. and P. Bazzurro (2007). "Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses?" In: *Earthquake Engineering & Structural Dynamics* 36.13, pp. 1813–1835. DOI: https://doi.org/10.1002/eqe.695.

Luco, N. and A. Cornell (2007). "Structure-Specific Scalar Intensity Measures for Near-Source and Ordinary Earthquake Ground Motions". In: *Earthquake Spectra* 23.2, pp. 357–392. DOI: 10.1193/1.2723158.

Mai, C., K. Konakli, and B. Sudret (2017). "Seismic fragility curves for structures using non-parametric representations". In: *Frontiers of Structural and Civil Engineering* 11.2, pp. 169–186. ISSN: 2095-2430, 2095-2449. DOI: 10.1007/s11709-017-0385-y.

Mai, C., M.D. Spiridonakos, et al. (2016). "Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions". In: *Int. J. Uncertainty Quant.* 6, pp. 313–339.

Mandal, T. K., S. Ghosh, and N. N. Pujari (2016). "Seismic fragility analysis of a typical Indian PHWR containment: Comparison of fragility models". In: *Structural Safety* 58, pp. 11–19. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2015.08.003.

Marrel, A., B. Iooss, and V. Chabridon (2021). "The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel—Applications in Thermal-Hydraulics". In: *Nuclear Science and Engineering* 0.0, pp. 1–21. DOI: 10.1080/00295639.2021.1980362.

Marrel, A., B. Iooss, S. Da Veiga, et al. (2012). "Global sensitivity analysis of stochastic computer models with joint metamodels". In: *Statistics and Computing* 22.3, pp. 833–847. ISSN: 1573-1375. DOI: 10.1007/s11222-011-9274-8.

Marrel, A., B. Iooss, F. Van Dorpe, et al. (2008). "An efficient methodology for modeling complex computer codes with Gaussian processes". In: *Computational Statistics & Data Analysis* 52.10, pp. 4731–4744. ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2008.03.026.

Matheron, G. (1962). *Traité de géostatistique appliquée*. Éditions Technip.

Mercer, J. (1909). "Functions of positive and negative type, and their connection the theory of integral equations". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209.441-458.

Meynaoui, A. (2019). "New developments around dependence measures for sensitivity analysis : application to severe accident studies for generation IV reactors". PhD thesis. Toulouse, INSA.

Mitropoulou, C. and M. Papadrakakis (2011). "Developing fragility curves based on neural network IDA predictions". In: *Engineering Structures* 33.12, pp. 3409–3421. ISSN: 0141-0296. DOI: https://doi.org/10.1016/j.engstruct.2011.07.005.

Morio, J. and M. Balesdent (2015). *Estimation of rare event probabilities in complex aerospace and other systems: a practical approach*. Woodhead Publishing.

Muré, J. (2018). "Objective Bayesian analysis of Kriging models with anisotropic correlation kernel". PhD thesis. Université Sorbonne Paris Cité.

Nalisnick, E. and P. Smyth (2017). "Learning Approximately Objective Priors". In: arXiv: 1704.01168 [stat.ML].

Oh, M. and J. Berger (1992). "Adaptive importance sampling in Monte Carlo integration". In: *Journal of Statistical Computation and Simulation* 41.3-4, pp. 143–168.

Owen, A. and Y. Zhou (2000). "Safe and Effective Importance Sampling". In: *Journal of the American Statistical Association* 95.449, pp. 135–143.

Padgett, J. E., B.G. Nielson, and R. DesRoches (2008). "Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios". In: *Earthquake Engineering & Structural Dynamics* 37.5, pp. 711–725. DOI: 10.1002/eqe.782.

Papaioannou, I., S. Geyer, and D. Straub (2019). "Improved cross entropy-based importance sampling with a flexible mixture model". In: *Reliability Engineering & System Safety* 191, p. 106564.

Park, Y.J., C.H. Hofmayer, and N.C. Chokshi (1998). "Survey of seismic fragilities used in PRA studies of nuclear power plants". In: *Reliability Engineering & System Safety* 62.3, pp. 185–195. ISSN: 0951-8320. DOI: https://doi.org/10.1016/S0951-8320(98)00019-2.

Paulo, R. (2005). "Default priors for Gaussian processes". In: *The Annals of Statistics* 33.2, pp. 556–582. DOI: 10.1214/009053604000001264.

PEER (2004). "Proceedings of the International Workshop on Performance-Based Seismic Design - Concepts and Implementation". In: BLED, SLOVENIA: PEER Center.

Porter, K., R. Kennedy, and R. Bachman (2007). "Creating Fragility Functions for Performance-Based Earthquake Engineering". In: *Earthquake Spectra* 23.2, pp. 471–489. DOI: 10.1193/1.2720892.

Pronzato, L. and W. G. Müller (2012). "Design of computer experiments: space filling and beyond". In: *Statistics and Computing* 22.3, pp. 681–701. ISSN: 1573-1375. DOI: 10.1007/s11222-011-9242-3.

Pronzato, L. and A. Pazman (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Lecture Notes in Statistics. Springer, p. 399. DOI: 10.1007/978-1-4614-6363-4.

Quilligan, A., A. O'Connor, and V. Pakrashi (2012). "Fragility analysis of steel and concrete wind turbine towers". In: *Engineering Structures* 36, pp. 270–282.

Rasmussen, C. and C. Williams (2005). *Gaussian Processes for Machine Learning*. MIT Press. ISBN: 978-0-262-18253-9.

Rezaeian, S. and A. Der Kiureghian (2008). "A stochastic ground motion model with separable temporal and spectral nonstationarities". In: *Earthquake Engineering & Structural Dynamics* 37.13, pp. 1565–1584. DOI: https://doi.org/10.1002/eqe.831.

— (2010). "Simulation of synthetic ground motions for specified earthquake and site characteristics". In: *Earthquake Engineering & Structural Dynamics* 39.10, pp. 1155–1180.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer New York.

Rodolfo Sargoni, G. and G.C. Hart (1973). "Simulation of artificial earthquakes". In: *Earthquake Engineering & Structural Dynamics* 2.3, pp. 249–267. ISSN: 1096-9845. DOI: 10.1002/eqe.4290020305.

Sacks, J. et al. (1989). "Design and Analysis of Computer Experiments". In: *Statistical Science* 4.4. Publisher: Institute of Mathematical Statistics, pp. 409–423. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177012413.

Sainct, R. et al. (2020). "Efficient methodology for seismic fragility curves estimation by active learning on Support Vector Machines". In: *Structural Safety* 86, p. 101972. ISSN: 0167-4730. DOI: 10.1016/j.strusafe.2020.101972.

Saitoh, S. and Y. Sawano (2016). *Theory of Reproducing Kernels and Applications*. Developments in Mathematics. Springer Nature Singapore. ISBN: 9789811005305.

Saltelli, A. et al. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley. ISBN: 978-0-470-87094-5.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer series in statistics. Springer. ISBN: 978-0-387-95420-2.

Schölkopf, B., A. Smola, and K.-R. Müller (1998). "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation* 10.5, pp. 1299–1319. ISSN: 0899-7667. DOI: 10.1162/089976698300017467.

Schölkopf, B., B.S.A.J. Smola, et al. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive computation and machine learning. MIT Press. ISBN: 978-0-262-19475-4.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Shinozuka, M. et al. (2000). "Statistical Analysis of Fragility Curves". In: *Journal of Engineering Mechanics* 126, pp. 1224–1231. DOI: 10.1061/(ASCE)0733-9399(2000)126:12(1224).

Sobol', I.M. (1993). "Sensitivity estimates for non linear mathematical models". In: *Mathematical Modeling and Computer Experiments* 1, pp. 407–414.

— (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation*. The Second IMACS Seminar on Monte Carlo Methods 55.1, pp. 271–280. ISSN: 0378-4754. DOI: 10.1016/S0378-4754(00)00270-6.

Soize, C. (2005). "A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics". In: *Journal of Sound and Vibration* 288.3. Uncertainty in structural dynamics, pp. 623–652. ISSN: 0022-460X. DOI: https://doi.org/10.1016/j.jsv.2005.07.009.

Solomos, G., A. Pinto, and S. Dimova (2008). *A review of the seismic hazard zonation in national building codes in the context of eurocode 8*. EUR 23563 EN-2008. Technical report. JRC European Commission.

Stan Development Team (2018). *The Stan Core Library*. Version 2.18.0. URL: http://mc-stan.org/.

Steen, N. M., G. D. Byrne, and E. M. Gelbard (1969). "Gaussian quadratures for the integrals $\int_0^\infty e^{-x^2} f(x)dx$ and $\int_0^b e^{-x^2} f(x)dx$". In: *Mathematics of Computation* 23.107, pp. 661–671. DOI: 10.1090/s0025-5718-1969-0247744-3.

Stein, Michael L. (1999). *Interpolation of Spatial Data*. Springer My Copy UK.

Steinwart, I. and A. Christmann (2008). *Support Vector Machines*. Springer Science & Business Media. ISBN: 978-0-387-77242-4.

Steinwart, I. and C. Scovel (2012). "Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs". In: *Constructive Approximation* 35, pp. 363–417.

Stenger, J. (2020). "Optimal uncertainty quantification of a risk measurement from a computer code". PhD thesis. Toulouse 3.

Straub, D. and A. Der Kiureghian (2008). "Improved seismic fragility modeling from empirical data". In: *Structural Safety* 30.4, pp. 320–336. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2007.05.004.

Stroh, R. (June 2018). "Planification d'expériences numériques en multi-fidélité : Application à un simulateur d'incendies". PhD thesis. Université Paris-Saclay (ComUE).

Stroock, D. (1993). *Probability Theory, an Analytical View*. 1st ed. Cambridge University Press. ISBN: 0521431239.

Sudret, B. (2007). "Uncertainty propagation and sensitivity analysis in mechanical models – Contributions to structural reliability and stochastic spectral methods". HDR (French accreditation to supervise research). Université Blaise Pascal - Clermont II.

Tonon, F. and Bernardini. A. (1998). "A random set approach to the optimization of uncertain structures". In: *Computers and Structures* 68.6, pp. 583–600. ISSN: 0045-7949. DOI: https://doi.org/10.1016/S0045-7949(98)00079-0.

Touboul, F., N. Blay, et al. (2006). "Enhanced seismic criteria for piping". In: *Nuclear Engineering and Design* 236.1, pp. 1–9.

Touboul, F., P. Sollogoub, and N. Blay (1999). "Seismic behaviour of piping systems with and without defects: experimental and numerical evaluations". In: *Nuclear Engineering and Design* 192.2, pp. 243–260.

Trevlopoulos, K., C. Feau, and I. Zentner (2019). "Parametric models averaging for optimized non-parametric fragility curve estimation based on intensity measure data clustering". In: *Structural Safety* 81, p. 101865. ISSN: 0167-4730. DOI: 10.1016/j.strusafe.2019.05.002.

Vaart, A. W. van der (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Vazquez, E. and J. Bect (2009). "A sequential Bayesian algorithm to estimate a probability of failure". In: *IFAC Proceedings Volumes* 42.10, pp. 546–550. ISSN: 14746670. DOI: 10.3182/20090706-3-FR-2004.00090.

Villemonteix, J., E. Vazquez, and E. Walter (2009). "An informational approach to the global optimization of expensive-to-evaluate functions". In: *Journal of Global Optimization* 44.4, pp. 509–534. ISSN: 0925-5001, 1573-2916. DOI: 10.1007/s10898-008-9354-2.

Wang, F. and C. Feau (2020). "Influence of Input Motion's Control Point Location in Nonlinear SSI Analysis of Equipment Seismic Fragilities: Case Study on the Kashiwazaki-Kariwa NPP". In: *Pure and Applied Geophysics*. ISSN: 1420-9136. DOI: https://doi.org/10.1016/j.engstruct.2018.02.024.

Wang, Z. et al. (2018). "Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment". In: *Engineering Structures* 162, pp. 213–225. ISSN: 0141-0296. DOI: https://doi.org/10.1016/j.engstruct.2018.02.024.

Wasserman, L. (2004). *All of statistics : a concise course in statistical inference*. 2nd ed. Texts in Statistics. Springer. ISBN: 978-1-4419-2322-6.

Zentner, I. (2010). "Numerical computation of fragility curves for NPP equipment". In: *Nuclear Engineering and Design* 240.6, pp. 1614–1621. ISSN: 0029-5493. DOI: 10.1016/j.nucengdes.2010.02.030.

— (2017). "A general framework for the estimation of analytical fragility functions based on multivariate probability distributions". In: *Structural Safety* 64, pp. 54–61. ISSN: 0167-4730. DOI: https://doi.org/10.1016/j.strusafe.2016.09.003.

Zentner, I., M. Gündel, and N. Bonfils (2017). "Fragility analysis methods: Review of existing approaches and application". In: *Nuclear Engineering and Design* 323, pp. 245–258. ISSN: 0029-5493. DOI: 10.1016/j.nucengdes.2016.12.021.

Zhao, C., N. Yu, and Y.L. Mo (2020). "Seismic fragility analysis of AP1000 SB considering fluid-structure interaction effects". In: *Structures* 23, pp. 103–110. ISSN: 2352-0124. DOI: https://doi.org/10.1016/j.istruc.2019.11.003.

Zhao, Y. et al. (2021). "Seismic fragility analysis of nuclear power plants considering structural parameter uncertainty". In: *Reliability Engineering & System Safety* 216, p. 107970. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2021.107970.

Zuniga, M. Munoz, A. Murangira, and T. Perdrizet (2021). "Structural reliability assessment through surrogate based importance sampling with dimension reduction". In: *Reliability Engineering & System Safety* 207, p. 107289.