Developing and applying the concept of Value of Information to optimise data collection strategies for seismic hazard assessment

By

Haifa Tebib

Thesis submitted in fulfilment of the
requirements for the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

Supervised by Dr. John Douglas and Dr. Jennifer Roberts

September 2023

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Haifa Tebib

Date: 26/09/2023

Research outputs

- Research paper published to peer-reviewed journal:
  - Extended version found in Chapter 4:

  ‘Using the value of information to decide when to collect additional data on near-surface site conditions.’

  Tebib Haifa, Douglas John and Roberts Jennifer

  Soil Dynamics and Earthquake Engineering, 165(9), https://doi.org/10.1016/j.soildyn.2022.107654

  Submitted June 2020, Revised September 2022, Accepted November 2022, Published February 2023.

  My contribution to this work: Author, Conceptualisation, Methodology, Writing - original draft, Writing – review and editing. John Douglas contributed to the Conceptualisation and Methodology. All authors provided feedback and reviewed the manuscript before submission and during revision.

- Conference paper:

  ‘Assessing the Value of Information in Site-Response Analyses.’

  Tebib Haifa, Douglas John and Roberts Jennifer

  Proceedings of the Third European Conference on Earthquake Engineering, 4-9 September 2022, Bucharest, Romania.

Abstract

In seismic hazard assessments the importance of knowing different input parameters accurately depends on their weight within the hazard model. Many aspects of such assessments require inputs based on knowledge and data from experts. When it comes to decisions about data collection, facility owners and seismic hazard analysts need to balance the possible added value brought by acquiring new data against the budget and time available for its collection. In other words, they need to answer the question “Is it worth paying to obtain this information?”. Assessing the value of information (VoI) before data collection should lead to optimising the time and money that one is willing to invest.

This thesis presents a method that combines available data and expert judgment to facilitate the decision-making process within the site-response component of seismic hazard assessments. The approach integrates influence diagrams and decision trees to map the causal-relationships between input parameters in site-response analysis, and Bayesian inference to update the model when new evidence is considered. Here, the VoI is assessed for univariate, bivariate and multivariate uncertain parameters to infer an optimal seismic design for typical buildings and critical facilities. For the first time in the field of seismic hazard assessment and earthquake engineering, a framework is developed to integrate prior knowledge, ground investigation techniques characteristics and design safety requirements.

The consistent findings across different applications show that VoI is highly sensitive to prior probabilities and to the accuracy of the test to be performed. This highlights the importance of defining those from available data as well as only considering tests that are suitable for our needs and budget. The developed VoI framework constitutes a useful decision-making tool for hazard analysts and facility owners, enabling not only the prioritisation of data collection for key input parameters and the identification of optimal tests, but also the justification of the associated decisions. This approach can enhance the accuracy and reliability of seismic hazard assessments, leading to more effective risk management strategies.

Acknowledgments

I am very grateful to my supervisor Dr. John Douglas who has given me the opportunity to enrol in this PhD and to make a contribution in my field of predilection. Dr. Douglas was extremely supportive, which was much needed during the Covid pandemic, and responsive throughout my PhD journey. I have benefited from his unparalleled expertise, his technical skills and analytical abilities. He encouraged me in all aspects of my research and helped me to enhance my skills as a researcher. I would also like to thank my second supervisor, Dr. Jennifer Roberts, for her insights on many aspects of my research, her valuable suggestions and her support.

Throughout this journey, I had the chance to collaborate and discuss with several important actors in the research and industry field. First, many thanks to Prof. John Quigley for his guidance on the value of information concept and its probabilistic theory. Second, I would like to thank both David Hamilton (EDF) and Jacobs seismic hazard team (Dr. Iain Tromans, Dr. Guillermo Aldama-Bustos, Dr. Angeliki Lessi-Cheimariou) for agreeing to be interviewed in the scope of my PhD. The discussions outputs contributed to optimise my research goals. Jacobs seismic team have supported me throughout my PhD. Particularly, I would like to thank Dr. Angeliki Lessi-Cheimariou for her guidance and help in understanding seismic hazard approaches.

I wish to thank the University of Strathclyde for awarding me the fundings to conduct my PhD (University Excellence Award Studentship). Being part of the department of Civil and Environmental Engineering gave me the opportunity to meet beautiful souls who motivated and supported me in hard times but also celebrated my highlights. Thank you, Georgia and Kitsos.

Finally, my special thanks go to my mum who has always been my number one fan. I dedicate all my achievements to you. I want to express my gratitude to my partner Karthik, who has been of tremendous support and motivation, and who was so patient during my endless talks about value of information that he could write a thesis himself. Many thanks to my second mum, my aunt, and to my dad, cousins, Dhafoura, my brothers and sister, my friends and everyone who believed in me.

Table of Contents

List of Figures

List of Tables

Acronyms and Symbols

AF	Amplification Factor
AFoE	Annual Frequency of Exceedance
BN	Bayesian Network
CBA	Cost Benefit Analysis
cov	Covariance
$C_{50\%}$	Median capacity
C_d	Construction cost
β	Composite logarithmic standard deviation
D	Damping ratio
DBE	Design Basis Earthquake
EAL	Expected Annual Loss
EL	Expected Loss
ENGS	Expected Net Gain of Sampling
EOL	Expected Opportunity Loss
EVI	Expected Value of Information
EVII	Expected Value of Imperfect Information
EVPI	Expected Value of Perfect Information
EVPPI	Expected Value of Partially Perfect Information
EVSI	Expected Value of Sample Information
σ_e	Error function
f_0	Fundamental frequency
G	Shear modulus
GI	Ground Investigation
GMM	Ground Motion Model
GMPE	Ground Motion Prediction Equation
H	Thickness
HC	Hazard curve

ID	Influence Diagram
M_w	Moment magnitude
μ_x	Mean or median of x
NPP	Nuclear powerplant
o	Outcome function
pdf	Probability Density Function
P_F	Seismic risk
P_f	Fragility curves – probability of damage/failure
PGA	Peak Ground Acceleration
PSHA	Probabilistic Seismic Hazard Assessment/Analysis
PoV	Posterior Value
PV	Prior Value
RC	Reinforced Concrete
ρ	Density
SA or S_a	Spectral acceleration
SCC	Systems, Structures and Components
SHA	Seismic Hazard Assessment/Analysis
SRA	Soil Response Analysis
SSM	Seismic Source Model
σ_x	Standard deviation of x
T	Spectral period
TF	Transfer Function
U	Utility function
UHS	Uniform Hazard Spectrum
VoI	Value of Information
V_s	Shear-wave velocity
Vs_{30}	Average shear-wave velocity in the first 30m

1 Introduction

"*Nothing is more difficult, and therefore more precious, than to be able to decide*."

– Napoleon Bonaparte-

1.1 Background and Motivations

Decision-making is an essential process that permeates all aspects of our daily lives and the lives of those around us. From the mundane decisions of choosing what to wear, to more consequential decisions that can have far-reaching impacts on our careers, relationships, and even the world at large, we are consistently presented with alternatives that necessitate the evaluation of the advantages and disadvantages. While it would be ideal if we could always make decisions based on a complete understanding of the potential outcomes of our choices, the reality is that most decisions are made under conditions of uncertainty, ambiguity and limited information. To make informed decisions, it is important to reduce these uncertainties by increasing our knowledge and gathering as much relevant information as possible. In this way, we can enhance our decision-making abilities, resulting in choices that are grounded in logical and defensible principles.

In all domains of study and practice, the use of measurements as means of gathering and evaluating information is of paramount importance. The allure of measurement lies in its ability to provide a comprehensive and nuanced understanding of complex phenomena. One measures for three reasons. The first reason is to facilitate the decision-making process by furnishing decision-makers with a rich trove of information to draw from, thereby enabling them to make more informed and prudent decisions. The second reason is to estimate the market value of a product. For instance, market research surveys designed to elicit consumer preferences, feedback, and sentiments can help gauge the value and potential of a product or service, thus

informing future innovations or creations. In some cases, this data can also be monetised by selling it to third parties. Finally, measurements can satisfy the intellectual curiosity of scholars and researchers seeking to expand our understanding of the world around us, pushing the boundaries of knowledge and inquiry. In short, measurements offer a versatile and indispensable means of acquiring information in a variety of contexts, serving as a critical tool for informed decision-making, market research, and scientific exploration.

In some cases, the immediate benefits of measurements might not be readily apparent. Nonetheless, the acquisition of a piece of information can be a valuable asset, providing a lasting resource that can be leveraged across a variety of applications. For example, observing butterflies might seem free from any practical purpose; however, observing butterflies' wings and measuring their interaction with water droplets enhanced the development of water-shedding technology (Bird *et al.*, 2013). Today, this technology is applied to develop super-hydrophobic surfaces, revolutionising rain jackets and aircraft wings.

One of the most used decision-making strategies is cost-benefit analysis (CBA) (Griffin, 1998; Boardman *et al.*, 2017). CBA involves weighing the costs of a particular decision against the benefits it is expected to yield. This is done by quantifying the costs and benefits in monetary terms and comparing them to determine whether the benefits outweigh the costs. In a nutshell, the purpose of CBA is to assess the economic feasibility of a given decision or project and to implement the most cost-effective option. CBA has been used in more than 140 research areas since the 1950s (Jiang and Marggraf, 2021), from which most applications are found in engineering (e.g., Van de Poel, 2009; Koopmans and Mouter, 2020), followed by environmental sciences ecology (e.g., Hammond, 1960; Hanley et al., 2009), energy (e.g., Clinch, 2004), business economics (e.g., Campbell and Brown, 2003; Tsiboe, 2015) and healthcare sciences services (e.g., Weisbrod, 1961; Machado, 1999).

While CBA has been widely used as a decision-making strategy across various fields, an alternative method called Value of Information (VoI) analysis can also serve as a useful tool for decision-making. VoI was first introduced by Raiffa and Schlaifer (1961) and has been developed and increasingly used in the last decades (Keisler *et*

al., 2014), especially in the medical and healthcare field (e.g., McFall and Treat, 1999; Bindels et al., 2016; Jackson et al., 2022). VoI involves assessing the potential value of additional information that could be obtained to inform a decision. It evaluates the expected benefits of acquiring new information against the costs of obtaining it.

If both VoI analyses and CBA are ultimately performed to inform better decisions, the type of decision-making that these approaches serve are slightly different. While CBA evaluates the benefits of selecting one option over another in light of uncertainty, VoI is concerned with addressing the underlying cause of the uncertainty. Specifically, VoI enables the assessment of the value of obtaining a particular piece of information in reducing uncertainty, before conducting the measurements, and provides a quantitative measure of the potential benefits that may arise from updating a decision based on new evidence. In other words, VoI consists in assessing the difference between the outcomes of making a decision under current uncertainties and the expected outcome after obtaining a piece of information and choosing an optimal decision. Ultimately, VoI estimates are to be compared to the cost of obtaining the information (e.g., through purchasing).

While VoI has traditionally found use in the domains of medicine (Willan and Pinto, 2005; Tuffaha et al., 2016; Heath et al., 2016) and economics fields (Ducoffe and Curlo, 2000; Levitt and Syverson, 2008), its potential application can extend to other fields, such as hazard assessments and civil engineering for the evaluation of seismic risk. The safety of structures must be guaranteed against external hazards, including earthquakes. To this end, hazard assessments are used to assess existing structures as well as to inform the design of new structures. The importance of hazard assessments cannot be overstated, especially for critical structures such as nuclear facilities, the failure of which could result in catastrophic consequences (ONR, 2017; IAEA, 2022). Accurately assessing the seismic hazard at a particular location requires a wide range of data, which are often associated with significant uncertainties.

Uncertainties are generally categorised into two types:

1. Epistemic: due to lack of knowledge of a parameter or a process
2. Aleatory: the variability inherent to the probabilistic nature of a random event

Epistemic uncertainties can be reduced through collection of new data or information; however, approaches to gather such data can be costly and time consuming. As a result, there is a pressing need to estimate the value of additional measurements used as inputs to seismic hazard assessments (SHAs). SHA is an essential step in defining the appropriate seismic design of a structure, which is critical in preventing substantial damage and potential collapse. Typically, higher seismic design levels translate to costlier designs. However, new information can help reduce the design requirements, potentially lowering the costs associated with constructing new facilities or retrofitting existing ones (Giordano *et al.*, 2022; Iannacone *et al.*, 2022). An appropriate seismic design must balance two potentially conflicting purposes: safety and economics, leading to a potential trade-off between construction costs and the acceptable target levels of safety. Risk-targeted and minimum-cost design procedures are attractive methods to balance these purposes (Gkimprixis *et al.*, 2020).

Furthermore, SHAs rely on various parameters, whose significance in the assessment determines the degree of knowledge required for each. Previous studies have found that the sensitivity of SHA outputs (e.g., hazard curves, uniform hazard spectra, and eventually, the seismic design) to different inputs can significantly vary (Aguilar-Meléndez *et al.*, 2018). As a result, it is essential to determine the extent to which we must pursue new information to constrain key SHA parameters.

One of the most important steps in SHA is site-response analysis, which relies on the characterisation of the near surface (often the top ~100m) below the proposed or existing structure. Site-response analyses can vary in complexity based on the available data and importance of the project. Key inputs include shear-wave velocity (V_s) profiles; depth to bedrock and its V_s; average V_s in the first 30 meters $V_{s_{30}}$; and fundamental resonance frequency of the site, f_0. If information to constrain these inputs is not available or not known precisely, uncertainties must be integrated into site-response analyses (e.g., McGuire and Shedlock, 1981; Abrahamson and Bommer, 2005; Ordaz and Arroyo, 2016). These uncertainties can have a considerable impact on the overall results of the SHA (e.g., Barani et al., 2013).

As for epistemic uncertainties, these can be reduced by collecting new information through geophysical and geotechnical surveys and/or by installing on-site

seismometers. Some surveying techniques (e.g., crosshole measurements) can accurately characterise the near-surface but are intrusive, costly, and take considerable time, potentially causing the overall project to run over time and over budget. On the other hand, some surveying techniques (e.g., using ambient vibrations to estimate horizontal-to-vertical spectral ratios) do not characterise the near-surface as accurately but are non-intrusive, cheaper and quicker to undertake. In this context, performing VoI analyses could be useful for seismic hazard analysts, investors, insurance providers and facilities owners. Indeed, VoI can aid in prioritising the most effective methods, or combination of methods, for characterising a site or location of interest, ultimately leading to more cost-effective and informed decision-making. With its ability to quantify potential gains from new information, VoI analysis as a tool could be a successful candidate for enhancing seismic hazard assessment strategies, leading to improved safety and risk management outcomes.

Most decision-making strategies in this context rely on CBA to determine optimal decisions under uncertainty and are based on current knowledge (FEMA, 1992; Ketchum *et al.*, 2004; Williams *et al.*, 2009). There is a paucity of applications of VoI in the field of earth sciences and civil engineering. While there have been some recent advances such as using VoI for assessing the benefits of geophysical measurements in drilling decisions (Eidsvik et al., 2008; Bhattacharjya et al., 2010; Eidsvik et al., 2015), geotechnical investigations in reliability-based design (Ching and Phoon, 2012) and more recently, structural health monitoring (Kamariotis *et al.*, 2023), there is still an urgent need to develop a method that allows for optimal decision-making not only during the design stage but also at the data collection stage. Incorporating VoI analysis at the data collection stage would allow for more efficient allocation of resources, leading to more informed and cost-effective decision-making, ultimately improving the overall hazard assessment, and subsequently, the risk and design assessments.

1.2 Aims and potential challenges

In civil engineering applications where SHA are to be performed, there are several regulatory requirements that need to be satisfied. Contractors and hazard analysts need to be rigorous in assessments and calculations, as well as capable of justifying all decisions throughout the different stages of the design process. On one hand, an

optimal decision strategy ought to consider all sources of uncertainties in the available data, the calculation process and the uncertainty in each of the decision outcomes. One the other hand, a highly uncertain decision with significant consequences might benefit from data collection to reduce those uncertainties. Currently, the state-of-the-art when dealing with these uncertainties is relying on expert judgement and published guidelines to justify data collection. However, guidelines are rarely study-specific whereas expert judgement is based on experience and beliefs, which can increase the risk of potential biases.

We propose to develop an innovative decision-making strategy for data collection that will be based on the fusion of well-established concepts, and statistical and probabilistic techniques. Our approach will primarily use the VoI concept and Bayesian methods such as influence diagrams and Bayesian inference. By doing so, we aim to address critical challenges such as the justification and prioritisation of ground investigations for near-surface site conditions, while ensuring the integration of safety requirements, modern-practice approaches and expert judgement in the field of SHA applied in seismic design. This will enable us to develop a comprehensive and reliable framework that will not only effectively guide decision-makers in the selection of an optimal data collection strategy, but also improve cost-effectiveness by avoiding unnecessary data collection.

The accurate estimation of VoI hinges on accounting for the complex web of causal relationships and interdependencies amongst parameters, along with the corresponding probabilities assigned to each of them. It is also essential to ensure that the procedure for the collection of additional information allows for the updating of these probabilities based on new evidence. Moreover, to address effectively the question of whether a parameter is worth investigating further, the framework should take into consideration the consequences of decisions, as well as the monetary costs that such a decision would entail.

While VoI has been applied in other fields of study, its application in the context of SHA presents several unique challenges such as:

a) defining the key requirements of VoI specific to this field

b) accurately propagating uncertainties in measurable parameters to estimate the variability in decisions and building probabilistic models
c) quantifying in monetary units the outcomes of the coupling of parameter uncertainties and decision variability.

Additionally, the proposed method requires the quantification of the variability in future measurements and its dependence on the type of data collection to be used. The significance of this work lies in demonstrating the potential usefulness of VoI in this field, where epistemic uncertainties in inputs might lead to high uncertainties in the final results. Seismic hazard analysts, facilities owners and insurance companies can derive significant benefits from evaluating the VoI, especially when uncertainties can be reduced through data collection. In the field of SHA, the gap analysis stage can often lead to extended debates between clients, stakeholders and analysts on whether data collection should be carried out. In such cases, the question of "should we collect data? and if so, on which parameter?" deserves a more quantifiable answer.

This thesis represents a contribution to the field of SHA by presenting a methodology for evaluating the VoI of critical parameters in SHA. For the first time, the proposed approach allows for a systematic assessment of the VoI, enabling more informed and accurate decisions regarding seismic design. More specifically, this thesis aims to:

- Examine the theoretical methodology of VoI, analyse its implementation in existing decision-making scenarios, and identify the key components that must be considered;
- Assess the suitability of Bayesian graphical and probabilistic methods and their relevance for the developed VoI approach;
- Emphasize the prevalence of decision-making in SHA and underscore the necessity of VoI as a supportive tool to rationalise and prioritise data collection;
- Specify and develop VoI approaches, from using relatively simple assumptions to more realistic applications, to decide when and what additional data to collect on near-surface site conditions aiming at inferring an optimal seismic design for a given structure;
- Identify the inputs, parameters and configurations that exert the greatest influence on VoI estimations; and

- Propose a general framework, detailed guidelines and recommendations for the use of VoI for the assessment of key parameters of seismic design and in other applications.

1.3 Outline of the thesis

The first part of the thesis consists of a thorough review of VoI theory and of previous and current applications, as well as essential statistical concepts and probabilistic tools crucial for the field-specific application. In addition, we provide an extensive evaluation of the key components of SHA and an overview of decision-making situations. The main part of the thesis consists of building and applying the VoI approach for applications of three increasing levels of complexity, from a simplified approach to a more realistic implementation withiin a state-of-the-art site-specific seismic hazard assessment via and intermediate step. Finally, we provide a general framework and guidelines as a support to decision-makers when deciding on data collection.

More specifically:

- *Chapter 2 – The Value of Information: Theory, Applications and Bayesian Analysis*

  This chapter introduces the VoI concept as a statistical decision strategy and highlights its various applications to facilitate informed decision-making. The different types of VoI are described, and their respective advantages and disadvantages are discussed. The rising use of VoI in diverse domains is emphasised, while the lack of its implementation in SHA and seismic design is highlighted. Furthermore, Bayesian methods are discussed, and their ability to respond to VoI requirements is evidenced.
- *Chapter 3 – The need for VoI within the gap analysis phase in seismic hazard assessment*

  This chapter provides a comprehensive overview of standard practice methods for evaluating seismic hazard, detailing the various stages, components, inputs and applications, while also examining their limitations. The chapter particularly focuses on the presence of uncertainty in most inputs, which can

result in biased estimations if not accounted for and at best, lead to high variabilities when considered. Furthermore, the chapter delves into the process of designing critical facilities, such as nuclear powerplants, in the UK, thoroughly evaluating the guidelines and requirements for both SHA and seismic design,and discussing important decision-making steps. To shed further light on this issue, the chapter presents a summary of two semi-structured interviews, conducted with representatives from a seismic hazard team (i.e., a consultant) and utility operator (i.e., a client). The interviews highlight the need for decision strategies to quantitatively assess the benefits of data collection.

- *Chapter 4 – Assessing VoI in a seismic design application: single parameter, Vs_{30}*

  The chapter details a VoI approach developed to estimate the added value of obtaining new evidence for a single uncertain parameter in optimising the selection of an appropriate seismic design for a particular building, at a given location. The design criterion is assumed to be only dependent on the peak ground acceleration (PGA) of the estimated hazard at the location. The approach for estimating the seismic hazard is based on the product of the hazard at the reference rock and a frequency-independent site amplification factor. The method is tailored to a (a) discrete then (b) continuous variable with a prior probability distribution. Several VoI aspects and configurations are explored such as: (a) the identification and integration of prior knowledge, (b) the characterisation of the variability of future measurements through error functions and (c) the construction of conditional dependencies between the input parameter uncertainty and the variability in the overall outcomes. Outcomes are defined as expected losses when deciding under uncertainty. The method is supported by several sensitivity analyses to identify the factors that most control the obtained VoI estimates.

- *Chapter 5 – VoI for bivariate uncertain parameters within site-specific probabilistic seismic hazard assessment*

  This chapter presents an upgrade of the application in Chapter 4, where VoI is assessed for a bivariate uncertain parameter describing site conditions for a

single layer over bedrock and for designing the same building as in Chapter 4 based on PGA. The site amplification function is obtained analytically through 1D linear site-response analysis. Joint probability distributions are introduced in the approach, allowing the estimation and comparison of the VoI of obtaining new evidence on one of the two uncertain parameters, but also on both. Sensitivity analyses using different types of measurements with specific variabilities evidence: (a) the contribution of each parameter to the overall hazard variability and to the associated decision outcomes, allowing prioritisation of data collection on one parameter over another, and (b) the influence of the type of measurements (quality and/or quantity) on VoI estimation, allowing derivation of an optimal combination of measurement techniques. Finally, the latter results are compared to those of a second case study defined by a more complex soil profile and a univariate uncertain parameter, where amplification functions are numerically computed through 1D linear site-response analyses.

- *Chapter 6 – VoI for multivariate uncertain parameters within site-specific probabilistic seismic hazard assessment: Full convolution method*

  This chapter expands the previous approach to a more realistic scenario involving a hypothetical critical facility, specifically a nuclear powerplant. This facility is situated in a moderate to high seismicity region underlain by a complex six-layer 1D profile. 1D linear-equivalent soil-response analyses are performed as well as state-of-the-art site-specific probabilistic SHAs in accordance with modern-practice. Furthermore, the approach is enhanced to incorporate design criteria that are based on a uniform hazard spectrum defined for a range of spectral periods. The VoI approach is further developed to include current published guidelines recommendations and requirements for the safety of critical facilities. The process is carefully documented, providing a comprehensive step-by-step approach. Multiple sensitivity analyses are conducted for a wide range of input parameters and assumptions, and the resulting estimates are compared to those from previous applications (Chapter 4 and 5).

- *Chapter 7– Conclusions and recommendations*

  This chapter outlines the key findings of this thesis and provides guidelines and recommendations for conducting VoI within SHA for seismic design applications and beyond. A general conceptual framework is presented, which elucidates all steps, inputs and interdependencies, along with guidance for implementing VoI in diverse contexts. Additionally, this chapter sheds light on the current challenges and limitations encountered in this field, and proposes potential directions for future applications.

2 The Value of Information: Theory, Applications and Bayesian Analysis

In this chapter, a concept in decision theory called Value of Information (VoI) is introduced. This concept is used as a tool to analyse the value of obtaining additional information in the presence of uncertainties. First, the origin and analytical definition of VoI is established. Second, the different types of VoI measurements and their purpose are described. Then, some previous applications of VoI in several fields are summarised and discussed, highlighting the lack of current developed approach for seismological information. Finally, we justify the need for Bayesian analysis in assessing VoI, where Bayesian methods are described and their utility in handling uncertain frameworks and decision-making situations are listed. This chapter gives an overview of the advantages that an approach such as VoI might have in quantifying the reduction of uncertainties through the collection of additional data and, therefore, in aiding decision-making.

2.1 Origins and definition

There are three major advantages in introducing VoI in practical applications (Eidsvik et al., 2015):

- Implementing VoI contributes to the evaluation of current knowledge, the quantification of existing uncertainties and the characterisation of potential consequences.
- VoI enables the probabilistic assessment of the measurement outcomes before measuring or purchasing a piece of information.
- VoI can be expressed quantitatively, depending on the type of expenditure to be optimised (e.g., monetary: profits, optimising costs, avoided losses; time: reducing or extending a duration; safety; health), This allows comparison of estimated value and the cost (i.e., in same units) of obtaining the information.

Raiffa and Schlaifer (1961) pioneered the use of VoI and provided much of the mathematical background concerning decision-making in an uncertain world. Their book provides an introduction to the mathematical analysis of decision-making in the presence of uncertainty. Their work represents an important contribution to several fields and domains.

To evaluate the VoI, Raiffa and Schlaifer define utility functions U (considered here as payoffs), for each action a based on observations z from an experiment e and within a given "*state of the world''* Θ that cannot be predicted, as $U(e, z, a, \Theta)$. Since the decision-maker would like to maximize the expected utility, the decision a (or decisions) should maximize U. The idea is that when it comes to adding new information z', an appropriate experiment e' should be conducted to increase the expected utility U and hence, update decision a'. The VoI is then simply defined as the increase in utility when we have that additional information:

$$VOI = U(e', z', a', \Theta) - U(e, z, a, \Theta) \quad (2.1)$$

Acknowledging that more information generally leads to a reduction in uncertainty, the key question is whether a decision should be made based on current information or whether it is best to invest in additional information by considering its potential impact on the payoff that, as a result, could lead to revisiting the original decision. The VoI is a key tool to prioritise research and the collection of more information to reduce uncertainty. This should lead to more accurate and less risky decisions. The updated decision might remain the same as the original decision after obtaining additional information (i.e., a=a'). This outcome might result from collecting data that do not inform the source of uncertainty or that might be relevant, reinforcing the probability of the original decision being optimal,

If we could know the VoI itself, we would also acknowledge the value that should be spent in conducting measurements for this information. Similarly, we would know if it is worthwhile and if we should prioritise the collection of one piece of information over another one. Eidsvik et al. (2015) contributed in identifying and representing the essential criteria for determining if a piece of information is valuable in a *pyramid of conditions* (Figure 2.1). The conditions should be satisfied from base to top.

The conditions are the following.

- Relevant (base): The measurement or information to be collected should be relevant to constraining the variable. Obtaining a piece of information should affect the uncertain parameter and have an impact on the decision-maker's beliefs and strategies.
- Material (middle): Observing the information should be able to shape the final decision. A direct or indirect link between the information and the outcomes should exist leading to a more adapted decision thanks to reduced uncertainties.
- Economic (top): The cost of obtaining the information should not exceed its value. If VoI is assessed thoroughly, it is possible to invest in conducting measurements or surveys if the associated costs are lower than the amount gained or losses avoided from a better-informed decision.

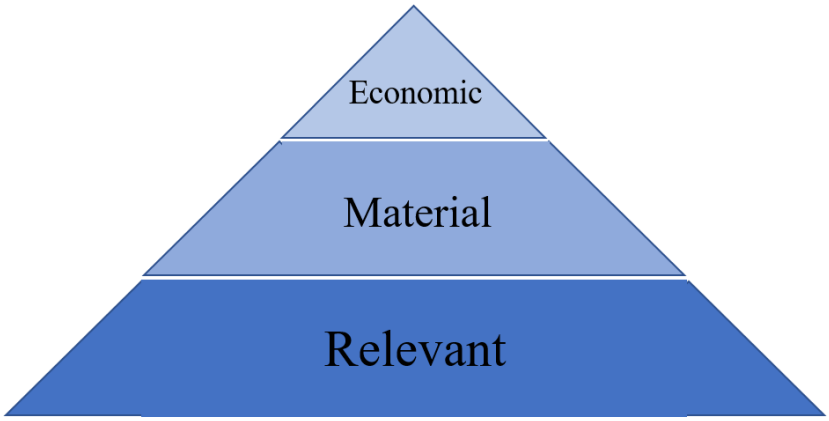

Figure 2.1: Pyramid of conditions to justify obtaining a piece information (modified from Eidsvik et al., 2015)

VoI can be considered as the amount that someone would be willing to pay to obtain a piece of information. In several fields, the decision is often based on the available information and, in the presence of uncertainties, decisions are based on expert judgement. VoI is used to reduce the need for expert judgement, not only as it emphasises the importance of understanding the uncertainty and taking it into account when making a decision, but also as it explicitly justifies the decision. Indeed, VoI can remove issues of subjective influence of personal or group interests that can bias decisions and make wholly transparent decisions that might otherwise be quite opaque. VoI is used to tackle the uncertainty as it helps show the importance of its characterisation and ways to reduce it.

VoI supports information-gathering schemes when used within a well-defined and informed framework. The general framework required to calculate VoI and employ it for decision-making is represented in Figure 2.2.

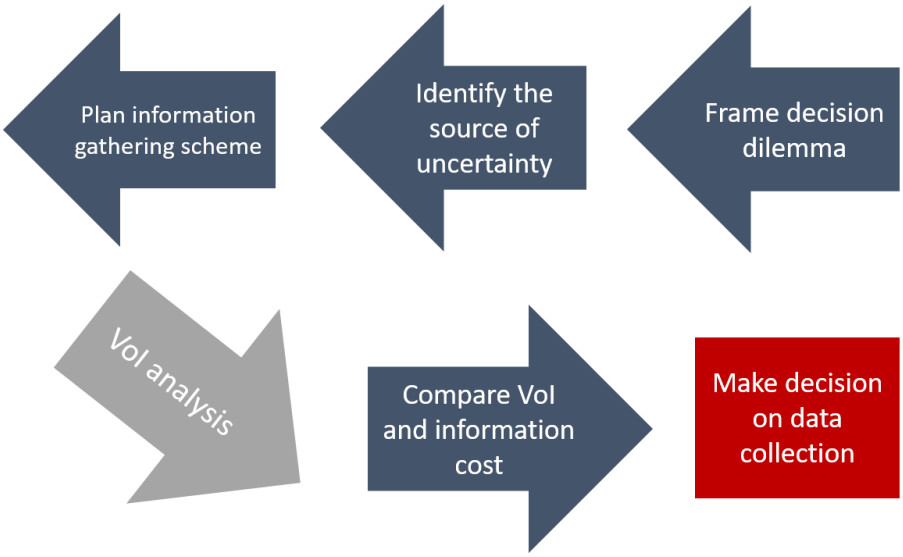

Figure 2.2: Key steps in assessing VoI within deciding on data collection

This thesis aims at defining, developing, and applying the above framework to assess the value of information within seismic hazard assessment and civil engineering.

2.2 Types of value measurements

How VoI is modelled is both field and application specific. A utility function and a unit of measurement must be defined based on the stakeholders or the decision-maker's interests. The utility function will either help estimate avoided losses, which is often used in the health care field (Tuffaha et al., 2016; Antoine-Moussiaux et al., 2019) and the evaluation of losses due to external hazards (Williams et al., 2009), or estimate the maximization in gains, which is often used in marketing and pharmaceutical applications (e.g, Wilson, 2015). The unit of this function can be monetary (e.g., representing profit or revenue) or non-monetary (e.g., time, happiness, welfare, reputation or equality). In earthquake engineering, it is more common to work toward minimising the *Expected Losses* (e.g., Gkimprixis et al., 2020).

In addition to the utility function, other choices are important in VoI calculations, e.g., the number of alternative decisions, the number of parameters considered and their types of uncertainty. The VoI method is firmly based on a Bayesian statistical

framework in which probabilities represent the degree of belief in the values that a parameter can take.

Key measurements for the VoI are numerous. As an example, VoI can be calculated by quantifying the Expected Opportunity Loss (EOL) which represents the cost of being wrong when making a decision (Hubbard, 2007). Where uncertainties are translated into probabilities, we define the EOL as follows:

$$EOL = chances\ of\ being\ wrong \times cost\ of\ being\ wrong \tag{2.2}$$

Generally, the more information we collect, the lower the uncertainties. As a consequence, facing less uncertainties might decrease the chances of being wrong which reduces the EOL. When faced with different alternatives with probabilities for different states for each one, the decision is made to follow the alternative that minimises the EOL.

Now that we have defined the EOL, the Expected Value of Information (EVI) is expressed as follows:

$$EVI = EOL_{Before\ Info} - EOL_{After\ Info} \tag{2.3}$$

The EVI simply represents the reduction in risk after considering extra information.

When the information is perfect, i.e., complete elimination of uncertainty, the associated EOL is zero and the EVI will simply be the initial EOL without that information. Under these conditions, EVI is called Expected Value of Perfect Information (EVPI) (Table 2.1).

The EVPI represents the best-case scenario, i.e., acquiring high-quality information that would lead to the best decision. Importantly, if the EVPI is less than the cost of obtaining the information, it is not worthwhile collecting that information because, even though the information completely eliminates the uncertainty and leads to a less risky choice, it is not worthwhile in terms of the unit considered (e.g., financial cost) when compared to the cost of obtaining the information. Hence, EVPI is a helpful tool when it comes to rejecting data-collection proposals as it defines an upper limit on the budget for data collection.

In most fields, and particularly in seismic hazard assessment, perfect information does not exist and thus uncertainties will always remain. As a result, the Expected Value of Imperfect Information (EVII) is a more practical concept compared to EVPI. EVII requires Bayesian updating of current information in light of new data.

When considering the problem of size sampling (e.g., number of samples, boreholes and sensors), the Expected Value of Sample Information (EVSI) is assessed to determine the optimal number of samples needed to maximise the Expected Net Gain of Sampling (ENGS) that considers the cost of obtaining the information.

Hubbard (2009) discusses VoI and its importance in deciding which variables to measure and how much we should pay to measure them. Hubbard (2009) calculated the EOL for about 100 variables in 60 different models and noticed *patterns*. Most of the time, the variables that have the highest EVPI are the least measured by organizations and on the contrary, those who have the lowest EVPI are the most measured. This is termed the *measurement inversion* and it is seen in various fields, e.g.: environmental policy, military logistics and market forecasting.

Table 2.1: Description of the types of VOI measurements

Parameter	Definition	Equation	Use
Expected Opportunity Loss (EOL)	Cost of being wrong when making a decision	Cost of being wrong × chances of being wrong	Helps determine optimal courses of action
Expected Value of Information (EVI)	Reduction in risk after considering extra information	$EOL_{Before\ Info} - EOL_{After\ Info}$	Estimates benefits of additional information
Expected Value of Perfect Information (EVPI)	EVI when the information is perfect	$EOL_{Before\ Info}$	Ideal for the rejection of proposals
Expected Value of Imperfect Information (EVII)	Gain in utility when purchasing an uncertain information	$EOL_{Before\ Info} - EOL_{After\ Info}$	Estimates the improvement in knowledge before obtaining an uncertain piece of information
Expected Value of Sample Information	Gain in utility when purchasing a number n of sample information	$EOL_{Before\ Info} - EOL_{After\ Info}(n)$	Estimates the improvement in knowledge if n samples of information are obtained
Expected Net Gain of Sampling (ENGS)	Takes into account the cost of sampling	EVSI – Cost of information	Helps the determination of optimum size sampling

2.3 Example of VoI applications

2.3.1 VoI in the literature

One of the first studies reviewing the use of VoI in the peer-reviewed literature was performed by Keisler et al. (2014). This work analyses the prevalence of VoI applications in peer-reviewed articles from 1990 to 2011. They find that VoI assessment is typically used to serve three purposes: to guide decision-makers to focus

on the information that has the most impact on a decision, to reduce unwanted consequences, and to increase the robustness of the decision-making process. These purposes can be seen as a direct consequence, in a practical manner, of the combination of the three main advantages of VoI listed in section 2.1.

Keisler et al. (2014) also evidences fields of research and engineering where the VoI is mostly used and where it needs to be more considered. They listed almost 260 applications of VoI. Figure 2.3 shows the number of applications by area and grouped in three-year periods. It is evident that amongst all fields of research, the medical field is where the application of VoI has increased the most. In medical applications, VoI methods are seen as a useful tool for handling decision uncertainties and for allowing optimal design of clinical trials and assessments by predicting the expected benefits of undertaking trials of new healthcare interventions (McFall and Treat, 1999; Eckermann and Willan, 2007; Bindels *et al.*, 2016). Commonly-developed analyses aimed at choosing the optimal number of surveys, target measurements and optimal number of samples.

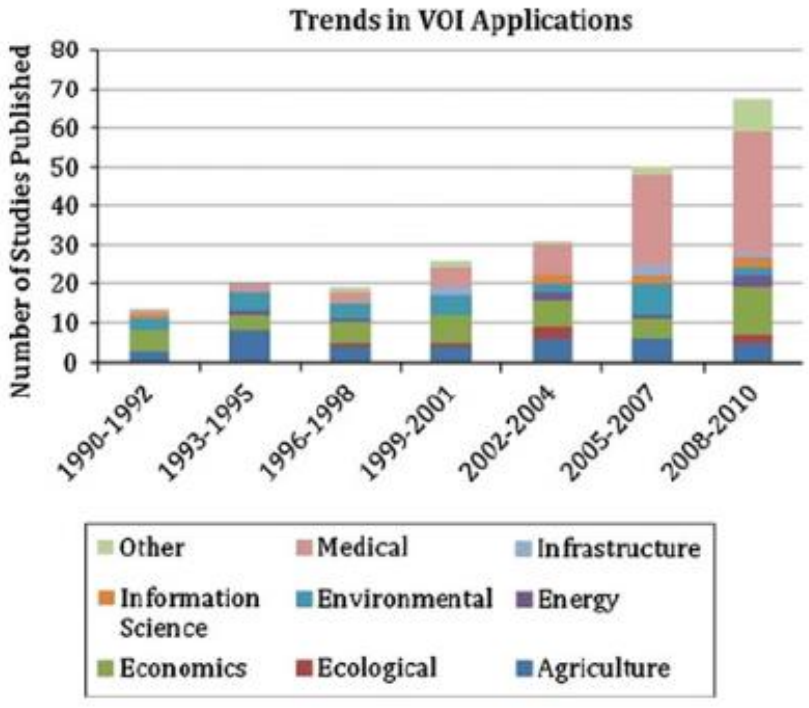

Figure 2.3 : Twenty-year trend in the use of the Value of Information in several fields of study (Keisler et al., 2014)

The second major field where VoI is commonly used is in economics and marketing. VoI can help define the benefits of purchasing information on consumers behaviour, needs and demands for specific markets (Copeland and Friedman, 1992; Chen *et al.*, 1999). As a subsequent application, VoI can be used to quantify the benefit of targeted advertisement and to decide on the number and level of investments to ensure a positive turn-over. Numerous studies focused on estimating the impact of advertising on consumers (Ducoffe, 1995; Ducoffe and Curlo, 2000; Van-Tien Dao *et al.*, 2014). Recognizing the challenge and intricacy of predicting behaviours, VoI methods enable the incorporation of multiple sources of uncertainty.

While VoI is becoming more widespread, there are currently few applications within policy and risk, or geotechnical and civil engineering. In earth sciences, VoI started becoming a popular tool within the petroleum industry to aid drilling decision-making. Recent examples for petroleum applications include Eidsvik et al. (2008), Bhattacharjya et al. (2010) and Martinelli et al. (2013). For petroleum applications, VoI is assessed using utility functions translating the expected oil/gas volume and their associated net benefits for different borehole configurations. These predictions are developed based on prior knowledge (e.g., seismic and geophysical data, lithology, porosity and saturation) and the expected added value of future data collection. The resulting VoI is then compared to the price, y, of a given experiment. If y>VoI, the experiment is not economically viable to undertake. This method enables calculations for various types of experiments over different price ranges and helps decide which ones should be undertaken. Eidsvik et al. (2015) presents extensive work in developing and implementing VoI approaches that integrate several types of uncertainties and measurements as a decision-making strategy to address spatial uncertainties and determine alternatives for gathering information.

In the last decade, VoI and similar approaches have started to be further developed in earth sciences. We can cite applications in remote sensing (Macauley, 2006; Brathwaite and Saleh, 2013), structural health monitoring (Cantero-Chinchilla *et al.*, 2020; Giordano *et al.*, 2022; Iannacone *et al.*, 2022; Kamariotis *et al.*, 2023) and geotechnical site investigations (Ching and Phoon, 2012). The increase in application

of VoI reflects the need to develop more quantitative, objective, and rigorous decision-making methods.

One of the methods for assessing VoI involves estimating the extent to which uncertainties vary with the acquisition of additional information. The presence of uncertainties during experiments, which is common, can result in considerable variability in the overall results, especially within risk-related civil engineering branches (e.g., earthquake engineering). Particularly, the method used to estimate a specific feature can be highly sensitive to the quality and amount of newly collected data. This was evidenced by Ching and Phoon (2012) who demonstrate how more or less extensive geotechnical site investigations can have an impact on the final design dimension of a geotechnical structure (i.e., pad foundation design) by considering three types of geotechnical information, separately and combined, and the results from three different design methods. They state that replying to the question "Is it worth the money/time to collect more information?" cannot be answered theoretically. Instead, it can be answered by applying design methods and varying systematically the amount of site information. Their aim was to demonstrate that more information is not only a "cost" item but an "investment" item as the reduction of uncertainties through more and higher accuracy tests can lead to design savings. The conclusion of this paper is that only one out of the three design methods (Reliability-Based Design) was able to link, in an efficient way, site investigation efforts to the design outcome in a rational way. In other words, more information about site conditions systemically leads to an increase/reduction in geotechnical structures dimensions which monetise the value of geotechnical information within the increase/reduction of the dimensions.

2.3.2 Similar approaches incorporating seismic hazard

The VoI analysis is not well exploited in the fields of seismic hazard assessment and civil engineering, regardless of its clear advantages. Nevertheless, approaches similar to VoI, used to make informed decisions, can be found under different forms and names. To cite a few, alternative decision-making approaches can be referred to as *decision analysis* or *cost-benefit analysis*.

As an example, Williams et al. (2009) developed a methodology to make informed decisions on whether to retrofit a structure that experienced previous earthquakes. The

method was based on the economic benefits of retrofitting structures in moderate and high seismic hazard areas. When it comes to such decisions, the consequences of an underestimation of the vulnerability of the structures and/or the seismic hazard at the site are a threat to general safety (i.e., risk of injuries and fatalities) and economic losses. A framework was built based on assessing the annual probability of failure of buildings P_F. P_F is computed by convolving the structure fragility curve with the seismic hazard at the site:

$$P_F = \int_{Sa} F(S_a) f(S_a) dS_a \tag{2.4}$$

where $F(S_a)$ is the seismic fragility of the structure defined as the conditional probability of attainting or exceeding a given performance level S_a and $f(S_a)$ is the annual probability density of S_a estimated from the hazard curves at the site.

The type of utility function within the approach has been defined as the Estimated Annual Loss (EAL). Results shows that EAL systemically decreases after the retrofit where the economic benefits of retrofitting are estimated using the approach of Porter et al. (2006). The retrofit is then worth undertaking if its economic value is greater than the retrofit cost. While this method includes seismic hazard assessment outputs to inform better retrofitting decisions by comparing the outcomes of each decision, it is not quite a VoI application. VoI primarily helps decide on data gathering-schemes for a specific application. In this context, a VoI application could help decide on conducting a specific measurement aiming at defining the optimal level of retrofit for a structure (e.g., structural health monitoring).

Losses due to severe damage or building collapse are not limited to material and economic losses. Other consequences such as human injuries or mortality in the case of seismic event should also be considered, which has not been the case in the contribution of Williams et al. (2009). The risk of casualties decreases when appropriate retrofitting based on a rigorous seismic hazard assessment is performed. However, estimating the statistical value of a human life remain a difficult and controversial task. The work of Galanis et al. (2018) is an example of including earthquake casualties in a complex framework to estimate the benefit of seismic upgrading. The statistical value of life was estimated between 1 M and 10 M euros and

the level of injury was linked to the building damage state and its probability of occurrence.

Clotaire et al. (2019) further developed the approach of Galanis et al. (2018) and applied it to a case study in Switzerland by defining the risk of casualties as the product of individual risk, occupancy risk and statistical value of life (10,000,000 CHF). Kappos et al. (2008) expanded the costs in their cost-benefit analysis of pre-earthquake strengthening by including the expense of indemnities related to fatalities. FEMA-227 (FEMA, 1992) models were used to infer costs of indemnities and were integrated in a life cycle-cost analysis to estimate the optimal retrofit level for mitigating seismic risk.

When it comes to life threatening hazards, a consistent cost-benefit analysis should include the estimation of both casualties and economic losses in order to make rational decisions. Therefore, a robust VoI assessment within an application that includes seismic hazard should accounts for both material losses and threat to safety.

2.4 VoI as a Bayesian analysis

To estimate the VoI, it is crucial to build a methodology that would consider the causal-relationships and the dependencies between parameters as well as the probabilities (beliefs of accuracy) assigned to each of them. When collecting more information, the built framework should allow these probabilities to be updated based on new evidence. Moreover, the method should be able to answer the question “Is this parameter worth investigating?”. To successfully answer the question, decision consequences should be accurately expressed within the framework.

Assuming that we would need to consider a method in which probabilities and uncertainties are expressed and updated, it is worth investigating the advantages of Bayesian reasoning and learning.

2.4.1 Bayesian networks and Influence diagrams

Bayesian networks

A Bayesian network (BN) is a method for modelling complex and uncertain problems (Pearl, 1988). There is an increasing use of BNs especially in medical decision support

tools (Lucas *et al.*, 2004), in banking for fraud spotting (Lev, 2008), in earth sciences (Eidsvik et al., 2015), in decision analysis (Howard and Matheson, 1984) and in statistics (Smith, 1987). BNs constitute tools for reasoning under uncertainty.

A BN is a fusion of two models: graphical and probabilistic. This combination ensures capturing both conditional dependencies and independencies between parameters. The graphical aspect comprises several variables called nodes, linked to each other when directly dependent by arrows called *Edges*. A set of probability distribution functions (*pdf*s) are assigned to these variables, constituting the probabilistic model.

To illustrate the concept, let's consider two variables A et B, conditionally independent. Both variables depend on a third variable C. Graphically, A and B will be linked to C using edges, whereas no edge links A and B as shown in Figure 2.4.

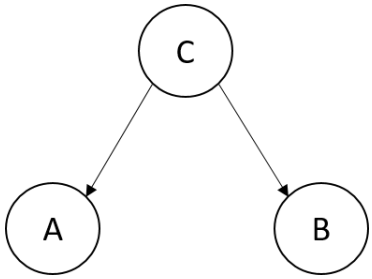

Figure 2.4: BN example of a dependence on two variables

Given that A and B are not necessarily totally independent but independent given C, the associated probability of A and B conditioned on C are expressed as follows, respectively:

$$P(A|C,B)=P(A|C) \quad (2.5)$$

$$P(B|C,A)=P(B|C) \quad (2.6)$$

Each edge within a BN encodes a specific factorisation of the joint distribution. The joint distribution of all the variables in this case is expressed as follows (Stephenson, 2000):

$$P(A,B,C)=P(A|C)P(C)P(B|C) \quad (2.7)$$

Another configuration illustrated in Figure 2.5 has a joint distribution as follows:

$$P(A,B,C)= P(A)P(B|A)P(C|B) \quad (2.8)$$

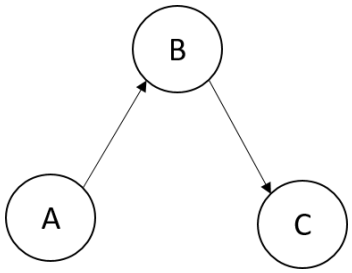

Figure 2.5: Simple example of a BN with 3 variables

This simple BN expresses a suite of dependences where node C is directly dependent on B, which in turn depends on node A.

In general, if we assume having N nodes $X = X_1, \dots, X_N$, the factorisation of the joint distribution of all the variables is expressed as follows:

$$P(X) = \prod_{i=1}^{N} P(X_i|parents(X_i)) \quad (2.9)$$

The joint distribution of all nodes is simply the product of the probabilities of each variable given its parents' values. A parent node of a variable X_i is a node which one of its outgoing edges directs toward the variable X_i, called the *child*. The key feature of BNs is that models are directed acyclic graphs (Neapolitan, 1989), meaning that all the edges point in a specific direction and there are no cycles or loops. This property allows BN to efficiently represent probabilistic dependencies between variables.

Influence diagrams

Influence diagrams (ID) are acyclic directed graphs representing decision problems (Howard and Matheson, 1984). Similar to BNs, IDs are useful in describing the structure of a decision problem, along with all variables directly or indirectly linked to the decision. The primary purpose of using IDs is to identify the decision alternative that would maximise the outcome of interest (i.e., utility). IDs can be seen as an extension to BNs as they allow representing various types of variables:

- Chance nodes: Identical to BNs nodes, they are usually drawn as circles and represent uncertain quantities defined by conditional probability distributions.
- Value nodes: Usually drawn as diamonds and represent the utility of interest dependent on the state of the parent nodes
- Decision nodes: Drawn as rectangles, they model a set of alternative decisions

To simply illustrate the structure of an ID, we give an example of a decision problem in Figure 2.6. In this example, the decision-maker must decide whether to invest in obtaining additional data based on existing knowledge. This causality is expressed with an arrow starting from the uncertain chance node 'Prior knowledge' and directed towards the decision node 'Invest in data collection'. The consequence of investing in the new state of knowledge is represented by a new chance node, 'Actualised knowledge', translating an update of the prior knowledge. The ID then allows the estimation of the outcomes (i.e., utility) for both decisions (i.e., invest or not) within the value node 'Outcomes'.

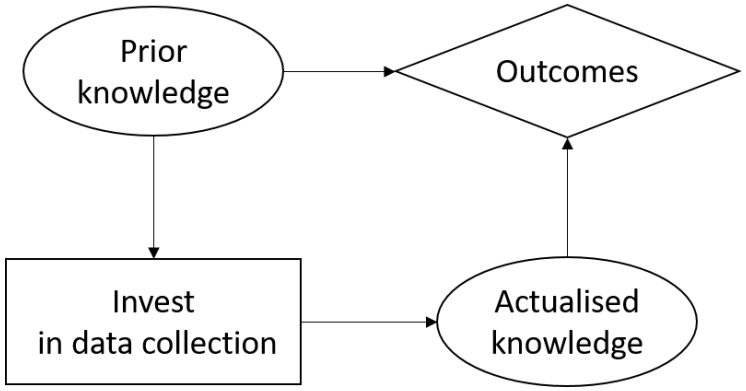

Figure 2.6: Simplified representation of an influence diagram

BNs are identical to IDs when they are only constituted by chances nodes. IDs have the same advantages as BNs, which are detailed in the following sections. They represent a suitable tool to define and develop a VoI approach that could integrate prior knowledge, allow updates from new evidence and select an optimal decision amongst several alternative options.

There are two types of decision alternatives: *test decisions* and *action decisions* (Raiffa and Schlaifer, 1961; Friis-Hansen, 2000). *Test decisions* represent the decisions of collecting more evidence to add to the model and *action decisions* are made based on the chosen *test decisions*. To give an example, a test decision can be a decision to collect more data on the vulnerability of a structure through structural health monitoring, and an action decision might represent retrofitting a structure after obtaining these measurements.

In Chapter 3, we show that within seismic hazard assessment, several steps usually have a determined order where multiple decisions have to be made. Experts are called to give their insights and judgments to determine the approach that best satisfies stakeholders, regulators and/or public safety. When a study involves constructing a nuclear powerplant or retrofitting an existing one, there is even more pressure in making an appropriate decision that adheres to safety requirements and assurance compliance. All available information and *test decisions* are carefully considered and assessed to have a solid base when taking *action decisions*. Action decisions are, by definition, dependent on the expected assessment results and implied cost.

2.4.2 Integration of prior knowledge

When building a BN or an ID, both the structure and the variables' probability distributions are not always known. In this scope of research, the structure is well-known, which acknowledges that probabilistic seismic hazard assessment (PSHA) has been performed for several decades and that the conditional dependences of the parameters and their use have a minimum of constraints. However, several parameters can be poorly constrained, and carry different types and levels of uncertainties. The case when the structure is well-known with partially observed parameters is discussed in Stephenson (2000), using the introduction of Heckerman (1999) to some of the issues involved in learning within a BN.

Some prior knowledge on parameters values is usually available because PSHAs rarely start from scratch. The parameters can represent physical measurements such as shear-wave velocity (V_s) profiles, strong-motion data, and earthquake catalogues including historical and instrumentally-recorded events. Prior knowledge can also be weights assigned to models such as seismic source models (SSMs) and ground-motion models (GMMs), which are themselves based on measurements, but are usually not site-specific, and thus, can be subjected to uncertainties.

Types of prior knowledge

Prior uncertain knowledge is often translated into prior probabilities within Bayesian analysis, representing probability distributions given to uncertain parameters and based on current knowledge before acquiring further evidence. While uniform

distributions are used as noninformative priors when a parameter is fully unknown, informative priors such as beta and gamma distributions can be constructed when some data are available.

It is important to clarify that some of the probabilities assigned within Bayesian analysis are not only physical probabilities (Heckerman, 1997). A physical probability is, for example, the probability of having a six when rolling one die. Physical probabilities are built from data and probability theory. However, some Bayesian probabilities represent degrees of belief and are often built on knowledge elicitation from experts in the domain (Woodberry et al., 2004).

Probability models from data

In general Bayesian analysis, priors regarding parameters can be informed based on available data. Data can be obtained from observations or measurements from experiments and can be quantified by individual probability models or probability models conditioned on other variables. Such models can be built using Empirical Bayes methods to estimate priors using available data (Robbins, 1956; Casella, 1985; Carlin and Louis, 1996; Efron, 2010). Empirical Bayes methods approximate more exact Bayesian methods.

There are three major steps in Empirical Bayes estimation:

- Estimate the overall distribution of data
- Infer the distribution model that represents the best fit
- Use the distribution as a prior to estimate the probability of each value

Translating a set of data into a probability distribution is achieved through building the histogram of the data, identifying the best distribution that describes the data and estimating the associated parameters. Programming languages such as MATLAB and Python enable the identification of such distributions (e.g., beta, binormal, gaussian or gamma distributions) and offer a means to estimate the best fit to the data using methods such as the Maximum Likelihood Estimation (MLE) and the Method of Moments (MM). Noise in available data, for example, can also give a level of

confidence and reliability to the data and enable construction of conditional probabilities for the parameters involved.

The advantage of fitting a model distribution is to describe large sets of data using a limited number of parameters (e.g., median, mean, variance, standard deviation). Besides learning about the distribution of parameters from observed data, empirical Bayes methods allow the estimates from future observations to be improved and decision-making to be enhanced.

Probability models from expert elicitation

Expert elicitation is conducted with experts whose knowledge and experiences can support informed judgments on a decision, a prediction or an outcome. Expert elicitation is often undertaken for problems that carry a significant number of uncertainties, which are usually difficult to constrain. The results of such elicitation are probabilistic, meaning that subjective probability distributions are inferred for the considered parameters or phenomenon. When information is severely lacking, experts intervene to propose the most suitable prior distributions and models, according to their experience and knowledge, with a certain level of confidence and belief. Following the elicitation process, encoding procedures are put in place to translate the various judgments as probabilities using methods such as the fractile technique that requires each expert to assess specified percentiles of the distribution (e.g., Walls and Quigley, 2001).

Several guidelines have been published that detail the recommended steps for conducting rigorous elicitations (Meyer and Booker, 2001; EPA, 2009; EFSA, 2014). Recommendations on the involved steps such as the identification of elicitation variables, selection of experts, conducting the elicitation and the post elicitation analysis vary depending on the type of guidance and whether it is domain-specific (Bojke *et al.*, 2021). Relying on expert judgments is common practice in several applications such as in medicine and pharmacology (e.g., Bennett et al., 2005; Grigore et al., 2013; Walley et al., 2015), environmental studies (e.g., Kotra et al., 1996; Choy et al., 2009; Nemet et al., 2017), economics (e.g., Leal et al., 2007; Iglesias et al., 2016) as well as in PSHA (Budnitz *et al.*, 1997).

Expert elicitation usually answers this type of question: *What is the probability that a variable X takes this state knowing the values of its parents?* Within the current application in PSHA, the question might be '*What is the probability of a V_s profile being accurate knowing that measurements from ambient vibration gives a site fundamental frequency f_0 with uncertainty σ_{f0}?*

The level of accuracy in the expert elicitation process can be assessed by using different methods: comparing elicited values to statistical values, comparing values from multiple elicitations and double-checking cases where probabilities are extreme (Woodberry et al., 2004; McBride et al., 2012). Nevertheless, expert elicitation can still be biased by incomplete knowledge or a misunderstanding of the causal relationships and dependency between variables (e.g., Tversky and Kahneman, 1974). Hence, there is an increase in interest for developing automated methods for building BNs from data (Heckerman and Geiger, 1995; Wallace and Korb, 1999).

Building the prior knowledge

Building the prior model is the first step for using Bayesian inference, which is to update the model in light of additional evidence. The ideal way to build a prior model reflecting the current state of knowledge is to combine both expert elicitation and knowledge from data.

Woodberry et al. (2004) proposed a methodology for a quantitative Knowledge Engineering Bayesian Network. In this approach elicitation from expert as well as machine learning using data are combined to proceed to parameter estimation and quantitative evaluation to study the sensitivity of a BN to the structure, to the parameters' prior and conditional probabilities as well as to adding new evidence. Parameter estimation is the first step in building the model by integrating the current knowledge. It is performed by assigning conditional probabilities to all variables that have at least one parent either from expert elicitation, from data or based on both. Combining both comes with a condition: the expert elicitation needs to be calibrated to the available data.

2.4.3 Bayesian inference

Bayesian inference is a statistical method for updating the probability at a node when more evidence is available about its parent nodes. Probabilities in nodes are called beliefs (Heckerman, 1999) and can be propagated (Pearl, 1988). This means that when a new observation is made (e.g., measurements, surveys) on one or more variables, it can be added to the BN or ID, leading to updating the beliefs of some variables that are directly or indirectly linked to the nodes whose values have been observed.

This concept and methodology of computing the updated beliefs is essential for estimating the VoI. Assessing the VoI within the current field of application (i.e., seismic hazard assessment) requires considering all parameters, their values and uncertainties as well as the possibility of obtaining new measurements. Therefore, Bayesian inference will be an essential component within the developed VoI approach to compute the impact of enhancing information about a parameter and therefore assessing its value regarding a range of beliefs.

Bayesian updating

When collecting a piece of information *y* on a variable of interest *x*, *y* can be *perfect*, meaning that it perfectly informs about *x*, or imperfect/partial mainly because of noise or because it represents one variable of a multivariate set. For example, in seismic hazard assessments, a 1D shear-wave velocity profile (V_S) could be the variable of interest *x*. An 'almost' perfect information *y* could be obtained from borehole measurements if it is deep enough to reach the bedrock. Partial information could be measurements from a shorter borehole or using another non-invasive technique that would carry some variability in the results.

The model for observed data *y* can be represented by a conditional *pdf* that expresses the probability of obtaining measurements *y* if the true state of the value of interest is *x*. This model is called the likelihood, $p(y|x)$. Moreover, prior knowledge on the state of *x* allows defining a prior probability model $p(x)$. Ultimately, we aim at computing the updated probability of *x* in light of the new evidence *y*, $p(x|y)$, called the posterior probability.

The posterior probability can be computed using Bayes' rule (2.10), allowing a transition from a straightforward problem of predicting the probability of some measurements based on prior knowledge to updating the knowledge given evidence. Both forward and inverse model transitions are illustrated in Figure 2.7.

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} \tag{2.10}$$

where $p(y)$ is the marginal or pre-posterior probability model translating the probability of obtaining a given observation:

$$p(y) = \int p(x)p(y|x)dx \tag{2.11}$$

The posterior can be viewed as a combination of prior knowledge and information brought by the data.

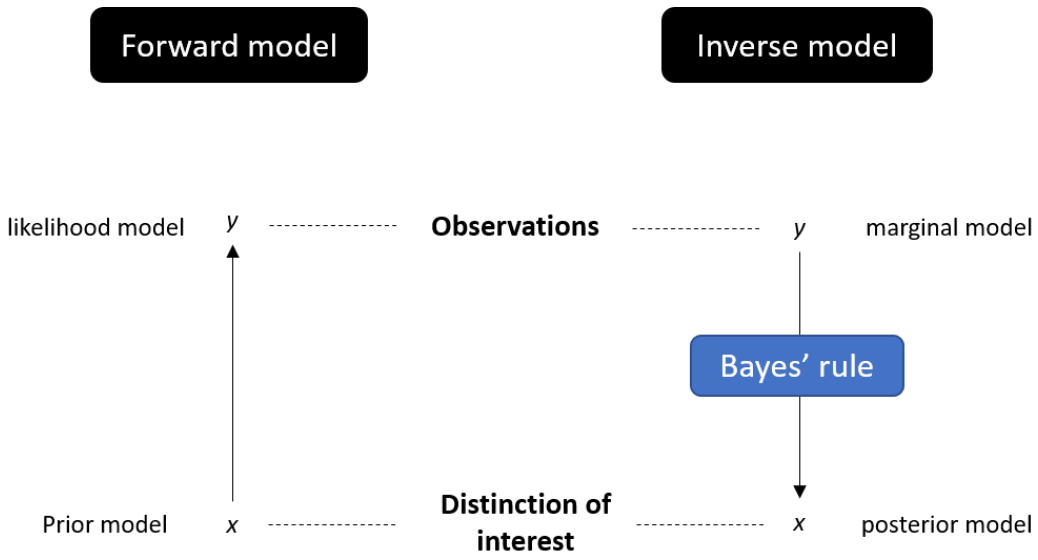

Figure 2.7: Framework to compute the posterior model: forward and inverse Bayesian modelling using Bayes' rule (modified from Eidsvik et al., 2015)

Bayes' rule is used to construct posterior distributions that are essential for VoI calculations. It is often used in decision trees (Gilbert and Habibi, 2015), in Bayesian analysis for learning and updating geotechnical parameters from measurements (Straub and Papaioannou, 2015) and in BNs within various examples in earth sciences (Eidsvik et al., 2015).

Quantitative evaluation

BNs and IDs are excellent tools for analysing the sensitivity of a framework and associated outcomes to observations, by updating probabilities when input parameters are varied. The quantitative evaluation gives a measure of the degree of independency between various nodes. While using the link structure (i.e., visually analysing the edges in a graph) can be an option, it is not always feasible with complex frameworks. To tackle this, using the d-separation rule can help identify whether the evidence (observations) in one variable can influence the belief in a query variable (Pearl, 1988). The d-separation rule is a criterion used to identify conditional independencies in a BN or ID. In other words, it helps decide whether a set of variables X is independent on a set Y given a set Z. It can be quantified by two types of measurements:

- The entropy H of a multivariate variable X used to evaluate the uncertainty or randomness of a probability distribution before an observation Z:

$$H(X) = -\sum_{x \in X} p(x).\log\,(p(x) \tag{2.12}$$

- The Mutual Information (MI) to measure the effect of one variable Y to another variable X after observation Z:

$$MI(X|Y) = H(X) - H(X|Y) \tag{2.13}$$

These measurements help to identify variables that can affect the MI and enable the quantification of the degree of dependency between variables. Particularly, quantitative evaluation allows us to determine whether the variables are insensitive or on the contrary, highly sensitive to certain changes. This focuses efforts towards refining conditional probabilities for certain variables when they are highly sensitive and identifying errors in the structure. By providing insights into the relative importance of different variables, this analysis can assist experts in identifying which factors have the greatest impact on the overall outcomes.

In site-specific seismic hazard assessment, for instance, information about the soil materials and stratification is known to have an important impact on the estimated hazard. Hence, site properties are investigated and should be wisely characterised to reflect, as much as possible, the site and path conditions. Experts need to consider all

potential possibilities and integrate them into the calculations. Sensitivity analyses will determine the contribution of each component to the model and will then help focus on uncertain parameters with the highest impact.

Alternative sensitivity analyses may require using an empirical approach consisting in changing, one by one, a parameter's value in a node and studying the changes in the posterior probabilities of a target node. However, this can be time consuming, especially when networks are large. Coupé and van der Gaag (2002) tackled this issue by identifying a set of variables (using d-separation) that, from evidence, could have an impact on a target node. They demonstrated, from systematic changes of a parameter value, that the changes in posterior probability distribution given evidence can be given a functional representation, either linear or hyperbolic. This helped identify and characterise a certain number of variables with high impact on the outcomes.

Similarly, Woodberry et al. (2004) used this approach to calculate sensitivity functions defined by specific coefficients for different parameters. Results were used as a selection criterion for expert elicitation inputs to avoid errors in the structure or in the definition of conditional probabilities. In essence, this sensitivity analysis gives guidance in determining what data should be prioritised and collected to infer a specific parameter.

2.4.4 Bayesian analysis as a tool for handling uncertainties and optimising decision-making involving seismic risk assessment

The occurrence of many damaging earthquakes throughout history highlighted the need to better consider the vulnerability of structures as well as the importance of seismic retrofitting. The decision-making process for property owners is not always straightforward when it comes to seismic retrofitting, mainly due to economic limitations. There is a need to perform risk-based prioritisation of buildings at risk (Ellingwood, 2001; Tesfamariam and Saatcioglu, 2010).

Uncertainties in available data and expert judgments can complicate seismic risk assessment and subsequent decision-making (Wen *et al.*, 2003). Several methods offer ways of taking into account these uncertainties, such as soft computing including fuzzy-based methods, neuro-computing, probabilistic theories, BNs or learning theory

(Zadeh, 1994; Knezevic et al., 2018). Especially, when historical data are sparse and information is incomplete, probabilities can be imprecise. Methods such as second order distributions, lower and upper bounds and propagation of variances can allow the propagation of imprecise probability through dependency structures such as BNs and IDs to handle the uncertainties on probabilities and data-to-data relationships and dependencies (Kleiter, 1996; Tesfamariam, 2013). They can be highly beneficial when physical models are uncertain but expert intuitive knowledge is available (Li *et al.*, 2010). There are increasing applications of the use of BNs and IDs in regional risk assessments (Cockburn and Tesfamariam, 2012), seismic hazard assessments (Bensi *et al.*, 2009; Bayraktarli *et al.*, 2011), risk assessments for buildings (Faizian *et al.*, 2004; Tesfamariam *et al.*, 2010; Bayraktarli and Faber, 2011) and loss assessments (Li *et al.*, 2010).

In decision-making problems, decision trees (e.g., IDs) can be highly beneficial in the decision-making process under uncertainty as they allow (depending on a set of alternative decisions, observations, associated probabilities and outcomes) to compute the expected utility for each set of parameters and decisions called a *sequence*. An optimal decision strategy will be the sequence(s) associated with the maximum expected utility. More details about decision trees are available in the literature (Benjamin and Cornell, 1970; Raiffa, 1997; Jordaan, 2005). When decisions can be made about several components (nodes) without knowing the exact order, all the possible permutations should be expressed in the tree. By computing the expected utility for all of them, a prioritisation strategy is then possible through IDs, containing a probability model and a decision problem structure. That is at the heart of the problem addressed in this PhD thesis: knowing which information should be collected first, second and so forth to have the most benefit and the least expected loss.

Limited memory influence diagrams (LIMIDs) can be used as an approach to an asymmetric problem in the ID. In other words, when the order of decisions is not known at the time of the ID construction. LIMIDs assume that only the nodes that are represented as a parent to a decision are known at the time the decision is made. LIMIDs are supported in popular commercial BN Software such as Hugin Expert A/S 2008.

Bensi et al. (2011) applied LIMIDs for optimising a post-earthquake decision strategy for inspections and closure. In this application, there are several components that could or could not be damaged after the occurrence of an earthquake. LIMIDs were used in the definition of an ID to help the decision-maker identify the components that should be inspected and to determine an optimal order of inspection. This is achieved by calculating the expected utility of inspecting a component and the loss in case it is damaged but has not been inspected. Two main decisions, with a specific order should be taken here: inspection or not and shutdown or not. The proposed diagram for each component i is built as follows:

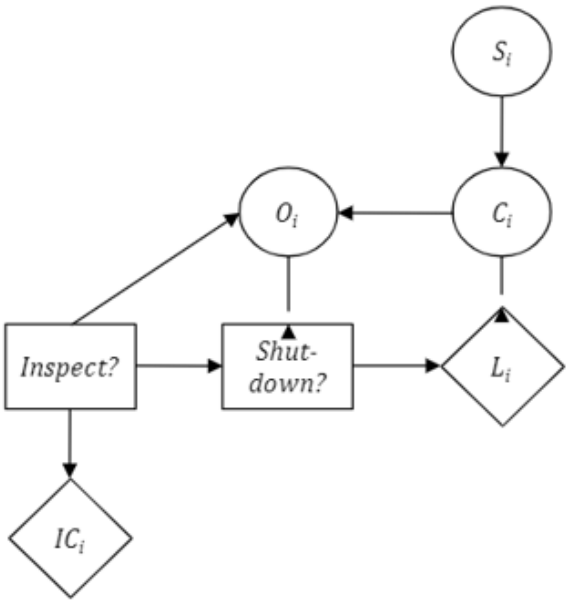

Figure 2.8: ID modelling an inspection-shutdown decision at component level (Bensi et al., 2011) S_i: seismic demand - C_i: actual damage state – O_i: observed damage after inspection – Inspect?: decision on inspection – Shutdown?: decision of shutting down the component – IC_i: cost of inspection – L_i: losses

S_i represents the seismic demand/ground motion intensity, C_i is the actual damage (states of very low/low/moderate/high/very high) and O_i is the observed damage that takes a state after the decision to inspect (Inspect?) has been made. The inspection has a cost IC_i. If an inspection is performed, the observation will be made and this represents a piece of evidence to update the decision 'shutdown', reducing or leaving the component in full operation. L_i represents the losses, which is a function of the shutdown decision and the real state of the component (damaged or not). The link

between the two decisions ‘inspect?’ and ‘shutdown?’ sets an order on these decisions. Defining this order encodes the VoI obtained from the inspection.

The particular case presented by Bensi et al. (2011) contained eight components, the VoI of each component is computed in case of no evidence (no inspection), imperfect inspection and perfect inspection. The EVPI has been computed as well as the Expected Value of Sample Information (EVSI) based on the definitions in section 2.2. The study indicated that VoI varies depending on the configuration of the system, the dependency between components and the reliability and liability of each one. Similarly, the prioritisation of inspecting a component depends on these latter configurations.

2.5 Conclusion

In seismic hazard assessment and earthquake engineering there is still a gap in decision strategies when it comes to justifying data collection. There are no existing approaches quantifying the benefits of data collection within seismic hazard assessments and as a consequence, in earthquake engineering. Currently, there is a challenge in determining the potential usefulness of collecting a specific piece of information for seismic hazard assessment. Particularly, it is unclear whether this information can significantly impact decisions related to structural design or retrofitting. Although cost-benefit analyses have been carried out for many different applications, including in earthquake engineering, VoI not only estimates the benefits of making one decision over another but, most importantly, VoI estimates the benefits of data collection before collecting the information.

This literature review finds no clear evidence that uncertainties reduction leads automatically to cost savings. Nevertheless, uncertainties reduction enhances the reliability and accuracy of the outcome’s estimations given a decision.

From the current understanding of VoI decision theory and the motivations of building an approach tailored to the problem and to the scope of this PhD, a number of must-haves have been identified. These represent the main features and components that this innovative approach, applied in assessing the value of seismological information in hazard assessment applications, should include:

- Conditional probabilities

It is essential, within the framework to be built, to express the causal-relationships and dependencies between the various variables. A variable value and uncertainty might be conditioned on other variables value and uncertainties. The conditional nature of these variables can be defined through estimating conditional probabilities. Conditional probabilities are important for VoI analysis as conditioning an observation from information could lead to improvement in the decision making.

- Graphical models

To understand how variables are linked, connected and influenced by each other, or none of these, graphical models are a powerful tool to serve this purpose. Such graphical models are referred to as Bayesian networks, Bayes nets or belief and influence diagrams. Besides these models being a visual tool, they allow defining individual probabilities, conditional probabilities and propagating new information to update these probabilities. These properties and advantages align with the core of the VoI approach to be developed.

- Updating

The approach should include the definition of priors and likelihood functions to estimate posterior distributions in light of new evidence. Bayesian updating is a major step in assessing the impact of added information on the overall conditional probabilities and defined outcomes.

This chapter highlights that Bayesian analysis, BNs and decision trees include these must-haves and could be an essential tool to assess VoI within uncertain parameters in seismic hazard assessment as uncertainties are present in almost every component.

In the next chapter, we describe the key components and main uncertainties within seismic hazard assessment. Then, we highlight the important steps and decisions for a rigorous assessment that is compliant with regulations and safety guidelines for design purposes. In the following, we develop the understanding of one of the first stages in seismic hazard assessment to decide on data collection; the gap analysis. Finally, we provide a summary of interviews conducted with seismic hazard analysts and a

representative of nuclear facilities operator confirming the need for developing an approach like VoI to aid in the decision-making process.

3 The need for VoI within the gap analysis phase in seismic hazard assessment

This chapter introduces the seismic hazard assessment (SHA) framework, the key components at play and their various sources of uncertainties, particularly within site-specific characterisation. We present the most common ground investigation techniques and highlight the importance of evaluating their cost-benefit before conducting them. Then, an overview of the current regulations and safety guidelines concerning the design of critical facilities is given, using nuclear installations as a case study. We provide a detailed overview of the decision-making process in SHAs performed by seismic hazard analysts, as well as in seismic design studies conducted by structural and civil engineers. Finally, we present a summary and outcomes of two interviews (detailed in Appendix A): one interview conducted with a critical facilities operator, and another with a group of three seismic hazard analysts, which were focused on SHA, regulations and strategies currently in place to justify data collection. The main motivation behind conducting these interviews is to identify where VoI could be used as a support tool in the decision-making process.

3.1 Seismic hazard assessment

SHA is a key concern in earthquake engineering. Quantifying and characterising the rate of ground-motion occurrence and its intensity are essential in seismic design, especially for critical facilities (e.g., nuclear installations). There are two representative methods aiming to cope with complicated and random earthquake processes: Deterministic and Probabilistic Seismic Hazard Assessment, respectively referred to as DSHA and PSHA (Cornell, 1968; McGuire et al., 2001). The former relies on geological features and the strongest past events regardless of how often they occur. The latter estimates the probability of exceeding some levels of ground shaking at a specific site or region. In fact, PSHA considers all historical and instrumentally

recorded events, enabling the incorporation of both the most probable (smaller) events and the rarer (largest) events. PSHA is used to produce seismic hazard maps, evaluate the seismic risk of a structure or region, define building codes requirements and assess the safety of critical facilities.

3.1.1 PSHA components

Typically, PSHA is performed following the four-stage process illustrated in Figure 3.1.

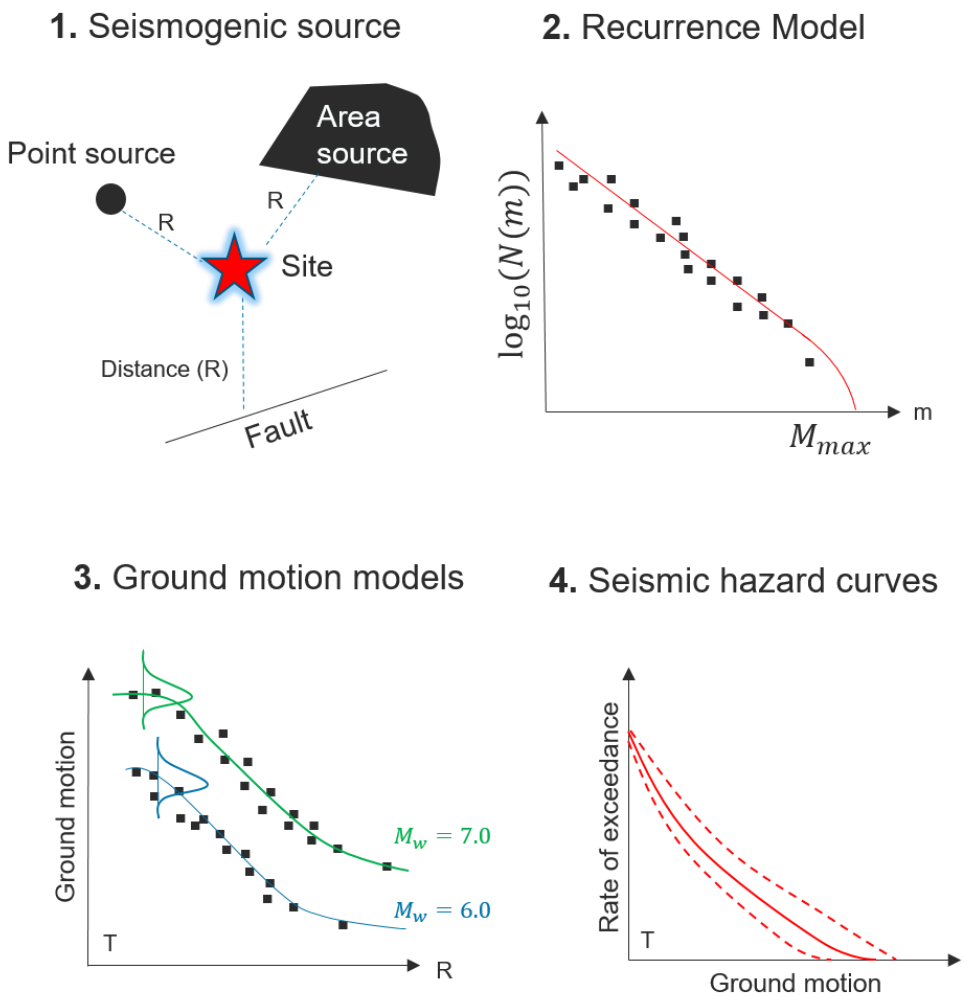

Figure 3.1: General PSHA framework. T is the spectral period and N the number of earthquakes of magnitude exceeding m. M_w is the moment magnitude and M_{max} the moment magnitude at which the recurrence model is truncated

The first and the second stage fall within the characterisation of seismic source models (SSM). The first stage requires the identification of all seismic sources within a region or in the vicinity of the site of interest depending on the level of available information, the definition of source types and the characterisation of the rate of seismicity. The sources can be either included within the framework as point or areal sources, where the seismicity is uniformly distributed and homogeneous in space and time, or as fault sources when enough data are available to characterise fault rupture mechanisms and seismic activity.

The second part of source model characterisation consists of building recurrence models, for which the Gutenberg-Richter model (Gutenberg and Richter, 1944) is the best known. The Gutenberg-Richter model describes the relationship between the seismicity rate and the magnitude of earthquakes in a specific region. The rate of occurrence of events with magnitude M exceeding a threshold value *m* is expressed as a linear function in logarithmic scale as follows:

$$\log_{10} \lambda_m = a - bM \tag{3.1}$$

where a and b are the coefficient of the linear regression. An example of a recurrence model is shown in Figure 3.1. This relationship indicates that the stronger the seismic event, the lower is its probability of occurrence.

The third stage of PSHA is establishing/selecting ground motion models (GMMs) that describe the level of intensity and the attenuation of ground motion with distance. These models are also called attenuation relations (Abrahamson and Silva, 1997) and, most commonly, ground motion prediction equations (GMPEs). GMPEs are derived from empirical data, simulated data or a combination of both. They relate ground-motion parameters to variables describing the earthquake's source, path and site effects for a specific earthquake scenario. Using GMPEs for different spectral periods can result in a response spectrum, which is a plot of the peak response of a series of damped oscillators of various natural frequencies forced by a single motion. They are essential and are being increasingly developed within engineering seismology (Douglas, 2017). The general form of GMPEs is expressed as a combination of a source function f_{source}, the path f_{path} and the site f_{site} as follows:

$$\ln(Y) = f_{source}(M, fault\ mecanism) + f_{path}(M, R) + f_{site}(V_{s30}\ or\ site\ class, f_0 \dots) \pm \varepsilon\sigma \quad (3.2)$$

Y: ground motion intensity at spectral period T

M: Magnitude

R: source-site distance

V_{s30}: Average shear-wave velocity in the first 30 meters

$site\ class$: Site classification from Eurocode 8 (A, B, C, D, E, S1, S2)

f_0: Fundamental resonance frequency of the site

σ: standard deviation (residuals) of the ground motion logarithmic distribution

ε: number of standard deviations above or below the median.

It is of upmost importance to consider uncertainties within the GMPEs. In generic approaches of SHA, σ represents an aleatory variability, which is often assumed to be unreducible. It captures the inherent randomness of a model or a natural process. In addition, a part of the source, path and site functions might suffer from epistemic uncertainties (i.e., those due to a lack of knowledge). These uncertainties can be reduced depending on the level of knowledge, the amount and processing of available information as well as the acquisition of new data. Epistemic uncertainties can be captured and propagated in the framework through the use of logic trees (Reiter, 1990), where alternative seismogenic models, recurrence models and GMMs are taken into account according to experts' level of confidence (e.g., Bommer & Scherbaum, 2008; Delavaud et al., 2012).

3.1.2 Seismic hazard calculations

By combining a SSM and a GMM, it is possible to assess seismic hazard by calculating the probability that a ground motion Y exceeds a specific value y^* for a given earthquake occurrence.

The probability of exceedance in t number of years is expressed as follows (e.g., Baker, 2013):

$$P[Y > y^*|t] = \iint P[Y > y^*|m, r] f_M(m) f_R(r) dm dr \quad (3.3)$$

where $P[Y > y^*|m, r]$ is retrieved from the GMPE and $f_M(m)\ and\ f_R(r)$ are the probability density functions associated to the magnitude and distance, respectively.

Similarly, the annual rate of exceeding y^* can be written as follow:

$$\lambda[Y > y^*] = -\frac{\ln(1-P[Y > y^*|t])}{t} \tag{3.4}$$

It is then possible to define the return period *RP* as:

$$RP = \frac{1}{\lambda[Y>y^*]} \tag{3.5}$$

The main outputs of PSHA are Hazard Curves (HC), Uniform Hazard Spectra (UHS) and hazard disaggregation. HCs are computed for each spectral period of interest and express the rate or probability of exceeding a given ground motion intensity. A UHS is defined by a specific return period where input intensities are retrieved from HC at each spectral period at a fixed rate of exceedance. Finally, the disaggregation of the hazard consists in identifying the magnitude-distance pairs that most control the hazard at a specific return period. Figure 3.2 gives an example of a hazard disaggregation at a return period of 475 years showing that earthquakes of magnitude 6.1 occurring at a distance of around 70 km have the highest contribution to the overall hazard results.

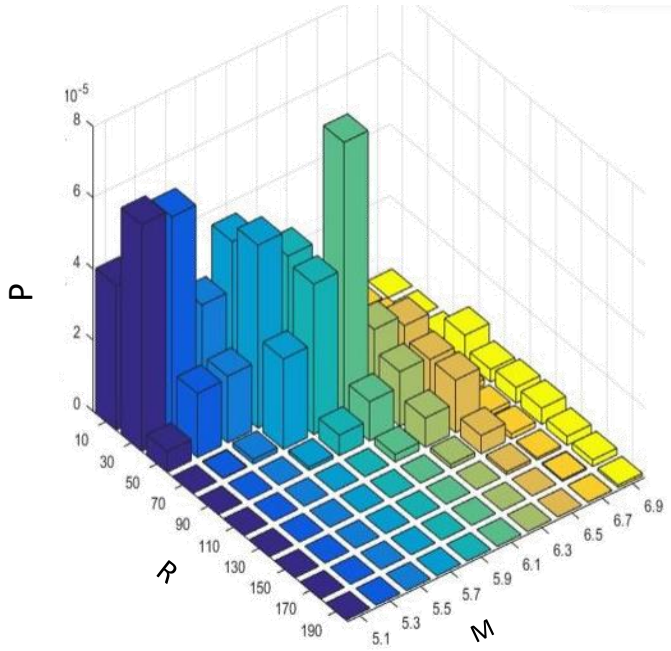

Figure 3.2: Disaggregation: magnitude and distance contribution to the overall seismic hazard for a return period of 475 years (Tebib, 2017)

PSHA enables specific rates of occurrence at a given site and for a chosen period of time to be targeted, thereby permitting seismic design guidelines and codes to be adapted to the nature of a structure and its exposure. Indeed, most engineering applications have a defined hazard level where ground motions are evaluated. Most building codes use return periods of 475 years (10% probability of exceedance over 50 years) for typical buildings (e.g., CBC 2019; Eurocode 8, EC8 2004) while for critical facilities (e.g., nuclear installations) a return period of 10,000 years is preferred (e.g., HSE, 2009).

3.1.3 Sources of uncertainties

Results from conducting PSHA can exhibit significant uncertainties emanating from the major components of the process. These results are used in earthquake engineering for the definition and selection of appropriate input motions for design and assessment of structures. These analyses are essential to study the structural vulnerability of a given structure, evaluate whether a structure conforms to the design codes and regulations and, finally, to assess the seismic risk. Hence, it is of upmost importance to identify these uncertainties, characterise them and incorporate them within PSHA.

Seismic Source Models (SSM)

The degree of uncertainty within SSMs usually depends on the seismicity of the region of interest. While for regions of high seismicity, data are generally available and seismic sources can be identified with some degree of confidence, for low-to-moderate rates the small numbers of events may lead to large uncertainties in accurately identifying seismic sources and rates, which can be challenging for developing SSM and for deriving site or region-specific GMPEs.

Earthquake catalogues need to be tested for completeness through identifying a minimal magnitude level above which all local events can be reliably detected and located. This magnitude of completeness (M_c) is related to the degree of understanding the seismicity of a region and is subjected to change in space and time (Wiemer and Wyss, 2000). This is a source of uncertainty that can affect recurrence models. While M_c is essential to truncate the recurrence model lower magnitude bound (Woessner and Wiemer, 2005), another high bound truncation parameter M_{max}, representing the

maximum regional magnitude, can affect significantly the overall hazard (e.g., Ake, 2008). The latter uncertainties can induce biases within the estimation of the Gutenberg-Richter parameters, the seismicity rate and so forth.

Such uncertainties could be reduced by enhancing instrumental seismic monitoring, investing in paleo-seismological surveys to characterise historical events and developing better faults and earthquake rate models (Erdik *et al.*, 2004; Chartier *et al.*, 2021).

Ground motion models (GMMs)

The SHA process includes a multitude of parameters that need to be adequately tuned to obtain more accurate results, such as GMMs. GMMs are key for the prediction of the ground motion at a site of interest. Abrahamson et al. (2019) investigated the impact of using non-ergodic GMM instead of ergodic GMM in seismic hazard assessment for California. Non-ergodic GMMs include systematic source, path and site effects, whereas in the ergodic assumption the average over all sources, paths and sites is taken into account. In other words, the ergodic assumption assumes that the ground motion variability observed over many stations from a multitude of earthquakes is comparable to the one for a single site-source combination over time (Anderson and Brune, 1999). Fully ergodic GMMs can be used for locations with similar tectonics, e.g., NGA-West1 GMMs (Power *et al.*, 2008). However, larger data sets (e.g., NGA-West2) clearly put in evidence the presence of strong regional differences to the ground-motion scaling (Bozorgnia *et al.*, 2014). Several researchers have shown that by adding specific information regarding source, path and site, the aleatory variability of the ground motion is smaller than the one in global models based on ergodic assumption (Atkinson, 2006; Anderson and Uchiyama, 2011; Lanzano *et al.*, 2017).

Abrahamson et al. (2019) demonstrate that using non-ergodic GMMs could potentially reduce the aleatory variability. However, this does not necessarily guarantee a reduction in seismic hazard levels but adopting non-ergodic GMMs will remove systematic over- or underestimation of the seismic hazard at different locations. It should be noted that the partially non-ergodic approach is becoming one of the

preferred methods for critical facilities and has been adopted for several major projects around the world, such as the Thyspunt Nuclear Power Plant in South Africa (Rodriguez-Marek *et al.*, 2014) and the Hanford Nuclear Production Complex in USA (PNNL, 2014).

Improvements have been made in constraining the site term within GMPEs (f_{site}). As f_{site} represents the average deviation at a site from the predictions of the GMPE, quantifying and reducing its epistemic uncertainty is possible if instruments are located at the site and sufficient measurements are made. Where this is not possible, numerous approaches have been developed to estimate the site term. f_{site} must account for the site amplification for a given input motion and should capture the effects of deep shear-wave velocity (V_s) and site attenuation known as κ_0. As most GMPEs are defined for standard rock ($V_{s30} \approx 800m/s$) and often for regions rather than site-specific, adjustments should be made. These adjustments are called $V_s - \kappa$ corrections, where the effect of the two latter parameters is captured by accounting for differences between the target site profile and the reference (host) profile used for the GMPE (e.g., (Boore and Joyner, 1997; Cotton et al., 2006). κ represents the full path attenuation including κ_0 (characterising site attenuation) and the quality factor Q (characterising path attenuation) (e.g., Campbell, 2003; Van Houtte et al., 2011).

Site characterisation

Irrespective of the scale of the study (e.g., national, regional or site-specific), site conditions and associated uncertainty have to be considered to select the appropriate GMPEs, infer suitable input motions, quantify the dynamic properties and their uncertainties (e.g., shear-wave velocity, damping curves, nonlinear shear modulus reduction curves), and obtain site amplification factors through site-response analyses.

Site properties such as V_s, stratigraphy and layering, depth of the bedrock, dynamic properties and the nature of geological formations (2D-3D effects) significantly influence the ground motion at the surface and the overall hazard (e.g., Makra et al., 2005; Bard, 2011; Barani et al., 2013; Lacave et al., 2014). The site response's sensitivity to these superficial geological layers is commonly called *site effects*. Site conditions are often characterised by one or more site attributes, known as proxies

(Pilz *et al.*, 2011; Trifunac, 2016). Proxies are useful indicators when developing or using GMMs (Douglas and Edwards, 2016; Kotha *et al.*, 2020), for the evaluation of local amplification (e.g., Bindi et al., 2014), the assessment of site-specific seismic hazard (e.g., Rathje et al., 2015; Aristizábal et al., 2022) and for defining soil classification within building regulations such as Eurocode 8 (EC8, 2004), NEHRP (2015) and the International Building Codes (e.g., IBC, 2012).

In the absence of sufficient data, proxies are often inferred and used to facilitate structural and design engineering applications. Site proxies can be better characterised by carrying out site and ground investigations (GIs). Direct measurements can be performed by conducting geological and geophysical surveys. Some examples of site proxies as well as some commonly associated methods of measurements are given in Table 3.1.

Cultrera et al. (2021) conducted a large survey amongst experts in several fields that aim to identify the most useful indicators for site characterisation at seismic stations. This survey was carried as an online questionnaire and had the purpose of identifying the most useful indicators to describe site effects, the methods and level of feasibility of measuring them as well as the costs of conducting those measurements. This work was complemented by Di Giulio et al. (2021) who defined a quality index for each proxy or combinations of proxies that describe the reliability of such measurements for site characterisation. Seven site indicators were identified to be of major importance and reliability in site characterisation; $f_0, V_s, V_{s30}, H_{bedrock}$(seismological, engineering), surface geology and soil class. While surface geology and soil class are easier and cheaper to infer, the other indicators are usually obtained from in-situ geophysical measurements and can be costly, especially when using invasive GI techniques. Furthermore, a rigorous site characterisation often requires using a combination of those proxies (e.g., Derras et al., 2017) which could result in higher GI costs with prolonged durations. While f_0 may be obtained through non-invasive geophysical prospections where costs of acquisition are relatively cheap, V_s profiles might require wider and longer deployments of non-invasive or invasive (e.g., drilling boreholes) measurement campaigns.

Table 3.1: Description of commonly used proxies and measurement methods

Proxies	Full name	Type of measurements
f_0	Fundamental resonance frequency	Horizontal-to-Vertical spectral ratio on earthquake (HVSR), on noise (HVN) Standard Spectral Ratio (SSR) to a reference station Ambiant Vibration Arrays acquisition (AVA)
STF	Site transfer function describing the amplification function	
Vs_{30}	Average shear-wave velocity over the top 30 meters	Inversion of surface-wave dispersion curves from Multi-channel Analysis of Surface Waves (MASW) Crosshole and Downhole measurements Seismic Cone Piezocone Test (SCPT)
$V_s(z)$	Shear-wave velocity profile as a function of depth z	
$H_{bedrock}$	Depth of seismological/engineering bedrock	HVSR – MASW- Boreholes- Microtremor Array Measurement (MAM)
Surface Geology	Lithological and geological information	Geological surveys – available cartography
Non-linear curves	Dynamic properties degradation (Damping ratio, D; Shear modulus, G)	Established models (Seed and Idriss, 1970; Darendeli, 2001) Standard Penetration test (SPT) Laboratory testing (SLAB)

The better the site characterisation, the lower the epistemic uncertainties affecting the ground motion analyses and, consequently, the overall seismic hazard estimates. Hence, wise choices should be made to prioritise measurements for data collection in light of the available budget and potential investments.

3.2 Decisions and regulations within SHA for nuclear powerplants in the UK

Designing and building new structures require considering external hazards (e.g., flood, wind and earthquakes) to ensure acceptable levels of safety for the public and to avoid damage and more importantly, avoid collapse. When it comes to earthquake risk, seismic hazard analysts and civil engineers work toward minimising future damage by performing advanced analyses based on guidelines and specific procedures. Whether it is for typical buildings, bridges or sensitive facilities such as nuclear powerplants (NPPs), seismic hazard analysts and engineers need to follow and be updated on safety requirements to ensure an acceptable level of safety. In both stages of hazard analysis and seismic design, several steps and procedures should be respected to ensure conformity to current provisions. They require expert judgments to make the optimal decisions.

In this section, we present the different steps and requirements when performing seismic hazard analysis as input for the structural design of NPPs in the UK.

3.2.1 Overview of the process of inferring an optimal seismic design decision

Based on the above literature review, we created an influence diagram that gives an overview on the variables that influence decisions regarding data collection, assigning weights in the logic tree for PSHA calculations and the choice of a seismic design at an acceptable level of safety (Figure 3.3). The influence diagram was built using GenIE Modeller, a software package that enables building graphical decision-theoretic models. This software not only allows the probabilistic dependence between random variables and its updating through Bayesian networks to be modelled through influence diagrams but also performs decision analyses in light of uncertainties and obtaining new information.

The diagram regroups most of the required inputs within PSHA calculations along with the essential decisions within the process for seismic hazard analysts and ultimately, for structural engineers in inferring a design spectrum for a new-build or to retrofit an existing structure. Variables in pink oval nodes represent the main inputs to

PSHA. This diagram gives a rough idea of the number of uncertain inputs. While most models carry inherent aleatory variability, one must also properly capture the epistemic uncertainty to avoid biased estimates of the mean hazard.

VoI analysis (yellow oval node) has been included as a preliminary analysis to decide on data collection. Indeed, assessing VoI is best performed in the early stages of SHA. These early stages include conducting a "gap analysis". The gap analysis should be directed toward identifying the main sources of uncertainties and deciding, supported by peer review and expert elicitation, on the necessity of data collection to reduce targeted uncertainties. There is a trade-off between the time and effort necessary to compile a relevant, reliable and detailed database, and the degree of uncertainty that the analyst should take into consideration at each step of the process. This is where assessing the VoI is believed to be highly valuable.

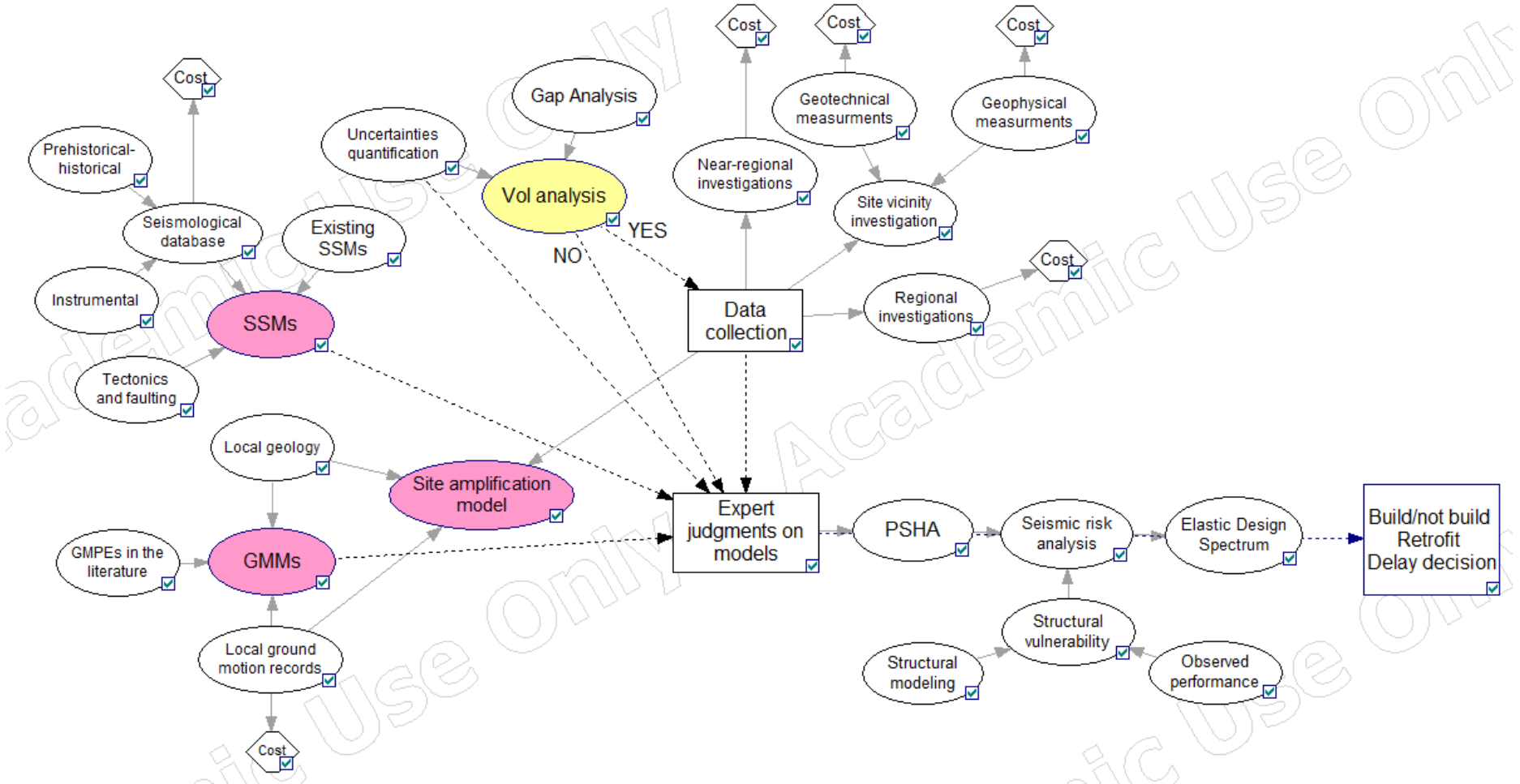

Figure 3.3: Influence diagram describing the general framework of inferring a seismic design. Oval nodes represent uncertain variables, hexagon nodes the cost of evaluating some variable and rectangle nodes are decisions made based on the values of variables. Variables in pink oval nodes represent the main inputs to PSHA. Each parameter that must be considered in a variable or a decision node is linked to it with an arrow.

3.2.2 SHA best practice for Nuclear Powerplants in the UK

As of December 2022, NPPs in the UK provide about 15% of its electricity (Matthew, 2022). Currently, five NPPs are operating throughout the UK by EDF Energy. To develop new nuclear facilities, there are regulatory requirements that ensure that the utility operators undertake the necessary assessments for external hazards, including seismic hazard.

PSHA has been used in the UK for the evaluation of seismic hazard for the sites of NPPs since the 1980s (e.g., SHWP, 2001; Musson, 2014; Tromans et al., 2019). Along with these assessments, safety cases must be put in place and are reviewed by the Office for Nuclear Regulation (ONR). The ONR also makes sure that these assessments are robust, follow good practice and protect the general public. The ONR is a non-prescriptive regulator, meaning that the utility operator is responsible for developing the safety cases adequately.

When assessing seismic hazard and establishing safety cases, several actors take part in the process. Generally, three major parties are involved:

- The nuclear regulator: ONR
- The seismic hazard team: responsible for performing the PSHA
- The client and stakeholders: utility operators

Figure 3.4 provides an overview of the interactions between and within the involved parties.

The seismic hazard team forms a project management team that will select and supervise the technical delivery. The project management team is mainly responsible for defining and refining the several components that constitute the SHAs and ultimately the hazard calculations. Throughout the project, selected members of the ONR, the seismic hazard team and the utility operator meet regularly to update each other on the decisions made. These decisions often concern the type of approach chosen, the type of parameters considered, their current uncertainties and whether there is a need for more GIs. During these discussions, often called workshops (Aldama-Bustos *et al.*, 2019), selected ONR members and client representatives are

invited as observers, along with some experts chosen by the ONR to assure the compliance of the stakeholders and the client to the regulatory assurance and project objectives.

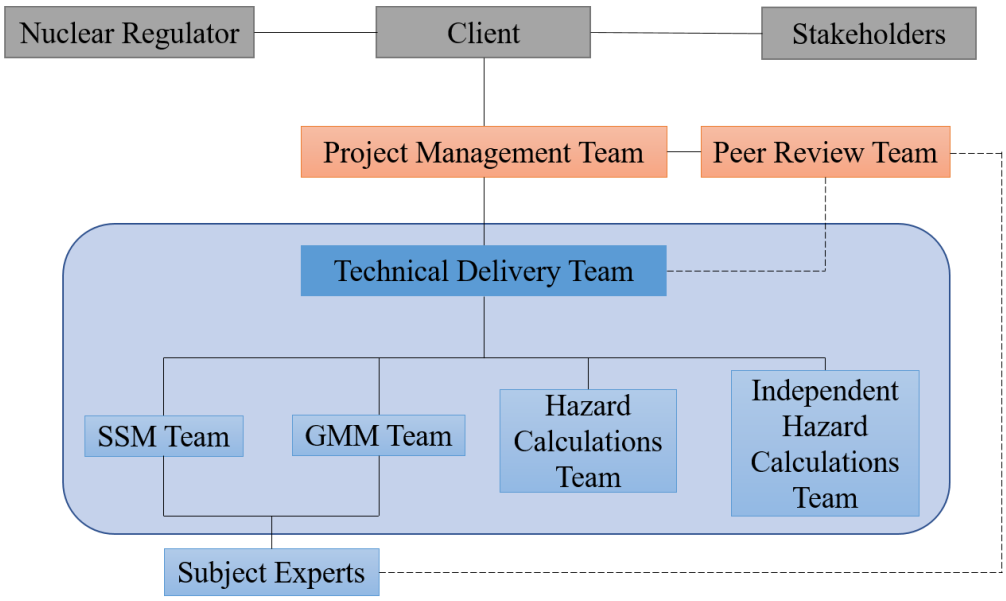

Figure 3.4: Overview representation of the different interactions between and within parties involved in SHA and in establishing safety cases for the design of nuclear powerplants in the UK. Solid lines represent direct formal interactions, dashed lines represent close interactions between the various groups (Modified from Aldama-Bustos et al., 2019)

There are several phases in the overall project. Aldama-Bustos et al. (2019) summarise the process in three major phases based on NUREG-2117 guidelines (USNRC, 2012):

Phase 1: Study definition stage and gap analysis

The first phase consists of organising meetings between the seismic hazard team and the client where a review of previous studies in the region of interest is conducted and a gap analysis is thoroughly evaluated. Initially, a preliminary gap analysis is performed by the seismic hazard analysts and the client experts separately. Then, these analyses are shared and discussed in a common meeting. At this stage, decisions regarding collecting further data are taken on behalf of both parties. In most cases, the latter decisions contribute to the regulatory assurance defined in NUREG-2117 (USNRC, 2012). The team of analysts refer to several guidelines such as TAG-13 issued by the ONR (2017), Safety Standard Series nº SSG-9 by the International

Atomic Energy Agency (2022) and United States Nuclear Regulatory Commission reports (USNRC, 2007).

Finally, the project methodology is discussed and agreed upon. This involves recruitment of external subject experts, creating a preliminary database and identifying potential future measurements for collecting additional data. The gap analysis phase is usually continuously assessed throughout the whole project until agreeing on the final hazard calculations.

Phase 2: Hazard calculations

A peer review team is selected depending on the budget and the followed guidelines, such as the Senior Seismic Hazard Analysis Committee (SSHAC) guidelines (e.g., Budnitz et al., 1997), which provides guidance when multi-disciplinary experts are involved. The review team is then invited to take part in meetings held with the client's representative and the technical delivery team. Other experts are also solicited to share their insights within the decision-making process on the collection of more data as well as to evaluate their relevance and supervise their interpretation and implementation of prior knowledge.

Following that, a preliminary PSHA is performed using preliminary SSMs and GMMs (e.g., Reiter, 1990; McGuire, 2004). Then, members of the technical delivery team discuss the hazard calculation results and agree on updated SSMs and GMMs based on the available data and expert judgments. Consequently, the decision on the chosen model's implementation is presented to the peer review team and their comments are considered. Finally, the final hazard calculations are performed and results are reported to the client and the ONR.

Phase 3: Safety Case support and submission

This phase mainly focuses on providing support and guidance to the client on the Safety Case submission to the ONR.

As evidenced by Aldama-Bustos et al. (2019) and within most of official guidelines, decisions need to be made at every step and in every phase of the process. Whether PSHA is site-specific or performed at larger scales, following guidelines with caution

is crucial as the associated outputs (i.e., HCs, UHSs) serve primarily as a basis for civil engineers to design seismically-sustainable structures.

3.3 From seismic hazard analysts to civil engineers

The common purpose of SHAs is to provide a solid evaluation of the hazard for seismic risk assessment representing a critical consideration for structural design and assessment. Designing a structure to withstand seismic activity is a crucial step to ensure an acceptable level of safety, considering the seismic hazard in the region or site of interest. The basis of seismic design is to analyse the response of a structure to the level of expected ground motions indicated by PSHA outputs at a specific design return period. Moreover, seismic design enhances the safety of civilians by targeting low probabilities of damage over relatively long periods. Finally, a rigorous seismic design works towards preventing economic losses and ensuring operational continuity during and after construction.

3.3.1 General comments on structural seismic design and building codes

It is on the overall seismic risk evaluation and not only seismic hazard results that structural engineering, insurance premiums and other policies are based (Wang, 2011). All decisions are made under a certain degree of uncertainty.

Seismic risk is quantified by four parameters (Wang, 2009):

- Probabilities
- Level of severity (physical measurements)
- Spatial measurements (where)
- Temporal measurements (when and how often).

The level of seismic risk depends on the interaction between the seismic hazard and the vulnerability of a structure (Equation 3.6). Analytically, the seismic risk is evaluated by convolving the seismic hazard (probability of exceeding a certain level of ground motions) with the fragility curves that expresses the probability of

occurrence of a given level of damage given a range of load levels, and including the potential economic, social and environmental consequences.

$$Seismic\ risk = Seismic\ hazard\ \times Vulnerability \tag{3.6}$$

The level of damage is related to the definition of a suite of Engineering Demand Parameters (EDPs) for the different structural and non-structural components of the structure such as the Damage Index (DI), the Roof Drift Ratio (RDR) and the inter-storey drift ratio (IDR). Design codes such as EC8 (2004) and FEMA 356 (2000) provide definitions for such EDPs.

For example, Douglas et al. (2013) gives useful inputs for France when assessing fragility curves and determining design values.

Building codes

Seismic hazard maps are often a basis for the development of building codes, such as the International Building Code (IBC), Eurocode 8 (EC8, 2004) and NTC18 Italian code (NTC, 2018). Technical guidance and requirements for seismic design for constructions in Europe, for example, are explained with a high-level detail in Eurocodes. EC8 is now widely adopted and is an advanced update of previous Eurocodes, which ensures a coverage of seismic design requirements for several type of structures and near-surface geologies.

Current building codes, such as EC8, combine parameters from seismic zonation maps and the classification of near-surface geology to construct an elastic design response spectrum at any site. Following EC8 guidelines consists of identifying the structure's location, the level of seismicity (i.e., type1=high, type2=low) and the soil classification. Depending on these inputs, EC8 proposes an elastic design spectrum. In fact, the structure's design is defined by its dynamic properties and assessed by its response to a particular seismic load. Figure 3.5 shows EC8's elastic spectra for both seismicity level types and different soil classifications (A: rock; B: very dense sand or gravel, very stiff clay; C: dense sand or gravel, stiff clay; D: loose-to-medium cohesionless soil, soft-to-firm cohesive soil; E: soil profiles with a surface layer of alluvium of thickness 5 to 20m).

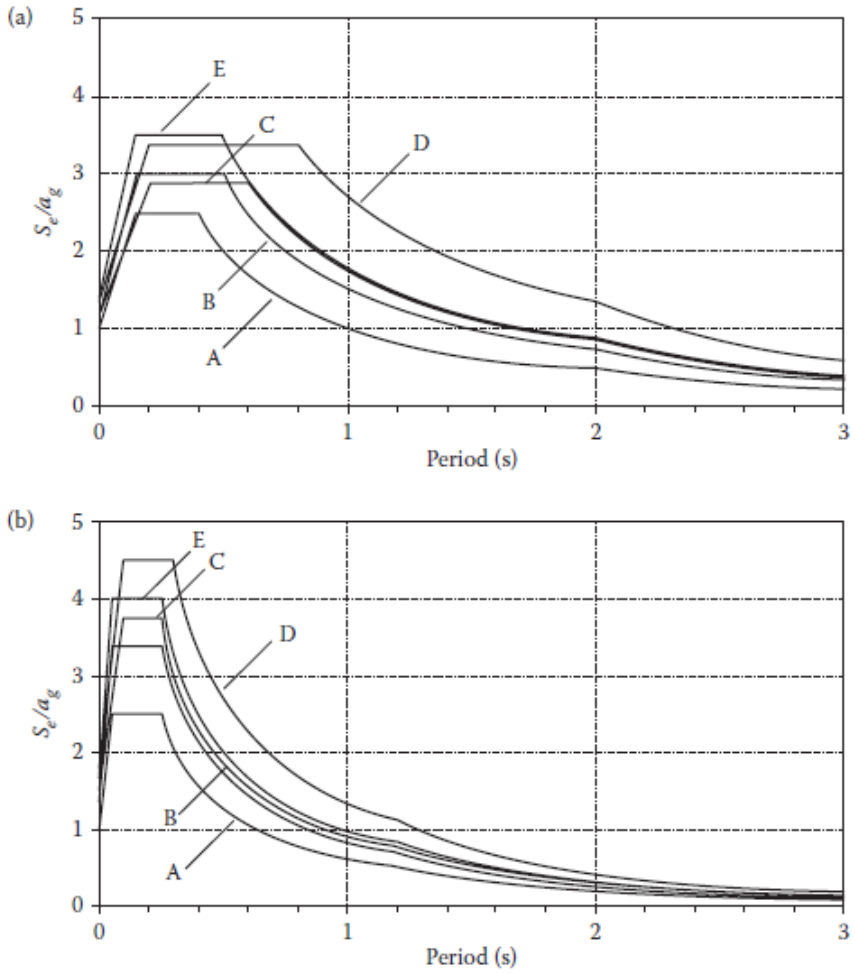

Figure 3.5: EC8 5% damped elastic spectra. a) Type1 b) Type 2. a_g is the PGA for soil class A (from Williams, 2016)

The noticeable difference between the two types is that spectral amplifications in the "high seismicity" spectra (i.e., type 1) tends to occur at longer periods and over a wider period range, compared to the "low-to-moderate seismicity" spectra (i.e., type 2). Moreover, the soil class affects the level of spectral amplification and the spectral period at which it occurs. Usually but not always, a denser soil gives rise to lower amplifications.

Limitations and alternatives

Although seismic codes such as EC8 offer useful guidance for earthquake-resistant design for many structures, they typically do not provide options for choosing the level of risk. EC8 design spectra are anchored to a PGA value obtained from the hazard

curve at a predefined probability of exceedance, and whose shape depends on the local site conditions. This approach may lead to a non-uniform level of risk depending on the structure's location (Gkimprixis *et al.*, 2019). An approach that can be used as an alternative to tackle this issue is the risk-targeting approach (Kennedy, 2011). This structure-specific approach aims at designing a structure based on a predefined acceptable and controlled level of risk called the mean annual frequency (MAF) of collapse. Iterations are performed by fixing the MAF and testing trials of design PGAs using the hazard curve to determine the appropriate design acceleration for the building to withstand.

We can cite other cases where building codes might not be sufficient (Bommer and Stafford, 2016):

- Projects on a site with deep and/or soft soil. The effect of near-surface geology on the ground motion is then being captured in the basic soil classification provided or in the spectral shapes.
- Projects on a site located in the vicinity of active faults where directivity effects need to be considered.

In these cases, a site-specific seismic hazard analysis is required supported by expert judgements to infer decisions regarding the different models and parameters within SHA as well as to incorporate associated uncertainties.

3.3.2 Seismic design for nuclear powerplants in the UK

Critical facilities such as NPPs are subjected to even tighter constraints and requirements to ensure a high-level of safety. Regulators and regulation practices differ depending on the country and whether regulators are part of the government (e.g., Switzerland and Japan) or independent such as in France, USA and UK (Bredimas and Nuttall, 2008).

In the UK, the Nuclear Safety Technical Assessment Guides by the ONR provides useful guidelines to meet the Safety Assessment Principles (SAP). These guidelines serve as the basis for seismic analysts and civil engineering decisions in NPPs design. The latest update, known as TAG-13, states that "*The design basis ground motion for*

a site, often referred as Design Basis Earthquake (DBE), is the fundamental input for deriving structural design loads and assessing plant effects as part of the Design Basis Analysis (DBA)". A sufficient level of conservatism should be incorporated in the DBE not just for one or two elements of the seismic analysis and design process but for the process as a whole.

The main steps to determine the DBE are as follows:

Step 1: UHS should be derived from site-specific PSHAs, including propagating uncertainties to take into account the overall hazard uncertainty.

Step 2: Conservatism should be included by considering the UHS at a target annual frequency of exceedance (AFoE) of 10^{-4}, equivalent to a return period of 10,000 years. Moreover, TAG-13 states that a DBE should envelop the 84^{th} percentile (i.e., one standard deviation above the mean UHS). The conservatively defined site-specific ground motion hazard is then viewed as the seismic demand spectrum.

Step 3: A DBE is then developed and considered as the seismic design spectrum. DBE is an extremely useful tool in the design process especially in the concept design stage. Within the UK and several other countries, the DBE is often represented as a standardised response spectral shape scaled to a PGA value from a PSHA that is indicative of the site of interest. The DBE provides constraints on potential design solutions.

Defining and reviewing the suitability of the DBE often requires appropriate expert judgments. The ONR makes sure that it meets or exceeds the seismic demand guaranteeing that the NPP is designed to resist a minimum level of strong ground motion. The minimum level of requirements in the UK is defined by the International Atomic Energy Agency (IAEA) and the Western European Nuclear Regulators' Association (WENRA) as the "*horizontal free-field standardized response spectrum anchored to a PGA of 0.25g*".

Seismic hazard analysts and clients work together in selecting the appropriate seismic design. There are cases where the DBE respects the requirements at PGA but over- or underestimates the UHS at other frequency ranges. We can find an example of tackling

this issue in the work of Villani *et al.* (2020) where the design response spectrum is tuned to an acceptable level of safety using alternative regulatory guides. Another possible strategy is to reduce the uncertainties within the inputs to the PSHA. This aims at better characterising the overall hazard and, thereby, increasing the likelihood of falling within the acceptable level of safety. Reducing uncertainties may not guarantee this latter goal but using VoI might assess the feasibility of such an approach. In case of an under-conservative design spectrum, a higher DBE than the initial one is often considered. This can induce higher design costs. It can be worth computing the VoI to assess whether reducing uncertainties might lead to considering a lower design level given the level of seismic hazard at the site.

3.4 The need for VoI in the industry - Interviews

3.4.1 Scope and involved parties

Understanding current nuclear industry approaches regarding data collection for seismic hazard assessment provides the context in which the decision tools that we develop will be applied. For this reason, semi-structured interviews were conducted with seismic hazard analysts (specifically a team from Jacobs) and nuclear facilities owners (specifically a high-level representative from EDF Energy) in summer 2020. The interviews aimed to assess the current state of practice for data collection within SHAs for UK nuclear facilities and to identify gaps within the justification and prioritisation of data collection, so as to tailor decision-making tools to bring maximum value to industry.

Table 3.2 provides details about both groups as well as the main purpose of the interviews. The outcomes of both interviews have been summarised and partly transcribed in Appendix A of this thesis and were considered when building the overall approach to assess the Value of Information.

Table 3.2: Selected interviews participants and main purpose for interview

Group	Role	Company	Interview purpose	Number of participants
Consultant	Seismic hazard analysts	Jacobs	Insights on strategies for prioritising and justifying data collection	3
Client	Utility operator	EDF Energy	Insight on the decision-making process for financing data collection	1

Interview questions were carefully selected to ensure that the interview purpose was achieved. Some of the inquiries that were examined include:

- Which parameters most control PSHA outputs?
- What criteria are used to determine when and why new data collection should be funded?
- How is the prioritisation of data acquisitions carried out?

The interviews outcomes support what has been mentioned in the previous sections. After analysing the outputs of each interview, we find a great consistency between the two groups, i.e., the client and the consultant.

3.4.2 Main interview outputs

From our interviews, we provide a summary of insights on the identification of key SHA parameters, the gap analysis stage and the steps leading to making the decision on the collection of additional data. Finally, we share some observations from the participants about the lack of established decision-making strategies and the challenges that can be encountered in developing approved ones.

Insights into key parameters in PSHA

Identifying the parameters that play a major role in PSHA helps refine the scope of this PhD to focus on assessing their value when it comes to decision-making. In both interviews, there was a focus on the choice of GMMs and the estimation of their weights within the logic tree. The seismic hazard team of Jacobs highlighted the

contribution of site effects to the overall hazard and insisted on providing a well characterisation of site properties such as V_s profiles, the bedrock V_s and depth as well as the high-frequency attenuation parameter κ_0. Moreover, uncertainties should also be well characterised and incorporated within the calculations through Monte Carlo simulations (e.g., within soil-response analyses) and through the logic tree for the estimation of the mean hazard and associated percentiles.

Gap analysis and data collection

The answers to questions concerning gap analysis indicate that it is a continuous process throughout the PSHA study. Where data are usually available from previous projects, there is always a will to upgrade the PSHA to ensure that the safety requirements are met as well as to be up-to-date with modern practice. Both parties confirmed that new NPPs are generally planned to be located within the vicinity of existing facilities where GIs have been conducted in the past. Nevertheless, additional GIs are often required as guidelines and safety standards are periodically reviewed and updated. Common practice first consists of the client reviewing available data and identifying gaps. Following that, a discussion within the consultant's team (i.e., the seismic hazard team) takes place to decide whether additional measurements are needed. This transitional phase is of great interest within this research, where VoI calculations might represent a key support tool that aims to facilitate such decisions.

Decision-making process for data collection

Both interviews confirmed that the decision-making process for collecting new data is not always straightforward. Published guidelines are usually used to decide and prioritise the collection of data. These guidelines can also be a support in choosing the most cost-effective methods for data acquisition. Jacobs added that cost-benefit strategies are being developed but they believe that regulatory assurance is the primary concern. In addition, the EDF Energy representative made clear that the client usually relies on the consultant's and other expert opinions towards making sure modern practices are adopted and safety requirements are applied and adhered to.

3.4.3 Conclusion

Insights from interviews identify a clear gap when it comes to strategies and workflows adopted to justify, objectively and numerically, decisions on gathering a specific piece of information. It is currently difficult to estimate beforehand whether a particular data acquisition will be beneficial to the PSHA and, eventually, to the design process. Both of these processes are full of uncertainties that should be incorporated within the decision-making process. Although guidelines for both of these important steps exist and are subject to continuous updates and improvements, there is a prominent decision-making component relying on expert judgments, beliefs, experiences and insights. Some of these decisions could be hard to justify in light of uncertainties and budget limitations, such as decisions regarding data collection. This highlights the need to develop a robust approach integrating both modern practice and safety guidelines to quantify the VoI within SHA and the seismic design stage and to reduce the decision-making burden.

4 Assessing VoI in a seismic design application: single uncertain parameter, Vs_{30}

This chapter presents an extended version of the following journal article:

Tebib, H., Douglas, J. and Roberts, J. J. (2023) 'Using the value of information to decide when to collect additional data on near-surface site conditions', Soil Dynamics and Earthquake Engineering, 165, p. 107654. doi: 10.1016/j.soildyn.2022.107654

4.1 Introduction

In this chapter, we define a case study aimed at developing a methodology for calculating the value of information (VoI). Specifically, we focus on evaluating the VoI of a single parameter commonly used in site-response analysis, with the ultimate goal of informing seismic design decisions for a specific building. This relatively simple case study using real data to compute outcomes, supports the elaboration of a general method for VoI calculations that can be applied in the fields of soil-response analysis, seismic hazard assessment and seismic design. The goal is to validate the method and study its sensitivity to various input parameters. Different VoI definitions are considered, from the least to the most realistic.

First, we define a case study aiming to determine an appropriate seismic design for a hypothetical building. This particular building has been chosen, for convenience, from the work of Gkimprixis et al. (2020). Their study provides some of the inputs needed to run the analysis of VoI, including fragility curves and probabilities of damage as well as different construction costs and expected losses from damage. The main goal of this case study is to build a method for VoI assessment for one of the uncertain parameters that plays a major role in site characterisation and, hence, site-response analysis: the average shear-wave velocity in the top 30m, Vs_{30} (Cultrera *et al.*, 2021).

Second, we apply the theory for VoI in the case of discrete possible values for the chosen parameter to compute the Expected Value of Perfect and Imperfect Information

(EVPI or EVII, respectively). We highlight that in this field of study, obtaining perfect information is rare and thus, computing the EVPI has the purpose of training and partially validating the developed method. The EVII is the most appropriate and realistic value to compute.

Third, we conduct a comprehensive analysis of the results, leveraging sensitivity analyses to better understand the impact of various factors on VoI. Specifically, we explore how changes in prior probabilities for our key parameter of interest and other relevant factors within the methodology may affect the overall VoI.

The same steps are performed for a more realistic case where Vs_{30} is defined by continuous possible values. For this situation, Monte Carlo simulations and approximations are used to infer EVPI and EVII.

Finally, we proceed to interpreting the results and drawing out key conclusions.

4.2 Overview

We aim to determine an appropriate seismic design for a hypothetical four-storey three-bay reinforced concrete frame building located in Patras, Greece. The building is symmetrical in plan and elevation, with span length and column height equal to 5 m and 3 m, respectively. In Figure 4.1, the location and the 3D building design model are shown on the seismic hazard map indicating peak ground acceleration (PGA) values that correspond to 10% probability of exceedance in 50 years, obtained from the 2013 European Seismic Hazard Model (ESHM13, Giardini, 2013; Woessner et al., 2015). The seismic design should ensure an acceptable level of safety regarding the seismic hazard along with respecting a reasonable budget. The decision-maker should take into account these variables with a certain level of scrutiny as both tangible and intangible losses can represent a risk.

This building was designed for different levels of design peak ground acceleration, PGA_d, according to Eurocodes (CEN, 2004; EC8, 2004) and using the Type 1 horizontal design acceleration spectrum of Eurocode 8 (Gkimprixis *et al.*, 2020). The seismic design for the different PGA_d levels was carried out by Gkimprixis *et al.*, 2020 using an importance class II and a medium ductility class that corresponds to a

behaviour factor q=3.9 (EC8-1-5.5.5.5) along with a class B assumed for the soil conditions and a 5 % damping ratio. The fundamental vibration periods of the buildings are 0.36s, 0.32s, 0.25s and 0.20s for PGA_d of 0.0 g, 0.1 g, 0.3 g and 0.5 g, respectively.

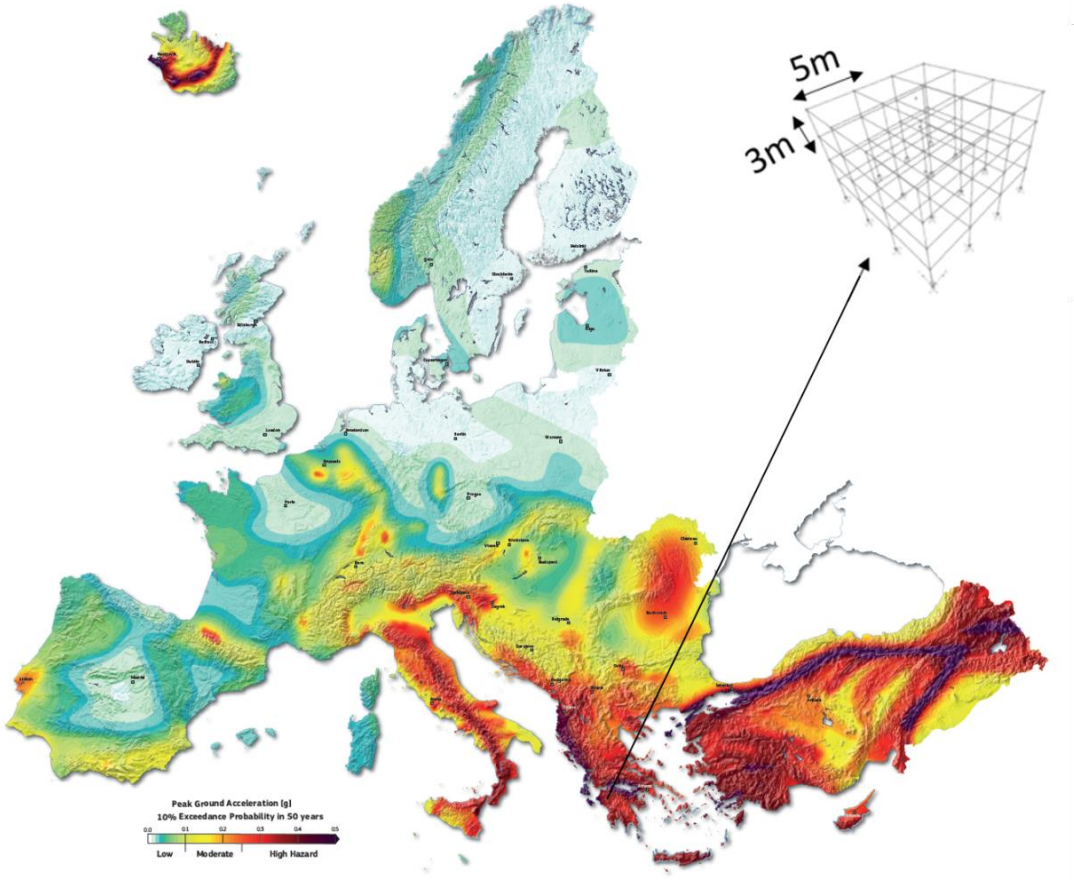

Figure 4.1: ESHM13 hazard map of peak ground acceleration [PGA] for 10% probability of exceedance in 50 years (average return period of 475 years) (Danciu et al., 2021). The 3D building design model is shown inset, and its location in Patras indicated by the black arrow (Gkimprixis et al., 2020)

In order to determine the PGA level at which the building should be designed, site-specific seismic hazard assessment should be performed. Depending on the type of building, a return period is fixed and hazard curves are produced to retrieve the appropriate design PGA (e.g., via a risk-targeted approach).

We assume that available data are all reliable and that the only uncertain variable is the shear-wave velocity in the first 30m at the site location, Vs_{30}. Vs_{30} is commonly

used in ground motion prediction equations (GMPE) to predict ground motion at the surface as well as to estimate the ground motion amplification compared to a reference rock site.

At this stage, the decision-makers face an important dilemma:

1- Choose a particular seismic design based on the available information. This involves taking the risk of choosing: (a) a higher, and more costly seismic design than needed; or (b) a lower and less-resistant design where the bedrock hazard and the site amplification could result in building damage or even total collapse.
2- Conduct geophysical/geotechnical tests to decrease the uncertainties on Vs_{30}. This will reduce the risk of choosing an inappropriate seismic design. However, the test will have a cost that depends on their type, the company hired to perform them and the price of buying or renting the testing equipment.

In some cases, decisions might not have high stakes or large risks. In others, consequences can be significant. In this study case, people's safety is the main concern. Damage or collapse of the building might result in human losses (deaths or injuries).

We aim to build a method for VoI calculation that includes all currently available information and estimates the losses for each possible value of the uncertain parameter, each test outcome and each decision.

4.3 Scenario 1: VoI for a discrete uncertain variable

In our first scenario, the uncertain parameter (Vs_{30}) is assumed to have discrete values. This simplification is useful to understand the process of building the VoI method as well as to have a clear comprehension of the impact of different variables on the results.

In this scenario, Vs_{30} is assumed to equal either V_1 or V_2. Available data and expert knowledge will help assign prior probabilities to V_1 and V_2. We clarify that this is a simple case where the parameter of interest has binary possible values. A second scenario (section 4.4) considers the more realistic case of continuous distributions.

4.3.1 Methodology

While collecting information can be useful for reducing uncertainties, its primary purpose is to support more informed decision-making. There is no intrinsic monetary value of reducing uncertainties. However, there is a need to formulate the main goals of reducing uncertainties such as to avoid losses by reducing the risk of making the wrong decision. In this case, the decision is to apply the appropriate seismic design.

Setting a clear framework and including all available data is necessary to increase the reliability of results. This ensures a more solid analysis and interpretations to choose the adequate investigation method while sticking to a reasonable time/money budget.

Inputs and parameters

The purpose of the simplified site-response analysis performed in this study is to estimate the resulting ground motion at a theoretical site, in terms of peak ground acceleration (PGA on soil), based on the PGA on a reference outcropping rock and the site-amplification factor. The obtained PGA on soil would then constitute an indicator of the ground motion level to which the building should be seismically designed to resist.

According to previous studies (Trainor-Guitton *et al.*, 2011) and the literature review presented in Chapter 2, the three components involved in determining VoI are:

1- Prior uncertainties on the parameters involved to make the decision
2- Reliability of available data
3- Expressing the outcomes in terms of value

In our present study, fragility curves, costs and site-response parameters beside Vs_{30} are assumed to be reliable. We will focus on components 1 and 3 to estimate VoI.

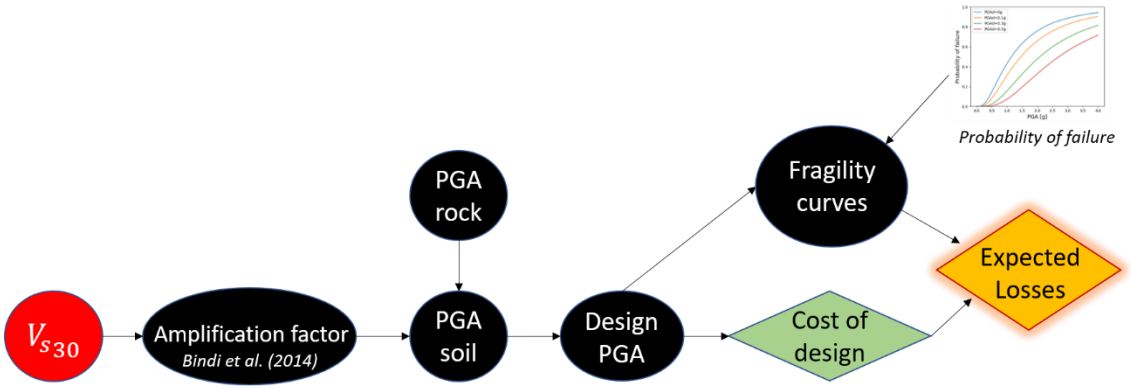

Figure 4.2: Framework for site-response analysis and the estimation of the expected costs and losses. Black circular nodes represent known parameters, and the red circular node (Vs_{30}) is the uncertain parameter. The lozenge green node is the initial cost of design, the yellow node is the outcome (estimated costs and losses).

Figure 4.2 shows an influence diagram that summarises the parameters that are computed and/or used to estimate the appropriate PGA_d. The components are:

- Vs_{30}: Average shear-wave velocity in the first 30m. This is considered here to fully represent the site characterisation.
- Site amplification factor, Fs. (Equation 4.1)
- PGA on rock, PGA_r: Peak ground acceleration at a reference rock site
- Resulting PGA on soil: Simple multiplication of PGA on rock and the amplification factor
- Design PGA: PGA to which the building is seismically designed
- Expected losses: Considered as the outcomes for VoI calculations and detailed in the next subsection

The hazard curves from the ESHM13, associated to the case study location, were used to retrieve the expected losses and the PGA on reference rock (PGA_r). The PGA_r is estimated at 0.43g, which has been estimated for a 50-year lifetime and a probability of exceedance of 10% (corresponding to a return period of 475 years). This value is for a rock site with Vs_{30}>800 m/s.

In this case study, site-response analysis is simplified by neglecting non-linear effects and by assuming that Vs_{30} completely controls the near-surface site amplification. Thereafter, the frequency-independent amplification factor F_s of the site is assumed to be (Bindi *et al.*, 2014):

$$F_s = \gamma \log_{10}(\frac{Vs_{30}}{V_{ref}}) \quad (4.1)$$

where V_{ref} is fixed to 800 m/s and γ= -0.3019 (associated to PGA values)

Let's assume that $V_1<V_2$, then the dilemma becomes the following:

Before additional data collection

1- Choose *Design1*, associated to V_1. The lower the Vs_{30}, the higher the site amplification of the ground motion on bedrock (provided all other variables are kept the same and AF computed following Equation 4.1), and so, the higher the structural performance of the building, which needs to be in terms of strength (i.e, design PGA) as described later and behaviour factor q as defined in Eurocode 8 (EC8, 2004) and ductile response (i.e., extent of inelastic deformation). As such, *Design1* has a better structural performance than *Design2*, associated to V_2, assuming that $V_1<V_2$ and that the AF is computed following Equation 4.1. If the site Vs_{30} is V_2, the building is likely to be "over-designed" or, in other words, "unnecessarily resistant" for the actual seismic hazard. There are no drawbacks in terms of safety in over-estimating a building's seismic design. However, this design will cost more than a cost-optimised design, due to additional materials and construction time.

2- Choose *Design2*, associated to V_2. If the Vs_{30} is V_1, we would be underestimating the seismic hazard. This will result in a higher risk of building damage. Damage can cause injuries and fatalities as well as requiring repair or re-construction. Depending on the level of injury and the situation, those harmed (or their family in case of death), are financially compensated. It is ethically difficult to put a price on a human life, but the cost of possible compensation could be considered when computing the expected losses when taking the wrong decision, i.e., using a seismic design that underestimates the seismic hazard.

After additional data collection

Conduct tests, perfect or imperfect, to know the value of Vs_{30} more accurately (i.e., with lower uncertainties) for the case of an imperfect test or know it

exactly in the case of a perfect test. A (near) perfect test could be considered as crosshole or downhole tests, where a geophone is lowered into a borehole and shots are fired to estimate the V_s within the vertical soil column under the site. An imperfect test would be geophysical survey techniques such as Multi-Channel Analysis of Surface Waves (MASW) or ambient vibration measurements.

These three different decisions need to be considered when assessing VoI in this first scenario.

Prior probabilities

Prior probabilities are defined for each possible value of Vs_{30}.The prior probability, *p*, is the probability that V_1 is believed to be the true Vs_{30} at the site. Similarly, the prior probability *1-p* is the probability that V_2 is believed to be the true Vs_{30}. These probabilities are often inferred by the seismic analyst from available data or from expert elicitation. In some cases, the available data coming from geological studies, past prospections or Vs_{30} values in nearby locations may suggest a tendency towards choosing V_1 or V_2 as the Vs_{30} of the site of interest. As previously mentioned in Chapter 2, Empirical Bayes methods can also be an effective way to infer prior probabilities.

For either of the three decisions, prior probabilities remain effective. In fact, when making the decision to apply *Design1* or *Design2* without conducting the perfect test, there is a probability of being wrong. Similarly, when deciding to conduct a test, the posterior probability depends on the prior probabilities.

Design PGA

Several approaches can be used to infer the appropriate seismic design such as design based on “uniform hazard” which are the basis of many modern seismic codes (e.g., Eurocode 8). This spectrum in anchored to a PGA that corresponds to a specific probability of exceedance (e.g., 10% probability of exceedance in 50 years) from the hazard spectrum at the site’s building. This PGA is henceforth referred to as the design PGA, PGA_d.

Another approach is to consider a risk-targeted approach that is specific to the considered type of building. This approach is an iterative process to obtain the design PGA (PGA_d) by expressing the mean annual frequency (MAF) of collapse λ_f that will secure the building to an acceptable and controllable risk level (Kennedy, 2011).

$$\lambda_f(PGA_d) = \int P(C|IM).|dH(IM)| \tag{4.2}$$

where $P(C|IM)$ represents the probability of collapse given an intensity IM and H is the hazard curve obtained from PSHA.

In this case study, we assume that one of these methods is used to infer the design PGA on rock. The design PGA on soil is then simply the retrieved PGA on rock multiplied by the site amplification factor.

Probability of failure

The fragility curves for different PGA_d are obtained from Gkimprixis et al. (2020). The curves were derived from Incremental Dynamic Analysis (IDA) (Vamvatsikos and Cornell, 2002) by considering explicitly structural (S), non-structural drift-sensitive (N/D) and non-structural acceleration-sensitive (N/A) components at each storey. They are assumed to have a lognormal distribution and IDA allows the estimation of the median capacity $C_{50\%}$ and the composite logarithmic standard deviation β for each PGA_d. Gkimprixis et al. (2020) consider several limit states of damage defined by a specific Engineering Demand Parameter (EDP), in this case the inter-story drifts (ISDs). The damage limitation, or limit state, is when the ISD levels are above a certain threshold. The fragility curves considered in this study indicate the probability of the limit state of global collapse for every possible value of PGA resulting from a future earthquake.

In Figure 4.3, the fragility curves give the probability of total collapse of the building designed for a specific PGA_d at a given seismic intensity measure, here PGA. In order to use these fragility curves for a more extensive set of PGA_d, the median and β have been interpolated to produce the probabilities of failure for wider range of PGA_d.

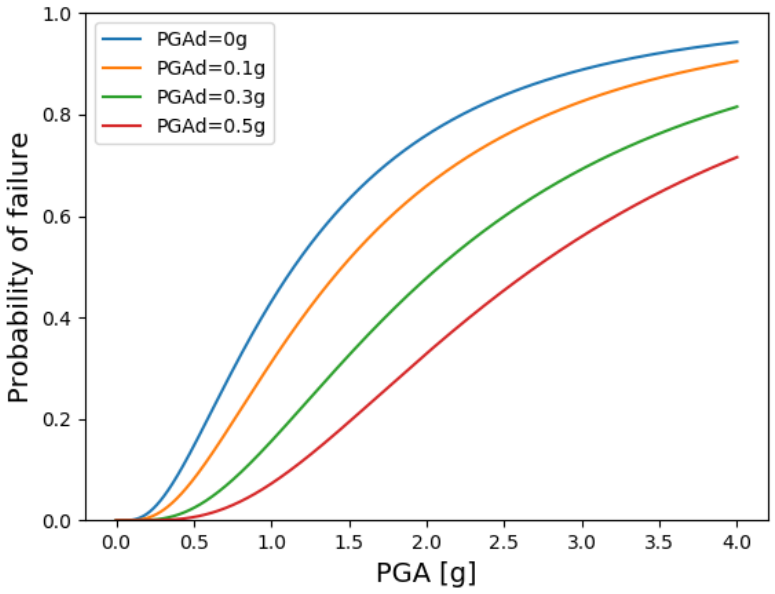

Figure 4.3: Fragility curves for the limit state of global collapse for different PGA_d

Outcomes: Expected Losses

As the VoI is estimated in monetary units in this study, we need to estimate the losses associated to designing the building for a specific PGA_d. The losses due to possible future earthquakes are a function of the hazard at the building's location and the vulnerability of the structural and non-structural components. The potential outcomes are the expected consequences for each of the possible decisions.

The fragility curves are converted to vulnerability curves using data relative to the cost of each structural and non-structural components and the damage percentage for each limit state. These vulnerability curves are convolved with the hazard curves to provide the Expected Annual Losses (EAL) (Gkimprixis *et al.*, 2020):

$$EAL = \int_0^\infty E(AL|IM)dH(IM) \tag{4.3}$$

where $E(AL|IM)$ is the expected annual losses conditional on the intensity level IM and $H(IM)$ is the hazard for the intensity IM.

These losses represent the cost of repair or replacement of each structural and non-structural component due to damage or total collapse. In case of total collapse, the cost of replacement represents the initial construction cost, C_d. Additional losses are loss of function of the building as well as injuries and fatalities (i.e., minor and major

injuries and cost of human fatalities). Other losses can be added to the expected losses (Lagaros, 2007), such as the loss of contents and rental loss (i.e., the loss of rental income while the building is being restored) and income loss, which is applicable to buildings that are used for commercial reasons, such as assumed for this building.

Depending on the time period, the expected future losses for a period t is as follows:

$$E[FL] = EAL.(1 - e^{\lambda t})/\lambda \tag{4.4}$$

where λ is a constant discount rate/year, which converts the future losses into present monetary value.

Finally, the life-cycle cost in case of total collapse is then simply:

$$E[LCC] = C_d + E[FL] \tag{4.5}$$

where C_d being the replacement cost of the building for a specific PGA_d.

Gkimprixis et al. (2020) applied this method to estimate the expected losses for the same four-storey three-bay reinforced concrete building at the same location. In Figure 4.4 are displayed the different costs computed from Equations (4.4) and (4.5) along with C_d for different PGA_d are displayed. These costs will be used within the following VoI calculations.

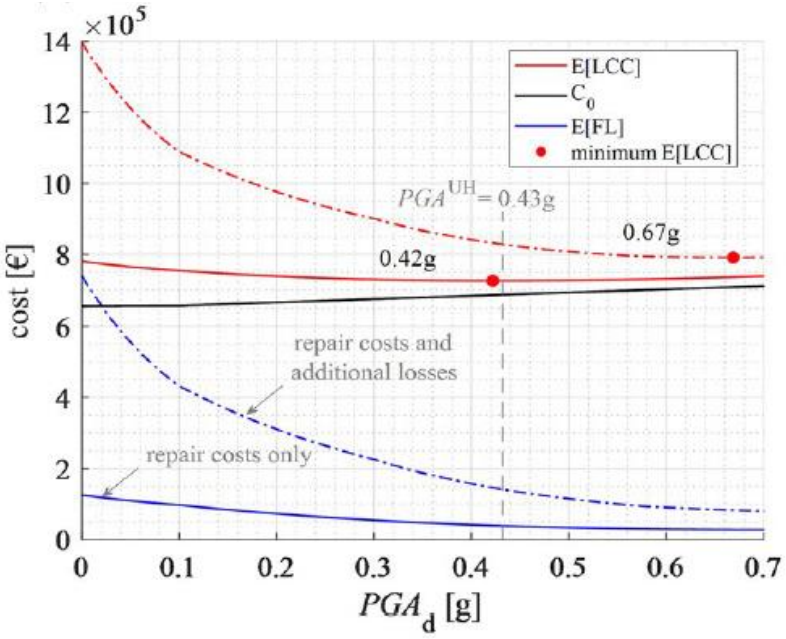

Figure 4.4: Different costs of expected losses for Patras including the initial construction costs. Dashed red line represents the total cost of construction, repair and the additional losses (from Gkimprixis et al., 2020)

The decisions outcomes here are the expected consequences for each of the possible decisions. These expected consequences are defined to be the sum of the initial construction cost C_d and the Expected Life-Cycle Losses, E[LCC], in case of total collapse, $o(x, d)$.

For one decision d, the expected outcomes are as follows:

$$E\big(o(x,d)\big) = \sum_x o(x,d)p(x) \tag{4.6}$$

where $o(x,d)$ represents the outcomes when choosing d and Vs_{30} is in the state x. x is the measure of interest and $p(x)$ is the prior probability of the state x. The outcomes for a measure x and a decision (i.e., design) d are expressed as follows:

$$o(x,d) = C_d + E[LCC](d).f_c(d,x) \tag{4.7}$$

where $f_c(d,x)$ is the probability of failure for a PGA associated with state x and extracted from the fragility curve of a design d.

4.3.2 Expected Value of Perfect Information

The Expected Value of Perfect Information (EVPI) represents the upper bound in monetary, utility or time unit that one should be willing to spend or consider

conducting surveys, tests, experiments or any action to obtain a piece of information without uncertainties.

Here, we assume performing a perfect test to infer the value of Vs_{30}. This *almost* perfect test could be borehole measurements up to the first 30m depth at the site of interest in Patras, in this case. This would eventually lead to no uncertainties, resulting in perfect information. It is important to stress that such a situation is quite unlikely. In fact, when it comes to geophysical and geotechnical information, uncertainties are generally always present. Nevertheless, computing EVPI has benefits as it fixes an upper limit value that should not be exceeded.

Building the method starting from a simple case for the EVPI estimation is necessary to partially validate the framework that is being constructed.

Decision tree

Decision trees are believed to be an excellent way of proceeding to a straightforward construction and description of the different important parameters and their causal dependency. In the decision tree illustrated in Figure 4.5, decisions are represented as rectangular nodes and probabilities as circular nodes. This decision tree depicts three possible decisions:

Action decisions

- Applying *Design 1* (associated with a $Vs_{30}=V_1$) with current information.
- Applying *Design 2* (associated with a $Vs_{30}=V_2$) with current information.

Test decisions

- Conduct a perfect test to obtain information about Vs_{30} without uncertainties.

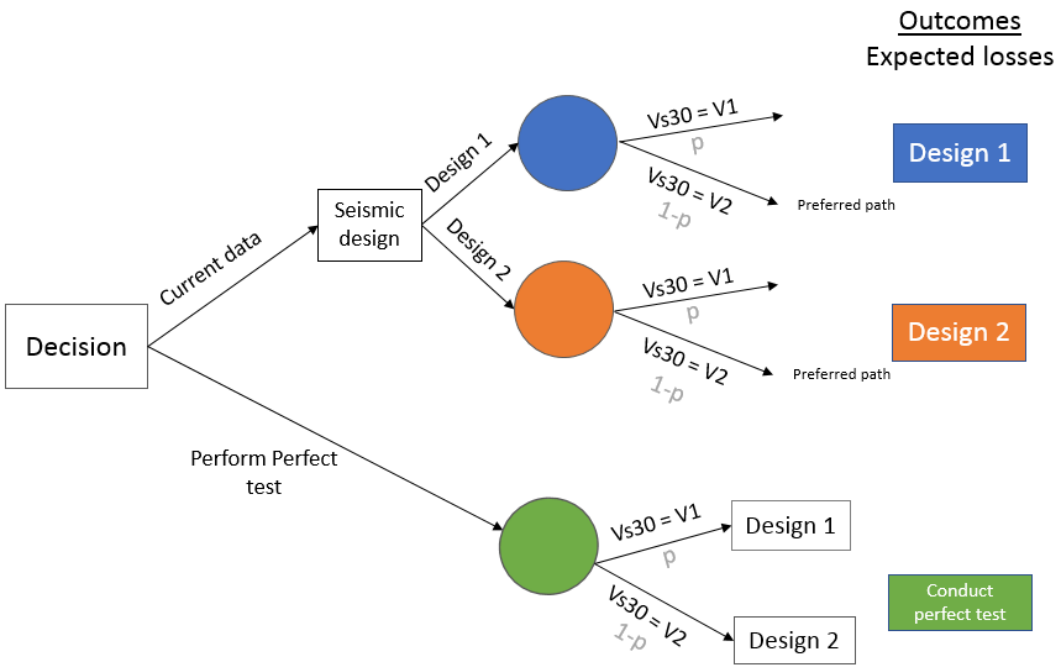

Figure 4.5: Decision tree for the computation of EVPI

The decisions relative to the case of using only current knowledge takes into consideration the possibilities of having V_1 and V_2 by computing the expected outcomes in both cases. The expected outcomes are computed for both decisions to identify the decision that minimises the expected losses when Vs_{30} is uncertain. This value represents the *Prior Value.*

If the decision of performing a perfect test is considered, the probability of the measurement outcome is equal to the prior probabilities of Vs_{30} and the expected losses are called the *Posterior Value*.

The above decision tree is an effective tool for computing the expected consequences of each decision and taking into consideration the prior probabilities of the parameter of interest.

Value of Information

Prior Value

In the case of *Before additional data*, two choices are possible, applying *Design 1* or *Design 2*. The expression of the associated *Prior Value*, *PV*, is as follows:

$$PV = \max_{d \in D}\{E(o(x,d))\}$$
$$= \max_{d \in D}\{\sum_x o(x,d)p(x)\} \quad (4.8)$$

where $\mathcal{D}$ is the domain of decisions d, $o(x,d)$ represents the outcomes for a decision d if Vs_{30} is in the state x and, $p(x)$ is the prior probability of the state x.

Posterior Value

The *Posterior Value, PoV*, is the resulting outcome of conducting a perfect test and thus, obtaining perfect information about x (in this case the true value of Vs_{30}) and applying the appropriate design.

$$PoV = \sum_x o(x, d_x)p(x) \quad (4.9)$$

where d_x is the appropriate design for state x.

Value of Information

The VoI is then simply the difference between the *PoV* and *PV*:

$$EVPI = PoV - PV \quad (4.10)$$

Irrespective of its type, *PV* is always constant. VoI is never negative as adding information always has benefits or no impact on the decision-making process. The VoI is then compared to the cost of the test to decide whether to proceed with data collection. The EVPI is then the maximum amount the decision-maker is willing to invest to have perfect information. In this case, EVPI is acting as a useful upper-bound to prevent the decision-maker from investing in expensive data-gathering schemes, methods or techniques but also from investing in a data collection process that would not be beneficial for decision-making.

Calculations and results

In this section, we proceed to the estimation of the VoI regarding the parameter Vs_{30}. We acknowledge that the VoI definition is used, and the expressions are developed for

the first time in this type of application. Thus, several sensitivity analyses are performed, and key values computed to validate the method in this discrete case.

EVPI sensitivity to prior probabilities

Prior probabilities are set at the beginning of the calculations, for the couple $[V_1,V_2]$. This translates beliefs, past experiences and available measurements on the state of Vs_{30}. For example, assuming 70% chance of Vs_{30} being V_1 is also assuming 30% of chance of Vs_{30} being V_2.

Here the Vs_{30} couple is fixed, and the prior probabilities are varied. V_1 is fixed to 100 m/s and V_2 to 500 m/s. Figure 4.6-a displays the expected outcomes in euros (i.e., losses) combining the construction costs and the expected losses from total collapse for the three main branches of our decision tree (Figure 4.5). In the legend, *Ec1* refers to the expected consequences computed from the branch associated to applying *Design1*, *Ec2* to *Design2*. Finally, *Ect* represents the expected losses after obtaining the perfect information and choosing the associated optimal design. The outcomes for each decision are computed for a range of all possible prior probabilities assigned to V_1 (similarly to V_2). Figure 4.6-b represents the VoI for several V_1 prior probabilities.

In Figure 4.6-a, we notice that the preferred decision before information depends on the prior probability. When giving to V_1 a prior probability $p<0.42$, the decision to apply *Design2* is preferred because of the least expected losses. Above that prior probability, *Design1* is preferred. Intuitively, the highest the probability of a particular value, the more probable it is to go with the associated decision.

The intersection between *Ec1* and *Ec2* at $p(V_1)=0.42$ is called the *indifference point* (Gilbert and Habibi, 2015), where outcomes for both decisions are equal. At this prior probability, the EVPI is at its maximum (Figure 4.6-b). EVPI is simply the difference between the preferred decision's outcomes before information and the outcomes after the information (*Ect*). At this indifference point, EVPI is about 12 000 euros. This represents the maximum amount to invest in a test that would completely remove the uncertainties.

It is interesting to notice that when V_1 is given a prior probability of 0 or 1 (meaning we already have certainty about the state of Vs_{30}), EVPI is equal to zero. This partially validates the method.

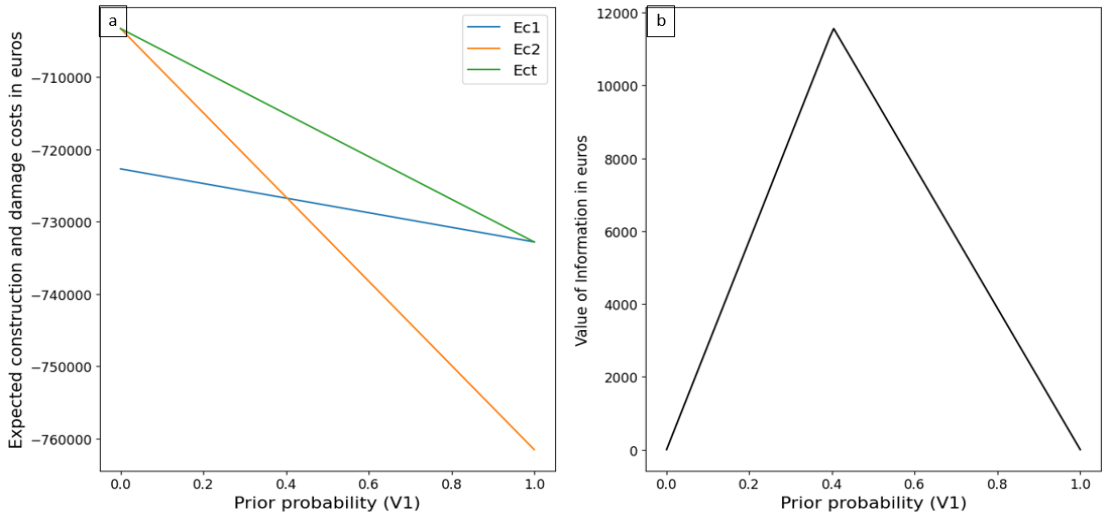

Figure 4.6: Sensitivity to prior probabilities for (a) the expected outcomes and (b) EVPI for the couple [100,500] m/s

Sensitivity to [V_1,V_2] couples

The same process is repeated for several [V_1,V_2] couples in order to assess the sensitivity of EVPI to the values of the binary couples and the prior probabilities.

Figure 4.7 demonstrates that the point of indifference, i.e., the prior probability associated with the highest EVPI, depends on the [V_1,V_2] couple. This means that the point of indifference depends on the outcomes associated to each decision. In the examples provided above, the point of indifference seems to be in the range of [0.4,0.6].

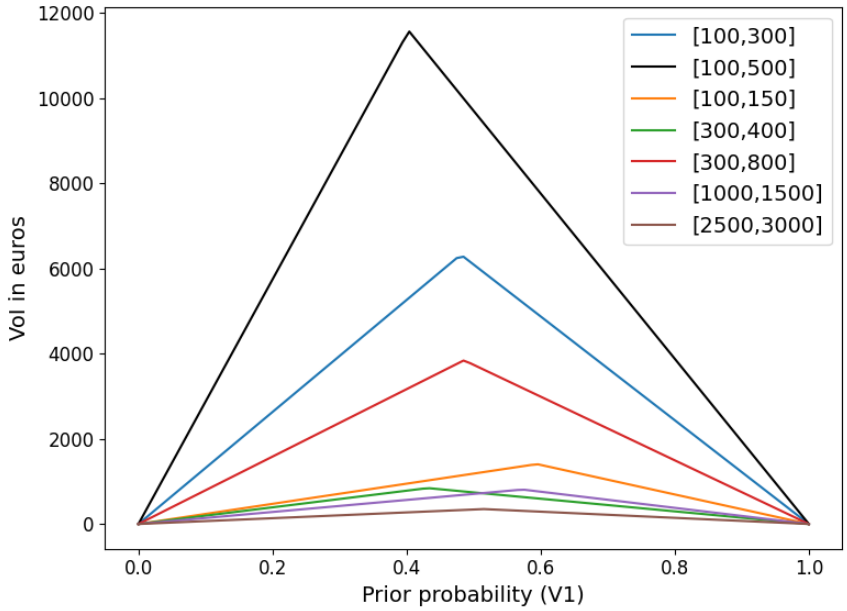

Figure 4.7: EVPI sensitivity to prior probabilities for several $[V_1, V_2]$ couples

A prior probability of 0.5 would suggest that there is no prior knowledge on the state of Vs_{30} and that VoI should be, intuitively, at its maximum. While this is not always the case, the range [0.4,0.6] of prior probability still represents high uncertainties about the Vs_{30}. An explanation to why the point of indifference is not always at a probability of 0.5 when we consider binary values can be found when looking at the expected outcomes for the different decisions (Figure 4.6-a). By changing the construction costs or the probability of failure, the point of intersection of *Ec1* and *Ec2* changes as well. Expressing a high belief for a particular state might suggest choosing the associated seismic design but this is not always the case as it depends on the expected losses for a particular decision. For example, if the construction costs for *Design1* were significantly higher than for *Design2* and that a prior probability of 0.7 is given to V_1, the expected losses will still be higher when choosing *Design 1* over *Design2*. In this case, *Design2* will still be the preferred decision and the point of indifference will be greater than 0.7.

To study further this dependence between $[V_1,V_2]$ and the prior probability that maximises the EVPI (i.e., *indifference point*), the EVPI was computed for $[V_1,V_2]$ couples for different probabilities and the *indifference point* relative to V_1 has been identified and represented in Figure 4.8. The *x* and *y* axis represent V_2 and V_1,

respectively. The colourbar, with midpoint at p=0.5, translates the indifference point to a colour. The white diagonal bar is an area that is not defined as VoI=0 when V_1=V_2.

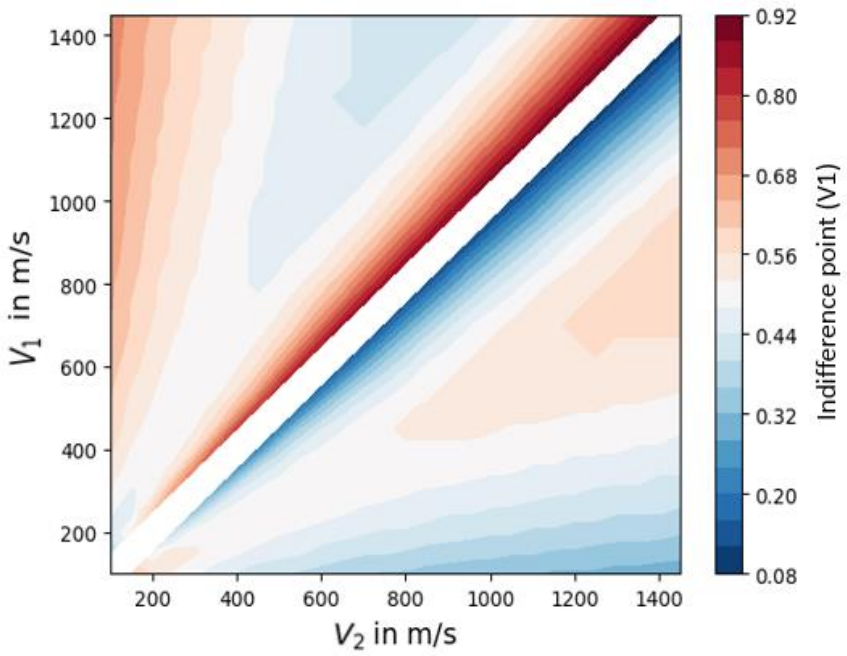

Figure 4.8: Impact of [V_1, V_2] couples on the prior probability that maximises the EVPI (i.e., the indifference point)

The point of indifference is mostly between 0.3 and 0.7. The closer V_1 and V_2 are, the further the point of indifference is from p=0.5. The lower triangle below the undefined area represents the point of indifference when V_1<V_2. The lower V_1, the more sensitive is the point of indifference to [V_1,V_2] gap increasing. This is the same for the upper triangle. Indeed, the patterns are symmetric relative to the non-defined area because the probabilities of V_1 and V_2 are complementary. This clearly shows that when the gap between V_1 and V_2 is large, the EVPI is maximum when we give a higher belief to the highest value. This confirms that the EVPI is not necessarily the highest when we are in complete uncertainty (0.5) and that it depends strongly on the range of uncertainty regarding the Vs_{30} couple (gap).

This is also confirmed when analysing the EVPI for each [V_1,V_2] couple shown in Figure 4.9. We see that the bigger the gap between the two possible values of Vs_{30}, the higher is the EVPI. This makes sense as the decision of going ahead by choosing one of the two values without perfect information greatly increases the chances of making

the wrong decision and thus, increases the expected losses. In contrast, the higher are the Vs_{30} values in the couple (i.e., towards rock conditions), the lower is the EVPI. This is explained because the amplification and thus, the PGA on soil, have a logarithmic behaviour.

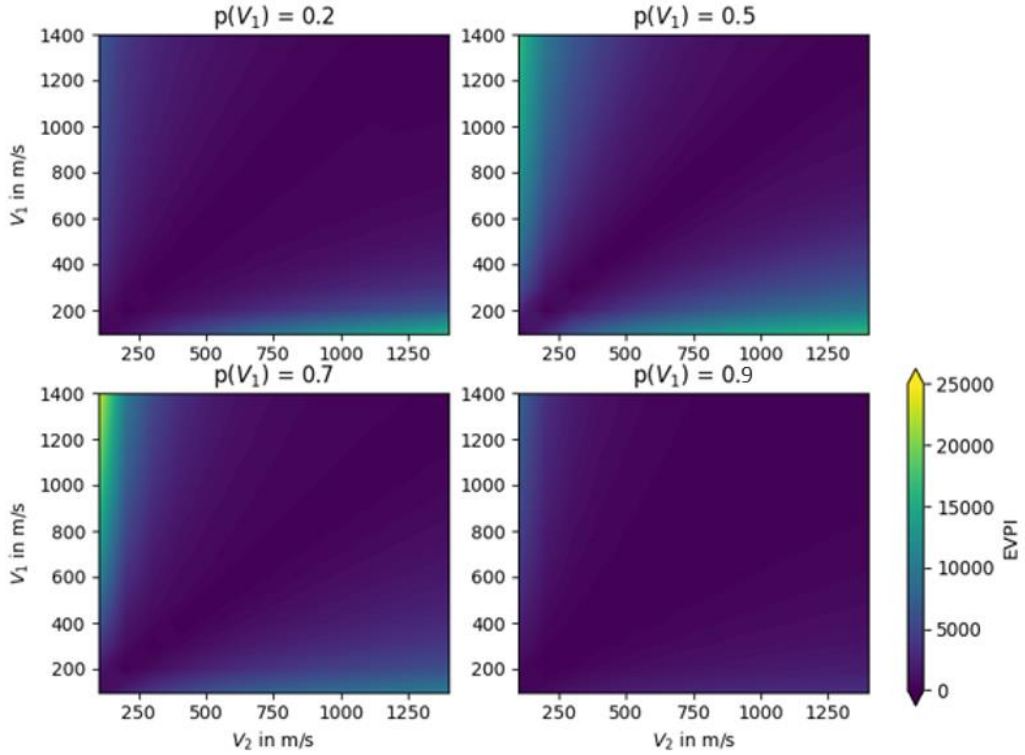

Figure 4.9: Sensitivity of the maximum EVPI (i.e., indifference point) to $[V_1, V_2]$ gap for V_1 prior probability of 0.2, 0.5, 0.7 and 0.9.

Moreover, intuitively, the more confidence we give a Vs_{30} value, the lower is the EVPI. This is not always true and depends on how the amplification function is defined. In Figure 4.9, for low V_2 and $V_2<V_1$, fixing $p(V_1)$ at 0.7 gives the highest EVPI values. Overall, the highest EVPI are noticed for prior probabilities between 0.3 and 0.7, while the lowest are noticed for very high probabilities (e.g., $p(V_1)>0.9$).

A method for VoI estimation has been created to infer the optimal seismic design for a structure when a single parameter is uncertain and discrete. In the next section, we aim to provide more realistic VoI estimates by computing the expected value of imperfect information (EVII). We assess the difference between the EVII and EVPI as well as its co-dependency on additional inputs.

4.3.3 Expected Value of Imperfect Information

Realistically, geotechnical or geophysical information is rarely completely free of uncertainties, i.e., perfectly accurate. Because most surveys would need analysis and interpretation to infer the measurement of interest, results are likely to have dispersion, characterised, for example, by a normal distribution with a given standard deviation. It is common to combine two or more tests to decrease the standard deviation (Long and Donohue, 2010).

When it comes to Vs_{30} measurements, the methods of prospection used will result in numerous possible values. Thus, the information is imperfect and incomplete. Although EVPI is a useful tool to fix an upper bound to not exceed, the EVII is the most likely VoI to be used for actual applications.

In this section, a EVII method is built using decision trees and Bayesian updating. Several sensitivity analyses are performed to understand the impact of the inputs on the results. Extreme values are tested to validate the method.

Updated decision tree

To compute the EVII, the use of a decision tree is still very convenient to set the causality and dependence between the different components, decisions and parameters as well as to infer the associated probabilities. Similarly, the imperfection of the test is expressed in the decision tree using marginal and posterior probabilities.

"Flipping" the tree

In this binary discrete case, a probability of accuracy is assigned to the test. This probability is set by experts from available information about the specific test or by the manufacturer, and expresses the level of confidence given to the measurements and interpretations.

For data y obtained from the imperfect test and x being the measure of interest (i.e., real Vs_{30}), the probability of the test being truthful to the real state of x is $p(y|x)$. In other words, $p(y|x)$ translates the level of accuracy of the test, called the *likelihood*, and is equal to 1 in case of perfect information. $p(x)$ is the prior model for the distinction of interest before the data y are made available. We define the *posterior*

model of *x* conditioned on the data *y*, *p(x|y)*, using Bayes' rule. We recall that Bayes' rule is used to perform Bayesian updating as follows:

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)} \tag{4.11}$$

where $p(y)$ is called the *marginal* or *pre-posterior probability model* and translates the probability that *y* is observed.

$$p(y) = \sum_x p(x)p(y|x) \tag{4.12}$$

We note that the straightforward problem allows going from *x* to *y* (Figure 4.10, left), Bayes' rule permits processing the problem from y to x and, thus, flipping the initial configuration. This is needed as the decision depends first on the test results to infer the distinction of interest. The reverse problem is illustrated in Figure 4.10 (right) and uses the marginal and posterior probabilities defined above.

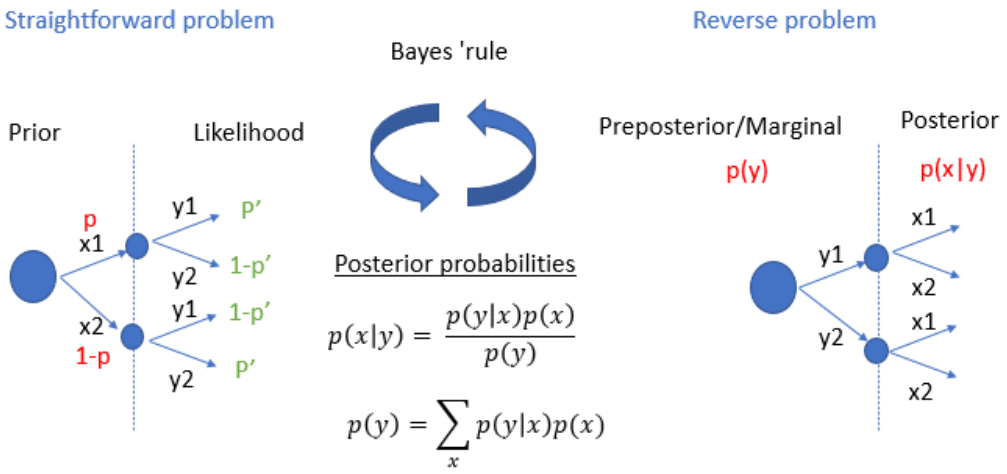

Figure 4.10: The use of Bayes' Rule to flip the decision tree for EVII computation. Left: Straightforward problem Right: Reverse problem. p is the prior and p' the likelihood probability

Updated tree

The full updated decision tree is illustrated in Figure 4.11. Similar to the previous decision tree, circles translate the probabilities and the rectangle nodes the decisions. The decision of conducting an imperfect test is made more complex, relative to the perfect test case.

In the continuity of the binary assumption, the imperfect test can give two possible values *Tv1* (V_1 from test) or *Tv2* (V_2 from test) with marginal probabilities as follows, respectively:

$$p(y) = \begin{cases} p(Tv_1) = p_m & if\ y = V_1 \\ p(Tv_2) = 1 - p(Tv_1) = 1 - p_m & if\ y = V_2 \end{cases} \quad (4.13)$$

The branches used to infer decisions with current information remain unchanged, as well as the associated expected losses. Hence, the *PV* remains the same. The decision branch relative to conducting an imperfect test is further developed. The uncertain circular node named *Marginal* translates the marginal probabilities of the test results.

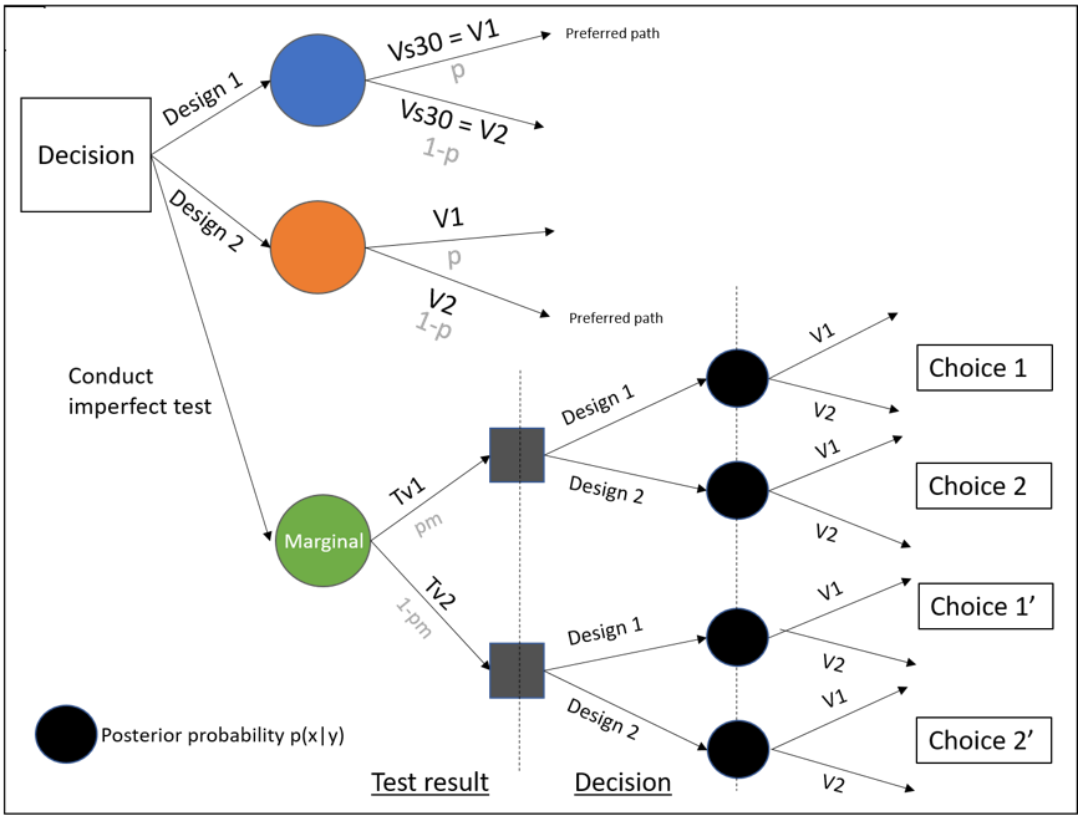

Figure 4.11: Updated decision tree for EVII computation. Circle nodes represent the uncertain parameter and rectangle nodes the decisions. Probabilities are displayed in grey with pm as the marginal probability of the test result being V_1

Value of information

The test results being uncertain, the decision-maker is faced again with the two same decisions but with updated probabilities. The expected losses for each {result, decision} pair are then computed by including the chance of the result being inaccurate. These outcomes will define the *Posterior Value*.

Posterior Value

There are four posterior probabilities associated with four possibilities for this binary case with *y* being the test result and *x* the real value of the measure of interest. The decision tree is read from right to left. The probability of *x* being V_1 if the test result is V_1 is inferred from Bayes' rule as follows:

$$p(x = V_1|y = V_1) = \frac{p(x=V_1)p(y=V_1|x=V_1)}{p(y=V_1)} \tag{4.14}$$

The probability of *x* being V_1 if the test result is V_2

$$p(x = V_1|y = V_2) = \frac{p(x=V_1)p(y=V_2|x=V_1)}{p(y=V_2)} \tag{4.15}$$

The probability of *x* being V_2 if the test result is V_2

$$p(x = V_2|y = V_2) = \frac{p(x=V_2)p(y=V_2|x=V_2)}{p(y=V_2)}$$

$$= 1 - p(x = V_1|y = V_2) \tag{4.16}$$

The probability of *x* being V_2 if the test result is V_1

$$p(x = V_2|y = V_1) = \frac{p(x=V_2)p(y=V_1|x=V_2)}{p(y=V_1)}$$

$$= 1 - p(x = V_1|y = V_1) \tag{4.17}$$

We note that in this binary case, we can limit the calculations to two posterior probabilities as the posterior probabilities are complementary when *y* is fixed.

As illustrated in Figure 4.11, each test result and subsequent decision is called *Choice n°*. After observing a measurement, the decision-maker will choose the decision that minimises the expected losses.

The *Posterior Value* (*PoV*) is then

$$PoV = \sum_y p(y).\max_{\{1,2\}}\{Choice^{(y)}\}$$

$$= \sum_y p(y).\max_{d\in D}\{\sum_x o(x,d)p(x|y)\} \quad (4.18)$$

We recall that the *max* operator is used instead of *min* as the expected losses are assumed to be negative values.

Value of information

The EVII is the difference between the *PoV* and the *PV:*

$$EVII = PoV - PV$$

$$= \sum_y p(y).\max_{d\in D}\{\sum_x o(x,d)p(x|y)\}$$

$$- \max_{d\in D}\{\sum_x o(x,d)p(x)\} \quad (4.19)$$

Note that the perfect information can be viewed as a special case of the imperfect information when there are no uncertainties in measurements (i.e., $p(x|y) \in \{1,0\}$).

Similar sensitivity analyses to the EVPI case are performed, analysed and interpreted in the next section. In addition. sensitivity analyses over the likelihood are also presented which permit validation of the methodology and putting in evidence the importance of thoroughly assigning the accuracy of a test.

Calculations and results

In this section, the EVII is computed using the method and associated expressions above for the same PGA_r. The EVII is expected to be lower than the EVPI assuming that the information is imperfect.

Sensitivity to prior probabilities

EVII is computed for the same $[V_1,V_2]$ couple as Figure 4.6 and a test accuracy (i.e., likelihood) of 80%. The results are compared to the EVPI and displayed on Figure 4.12.

The expected losses for *Design1* and *Design2* remain unchanged whereas the expected losses in Figure 4.12-a after the imperfect test (dashed green line) are higher than the

perfect test (solid green line). This results in a decrease in the VoI, shown in Figure 4.12-b. We notice that, unlike the EVPI, the EVII can be equal to zero for prior probabilities different from 0 and 1. In a further analysis, we show that this result is correlated with the likelihood.

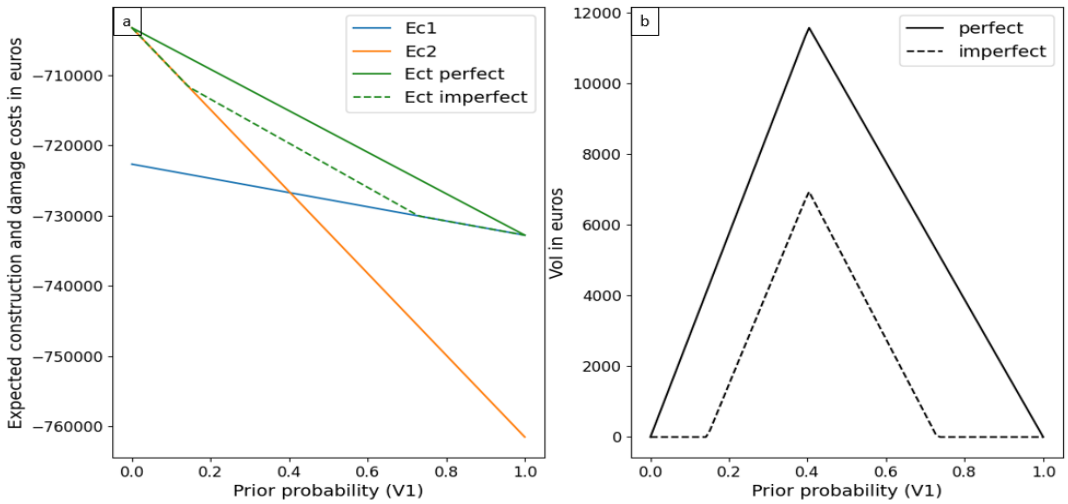

Figure 4.12: Sensitivity to prior probabilities for the (a) Expected outcomes and (b) EVPI (solid line), EVII (dashed line) for the couple [100,500] m/s

Indeed, the likelihood represents an input that needs to be carefully estimated and incorporated into the calculations. We suggest studying the sensitivity of the EVII to the likelihood when prior probabilities are fixed.

Sensitivity to the likelihood

The prior probabilities are fixed to 0.5 and the Vs_{30} couple to V_1=100 m/s and V_2=500 m/s. The EVII associated to a range of likelihood probabilities from 0 to 1 is computed and displayed on Figure 4.13.

The results suggest that EVII equals zero when the test accuracy is 50% (i.e., likelihood equals 0.5). A test with 50% accuracy is of no help in this binary case as there is 50% chance that the test is right and 50% chance it is wrong. On the other hand, a test accuracy of 100% is equivalent to a perfect test and EVII is at its maximum, and equal to the EVPI. We notice that the graph is symmetric. For example, a test accuracy of 0% means that if the test result is V_1, there is a 100% probability that the soil Vs_{30} is equal to V_2. The test becomes equivalently perfect.

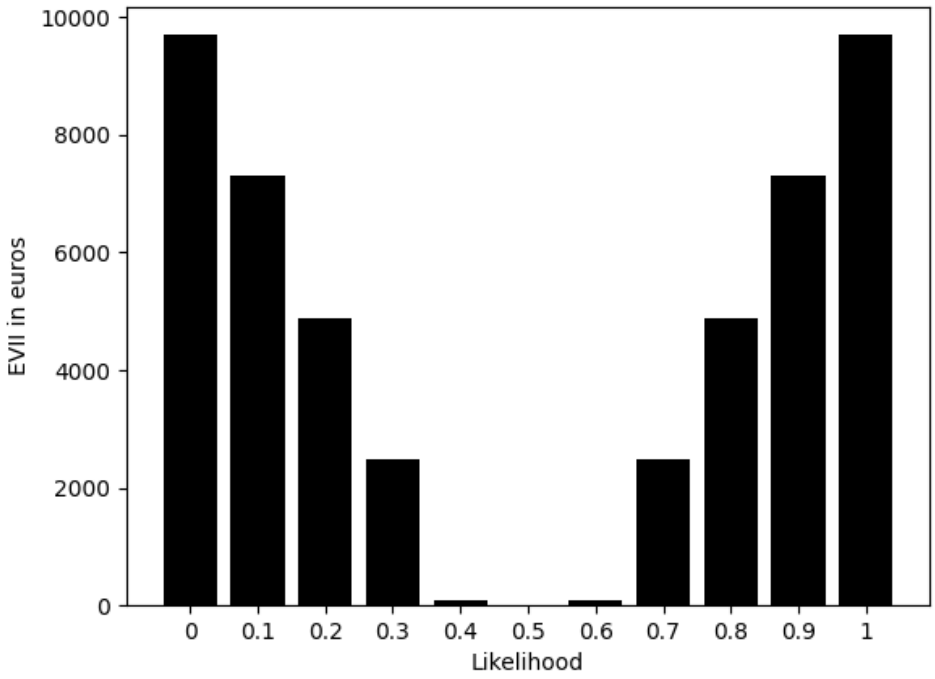

Figure 4.13: Sensitivity of EVII to the likelihood probability (test accuracy) for prior probability = 0.5

Because of the symmetrical nature of the graph, studying the right side (likelihood>0.5) is sufficient. In our example, increasing the likelihood by 10% increases the EVII by around 2000 euros. The increase appears to be linear, which is confirmed by Figure 4.14, which shows the EVII for several prior probabilities and different likelihoods. The results confirm that EVII is equal to the EVPI when the likelihood is set to 100%. This is an additional validation of the method.

This sensitivity analysis denotes the impact of both the prior and the likelihood on the EVII. Indeed, the less accurate the test, the smaller the range of prior probabilities given to V_1 and V_2 where there is a non-zero VoI. If we take the example of a test that is 90% accurate (Figure 4.14), we find that beyond a V_1 prior probability of about 85%, there is no value to the additional information. In fact, there is no VoI when the prior probability is equal or above the likelihood. In other words, there are no benefits in conducting a test if we are more confident about the value of the measure of interest than the test itself. For example, if we are certain at 60% about the value of the true Vs_{30}, the surveys that need to be conducted to obtain a non-zero VoI should give results

with an accuracy of more than 60%. This interpretation confirms that the decision-maker could filter the list of tests that could be beneficial based on the current knowledge before performing the VoI assessment.

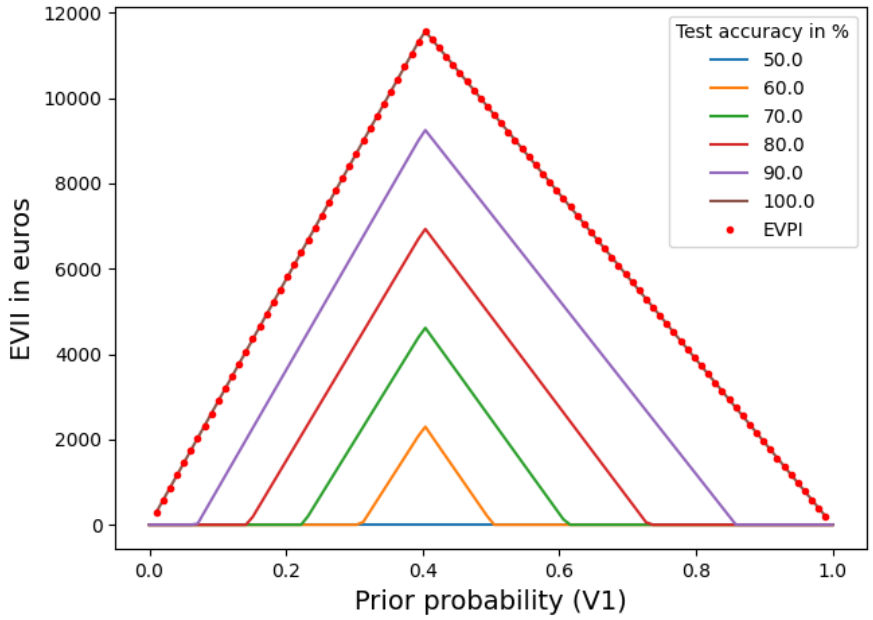

Figure 4.14: Sensitivity of EVII to V_1 prior probability for different test accuracy probabilities

4.3.4 Conclusion

These sensitivity analyses highlight the parameters and inputs that influence most the VoI when the uncertain parameter is discrete. This shows that some inputs should be estimated and chosen carefully to obtain a high reliability in the VoI estimates. The prior probability given to the possible values of the uncertain parameter Vs_{30} was shown to greatly influence the EVPI and EVII. Thus, it is important to wisely and thoroughly use the available information to infer the prior probabilities.

The uncertainty about the parameter is also expressed through the gap between V_1 and V_2. In a more realistic and complex case study in the next section, these uncertainties are considered through probability distributions assigned to Vs_{30}. Another important finding is that for site rock conditions, uncertainties regarding the Vs_{30} that have little influence on the EVPI and it might not be worth conducting further measurements.

The developed method has shown to validate the intuitive behaviour of VoI regarding the inputs. VoI decreases when information is imperfect which makes EVII always smaller than EVPI. The level of confidence given to a test has a strong correlation with the prior probabilities when it comes to VoI. A test with less accuracy than the prior probabilities set by the available data and experts' knowledge is not worth conducting. Moreover, the more accurate the test, the higher the benefits of obtaining the information. Calculations using extreme and specific values have validated the present method. It has been shown that gaps between the outcomes of either of the decisions are strongly correlated with VoI estimates.

Recall that this binary uncertainty for Vs_{30} is not realistic, especially when V_1 and V_2 have a large gap. Uncertainties for a parameter such as Vs_{30} should ideally be expressed through continuous probability distributions. However, this binary approach may be useful in other scenarios, such as when making decisions about whether a fault is active or inactive (i.e., it does not generate earthquakes). In this case, this method could be useful in justifying in-situ or satellite remote sensing data collection as well as estimating the maximum investment in time and resources.

4.4 Scenario 2: VoI for a continuous uncertain variable

In this second scenario, the assumption is made that Vs_{30} is continuous rather than discrete, which is a more realistic representation of the parameter. We consider the same case study and inputs as scenario 1 (Section 4.3) with Vs_{30} uncertainties expressed through probability density functions.

4.4.1 Input and parameters

If some inputs such as the PGA on rock, the fragility curves and the definition of expected outcomes remain unchanged, the definition of probabilities and the approach for VoI computations are relatively easily adapted for continuous parameters.

Continuous prior probability distribution $p(x)$

Integrating available data and consulting experts are useful in estimating the probability distribution to be used. While available data may indicate complex probability distributions, expert elicitation is more straightforward and often capture

the uncertainty through standard distributions. Here we assume that Vs_{30} uncertainties can be expressed by a normal distribution with mean μ and standard deviation σ. It is assumed that the current prior information only defines a range of possible Vs_{30} along with extreme possible values, which makes the normal distribution more realistic than binary values. Figure 4.15 displays the prior probability distribution in blue for Vs_{30} with mean 500 m/s and a standard deviation of 120m/s. This probability distribution allows Vs_{30} to have values comprised between 100 and 900m/s with a maximum belief for Vs_{30}=500m/s. Calculations using a continuous variable rely on integrals to capture the spectrum of all possible values. Nevertheless, some functions might require approximations to evaluate the integrals.

Decisions

The decision-maker must make a choice on applying a particular seismic design to the building based on available information or conducting a perfect/imperfect test to make the decision under lower uncertainties. Besides the decision regarding data collection, the decision-maker has to choose from different seismic designs. This finite number of possible seismic designs $\mathcal{D} = \{d_1,\ldots, d_i,\ldots d_M\}$ is set *a priori* from the range of possible Vs_{30} based on current information. Specifically, both optimal designs d_1 and d_n are based on both extreme values of Vs_{30}. The range d_1-d_n is then discretised for additional decisions on the seismic design.

Outcomes

The outcomes are defined as follows:

$$o(x, d_i) = \begin{cases} C_{d_i} & PGA_{opt}(x) > PGA(d_i) \\ C_{d_i} + E[LCC](x).\, f_c(d_i, x) & PGA_{opt}(x) < PGA(d_i) \end{cases} \quad (4.20)$$

$PGA_{opt}(x)$ is the PGA associated to the optimal design for a value x of Vs_{30} and $PGA(d_i)$ is the design PGA relative to the decision d_i. If the inferred optimal design for x is more robust than the chosen seismic design from $\mathcal{D}$, only the construction costs are considered. Otherwise, expected losses from total collapse are included in the outcomes. $PGA_{opt}(x)$ is computed by minimising Equation. (4.7).

4.4.2 Prior Value before information

The prior expected losses before performing a test are expressed as follows:

$$PV = \max_{d_i \epsilon D} \{ \int o(d_i, x)\, p(x) dx \} \tag{4.21}$$

$o(d_i, x)$ is the cost of designing the building for the decision d_i when the real distinction of interest Vs_{30} is equal to *x*. $p(x)$ is the probability density function of Vs_{30}.

To approximate the integral in Equation (4.21), Monte Carlo simulations (Gelman *et al.*, 1995) are performed to infer samples x_k from the probability distribution $p(x)$. In particular, Markov Chain Monte Carlo (MCMC) methods are used to generate *n* samples according to the chosen probability distribution. This translates into a higher number of samples where the probability is high and vice-versa. In Figure 4.15, dots in red represents 10,000 samples from *p(x)*. Indeed, the higher the probability density function, the higher the number of samples.

The decision tree for the computation of the prior value is presented in Figure 4.16 for an arbitrary number of five alternative seismic designs. The structure of the decision tree after performing MCMC simulation is similar to the discrete case in section 4.3. The number of branches is significantly higher when all samples are included. Expected outcomes are computed for each sample x_k and decision d_i.

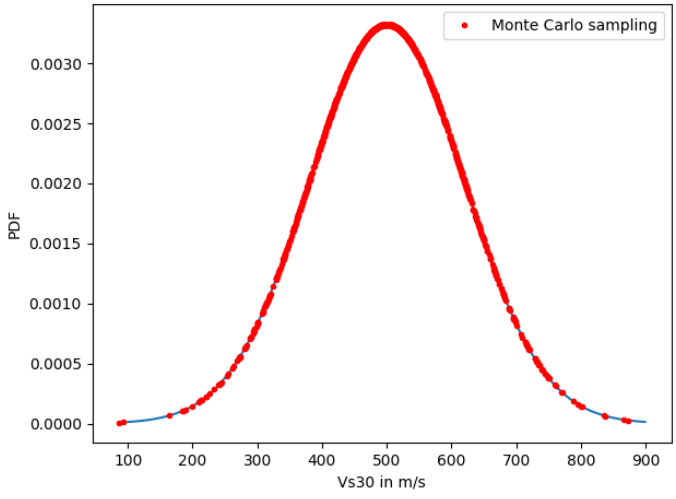

Figure 4.15: Vs_{30} Prior probability distribution (Blue) and Monte Carlo simulation (10,000 samples) (red). Mean 500m/s and standard deviation of 120m/s

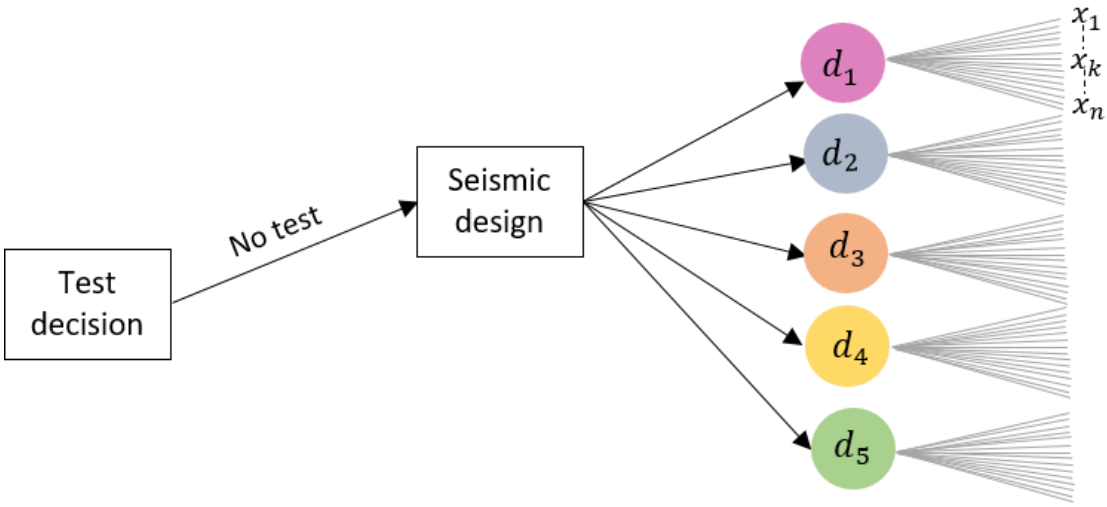

Figure 4.16: Decision tree to compute the prior value, PV. d_i represents the alternative decisions and x_k are samples obtained from Monte Carlo simulations on $p(x)$

The Monte Carlo integral approximation is used to evaluate the *prior value PV* before information as follows:

$$PV = \max_{d_i \epsilon D}\{\int o(d_i, x)\, p(x) dx\} \sim \max_{d_i \epsilon D}\{\frac{1}{n}\sum_{k=1}^{n} o(d_i, x_k)\} \tag{4.22}$$

This approximation estimates the integral by computing the average of the outcomes for each decision if and only if x is sampled according to the probability distribution $p(x)$. *PV* is then obtained by choosing the decision that will minimise the expected losses. We recall that the operator max is used since the losses are expressed as negative values.

4.4.3 Expected Value of Perfect Information

EVPI is computed in case of a perfect test where the information is equal to the true value of Vs_{30}. The Posterior Value (*PoV*) is computed by simply reversing the integral and the *max* operator in Equation (4.22).

$$PoV = \int \max_{d_i \epsilon D}\{ o(d_i, x)\} p(x)\, dx \sim \frac{1}{n}\sum_{k=1}^{n} \max_{d_i \epsilon D}\{o(d_i, x_k)\} \tag{4.23}$$

The calculation requires no additional steps if we have computed it for all samples and all alternatives for the prior value approximation in Equation (4.22).

EVPI is then simply:

$$EVPI = PoV - PV$$

The assigned prior distribution has a strong impact on VoI. To illustrate this, EVPI is computed using prior distributions of same mean μ but of different standard deviations σ (Figure 4.17). On one hand, the EVPI is shown to have a linear increase with σ. Indeed, a higher prior standard deviation suggest higher uncertainties, equivalent to lower prior knowledge. The less available data, the more chances a perfect test will be beneficial. On the other hand, the EVPI in this continuous case are lower than the values estimated in the configuration of the discrete case. This is explained by the difference in the prior probabilities but mostly, by their definition. When the uncertain parameter is described by a continuous probability distribution, the number of possible true values are higher (*n* possible values), whereas they are only two in the discrete case. Here, the gap between the two extremities of Vs_{30} range can be seen as the gap between V_1 and V_2 in the first case, only here, all the possible values in between are considered. Moreover, the number of alternative decisions is higher, which results in a lower EVPI.

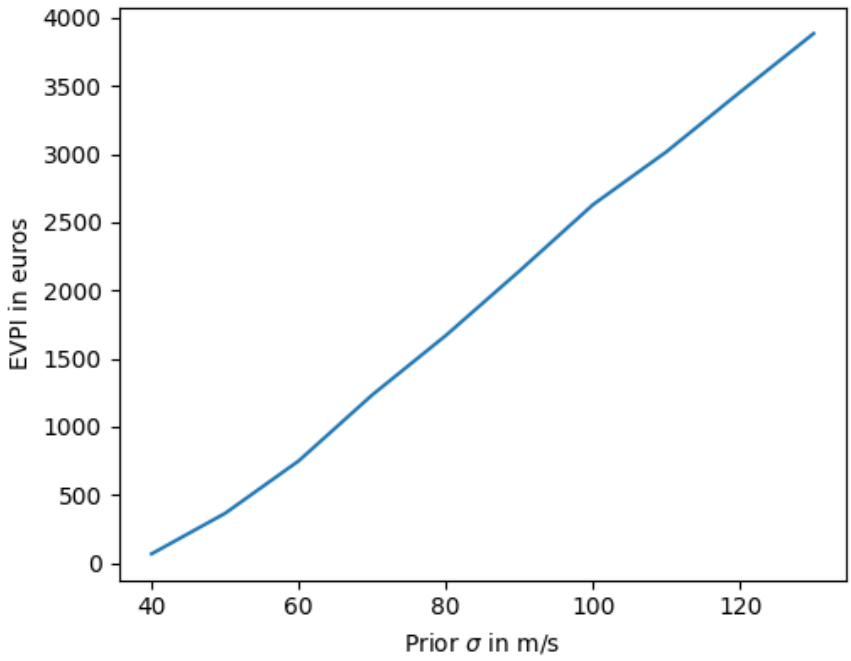

Figure 4.17: Sensitivity of EVPI to prior standard deviation

The stability of the results depends on the number *n* of random samples generated. Monte Carlo simulation is efficient when sampling according to a probability distribution, but sampling remains random. On one hand, a low number of samples may result in unstable VoI values when sampling and calculations are repeated. On the other hand, a high number of samples might have a high computational cost.

To assess the optimal number of samples needed to obtain stable results, EVPI is computed for various numbers of samples and two different prior standard deviations. Results are shown in Figure 4.18. For σ= 120 m/s, EVPI values fluctuates when *n*<10,000 but are stable for higher number of samples. Whereas for a tighter distribution, σ=60 m/s, EVPI is stable for *n*> 4 000. Fewer samples are needed when the range of values is smaller. This insight is beneficial in order to reduce the number of unnecessary calculations. For the following calculations, σ is fixed at 120 m/s and the number of samples is set to 10,000.

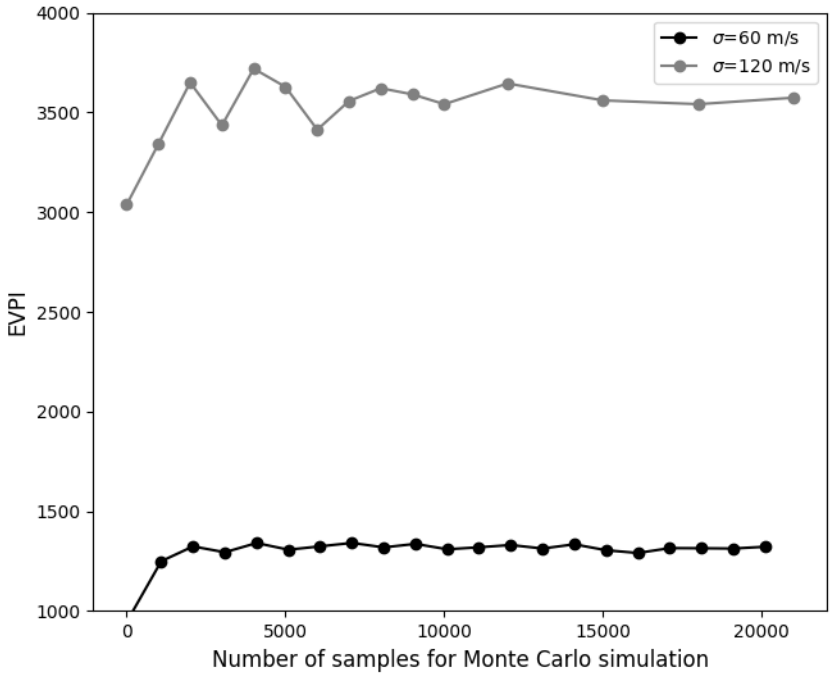

Figure 4.18: Stability of VoI to Monte Carlo number of samples *n*

4.4.4 Expected Value of Imperfect Information

To compute EVII, we assume that the test to be performed is imperfect. To translate this imperfection, we define a test error function denoted $e(x)$. The probability density

function is normally distributed with a mean of μ_t (=0) (i.e., the test is unbiased) and standard deviation σ_t. The lower the standard deviation, the more accurate is the test. An example of the function $e(x)$ is shown in Figure 4.19 when σ_t = 30.

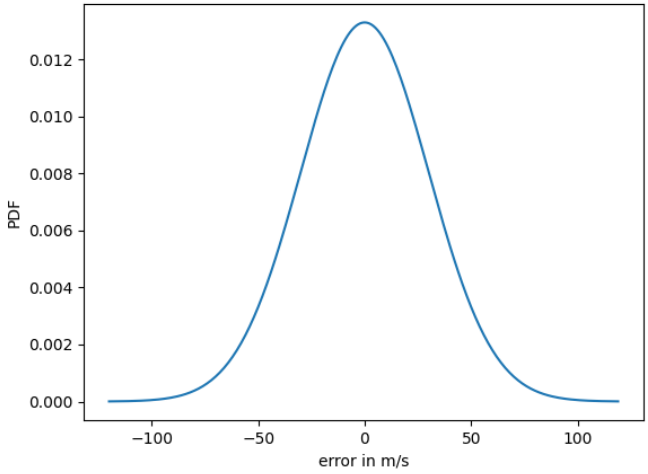

Figure 4.19: Test error function with mean 0 and a standard deviation of 30 m/s

General workflow for EVII assessment

The adopted workflow to compute VoI for a continuous prior and likelihood is detailed as follows:

1- *Simulation of test observations*

- Pick x^* as random Vs_{30} value from the prior probability distribution $p(x)$
- Pick random error e^* from the test error function $e(x)$
- Compute $y^* = x^* + e^*$
- Repeat

2- *Construction of observations probability distribution: Marginal distribution p(y)*

- Plot histogram of simulated y^* and fit a distribution to construct $p(y)$
- Monte Carlo sampling of $\boldsymbol{N}$ observations from the probability distribution $p(y)$:

$$y = \{y_1, .., y_j, .. y_N\}$$

3- *Computation of expected outcomes conditioned on each observation y_j*

- Compute the posterior distribution for an observation y_j using Bayes rule, $p(x|y_j)$

The posterior probability distribution $p(x|y_j)$ is constructed by computing $p(x_k|y_j)$ for each sample x_k from the definition domain of $p(x)$ using Bayes' rule:

$$p\left(x_k\middle|y_j\right) = \frac{p(x_k)p(y_j|x_k)}{p_m(y_j)} = \frac{p(x_k)e(y_j-x_k)}{p_m(y_j)} \quad (4.24)$$

The marginal probability $p_m(y_j)$ is a factor used to normalise $p(x|y_j)$

$$p_m(y_j) = \sum_k p(x_k)p(y_j|x_k) \quad (4.25)$$

- Perform a Monte Carlo sampling from the posterior distribution $p(x|y_j)$ to infer m samples of $x = \{x^1,..,x^l,..x^m\}$ and apply the Monte Carlo integral approximation to compute the Expected Outcomes, EO, conditioned on y_j for a decision d_i:

$$EO(d_i|y_j) = \int o(d_i,x)p(x|y_j)dx \sim \frac{1}{m}\sum_{l=0}^{m} o(d_i,x^l) \quad (4.26)$$

Note that the samples x^l are different from the previous samples x_k as they are sampled from the posterior distribution.

- Repeat for each decision d_i
- Pick the highest EO, $\max_{d_i \in D}\{EO(d_i|y_j)\}$

4- Value of information

Consequently, VoI can be computed using this equation:

$$VoI = \int_y p(y).\max_{d_i \in D}\{\int_x o(d_i,x)p(x|y)dx\}\,dy - PV$$
$$\sim \frac{1}{N}\sum_{j=0}^{N}\max_{d_i \in D}\left\{\frac{1}{m}\sum_{l=0}^{m} o(d_i,x^l)\right\} - PV \quad (4.27)$$

The workflow to compute *PoV* is further illustrated in Figure 4.20 through a decision tree.

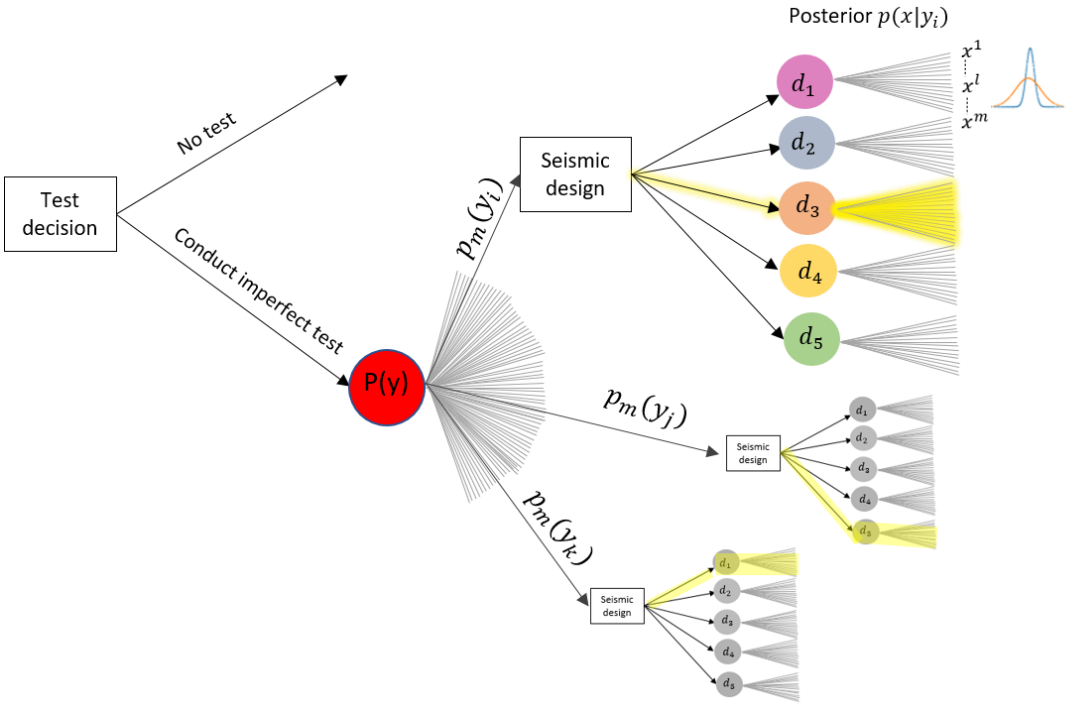

Figure 4.20: Posterior decision tree when conducting imperfect test. The red node represents the observations marginal probability. Examples of decision trees for three different observations are given and the optimal decision is highlighted in yellow

This method allows simulating a large number of probable observations that include the variabilities inherent to the test. The first probability node (red) in Figure 4.20 represents the marginal probability of the various simulated observations. Using the posterior distribution, the expected outcomes for each possible decision are calculated, conditioned on the observation. Then, the decision that results in the lowest losses is selected. Monte Carlo sampling from the simulated observations allows the *PoV* to be computed by averaging the expected losses relative to the optimal decision for each observation.

Results and sensitivity analyses

EVPI and EVII have been computed for the parameter Vs_{30} of prior probability distribution defined by $Vs_{30} \approx N(\mu = 500, \sigma = 120)\ m/s$. The computation of EVII is more complex and time consuming than EVPI. The calculations were performed using Python. For a fixed prior and test error distributions, EVII can be computed in approximately 15 minutes if N =1,000 observations are simulated from $p(y)$.

Posterior distribution behaviour

We would like to observe how posterior distributions vary for a given observation and prior probability distribution $p(x)$. Figure 4.21 displays the prior probability distribution $p(x)$ and posterior probability distributions for three different observations *y**, 300, 500 and 700 m/s. These values are carefully chosen to study the impact of the prior probability $p(x)$ on the posterior $p(x|y^*)$. The observations of 300 and 700 m/s are lower and higher, respectively, than the mean μ of $p(x)$ and the observation *y*=500 m/s equals μ. The vertical grey lines are added to each posterior distributions to indicate the associated mean observation value.

The posterior conditioned on observing *y*=300 m/s has a mean that is slightly higher than the value of the test result (observation). Whereas for *y*=700m/s, the mean is slightly lower. For *y*=500m/s, the mean of the posterior corresponds exactly to the observation. These results show the impact of the prior distribution. In fact, for *y* lower or higher than μ, the associated posterior's *mean* tends to get closer to the highest prior probability, meaning that the credibility of the observation is not at its highest and suggest more closer values to the actual highest prior. This is confirmed for the case where the observation equals the highest prior probability value 500 m/s.

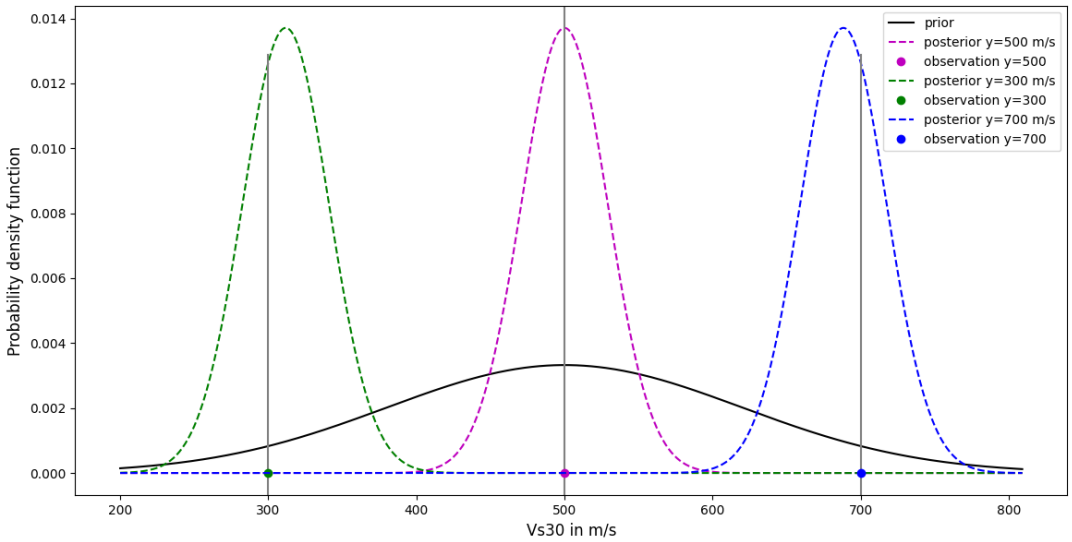

Figure 4.21: Prior distribution (black) and posterior distributions (dashed) for observations 300, 500 and 700m/s

The posterior is shown to be a compromise between the prior and the likelihood. To confirm this, we study further this relationship by observing the behaviour of the posterior when the prior standard deviation changes. The indicator chosen for this study, I, is the percentage of the difference between the posterior distribution mean μ_{post} and the observation. Its value is shown on Figure 4.22 for different prior σ and observations y_{obs}.

The shift in percentage of the posterior mean μ_{post} from the observation is expressed as follows:

$$I = \frac{\mu_{post} - y_{obs}}{\mu} . 100 \tag{4.28}$$

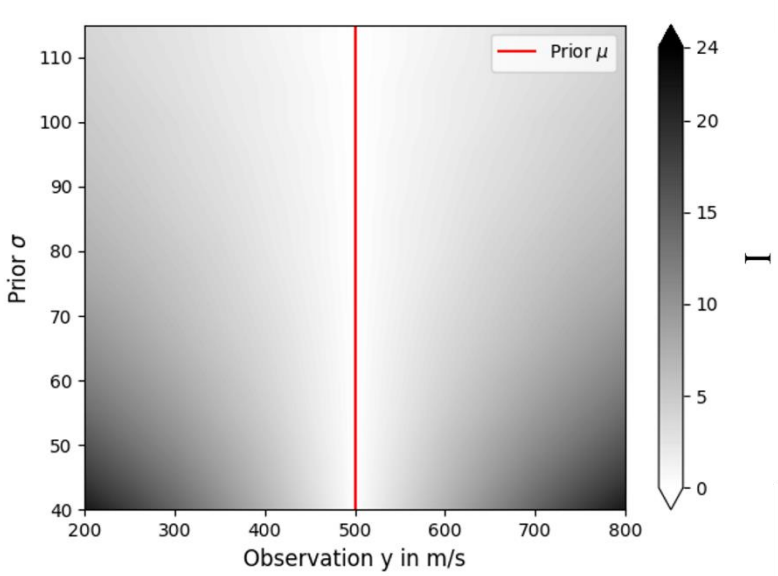

Figure 4.22: Impact of prior standard deviation σ on posterior distribution mean μ_{post}

The x and y axis represent the observation after test and the prior standard deviation σ, respectively. The prior mean μ is fixed and represented by a red vertical line. We observe that the further the observation from the prior mean, the more the posterior's mean is shifter toward the prior. This behaviour is valid for all σ. In other words, the further the observation from μ, the less credibility is given to the observation. The chosen prior σ also has an impact on the sensitivity to the posterior. Indeed, the tighter the prior distribution is (lower σ), the more the posterior is influenced and shifted toward the prior mean. This shows that when prior knowledge is high, the posterior is more tuned to fit the prior.

Sensitivity to likelihood

The workflow described above was applied to compute EVII for several likelihood functions defined by different error functions and associated standard deviations. Figure 4.23 displays the sensitivity of EVII to different likelihood standard deviations. The result is in accordance with the intuition that more accurate tests (lower error standard deviation) lead to higher VoI. Moreover, when the test standard deviation tends to zero, the EVII tends to the EVPI.

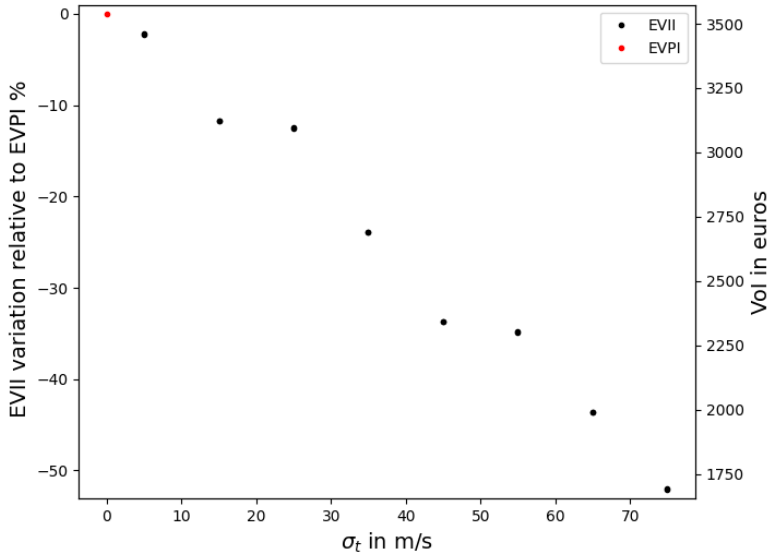

Figure 4.23: EVII sensitivity to test's error standard deviation. Variation relative to EVPI (left axis), VoI (right axis)

Whether using a discrete or a continuous definition for an uncertain parameter, the likelihood function has a significant effect on the estimated EVII. We conclude that it is important to carefully examine and establish the reliability of any testing procedures, based on past experiences.

In general, we can say the following about the prior-likelihood-posterior relationship:

- The posterior distribution is a compromise between the prior and the likelihood. The higher the data quality and/or quantity, the higher the influence of the likelihood on the posterior.
- For a given set of data, the greater the certainty in the prior, the higher the influence of the prior mean over the posterior.

- Conversely, for a given set of data, the less prior knowledge, the more the likelihood controls the posterior.

In other words, the expected paucity or abundance of observed data have an impact on how priors might be defined. A non-informative prior (e.g., uniform distribution) might be sufficient to estimate the posterior distribution when enough data can be collected. However, when data provide little constraint on the target parameters, a more informative prior should be considered.

An interesting application of the method would be to assess EVII for different types of tests and choose the type with the highest EVII. Nevertheless, the EVII should also be compared to the cost of the test. The computation of the net benefit, EVII(test)-cost(test), is an efficient way to choose the type of test (or not to test at all).

Other prior distributions

Thus far, only normal distributions have been considered for prior definitions. Some of the limitations encountered when a mean is fixed and the standard deviation is increased is that Vs_{30} could become negative, which is physically impossible. To overcome this, it is often preferable to use lognormal distributions. Lognormal distributions are often utilised to model uncertainty in Vs_{30} obtained from datasets of ground investigations or strong-motion recordings (Mital et al., 2021; Mori et al., 2020; Seyhan et al., 2014; Boore et al., 2011). Lognormal distributions are suitable as they are bound by zero and thus defined for only positive values. This feature makes the distribution often skewed to the right. This helps grasp higher Vs_{30} values and consider rock conditions.

To examine the impact of lognormal distributions on the sensitivity of VoI, we computed EVPI for four different lognormal distributions. The scale parameter was held constant while the shape parameter varied between distributions, $\sigma_{\ln(Vs30)}$ (0.2, 0.3, 0.4 and 0.5). Figure 4.24 displays (a) the associated probability density functions and (b) EVPI. Results are also compared to a normal distribution of $\sigma = 120\,m/s$. EVPI increases with shape parameter as the shape parameter leads to a larger range of possible values which translates lower prior knowledge.

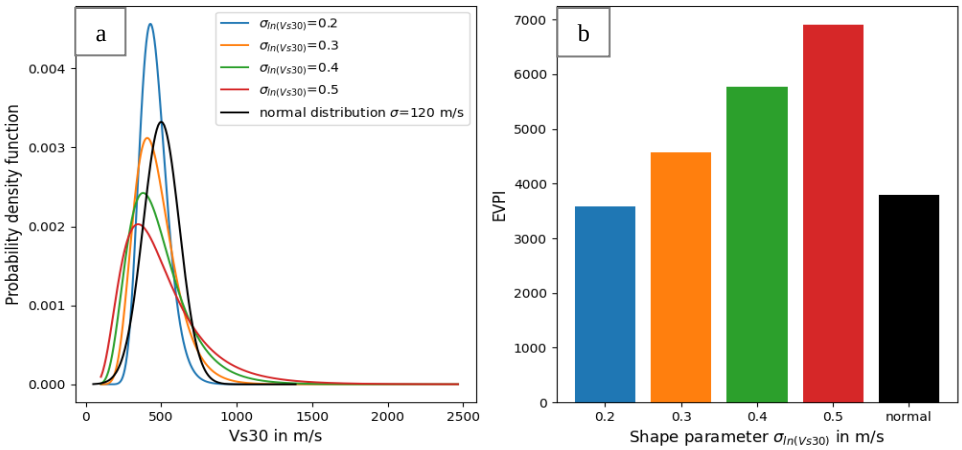

Figure 4.24: VoI for lognormal distributions of different shape values (a) probability density functions (b) EVPI

4.4.5 Conclusion

Using continuous variables to compute the VoI is more complex and has higher computational cost than using discrete variables. Nevertheless, assuming the Vs_{30} prior distribution as continuous is more realistic because it considers a large set of possible values.

The definition of the prior distribution and likelihood have shown an impact on the posterior distributions, where the level of confidence in an observation is dependent on the marginal probability of that same observation. This stresses the requirement to carefully estimate the latter based on solid evidence to insure reliable results. By utilising this approach, it becomes possible to establish a threshold level of uncertainty below which further investigation is not required before making a decision. This threshold can help streamline decision-making processes and reduce unnecessary expenditures of time, effort and resources on collecting additional data.

Tests with higher accuracy and precision are shown to increase the VoI. It is crucial to bear in mind that more efficient tests and surveys are usually more costly. The VoI permits consideration of all aspects of a test's expenses in terms of time, budget and resources, to help the decision-maker choose an optimal decision. The optimal decision does not only represent the collection of information but also the main goals of a project (e.g., choosing an appropriate seismic design, deciding on a post-earthquake evacuation, retrofitting).

4.5 Main findings and discussion

A VoI estimation method has been developed to infer the optimal seismic design for a structure when one key site parameter is uncertain. The framework was developed progressively, starting with simpler to more complex configurations, where we first consider a single input parameter to have discrete prior probabilities and then defined by continuous probability distributions. While the basic definition of VoI in the literature is simple and intuitive, implementing it within earth science can be challenging. Here, we built and implemented a method to compute VoI but also to understand the sensitivity of VoI to different variables and parameters. This study was essential to implement the theory and the computation, via Python, as well as to identify the parameters that mostly influence the VoI results.

Sensitivity analyses showed that EVPI is closely tied to the inputs and that the measure represents the maximum amount that a decision-maker would be willing to pay in order to obtain a perfect information. However, in seismic hazard assessment and earth sciences in general, the EVII is more likely to be used as it accounts for the variabilities associated with the type of investigation to be conducted to infer information. Sensitivity analyses were a useful tool to validate the approaches and to highlight the importance of correctly defining some input parameters. Overall, the prior and likelihood distributions were identified to most impact the VoI estimates.

Although defining Vs_{30} as a continuous parameter is more realistic, the discrete case methodology can be used in other applications such as computing VoI for the identification of soil classes and considering the activity or not of a neighbouring fault within a seismic source model.

In this chapter, we assumed a simple approach to estimate VoI by introducing the soil properties through a single proxy, $Vs_{30,}$ to estimate a frequency independent site amplification. This proxy has been at the subject of active debate with regards its use in the field of SHA and ground-motion prediction. This discussion is related to the fact that it represents the effects of shallow (<30m) geology. For multi-layered and/or deep profiles, Vs_{30} fails to capture the different impedance contrasts and nonlinear properties (e.g., Boore et al., 2011). Nevertheless, these computations are included

here to develop a step-by-step approach and to provide the reader with a preliminary understanding of the sensitivity of VoI to the different input parameters.

In the next chapter, the site amplification factor is obtained from 1D linear soil-response analysis, first performed analytically for a single layer profile and then using an appropriate software for a more complex profile. Based on real data, 1D V_s profiles are randomised to reflect uncertainties. Finally, the VoI approach is improved to tackle a less straightforward configuration: the presence of uncertainties in more than one parameter.

5 VoI for bivariate uncertain parameters within site-specific probabilistic seismic hazard assessment

5.1 Introduction

We are interested in direct applications of the Value of Information (VoI) regarding parameters within site-specific soil-response analyses (SRA). The scope of this chapter is similar to the previous chapter, where the decision-maker wants to decide on collecting additional information to reduce uncertainties, and to ultimately determine an optimal design for a particular building.

In the previous chapter, site effects are captured through only a single proxy, Vs_{30}, used to compute a frequency-independent amplification factor, *AF*. The hazard on soil was simply estimated by multiplying *AF* by the hazard on rock at zero spectral period (i.e., peak ground acceleration, PGA) obtained from previous Probabilistic Seismic Hazard Assessments (PSHAs). While Vs_{30} can be a useful indicator on soil properties to include in ground motion prediction equations (GMPEs) or in estimating *AF*, either directly (e.g., Bindi et al., 2014) or in defining the soil class (e.g., Berge-Thierry et al., 2003), it fails to capture all the features of site conditions such as impedance contrasts and nonlinear properties (e.g., Boore et al., 2011).

In this chapter, we improve the previous VoI approach to enable performing a more standard approach in site-specific PSHA. The methodology includes several improvements:

- Site effects are captured through 1D shear-wave velocity profiles (V_s), layer density ρ, including V_s uncertainties (σ_{V_s}) and thickness (σ_H) at each layer.

- Linear elastic SRA are performed both analytically, using a single-layer soil profile, and numerically, for a multi-layer profile through the vertical propagation of shear-waves.
- Site-specific PSHA is conducted using a hybrid approach including a frequency-dependent *AF* (Cramer, 2003).
- VoI is computed for two uncertain variables (V_s and H) by introducing bivariate joint distributions.

The chapter starts by an overview of state-of-the-art approaches to incorporate site effects in PSHAs. Then, the site-specific seismic hazard approach to be used is introduced and its inputs outlined. The following two sections present the methodology and results of VoI calculations when performing: (a) analytical linear SRA using a single-layer soil profile incorporating two sources of uncertainties and considering several types of observations, and (b) numerical linear SRA for a multi-layer uncertain profile.

5.2 Overview on methods to incorporate site effects in site-specific PSHAs

The level of complexity of the VoI method strongly depends on the approach used to integrate site effects. Some approaches have frequency-independent outputs, for some they are frequency-dependent and, others are dependent on both frequency and input motions (i.e., accounting for nonlinear site effects). Moreover, the number of uncertain variables to be included can make VoI less straightforward, requiring a more complex approach to isolate the contribution of the target parameters (i.e., that aimed to be better constrained through data collection) on the overall variability in the results. For VoI to be applied within modern practices, rigorous site-specific PSHAs should be included when building the approach.

Several authors have worked toward developing methods to estimate hazard curves (HC) and Uniform Hazard Spectra (UHS) that include site effects (Kramer, 1996; Bazzurro and Cornell, 2004; Stewart *et al.*, 2006; Rathje *et al.*, 2015; Barani and Spallarossa, 2017; Aristizábal, 2018; Aristizábal *et al.*, 2018). These methods have

been split into different categories depending on their level of complexity. These categories are summarised in Table 5.1. A more detailed classification can be found in the doctoral thesis of Aristizábal (2018).

The approaches are separated into the following two main categories:

Generic or partially site-specific approaches (level 0)

Classic PSHA is performed by using GMPEs or *AF*s that integrate averaged site effects. These site effects are usually approximatively described through one or several proxies (e.g., Vs_{30}, site correction factors). This approach suffers from high aleatory variability emanating from the selected GMPEs.

Site specific approaches (level 1 and 2)

These approaches consider a more refined estimation of the site amplification, based on measurements or numerical analyses. They allow the integration of more detailed site properties (e.g., soil V_s profiles, degradation curves, and 1D, 2D or 3D site effects) as well as their associated uncertainties. Site-specific approaches are categorised in two main levels:

- Level 1 uses a frequency dependent amplification factor, $AF(f)$, estimated through instrumentally assessing site-specific residuals (e.g., Ktenidou et al., 2015) or Standard Spectral Ratios (SSR) (e.g., Raptakis et al., 1998), or through numerical simulations of weak ground-motions by performing linear SRA. At this level, soil nonlinearity effects are neglected.
  The hazard on soil is computed using the Partially Probabilistic Hybrid Method (PPHM) based on disaggregation (Cramer, 2003).
- Level 2 is generally based on numerical simulations and considers soil nonlinearity effects. The hazard on soil can be estimated using PPHM or the full convolution method (Bazzurro and Cornell, 2004). This requires computing the frequency- and ground-motion dependent *AF*, $AF(f, S_a)$, from linear equivalent/nonlinear SRA. Other fully probabilistic approaches can be found in Table 5.1 (e.g., level 2b-d-e).

In the next section, we detail Level 1c (Table 5.1). This approach is used in this VoI application to better capture the site-effects and their variability within the estimation of the hazard on soil (PGA).

Table 5.1: Methods for the integration of site-effects in PSHA (Modified from Aristizabal et al., 2022)

	Generic or partially Site-specific approaches Level 0		**Site-specific approaches**							
			Level 1: Linear site-response $AF(f)$ Ground-motion independent			**Level 2: Nonlinear site-response $AF(f, S_a)$ Ground-motion dependent**				
	0a	0b	1a	1b	1c	2a	2b	2c	2d	2e
Site effect integration method	Proxy in GMPE	Proxy in GMPE and AF	Partially probabilistic Hybrid method (Cramer, 2003) $UHS_{rock} \times AF(f)$			Hybrid method	Full probabilistic stochastic method	Analytical approximation of the full convolution method	Full probabilistic Classical PSHA with site-specific GMPE	Full Probabilistic integration-based Method
			$AF(f)$ from site-specific residuals	$AF(f)$ from SSR with rock reference station outcropping or downhole	$AF(f)$ from numerical simulation of weak ground motion	$AF(f, S_a)$ from numerical simulation of strong ground motion	Full PSHA	Classic PSHA rock * $AF(f, S_a)$	Classic PSHA soil	Stochastic PSHA soil

5.3 Site-specific PSHA approach

In this chapter, the incorporation of site effects in PSHA is performed using the Partially Probabilistic Hybrid Method (PPHM) where a deterministic site-response is combined with the probabilistic rock hazard (Cramer, 2003). The site-specific hazard is numerically computed with the approach in Level 1c and detailed in Aristizábal (2018). This approach, also referred to as the traditional approach (e.g., Rathje et al., 2015), consists of the multiplication of the UHS on a reference rock, obtained from previous PSHA studies for rock conditions, by a frequency-dependent amplification function resulting from a 1D, 2D or 3D SRA (Equation 5.1).

Within this approach, the soil response is usually considered linear (i.e., independent of the rock-motion level). This is equivalent to neglecting nonlinear site effects and assuming that the soil dynamic properties are not affected by the shear strains (e.g., input-motion intensity). This assumption is only valid for weak ground motions.

The UHS on soil, UHS_{soil}, is computed using the hybrid method as follows:

$$UHS_{soil}(f) = UHS_{rock}(f) \times AF(f) \tag{5.1}$$

where UHS_{rock} is the UHS on the reference rock at a particular annual frequency of exceedance and $AF(f)$ the linear amplification function at a frequency f representing the ratio of the spectral response between the soil surface and the outcropping reference rock.

In this application, the considered building is the same as in the previous application in Chapter 4: a four-storey three-bay reinforced concrete (RC) building. The hypothetical building location is chosen to be in the city of Mirandola, Italy. The city, situated in northern Italy in the region of Emilia Romana, has moderate to high seismic activity (Figure 5.1). This region was shaken by the Emilia earthquake of magnitude 6.1 in May 2012, which resulted in several fatalities and severe structural damage. This location is selected because of the availability of geophysical, geotechnical and geological studies aiming at characterising several sites in the region. Associated to this, hazard curves for rock sites are available in the 2020 updated European Seismic

Hazard Model (ESHM20) database (Danciu *et al.*, 2021), enabling us to carry out a site-specific PSHA.

The ESHM20 database provides hazard curves for rock conditions (Vs_{30}=800m/s), indicating probabilities of exceedance in 50 years. These curves are converted to annual frequency of exceedance (AFoE) using the assumption of earthquakes occurrence being a Poisson process (Cornell, 1968; Reiter, 1990; Stein and Wysession, 2003). The probability of exceeding a ground motion intensity *IM* in t years is expressed as follows:

$$P(z > IM) = 1 - e^{-\lambda t} \tag{5.2}$$

where λ is the annual rate/frequency of exceedance and is then calculated as follows:

$$\lambda = -\frac{\ln(1-P(z>IM))}{t} \tag{5.3}$$

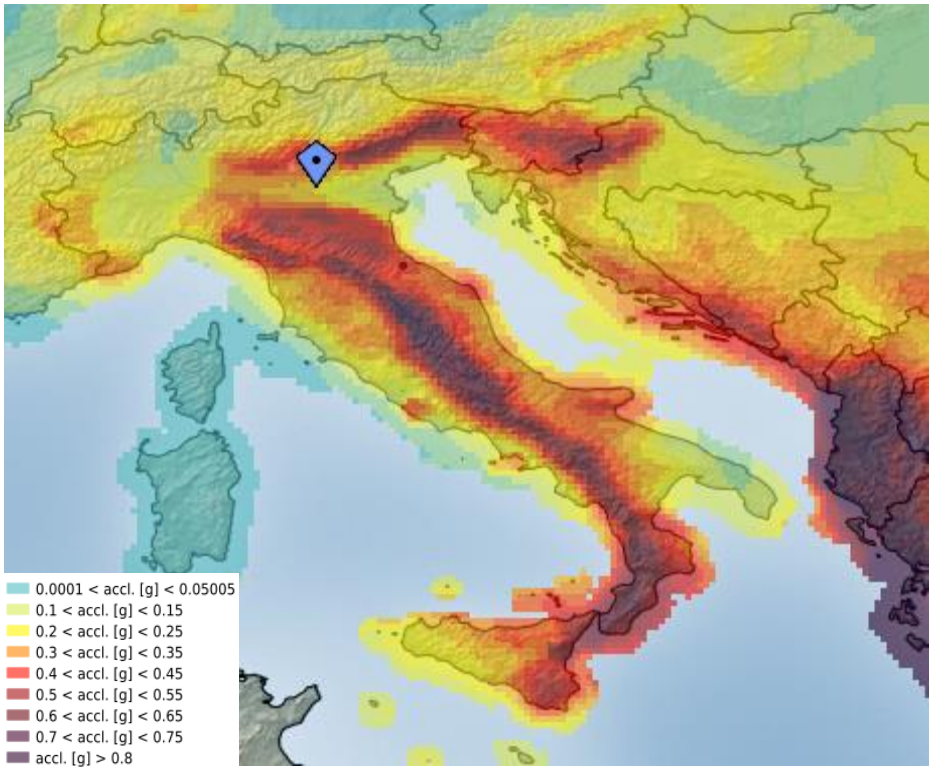

Figure 5.1: Hazard map from ESHM20 database showing Mirandola's location and the spectral accelerations at 0.2s with 10% in 50 years of probability of exceedance (Return period of 475 years) (from Danciu et al., 2021)

The PGA hazard curve for Mirandola converted to AFoE is shown on Figure 5.2. Similar to Chapter 4, the hazard on soil is only expressed in terms of PGA. The considered building is aimed to be designed for a higher target hazard level, associated with an AFoE= 10^{-4} (i.e., return period of 10,000 years). This AFoE is usually used when considering critical facilities (HSE, 2009).

Finally, the UHS_{rock} to include in Equation (5.1) is then simply the PGA value reported on the rock hazard curve for AFoE= 10^{-4}, called hereafter PGA_{rock} and estimated at 1.09g.

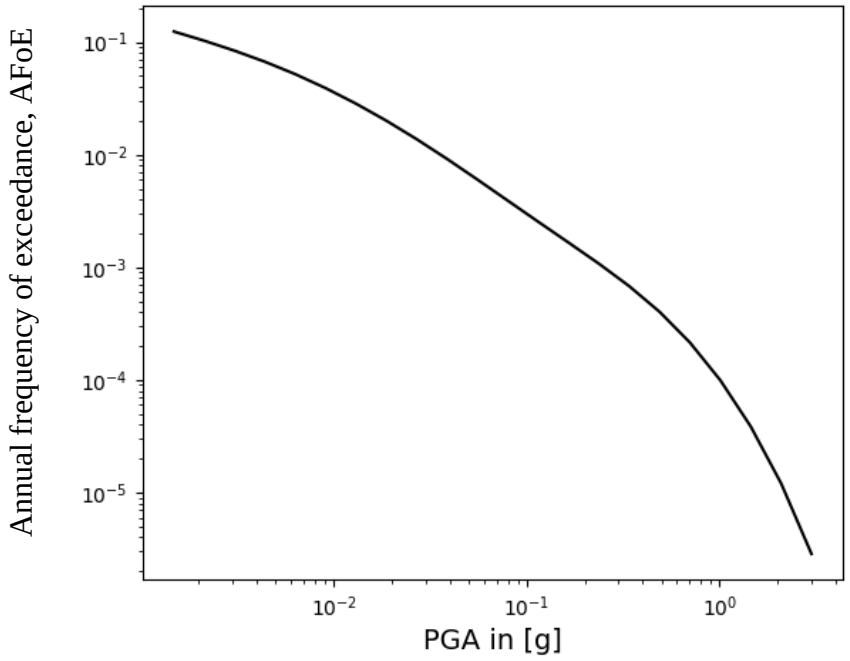

Figure 5.2: Hazard curve at PGA for the Mirandola site, Italy

The next step is to estimate AF at PGA and to quantify its variability. Soil properties are considered to represent the only source of uncertainties within this application. VoI assessments are targeted to estimate the benefits of reducing such uncertainties.

5.4 VoI for single-layer profile: Analytical Soil Response Analysis

In this section, uncertainties in soil properties are integrated by analytically performing linear site-specific SRA for a single layer over an elastic bedrock profile. VoI is computed for uncertain parameters such as V_s and thickness (H), considered to

represent the single layer properties. The results are compared to VoI estimates from direct measurements of the amplification factor.

5.4.1 Analytical linear elastic soil-response analysis

To study the impact of reducing uncertainties of specific soil properties such as the V_s or the thickness (H), we perform 1D linear elastic SRA for a single layer over an elastic bedrock. Ultimately, the influence of reducing these epistemic uncertainties is shown through the computation of VoI.

Calculation of the amplification factors

Linear elastic SRA carry the assumption that weak ground motions do not affect the soil dynamic properties and that the resulting amplification factor is independent of the ground-motion intensity. Therefore, only one input motion is needed. This motion is vertically propagated through the uniform soil layer profile to estimate the spectral acceleration at the surface.

First, the transfer function, *TF*, is analytically computed at each angular frequency (ω), assuming a 5% damping in the single layer over an elastic bedrock considered as an infinite half space:

$$|TF(\omega)| = \frac{1}{|\cos(k_z H) + i k_a \sin(k_z H)|} \tag{5.4}$$

with $k_a = \frac{\rho_L V_{s_L}}{\rho_b V_{s_b}}$, translating the impedance ratio including the bedrock, ρ_b, and the layer, ρ_L, densities. k_z is the wave number and is expressed as follows:

$$k_z = \frac{\omega}{Vs_b} \tag{5.5}$$

The density ρ_b and the shear-wave velocity of the bedrock Vs_b, considered as a sedimentary rock, are fixed to 2.25 g/cm^3 and 1500 m/s, respectively (Hu et al., 2021). ρ_L is computed according to the Al Atik and Abrahamson (2021) ρ-Vs correlation:

$$\rho = 1.742 + 0.2875\, V_s(km/s) \tag{5.6}$$

Introducing the damping, ξ, is achieved by considering a complex expression of bedrock (V_{s_b}) and soil layer (V_{s_L}) velocities as follows:

$$Vs_b^* = Vs_b(1 + i\xi) \quad (5.7)$$

$$Vs_L^* = Vs_L(1 + i\xi) \quad (5.8)$$

Damping can be introduced into the SRA by replacing the layer V_s by their complex expressions.

Finally, the soil response is analytically computed as follows:

- The input motion is transformed in the Fourier domain using the Fast Fourier Transform (FFT).
- For each frequency, the Fourier series of the input motion is multiplied by the appropriate value of the transfer function to obtain the output Fourier series
- The output is transformed to the time domain by using the inverse FFT.

The amplification function, $AF(f)$, is then the ratio, at each frequency, of the response spectrum at the soil surface and the response spectrum at the top of the bedrock, assumed here to be the same as the outcropping rock.

Following Level 1c, the site-specific seismic hazard using the hybrid method simply consists of multiplying $AF(f)$ by the hazard curve on rock at the same spectral frequency f and for a particular annual probability of exceedance.

Uncertain parameters

Unlike the previous chapter where only one parameter was uncertain (Vs_{30}), we consider here both the soil layer V_s and thickness H to be uncertain. We define the associated hypothetical normal prior probability distributions as follows:

$$V_s \approx N(\mu_{Vs}, \sigma_{Vs}^2) = N(400, 60^2) \text{ in m/s} \quad (5.9)$$

$$H \approx N(\mu_H, \sigma_H^2) = N(20, 5^2) \text{ in m} \quad (5.10)$$

where μ and σ represent the mean and the standard deviation of the considered parameter, respectively.

The uncertainties in parameters V_s and H induce a variability in the overall AF, which in turn is reflected in the final hazard on soil estimates. To capture this variability, AF is computed for all possible V_s-H combinations.

To do so, 1,000 values are uniformly sampled from each uncertain soil property, V_s and H, and AF is computed following the previous framework for each possible (V_s, H) combinations. The resulting AF at PGA is shown in Figure 5.3, where the x and y axis represent H and V_s, respectively. Note that the thickness H and the bedrock depth (x axis) are equivalent for a single layer profile. AFs values are indicated by a greyscale, ranging from 0.8 to 1.8.

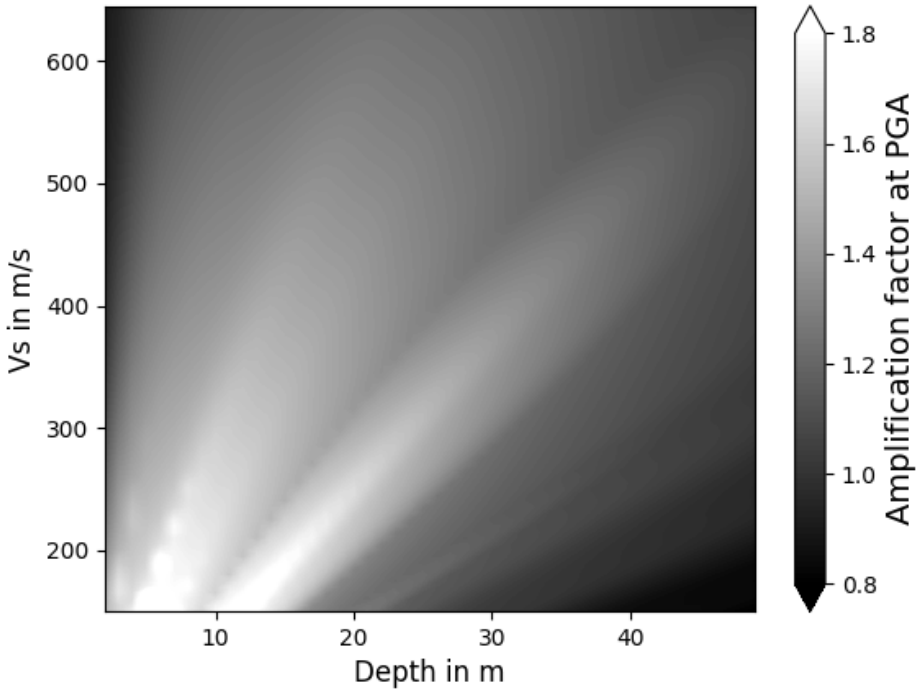

Figure 5.3: Amplification factor matrix at PGA computed for all combinations of (V_s, H) couples from analytical 1D linear elastic soil-response analysis

At PGA, the highest amplification mostly occurs for thin layers and low V_s. In other words, shallow sites composed of soft materials amplify most the ground motion at the surface. Conversely, the site amplification decreases when the thickness and/or V_s increases. Moreover, de-amplification (i.e., AF<1) is noticed for layers where V_s >600 m/s and 2m<H<5m as well as for H>40m and low V_s (< 200 m/s).

The calculation of AF prior to applying the VoI methodology is essential for reducing the computational cost. This grid/matrix of computed AFs will be used next to compute VoI for two different types of measurements.

5.4.2 Value of Measurements of Profile characteristics

As demonstrated above, *AF*s are strongly dependent on the single layer properties, V_s and *H*. *AF* variability induced by these uncertain properties can be linearly translated in the estimation of the PGA hazard on soil when computed using the Level 1 approach for site-specific PSHA (i.e., the hybrid method). The variability in the hazard estimates gives rise to less straightforward decision-making when choosing an optimal design for the structure. This variability might be reduced when additional measurements on the sources of uncertainty are obtained.

We want to estimate the value, in monetary units, of collecting data on V_s, H or both. The challenge is to develop the approach in Chapter 4, defined for a single uncertain parameter, to a more complex situation of having two uncertain parameters. This will allow analysing, in terms of VoI, the impact of inferring V_s and H on the overall calculations and will enable deciding on prioritising the measurement of one parameter over another.

Prior joint probability density function

While Chapter 4 solely focused on univariate distributions due to the identification of only one source of uncertainty, the present analysis requires the consideration of prior probability distributions for two parameters. Because both parameters are simultaneously used within the framework, it is essential to define a common joint distribution corresponding to all possible pairs of V_s and H, respecting each individual distribution. Moreover, a joint distribution helps to perform the sampling of (V_s, H) couples through Monte Carlo simulations to approximate the integrals within the VoI calculations.

As V_s and H are described by normal distributions, the prior joint distribution results in a bivariate normal distribution. The prior joint probability density function (*pdf*) of the normally distributed variables V_s and H, denoted as $f_{prior}(V_s, H)$, is expressed as follows:

$$f_{prior}(V_s, H) = \frac{1}{2\pi\sigma_{Vs}\sigma_H} e^{\left(\frac{Vs-\mu_{Vs}}{\sigma_{Vs}}\right)^2+\left(\frac{H-\mu_H}{\sigma_H}\right)^2} \quad (5.11)$$

f_{prior} is parametrised by a mean vector, *mean*, and a covariance matrix, *cov*. The two random variables V_s and H being uncorrelated, the covariance matrix is diagonal and expressed as follows:

$$cov = \begin{bmatrix} \sigma_{Vs}^2 & 0 \\ 0 & \sigma_H^2 \end{bmatrix} \quad (5.12)$$

and the mean as

$$mean = \begin{pmatrix} \mu_{Vs} \\ \mu_H \end{pmatrix} \quad (5.13)$$

V_s and H joint *pdf* is shown in a 3D plot in Figure 5.4-a, where x and y axes indicate H and V_s, respectively. Since the two variables have no covariance, there is no clear trend here. The distribution is a circle with a centre point at coordinates (μ_{Vs}, μ_H).

Next, Monte Carlo simulations are performed to generate samples according to the joint *pdf*. An example of a sampling of 10,000 (V_s, H) couples is shown in Figure 5.4-b, where each red dot represents a (V_s, H) couple. We can clearly see that the number of samples depends on the value of the joint *pdf*. Indeed, the highest number of samples is at the centre and decreases for lower joint *pdf* values (i.e., towards the dark background).

Figure 5.4 Figure 5.4-c represent a comparison between the prior (red dots) and the posterior sampling when uncertainties on V_s and H are reduced by 50% for two different observations of the (V_s-H) couple. The posterior estimation is further developed in the following subsection.

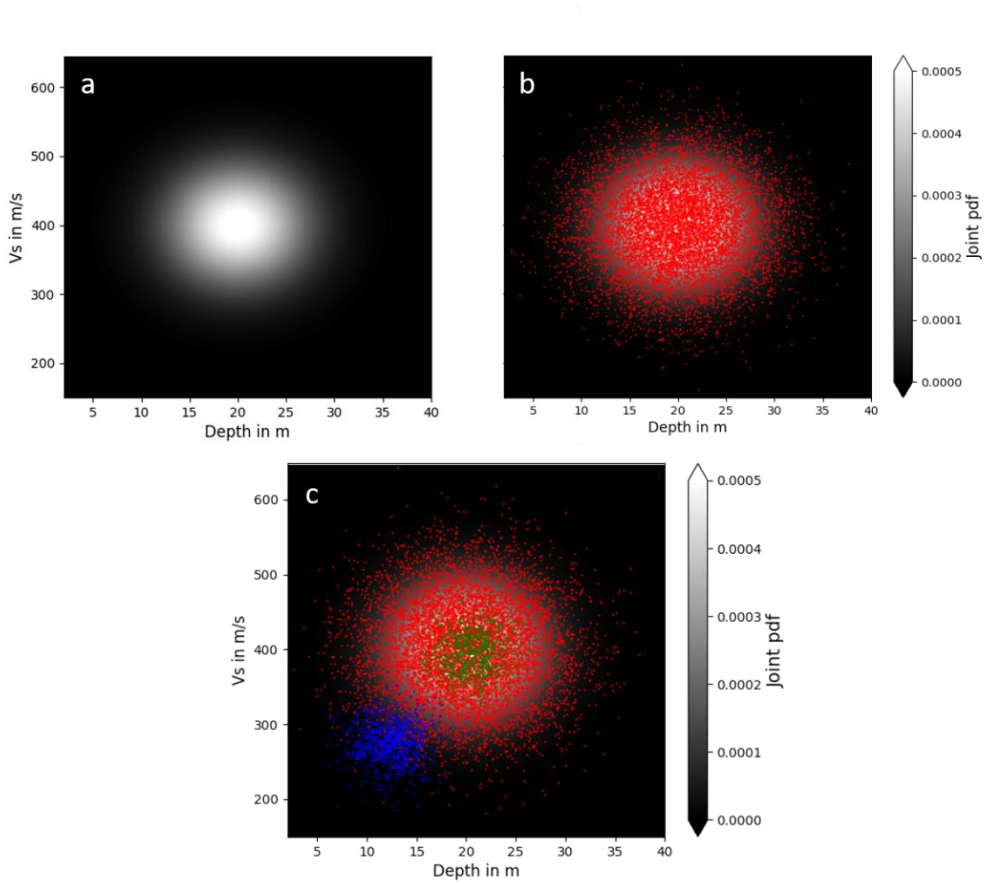

Figure 5.4: Monte Carlo sampling process. (a) Prior joint pdf, (b) Prior sampling (red dots), (c) Posterior sampling for 50% uncertainty reduction and 2 different observations (blue and green dots)

VoI methodology

The purpose of this application is to study the impact of uncertainty reduction on each of the uncertain parameters, but also on both simultaneously. We assume that geophysical/geotechnical investigations might infer different degrees of knowledge on V_s and/or H. This degree of knowledge is considered here as a percentage, X, of the uncertainty reduction. For instance, if a test is performed with a likelihood of reducing the uncertainties on V_s by X %, the error distribution function, e, assigned to the test is expressed as follows:

$$e \approx N\left(0, \left(\sigma_{V_s}(1-X)\right)^2\right) \tag{5.14}$$

e is chosen as a normal distribution with mean 0 (i.e., unbiased test) and standard deviation $\sigma_{V_s}(1-X)$.

The following VoI approach is developed for three possible configurations:

1- Conducting a perfect/imperfect test only on V_s to reduce uncertainties by X %.
2- Conducting a perfect/imperfect test only on H to reduce uncertainties by Y %.
3- Conducting a perfect/imperfect test on both V_s and H to reduce uncertainties by X % and Y%, respectively.

The decision tree in Figure 5.5 illustrates the workflow of inferring the expected outcomes resulting from these different scenarios, along with the representation of the decision-making in applying a design based on the available information. The marginal probabilities of the three types of observations are shown as red chance nodes (circles) along with an example of the used posterior probabilities given a random observation. Finally, the expected outcomes to be used for VoI calculations are computed using the joint posterior distributions (black nodes) for each decision alternative d_i.

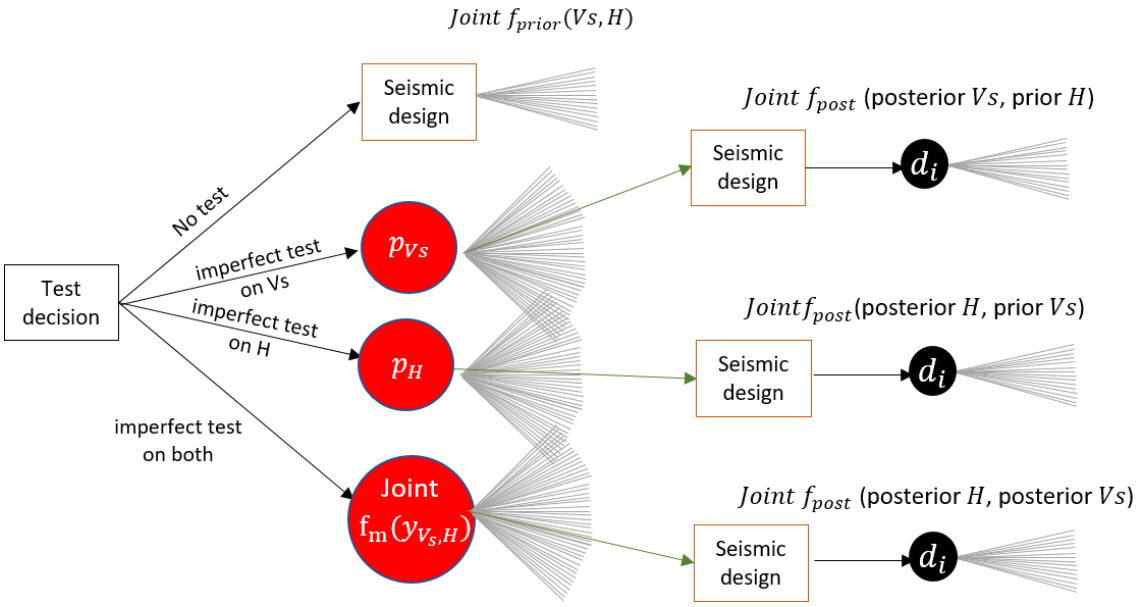

Figure 5.5: Decision tree for the estimation of VoI when collecting data on V_s, H or both. Red chance nodes represent marginal probabilities, black chance nodes represent joint posterior probabilities. Rectangle nodes represent the decisions. d_i is the decision alternative on seismic design

While the definition of the expected losses remains unchanged (see Chapter 4, section 4.3.1), the methodology is further developed to integrate the prior as a joint distribution. The detailed steps to compute the prior value (*PV*), the Expected Value

of Perfect Information (EVPI) and the Expected Value of Imperfect Information (EVII) on V_s and/or H are as follows:

1- Prior Value (PV)

- Compute prior joint distribution f_{prior} and the associated *mean* and *covariance*
- Monte Carlo sampling from f_{prior} (*n* samples)
- Associate each sample (couple) with the previously computed AF matrix (Figure 5.3)
- Estimate the hazard on soil, PGA_s, according to Equation (5.1) for each sample. Determine the probability distribution of PGA_s, f_{PGAs}.
- Select $d \in D$ potential seismic designs. This can be done by observing the variability of the resulting PGA_s and selecting alternative seismic designs.
- Calculate the *PV* as follows:

$$PV = \max_{d_i \in \mathrm{D}} \{ \int o(d_i, x)\, f_{PGAs}(x) dx \} \sim \max_{d_i \in \mathrm{D}} \{ \frac{1}{n} \sum_{k=1}^{n} o(d_i, x_k) \}$$

$x = \{x_1, .., x_k, .. x_n\}$ represent the values of the computed PGA_s describing f_{PGAs}. $o(d_i, x)$ is the expected losses function for a true state x when applying design d_i. The function is defined in Chapter 4, Equation (4.20).

2- Assessing the EVPI

To compute the EVPI on V_s and H (i.e., conduct perfect test to infer V_s and H), we estimate the posterior value (*PoV*) as follows:

$$PoV = \int \max_{d_i \in D} \{ o(d_i, x) \} f_{PGAs}(x)\, dx \sim \frac{1}{n} \sum_{k=1}^{n} \max_{d_i \in D} \{ o(d_i, x_k) \}$$

EVPI is then simply

$$EVPI = PoV - PV$$

3- *Assessing the EVII*

a- Test error functions

- Define the reduction in uncertainties, that could be achieved from conducting a given test, for each of V_s and H distributions and compute the new standard deviations, σ_{Vs_T} and σ_{H_T} respectively.
- Use σ_{Vs_T} and σ_{H_T} to define test error functions for V_s measurements, e_{Vs} , and H measurements, e_H, respectively. (see Equation 5.14)

b- Observations

- Simulate potential observations for each of V_s and H:

  Pick random error e^*_{Vs} from e_{Vs} (similarly e^*_H from e_H)

  Pick random V_s^* from $N(\mu_{Vs}, \sigma^2_{Vs})$ (similarly H^* from $N(\mu_H, \sigma^2_H)$)

  Compute the observation V_s^*+e^*_{Vs} (similarly $H^* + e^*_H$)

  Repeat to obtain enough observations describing the overall distribution, representing the associated marginal probability distribution

  $$p_{Vs} \sim N(\mu_{Vs_m}, \sigma_{Vs_m}) \text{ and } p_H \sim N(\mu_{H_m}, \sigma_{H_m})$$

- Define the joint marginal probability density function f_m of mean $mean_m$ and covariance cov_m as follows:

$$mean_m = \begin{pmatrix} \mu_{Vs_m} \\ \mu_{H_m} \end{pmatrix} \tag{5.15}$$

$$cov_m = \begin{bmatrix} \sigma^2_{Vs_m} & 0 \\ 0 & \sigma^2_{H_m} \end{bmatrix} \tag{5.16}$$

- Perform Monte Carlo simulations to sample N observations of (V_s, H) couples from f_m.

c- Posterior Value (PoV)

- For each bivariate observation $y_{(Vs_j, H_j)}$ $(j \in \{1, .., N\})$:

  Compute posterior *pdf*s (using Bayes' rule) for each of the observations Vs_j and H_j.

  Identify the mean and the standard deviation of each *pdf*.

Calculate the posterior joint distribution $f_{post}(y_{(Vs_j,H_j)})$ using the identified means and standard deviations. Estimate the mean, $mean_{post}$, and the covariance matrix, cov_{post}.

Sample m posterior couples (V_s^l, H^l) from f_{post} where $l \in \{1,..,m\}$. An example of posterior sampling is given in Figure 5.4-c for two different observations.

Compute PGA_s values for each of the m posterior samples by using the amplification factor matrix in Figure 5.3 and applying Equation (5.1).

Estimate the expected outcomes, EO, conditioned on the observation $y_{(Vs_j,H_j)}$ ($j \in \{1,..,N\}$) for each decision d_i:

$$EO\left(d_i|y_{(Vs_j,H_j)}\right) = \int o(d_i,x)p\left(x|y_{(Vs_j,H_j)}\right)dx \sim \frac{1}{m}\sum_{l=0}^{m} o(d_i,x^l) \quad (5.17)$$

where x^l is the PGA_s corresponding to the posterior sample (V_s^l, H^l).

- The posterior value is then computed for each of the N observations to estimate the VoI as follows:

$$PoV = \int_y p(y).\max_{d_i \in D}\{\int_x o(d_i,x)p(x|y)dx\}\,dy$$

$$\sim \frac{1}{N}\sum_{j=0}^{N}\max_{d_i \in D}\left\{\frac{1}{m}\sum_{l=0}^{m} o(d_i,x^l)\right\}$$

d- Value of information

The EVII is then simply:

$$EVII = PoV - PV$$

Results and sensitivity analyses

We choose an arbitrary number of ten alternative decisions, corresponding to the PGA at which the building can be designed, by uniformly sampling from the obtained PGA_s in the *PV* calculations (i.e., initial variability from available knowledge). We want to study the sensitivity of VoI estimations to several types of measurements and different combinations. To achieve this, we assess the VoI by considering all possible test

accuracies (i.e., the standard deviation in the error functions and percentage of uncertainty reduction) for both uncertain parameters, V_s and H.

Practically, for a fixed e_{Vs}, e_H is varied to cover a range starting from no uncertainty reduction on H to a test that gives perfect measurements. Subsequently, VoI is assessed for each e_H. e_{Vs} is then modified and the process is repeated. The resulting VoI for all possible configurations and combinations of tests on V_s and H is shown in Figure 5.6.

This VoI sensitivity analysis is presented as a heatmap with x and y axis representing the different test accuracies and translating the percentage of expected uncertainty reduction for H and V_s, respectively. Uncertainties reductions are expressed by a percentage relative to the initial uncertainty (σ_{Vs} and/or σ_H). The estimated VoI at each point of the grid is indicated by a greyscale, ranging from 2400 to 6700 euros, where lighter colours indicate higher values. Specifically, VoI estimates at the coordinates (100,0) and (0,100) correspond to the Expected Value of Partially Perfect Information (EVPPI) on H and V_s, respectively. EVPPI is an estimate of the value of obtaining perfect information on one parameter from a subset of uncertain parameters. This plot is useful as it enables the visualisation of the impact on VoI of reducing the uncertainties of one parameter at a time but also both simultaneously.

VoI smoothly increases when there is additional information about one of the two variables. We notice that VoI is slightly more sensitive to V_s uncertainty reduction than reduction in uncertainty in H. Overall, reducing the uncertainties for V_s and H increases VoI. The red dot at the upper right corner is the EVPI previously computed when assuming perfect information about both parameters. The EVII framework applied when assuming a 100% reduction in uncertainty in both parameters gives results equal to the EVPI. This result validates the developed approach.

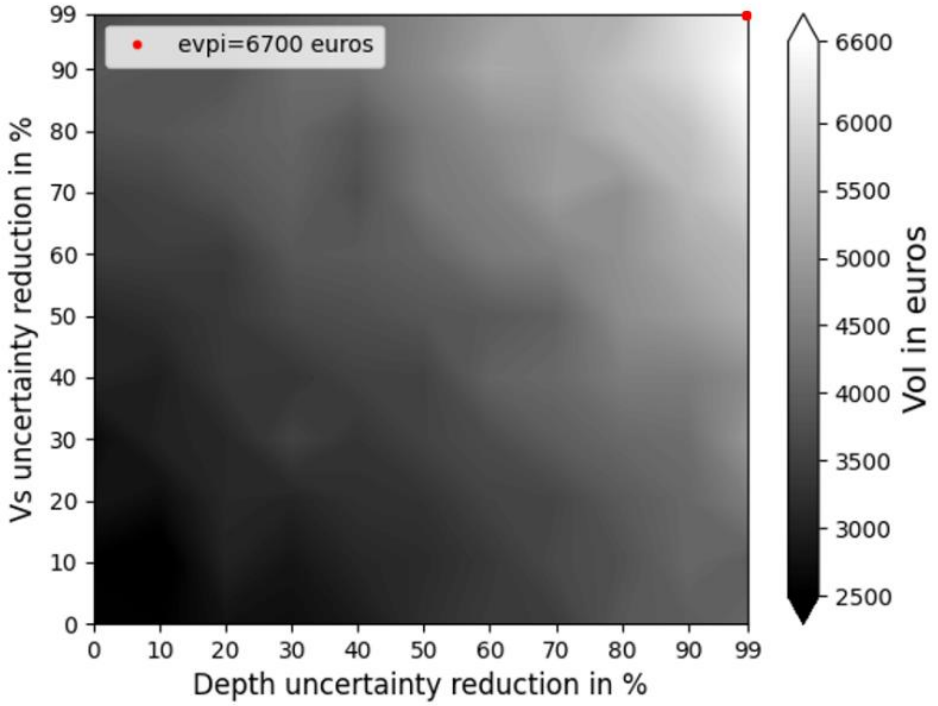

Figure 5.6: Value of Information on V_s and/or H

Another interesting finding is that in some cases, investing in better quality of measurements for only one parameter is more valuable than obtaining lower quality measurements on both parameters. For example, reducing the uncertainty only on H by 60% (i.e., no reduction on V_s) results in a higher VoI than investing in measurements that would reduce H uncertainty by 30% and V_s by 20%.

It is worth mentioning that in practice, V_s and H are not necessarily uncorrelated. In fact, most invasive and non-invasive measurements (e.g., crosshole, dowhole, multi-channel analysis of surface waves) provide information on both parameters, simultaneously. This degree of dependence of a measurement on the two parameters can be considered within this framework by defining a joint error function distribution that translates the level of correlation within its covariance matrix.

This approach can be applied to other parameters with uncertainties as a tool for estimating the value of combining different ground investigation techniques, but also as an indicator for prioritising some measurements over others.

5.4.3 Value of Direct Measurements of the Amplification Factor

In this subsection, we assume that we target measurements that would directly reduce the *AF* variability. Indeed, if geophysical/geotechnical investigations can directly or indirectly reduce the uncertainties around soil properties, some other measurements are able to directly infer the characteristic of interest here, the amplification factor. *AF*s can be instrumentally evaluated through recordings of ground motions during earthquakes using the reference site method (Borcherdt, 1970), where an experimental transfer function is estimated from spectral ratios of records from reference rock and sediment site (simultaneously), also known as the Standard Spectral Ratio (SSR) (Tucker and King, 1984; Mittal *et al.*, 2013; Gelis *et al.*, 2022). Other techniques are used to estimate site amplification such as the spectral inversion method (Iwata and Irikura, 1988) using shear-waves data from the observed ground motions during earthquakes, and the Horizontal to Vertical Spectral Ratio (HSVR) obtained from microtremor data as well as from weak and strong ground motions (Nakamura, 1989; Theodoulidis *et al.*, 1996; Carpenter *et al.*, 2018).

Here, the VoI is assessed for direct measurements on *AF*. Hence, *AF* is the main uncertain parameter to be reduced to better constrain the site hazard and enhance the decision-making process on choosing the building design level at PGA. To do so, the prior distribution of *AF* should be determined.

Prior AF distribution

The value of direct measurements on AF is assessed using the same single-layer profile, the same uncertainties on V_s and H and the same design alternatives. AF should express the variability induced by both parameters' uncertainties. To estimate this variability, we sample n=10,000 (Vs, H) couples through performing Monte Carlo simulations using the joint bivariate probability distribution of V_s and H computed in the previous section and denoted f_{prior}.

For each of these n couples, associated *AF* values are reported from the amplification matrix previously computed and shown in Figure 5.3. Figure 5.7 shows the normalised histogram of the resulting n *AF*s. *AFs* is shown to be lognormally distributed with a

median estimated at 1.18. On the same figure, we superimpose the best-fit lognormal distribution (red) estimated using Maximum Likelihood Estimation (MLE) and considered as the prior *pdf* to be used within the VoI approach.

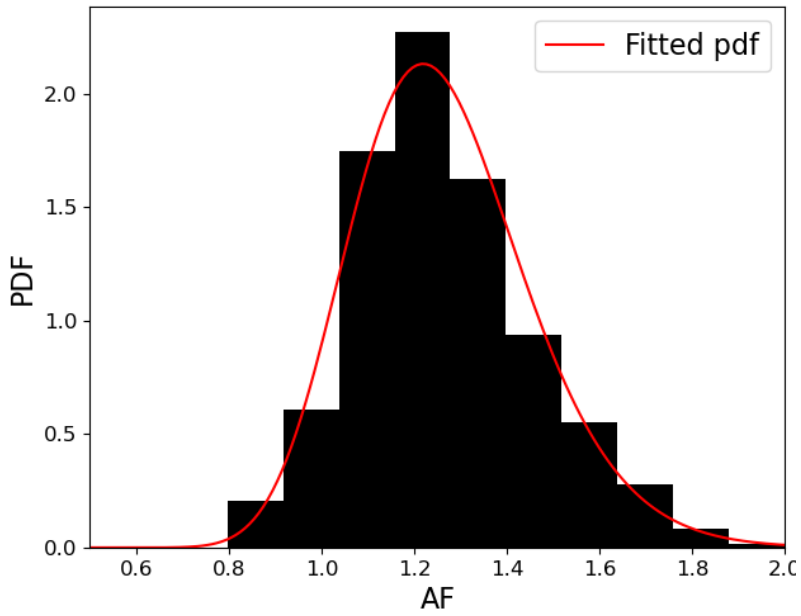

Figure 5.7: Normalised prior AFs histogram for PGA and best-fit distribution (red)

*AF*s and their distributions are considered as one of the most important inputs in estimating the hazard at the surface (e.g., Bazzurro and Cornell, 2004). Generally, *AF* is assumed here to be lognormally distributed (e.g., Li and Assimaki, 2010; Rathje et al., 2010; Bahrampouri et al., 2018). The variability in the results, obtained from analytically computing *AF* using uncertain parameters within this study, is in line with the common practice that assumes a lognormal distribution for *AF*.

Measurements error functions

Similarly, the variability in *AF* measurements obtained from techniques such as SSR and HVSR is generally assumed to follow lognormal distributions (e.g., Edwards et al., 2013). This assumption is used to define lognormal test error functions, describing the variability in the future measurements. We construct the test error distribution, *e*, based on the prior *AF* distribution which is defined by a median μ_{AF} and a shape σ_{AF} (i.e., standard deviation) as follows:

$$\mu_e = 0 \tag{5.18}$$

$$\sigma_e = \sigma_{AF}(1 - X) \tag{5.19}$$

where X translates the reduction, in percentage, of the prior AF variability after performing the test. Moreover, the obtained lognormal distribution e is shifted with the use of the *location* parameter in Python to force its median to be 0, assuming a non-biased test.

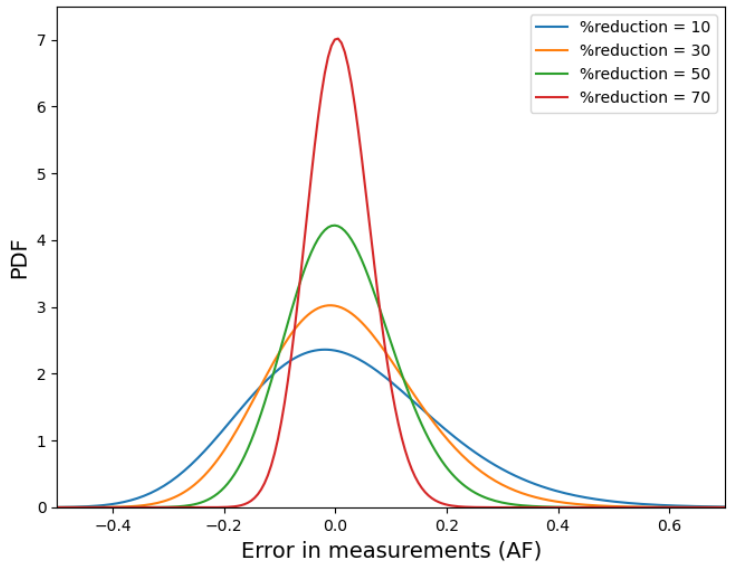

Figure 5.8: Example of the error functions used to describe measured AF variabilities from different tests

We show in Figure 5.8 four examples of error functions describing different measurement variabilities. These distributions are associated with measurements that would reduce the prior AF variability by 10, 30, 50 and 70%. Measurements with higher accuracy (i.e., higher X) have narrower and more constrained variabilities. Defining the error functions are not only important for expressing the decrease in the observed variability. Indeed, they are used as 'vessels' within the definition of the likelihood, which enables consideration of the possibility of measurements with medians that are different than the prior AF median. In the following, we aim at estimating the sensitivity of VoI to these different error functions.

Results and sensitivity analyses

The framework applied to estimate VoI is equivalent to the framework detailed in Chapter 4 when a single uncertainty is considered as continuous. Here, we want to estimate the value of obtaining direct measurements on *AF* by assessing the EVPI and EVII.

The sensitivity of VoI to the test accuracy is analysed by computing VoI for a range of potential uncertainty reductions X, ranging from 0% (i.e., same variability as prior *AF*) to 100% (i.e., measurements that perfectly estimate *AF*). The obtained estimations of VoI are displayed on Figure 5.9.

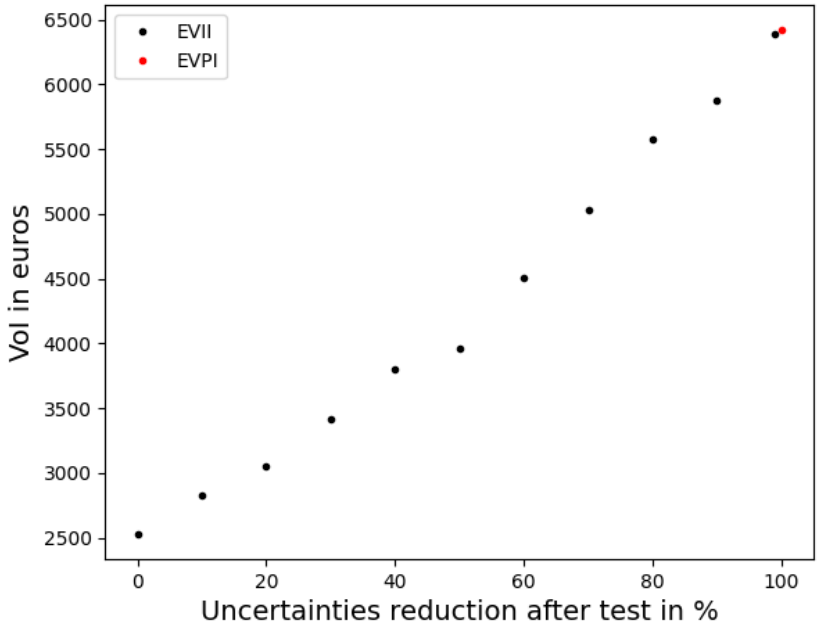

Figure 5.9: VoI sensitivity analysis to the percentage of uncertainties reduction after data collection

Similar to the previous results presented in Chapter 4, VoI increases linearly when the quantity and/or quality of measurements is higher, converging asymptotically towards the EVPI (red dot). Moreover, it is interesting to notice that for X=0% (i.e., obtained measurements have the same variability as the prior *AF*), VoI is different from zero. Intuitively, we tend to think that this measurement would be of no use in reducing the variability. Bayesian inference tells us otherwise. The explanation lies within the definition of the posterior probability, which is used in the computation of the expected outcomes. As defined in Chapter 4, the posterior probability is a combination of the

likelihood and the prior estimates. VoI would be zero if the posterior was equal to the prior. However, the likelihood of a random observation when X=0% has the same standard deviation as the prior. Hence, the likelihood brings additional information, suggesting that *AF* has the highest probability of being equal to the observation along with carrying the same variability. This is still additional information that gives value to this specific measurement.

To better illustrate this case, we compute the posterior distribution of *AF* when measurements (observations) have a median of 1.4 and the same variability as the *AF* prior variability (i.e., X=0%). Figure 5.10-a shows the posterior along with the likelihood and prior distribution. This result, valid for any median observation, shows that even when the observed variability is not reduced with additional measurements compared to the prior knowledge, the posterior distribution has lower variability than the prior, resulting in a VoI that is greater than 0.

Moreover, it would be interesting to study the posterior when an error function has a higher variability than the prior distribution. In other word, what is the impact of obtaining measurements that carry more uncertainty than the prior knowledge? We compute the posterior for the same median observation but instead, the error function has twice the prior standard deviation (i.e., $\sigma_e = 2\sigma_{AF}$). The obtained posterior distribution is shown on Figure 5.10-b and is compared to the likelihood and prior distributions. The posterior distribution has a large variability, comparable to the prior distribution but slightly lower. This type of measurement appears to slightly enhance the prior knowledge. Larger variabilities have been tested, showing that the posterior, by definition, cannot express larger uncertainties than the prior. In other words, obtaining additional data can never decrease the prior knowledge.

In reality, additional measurements with high variability might suggest a more complex state of knowledge such as specific geological features, 2D or 3D site effects and strong soil nonlinearity. Such features, which associated uncertainties have not been included in the definition of prior distributions, must be considered and VoI should be reassessed. To conclude, additional information seems to always be valuable when relevant to the characteristic of interest. However, its value might be negligeable in some cases.

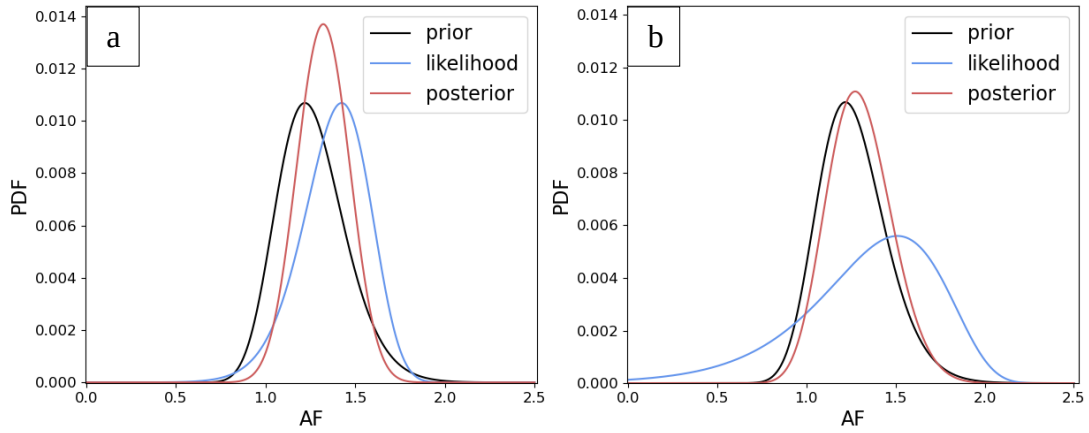

Figure 5.10: AF prior, posterior and likelihood distributions for a median observation of *AF*=1.4 when (a) X=0% and (b) $\sigma_e = 2\sigma_{AF}$

This VoI sensitivity analysis shows values similar to the results in section 5.4.2. In fact, both direct measurements on *AF* and measurements to infer V_s and H result in VoI estimations ranging from 2 500 to 6 500-6 700 euros, depending on the test accuracy. However, direct *AF* measurements are more effective when compared to measurements obtained solely on V_s or H. Nevertheless, measuring accurately *AF* from recordings requires instrument installations and a prolonged monitoring period to correctly characterise the amplification at the site of interest. VoI should be compared to the cost of such instrumentation and monitoring to make a decision.

5.5 VoI for multi-layer profile: Numerical Soil Response Analysis

In this section, the VoI is assessed to decide on investing in a technique to directly measure *AF* when the site of interest is characterised by a complex and uncertain multi-layer 1D profile. The scope of VoI remains the same as the previous section, which is minimising the expected losses when deciding on the seismic design of the same four-storey three-bay reinforced concrete building located in Mirandola, Italy.

We consider a multi-layer V_s profile obtained from site characterisation through geophysical and geotechnical measurements at a specific location in Mirandola. Barani and Spallarossa (2017) provide estimates of the soil properties, which are detailed in Table 5.2. The soil model is a six-layer profile over an elastic bedrock. V_s is defined by a lognormal distribution of median μ_{V_s} and standard deviation σ_{V_s} at each layer's mid-depth.

Table 5.2: Mirandola soil model properties (Barani and Spallarossa, 2017)

	Material	**Median Thickness** (m)	**Unit Weight** (kN^3/m)	**Median** V_s μ_{V_s} (m/s)	σ_{V_s}
Layer 1	Clay	12	17.5	180	11
Layer 2	Sand	18	18.2	270	17
Layer 3	Sand	10	18.2	475	30
Layer 4	Sand	25	18.2	288	18
Layer 5	Sand	35	18.2	400	25
Layer 6	Softer rock	20	19.1	775	48
Bedrock	Rock	∞	20	800	50

5.5.1 Building the prior

In this study, the prior distribution of *AF* at PGA is computed numerically through 1D linear elastic SRA performed with STRATA software (Kottke *et al.*, 2013). This analysis requires a single weak input motion, vertically propagated through the profile to respect the linear assumption. As a reminder, the linear assumption considers that for weak input motions, *AF* is ground-motion independent.

Soil parameter uncertainties are directly introduced in STRATA, via Monte Carlo simulations. This consists in varying the shear-wave velocity and the layer thickness following lognormal distributions using the Toro (1995) model, while the dynamic properties remain fixed (i.e., linear assumption). Performing these simulations are an essential step in considering alternative profiles and incorporating the variability in site effects.

Figure 5.11 shows the Mirandola soil model (median profile in Table 5.2) as well as 200 random profiles realisations generated using V_s uncertainties as σ_{lnVs} and a thickness uncertainty increasing with depth (Toro, 1995).

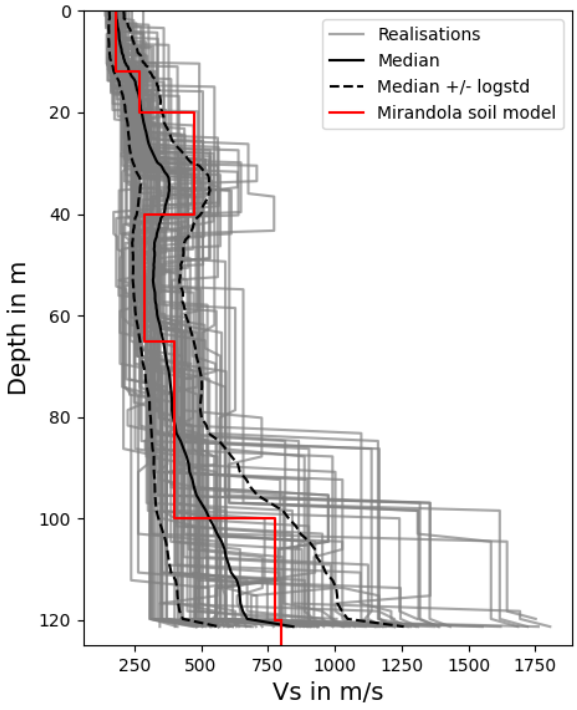

Figure 5.11: V_s profile at the Mirandola site (red) (Barani and Spallarossa, 2017) and 200 Monte Carlo realisations (grey)

1D linear elastic SRA is performed using a single weak motion for each of the profile realisations to obtain the acceleration response spectra at the surface for 5% damping. The spectral amplification functions are simply estimated by dividing each of the obtained response spectrum at the surface by the response spectrum of the input motion defined at the top of bedrock. A histogram of the 200 obtained amplification factors at PGA is displayed in Figure 5.12. AF is shown to have a moderate variability (0.65-1.85) and tends to follow a normal distribution ($i.e., N(1.37,0.26^2)$). The blue curve represents the best-fit distribution for the histogram estimated using the MLE method and is considered as the prior probability density function for AF, $f_{AF_{prior}}$.

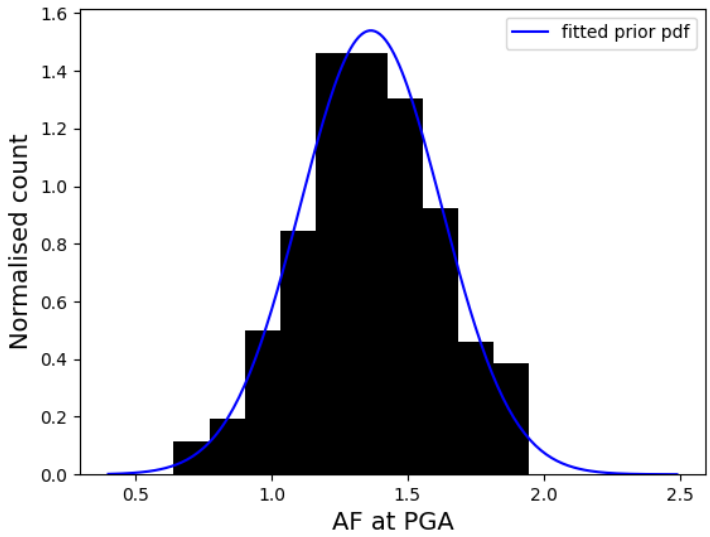

Figure 5.12: Normalised prior AF histogram at PGA (black), best-fit *pdf* (blue)

To estimate the integrals within VoI expressions, $f_{AF_{prior}}$ is used to generate $n = 10{,}000$ samples. The hazard on soil, PGA_s, is computed for each sample using Equation (5.1). The sampling is iterated to ensure a stable estimate of the prior distribution. Figure 5.13 displays the histograms of PGA_s, resulting from each Monte Carlo sampling procedure along with the best-fit distribution in black, f_{prior}. This step is important to verify that the histograms are well fitted to f_{prior}. The PGA_s variability is then used to sample, uniformly, ten values to be considered as decisions alternatives for seismic design, PGA_d.

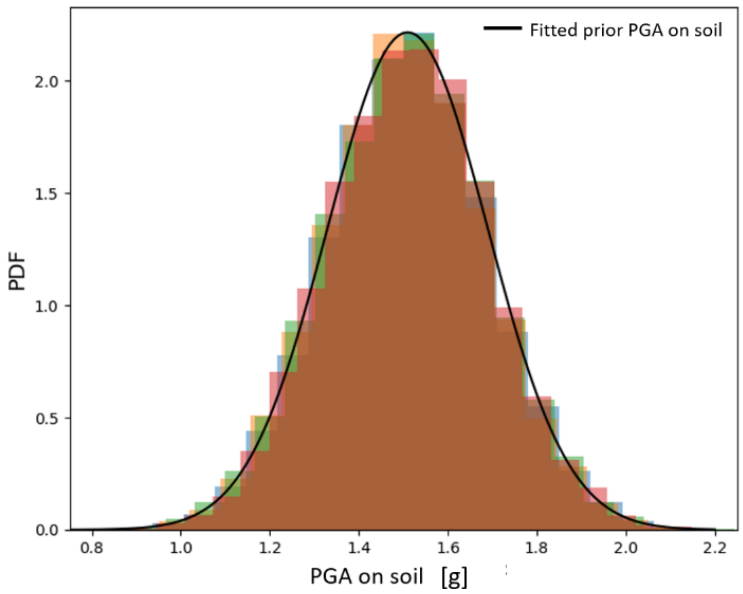

Figure 5.13: Several PGA_s histograms from Monte Carlo sampling and best-fit prior PGA_s *pdf* (f_{prior})

5.5.2 VoI results and sensitivity analyses

VoI is estimated for several potential tests, aimed at directly measuring *AF*. The *PV* and EVPI are computed using the prior PGA_s distribution, f_{prior}, and based on the approach detailed in Chapter 4 for a single parameter defined by a continuous probability distribution (here *AF*). It is worth noting that the prior used in the calculations represent the distribution of the hazard at the site, PGA_s. Although the observed variable is *AF*, the shortcut in directly using f_{prior} is correct as the latter was built from $f_{AF_{prior}}$ using the linear expression in Equation (5.1). Moreover, this helps reduce the computational costs.

Regarding the EVII, it is estimated by considering measurements defined by a normal error function $N(0, (\sigma_{PGA_s}(1-X))^2)$, where X represents the percentage in PGA_s variability reduction and σ_{PGA_s} the standard deviation of the distribution f_{prior}.

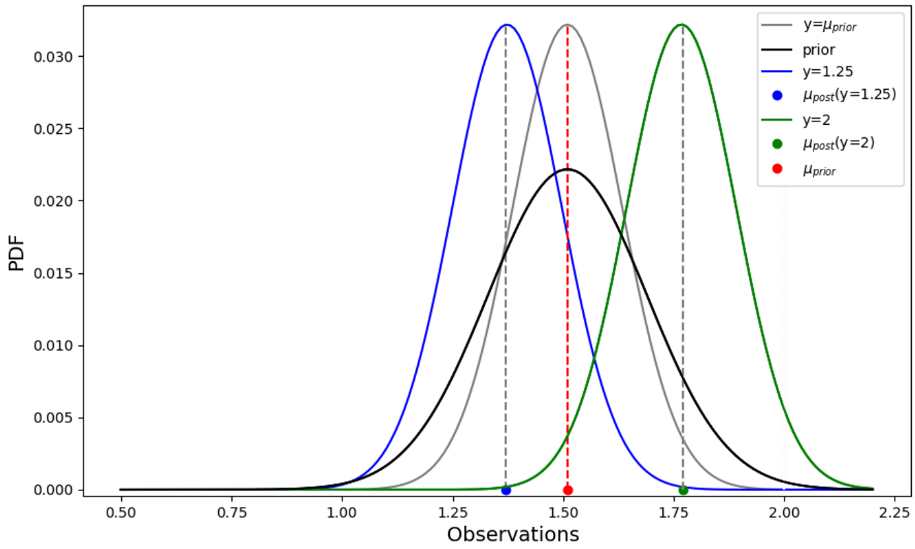

Figure 5.14: f_{prior} and posterior *pdf*s of *AF* for 3 different indirect PGA_s mean observations from measurements that would reduce the prior uncertainties by X=5%

Figure 5.14 displays the normalised prior and posterior PGA_s *pdf*s for 3 different PGA_s mean observations and for $X = 5\%$. Observations, y, are simply the measured AFs multiplied by the hazard on rock, PGA_r. The displayed posteriors are computed for a mean observation equal to μ_{prior}, 1.25 and 2. Vertical dashed lines are associated to each posterior to indicate the distributions mean $\mu_{post}(y)$. Comparing y and the obtained $\mu_{post}(y)$ shows the influence of the prior and likelihood definition on the posterior *pdf*s. In fact, the estimated $\mu_{post}(y)$ tend to shift towards μ_{prior} when $y \neq \mu_{prior}$, suggesting a 'correction' to the observation in light of the prior knowledge. Conversely, $y = \mu_{prior}$ produces no shift to the estimated posterior mean that equals μ_{prior}. These results are in accordance with the outcomes of Chapter 4.

Finally, EVII is assessed for different X, ranging from measurements having the same prior variability (i.e., X=0%) to measurements that perfectly infer AF. The sensitivity results are shown in Figure 5.15. Comparable to Figure 5.9, EVII shows a roughly linear tendency where values increase when the percentage of uncertainty reduction is increased. Similarly, EVII tends to asymptotically converge towards the EVPI, which is indicated by a red dot.

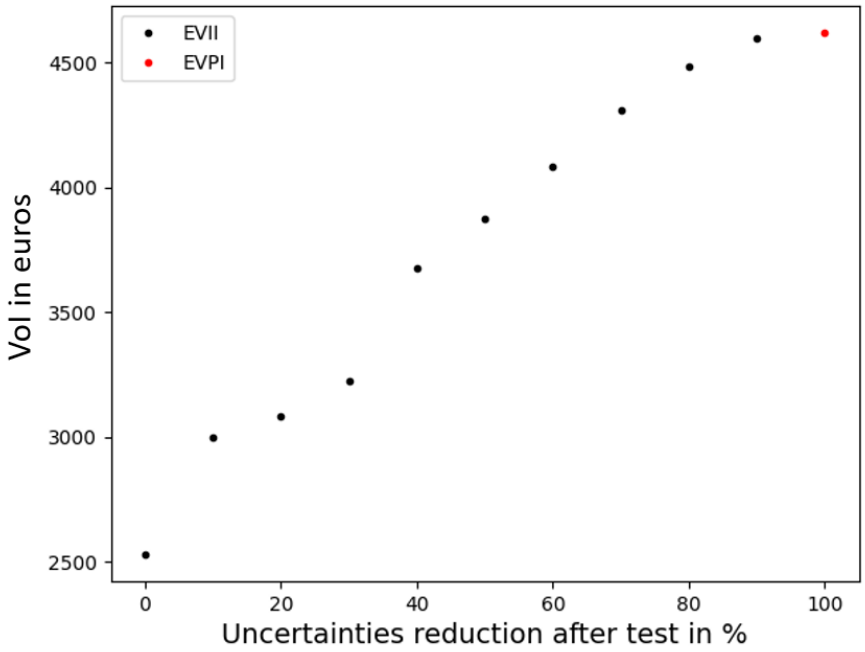

Figure 5.15: VoI sensitivity analysis to the percentage of uncertainty reduction after data collection

The results depend on the Monte Carlo sampling regarding the number of simulated observations (N), as well as the sampling from posteriors (m). An acceptable stability (i.e., negligeable variations) is reached for N=5,000 and m=100.

Moreover, we notice that when computing EVII for $X = 0\%$ (i.e., collecting data with no improvement in data quality), VoI is different from 0. As demonstrated in section 5.4.3, collecting twice the same data strengthens the prior distribution, where the posterior becomes the new prior. Similarly, hearing a piece of information twice might increase the chances of it being true.

When comparing the EVII results with the estimates from direct measurements on *AF* for a single-layer profile in Figure 5.9, we notice that in this study case the slope translating the increase of EVII when better quality and/or quantity data are collected is smaller than the study case in section 5.4.3. In addition, the EVPI is also lower. These differences suggests that within these specific examples, direct *AF* estimations are more valuable for a single-layer profile than a multi-layer profile. This is primarily due to a smaller prior *AF* variability than for the single-layer profile case.

5.6 Conclusion

In this chapter, we assessed the VoI for different site characterisations and several types of measurements, aiming at reducing uncertainties regarding one or more parameters. The VoI application has been designed to help in the decision-making process of inferring an optimal seismic design for a building located in northern Italy. VoI was first estimated when considering uncertain properties in a single-layer profile, such as the shear-wave velocity V_s and the thickness H, where analytical calculations of amplification factors (*AF*s) were used to estimate the seismic hazard on site. The second profile configuration consisted of a multi-layer complex profile defined by uncertainties on both V_s and H at each layer. *AF*s were computed by performing 1D numerical elastic linear site-response analysis. The goal of these assessments is not only to compute VoI, but also to study its variability to the different inputs (i.e., site characterisation, prior knowledge and likelihood).

The single layer-profile case was applied to develop a VoI approach applicable when we have two sources of uncertainty, V_s and H, by introducing bivariate joint distributions. The developed framework permits assessment of the value of additional data on solely V_s, solely H and both simultaneously with different degrees of data quality and/or quantity. Results show that in most cases, combining measurements of both parameters is more valuable (i.e., increase in VoI), for the condition that both measurements are reliable and of high quality. In fact, we notice that often, high quality data for one parameter results in a higher VoI than using poorly constrained data for both parameters. Moreover, the value of direct measurements on *AF* (e.g., SSR, HVSR) gives similar estimates of VoI compared to obtaining measurements on both V_s and H. These results are mainly due to the assumption of a linear elastic soil response and cannot be generalised for regions with moderate to high seismicity where the soil is likely to display nonlinear behaviour, rendering computational AFs from linear site-response analyses more difficult to match direct AF measurements. Finally, deciding on the target parameter and the type of test to conduct would depend on the cost of the chosen ground investigation technique.

In this chapter, comparisons have been made between the different results of the two site profiles (i.e., single and multi-layer) and the three measurement types (V_s, H, and AF). While these interpretations are true for these specific applications, they cannot be generalised. In the last case (section 5.5), we concluded that measuring AF for a single layer is more valuable than for a multi-layer profile. While it has been proven in these specific cases to be true, it is important to keep in mind that the definition of the input prior knowledge is, however, different. While uncertainties on V_s an H have been assigned arbitrarily, the variability expressed by the multi-layer profile is defined differently and is incorporated using the Toro (1995) model. Specifically, the obtained AF variability for each of the single and multi-layer profile is different. Indeed, prior variability in the AF in the multi-layer case is smaller than the single-layer profile case suggesting more constrained prior knowledge. As a result, we obtain lower VoI estimates. This finding aligns with the sensitivity analyses performed in Chapter 4 to study the impact of prior knowledge on VoI where we show that the less we know, the more we benefit from additional information.

Analysing and comparing the results within the case of single-layer profile is more straightforward. The findings from VoI estimation for the three types of measurements can be more readily compared as they share the same prior knowledge, same profile configuration and same method for the computation of AF.

The next chapter constitutes a significant upgrade to the overall approach. While VoI is still used to make informed decisions on seismic design, the structure of interest is now considered to be a critical facility, specifically a nuclear powerplant (NPP). This NPP is assumed to be located in the same city, Mirandola, for convenience when applying the method. The seismic design criteria are no longer only based on PGA, as in the current and previous chapter. Instead, the UHS over a large structural frequency range is considered. Moreover, the site-specific PSHA is computed according to modern practice, allowing the incorporation of nonlinear site effects. The site-specific PSHA approach is modified to a non-ergodic approach focused of assessing the value of reducing epistemic uncertainties in the site characterisation.

6 VoI for multivariate uncertain parameters within site-specific probabilistic seismic hazard assessment: Full convolution method

6.1 Introduction

The state-of-the-art for determining a robust seismic design for a structure requires considering the full Uniform Hazard Spectrum (UHS) computed from a Probabilistic Seismic Hazard Assessment (PSHA) at a particular return period. In the previous chapters (4 and 5), the Value of Information (VoI) approach was developed by computing the expected losses for the failure of a hypothetical building based solely on the site response at zero spectral period (i.e., the peak ground acceleration, PGA).

We firstly assumed a single uncertain parameter, Vs_{30}, characterised by a discrete probability function (Chapter 4). The resulting site response was simply computed by the multiplication of a frequency-independent amplification factor, which is also only dependent on Vs_{30}, by a rock motion characterised by PGA. Following that step, the VoI approach was further developed to consider Vs_{30} as a continuous uncertain variable using Monte Carlo simulations.

The approach evolved in Chapter 5, where 1D linear site-response analyses (SRA) were performed to estimate the site amplification functions. A first case described a site characterised by a single layer over an elastic bedrock where the associated shear-wave velocity (V_s) and thickness are assumed uncertain. Joint probability distributions were introduced to define the prior probabilities when computing VoI. SRA were analytically performed and the Level-1c site-specific PSHA approach (Table 5.1) was applied to estimate the expected PGA (i.e., the ordinate of the UHS for zero spectral period). The second case consisted of a complex six-layer V_s profile where numerical linear SRA were run using random profiles generated through Monte Carlo

simulations. Through sensitivity analyses, the developed VoI approach showed coherency with previous results. Nevertheless, only the response or hazard representing the PGA was accounted for when evaluating the probability of a building's failure and the associated expected losses.

Only considering the hazard at PGA for a fixed annual frequency of exceedance (AFoE) might underestimate the potential hazard at other structural periods and would possibly put the structure at risk and general safety in peril. The design response spectrum, when constructed to envelop the site-specific UHS, helps evaluate the possible earthquake lateral loads that a structure can be expected to withstand during its design lifetime. Several guidelines and procedures have been developed in the last decades to infer seismic designs at a specific location depending on the structure's characteristics, such as the National Earthquake Hazards Reduction Program provisions (e.g., NEHRP, 2015), the International Building Code (IBC, 2012) and Eurocodes (e.g., CEN, 2004). Nevertheless, site-specific PSHA is still encouraged to take into account all uncertainties, such as soil properties and non-linearity effects as well as catalogue incompleteness and path attenuation, in order to infer a seismic design with an acceptable level of seismic performance.

In this chapter, we propose a novel approach for VoI calculations that aims to respect the state-of-the-art regarding 1D site-specific SRA, site-specific PSHA and the guidelines and requirements of seismic design codes. The VoI approach is developed to estimate the value of reducing V_s profile uncertainties applied in the selection of an appropriate seismic design for a critical facility, a hypothetical nuclear powerplant (NPP) located in Mirandola, Italy. Although this location might not be suitable for NPP installations, it has been selected as of the available geophysical and geotechnical data, for comparison purposes with the results in Chapter 5 and to develop the VoI methodology.

This framework consists of:

- Performing 1D linear-equivalent SRA
- Incorporating SRA results within a full convolution site-specific PSHA approach (Bazzurro and Cornell, 2004)

- Proposing and developing a method to capture and differentiate the epistemic uncertainties from the motion-to-motion variability
- Inferring hazard curves and UHS
- Developing an approach to estimate a multivariate joint distribution describing the UHS variability over a range of spectral periods
- Constructing potential nuclear seismic designs and associated fragility curves at specific spectral periods
- Implementing VoI input parameters and defining different test error functions

The full VoI framework is then detailed and the expected value of perfect information (EVPI) and imperfect information (EVII) for different hypothetical tests measurements are estimated, followed by comparison and interpretation of the results.

6.2 Site-specific PSHA

PSHA accounts for all possible earthquakes (source models) as well as the variability of the ground motion (source, path, site etc.) to obtain hazard curves expressing the probability of exceeding a certain ground motion. PSHA has become the standard approach for developing probabilistic estimates of ground motion levels and it is often used to infer seismic design loads for structures or to derive inputs for dynamic analyses. It incorporates a wide range of sources of epistemic uncertainties and aleatory variabilities affecting both the seismic-source and the ground-motion models. PSHA studies are usually undertaken for rock sites to permit consistency between studies and to enable the estimation of input rock motions used in SRA.

In this chapter, we apply the full convolution method (Bazzurro and Cornell, 2004) to estimate the site-specific PSHA at a particular location and to produce the hazard curves at several spectral periods as well as the associated UHS. In the following, all input parameters are defined to implement SRA. SRA outputs are then incorporated within the full convolution method to infer the output of interest: the UHS at a target hazard level. Moreover, we propose an alternative approach to the traditional convolution method to retrieve the contribution of the epistemic uncertainty. Finally, we show that the alternative method is equivalent to the traditional one and valid to be used within the proposed VoI framework (in section 6.3).

6.2.1 Site characterisation

The hypothetical NPP is assumed to be constructed at a site in the city of Mirandola, Italy (same location as in Chapter 5). Site-specific ground motion analyses require detailed measurements of the site soil properties in order to quantify, as accurately as possible, the local amplification/de-amplification due to the underlined geology. Barani & Spallarossa (2017) provide a six-layer 1-D numerical soil model for a site in Mirandola, which is inferred from available geophysical and geotechnical data. The profile properties are detailed in Table 5.2 (Chapter 5).

The shear-wave velocity, V_s, is defined by a lognormal distribution of median μ_{V_s} and standard deviation σ_{V_s} at each layer's mid-depth. Within STRATA, the Toro (1995) model allows the generation, through Monte Carlo simulations, of random V_s profiles following defined lognormal distributions. Performing these simulations are an essential step in integrating measurement uncertainties. Regarding the uncertainties in the layer thicknesses, the Toro (1995) model employs a generic depth dependent rate model $\lambda(d)$:

$$\lambda(d) = a.(d + b)^c \tag{6.1}$$

The coefficients a, b and c, estimated by Toro (1995) using layering from 557 sites in California, describe the increase in thickness uncertainties with depth. These coefficients have been modified in our present study to assume a constant thickness uncertainty with depth. Therefore, a, b and c have been fixed to 1, 1 and 0, respectively. These correlation models are directly implemented in the software STRATA (Kottke *et al.*, 2013) to allow the layering and shear-wave velocities V_s to vary through Monte Carlo simulations. STRATA allows the user to define the median V_s (μ_{V_s}) and the standard deviation ($\sigma_{\ln(V_s)}$).

For the purpose of the current application, the values of σ_{Vs} have been increased by doubling the value obtained from the initial measurements (Table 5.2) to increase the prior uncertainties. The associated $\sigma_{\ln(V_s)}$ is approximated as follows for all layers:

$$\sigma_{\ln(V_s)} \approx \frac{\sigma_{Vs}}{\mu_{Vs}} \approx 0.12 \tag{6.2}$$

Constraints have been added to STRATA to truncate the generated V_s profiles at an upper and lower bound of +/- $2\sigma_{\ln(V_s)}$ around the median to avoid unrealistic profiles, following the recommendation of EPRI (2013).

The VoI framework requires quantifying the overall impact of the V_s profile uncertainties on the obtained UHS before and after uncertainty reduction through additional measurements. SRA should then be performed for different values of $\sigma_{\ln(V_s)}$. 50 randomised models are generated through STRATA for different $\sigma_{\ln(V_s)}$: 0.12, 0.06, 0.03, 0.015. The obtained V_s profiles realisations are presented in Figure 6.1.

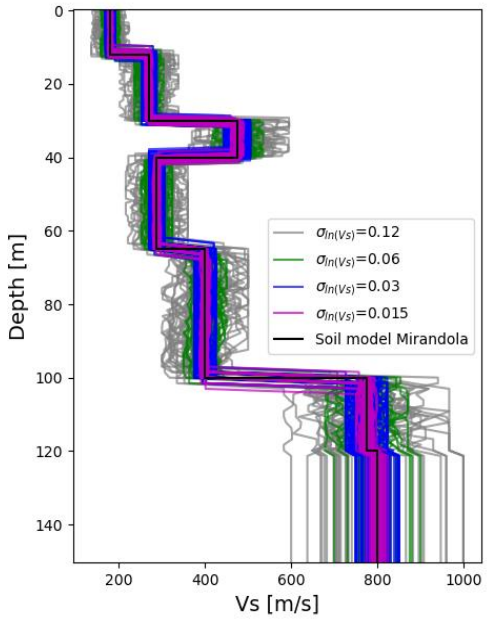

Figure 6.1: Mirandola soil model and Monte Carlo realisations with different $\sigma_{ln(V_s)}$

The modification of the Toro (1995) layering model results in thickness and V_s variations coherent with the median soil model at Mirandola and the defined truncated lognormal distributions. The interlayer shear-wave velocities show minimal variations for some profiles with negligible effects on the overall soil response. These profiles, corresponding to each of the configurations, are used in SRA to retrieve the site amplification functions considered as input to the site-specific PSHA.

6.2.2 Input motions

STRATA enables 1D linear or linear-equivalent SRA by using either input time series or the Random Vibration Theory (RVT) approach (Silva *et al.*, 1997; Boore, 2003). RVT has the advantage of solely requiring an input motion described by its Fourier Amplitude Spectrum (FAS) and a motion duration. The other advantage is that the method does not suffer from time series non-uniqueness and spectral matching methods. Nevertheless, this method has some drawbacks as it might overestimate the predictions in terms of site amplification around the natural site frequencies by as much as 20-50% (Kottke & Rathje, 2013).

The alternative is using recorded time-series as input motions for 1D linear-equivalent SRA. A suite of input time-series is used to simulate the motion-to motion variability to generate stable estimates of the expected response at the reference outcropping rock. Input motions should be matched to the spectrum at the reference outcropping rock (e.g., the UHS) and at the target hazard level over a range of spectral periods or a single period (Stewart *et al.*, 2014). In the present study, the target hazard level according to UK regulations for NPPs (HSE, 2009) corresponds to an annual frequency of exceedance (AFoE) of 10^{-4}, equivalent to a UHS at a return period of 10,000 years. The issue in considering matching records to a single hazard level is that the amplification functions are defined for a narrow ground motion intensity range. In case of nonlinearity, this range is often insufficient and fails to fully capture nonlinearity effects. The U.S. Nuclear Regulatory Commission (USNRC, 2007) requires disaggregation of the mean probabilistic hazard (Bazzurro and Cornell, 1999) at ground motion levels corresponding to AFoEs of 10^{-4}, 10^{-5} and 10^{-6} for NPP applications. Disaggregation is an efficient tool to infer the most controlling earthquake scenarios as input to the SRA.

It was preferred to find matching records for a single target hazard level: 10^{-4}. This was achieved using the online tool RexelWeb developed by Sgobba et al. (2019). This tool enables the user to find and scale matched records to a particular spectrum. To do so, hazard curves at the reference rock site in Mirandola are retrieved from the ESHM20 database (Danciu *et al.*, 2021) and converted to AFoE for several spectral periods in the range [0.01-3]s. This permits the construction of the UHS at AFoE=10^{-}

[4]. RexelWeb considers the UHS as input along with other specifications on the record selections (e.g., Eurocode 8 site classification of the station, magnitude-distance range, number of record combinations). We fix the criteria to obtain recordings at stiff soil stations (Eurocode 8 site class B), from events of moment magnitude (M_w) between 5 and 6.5 and at epicentral distances ranging from 10 to 50 km. Recordings at sites of class A (Vs_{30}>800 m/s) would be more suitable, considering that the used hazard curves on rock are defined for Vs_{30}=800 m/s. However, insufficient recordings were obtained from RexelWeb for this site class and the target magnitude-distance range. For similar reasons, we obtain a total number of six records, slightly below the minimum number of seven input motions recommended by Stewart *et al.*, (2014). For non-linear response analysis, it is recommended to use 11 records. In this study, we intend to perform a linear-equivalent SRA and express the soil epistemic uncertainties to include within the VoI framework. For the scope of this chapter, six records motions are considered sufficient to capture the variability of the hazard at the site and to incorporate in the framework.

Table 6.1 lists key parameters for the six input motions from the Engineering Strong Motion (ESM) database (Luzi *et al.*, 2015):

Table 6.1: Input motions from site class B stations used in soil-response analyses

Station name	**Event**	**Component**	M_w	**PGA (g)**	**Vs_{30} (m/s)**	**Epicentral distance (Km)**
BRN	Irpinia (23/11/1980)	E	6.9	0.17	403	42.6
SLO	Sellano (26/10/2016)	N	5.9	0.07	N/A	14.4
CSC	Norcia (30/10/2016)	E	6.5	0.17	698	14.9
T1214	Norcia	E	6.5	0.60	N/A	11.4
MZ29	Norcia	E	6.5	0.69	N/A	26.9
MZ29	Norcia	N	6.5	0.41	N/A	26.9

The associated time-series are obtained along with an associated scaling factor. The unscaled and scaled acceleration spectrum of each recording are shown in Figure 6.2. The scaled spectra match well, on average, the UHS at AFoE=10^{-4}. However, the scaled recordings have been automatically normalised through RexelWeb to the target spectral acceleration, the PGA. This prevents the characterisation of nonlinear effects in the spectral period range 0.01s to 0.04s. Using these scaled records limits the input motion range to apply the full convolution method within site-specific PSHA. Therefore, the unscaled records are preferred as input for SRA. For purpose of comparison, the UHS at 10^{-3} is also shown in Figure 6.2.

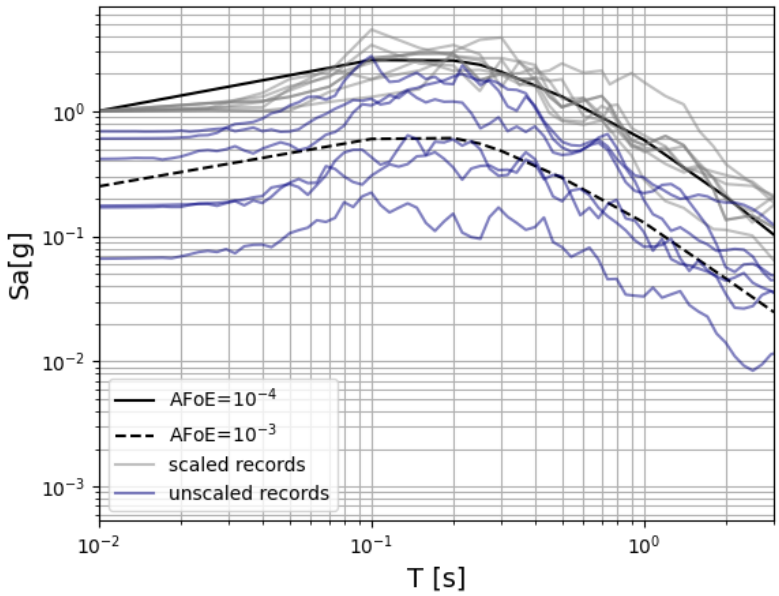

Figure 6.2: Target UHS at reference rock (AFoE= 10^{-4} and 10^{-3}) and scaled/unscaled records response spectra

The target UHS being defined at the reference outcropping rock, the matched accelerograms should be deconvolved to obtain the input motion at the bedrock level through using generic profiles such as Cotton et al., (2006), and then to be propagated through the soil column. We simplified the calculations, however, by assuming here that the response at the reference rock is comparable to the response at the bedrock level. Therefore, the selected accelerograms are vertically propagated from the base (bedrock) of each of the 50 profiles generated with the Monte Carlo approach outlined in the previous section to assess the ground motion at the surface.

6.2.3 Site Response Analyses

1D equivalent-linear response analysis is carried out with STRATA to take into account the dependence of the ground shaking on the dynamic soil properties and thus, on the shear strain. The nonlinearity of the soil response is then included in the response analysis by simplifying it into a linear system that uses strain compatible dynamic properties. The strain dependence on the nonlinear properties is described in several different models in the literature (Seed and Idriss, 1970; EPRI, 1993; Darendeli, 2001). The model developed by Darendeli (2001) is used within STRATA and assumes that both the damping ratio (D) and shear modulus reduction (G) are normally distributed. This model creates the shear modulus reduction and damping curves, which are assumed to be known.

The 5% damped response spectrum at the surface (0m) of the site is computed by vertically propagating each of the six input motions through the 50 simulated soil profiles using μ_{Vs} and $\sigma_{\ln(V_s)}$={0.12,0.06,0.03,0.015} denoted respectively $\{\sigma_{\ln(V_s)}, \frac{1}{2}\sigma_{\ln(V_s)}, \frac{1}{4}\sigma_{\ln(V_s)}, \frac{1}{8}\sigma_{\ln(V_s)}\}$ henceforth. The amplification factors are then calculated as the ratio of the 5% damped spectral acceleration at the surface of the soil profile (SA_{soil}) and the 5% damped spectral acceleration at the bedrock (SA_r). The period-dependent amplification factor is defined as follows:

$$AF(T) = \frac{SA_{soil}(T)}{SA_r(T)} \tag{6.3}$$

We obtain 300 AF curves corresponding to the propagation of each input record (six) through the 50 Monte Carlo generated soil profiles where $\sigma_{\ln(V_s)}$=0.12. AFs are represented in Figure 6.3 along with the corresponding median curve.

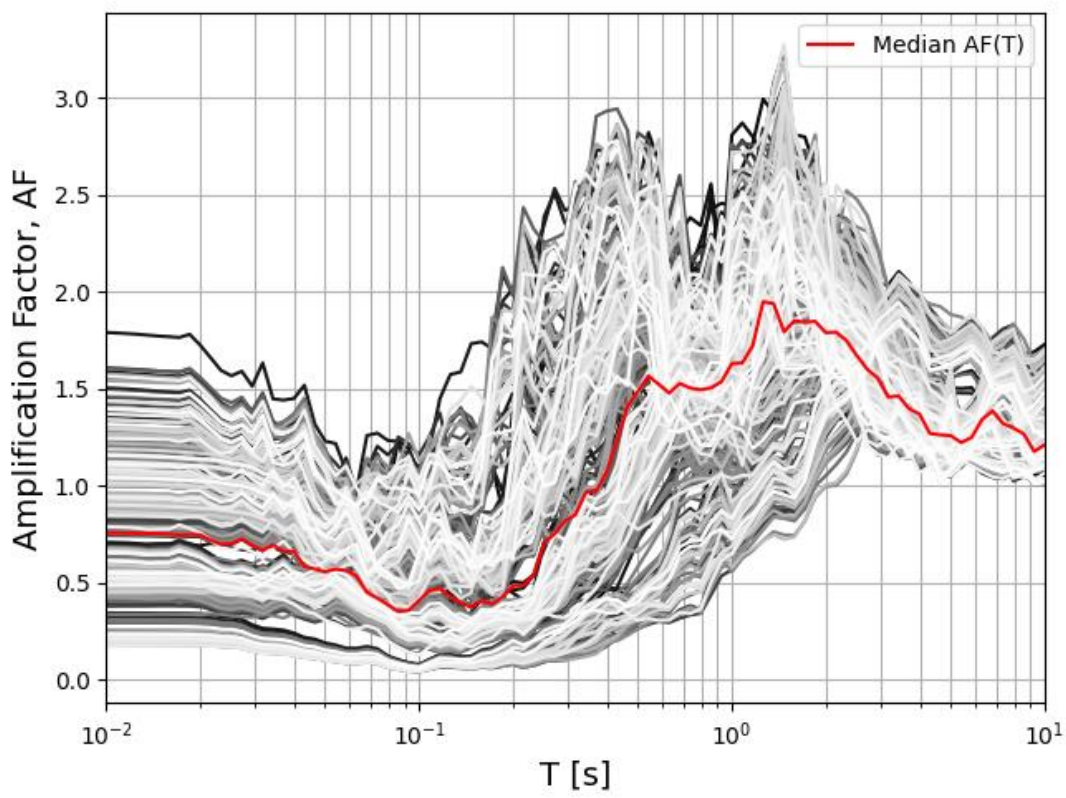

Figure 6.3: Computed Amplification Factors (*AF*s) for 50 Monte Carlo V_s profiles simulations ($\sigma_{ln(V_s)}$=0.12) and six input records

The highest *AF* peak can be seen around the spectral period 1.4s, corresponding to the soil model's fundamental period T_0. A second peak is noticeable in the range 0.3-0.7s which can be due to the second and third modes of the soil model, which are identified as 0.7s and 0.35s, respectively.

The median *AF* for the two peaks are 1.9 and 1.6, respectively, and decrease after the second peak, tending towards 1.0. The site *AF* presents a strong de-amplification at around 0.1s which could be due to the shear-waves reversal at around 40 meters depth. The soil nonlinearity might also affect the amplification as ground motions are strong around these spectral periods. Moreover, the site amplification depicts the highest variability around the spectral periods corresponding to PGA and the two first natural frequencies of the soil model.

AF curves and the associated standard deviation are inputs of high importance to perform site-specific PSHA using the full convolution method of Bazzurro & Cornell (2004).

6.2.4 Site-specific PSHA approach: Full convolution method

The full convolution method introduced by Cramer (2003) and analytically developed by Bazzurro & Cornell (2004) permits the estimation of the site-specific seismic hazard by accounting for site-effects in a fully probabilistic framework. This method is equivalent to the Approach 3, used in the U.S. nuclear industry and introduced within NUREG 6728 (McGuire et al., 2001). The soil hazard curves, HC_{soil}, are computed through the convolution of the reference rock hazard curves and the amplification function as follows:

$$HC_{soil}(Z) = \int_o^\infty P\left[Y > \frac{Z}{x}\middle|x\right] f_x(x)dx \tag{6.4}$$

where $Z = SA_{soil}[T]$ and $x = SA_r$[T]. Y is the input dependent site amplification $AF(T, SA_r)$ and $P[Y]$ its distribution conditioned on the input rock motion SA_r. Finally, $f_x(x)$ is the probability density function (*pdf*) of $x = SA_r$.

This equation was made readily useful by Bazzurro & Cornell (2004) by discretizing the integral as follows:

$$HC_{soil}(Z) = \sum_{all\ x_j} P\left[Y > \frac{Z}{x_j}\middle|x_j\right] p_x(x_j) \tag{6.5}$$

The term $p_x(x_j)$ represents the probability of the input rock motion being equal or higher than x_j. Its value can be calculated by differentiating the rock hazard curve in discrete or numerical forms. In other words, $p_x(x_j)$ is the annual probability of occurrence of $SA_r = x_j$ and $P\left[Y > \frac{Z}{x_j}\middle|x_j\right]$ translates the probability that AF is greater than $\frac{Z}{x_j}$ given $SA_r = x_j$.

Equation (6.5) could be analytically approximated by making the following assumptions:

- The amplification factor $AF(T, SA_r)$ is lognormally distributed given SA_r: The median is described by a linear piece-wise dependence on SA_r (log-log scale) and its variability is described by a lognormal distribution.
- The rock hazard can be represented by a power law around the return period of interest.

This approach respects the probabilistic aspects of the hazard and subsequently the UHS and considers the amplification function as an *a posteriori* corrector to the rock hazard calculations.

Considering the above assumptions, $P\left[Y > \frac{z}{x_j}\middle| x_j\right]$ can be written as follows:

$$P\left[Y > \frac{z}{x_j}\middle| x_j\right] = \hat{\phi}\left(\frac{\ln\left(\frac{z}{x_j}\right) - \mu_{\ln AF|x_j}}{\sigma_{\ln AF|x_j}}\right) \tag{6.6}$$

Where $\mu_{\ln AF|x_j}$ and $\sigma_{\ln AF|x_j}$ are respectively the median value and standard deviation of the natural logarithm of AF given $SA_r = x_j$. $\hat{\phi}(.)$ is the standard complementary Gaussian cumulative distribution function. $\mu_{\ln AF|x_j}$ and $\sigma_{\ln AF|x_j}$ are derived through SRA by regressing AF values computed for several soil profiles and input rock motions for each spectral period T.

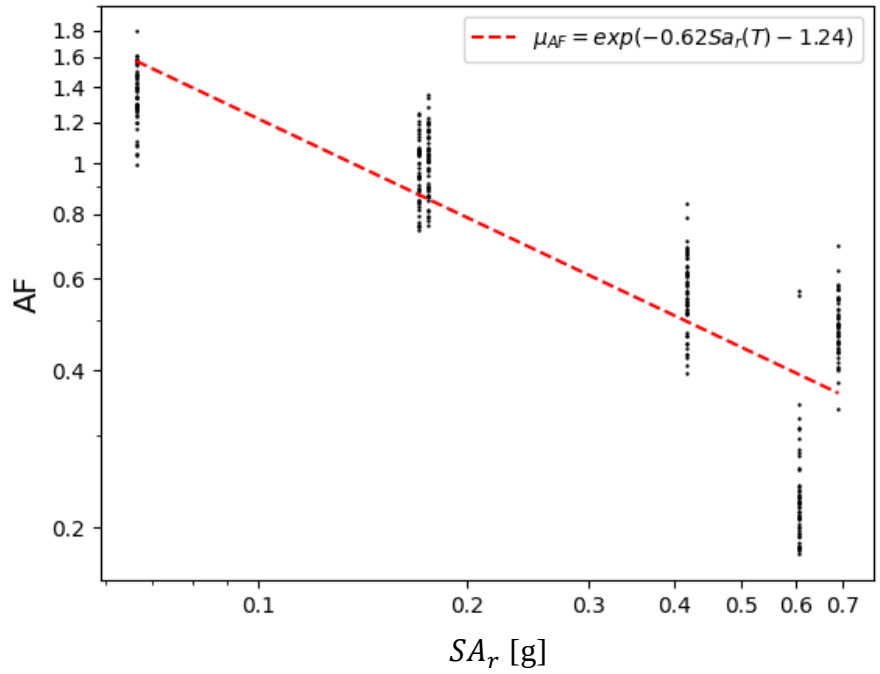

Figure 6.4: *AF* for six input records PGA and 50 random profiles. The linear regression curve representing μ_{lnAF} is a dashed red line.

The relationship describing the dependence of AF on SA_r can be expressed by a linear or a nonlinear relationship. An example of linear regression on AF for PGA inputs is given in Figure 6.4. Python can solve for the regression parameters and it can estimate

the associated standard deviation $\sigma_{\ln AF}$. Generally, the amplification factor decreases with increasing input motions intensity due to the soil's nonlinear properties. This can be seen through the negative slope of the $\mu_{\ln AF}$ function in Figure 6.4. The linear relationship has this form:

$$\ln \overline{AF}(T, SA_r) = a + b \ln(SA_r(T)) \tag{6.7}$$

a and b are regression coefficients with b representing the effect of nonlinearity.

When the range of input rock intensities is large enough, the regression could be in the form of a nonlinear relationship. This form is usually preferred, especially as the linear form can lead to non-physical values at low input intensities. In fact, $AF(T, SA_r)$ should flatten at the lowest input-motion intensities to respect linear soil conditions.

Stewart et al. (2014) modified the expression used by Abrahamson & Silva (1997) as follows:

$$\ln \overline{AF}(T, SA_r) = c_1 + c_2 \ln\left(\frac{SA_r(T)+c_3}{c_3}\right) \tag{6.8}$$

c_1 represents the weak-motion linearity, c_2 the nonlinearity and c_3 the reference input motion below which AF converges towards linearity. Higher degree polynomial forms can also be applied.

In the present study, the linear form is adopted because of the limited coverage of input motions, which are insufficient to capture both the linearity at very weak motions and the nonlinearity at very strong motions. From the following results, the nonlinearity effects appear to be moderate. We recall that the state-of-the-art requires matching records to several levels of hazard. However, as explained earlier, we only use six unscaled moderate to strong input records. This study focuses more on the application of VoI when site-specific PSHA is performed rather than providing the best engineering estimate of $AF(T)$.

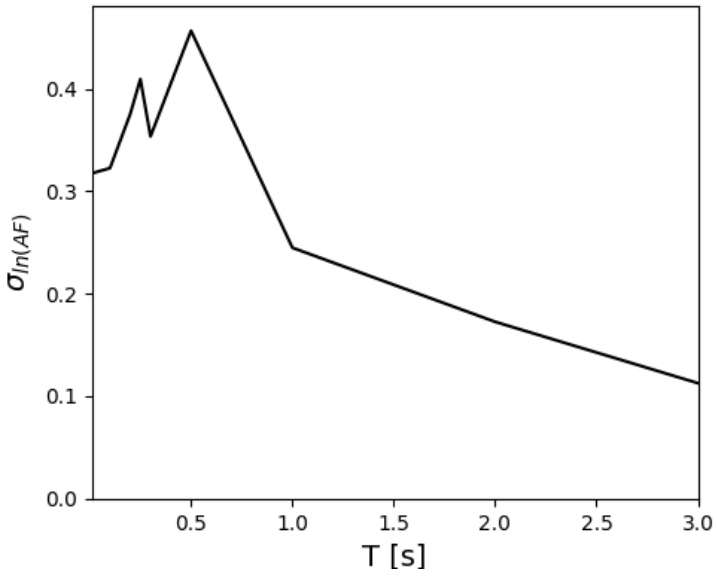

Figure 6.5: Standard deviation of the natural logarithm of *AF* ($\sigma_{ln(V_s)}$=0.12)

The advantage of this method is that the amplification variability is taken into account. $\sigma_{\ln AF}$ is computed for each spectral period and incorporated in Equation (6.6) to produce the hazard curves for the chosen spectral periods. The variation of $\sigma_{\ln AF}$ with spectral period T is shown in Figure 6.5 for $\sigma_{\ln(V_s)}$=0.12. The highest variability is observed around the second and third modes of the site model resonance periods then decreases gradually for the highest periods.

One of the limitations of the convolution approach is that it does not distinguish epistemic uncertainties from aleatory variability. $\sigma_{\ln AF}$ is considered in the full convolution method as aleatory (Rodriguez-Marek *et al.*, 2014) as a result of the motion-to motion variability but also because of the use of the Monte Carlo step to generate the soil parameters. Thus, $\sigma_{\ln AF}$ is ergodic. As a result, at each spectral period, the convolution of the amplification function and the rock hazard curve at the same spectral period results in a single soil hazard curve. However, to compute the VoI, it is important to be able to build prior distributions from available information but as equally important, to estimate the likelihood and posterior distributions to assess the outcomes in reducing uncertainties. Distributions cannot be constructed from a single soil hazard curve. To tackle that, it is necessary to separate the epistemic uncertainty from the aleatory, representing here the motion-to-motion variability.

Barani et al. (2013) have studied the influence of soil modelling uncertainties on site response by performing sensitivity analyses through varying one soil property at a time. The soil modelling uncertainties (e.g., V_s, layer thickness, unit weight, dynamic properties) and input motion variability (e.g., number of records and scaling) have been separately studied and analysed to quantify their individual impact on AF. This is useful in identifying both the epistemic and aleatory components but also to assess what parameter is predominant in the overall variability.

In this application, we opted for explicitly expressing the epistemic uncertainty by estimating $AF - SA_r$ regressions for each random profile. The standard deviation, associated to each regression, would then only capture the motion-to-motion variability. The convolution approach can then be applied for each random profile to obtain 50 soil hazard curves at each spectral period. This method allows the construction of the prior probability distributions that translate epistemic uncertainty, V_s profiles here, at each spectral period for a chosen AFoE.

Figure 6.6 shows AF versus SA_r for four spectral periods: 0.0s (i.e., SA_r is PGA), 0.1s, 0.3s and 1s for the reference case $\sigma_{\ln(V_s)}$=0.12. Regressions including all profiles are shown in red lines and those for each individual profile in grey. The slope of individual profiles linear regression is negative and consistent with the soil nonlinearity. From the spectral periods considered, 0.1s depicts the highest nonlinearity effect while at higher spectral periods (T=1s), the nonlinear effects decrease. $\mu_{lnAF}(SA_r, T)$ is then computed for each profile as well as the associated standard deviation, representing the motion-to-motion variability. The convolution has been applied and the resulting soil hazard curves are shown in Figure 6.7 for the same four periods.

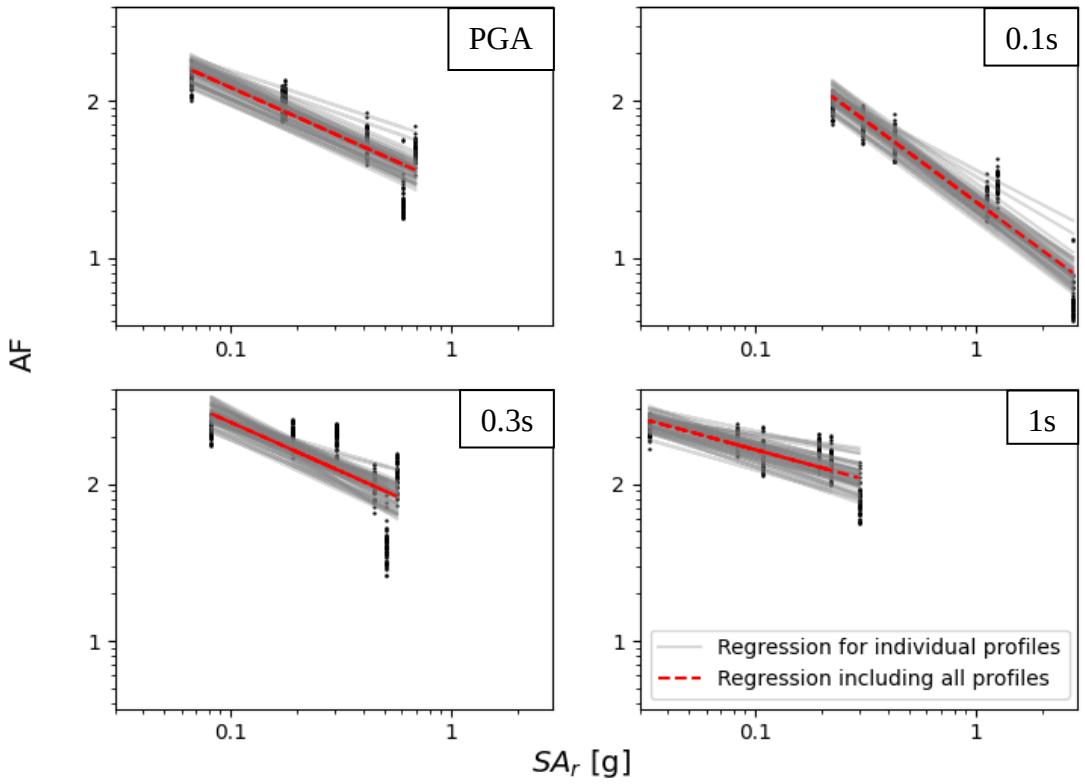

Figure 6.6: *AF* versus SA_r for T=0.0s (PGA),0.1s, 0.3s and 1s. Linear regressions including all 50 random profiles (red line) and for each individual profile (grey lines) are shown.

In Figure 6.7, we compare the single soil hazard curve computed by including the total standard deviation (i.e., considering all profiles) and the median hazard curve resulting from all individual profile hazard curves. These are almost superimposed, suggesting that the above standard deviation decomposition method is valid and can be a substitute to the traditional implementation of the full convolution method.

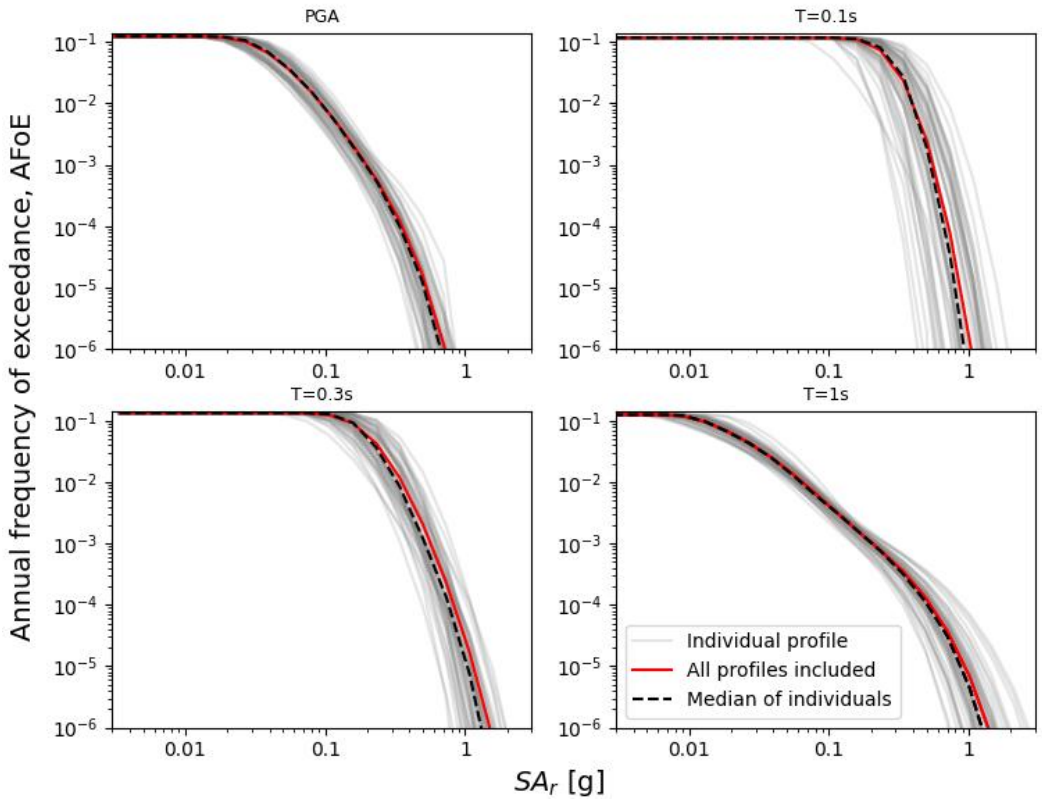

Figure 6.7: Soil hazard curves for T=0.0s (PGA),0.1s, 0.3s and 1s using the full convolution method including all profiles (red), individual profiles (grey) and the median of the individual profiles (black dashed).

Decomposing and isolating the epistemic uncertainties (layers V_s and thickness) enables to study the impact of new information on the hazard results and thus, on the site-specific UHS. The associated UHS at AFoE=10^{-4} is presented in Figure 6.8. This return period is chosen as the target level in our study case for an NPP. The spectral accelerations obtained from including all random profiles (red line) appear to be slightly higher than those obtained from calculating the median of individual profiles UHS in the period range [0.1s-0.5s]. Outside this range, both curves are almost superimposed. This validates the use of the decomposition method. This approach of site-specific PSHA enables building probability distributions, at chosen spectral periods, which define the prior knowledge part of the VoI calculation.

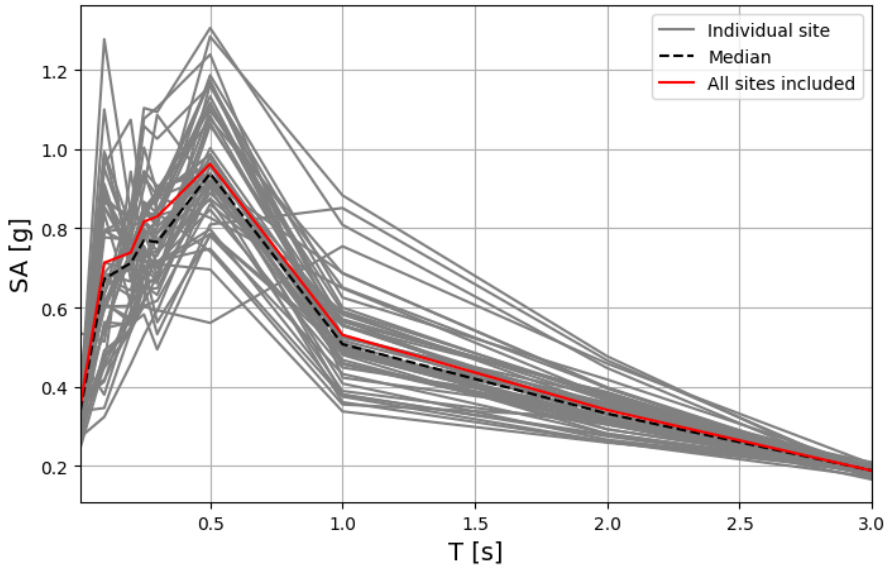

Figure 6.8: Soil Uniform Hazard Spectrum (UHS) at a return period= 10,000 years (AFoE of 10^{-4}) for $\sigma_{ln(V_s)}$=0.12

The next section details, step by step, the VoI framework. First, the chosen prior distributions are justified and built. Second, the different seismic design alternatives are constructed along with the associated fragility curves. Then, the expected outcomes estimations in case of total damage are defined and the associated values given. Finally, the expected posterior distributions after data collection are built based on results obtained from reducing σ_{VS}. Results of VoI estimation are then presented for several possible σ_{VS} reductions.

6.3 Value of Information approach

We are interested in predicting the reduction in expected losses in case of damage when inferring a seismic design for a NPP after data collection. We recall that the VoI assessment is to be performed before the data collection itself. In this study, we focus on one uncertain step leading to developing an appropriate seismic design, the seismic hazard assessment phase. Particularly, we aim at assessing the value, in monetary units, of reducing uncertainties regarding the site characterisation. The uncertain parameter of interest is the shear-wave velocity profile.

In this section, we propose a new approach resulting from a complex framework that uses the basic elements of the VoI method developed in the previous chapters (4 and

5). This approach is challenging as the uncertain variable is multivariate, unlike the previous studies (Vs_{30} and V_S-thickness of a single layer). The developed VoI approach is adapted and applied to a realistic situation, enabling seismic hazard analysts and facilities owners to understand how to implement the framework for specific applications.

First, we detail the preparation steps to define and construct the input parameters. Then, we explain their implementation within the VoI proposed framework. Finally, we analyse and interpret results of Prior Value (PV), Expected Value of Perfect Information (EVPI) and Expected Value of Imperfect Information (EVII) along with sensitivity analyses to understand the impact of various input parameters and to validate the results by comparing them with the results of previous studies (Chapter 4 and 5).

6.3.1 Building the framework input parameters

While the important elements and input parameters of the VoI method are theoretically known, the challenge lies in extrapolating the method for uncertain univariate and bivariate variables to a multivariate implementation. In previous chapters, VoI has been computed from the estimation of expected losses that are solely based on a univariate variable (e.g., expected PGA). Indeed, the fragility curves along with the expected loss function were only defined for PGA. The challenge in using a full UHS as an uncertain variable is the integration of prior probability distributions that translate the latter uncertainty over a range of spectral periods (T).

The following section is considered to be the most important step in carrying out consistent VoI estimations. This section details, step by step, the building of all crucial elements within the VoI calculation. From an attempt in predicting UHS variability with the input shear-wave standard deviations σ_{Vs} to building the requirements in the Expected Losses (EL) calculations, we also define the seismic design candidates and their associated fragility curves for specific spectral periods.

Distributions of UHS conditioned on σ_{Vs}

The idea is to study the impact of V_s uncertainty on the overall variability of the UHS and to consider the UHS as a prior conditional on σ_{Vs}. The prior chosen to be included in the VoI framework is not the V_s probability distribution. However, V_s could be the uncertain input prior given the condition that we can quantify its relationship to the resulting UHS. This condition is essential as UHS represents the judging criteria to compute EL. Attempts have been made to quantify the direct implication of σ_{Vs} variation on the UHS by analysing soil response results when σ_{Vs} is varied. The purpose being to better identify the impact of the different measurements on the resulting UHS. This would also simplify studying the sensitivity of VoI to this parameter.

Similar to Figure 6.6, Figure 6.9 depicts $AF - SA_r$ dependence and linear regression for different spectral periods for the reference $\sigma_{lnVs} = 0.12$ and when performing SRA using $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and $\frac{1}{8}\sigma_{Vs}$. The linear regressions computed at the same spectral period are all parallel, demonstrating similar nonlinearity effects. Table 6.2 shows the gradient (i.e., slope) of the linear regressions in log-log scale for all scenarios and at each spectral period. Median μ_{lnAF} appears to slightly increase when V_s uncertainties decrease. The variations being minimal, we focus on analysing the associated σ_{lnAF}.

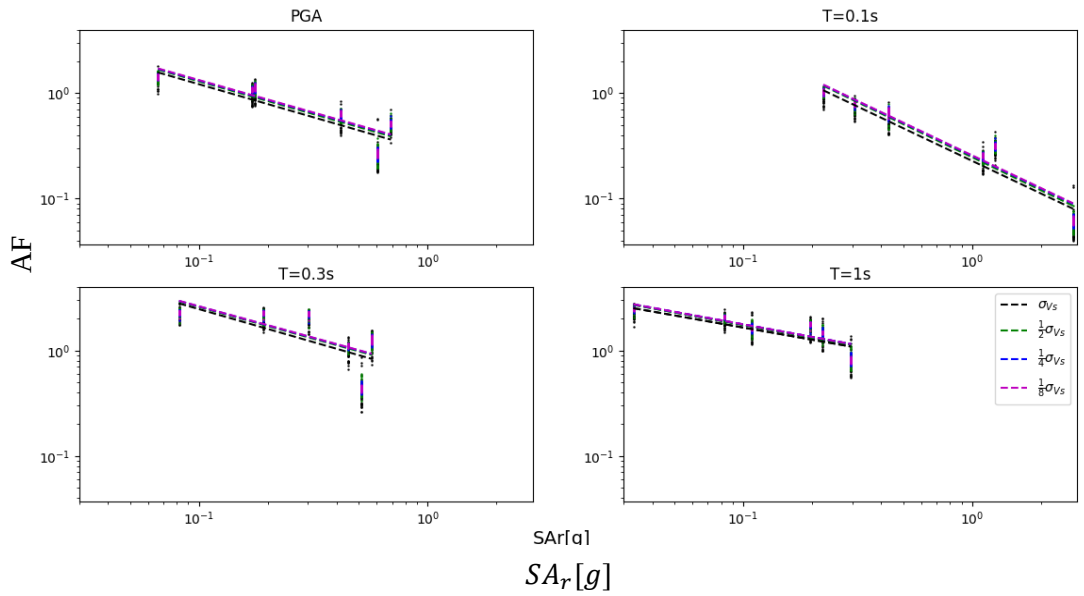

Figure 6.9: *AF* versus SA_r for T=0.0s (PGA) and T=0.1s, 0.3s and 1s. Linear regressions including all profiles for σ_{Vs}, $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and $\frac{1}{8}\sigma_{Vs}$ are represented as dashed lines

Table 6.2: Gradient regression coefficient of the linear regressions presented in Figure 6.9 in log-log scale

Linear regression coefficient / Scenario	gradient			
	T=0s	T=0.1s	T=0.5s	T=1s
σ_{Vs}	-0.60	-1.04	-0.62	-0.38
$\frac{1}{2}\sigma_{Vs}$	-0.60	-1.05	-0.60	-0.40
$\frac{1}{4}\sigma_{Vs}$	-0.60	-1.04	-0.60	-0.39
$\frac{1}{8}\sigma_{Vs}$	-0.59	-1.04	-0.60	-0.39

In Figure 6.10-a, we observe the total σ_{lnAF} values depending on T when implementing the soil model with the same median V_s and decreasing standard deviation (σ_{Vs}, $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and $\frac{1}{8}\sigma_{Vs}$). Generally, and for most spectral periods, decreasing σ_{Vs} results in lower associated σ_{lnAF}. To study in-depth the consequences, the percentage reduction in σ_{lnAF} compared to the reference σ_{Vs} is shown in Figure 6.10-b. There is a positive correlation between σ_{Vs} and σ_{lnAF} for all spectral periods except within the range of the 2nd and 3rd modes of the soil model (~0.35s-0.7s). The reduction in AF variability

is not homogeneous and seems difficult to quantify and predict for all periods. We can notice that for T=1s, the percentage reductions for $\frac{1}{4}\ \sigma_{Vs}$ and $\frac{1}{8}\ \sigma_{Vs}$ are almost equal.

Similar efforts were made by Barani et al. (2010) in deriving empirical predictive equations to relate the level of period-dependent amplification on Vs_{30} for a specific case study. It was found that Vs_{30} can be effective in predicting *AF*(T) only at short periods, but correlation was poor at longer periods. Best correlations were found when computing period-independent amplifications.

We conclude that attempts to extrapolate these results to predict the σ_{lnAF} associated to other V_s standard deviations without performing a proper soil response analysis would compromise the VoI framework and results.

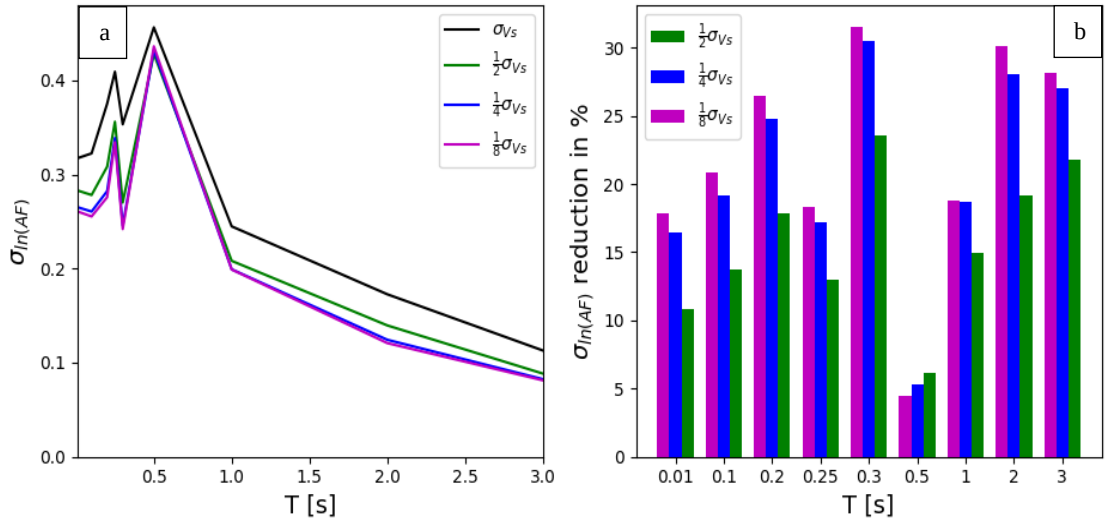

Figure 6.10: (a) Total $\sigma_{lnAF}(T)$ for σ_{Vs}, $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\ \sigma_{Vs}$ and $\frac{1}{8}\ \sigma_{Vs}$. (b) Bar plot showing the percentage of $\sigma_{lnAF}(T)$ reduction compared to the reference (σ_{Vs}) for each scenario

Further attempts to predict the impact of σ_{Vs} reduction is to directly study the resulting UHS when applying the decomposition method. Similar to the results in Figure 6.8, Figure 6.11-a shows all UHS at AFoE= 10^{-4} computed through the convolution method for each random profile and each of the reduced σ_{Vs} scenarios. We notice a decrease in the UHS variability, over all period ranges, when the uncertainty associated to the V_s profile decreases. In order to study further the impact of the uncertainty reduction, the UHS variability is analysed for each scenario at nine different spectral periods.

Figure 6.11-b shows an example of the histograms for the four scenarios at 0.1s. The histograms are superimposed to analyse the relative reduced UHS variability. From each histogram, we identify a corresponding probability density distribution (*pdf*), considered as the best-fit in describing the data variability. At T=0.1s, beta distributions show the best fit to the histograms. Beta distributions have great flexibility in shape (skewed right or left) and can be bounded to chosen intervals by keeping the same shape.

The analysis is performed for eight other spectral periods and shown in Figure 6.11-c. Most of variabilities could be explained by beta distributions, some by lognormal, and at long periods, normal distributions were preferred. Nevertheless, the behaviour of the median $\mu_{UHS}(T)$ and standard deviation $\sigma_{UHS}(T)$ conditioned on σ_{Vs} remain complex to characterise and predict for each spectral period. While at long spectral periods (>1s), $\mu_{UHS}(T)$ is the same for each σ_{Vs} scenario, shorter periods depict a chaotic evolution of the same parameter. For some periods, μ_{UHS} increases with σ_{Vs} (T=0.2s) and decreases for others (PGA, 0.25s and 0.5s). Similarly, $\sigma_{UHS}(T)$ depicts a nonlinear dependence on σ_{Vs}. However, despite the sensitivity analysis being inconclusive to build a predictive relationship, these distributions will be used to infer prior probability distributions and measurements error functions when estimating VoI.

We recommend performing a first SRA including prior knowledge and a second one including the same median soil model V_s but with a reduced σ_{Vs}. The latter σ_{Vs} would describe the approximate reduced σ_{Vs} after performing a specific test/measurements. It is encouraged to consult experts and review past measurements and data to estimate the test bias and associated variability.

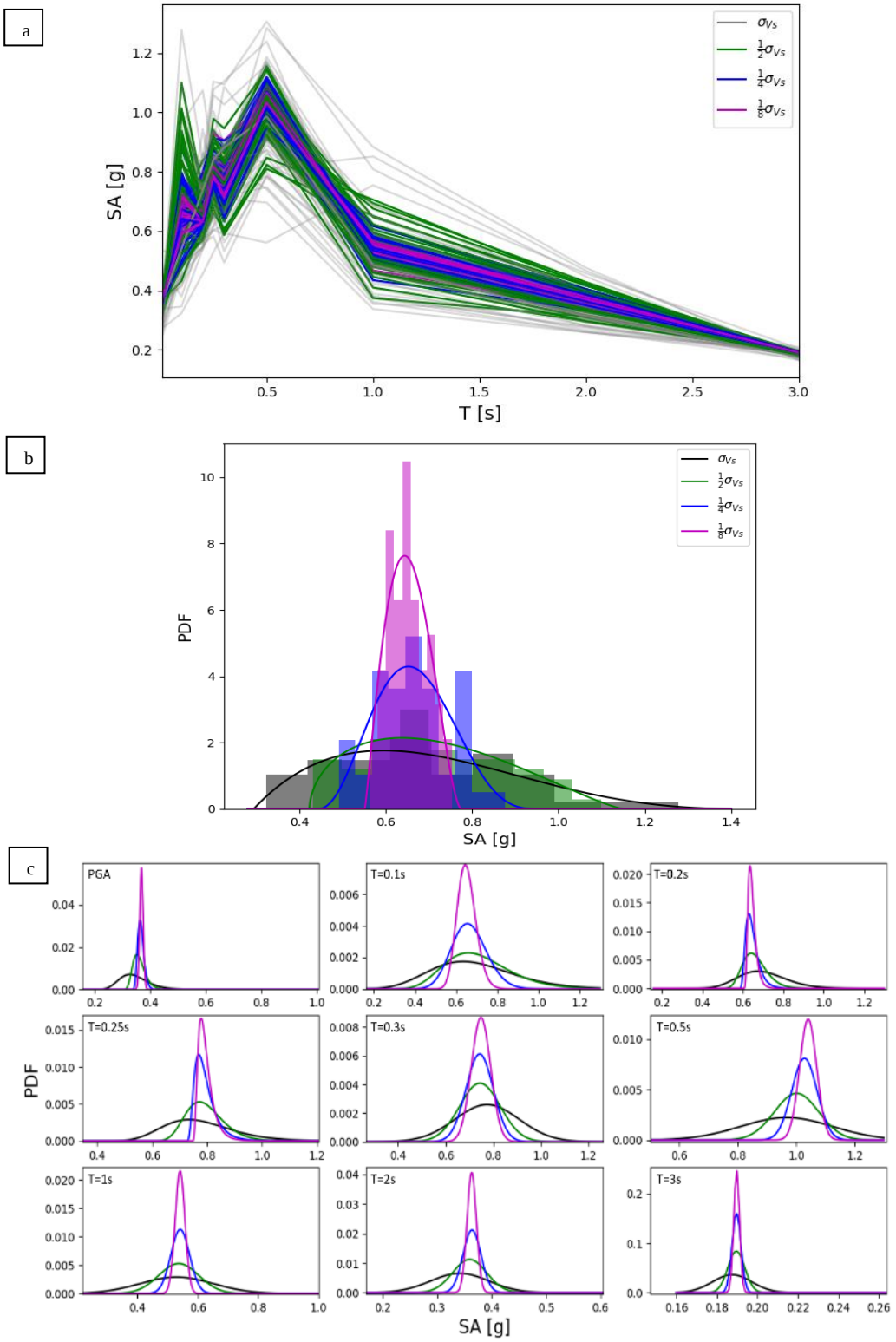

Figure 6.11: (a) UHS at AFoE=10^{-4} for σ_{Vs}, $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and $\frac{1}{8}\sigma_{Vs}$ (b) Associated histograms and probability density distributions at 0.1s (c) *pdf*s for nine spectral periods (unscaled axes)

Prior UHS distribution

In Chapter 4, VoI was estimated based on the site PGA when a single parameter (V_{S30}) is uncertain and is described by discrete binary probabilities. The VoI method was then further developed to consider a continuous prior univariate probability distribution.

In Chapter 5, bivariate joint distributions were introduced when both V_s and the thickness H were uncertain for a profile consisting of a single layer. Only the PGA was considered to describe the hazard and was included in the VoI framework as a univariate variable defined by a single distribution.

In this chapter, we complement the approach by taking into account the variability in the UHS defined over a range of spectral periods. While major components of VoI remain unchanged, defining these in light of standard PSHA practice and nuclear design requirements and regulations can be challenging as this is the first time the VoI concept is applied in this context.

The prior to be included as prior knowledge in the VoI framework is the UHS at AFoE=10^{-4} associated to the reference $\sigma_{lnVs} = 0.12$. As shown in Figure 6.11-a, the 50 UHS from the Monte Carlo simulations within STRATA are sufficient to describe the UHS variability at each spectral period T. However, to compute the PV within the VoI, a higher number of samples is needed to apply the Monte Carlo approach to approximate the integrals within the expression. The samples need to be generated from probability distributions.

The idea is to define a probability distribution that could respect the variability of the prior UHS at each spectral period. This distribution would then enable the user to generate n versions of the UHS that preserve the variability at each period. The UHS being a multivariate random variable, the suitable probability distribution achieving the latter condition would be a multivariate joint distribution.

The probability distributions shown in Figure 6.11-c in black are used to generate the prior UHS samples. Lognormal distributions were the best fit at PGA and T=2s whereas for the rest, beta distributions best describe the data histograms. A multivariate joint distribution is then built from these nine probability distributions using the Symbulate package in Python. Symbulate allows simulating samples based

on probability distributions. The multivariate joint distribution takes into account the marginal distributions of each random variable and respects the associated probability space.

Let's define x_T a random variable representing a UHS value at a spectral period T and f_T the associated probability distribution with $T \in \{0.0,0.1,0.2,0.25,0.3,0.5,1,2,3\}s$.

$$x_{PGA}, x_{0.1s}, x_{0.2s}, x_{0.25s}, x_{0.3s}, x_{0.5s}, x_{1s}, x_{2s}, x_{3s} = RV(\prod_{i \in T} f_i) \qquad (6.9)$$

RV is a class function within Symbulate that defines a random variable on a probability space. Equation (6.9) enables the simulation of a set of n samples of the UHS defined by nine variables associated to each chosen spectral period, T.

To build our prior samples, Equation (6.9) is applied to generate 5,000 UHS at AFoE=10^{-4} as shown on Figure 6.12, along with the original 50 UHS. Both median curves superimpose, except for a slightly higher median for the simulated UHS at around 0.5s. To further validate the simulated UHS, the resulting probability distributions for the nine spectral periods are compared to the computed ones in Figure 6.11-c, which shows the distribution is preserved after applying Equation (6.9). These simulated UHS represent our prior input within the VoI framework.

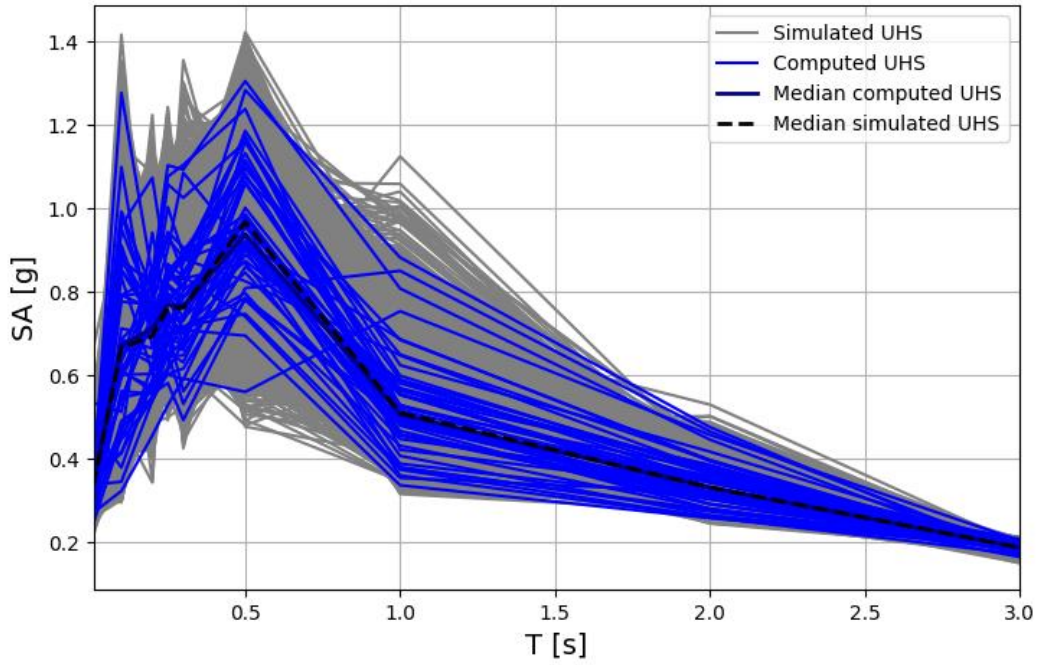

Figure 6.12: Computed (50) and simulated (5,000) prior UHS for an AFoE of 10^{-4}

Seismic design candidates

An important part of the characterisation of the prior knowledge is the definition of the decision alternatives. As part of the VoI framework, these alternatives are used to compute the expected losses. In this case study, these decisions represent the seismic design to be applied.

For NPPs, the UK guidelines (ONR, 2017) require considering a seismic design where the associated design spectrum envelops the 84th percentile of the UHS for an AFoE of 10^{-4}, equivalent to one standard deviation above the median UHS (TAG 13).

Inspired by the IBC (2012) guidelines for spectral design shapes, seismic designs are constructed in this application based on the conditional expressions in Equation (6.10) to compute SA at each T, where S_0 is the PGA and $S_{plateau}$ the spectral acceleration at the plateau value:

$$SA = \begin{cases} S_0 + \frac{S_{plateau} - S_0}{T_1 - 0.01}(T - 0.01) & 0.01 < T < T_1 \\ S_{plateau} & T_1 < T < T_2 \\ \frac{S_{1s}}{T} & T_2 < T < T_L \\ \frac{S_{1s} T_L}{T^2} & T_L < T \end{cases} \quad (6.10)$$

T_L is the long period transition spectral period which marks the transition between long and very long periods. It is fixed to 2s in this study. S_{1s} represents the spectral acceleration at T=1s and is equivalent to multiplying $S_{plateau}$ by T_2.

While design spectra are constructed using the IBC (2012) guidelines to allow flexibility of shapes and spectral values, the obtained seismic designs are subjected to nuclear design codes requirements. The NPP is assumed to belong to the seismic design category 1 (SDC1), corresponding to high hazard nuclear installations (Report SSG-67 IAEA, 2021). Depending on the hazard, the NPP should comply with the nuclear design codes and regulations by respecting a minimum target performance goal at the required hazard level (IAEA, 2019). These requirements are detailed in the next section and are implemented in the definition of the probability of failure and the expected losses.

To build the design alternatives, we consider low but also very conservative designs based on the UHS variability at AFoE=10^{-4}. Five design candidates are defined in Table 6.3. Design 1 represents the most conservative design, enveloping the overall variability of the prior UHS. Design 5 is the least resistant design with spectral accelerations below the median UHS at certain spectral periods (e.g., at T=0.5s). Figure 6.13 shows the five design spectra along with the prior UHS. These designs are used as alternative decisions in the decision tree for VoI calculations.

Table 6.3: Design spectra parameters for five seismic designs

	S_0	$S_{plateau}$	S_{1s}	T_1	T_2	T_L
Design 1	0.6	1.46	1.02	0.1	0.7	2
Design 2	0.5	1.33	0.93	0.1	0.7	2
Design 3	0.45	1.2	0.84	0.1	0.7	2
Design 4	0.4	1.0	0.8	0.1	0.8	2
Design 5	0.35	0.83	0.7	0.1	0.85	2

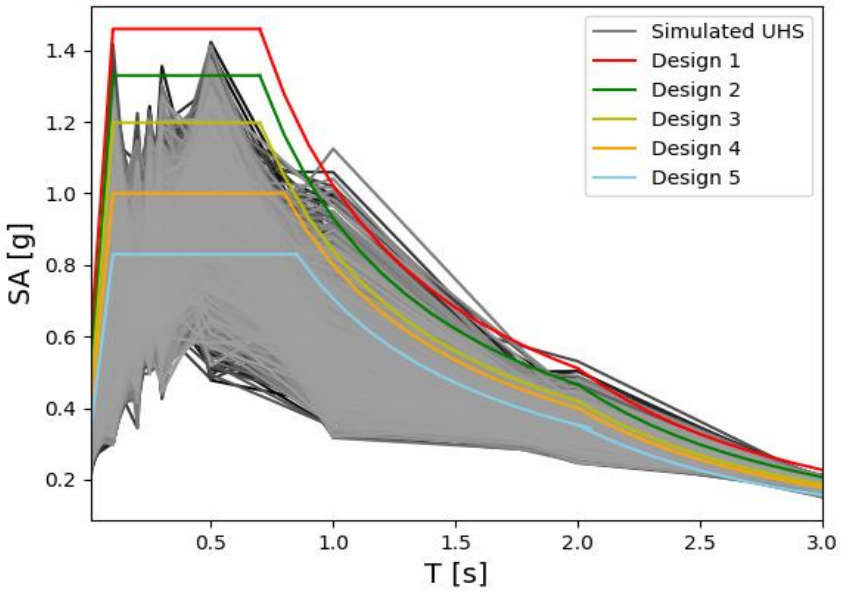

Figure 6.13: Prior UHS (5,000) and the defined five design spectra as decisions within VoI (Table 6.3)

Fragility curves and calibration to design spectra

Fragility curves need to be defined to estimate the probability of failure and the associated expected losses to incorporate in the VoI framework. These curves are a key point in assessing the seismic performance of a structure conditioned on the input motions. They are useful in making decisions regarding future structure characteristics or the improvement of an existent facility. Fragility curves are usually assessed for a structure in terms of PGA (Biswajit, 2017; Zentner *et al.*, 2017; Gkimprixis *et al.*, 2020) even if analytical methods allow the use of pseudo-spectral accelerations (PSA) for non-zero periods (NUREG/CR-3558). Fragility curves are often described by a lognormal conditional cumulative distribution with parameters representing the median capacity $C_{50\%}$, associated to the SA value at 50% probability, and the composite logarithmic standard deviation β. Because building the fragility curves from dynamic analyses for these seismic designs is out of the scope of this PhD, the parameters of the fragility curves have been inspired by the literature on this topic.

In this VoI framework, each prior UHS sample is compared to the different design spectra. Thus, considering only the fragility curves conditioned on PGA would disregard the SA at higher spectral periods and might over- or under-estimate the seismic risk at these periods. To be able to assess the overall UHS versus each seismic

design spectrum, it is important to define fragility curves for several spectral periods. Park et al. (2022) developed a method using conditional spectra (CS) to infer fragility curves at specific spectral periods from an original fragility curve conditioned on PGA. CSs are anchored to a reference UHS at specific control frequencies (f_c). That is, the spectral accelerations of the UHS, SA_{UHS}, and CS are equal at f_c. CS are usually used in the seismic design of buildings as an alternative to a reference UHS to predict a more realistic shape of response spectrum for different earthquake scenarios. The transition from a PGA-based fragility curve to a fragility curve conditioned on $SA(f_c)$ is implemented by calculating the ratio r between $SA(f_c)$ and the SA_{UHS} at 0s (i.e., PGA). In other words, the fragility curve conditioned on $SA(f_c)$ is equivalent to the PGA-based fragility curve conditioned on $SA_{UHS}(PGA) \times r$.

CS are not used in this study but the transition method is implemented to infer the fragility curves parameters at specific spectral periods T_{spec}. Besides 0.0s (corresponding to PGA), these specific periods are T=0.1s (T_1), T=0.5s and T=2s (T_L). The values of the median capacity ($C_{50\%}$) and β are inspired by the work of Zhao et al. (2020) for the collapse damage state. These values have been tailored and adjusted based on the NUREG/CR-6728 guidelines (McGuire et al., 2001) in defining the unacceptable seismic performance for a structure of seismic design category (SDC5) associated to NPPs. The performance-based method, described in ASCE/SEI Standard 43-05 (ASCE, 2005) provides criteria for the safety of nuclear facilities. The quantitative target annual probability of unacceptable performance is:

$$P_F = 10^{-5} \ /yr \tag{6.11}$$

A seismic performance P_F is acceptable for values lower than 10^{-5}/yr. Qualitatively, an accepted performance is to not exceed a specific limit state, considered as limit state D in ASCE (2005) Standard 43-05 (i.e., Essentially Elastic Behaviour). Higher limit states (C, B and A) are considered as significant inelastic deformation states inducing functioning problems that might lead to core damage. Their associated P_F are lower than 10^{-5}/yr. The risk target seismic goal is fixed in this study at 10^{-5}/yr and is used as a condition in the computation of the expected losses.

The seismic risk P_F is numerically computed as the convolution of the mean seismic hazard curve and the mean fragility curve by either of the following two equivalent equations (Kennedy, 2011):

$$P_F = -\int_0^{+\infty} p_f(a)\left(\frac{d\,H(a)}{da}\right)da \tag{6.12}$$

$$P_F = \int_0^{+\infty} H(a)\left(\frac{d\,p_f(a)}{da}\right)da \tag{6.13}$$

where $p_f(a)$ is the conditional probability of failure given the ground motion a and $H(a)$ is the mean hazard exceedance frequency for a. Equations (6.12) and (6.13) cannot be used in this study as we are unable to retrieve the hazard curves associated with the simulated UHS. As an alternative, a simplified seismic risk equation is used based on the assumption of linearity of the hazard curves when plotted on a log-log scale. Hence, over any 10-fold difference in exceedance frequencies, hazard curves can often be approximated by a power law (Kennedy, 1999; ASCE, 2005):

$$H(a) = K_1 e^{-K_H} \tag{6.14}$$

K_1 is a constant and K_H is a slope parameter defined by:

$$K_H = \frac{1}{\log(A_r)} \tag{6.15}$$

where $A_r = \frac{Sa_{0.1H}}{Sa_H}$. A_r is an amplitude ratio. Sa_H is the spectral acceleration at the target hazard level H equal to an AFoE of 10^{-4}. $Sa_{0.1H}$ is the spectral acceleration when $AFoE = 0.1 \times 10^{-4} = 10^{-5}$.

The simulated UHS samples represent the value of Sa_H at each T. In continuity of the assumption of linearity of the hazard curve on a log-log scale around H, $Sa_{0.1H}$(T) can be interpolated by inferring a linear regression using the computed hazard curves from the 50 profiles. Figure 6.14 shows the linear regression in log-log scale used in estimating $Sa_{0.1H}$ from the previously computed 50 hazard curves at PGA, 0.1s, 0.2s and 2s. This regression is performed for the four specific spectral periods to compute $A_r(T)$.

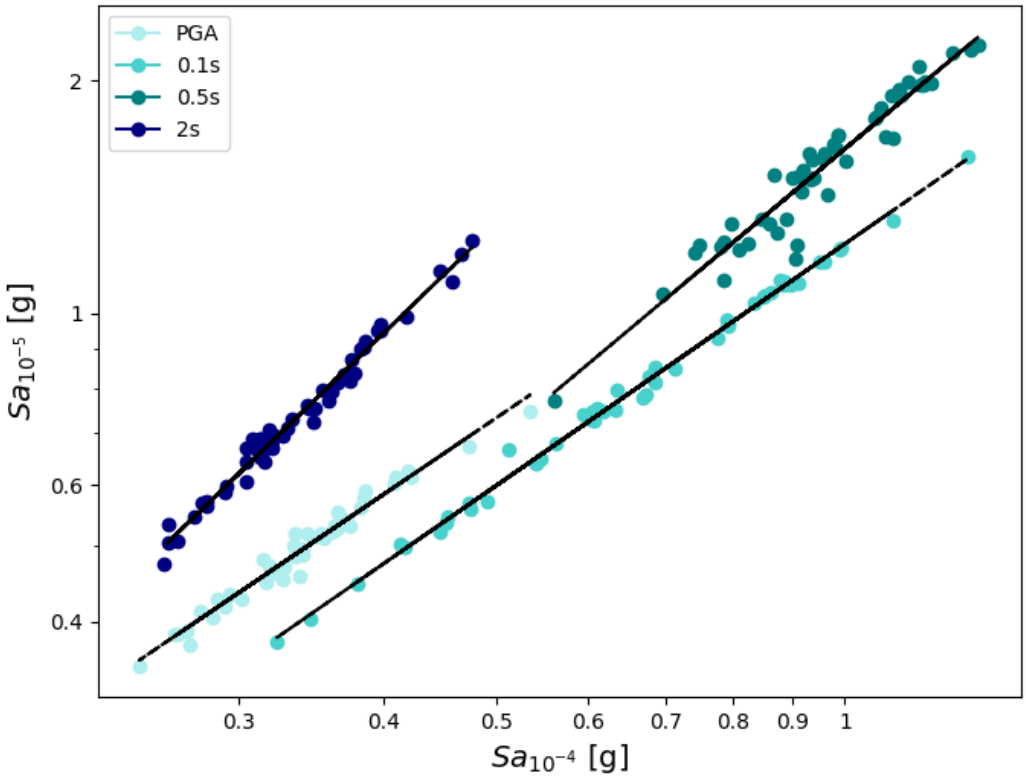

Figure 6.14: Sa at AFoE= 10^{-4} versus Sa at AFoE=10^{-5} scatter log-log plot and the associated linear regression at 0.0s (PGA), 0.1s, 0.5s and 2s. Data from SRA for 50 profiles.

The $Sa_{0.1H}$ estimation method is integrated within the seismic risk P_F function and automatised through a Python script.

With the assumption of a lognormally distributed fragility curve, p_f(a) and a hazard curve expressed as Equation (6.14), the closed form of Equations (6.12) and (6.13) to estimate the performance risk is then expressed as follows (Kennedy, 1999; McGuire et al., 2001; McGuire, 2004):

$$P_F = HF_{50\%}^{-K_H} e^{\alpha} \tag{6.16}$$

where $F_{50\%} = \frac{C_{50\%}}{Sa_H}$ and $\alpha = \frac{1}{2}(K_H\beta)^2$.

By fixing $C_{50\%}$ for PGA and using the conversion ratio *r* to infer it for the other spectral periods, the logarithmic standard deviation of the fragility curves, β, are tailored to all five seismic designs and at each specific spectral period by assuming the following:

If $SA_{UHS}(T) = SA_{Design}(T)$ then $P_F = 10^{-5}/yr$. In other words, the limit of acceptable risk is attained when a UHS is equal to the design spectrum at one of the specific spectral periods defined above. This will enable a P_F estimation aligned with the representation of the design spectra in Figure 6.13.

Using the conversion ratio r, $C_{50\%}$ at T=0.1s, 0.5s and 2s is calculated as follows for each seismic design (d) :

$$C_{50\%}(d) = \begin{cases} C_{50\%_{PGA}}(d).\frac{S_{plateau}(d)}{S_0(d)} & ;T = \{0.1,0.5\}s \\ C_{50\%_{PGA}}(d).\frac{S_{1s}(d)}{S_0(d)T_L} & ;T = 2s \end{cases} \quad (6.17)$$

The fragility curves parameters for the five designs at four specific periods are detailed in Table 6.4 and shown in Figure 6.15.

Table 6.4: Fragility curves parameters for each seismic design

	T=0.0s		T=0.1s		T=0.5s		T=2s	
	$C_{50\%}$	β	$C_{50\%}$	β	$C_{50\%}$	β	$C_{50\%}$	β
Design 1	1.80	0.485	4.38	0.410	4.38	0.510	1.53	0.330
Design 2	1.70	0.528	4.52	0.436	4.52	0.568	1.58	0.489
Design 3	1.60	0.542	4.25	0.442	4.25	0.584	1.49	0.545
Design 4	1.40	0.535	3.49	0.433	3.49	0.569	1.40	0.542
Design 5	1.20	0.528	2.87	0.422	2.87	0.550	1.20	0.547

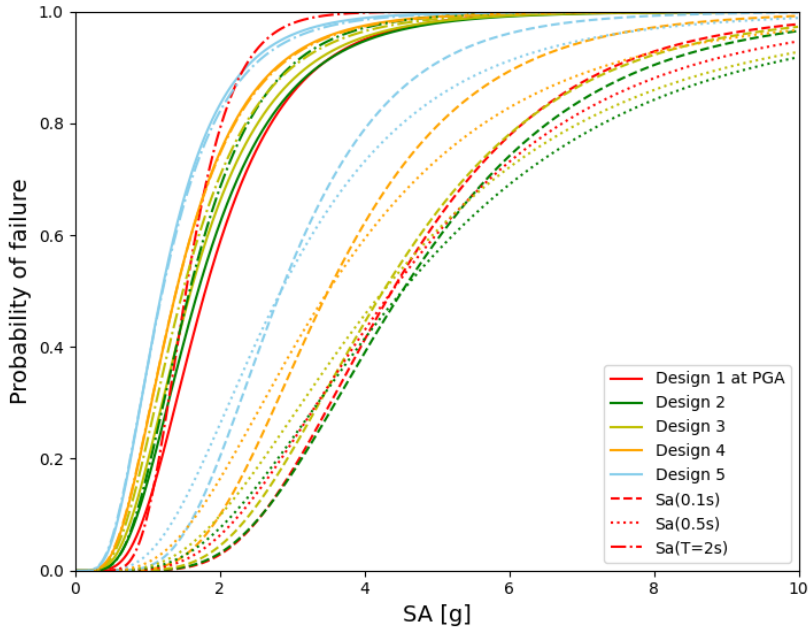

Figure 6.15: Fragility curves for five seismic designs conditioned on spectral accelerations at 0.0s (PGA), 0.1s, 0.5s and 2s

Expected Losses

The design of NPP traditionally adopts a two-level concept: A safety design that accounts for low probability of seismic exceedance for the design basis and a design for service using moderate seismic ground motion for operational limits (Katona, 2017). NPPs should be designed to withstand the Design Basis Earthquake (DBE), also referred to as Safe Shutdown Earthquake (SSE) according to U.S. terminology (USNRC, 1956).

The chosen design in our application should ensure the ultimate limit-state of safety-related systems, structures and components (SSC) accounted for in the DBE and the integrity of the reactor core. In other words, for a UHS above the chosen design spectra, the probability of SSC damages and radiological contamination is taken into account.

We consider the expected losses in case of severe damage. Some losses considered in the following are inspired by data and consequences of the Fukushima Dai-Ichi plant caused by the great Tohoku earthquake that was followed by a tsunami on 11 March 2011. Estimates of the total economic losses caused by the catastrophe are between $250 billion (IRSN, 2011) and $360 billion (Ferris and Solis, 2013), excluding the

human costs (victims compensation, houses decontamination, relocations and so forth). The manager of the powerplant, Takeshi Takahashi, estimated that the plant decommissioning could take between 40 and 50 years, making it the most complex and costly nuclear decommissioning in history (McCurry, 2013).

The total losses (EL) due to SCC damage, core melt, crisis management, decontamination, decommissioning, loss of function and potential victims' compensation are assumed to be €300 billion in this study. The duration of a decommissioning, around 50 years, is taken as the life-cycle duration in the calculation of the expected losses. The total expected loss is updated each year to account for inflation using a rate λ=3%.

The considered implementation costs (C_d) include the construction cost, the commissioning and operating phase (e.g., operation, fuel supply, insurance). The C_d affected to each seismic design follow on average the estimates of implementation costs for NPPs (Schlissel and Biewald, 2008; IAEA, 2017) and are shown in Table 6.5.

Table 6.5: Initial construction and operation costs for each seismic design

Design number	1	2	3	4	5
C_d in billion €	7	8	9	10	11

The Expected Losses function during the life-cycle (EL_{LC}) is defined as a condition of the seismic risk P_F when considering a specific UHS and a seismic design d:

$$EL_{LC}(d, UHS) = \begin{cases} C_d & if \quad \max_T P_F(d, UHS, T) < 10^{-5}/yr \\ C_d + EL\frac{1-e^{-\lambda t}}{\lambda} \max_T P_F(d, UHS, T) & else \end{cases} \quad (6.18)$$

where $EL = €300$ billion, t= 50 years and $\lambda = 3\%$.

Practically, for each UHS sample and a specific design spectrum, P_F is computed at each spectral period (0.0s, 0.1s, 0.5s and 2s). The highest P_F, associated to the highest seismic risk, is used to compute the expected losses to be incorporated within the VoI framework. If the highest P_F is below $10^{-5}/yr$, the design is accepted and considered

safe (i.e., acceptable level of risk). This also means, from the definition of the chosen design fragility curves, that the UHS is below the design spectrum. Therefore, the expected losses only represent the construction, functioning and maintenance costs. On the contrary, if the highest P_F is above $10^{-5}/yr$, the target annual probability of unacceptable performance is exceeded which puts the design at risk. This probability of failure is then multiplied by the expected losses due to failure and added to the initial design cost to constitute the total expected losses.

6.3.2 VoI framework implementation: Using Full UHS

In the previous subsection, all components and inputs of VoI have been defined based on available literature, engineering judgement and cost estimates. The scope of this chapter is to build a framework for the implementation of VoI within PSHA to estimate an upper budget value for data collection. Assessing the VoI is estimating the expected avoided losses from reducing the epistemic uncertainties of one or more parameters. The presented framework is built to estimate the expected avoided losses in reducing the V_s profile uncertainties within a six-layer profile on top of an elastic bedrock. Because of the complexity in predicting the expected losses when σ_{Vs} is varied, the considered prior within VoI is the UHS variability distribution associated to σ_{Vs} after performing SRA and site-specific PSHA.

In this section, the PV is calculated based on simulated UHS from the prior distributions in Figure 6.11-c. Then, the likelihoods for different σ_{Vs} and median observations are built to produce the posterior UHS distributions. Finally, potential observed UHS are simulated and used in assessing VoI for different hypothetical measurements.

Prior Value

The PV estimates the expected losses when deciding based on prior knowledge, i.e., before the collection of additional data.

We recall the expression of the PV:

$$PV = \max_{d \in D} \{ \int o(d,x)\, p(x) dx \} \tag{6.19}$$

where $\mathcal{D}$ is the domain of decisions d, $o(d,x)$ represents the outcomes for a decision d if UHS is equal to x and $p(x)$ is the prior probability density function of x.

The integral can be estimated by applying the Monte Carlo approximation when generating n samples from the prior distribution through Monte Carlo simulation. In the previous subsection, $n = 5{,}000$ prior UHS are simulated (Figure 6.12) to compute PV as follows:

$$PV \sim \max_{d\epsilon \mathrm{D}} \{\frac{1}{n}\sum_{k=1}^{n} EL_{LC}(d, UHS^k)\} \tag{6.20}$$

where n is the number of samples, $EL_{LC}(d, UHS^k)$ are the expected losses when design d is considered and the true site-specific hazard is equal to UHS^k.

Expected Value of Perfect Information (EVPI)

The EVPI is estimated by considering that future measurements might determine the exact site properties, implying that $\sigma_{Vs} = 0$. These measurements would infer the true V_s profile and result in a true UHS for an AFoE of 10^{-4}. In practice, the site-specific PSHA includes other sources of uncertainties and the UHS variability might only decrease. Nevertheless, this study case considers the V_s profile as the only source of uncertainty and assumes that a perfect test (one similar to high-resolution downhole measurements) could determine the exact hazard level at the site. Practically, the EVPI is only computed as a measure of an upper limit to not exceed when deciding on investing in ground investigation surveys.

The Posterior Value (PoV) after a perfect measurement is:

$$PoV = \int \max_{d\epsilon D}\{ EL_{LC}(d, UHS)\} p(UHS)\, dUHS$$

$$\sim \frac{1}{n}\sum_{k=1}^{n} \max_{d\epsilon D}\{EL_{LC}(d, UHS^k)\} \tag{6.21}$$

The EVPI is then calculated as follows:

$$EVPI = PoV - PV$$

Expected Value of Imperfect Information (EVII)

Let's recall the expression of the PoV in estimating the EVII:

$$PoV = \int_y p(y).\max_{d\in D}\{\int_x o(d,x)p(x|y)dx\}\,dy \qquad (6.22)$$

The variable y represents the possible median observations after hypothetical measurements from a specific test. As defined in Chapter 4, $p(y)$ is the observations marginal *pdf* and $p(x|y)$ the posterior probability of having x if y is observed. By making an analogy in our current VoI application, x and y are the true and 'observed' state resulting UHS at AFoE=10^{-4}, respectively.

a- <u>Likelihood</u>

To estimate the marginal and posterior probability distributions, the hypothetical test likelihood, $p(y|x)$, should be defined. This function expresses the probability of an observation being true to the real hazard level. In this present application, the observation is the UHS and is therefore considered *indirect*. In fact, the *direct* observed measurements are initially defined by a median V_s, μ_{Vs}, and an associated standard deviation, σ_{Vs}. Therefore, the UHS variability is considered as an *indirect* observation as it represents the expression of the *direct* measurements' uncertainties. Assuming the UHS as a potential observation is viable, following the modification of the traditional full convolution method to isolate the contribution of the epistemic uncertainties in site-specific PSHA. Hence, UHS variability becomes an articulation of the epistemic uncertainty.

A measurement likelihood can be estimated from the error function of the conducted test. We propose an approach to determine the error function of three hypothetical tests, aiming to reduce the prior σ_{Vs} by half, one quarter and one-eighth ($\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and $\frac{1}{8}\sigma_{Vs}$). 1D linear-equivalent SRA are run using the same μ_{Vs} and for each of the reduced σ_{Vs}. Site-specific PSHA is then performed to infer a median and UHS standard deviation. These calculations have been performed in section 6.3.1 and the probability functions at each spectral period T and for each test are shown in Figure 6.11-c. These probability functions are considered as likelihood functions for an observation with median μ_{Vs} and a reduced σ_{Vs}.

In order to define likelihood functions for any random observation, error functions are constructed as follows.

An example of an error function at PGA when the V_s standard deviation is reduced by half ($\frac{1}{2}\sigma_{Vs}$) is shown in Figure 6.16. The solid line *pdf*, labelled as *computed UHS*, represents the probability distribution of the UHS spectral accelerations at 0.0s (i.e., PGA) when estimating the site-specific PSHA based on the soil V_s model in Table 5.2 with $\frac{1}{2}\sigma_{Vs}$. This probability distribution is estimated following section 6.3.1 (also shown in Figure 6.11-c). The error function, $e(z)$, is built based on the same distribution parameters and shifted using the *location* parameter to align the maximum *pdf* value at 0 (dashed line). This is based on the assumption that the measurements have the highest probability of being true to the real V_s. For some tests, measurements can have a high probability of being biased by a certain value. To implement that, the error function can be shifted to translate this biased value.

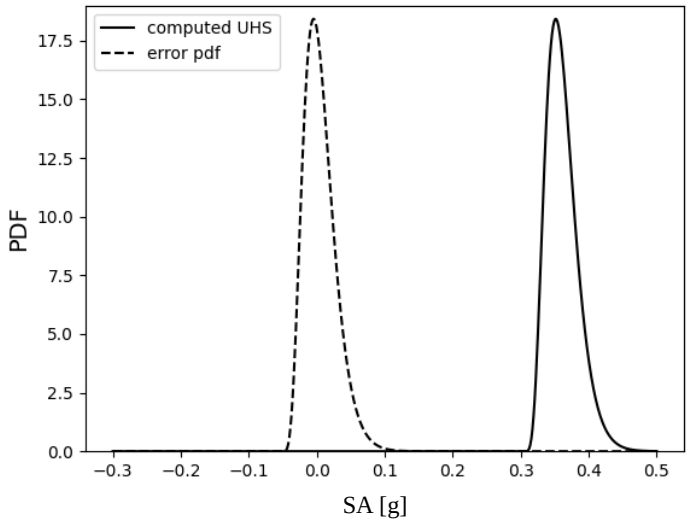

Figure 6.16: Probability distribution of computed UHS at 0.0s (PGA) and the associated test error function

The likelihood function is then simply calculated as follows:

$$p(y|x) = e(y - x) \tag{6.23}$$

This workflow needs to be repeated for each of the specific spectral periods (PGA, 0.1s, 0.5s and 2s) that define each UHS.

b- Observations

EVII calculations require the simulation of potential observations. These observations represent the resulting possible UHS after a high number of measurements using a chosen test. To simulate the observations, we proceed as follows for each spectral period $T \epsilon T_{spec}$:

- Pick a spectral acceleration from the prior probability distribution f_{prior}, $x_{prior}(T)$.
- Pick an error value from the error probability distribution e_T, $x_e(T)$
- The observed spectral acceleration is $y_{obs}(T) = x_{prior}(T) + x_e(T)$

This process is repeated a high number of times to construct the observed probability distributions for each T. Using Equation (6.9) enables a multivariate joint probability distribution defined on T_{spec} to be built. This can be used to simulate N observations. An example of 3,000 simulated observed UHS for measurements producing uncertainties of $\frac{1}{2}\sigma_{Vs}$ along with the prior UHS is shown in Figure 6.17.

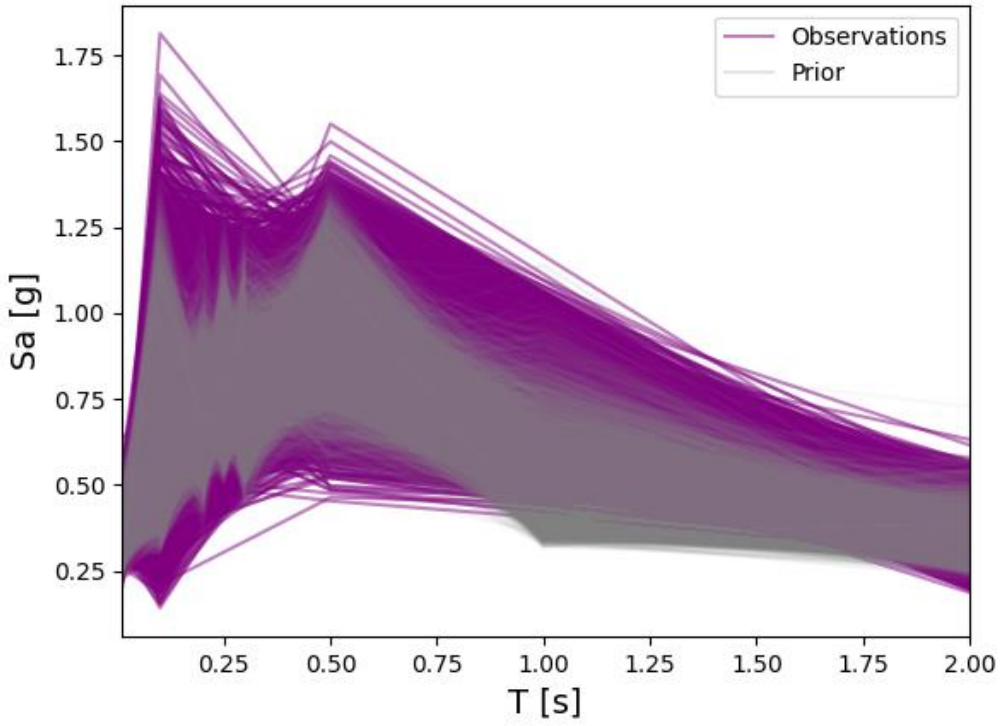

Figure 6.17: Prior and observed UHS at AFoE=10^{-4} for V_s uncertainty=$\frac{1}{2}\sigma_{Vs}$

Observations are meant to envelop the prior as a consequence of possible measurement errors. This step also allows to consider the possibility of a median soil model, μ_{Vs}, being different from the prior median.

c- Posterior

Posterior probability distribution functions combine the test likelihood and prior knowledge. The posterior probability for an observation y_{obs} at a spectral period T, $p(x|y_{obs})$, is constructed by computing $p(x|y_{obs})$ for each $x = x^k$ within the definition domain of f_{prior}(T):

$$p\left(x^k \middle| y_{obs}(T)\right) = \frac{f_{prior}(x^k,T)\, e_T(y_{obs}-x^k)}{p_m(y_{obs})} \quad (6.24)$$

$p_m(y_{obs})$ is the marginal probability of the observation y_{obs} at T:

$$p_m(y_{obs}) = \sum_k f_{prior}(x^k, T)\, e_T(y_{obs} - x^k) \quad (6.25)$$

An example of posterior *pdf* UHS at $T \epsilon T_{spec}$, $f_{post}(T)$, for the case $\frac{1}{2}\sigma_{Vs}$ is shown in Figure 6.18 along with the likelihood for an observed UHS (UHS_{obs}) and its prior distribution. The posterior is presented as a trade-off between the prior and the likelihood. We can notice that for the *pdf*s at T=0.5s, all probability distributions share the same mean/median. This case is unusual and is a consequence of a median observation being equal to the value of highest prior probability.

Equation (6.9) is then used to compute the multivariate joint distribution from all $f_{post}(T)$ to construct the posterior UHS for each observation UHS_{obs}.

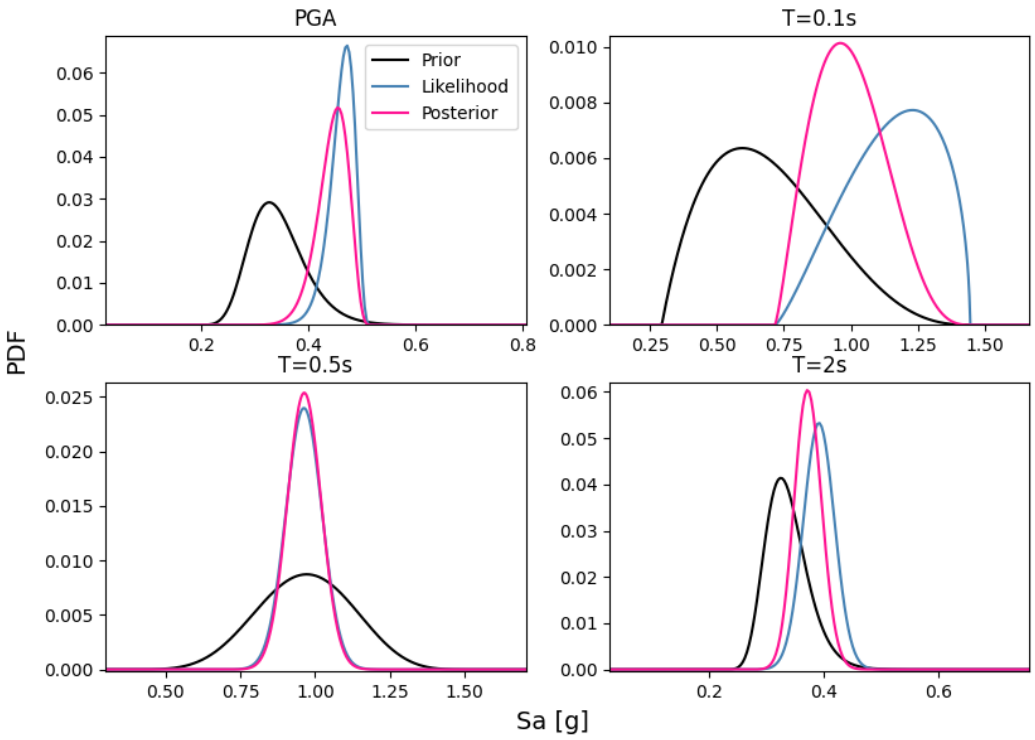

Figure 6.18: Prior and likelihood-posteriors UHS PDFs for four spectral periods for V_s uncertainty=$\frac{1}{2}\sigma_{Vs}$

d- <u>Final Implementation</u>

In this section, the full implementation of previously defined inputs is generalised. This represents the last step in VoI calculations. We start by giving notations to all variables. Some variables may have one, two or three subscripts and superscripts as we deal with multivariate variables. We then detail the steps leading to computing the EVII.

- *Notations*

Let the prior and observed UHS be defined as a multi-dimensional variable defined at spectral periods $T \in T_{spec} = \{T_1, .., T_s\}$:

$$UHS_{prior} = \{x_1, \dots, x_s\}$$

$$UHS_{obs} = \{y_1, \dots, y_s\}$$

If n is the number of simulated priors and N the number of observed UHS then each UHS_{prior} and UHS_{obs} samples are defined as follows:

$$UHS_{prior}^{k} = \{x_1^k, \dots, x_s^k\} \quad ; k \in \{1, \dots, n\}$$

$$UHS_{obs}^{j} = \{y_1^j, \dots, y_s^j\} \quad ; j \in \{1, \dots, N\}$$

f_{prior} is the prior probability distribution defined for each $T \in T_{spec}$:

$$f_{prior} = \{f_{prior_1}, \dots, f_{prior_s}\}$$

e is the error probability distribution function defined for each $T \in T_{spec}$:

$$e = \{e_1, \dots, e_s\}$$

f_{post} the posterior probability distribution defined for each $T \in T_{spec}$:

$$f_{post} = \{f_{post_1}, \dots, f_{post_s}\}$$

- *Implementation*

After computing the PV, we detail the steps in estimating the PoV. The first step is to generate *N potential* observed UHS using the prior and the chosen test error function as follows:

For each $T_i \in T_{spec}$:

- Sample x_i^* from f_{prior_i}
- Sample e_i^* from e_i
- Compute $y_i^* = x_i^* + e_i^*$

The sampling is repeated X times to identify a probability distribution for y_i.

The UHS_{obs} multivariate joint distribution is then assessed by coupling y_i probability distributions at each T_i (Equation 6.9). N number of UHS_{obs}^{j} are sampled using the same equation.

The next step is similarly iterative over the number of sampled observations. This sampling according to *pdf*s permits application of the Monte Carlo approximation to estimate the integrals in the VoI equations.

For each simulated observation UHS_{obs}^{j}:

1. For each $T_i \in T_{spec}$:

   Estimate $f_{post_i}(x|y_i^j)$ from the calculation of $f_{post_i}(x_i^k|y_i^j)$ for $k \in \{1, \dots, n\}$:

   $$f_{post_i}(x_i^k|y_i^j) = \frac{f_{prior_i}(x_i^k)e_i(y_i^j - x_i^k)}{p_m(y_i^j)} \quad (6.26)$$

   where $p_m(y_i^j) = \sum_{k=1}^{n} f_{prior_i}(x_i^k)e_i(y_i^j - x_i^k)$

   $f_{post_i}(x_i^k|y_i^j)$ assigns a *pdf* value at each x_i^k in the prior probability space to define $f_{post_i}(x|y_i^j)$.

2. Generate UHS_{post}^{j}:

   Use equation (6.9) to build the multivariate joint distribution as follows:

   $$f_{post_1}, \dots, f_{post_s} = RV(\prod_{i=1}^{s} f_{post_i}) \quad (6.27)$$

   Simulate m posterior UHS where each sample is denoted:

   $$UHS_{post}^{j,l} = \{z_1^{j,l}, \dots, z_s^{j,l}\} \quad ; l \in \{1, \dots, m\}$$

3. Estimate the expected losses $EL_{LC}(d, UHS_{post}^{j,l})$ using the expression (6.18) for each seismic design $d \in D$ for all m samples

Finally, the EVII is computed as follows:

$$EVII \sim \frac{1}{N}\sum_{j=0}^{N} \max_{d \in D} \left\{\frac{1}{m}\sum_{l=0}^{m} EL_{LC}(d, UHS_{post}^{j,l})\right\} - PV \quad (6.28)$$

When N and m are high enough to characterise the observations and posterior *pdf*s, respectively, the Monte Carlo approximation can estimate associated integrals by simply averaging the expected losses at each sample by the total number of samples.

6.4 Results and conclusions

The above framework is implemented to estimate the PV, EVPI and EVII using $n = 5{,}000$ simulated UHS_{prior} built from prior knowledge when considering the median V_s soil model in Table 5.2 and the reference V_s standard deviation $\sigma_{lnVs} = 0.12$. Results for total losses if failure EL= € 300 and € 200 billion are shown in Table 6.6.

For the initial case of EL= € 300 billion, PV has been estimated at € -7.306 billion, considered as a negative value to account for an outcome in terms of losses. The extremely high value is due to considering severe damage when underestimating the seismic design but is mainly explained by the initial construction costs where the less resistant design is estimated to cost € 7 billion.

Table 6.6: Multivariate uncertain parameter results of PV, EVPI and EVII for three different tests and two different EL inputs

Total Losses if failure, EL (billon €)	PV (billon €)	EVPI (million €)	EVII in million €		
			$\frac{1}{2}\ \sigma_{Vs}$	$\frac{1}{4}\ \sigma_{Vs}$	$\frac{1}{8}\ \sigma_{Vs}$
300	-7.3	31.9	8.3	18.6	30.6
200	-7.2	13.1	1.2	3.5	9.8

This EVPI has a high value compared to the costs of possible different measurements from non-invasive and even invasive ground investigations. This value is tightly related to the input value of losses (EL) when severe damages occur. We can notice a relatively decrease in PV, by € 100 million, when EL=€ 200 billion.

The EVII was assessed for three different types of measurements by assuming different possible uncertainty reductions. All values were calculated and reached stability for the numbers of simulation $N = 3{,}000$ and $m = 500$. It is interesting to notice that the larger the uncertainty reduction, the higher the EVII, and that this value converges towards the EVPI. We note that for EL= € 300 billion, reducing σ_{Vs} down to one eighth of its value results in an EVII averaging 96% of the EVPI. EVII for $\frac{1}{2}\sigma_{Vs}$, $\frac{1}{4}\sigma_{Vs}$ and

$\frac{1}{8}\sigma_{Vs}$, respectively represent 0.11%, 0.266%, 0.438% of the initial cost of designing Design 5, while it averages 0.456% for the EVPI. This result validates the framework and the results from the discrete and continuous studies in Chapter 4 and 5.

The number of samples $\{n, N, m\}$, related to the number of iterations, need to be chosen carefully. Depending on the distributions type and parameters, stable results can be achieved by studying the variability of the Monte Carlo estimator during simulation to insure its convergence.The Monte Carlo approximation often requires a high number of random samples, depending on the probability space and the sensitivity of the results to each sample. The higher the number, the higher the computational costs. It is not possible to advocate for a certain number of samples as it depends on the application, the nature of the outcomes and the elements at risk. Nevertheless, we can emphasise that the sampling should be high enough to fully describe the associated probability distribution.

Depending on the chosen ground investigation, one must be able to quantify the degree of uncertainty in the measurements. Some attempts have been made to study the uncertainties for different V_s measurement techniques (Passeri et al., 2019; Toro, 2022) such as invasive ground investigation (e.g., boreholes) or non-invasive (e.g., surface waves). Such quantifications are essential in defining the likelihood functions within the VoI calculations. During this research, it was challenging to obtain data from a particular ground investigation with consistent estimation of uncertainties. This is due to the variability in size and extent of ground investigation techniques, the incorporation of other types of measurements, the use of proxies and so forth. Nevertheless, the use of past data from similar projects along with expert elicitation might give an insight on the potential uncertainty reduction that a particular ground investigation technique could provide.

The VoI approach applied in inferring an appropriate seismic design can be further developed by taking into account other sources of uncertainties (e.g., dynamic soil properties, hazard on rock, fragility curves). The challenge remains in decoupling each source of uncertainty to study its reduction on the overall outcomes. VoI seems to

constitute a promising tool in decision-making. It can be developed to integrate higher levels of complexity, to make it suitable for a wide range of applications.

While these calculations might be a helpful indicator in setting a budget limit when deciding to collect additional information, they remain rough estimates. The term ‘expected’, added to almost all expressions, indicates the uncertainties arising from the assumptions, approximations and attempts to predict measurements and systems behaviour. The VoI concept might not constitute an exact science, but it is based on approved decision strategies, prior knowledge and the science intrinsic to the specific application.

7 Conclusions and recommendations

7.1 Key findings

The Value of Information (VoI) concept has been developed and tailored to respond to realistic decision-making situations within seismic hazard assessment (SHA) and earthquake engineering. SHA inputs suffer from epistemic uncertainties and aleatory variabilities, often leading to a high total uncertainty in the overall hazard estimates. Generally, guidelines and requirements are used to design a structure to an acceptable level of safety. Nevertheless, large uncertainties in hazard estimates could lead to high design spectra, resulting in increased construction costs. These designs, however, might be over-conservative because of an over-estimation of the actual seismic hazard. VoI was used to assess the benefit of investing in data collection to better constrain the seismic hazard and, ultimately, to design accordingly.

This thesis presented a gradual evolution of the method, from basic and simplified applications to more complex implementations that follow modern practice. This gradual upgrade was essential to: (a) provide a clear understanding of the influence of the various input variables, (b) validate the methodology by means of comparison of outputs obtained using applications of different levels of complexity, and (c) construct a comprehensive framework and justified guidelines for industry and other stakeholders.

We defined a general application involving SHA to help decide on the collection of data to reduce uncertainties. The VoI concept was chosen as part of a decision-making strategy to make better decisions on the level of structural seismic design in light of uncertain seismic hazard estimates. More specifically, we assessed the benefit of additional information in reducing the expected losses and in inferring an optimal design, respecting safety requirements and budget limitations. The VoI methodology was tailored to several configurations, decision criteria and characterisations of parameter uncertainties. The various applications are summarised in the following.

In Chapter 4, the hazard at the site of interest was assumed to be described by the product of the expected peak ground acceleration (PGA) on a reference rock for a specific return period and a frequency-independent amplification factor, *AF*. *AF* was only assessed using a single site condition proxy, Vs_{30}, which represents the only source of uncertainty. The VoI concept was tailored to estimate the value of reducing the Vs_{30} uncertainty when defined by: (a) a discrete binary variable and (b) a continuous probability distribution. The design criterion for a four-storey three-bay reinforced concrete building was only based on the PGA, and VoI outcomes were expressed in monetary units representing expected avoided losses. Sensitivity analyses showed that VoI is strongly dependent on the prior probability distributions. Specifically, when conducting measurements to reduce uncertainties, the Expected Value of Perfect Information (EVPI) and Imperfect Information (EVII) increases linearly with increasing prior uncertainty. That is, the higher the prior uncertainty, the more benefits are gained from reducing it. Moreover, the method was further validated by observing that more accurate and precise measurements increase the EVII and that this value converges towards the EVPI.

In Chapter 5, a linear-elastic soil response analysis (SRA) was performed to analytically infer a frequency-dependent *AF*. Site conditions were no longer approximated by a single uncertain proxy but by the shear-wave velocity (V_s) and thickness (*H*) of a single soil layer overlaying an elastic bedrock. The VoI method was tailored to include a bivariate joint distribution translating the joint uncertainty of V_s and *H*. Several calculations were performed to assess the value of additional information on V_s only, *H* only and both. The results were compared to the value of obtaining direct measurements on *AF* (i.e., a different type of measurements). These showed a consistency between the two frameworks (i.e., univariate and bivariate). Similar to Chapter 4, we find a linear dependence of EVII on the measurement's degree of accuracy (i.e., standard deviation of the error function). Another interesting finding suggests that in some cases, it is more valuable to obtain high quality measurements on only one parameter than poor quality data on both parameters. In this chapter, a third application consisted in numerically performing linear-elastic SRA using the STRATA software for a six-layer V_s profile. VoI sensitivity analysis for direct measurements of *AF* using the univariate framework showed similar patterns

but lower values than the previous application. This was explained by a lower variability in the *AF* when compared to the single-layer application. The results are aligned with the findings in Chapter 4.

Finally, Chapter 6 presented an upgrade to the application and the method at several levels. These upgrades were made to propose an approach that is more consistent with modern practice in SHA and in earthquake engineering. The prototype structure to be designed was defined as a critical facility, specifically a nuclear powerplant (NPP). *AF*s were estimated from linear-equivalent SRA to take into account the variability resulting from both V_s profile uncertainties and soil nonlinearity. We proposed an approach to: (a) decouple the contribution of V_s from the overall site hazard variability and (b) obtain the hazard curves associated to each randomised V_s profile, through modification of the full convolution approach (Bazzurro and Cornell, 2004) for site-specific Probabilistic Seismic Hazard Assessment (PSHA). The design criterion was based on the obtained Uniform Hazard Spectra (UHS) over a wide range of spectral periods. Fragility curves were inferred for each design spectrum alternative (in the decision domain) and at several spectral periods to assess the expected losses due to severe damage by including current NPP-specific guidelines and requirements ensuring safety assurance. This study required a multivariate approach for the VoI calculations to consider decisions that are dependent on UHS defined by more than one variable (i.e., over several spectral periods). More complex VoI calculations were performed by characterising the UHS variability at each spectral period, constructing the associated multivariate joint distributions and integrating them to assess the EVPI and EVII when obtaining measurements that reduce the uncertainties on V_s (i.e., σ_{V_s}). Our results showed that EVII converges towards the EVPI when the quality and/or quantity of measurements increase. This finding is in accordance with the results in the previous chapters. Compared to the previous applications for a non-critical building, these results show higher VoI estimates irrespective of the quality of measurements. This was identified to be mainly due to the definition of the expected losses, including high construction costs (i.e., in billion €). Moreover, the losses in case of severe damage are high considering the life and environmental threats induced by severe damage to an NPP.

7.2 Guidance for future applications

A summary of the major steps and VoI analyses components are shown in Figure 7.1. This conceptual and analytical framework can be used as a guideline to perform EVPI and EVII assessments. Key inputs and computations are represented, with arrows indicating the timeline direction and the dependence amongst the different actions and quantities. This framework highlights a more complex EVII methodology than for EVPI. The approach for EVII is characterised by additional inputs and a two-level Monte Carlo simulation (loop within a loop) indicated by the incremental rectangular nodes and conditional diamond nodes. We believe that this framework representation can be used as a support for VoI assessment for various applications, and not limited to SHA and earthquake engineering.

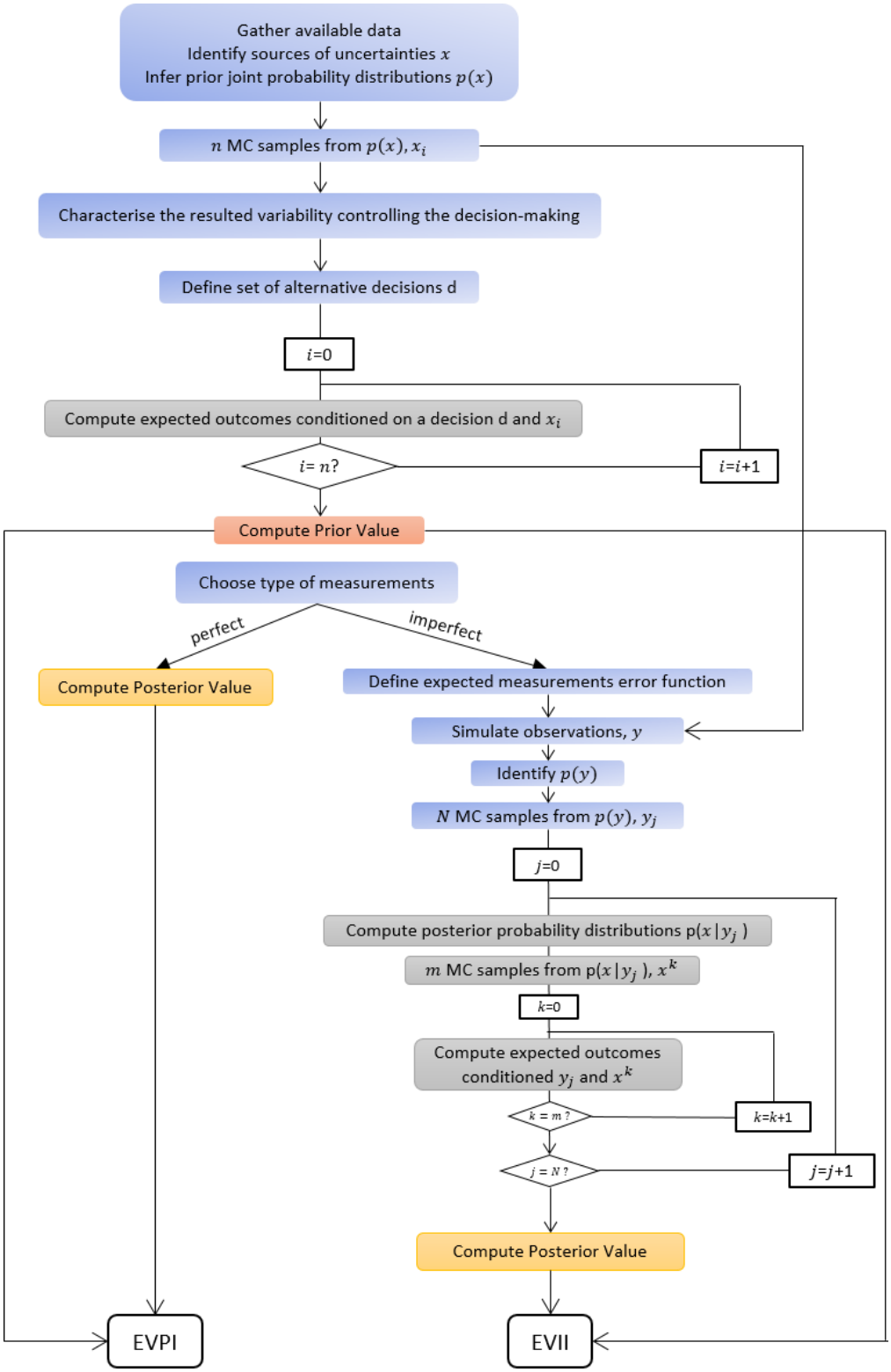

Figure 7.1: Conceptual and analytical framework of VoI approach to compute the EVPI and EVII. Coloured nodes represent actions: grey nodes represent iterative computations, and red and yellow nodes the final computations of the main elements of VoI (Prior Value and Posterior Value). Loops are represented by incremental rectangle nodes and conditional diamond nodes and indicate iterative computations from Monte Carlo simulations samples. Arrows translate the order of the actions, as well as the dependences between actions and results

The VoI assessments presented in this thesis demonstrate a high degree of consistency in their results, as well as in the sensitivity analyses conducted for various configurations and methodologies. Sensitivity analyses were tools of upmost importance to identify the elements, inputs and steps that contribute the most to VoI estimates. In particular, the findings highlight the following points:

- Prior knowledge: Sufficient amount of time should be dedicated to gather all available information to define the up-to-date prior knowledge based on the available data and/or expert elicitation. This knowledge needs to be translated into probability distributions that best fit the available data.
- Parameter estimation: Probabilistic sensitivity analyses should be performed using Bayesian Networks, influence diagrams, decision trees or simply Bayesian inference to identify parameters that most affect the decision criterion (e.g., PGA and UHS in this thesis).
- Relevant information: Ensure that the measurements to be obtained in the future are relevant to the characteristics of interest (i.e., the uncertain parameters).
- Targeting: Decouple the influence of the parameter of interest from the overall resulting variability. This permit obtaining VoI estimates that only reflect the value of reducing the measured parameter in the future.
- Error function: Ensure a good characterisation of the variability of future measurements from a given ground investigation (GI) technique. Identify if the technique is biased or not as well as its ability in constraining the characteristic of interest. Translate the characterisation into probability distributions to construct error functions.
- Outcomes: Frame the decision strategy in terms of a number of decision alternatives (based on the prior knowledge) and define the consequences of making one decision over another (e.g., in terms of gain or avoided losses).
- Integral approximations: Integrals in VoI equations can be estimated via Monte Carlo approximation by sampling from a probability distribution. The stability of VoI results is directly linked to the number of samples. It is important to sample a sufficient number to reach a stable value whose variability is negligeable compared to the mean results.

- Expected outcomes: The expected outcomes combine the outcomes for a decision and the probability of occurrence of this outcome. The decision-maker should be cautious in building the dependence relationship between the outcome relative to a decision, and the probability function of the characteristic of interest.

The 'Targeting' and 'Expected outcomes' points mentioned above are of upmost importance and require careful analysis. In fact, it is crucial to correctly characterise the relationship between the variability of the parameter that directly influences the decision (e.g., hazard level) and the sources of uncertainties that are aimed to be better constrained (e.g., V_s) through data collection. The relationship is not always straightforward when other uncertainties, that are irreducible, contribute to the overall variability. The latter uncertainties can be marginalised to eliminate their contributions and make sure to only consider the target parameters (e.g., Baio, 2018). In Chapter 6, two sources of uncertainties were affecting the site-specific PSHA results: uncertainty from V_s profiles and the motion-to-motion variability (i.e., nonlinear effects). The modification of the full convolution approach enabled decoupling each source and explicitly studying the contribution of V_s uncertainty on the UHS variability. The outcomes associated to each decision and resulting UHS were computed within the VoI framework, at each iteration.

Another approach is to study and characterise in a more direct way, the outcomes and their expected probabilities. That is, integrating in the calculations a probability distribution on the expected outcomes, conditioned on the prior probability of a characteristic of interest x. The main difficulty lies in predicting the potential consequences resulting from making a particular decision when x is uncertain. Several studies have tackled this issue (Briggs et al., 2006; O'Hagan & Stevens, 2001). Advances were mostly made within the healthcare and medical domains (e.g., Baio, 2018). These studies recommend characterising the variability in the expected outcomes given an uncertain parameter x in a preliminary analysis, outside of the loops within the VoI approach. In other words, a conditional probability distribution of the expected outcomes (EO), $p(EO, d|x)$, is constructed and directly included within the calculations. Prior and posterior probabilities of x are then translated to prior and

posterior probabilities of outcomes conditioned on new evidence. Some methods have proven to achieve the latter, such as using Gaussian process and Meta-modelling through generalised additive models (GAM) (e.g., Brennan et al., 2007; Jalal et al., 2015) or Integrated nested Laplace approximations (e.g., Heath et al., 2016). The main motivation of implementing these approaches is to reduce the computational cost.

As part of these guidelines, the 'Integral approximation' step is a requirement that needs to be fulfilled to obtain reliable and stable VoI estimates. On one hand, Monte Carlo simulations consist in sampling several values from a defined probability distribution. On the other hand, the Monte Carlo approximation enables calculation of the average computed quantities from the samples to estimate integrals. In Chapter 4, we put in evidence the strong dependence of this estimate on the number of samples, where the higher the number, the more stable the results. It is necessary, before using VoI estimates to justify decisions, to verify that the Monte Carlo simulations (a) best describes the probability distribution and (b) results in stable results with minimal variance. Some examples of currently used methods to infer the optimal number of samples are mentioned in the next section.

We have identified several situations and configurations where VoI can be used to inform better decisions on the collection of data. Table 7.1 gives a brief overview of three examples of situations for the decision-maker as well as the associated recommended workflow. We call the framework of a single VoI analysis (Figure 7.1) VoI_s, defined for a target source of uncertainty Y to be reduced when conducting a test T. The final decision is based on the Net Benefit (NB), representing the difference between VoI_s and the cost of the test T. NB represents a comparative measure that helps: (a) identify the best combination of tests and/or the parameters where data collection has the higher probability of leading to improved expected outcomes (e.g., less losses, maximum gain), and (b) ensure that the budget limitations are respected. Similarly, NB can be a key element in justifying future investments.

In the three cases presented in Table 7.1, VoI analysis aims to respond to these questions:

Case 1: Should I perform test T to reduce the only source of uncertainty Y?

Case 2: What is the optimal test to reduce the only source of uncertainty Y amongst T_s number of tests?

Case 3: What is the source of uncertainty Y_j amongst a set Y that I should prioritise better constraining from data collection? And what is the optimal test?

VoI is a polyvalent tool that can respond to various questions and can be adapted to multiple configurations of prior knowledge, number of uncertainties, type and number of tests as well as the aim of the decision-making. VoI helps to prioritise whether to conduct one test over another or whether to discard irrelevant tests. VoI also helps to justify decisions to infer information on a specific parameter over another.

Table 7.1: Description of the use of VoI for three situations that require different types of data collection

	Case 1	**Case 2**	**Case 3**
Number of sources of uncertainty X, N_X	$N_X = 1$	$N_X = 1$	$N_X > 1$
Target number of uncertain sources $Y=(Y_1,..,Y_j,..Y_{N_Y})$, N_Y	$N_Y = 1$	$N_Y = 1$	$N_Y < N_X$
Number of planned tests (T_s) of cost C_s targeting Y_j	$T_s = 1$	$T_s > 1$	$T_s(Y_j) \geq 1$ Each Y_j has a set of possible tests, T_i^j, $i \in \{1,..,T_s(Y_j)\}$
Marginalisation of non-target uncertainties	No	No	Yes. Over N_X-Y_j sources for $j \in \{1,..,N_Y\}$
Number of VoI analysis	1	T_s	$\prod_j T_s(Y_j)$
VoI analysis	$VoI_s(Y,T)$	$VoI_s(Y,T_i)$ for $i \in \{1,..,T_s\}$	$VoI_s(Y_j,T_i^j)$
Net Benefit $NB(Y,T,C_s)$	$VoI_s(Y,T) - C_s$	Select T_i associated to $\max_{T_i \in T} NB(Y,T_i,C_s(T_i))$ $= \max_{T_i \in T}\{VoI_s(Y,T_i) - C_s(T_i)\}$	Identify T^{Y_j} maximising NB for each Y_j The parameter to be prioritised is Y_j maximising: $NB(Y_j,T^{Y_j},C_s(T^{Y_j}))$

7.3 Challenges and future improvements

One of the main drawbacks of VoI analysis is the associated computational burden emanating from the Monte Carlo simulations and the computations within each iteration. As Figure 7.1 depicts, computing EVII requires a two-level Monte Carlo approach, commonly referred to as nested Monte Carlo integration, involving sampling parameters in an outer loop (e.g., observations) and sampling, for each parameter, in an inner loop (conditional parameters). The concerns regarding the stability of the results are due to errors in the estimations of the integrals, resulting from a chosen number of simulations R. Biases and errors can be estimated and an optimal number of simulations R can be determined through sensitivity analyses (Oakley *et al.,* 2010) or using quasi-Monte Carlo (Fang *et al.,* 2021). However, often thousands of samples are needed to correctly describe a distribution, and in some cases millions for complicated distributions. These simulations combined with the inner loop computations lead to very high computational demands.

Recently, there have been an increase in the number of studies aiming at reducing the computational costs resulting from VoI analyses. Most of these advances are made in the medical field where VoI methods are routinely applied to help inform decisions on healthcare and testing strategies. The currently used optimisation methods do not only target the number of Monte Carlo simulations but also aim to characterise the relationship between the input and the output. One of the approaches suggests defining a linear mathematical relationship linking the outcomes to the uncertain parameters (e.g., Chapter 4 and 5). This would allow the direct computation of integrals and skip the Monte Carlo simulations (Coyle and Oakley, 2008; Madan *et al.*, 2014). However, the relationship is often more complex and is generally nonlinear (e.g., Chapter 6). Some studies encourage defining a functional relationship between the input and the response, requiring a preliminary sensitivity analysis and the use of approximations (Strong *et al.*, 2014; Tuffaha *et al.*, 2016; Heath *et al.*, 2016).

To use the approach developed in this thesis and make optimal use of the results, it is highly recommended to improve the characterisation of potential future measurements. While the performed sensitivity analyses assumed theoretical ‘possible’ error functions relative to a ground investigation technique, there is a need

to better quantify the level of precision or biases obtained from measurements. Assigning a probability distribution to the error is an efficient way to consider the variability in measurements. Nevertheless, determining its standard deviation can benefit from further efforts to correctly estimate the level of accuracy of a technique, based on past measurements. For instance, further VoI assessments could consist of reducing the uncertainties on the V_s profile or the depth of bedrock through measuring the fundamental frequency f_0 from ambient vibration methods. Some guidelines such as those provided by the SESAME Project (SESAME, 2004) provide estimates of the threshold of variability below which horizontal to vertical ratios from ambient vibrations are valid for interpretation. This threshold can be a support in building the associated error functions. We can find other attempts in developing approaches to characterise the uncertainties and potential inherent inaccuracies associated to seismic investigation methods, their processing and interpretation (Anukwu *et al.*, 2018; Passeri *et al.*, 2019; Wotherspoon *et al.*, 2021).

A robust characterisation of the error functions of investigation techniques can lead to improvements in the VoI approach for SHA. For instance, a future interesting application would be to assess the value of combining two or more investigation techniques aiming at measuring the same parameter and to estimate the expected outcomes and NB. Doing so would permit the selection of an optimal combination that would represent a compromise between avoided losses and expected costs. In addition, a more complete approach would involve considering all sources of uncertainties within SHA, such as the hazard at the reference rock (e.g., including other percentiles, rather than only the mean), possible 2D-3D effects, the variability in the dynamic properties (e.g., through Monte Carlo simulations) and the risk of liquefaction.

Within the seismic design component of the proposed VoI application, more uncertainties could be included, e.g., those associated to estimates related to soil-structure interactions and vulnerability analyses. Finally, the expected outcomes could be based on more thorough studies (e.g., loss models) to consider both short- and long-term global consequences (Ordaz *et al.*, 2000; Bommer *et al.*, 2002; León *et al.*, 2022). While the outcomes of a decision within VoI are generally tangible estimates of the consequences, a decision can also have intangible consequences such as loss/gain in

reputation, increase/decrease of public's trust in a government or industry inducing gains/losses due to reliability concerns. These are, however, still challenging to quantify.

The incorporation of VoI into fields such as SHA and seismic design has only just started. Further improvements and considerations are needed to make the VoI approach a robust and reliable tool for decision-making under uncertainty. For now, we might ask the question: what will be the importance of VoI compared to approved guidelines and requirements, especially regarding the design of critical facilities? Despite building a VoI framework able to acknowledge strict requirements, such as the selection of a viable seismic designs that respects a defined acceptable level of safety or performing modern practice PSHA (e.g., Chapter 6), the optimal decision obtained from performing VoI still needs to satisfy approved best-practices, which are often mandatory.

In the short term, the use of VoI might be of great assistance to facilities owners, hazard analysts and insurance companies. In the long term, VoI might become a well-established tool capable of justifying sensitive decisions for governments and, perhaps more importantly, for the general public.

Bibliography

Abrahamson, N. A. and Bommer, J. J. (2005) ‘Probability and Uncertainty in Seismic Hazard Analysis’, *Earthquake Spectra*, 21(2), pp. 603–607. doi: 10.1193/1.1899158.

Abrahamson, N. A. and Silva, W. J. (1997) ‘Empirical Response Spectral Attenuation Relations for Shallow Crustal Earthquakes’, *Seismological Research Letters*, 68(1), pp. 94–127. doi: 10.1785/gssrl.68.1.94.

Abrahamson, N., Kuehn, N., Walling, M. and Landwehr, N. (2019) ‘Probabilistic Seismic Hazard Analysis in California Using Nonergodic Ground Motion Models’, *Bulletin of the Seismological Society of America*, 109(4), pp. 1235–1249. doi: 10.1785/0120190030.

Aguilar-Meléndez, A., Ordaz, M. G., Puente, J. D. la, Pujades, L., Barbat, A., Rodríguez-Lozoya, H. E., Monterrubio-Velasco, M., Escalante Martínez, J. E. and Campos-Rios, A. (2018) ‘Sensitivity Analysis of Seismic Parameters in the Probabilistic Seismic Hazard Assessment (PSHA) for Barcelona Applying the New R-CRISIS’, *Computación y Sistemas*, 22, pp. 1099–1122. doi: 10.13053/cys-22-4-3084.

Ake, J. P. (2008) ‘Uncertainties in Probabilistic Seismic Hazard Analyses for Regions of Low-to-Moderate Seismic Potential: The Need for a Structured Approach’. U.S. Nuclear Regulatory Commission.

Aldama-Bustos, G., Tromans, I. J., Strasser, F., Garrard, G., Green, G., Rivers, L., Douglas, J., Musson, R. M. W., Hunt, S., Lessi-Cheimariou, A., Daví, M. and Robertson, C. (2019) ‘A streamlined approach for the seismic hazard assessment of a new nuclear power plant in the UK’, *Bulletin of Earthquake Engineering*, 17(1), pp. 37–54. doi: 10.1007/s10518-018-0442-5.

Anderson, J. G. and Brune, J. N. (1999) ‘Probabilistic Seismic Hazard Analysis without the Ergodic Assumption’, *Seismological Research Letters*, 70(1), pp. 19–28. doi: 10.1785/gssrl.70.1.19.

Anderson, J. G. and Uchiyama, Y. (2011) ‘A Methodology to Improve Ground-Motion Prediction Equations by Including Path Corrections’, *Bulletin of the Seismological Society of America*, 101(4), pp. 1822–1846. doi: 10.1785/0120090359.

Antoine-Moussiaux, N., Vandenberg, O., Kozlakidis, Z., Aenishaenslin, C., Peyre, M., Roche, M., Bonnet, P. and Ravel, A. (2019) 'Valuing Health Surveillance as an Information System: Interdisciplinary Insights', *Frontiers in Public Health*, 7. doi: 10.3389/fpubh.2019.00138.

Anukwu, G., Khalil, A. and Abdullah, K. (2018) 'Evaluating the effectiveness of the MASW technique in a geologically complex terrain', *Journal of Physics: Conference Series*, 995, p. 012059. doi: 10.1088/1742-6596/995/1/012059.

Aristizábal, C. (2018) *Integration of Site Effects into Probabilistic Seismic Hazard Assessment. Integration of site effects into probabilistic seismic hazard methods [Doctoral Thesis]*. Universite Grenoble Alpes. Available at: https://theses.hal.science/tel-01825052.

Aristizábal, C., Bard, P.-Y. and Beauval, C. (2022) 'Site-Specific PSHA: Combined Effects of Single-Station-Sigma, Host-to-Target Adjustments and Nonlinear Behavior. A case study at Euroseistest', *Italian Journal of Geosciences*, 141(1), pp. 5–34. doi: 10.3301/IJG.2022.02.

Aristizábal, C., Bard, P.-Y., Beauval, C. and Gomez-Zapata, J. (2018) 'Integration of Site Effects into Probabilistic Seismic Hazard Assessment (PSHA): A Comparison between Two Fully Probabilistic Methods on the Euroseistest Site', *Geosciences*, 8(8), p. 285. doi: 10.3390/geosciences8080285.

ASCE (2005) 'Seismic Design Criteria for Structures, Systems and Components in Nuclear Facilities'. American Society of Civil Engineers, ASCE Standard 43-05.

Al Atik, L. and Abrahamson, N. (2021) 'A Methodology for the Development of 1D Reference VS Profiles Compatible with Ground-Motion Prediction Equations: Application to NGA-West2 GMPEs', *Bulletin of the Seismological Society of America*, 111(4), pp. 1765–1783. doi: 10.1785/0120200312.

Atkinson, G. M. (2006) 'Single-Station Sigma', *Bulletin of the Seismological Society of America*, 96(2), pp. 446–455. doi: 10.1785/0120050137.

Bahrampouri, M., Rodriguez-Marek, A. and Bommer, J. (2018) 'Mapping the uncertainty in modulus reduction and damping curves onto the uncertainty of site

amplification functions', *Soil Dynamics and Earthquake Engineering*, 126. doi: 10.1016/j.soildyn.2018.02.022.

Baio, G. (2018) 'Statistical Modeling for Health Economic Evaluations', *Annual Review of Statistics and Its Application*, 5(1), pp. 289–309. doi: 10.1146/annurev-statistics-031017-100404.

Baker, J. W. (2013) 'Probabilistic Seismic Hazard Analysis', *White Paper Version 2.0.1*.

Barani, S., Ferrari, R. and Ferretti, G. (2013) 'Influence of Soil Modelling Uncertainties on Site Response', *Earthquake Spectra*, 29, pp. 705–732. doi: 10.1193/1.4000159.

Barani, S., Ferrari, R., Ferretti, G. and Spallarossa, D. (2010) 'Calibration of soil amplification factors for real-time ground-motions scenarios in Italy', in *Fifth International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics*. San Diego.

Barani, S. and Spallarossa, D. (2017) 'Soil amplification in probabilistic ground motion hazard analysis', *Bulletin of Earthquake Engineering*, 15(6), pp. 2525–2545. doi: 10.1007/s10518-016-9971-y.

Bard, P.-Y. (2011) 'Seismic loading', *European Journal of Environmental and Civil Engineering*, 15(sup1), pp. 141–184. doi: 10.1080/19648189.2011.9695307.

Bayraktarli, Y., Baker, J. and Faber, M. (2011) 'Uncertainty treatment in earthquake modelling using Bayesian probabilistic networks', *Georisk*, 5, pp. 44–58. doi: 10.1080/17499511003679931.

Bayraktarli, Y. Y. and Faber, M. H. (2011) 'Bayesian probabilistic network approach for managing earthquake risks of cities', *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 5(1), pp. 2–24. doi: 10.1080/17499511003679907.

Bazzurro, P. and Cornell, C. (1999) 'Disaggregation of Seismic Hazard', *Bulletin of the Seismological Society of America*, 89. doi: 10.1785/BSSA0890020501.

Bazzurro, P. and Cornell, C. A. (2004) ‘Nonlinear Soil-Site Effects in Probabilistic Seismic-Hazard Analysis’, *Bulletin of The Seismological Society of America*, 94, pp. 2110–2123. doi: 10.1785/0120030216.

Benjamin, J. R. and Cornell, C. A. (1970) *Probability, Statistics, and Decisions for Civil Engineers*. New York: McGraw-Hill.

Bennett, P., Hare, A. and Townshend, J. (2005) ‘Assessing the risk of vCJD transmission via surgery: models for uncertainty and complexity’, *Journal of the Operational Research Society*, 56(2), pp. 202–213. doi: 10.1057/palgrave.jors.2601899.

Bensi, M. T., Der Kiureghian, A. and Straub, D. (2011) ‘A Bayesian Network Methodology for Infrastructure Seismic Risk Assessment and Decision Support’. Berkeley, CA: Pacific Earthquake Engineering Research Center.

Bensi, M. T., Kiureghian, Der. A. and Straub, D. (2009) ‘A Bayesian Network Framework for Post-Earthquake Infrastructure System Performance Assessment’, in *TCLEE 2009: Lifeline Earthquake Engineering in a Multihazard Environment*. Oakland (Proceedings), pp. 1–12. doi: doi:10.1061/41050(357)104.

Berge-Thierry, C., Cotton, F., Scotti, O., Griot-Pommera, D.-A. and Fukushima, Y. (2003) ‘New empirical response spectral attenuations laws for moderate European earthquakes’, *Journal of Earthquake Engineering*, 7(2), pp. 193–222. doi: 10.1080/13632460309350446.

Bhattacharjya, D., Eidsvik, J. and Mukerji, T. (2010) ‘The Value of Information in Spatial Decision Making’, *Mathematical Geosciences*, 42(2), pp. 141–163. doi: 10.1007/s11004-009-9256-y.

Bindels, J., Ramaekers, B., Ramos, I. C., Mohseninejad, L., Knies, S., Grutters, J., Postma, M., Al, M., Feenstra, T. and Joore, M. (2016) ‘Use of Value of Information in Healthcare Decision Making: Exploring Multiple Perspectives’, *PharmacoEconomics*, 34(3), pp. 315–322. doi: 10.1007/s40273-015-0346-z.

Bindi, D., Massa, M., Ameri, G., Pacor, F., Puglia, R. and Augliera, P. (2014) ‘Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 %-damped PSA at spectral periods up to 3.0 s using the

RESORCE dataset', *Bulletin of Earthquake Engineering*, 12. doi: 10.1007/s10518-013-9525-5.

Bird, J. C., Dhiman, R., Kwon, H.-M. and Varanasi, K. K. (2013) 'Reducing the contact time of a bouncing drop', *Nature*, 503, pp. 385–388. doi: https://doi.org/10.1083/nature12740.

Biswajit, D. (2017) 'Evaluation of methods used to calculate seismic fragility curves'. San Antonia, Texas: Center for Nuclear Waste Regulatory Analyses.

Boardman, A. E., Greenberg, D. H., Vining, A. R. and Weimer, D. L. (2017) *Cost-benefit analysis: concepts and practice*. Cambridge University Press.

Bojke, L., Soares, M., Claxton, K., Colson, A., Fox, A., Jackson, C., Jankovic, D., Morton, A., Sharples, L. and Taylor, A. (2021) 'Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study', *Health technology assessment*, 25(37), pp. 1–124. doi: 10.3310/hta25370.

Bommer, J. J. and Stafford, P. J. (2016) 'Seismic Hazard and Earthquake Actions', in Elghazouli, A. (ed.) *Seismic Design of Buildings to Eurocode 8*. 2nd Edition. London: CRC Press, pp. 7–40. doi: 10.1201/9781315368221.

Bommer, J. and Scherbaum, F. (2008) 'The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis', *Earthquake Spectra*, 24, pp. 997–1009. doi: 10.1193/1.2977755.

Bommer, J., Spence, R., Erdik, M., Tabuchi, S., Aydinoglu, N., Booth, E., Del Re, D. and Peterken, O. (2002) 'Development of an earthquake loss model for Turkish catastrophe insurance', *Journal of Seismology*, 6, pp. 431–446.

Boore, D. M. (2003) 'Simulation of ground motion using the stochastic method', *Pure and Applied Geophysics*, 160(3), pp. 635–676. doi: https://doi.org/10.1007/PL00012553.

Boore, D. M. and Joyner, W. B. (1997) 'Site amplifications for generic rock sites', *Bulletin of the Seismological Society of America*, 87(2), pp. 327–341.

Boore, D. M., Thompson, E. M. and Cadet, H. (2011) 'Regional Correlations of V S30 and Velocities Averaged Over Depths Less Than and Greater Than 30 Meters',

Bulletin of the Seismological Society of America, 101(6), pp. 3046–3059. doi: 10.1785/0120110071.

Borcherdt, R. D. (1970) 'Effects of local geology on ground motion near San Francisco Bay*', *Bulletin of the Seismological Society of America*, 60(1), pp. 29–61. doi: 10.1785/BSSA0600010029.

Bozorgnia, Y., Abrahamson, N. A., Al Atik, L., Ancheta, T. D., Atkinson, G. M., Baker, J. W., Baltay, A., Boore, D. M., Campbell, K. W. and Chiou, B. S.-J. (2014) 'NGA-West2 research project', *Earthquake Spectra*, 30(3), pp. 973–987.

Brathwaite, J. and Saleh, J. H. (2013) 'Bayesian framework for assessing the value of scientific space systems: Value of information approach with application to earth science spacecraft', *Acta Astronautica*, 84, pp. 24–35. doi: 10.1016/j.actaastro.2012.10.036.

Bredimas, A. and Nuttall, W. J. (2008) 'An international comparison of regulatory organizations and licensing procedures for new nuclear power plants', *Energy Policy*, 36(4), pp. 1344–1354. doi: https://doi.org/10.1016/j.enpol.2007.10.035.

Brennan, A., Kharroubi, S., O'Hagan, A. and Chilcott, J. (2007) 'Calculating partial expected value of perfect information via Monte Carlo sampling algorithms', *Medical Decisions*, 27, pp. 448–470.

Briggs, A., Sculpher, M. and Claxton, K. (2006) *Decision Modelling for Health Economic Evaluation*. Oxford, UK: Oxford Univ. Press.

Budnitz, R., Apostolakis, G., Boore, D., Cluff, L., Coppersmith, K., Cornell, C. and Morris, P. (1997) *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts: Main Report*, *NUREG/CR-6372*. Washington: U.S. Nuclear Regulatory Commission.

Campbell, H. F. and Brown, R. P. C. (2003) *Benefit-Cost Analysis: Financial and Economic Appraisal Using Spreadsheets*. Cambridge University Press. Available at: https://books.google.co.uk/books?id=b3_8Tapt2MYC.

Campbell, K. (2003) 'Prediction of Strong Ground Motion Using the Hybrid Empirical Method and Its Use in the Development of Ground-Motion (Attenuation) Relations in

Eastern North America', *Bulletin of the Seismological Society of America*, 93, pp. 1012–1033. doi: 10.1785/0120020002.

Cantero-Chinchilla, S., Chiachío, J., Chiachío, M., Chronopoulos, D. and Jones, A. (2020) 'Optimal sensor configuration for ultrasonic guided-wave inspection based on value of information', *Mechanical Systems and Signal Processing*, 135, p. 106377. doi: https://doi.org/10.1016/j.ymssp.2019.106377.

Carlin, B. P. and Louis, T. A. (1996) *Bayes and empirical Bayes methods for data analysis*. London: Chapman and Hall.

Carpenter, N. S., Wang, Z., Woolery, E. W. and Rong, M. (2018) 'Estimating Site Response with Recordings from Deep Boreholes and HVSR: Examples from the Mississippi Embayment of the Central United States', *Bulletin of the Seismological Society of America*, 108(3A), pp. 1199–1209. doi: 10.1785/0120170156.

Casella, G. (1985) *An Introduction to Empirical Bayes Data Analysis, Source: The American Statistician*.

CBC (2019) *2019 California Building Code, Title 24, International Code Council*. Sacramento, CA.

CEN (2004) 'Eurocode 2: Design of concrete structures—Part 1-1: General rules and rules for buildings', *London: British Standard Institution*, a.

Chartier, T., Scotti, O., Lyon-Caen, H., Richard-Dinger, K., Dieterich, J. H. and Shaw, B. E. (2021) 'Modelling earthquake rates and associated uncertainties in the Marmara Region, Turkey', *Natural Hazards and Earth System Sciences*, 21(8), pp. 2733–2751. doi: 10.5194/nhess-21-2733-2021.

Chen, C. J. P., Chen, S. and Su, X. (1999) 'Is Accounting Information Value Relevant in the Emerging Chinese Stock Market?', *SSRN Electronic Journal*. doi: 10.2139/ssrn.167353.

Ching, J. and Phoon, K. K. (2012) 'Value of geotechnical site investigation in reliability-based design', in *Advances in Structural Engineering*, pp. 1935–1945. doi: 10.1260/1369-4332.15.11.1935.

Choy, S. L., O'Leary, R. and Mengersen, K. (2009) 'Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models', *Ecology*, 90(1), pp. 265–277. doi: 10.1890/07-1886.1.

Clinch, J. P. (2004) 'Cost–Benefit Analysis Applied to Energy', in Cleveland, C. J. (ed.) *Encyclopedia of Energy*. New York: Elsevier, pp. 715–725. doi: https://doi.org/10.1016/B0-12-176480-X/00237-0.

Clotaire, M., Duvernay, B., Kölz, E., Jamali, N. and Lestuzzi, P. (2019) 'A framework to evaluate the benefit of seismic upgrading', *Earthquake Spectra*, 35(2), pp. 1045–1051. doi: 10.1193/061318EQS145D.

Cockburn, G. and Tesfamariam, S. (2012) 'Earthquake disaster risk index for Canadian cities using Bayesian belief networks', *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 6(2), pp. 128–140. doi: 10.1080/17499518.2011.650147.

Copeland, T. E. and Friedman, D. (1992) 'The Market Value of Information: Some Experimental Results', *The Journal of Business*, 65(2), pp. 241–266. Available at: http://www.jstor.org/stable/2353164.

Cornell, C. A. (1968) 'Engineering seismic risk analysis. Bulletin of the seismological society of America', *Bulletin of the seismological society of America*, 58(5), pp. 1583–1606.

Cotton, F., Scherbaum, F., Bommer, J. J. and Bungum, H. (2006) 'Criteria for Selecting and Adjusting Ground-Motion Models for Specific Target Regions: Application to Central Europe and Rock Sites', *Journal of Seismology*, 10(2), pp. 137–156. doi: 10.1007/s10950-005-9006-7.

Coupé, V. M. H. and van der Gaag, L. C. (2002) 'Properties of Sensitivity Analysis of Bayesian Belief Networks', *Annals of Mathematics and Artificial Intelligence*, 36(4), pp. 323–356. doi: 10.1023/A:1016398407857.

Cover, L. E., Bohn, M. P., Campbell, R. D. and Wesley, D. A. (1983) 'Handbook of nuclear power plant seismic fragilities, seismic safety margins research program. NUREG/CR-3558'. Livermore, CA: Lawrence Livermore National Laboratory. doi: https://doi.org/10.2172/5313138.

Coyle, D. and Oakley, J. (2008) ‘Estimating the expected value of partial perfect information: a review of methods’, *The European Journal of Health Economics*, 9(3), pp. 251–259. doi: 10.1007/s10198-007-0069-y.

Cramer, C. H. (2003) ‘Site-Specific Seismic-Hazard Analysis that is Completely Probabilistic’, *Bulletin of the Seismological Society of America*, 93(4), pp. 1841–1846.

Cultrera, G., Cornou, C., Giulio, G. Di and Bard, P.-Y. (2021) ‘Indicators for site characterization at seismic station: recommendation from a dedicated survey’, *Bulletin of Earthquake Engineering*, 19, pp. 4171–4195. doi: 10.1007/s10518-021-01136-7.

Danciu, L., Nandan, S., Reyes, C., Basili, R., Weatherill, G., Beauval, C., Rovida, A., Vilanova, S., Sesetyan, K., Bard, P.-Y., Cotton, F., Wiemer, S. and Giardini, D. (2021) *The 2020 update of the European Seismic Hazard Model: Model Overview. EFEHR Technical Report 001, v1.0.0*.

Darendeli, M. (2001) *Development of New Family of Normalized Modulus Reduction and Material Damping Curves [Doctoral Thesis]*. The University of Texas.

Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., Drouet, S., Douglas, J., Basili, R., Sandikkaya, M. A., Faccioli, E., Theodoulidis, N., Delavaud, E and Segou, M. (2012) ‘Toward a Ground-Motion Logic Tree for Probabilistic Seismic Hazard Assessment in Europe’, *Journal of Seismology*, 16, pp. 451–473. doi: 10.1007/s10950-012-9281-z.

Derras, B., Bard, P.-Y. and Cotton, F. (2017) ‘VS30, slope, H800 and f0: performance of various site-condition proxies in reducing ground-motion aleatory variability and predicting nonlinear site response’, *Earth, Planets and Space*, 69(1), p. 133. doi: 10.1186/s40623-017-0718-z.

Douglas, J., Ulrich, T. & Negulescu, C. (2013) ‘Risk-targeted seismic design maps for mainland France’. Natural Hazards, 65(3), 1999–2013 (2013). https://doi.org/10.1007/s11069-012-0460-6

Douglas, J. (2017) *Graph of ground motion predictions versus publication date. figshare. Figure.*, *www.gmpe.org.uk*.

Douglas, J. and Edwards, B. (2016) 'Recent and future developments in earthquake ground motion estimation', *Earth-Science Reviews*, 160, pp. 203–219. doi: 10.1016/j.earscirev.2016.07.005.

Ducoffe, R. H. (1995) 'How Consumers Assess the Value of Advertising', *Journal of Current Issues & Research in Advertising*, 17(1), pp. 1–18. doi: 10.1080/10641734.1995.10505022.

Ducoffe, R. H. and Curlo, E. (2000) 'Advertising value and advertising processing', *Journal of Marketing Communications*, 6(4), pp. 247–262. doi: 10.1080/135272600750036364.

EC8 (2004) 'Eurocode 8: Design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings', *European Standard EN 1998-1*. Brussels: European Committee for Standardization.

Eckermann, S. and Willan, A. (2007) 'Expected value of information and decision making in HTA', *Health economics*, 16, pp. 195–209. doi: 10.1002/hec.1161.

Edwards, B., Michel, C., Poggi, V. and Fäh, D. (2013) 'Determination of Site Amplification from Regional Seismicity: Application to the Swiss National Seismic Networks', *Seismological Research Letters*, 84, pp. 611–621. doi: 10.1785/0220120176.

Efron, B. (2010) 'Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction', *Large-Scale Inference Empirical Bayes Methods for Estimation, Testing, and Prediction*. doi: 10.1017/CBO9780511761362.

EFSA (2014) 'Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment', *European Food Safety Authority Journal*, 12(6), p. 3734. doi: https://doi.org/10.2903/j.efsa.2014.3734.

Eidsvik, J., Bhattacharjya, D. and Mukerji, T. (2008) 'Value of information of seismic amplitude and CSEM resistivity', *Geophysics*, 73(4), pp. R59–R69. doi: 10.1190/1.2938084.

Eidsvik, J., Mukerji, T., Bhattacharjya, D. and Dutta, G. (2015) 'Value of Information Analysis of Geophysical Data for Drilling Decisions', in. European Association of Geoscientists & Engineers, p. cp-456-00038. doi: 10.3997/2214-4609.201413621.

Eidsvik, J., Mukerji, Tapan. and Bhattacharjya, Debarun. (2015) *Value of Information in the Earth Sciences: Integrating Spatial Modelling and Decision Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139628785.

Ellingwood, B. (2001) 'Earthquake risk assessment of building structures', *Reliability Engineering & System Safety*, 74, pp. 251–262. doi: 10.1016/S0951-8320(01)00105-3.

EPA (2009) *Expert Elicitation Task Force White Paper*. Washington, DC.

EPRI (1993) 'Guidelines for Determining Design Ground Motions. EPRI TR-102293'. Electrical Power Research Institute.

EPRI (2013) 'Seismic evaluation guidance: screening, prioritization and implementation details (SPID) for the resolution of Fukushima near-term task force recommendation. 2.1: Seismic'. Palo Alto, California: Electric Power Research Institute.

Erdik, M., Demircioglu, M., Sesetyan, K., Durukal, E. and Siyahi, B. (2004) 'Earthquake hazard in Marmara region, Turkey', *Soil Dynamics and Earthquake Engineering*, 24, pp. 605–631.

Faizian, M., Schalcher, H.-R. and Faber, M. (2004) 'Consequence Assessment in Earthquake Risk Management Using Damage Indicators', in *First International Forum on Engineering Decision Making (IFED)*. Stoos, Switzerland.

Fang, W., Wang, Z., Giles, M. B., Jackson, C. H., Welton, N. J., Andrieu, C. and Thom, H. (2021) 'Multilevel and Quasi Monte Carlo Methods for the Calculation of the Expected Value of Partial Perfect Information', *Medical Decision Making*, 42(2), pp. 168–181. doi: 10.1177/0272989X211026305.

FEMA 356 (2000) 'Prestandard and Commentary for the seismic Rehabilitation of Buildings. Federal Emergency Agency. Washington, DC, USA

FEMA (1992) ‘A benefit-cost model for the seismic rehabilitation of buildings. FEMA-227’. Sacramento, California: VSP associated.

Ferris, E. and Solis, M. (2013) *Earthquake, Tsunami, Meltdown – The Triple Disaster’s Impact on Japan, Impact on the World*, *Brookings*. Available at: www.brookings.edu/blog/up-front/2013/03/11/earthquake-tsunami-meltdown-the-triple-disasters-impact-on-japan-impact-on-the-world/amp/.

Friis-Hansen, A. (2000) *Bayesian Networks as a Decision Support Tool in Marine Applications: Ph. D. Thesis*. Department of Naval Architecture and Offshore Engineering, Technical University of Denmark.

Galanis, P., Sycheva, A., Mimra, W. and Stojadinović, B. (2018) ‘A framework to evaluate the benefit of seismic upgrading’, *Earthquake Spectra*, 34(2), pp. 527–548. doi: 10.1193/120316EQS221M.

Gelis, C., Cauchie, L., Cushing, E., Froment, B., Franco, S., Jomard, H., Moiriat, D., Provost, L., Sariguzel, B. and Tebib, H. (2022) ‘Estimation of the Local Seismic Amplification on an Industrialized Site in the French Rhône Valley’, *Pure and Applied Geophysics*, 179, pp. 1–27. doi: 10.1007/s00024-022-03069-x.

Gelman, A., Carlin, J. B., Stern, H. S. and B., D. (1995) *Bayesian Data Analysis*. 1st Edition. Edited by Chapman and Hall/CRC. doi: https://doi.org/10.1201/9780429258411.

Giardini, D. et al. (2013) ‘Seismic Hazard Harmonization in Europe (SHARE): Online Data Resource’. doi: 10.12686/SED-00000001-SHARE.

Gilbert, R. B. and Habibi, M. (2015) ‘Assessing the value of information to design site investigation and construction quality assurance programs’, in Phoon, K.-K. and Ching, J. (eds) *Risk and Reliability in Geotechnical Engineering*. 1st Edition. Boca Raton: CRC Press Taylor and Francis Group, pp. 491–532.

Giordano, P. F., Iacovino, C., Quqa, S. and Limongelli, M. P. (2022) ‘The value of seismic structural health monitoring for post-earthquake building evacuation’, *Bulletin of Earthquake Engineering*. doi: 10.1007/s10518-022-01375-2.

Di Giulio, G., Cultrera, G., Cornou, C., Bard, P.-Y. and Bilal, A. T. (2021) 'Quality assessment for site characterization at seismic stations metadata · Seismic network database · Strong-motion database', 19, pp. 4643–4691. doi: 10.1007/s10518-021-01137-6.

Gkimprixis, A., Tubaldi, E. and Douglas, J. (2019) 'Comparison of methods to develop risk-targeted seismic design maps', *Bulletin of Earthquake Engineering*, 51(43), pp. 3727–3752. doi: 10.1007/s10518-019-00629-w.

Gkimprixis, A., Tubaldi, E. and Douglas, J. (2020) 'Evaluating alternative approaches for the seismic design of structures', *Bulletin of Earthquake Engineering*, 18(9), pp. 4331–4361. doi: 10.1007/s10518-020-00858-4.

Griffin, R. C. (1998) 'The fundamental principles of cost-benefit analysis', *Water Resources Research*, 34(8), pp. 2063–2071.

Grigore, B., Peters, J., Hyde, C. and Stein, K. (2013) 'Methods to Elicit Probability Distributions from Experts: A Systematic Review of Reported Practice in Health Technology Assessment', *PharmacoEconomics*, 31(11), pp. 991–1003. doi: 10.1007/s40273-013-0092-z.

Gutenberg, B. and Richter, C. F. (1944) 'Frequency of Earthquakes in California', *Bulletin of the Seismological Society of America*, 34, pp. 185–188.

Hammond, R. J. (1960) 'Benefit-Cost analysis and water pollution control', *American Journal of Agricultural Economics*, 42(4).

Hanley, N., Barbier, E. B. and Barbier, E. (2009) *Pricing nature: cost-benefit analysis and environmental policy*. Edward Elgar Publishing.

Heath, A., Manolopoulou, I. and Baio, G. (2016) 'Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation', *Statistics in Medicine*, 35(23), pp. 4264–4280. doi: 10.1002/sim.6983.

Heckerman, D. (1997) *Bayesian Networks for Data Mining, Data Mining and Knowledge Discovery*. Kluwer Academic Publishers.

Heckerman, D. (1999) ‘A tutorial in learning with Bayesian networks’, in Jordan, M. I. (ed.) *Learning in Graphical Models*. The Netherlands: Kluwer Academic Publishers, pp. 301–354.

Heckerman, D. and Geiger, D. (1995) ‘Learning Bayesian networks’, in Besnard, P. and Hanks, S. (eds) *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco (CA), pp. 274–284.

Van Houtte, C., Drouet, S. and Cotton, F. (2011) ‘Analysis of the Origins of κ (Kappa) to Compute Hard Rock to Rock Adjustment Factors for GMPEs’, *Bulletin of the Seismological Society of America*, 101(6), pp. 2926–2941. doi: 10.1785/0120100345.

Howard, R. and Matheson, J. (1984) ‘Influence diagrams’, in Howard, R. and Matheson, J. (eds) *The Principles and Applications of Decision Analysis*. 2nd edn. Menlo Park, CA: Strategic Decisions Group, pp. 720–762.

HSE (2009) ‘Technical Assessment Guide: External Hazards. T/AST/013, Issue 3’. UK: Health and Safety Executive-Nuclear Directorate.

Hubbard, D. W. (2007) *How to measure anything: Finding the Value of ‘intangibles’ in Business*. John Wiley & Sons.

Hubbard, D. W. (2009) *The failure of risk management*. Canada: John Wiley and Sons.

IAEA (2017) ‘Managing the Financial Risk Associated with the Financing of New Nuclear Power Plant Projects. No. NG-T-4.6’. Vienna: International Atomic Energy Agency. Available at: http://www.iaea.org/Publications/index.html.

IAEA (2019) ‘Approaches to Safety Evaluation of New and Existing Research Reactor Facilities in Relation to External Events, Safety Reports Series No. 94’. Vienna: International Atomic Energy Agency.

IAEA (2021) ‘Seismic Design for Nuclear Installations, IAEA Safety Standards Series No. SSG-67’. Vienna: International Atomic Energy Agency.

IAEA (2022) ‘Seismic Hazards in Site Evaluation for Nuclear Installations’, *IAEA Safety Standards Series No. SSG-9 (Rev. 1)*. Vienna: International Atomic Energy Agency.

Iannacone, L., Giordano, P. F., Gardoni, P. and Limongelli, M. P. (2022) 'Quantifying the value of information from inspecting and monitoring engineering systems subject to gradual and shock deterioration', *Structural Health Monitoring*, 21(1), pp. 72–89. doi: 10.1177/1475921720981869.

IBC (2012) *2012 international building code*. International Code Council, Country Club Hills, Ill. Available at: https://search.library.wisc.edu/catalog/9910971558202121.

Iglesias, C. P., Thompson, A., Rogowski, W. H. and Payne, K. (2016) 'Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations', *PharmacoEconomics*, 34(11), pp. 1161–1172. doi: 10.1007/s40273-016-0425-9.

IRSN (2011) 'Synthèse actualisée des connaissances relatives à l'impact sur le milieu marin des rejets radioactifs du site nucléaire accidenté de Fukushima Dai-ichi '. Institut de Radioprotection et de Surete Nucleaire.

Iwata, T. and Irikura, K. (1988) 'Source parameters of the 1983 Japan Sea Earthquake sequence', *Journal of Physics of the Earth*, 36, pp. 155–184.

Jackson, C. H., Baio, G., Heath, A., Strong, M., Welton, N. J. and Wilson, E. C. F. (2022) 'Value of Information Analysis in Models to Inform Health Policy', *Annual Review of Statistics and Its Application*, 9(1), pp. 95–118. doi: 10.1146/annurev-statistics-040120-010730.

Jalal, H., Goldhaber-Fiebert, J. D. and Kuntz, K. M. (2015) 'Computing Expected Value of Partial Sample Information from Probabilistic Sensitivity Analysis Using Linear Regression Metamodeling', *Medical Decision Making*, 35(5), pp. 584–595. doi: 10.1177/0272989X15578125.

Jiang, W. and Marggraf, R. (2021) 'The origin of cost–benefit analysis: a comparative view of France and the United States', *Cost Effectiveness and Resource Allocation*, 19(1), p. 74. doi: 10.1186/s12962-021-00330-3.

Jordaan, I. (2005) *Decisions under Uncertainty: Probabilistic Analysis for Engineering Decisions*. Cambridge: Cambridge University Press. doi: DOI: 10.1017/CBO9780511804861.

Kamariotis, A., Chatzi, E. and Straub, D. (2023) ‘A framework for quantifying the value of vibration-based structural health monitoring’, *Mechanical Systems and Signal Processing*, 184, p. 109708. doi: 10.1016/J.YMSSP.2022.109708.

Kappos, Andreas J, Dimitrakopoulos, A. E. G., Kappos, A J and Dimitrakopoulos, E. G. (2008) ‘Feasibility of pre-earthquake strengthening of buildings based on cost-benefit and life-cycle cost analysis, with the aid of fragility curves’, 45, pp. 33–54. doi: 10.1007/s11069-007-9155-9.

Katona, T. J. (2017) ‘Issues of the Seismic Safety of Nuclear Power Plants’, in Zouaghi, T. (ed.) *Earthquakes*. Rijeka: IntechOpen, p. Ch. 11. doi: 10.5772/65853.

Keisler, J. M., Collier, Z. A., Chu, E., Sinatra, N. and Linkov, I. (2014) ‘Value of information analysis: The state of application’, *Environment Systems and Decisions*, pp. 3–23. doi: 10.1007/s10669-013-9439-4.

Kennedy, R. P. (1999) ‘Overview of methods for seismic PRA and margin analysis including recent innovations’, in *Proceedings of the OECD-NEA Workshop on Seismic Risk*. Tokyo, Japan.

Kennedy, R. P. (2011) ‘Performance-goal based (risk informed) approach for establishing the SSE site specific response spectrum for future nuclear power plants’, in *Nuclear Engineering and Design*, pp. 648–656. doi: 10.1016/j.nucengdes.2010.08.001.

Ketchum, M., Chang, V. and Shantz, T. (2004) ‘Influence of Design Ground Motion Level On Highway Bridge Costs. Report 6D01’. Berkeley, California: Pacific Earthquake Engineering Research Center.

Kleiter D. G. (1996), ‘Propagating imprecise probabilities in Bayesian networks’, Artificial Intelligence, Volume 88, Issues 1–2, pp. 143-161, ISSN 0004-3702, https://doi.org/10.1016/S0004-3702(96)00021-5

Knezevic, M., Cvetkovska, M., Hanak, T., Braganca, L., & Soltesz, A. (2018). ‘Artificial Neural Networks and Fuzzy Neural Networks for Solving Civil Engineering Problems’. *Complexity, 2018*

Koopmans, C. and Mouter, N. (2020) ‘Chapter One - Cost-benefit analysis’, in Mouter, N. (ed.) *Advances in Transport Policy and Planning*. Academic Press, pp. 1–42. doi: https://doi.org/10.1016/bs.atpp.2020.07.005.

Kotha, S. R., Weatherill, G., Bindi, D. and Cotton, F. (2020) ‘A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe’, *Bulletin of Earthquake Engineering*. doi: 10.1007/s10518-020-00869-1.

Kotra, J. P., Lee, M. P., Eisenberg, N. A. and DeWispelare, A. R. (1996) *Branch technical position on the use of expert elicitation in the high-level radioactive waste program*. Oak Ridge, TN. doi: 10.2172/414310.

Kottke, A. R. and Rathje, E. M. (2013) ‘Comparison of Time Series and Random-Vibration Theory Site-Response Methods’, *Bulletin of the Seismological Society of America*, 103(3), pp. 2111–2127. doi: 10.1785/0120120254.

Kottke, A., Wang, X. and Rathje, E. (2013) ‘Technical manual for strata, geotechnical engineering center’. Austin: Department of Civil, Architectural, and Environmental Engineering: University of Texas.

Kramer, S. L. (1996) *Geotechnical Earthquake Engineering*. Prentice-Hall, Inc. New Jersey.

Ktenidou, O.-J., Abrahamson, N. A., Drouet, S. and Cotton, F. (2015) ‘Understanding the physics of kappa (κ): insights from a downhole array’, *Geophysical Journal International*, 203(1), pp. 678–691. doi: 10.1093/gji/ggv315.

Lacave, C., Koller, M., Lestuzzi, P. and Salameh, C. (2014) ‘The Significance of Site Effect Studies for Seismic Design and Assessment of Industrial Facilities’, in, pp. 475–483. doi: 10.1007/978-3-658-02810-7_40.

Lanzano, G., Pacor, F., Luzi, L., D’Amico, M., Puglia, R. and Felicetta, C. (2017) ‘Systematic source, path and site effects on ground motion variability: the case study of Northern Italy’, *Bulletin of Earthquake Engineering*, 15, pp. 4563–4583. doi: 10.1007/s10518-017-0170-2.

Leal, J., Wordsworth, S., Legood, R. and Blair, E. (2007) 'Eliciting Expert Opinion for Economic Models: An Applied Example', *Value in Health*, 10(3), pp. 195–203. doi: 10.1111/j.1524-4733.2007.00169.x.

León, J. A., Ordaz, M., Haddad, E. and Araújo, I. F. (2022) 'Risk caused by the propagation of earthquake losses through the economy', *Nature Communications*, 13(1), p. 2908. doi: 10.1038/s41467-022-30504-3.

Lev, M. (2008) 'Using Bayesian Belief Networks for credit card fraud detection', in *Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications*. Innsbruck, Austria.

Levitt, S. and Syverson, C. (2008) 'Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions', *The Review of Economics and Statistics*, 90(4), pp. 599–611.

Li, L., Wang, J., Leung, H. and Jiang, C. (2010) 'Assessment of Catastrophic Risk Using Bayesian Network Constructed from Domain Knowledge and Spatial Data', *Risk Analysis*, 30(7), pp. 1157–1175. doi: https://doi.org/10.1111/j.1539-6924.2010.01429.x.

Li, W. and Assimaki, D. (2010) 'Site and Motion-Dependent Parametric Uncertainty of Site-Response Analyses in Earthquake Simulations', *Bulletin of the Seismological Society of America*, 100(3), pp. 954–968. doi: 10.1785/0120090030.

Long, M. and Donohue, S. (2010) 'Characterization of Norwegian marine clays with combined shear wave velocity and piezocone cone penetration test (CPTU) data', *Canadian Geotechnical Journal*, 47(7), pp. 709–718. doi: 10.1139/T09-133.

Lucas, P., Van der Gaag, L. and Abu-Hanna, A. (2004) 'Bayesian networks in biomedicine and health-care', *Aritificial Intelligence in Medicine*, 3(30), pp. 201–214.

Luzi, L., Puglia, R., Russo, E. and ORFEUS WG5 (2015) 'Engineering Strong Motion Database (ESM), version 1.0', *Istituto Nazionale di Geofisica e Vulcanologia (INGV)*. Observatories & Research Facilities for European Seismology. doi: https://doi.org/10.13127/ESM.

Macauley, M. K. (2006) 'The value of information: Measuring the contribution of space-derived earth science data to resource management', *Space Policy*, 22(4), pp. 274–282. doi: 10.1016/j.spacepol.2006.08.003.

Machado, K. (1999) *Proving Clinical Efficacy to Create a Cost-benefit Analysis to Help Assess the Most Appropriate Method of Marketing Prealbumin Testing*. California State University, Stanislaus. Available at: https://books.google.co.uk/books?id=a-gTOAAACAAJ.

Madan, J., Ades, A. E., Price, M., Maitland, K., Jemutai, J., Revill, P. and Welton, N. J. (2014) 'Strategies for Efficient Computation of the Expected Value of Partial Perfect Information', *Medical Decision Making*, 34(3), pp. 327–342. doi: 10.1177/0272989X13514774.

Makra, K., Chávez-García, F., Raptakis, D. and Pitilakis, K. (2005) 'Parametric analysis of the seismic response of a 2D sedimentary valley: Implications for code implementations of complex site effects', *Soil Dynamics and Earthquake Engineering*, 25, pp. 303–315. doi: 10.1016/j.soildyn.2005.02.003.

Martinelli, G., Eidsvik, J., Sinding-Larsen, R., Rekstad, S. and Mukerji, T. (2013) 'Building Bayesian networks from basin-modelling scenarios for improved geological decision making', *Petroleum Geoscience*, 19(3), pp. 289–304. doi: 10.1144/petgeo2012-057.

Matthew, C. (2022) 'Nuclear energy in the UK. POSTnote 687'. UK Parliament POST.

McBride, M. F., Fidler, F. and Burgman, M. A. (2012) 'Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research', *Diversity and Distributions*, 18(8), pp. 782–794. doi: 10.1111/j.1472-4642.2012.00884.x.

McCurry, J. (2013) 'Fukushima two years on: the largest nuclear decommissioning finally begins', *The Guardian*, 6 March.

McFall, R. M. and Treat, T. A. (1999) 'Quantifying the Information Value of Clinical Assessments with Signal Detection Theory', *Annual Review of Psychology*, 50(1), pp. 215–241. doi: 10.1146/annurev.psych.50.1.215.

McGuire, R. K. (2004) *Seismic Hazard and Risk Analysis*. Earthquake Engineering Research Institute.

McGuire, R. K. and Shedlock, K. M. (1981) 'Statistical uncertainties in seismic hazard evaluations in the United States', *Bulletin of the Seismological Society of America*, 71(4), pp. 1287–1308. doi: 10.1785/BSSA0710041287.

McGuire, R. K., Silva, W. J. and Costantino, C. J. (2001) 'Technical Basis for Revision of Regulatory Guidance on Design Ground Motions: Hazard- and Risk-consistent Ground Motion Spectra Guidelines (NUREG/CR-6728)'. Washington: US Nuclear Regulatory Commission.

McGuire, Robin K, Silva, W. J. and Kenneally, R. (2001) 'New seismic design spectra for nuclear power plants', *Nuclear Engineering and Design*, 203(2), pp. 249–257. doi: https://doi.org/10.1016/S0029-5493(00)00345-9.

Meyer, M. A. and Booker, J. M. (2001) *Eliciting and Analyzing Expert Judgment*. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9780898718485.

Mital, U., Ahdi, S., Herrick, J., Iwahashi, J., Savvaidis, A. and Yong, A. (2021) 'A Probabilistic Framework to Model Distributions of VS30', *Bulletin of the Seismological Society of America*, 111(4), pp. 1677–1692. doi: 10.1785/0120200281.

Mittal, H., Kumar, A. and Kumar, A. (2013) 'Site Effects Estimation in Delhi from the Indian Strong Motion Instrumentation Network', *Seismological Research Letters*, 84(1), pp. 33–41. doi: 10.1785/0220120058.Mori, F., Mendicelli, A., Moscatelli, M., Romagnoli, G., Peronace, E. and Naso, G. (2020) 'A new Vs30 map for Italy based on the seismic microzonation dataset', *Engineering Geology*, 275, p. 105745. doi: https://doi.org/10.1016/j.enggeo.2020.105745.

Musson, R. M. W. (2014) 'UK seismic hazard assessments for strategic facilities: a short history', *Bollettino di Geofisica Teorica e Applicata*, 55(1), pp. 165–173.

Nakamura, Y. (1989) 'A method for dynamic characteristics estimation of subsurface using microtremor on ground surface'. Railway Technical Research Institute, Quarterly Reports, 30(1).

Neapolitan, R. E. (1989) *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. John Wiley. New York: A Wiley-Interscience Publication.

NEHRP (2015) 'Recommended provisions for seismic regulations for new buildings and other structures, FEMA 1050'. Washington, DC.: Building Seismic Safety Council, National Institute of Building Sciences.

Nemet, G. F., Anadon, L. D. and Verdolini, E. (2017) 'Quantifying the Effects of Expert Selection and Elicitation Design on Experts' Confidence in Their Judgments About Future Energy Technologies', *Risk Analysis*, 37(2), pp. 315–330. doi: 10.1111/risa.12604.

NTC (2018) 'NTC 18 - Aggiornamento delle Norme Tecniche per le Costruzioni'. Ministero delle Infrastrutture e dei Trasporti. Available at: https://www.gazzettaufficiale.it/eli/gu/2018/02/20/42/so/8/sg/pdf.

Oakley, J. E., Brennan, A., Tappenden, P. and Chilcott, J. (2010) 'Simulation sample sizes for Monte Carlo partial EVPI calculations', *Journal of Health Economics*, 29(3), pp. 468–477. doi: 10.1016/j.jhealeco.2010.03.006.

O'Hagan, A. and Stevens, J. W. (2001) 'Bayesian Assessment of Sample Size for Clinical Trials of Cost-Effectiveness', *Medical Decision Making*, 21(3), pp. 219–230. doi: 10.1177/0272989X0102100307.

ONR (2017) 'Analysis of Seismic Hazards for Nuclear Sites', *NS-TAST-GD-013 Annex 1 Reference Paper, Expert Panel on Natural Hazards report: GEN-SH-EP-2016-1*. Office of Nuclear Regulation.

Ordaz, M. and Arroyo, D. (2016) 'On Uncertainties in Probabilistic Seismic Hazard Analysis', *Earthquake Spectra*, 32(3), pp. 1405–1418. doi: 10.1193/052015EQS075M.

Ordaz, M., Miranda, E., Reinoso, E. and Pérez-Rocha, L. (2000) 'Seismic Loss Estimation Model for Mexico City', in *12th World Conference on Earthquake Engineering*. Aukland, New Zealand.

Park, J.-H., Dong-Hyun, S. and Jeon, S.-H. (2022) 'Seismic Fragility and Risk Assessment of a Nuclear Power Plant Containment Building for Seismic Input Based

on the Conditional Spectrum', *Applied Sciences*, 12(10), p. 5176. doi: https://doi.org/10.3390/app12105176.

Passeri, F., Foti, S., Cox, B. R. and Rodriguez-Marek, A. (2019) 'Influence of Epistemic Uncertainty in Shear Wave Velocity on Seismic Ground Response Analyses', *Earthquake Spectra*, 35(2), pp. 929–954. doi: 10.1193/011018EQS005M.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Edited by J. Pearl. San Francisco (CA): Morgan Kaufmann Publishers, Inc. doi: https://doi.org/10.1016/B978-0-08-051489-5.50008-4.

Pilz, M., Parolai, S., Picozzi, M. and Zschau, J. (2011) 'Evaluation of proxies for seismic site conditions in large urban areas: The example of Santiago de Chile', *Physics and Chemistry of The Earth - PHYS CHEM EARTH*, 36. doi: 10.1016/j.pce.2011.01.007.

PNNL (2014) 'Handford Sitewide Probabilistic Seismic Hazard Analysis. PNNL-23361'. Washington: Pacific Northwest National Laboratory.

Van de Poel, Ibo. (2009) 'Values in Engineering Design', in Meijers, A. (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: North-Holland, pp. 973–1006. doi: https://doi.org/10.1016/B978-0-444-51667-1.50040-9.

Porter, K., Scawthorn, C. and Beck, J. (2006) 'Cost-Effectiveness of Stronger Woodframe Buildings', *Earthquake Spectra*, 22. doi: 10.1193/1.2162567.

Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T. and Roblee, C. (2008) 'An overview of the NGA project', *Earthquake spectra*, 24(1), pp. 3–21.

Raiffa, H. (1997) *Decision Analysis: Introductory readings of choices under uncertainties*. McGraw-Hill.

Raiffa, H. and Schlaifer, R. (1961) *Applied Statistical Decision Theory*. Cambridge MA: M.I.T Press.

Raptakis, D., Theodulidis and Pitilakis, K. (1998) 'Data Analysis of the Euroseistest Strong Motion Array in Volvi (Greece): Standard and Horizontal-to-Vertical Spectral Ratio Techniques', *Earthquake Spectra*, 14. doi: 10.1193/1.1585996.

Rathje, E., Kottke, A. and Trent, W. (2010) ‘Influence of Input Motion and Site Property Variabilities on Seismic Site Response Analysis’, *Journal of Geotechnical and Geoenvironmental Engineering* , 136. doi: 10.1061/(ASCE)GT.1943-5606.0000255.

Rathje, E., Pehlivan, M., Gilbert, R. and Rodriguez-Marek, A. (2015) ‘Incorporating Site Response into Seismic Hazard Assessments for Critical Facilities: A Probabilistic Approach’, *Geotechnical, Geological and Earthquake Engineering*, 37, pp. 93–111. doi: 10.1007/978-3-319-10786-8_4.

Reiter, L. (1990) *Earthquake Hazard Analysis: Issues and Insights*. Columbia University Press. Available at: https://books.google.co.uk/books?id=Apg_nwEACAAJ.

Robbins, H. (1956) ‘An Empirical Bayes Approach to Statistics’, in *Proceedings of the 3rd Berkeley Symposium on Mathematics, Statistics and Probability. Vol, 1*. University of California Press, pp. 157–163.

Rodriguez-Marek, A., Rathje, E. M., Bommer, J. J., Scherbaum, F. and Stafford, P. J. (2014) ‘Application of Single-Station Sigma and Site-Response Characterization in a Probabilistic Seismic-Hazard Analysis for a New Nuclear Site’, *Bulletin of the Seismological Society of America*, 104(4), pp. 1601–1619. doi: 10.1785/0120130196.

Schlissel, D. and Biewald, B. (2008) ‘Nuclear Powerplants Construction Costs’. Cambridge, MA: Synapse Energy Economics.

Seed, H. and Idriss, I. (1970) ‘Soil Moduli and Damping Factors for Dynamic Response Analyses. Report EERC 70-10’. Berkeley: Earthquake Engineering Research Center, University of California.

SESAME (2004) ‘Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation’. SESAME European Research Project WP12., pp. 1–62.

Seyhan, E., Stewart, J. P., Ancheta, T. D., Darragh, R. B. and Graves, R. W. (2014) ‘NGA-West2 Site Database’, *Earthquake Spectra*, 30(3), pp. 1007–1024. doi: 10.1193/062913EQS180M.

Sgobba, S., Puglia, R., Pacor, F., Luzi, L., Russo, E., Felicetta, C., Lanzano, G., D amico, M., Barashino, R., Baltzopoulos, G. and Iervolino, I. (2019) 'REXELweb: a tool for selection of ground-motion records from the Engineering Strong Motion database (ESM)', in *7th International Conference on Earthquake Geotechnical Engineering (ICEGE)*. Roma, Italy.

SHWP (2001) 'Uniform risk spectra for Wylfa Power Station. Report for Magnox Electric plc'. Berkeley: Seismic Hazard Working Party.

Silva, W. J., Abrahamson, N. A., Toro, G. and Costantino, C. (1997) *Description and validation of the stochastic ground motion model, Report for Brookhaven National Laboratory*. Upton, New York.

Smith, J. (1987) 'Influence Diagrams for Statistical Modeling'. Coventry, England: Department of statistics University of Warwick, p. Technical report 117.

Stein, S. and Wysession, M. (2003) *An Introduction to seismology, Earthquakes, and Earth Structure*. Edited by S. Stein and M. Whysession. Blackwell Publishing.

Stephenson, T. A. (2000) 'An Introduction to Bayesian Network Theory and Usage'. Martiny-Valais Switzerland: Dalle Molle Institute for Perceptual Artificial Intelligence.

Stewart, J., Afshari, K. and Hashash, Y. (2014) 'Guidelines for performing hazard-consistent one-dimensional ground response analysis for ground motion prediction.' Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER).

Stewart, J., Goulet, C., Bazzurro, P. and Claassen, R. (2006) *Implementation of 1D Ground Response Analysis in Probabilistic Assessments of Ground Shaking Potential, GeoCongress 2006: Geotechnical Engineering in the Information Technology Age*. doi: 10.1061/40803(187)136.

Straub, D. and Papaioannou, I. (2015) 'Bayesian analysis for learning and updating geotechnical parameters and models with measurements', in Phoon, K. K. and Ching, J. (eds) *Risk and Reliability in Geotechnical Engineering*. 1st Editio. Boca Raton: CRC Press Taylor and Francis Group, pp. 221–263.

Strong, M., Oakley, J. E. and Brennan, A. (2014) 'Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample', *Medical Decision Making*, 34(3), pp. 311–326. doi: 10.1177/0272989X13505910.

Tebib, H. (2017) *Mise en oeuvre des modèles probabilistes d'aléa sismique: Application au territoire Français*. Engineering Thesis. Université de Strasbourg.

Tebib, H., Douglas, J. and Roberts, J. J. (2023) 'Using the value of information to decide when to collect additional data on near-surface site conditions', *Soil Dynamics and Earthquake Engineering*, 165, p. 107654. doi: 10.1016/j.soildyn.2022.107654.

Tesfamariam, S. (2013) 'Seismic risk analysis using Bayesian belief networks', in Tesfamariam, S. and Goda, K. (eds) *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*. 1st edn. Cambridge: Woodhead Publishing Limited, pp. 175–208.

Tesfamariam, S. and Saatcioglu, M. (2010) 'Seismic Vulnerability Assessment of Reinforced Concrete Buildings Using Hierarchical Fuzzy Rule Base Modeling', *Earthquake Spectra*, 26(1), pp. 235–256. doi: 10.1193/1.3280115.

Tesfamariam, S., Sadiq, R. and Najjaran, H. (2010) 'Decision Making Under Uncertainty—An Example for Seismic Risk Management', *Risk analysis: an official publication of the Society for Risk Analysis*, 30, pp. 78–94. doi: 10.1111/j.1539-6924.2009.01331.x.

Theodoulidis, N., Bard, P.-Y., Archuleta, R. and Bouchon, M. (1996) 'Horizontal-to-vertical spectral ratio and geological conditions: The case of Garner Valley Downhole Array in southern California', *Bulletin of the Seismological Society of America*, 86, pp. 306–319. doi: 10.1785/BSSA0860020306.

Toro, G. R. (1995) 'Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies. Technical Report No 779574'. Upton, New York: Brookhaven National Laboratory.

Toro, G. R. (2022) 'Uncertainty in Shear-Wave Velocity Profiles', *Journal of Seismology*, 26(4), pp. 713–730. doi: 10.1007/s10950-022-10084-x.

Trainor-Guitton, W., Caers, J. and Mukerji, · (2011) ‘A Methodology for Establishing a Data Reliability Measure for Value of Spatial Information Problems’, *Mathematical geosciences*, 43(8). doi: 10.1007/s11004-011-9367-0.

Trifunac, M. D. (2016) ‘Site conditions and earthquake ground motion – A review’, *Soil Dynamics and Earthquake Engineering*, 90, pp. 88–100. doi: 10.1016/j.soildyn.2016.08.003.

Tromans, I. J., Aldama-Bustos, G., Douglas, J., Lessi-Cheimariou, A., Hunt, S., Daví, M., Musson, R. M. W., Garrard, G., Strasser, F. O. and Robertson, C. (2019) ‘Probabilistic seismic hazard assessment for a new-build nuclear power plant site in the UK’, *Bulletin of Earthquake Engineering*, 17(1). doi: 10.1007/s10518-018-0441-6.

Tsiboe, F. (2015) *Cost-benefit Analysis and Potential Spillover Effects of Farmer Field Schools in Sub-Saharan Africa: The Case of Cocoa*. University of Arkansas, Fayetteville. Available at: https://books.google.co.uk/books?id=aW7otAEACAAJ.

Tucker, B. E. and King, J. L. (1984) ‘Dependence of sediment-filled valley response on input amplitude and valley properties’, *Bull. Seism. Soc. Am.*, 74, pp. 153–165. doi: 10.1785/BSSA0740010153.

Tuffaha, H. W, Gordon, L. G. and Scuffham, P. A. (2016) ‘Value of information analysis informing adoption and research decisions in a portfolio of health care interventions’, *MDM Policy & Practice*, 1(1). doi: https://doi.org/10.1177/2381468316642238.

Tuffaha, H. W., Strong, M., Gordon, L. G. and Scuffham, P. A. (2016) ‘Efficient Value of Information Calculation Using a Nonparametric Regression Approach: An Applied Perspective’, *Value in Health*, 19(4), pp. 505–509. doi: 10.1016/j.jval.2016.01.011.

Tversky, A. and Kahneman, D. (1974) ‘Judgment under Uncertainty: Heuristics and Biases’, *Science*, 185(4157), pp. 1124–1131. doi: 10.1126/science.185.4157.1124.

USNRC (1956) ‘Domestic Licensing of Production and Utilization Facilities’. 10CFR Part 50: U.S. Nuclear Regulatory Commission.

USNRC (2007) ‘A performance-based approach to define the site-specific earthquake ground motion. Regulatory Guide 1.208’. Washington: U.S. Nuclear Regulatory Commission.

USNRC (2012) ‘Practical Implementation Guidelines for SSHAC level 3 and 4 hazard studies. NUREG-2117, Rev. 1’. U.S. Nuclear Regulatory Commission.

Vamvatsikos, D. and Cornell, C. A. (2002) ‘Incremental dynamic analysis’, *Earthquake Engineering & Structural Dynamics*, 31(3), pp. 491–514. doi: https://doi.org/10.1002/eqe.141.

Van-Tien Dao, W., Nhat Hanh Le, A., Ming-Sung Cheng, J. and Chao Chen, D. (2014) ‘Social media advertising value’, *International Journal of Advertising*, 33(2), pp. 271–294. doi: 10.2501/IJA-33-2-271-294.

Wallace, C. S. and Korb, K. B. (1999) ‘Learning linear causal models by MML sampling’, in Gammerman, A. (ed.) *Causal Models and Intelligent Data Management*. Springer Verlag, pp. 89–111.

Walley, R. J., Smith, C. L., Gale, J. D. and Woodward, P. (2015) ‘Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study’, *Pharmaceutical Statistics*, 14(3), pp. 205–215. doi: 10.1002/pst.1675.

Walls, L. and Quigley, J. (2001) ‘Building prior distributions to support Bayesian reliability growth modelling using expert judgement’, *Reliability Engineering & System Safety*, 74(2), pp. 117–128. doi: 10.1016/S0951-8320(01)00069-2.

Wang, Z. (2009) ‘Seismic Hazard vs. Seismic Risk’, *Seismological Research Letters*, 80(5), pp. 673–674. doi: 10.1785/gssrl.80.5.673.

Wang, Z. (2011) ‘Seismic Hazard Assessment: Issues and Alternatives’, *Pure and Applied Geophysics*, 168, pp. 11–25. doi: 10.1007/s00024-010-0148-3.

Weisbrod, B. A. (1961) *Economics of public health: measuring the economic impact of diseases*. University of Pennsylvania Press.

Wen, Y. K., Ellingwood, B. R., Veneziano, D. and Bracci, J. (2003) *Uncertainty modeling in Earthquake Engineering*.

Wiemer, S. and Wyss, M. (2000) ‘Minimum Magnitude of Completeness in Earthquake Catalogs: Examples from Alaska, The Western United States, and Japan’, *Bulletin of the Seismological Society of America*, 90(4), pp. 859–869. doi: 10.1785/0119990114.

Willan, A. R. and Pinto, E. M. (2005) ‘The value of information and optimal clinical trial design’, *Statistics in Medicine*, 24(12), pp. 1791–1806. doi: 10.1002/sim.2069.

Williams, M. S. (2016) ‘Structural analysis’, in Elghazouli, A. (ed.) *Seismic Design of Buildings to Eurocode 8*. 2nd Edition. London: CRC Press, pp. 42–72. doi: 10.1201/9781315368221.

Williams, R. J., Gardoni, P. and Bracci, J. M. (2009) ‘Decision analysis for seismic retrofit of structures’, *Structural Safety*, 31(2), pp. 188–196. doi: 10.1016/j.strusafe.2008.06.017.

Wilson, E. C. F. (2015) ‘A Practical Guide to Value of Information Analysis’, *PharmacoEconomics*, (33), pp. 105–121. doi: 10.1007/s40273-014-0219-x.

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., Hiemer, S., Meletti, C., Musson, R. W., Rovida, A. N., Sesetyan, K. and Stucchi, M. (2015) ‘The 2013 European Seismic Hazard Model: key components and results’, *Bulletin of Earthquake Engineering*, 13(12), pp. 3553–3596. doi: 10.1007/s10518-015-9795-1.

Woessner, J. and Wiemer, S. (2005) ‘Assessing the Quality of Earthquake Catalogues: Estimating the Magnitude of Completeness and Its Uncertainty’, *Bulletin of the Seismological Society of America*, 95(2), pp. 684–698. doi: https://doi.org/10.1785/0120040007.

Woodberry, O., Nicholson, A. E., Korb, K. B. and Pollino, C. (2004) ‘Parameterising Bayesian networks’, in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. Springer Verlag, pp. 1101–1107. doi: 10.1007/978-3-540-30549-1_108.

Woodberry, O., Nicholson, A., Korb, K. and Pollino, C. (2004) ‘Parametrising Bayesian networks: A case study in ecological risk assessment. Technical Report

2004/159'. Monash University: School of Computer Science and Software Engineering.

Wotherspoon, L. M., Wentz, R., Cox, B. R. and Stolte, A. C. (2021) 'Assessing the quality and uncertainty of in-situ seismic investigation methods', in *NZGS Symposium 2021: Good Grounds for the Future*. Dunedin, New Zealand.

Zadeh, L. A. (1994) 'Fuzzy Logic, Neural Networks and Soft Computing', *Communication of the ACM*, 3(37), pp. 77–84.

Zentner, I., Gündel, M. and Bonfils, N. (2017) 'Fragility analysis methods: Review of existing approaches and application', *Nuclear Engineering and Design*, 323, pp. 245–258. doi: 10.1016/J.NUCENGDES.2016.12.021.

Zhao, C., Yu, N., Oz, Y., Wang, J. and Mo, Y. L. (2020) 'Seismic fragility analysis of nuclear power plant structure under far-field ground motions', *Engineering Structures*, 219, p. 110890. doi: https://doi.org/10.1016/j.engstruct.2020.110890

Appendix A

This appendix contains the summary of two semi-structured interviews with two group participants. Chapter 3 (section 3.4) outlines the aims and main motivations of conducting the interviews as well as the main outputs and conclusions.

A.1 Interview with the Chief Civil Engineer at EDF Energy

This section presents a summary of recorded interviews conducted via Skype For Business with David J Hamilton who works as the Chief Civil Engineer within the Design Authority Department at EDF Energy in East Kilbride. Mr Hamilton is responsible for the engineering governance and oversight of projects related to civil engineering. The interviews were split in two sessions (23rd July and 19th August 2020) with a duration of approximatively 1 hour each.

Mr Hamilton was willing to share his knowledge and experience regarding seismic hazard assessments performed for nuclear facilities owned by EDF Energy as a client for such assessments. During the discussion, Mr Hamilton discussed the various activities of EDF Energy within the UK and the current relationships with seismic hazard analysts and the Office for Nuclear Regulation (ONR). He also emphasised the concerns around nuclear safety and the importance of taking into account uncertainties regarding the seismic hazard. Finally, the question of the value of information was raised in order to learn more about the current practice in prioritising and justifying the collection of data as well as the possible gaps that the current PhD project can fill.

The following interview was divided into six main sections:

- Overview
- Required parameters for Probabilistic Seismic Hazard Assessment (PSHA)
- Life duration
- Budget
- Data collection
- Estimation of the value of information

The summary of the interviews with EDF Energy includes each question asked and a summary of their response. It should be noted that the following provides a summary of the answers and not a verbatim transcript of the interviews.

Overview

What are EDF Energy activities in the nuclear sector related to seismic hazard assessment?

EDF Energy is responsible for operating eight nuclear power plants throughout the United Kingdom. Some of the concerns of the Design Authority Department, where Mr Hamilton is the chief civil engineer, relate to natural hazards, especially when it comes to methods of assessing these hazards and their impacts in terms of structural response. Seismic hazard studies are important to EDF Energy because seismic safety cases have to be put in place for all nuclear sites and these are reviewed by the ONR. For the eight sites, some of them were commissioned in the 1970s, others in the 1980s and others in the 1990s, such seismic safety cases are required. Some of the nuclear power plants were seismically designed but half of the fleet were not originally assessed for seismic hazard. EDF Energy is making sure that the seismic hazard against which existing sites are assessed remains conservative. EDF Energy use modern probabilistic seismic hazard assessments (PSHAs) to quantify changes in the seismic hazard.

What is the frequency in investing in seismic hazard assessment for nuclear power plants?

There are 36 license conditions that are placed upon EDF Energy as a nuclear operator in the UK. One of these requires undertaking periodic safety reviews (PSR) which take place every 10 years to make sure that the hazard remains conservative. EDF Energy is interested in developing new-build power plants, which are generally built adjacent to existing sites. This is mainly because of the existing expertise in the area as well as the societal acceptance of nuclear facilities. In addition, they would be located near existing electricity connections and cooling water structures.

With the help of Jacobs and their team of seismic hazard analysts, sensitivity tests were performed for 4 different sites to assess how the hazard might change in the light of modern ground motion prediction equations (GMPEs) and other inputs. Sensitivity

studies are done on a rolling annual basis with the help of experts who have access to projects like SHARE and SIGMA and initiatives coordinated by the Organisation for Economic Co-operation and Development (OECD) and the International Atomic Energy Agency (IAEA). These guidelines are used to understand where PSHA is heading by testing the different modern practices for EDF Energy sites on an annual basis.

Relationship with the Office for Nuclear Regulation (ONR)

ONR has a legal duty to protect the general public and make sure that safety cases are robust. The ONR in the UK is not a prescriptive regulator unlike the regulator in other countries (e.g., USA). In the UK, nuclear licensees like EDF are responsible for developing safety cases adequately. The ONR reviews these safety cases with the help of hazard analysts within their team. The legal duty of EDF Energy is to reduce all risks to "As Low As Reasonably Practicable" (ALARP). This requirement sets the foundation of the work.

Meetings with ONR members take place on a regular basis and include 4 levels:

- Level 4: Lowest in term of technical exchange – Research plans exchanged and commented on
- Level 3: Actions are being undertaken and discussed
- Level 2 & 1: Represent the higher safety significance where members of the government are presented in case of a national safety interest.

ONR have developed guidelines called Technical Assessment Guides (TAGs). The number 13 (i.e., TAG13) covers the regulator guidance for reviewing seismic hazard safety cases.

Relationship with the seismic hazard analysts and level of involvement in the PSHA process

A great number of hazard assessments were undertaken in the 1990s by the Seismic Hazard Working Party (SHWP), which was founded by the Central Electricity Generating Board (CEGB). CEGB then became Nuclear Electric before becoming British Energy and finally EDF Energy. In the 1990s, this organisation did not undertake many hazard studies. However, now with the use of sensitivity studies, EDF tries to be as active as possible by involving experts in the hazard assessments such as John Douglas (University of Strathclyde) and Iain Tromans's team at Jacobs. EDF

Energy is acting as an “Intelligent Customer” to understand the significance of the parameters involved and keep informed so as to better communicate with the regulator.

Required parameters for PSHA

PSHA requires a lot of input parameters. What are the parameters that have the most impact on the results of PSHA and those which need to be known more accurately than others?

From EDF Energy perspective, seismic site characterisation (SSC) is a major contribution. There is a need to have an awareness and understanding of what is used to develop SSC models.

Ground motion characterisation (GMC) is of high interest as well. Back at the time when the SHWP was performing the analysis, a single GMPE from 1981 was used for peak ground acceleration (PGA); this GMPE was then expanded for response spectral accelerations between 1985 and 1988. Currently with the development of GMPEs, the practice is to include multiple models in a logic tree with different weights according to the level of confidence. This area is highly targeted by EDF Energy to determine which GMPEs are the most suitable for the UK. The company is also taking active interest in the Next Generation Attenuation (NGA) project in the United States where new models are being developed.

Another important input is the earthquake catalogue, which is generally being updated for EDF Energy by Roger Musson, formerly at the British Geological Survey (BGS). Mr Hamilton added that information about minimum magnitude and developments made for the use of Cumulative Absolute Velocity (CAV) is also important. EDF Energy is also interested in disaggregating the hazard to identify events that contribute the most to the overall risk. In addition to that, there is an ongoing project about developing early warning systems with Brian Babtie, a seismologist at the BGS.

Finally, spectral shapes (e.g., Uniform Hazard Spectra, UHS) are of major importance. Studies are being carried out on the conditional mean spectra for situations where the plant includes components sensitive to a particular frequency, in order to target their structural response.

As a part of developing the knowledge, it is important to know which inputs are important. EDF Energy has expanded its understanding in order to know what inputs could help in the future. If a limited budget was available to collect more information, there is a better understanding of where the most value can be obtained.

Life duration

What is the impact of the limited life duration of an existing nuclear facility on the decision of collecting new data?

Within the next five years, four of the eight existing nuclear sites will no longer be generating and no hazard studies will likely be conducted for those sites. When a nuclear power plant shuts down, 5 years are necessary for the nuclear fuel to be removed. During that period, there is still significant nuclear risk but a magnitude less because concerns will focus on safely handling the fuel. All the seismic safety requirements must be applied during fuel removal. For the other sites: one will operate until 2030, two until 2035 and the other until 2050. The proportion of efforts is directly related to the duration of the operation of the nuclear site. Once the fuel is off the site, the site will stay in maintenance for 100 years and will be considered as temporary depositary.

For an existing licensee like EDF, there is a reliance on ground investigation (GI) studies that were done prior to construction. Additional site investigations, for now, are not likely to be undertaken. Because of limitations regarding GIs, published shear-wave velocity (Vs) profiles, for example, can be used.

Budget

What is the budget for performing PSHA and what differs between old and new-builds?

PSHA studies are more costly for new-builds. The budget is conditioned with the level of rigour and expert elicitation required. For new sites, the budget is between £2 and £5 million and a magnitude less for existing sites as the risk is lower because of, as said earlier, the shorter operating duration. The greater the level of details required in a study, the higher the need to revisit the GI. Indeed, the GI for existing sites were more focused on standard civil engineering buildings, like standard rock conditions rather than looking at the dynamic properties of the ground.

Data collection

Gap analysis: Before performing any calculations, it is common to have an overview of the available data and proceed to a gap analysis, which is a review of the existing database and the data that need to be gathered. When and how does this procedure take place?

In the first instance, this is done within EDF Energy. It is part of the strategic siting assessment phase. It is common to look at a site X and then decide whether it is suitable to build a new nuclear facility. If there is already a station at that location, it is obvious that there will be a connection that can handle the level of power output that a new power station would generate. Knowing that 9 times out of 10 a new power plant will be adjacent to an existing power plant, EDF Energy will go through archive records to extract all the available geotechnical information. A database exists for all the sites with documented records of every geotechnical report (e.g., boreholes, ground and water table testing). Then, EDF Energy interacts with a consultant, shares what is available and discusses the gaps. A decision is made afterward regarding undertaking further GI.

How do you include and manage the uncertainties in the decision making?

First, it is important to make sure that the GI results are representative of the whole site. Then, design codes are considered like the ASCE (2005) code to give guidance on how to apply an upper and lower bound as well as a best estimate to capture a sensible range of what geotechnical data might be providing. When a site-specific UHS is derived for a new-build, the 84^{th} percent is considered and used as a comparison with the design spectrum used. Whereas on existing stations, the expected mean is often considered. This is mainly due to the fact that SHWP studies were known to include conservatism at source and deliberate bias within the choice of input parameters. The challenge for the operator is to derive a conservatist representation of the hazard for the return period of interest, 10,000 years typically. If new features regarding the site are discovered and PSHA studies already started, further measurements will be made.

Estimation of the value of information

Gathering more data can reduce epistemic uncertainties but it also costs money and takes time. What are the common strategies used for prioritising data and determining which ones must be collected?

Before undertaking a PSHA study for a new site, an outline specification is produced to set the requirements as a client. The PSHA process can be very detailed so flexibility should be given to the consultants. The decision for collecting more data is usually made while experts discuss parameters in detail to refine their understanding of the data. There are no clear guidelines currently but as a client, EDF Energy relies on the consultants by making sure modern practices are adopted. There is a constant need to challenge and make sure there are no limitations in the used data. If a specific parameter is needed according to academic studies, EDF Energy will invest to make sure the parameter is obtained. This will also be put through a study to make sure to have a value for investing.

Within EDF (this comes more from the French side of the business), expert panels are convened that consider all the evidence before making a decision with the help of a jury of experts along with independent people. They see PSHA not only as a probabilistic analysis but they also look at the long-term benefit and the value of investment. The quality of collected data is also very important because it can influence greatly the quality of the estimated hazard as well as the design of the power plant.

The philosophy within EDF Energy is to keep a balance between having the right amount of data and the capability to understand their meaning. Some of the new-build organisations, possibly including EDF Energy, suffer from the fact that the earthquake community is small and that some of them do not have the “intelligent customer” ability to understand current practices. The possible consequence is that the clients can be slightly led in different directions. Being surrounded by experts helps making sure that EDF Energy is proceeding in the right direction.

A.2 Interview with a seismic hazard team at Jacobs

This section presents a summary of recorded interviews conducted via Skype for Business with three members of the seismic hazard team at Jacobs in their London office: Iain Tromans (Technical Delivery Team lead on seismic hazard projects), Angeliki Lessi-Cheimariou (site characterisation and site response analysis lead) and Guillermo Aldama-Bustos (seismic hazard analyst and geotechnical earthquake engineer). The interviews were split into two sessions (14th and 28th July) with a duration of approximatively 1 hour each.

Jacobs Engineering Group Inc. is an international technical professional services firm headquartered in the USA but with branches worldwide. The company provides a large spectrum of expertise in technical, professional and construction services. It also performs scientific and specialty consulting for a broad range of clients including companies, organisations and government agencies.

The interviews summarised here focused on their activities in seismic hazard assessment, particularly for civil nuclear facilities. The aim of this discussion was to gain an overview of their activities in seismic hazard assessment as well as to identify the various inputs that have major impacts on their calculations. Also, these interviews provide important insights on the process of data collection and dealing with associated uncertainties. Finally, questions about the estimation of the value of the information were asked to identify the current state of practice and the possible gaps that this PhD thesis investigating the use of value of information calculations in this context could fill.

The following interview is divided into five main sections:

- Overview
- Required parameters for PSHA
- Data collection
- Site effects
- Estimation of the value of information

The summary of the interviews includes each question asked and a summary of their response. It should be noted that the following provides a summary of the answers and not a verbatim transcript of the interviews.

Overview

What are Jacobs activities in seismic hazard assessment?

Jacobs performs Probabilistic Seismic Hazard Assessment (PSHA) on conventional projects as well as nuclear projects. Within the UK, the majority of PSHA work is applied to nuclear facilities whether they are defence-related or related to civil power generation.

In the UK, PSHA has been performed since the 1980s. There was a big gap between the mid-1990s until about 2010 where not many PSHA were undertaken because of the lack of development of new nuclear power plants. Since the early 2010s, the Office for Nuclear Regulation (ONR) has developed its approach to the regulation of seismic hazard analysis with the help of experts in different task groups (e.g., Professor Julian J. Bommer and Professor Robert Holdsworth). This oversight from the ONR has led to significant advances in the state of practice of PSHA in the UK.

In the past several years, one of the biggest projects concerned a new nuclear power plant planned by EDF Energy. This was the first PSHA study that got accepted by the ONR following publication of an updated version of TAG13 (one of the ONR's Technical Assessment Guides).

Within Jacobs, PSHA is also performed for petrochemical facilities, new bridges and dams. However, the requirements for bridges and petrochemical facilities are less severe, compared to nuclear facilities.

Locations of PSHA studies

PSHA is performed by the company all over the world, including high seismicity regions such as Kashmir, Trinidad, Venezuela and the USA. The requirements of the PSHAs differ according to the region of study.

Relationship with the regulators (ONR)

Regulators are keen to have an ongoing relationship with the client and the consultants working with the client like Jacobs. They want to be aware of the work that is done as

the study progresses. However, they retain their independence on the process. Therefore, they will comment during workshops on the presented work but will always retain the right to comment on the final submitted documentation.

What is the most performed type of seismic hazard assessment?
PSHA is the most used approach. However, Deterministic Seismic Hazard Assessment (DSHA) is sometimes performed for comparison purposes.

Required parameters for PSHA

PSHA requires a lot of parameters. What are the parameters that have the most impact on the results of PSHA and those which need to be known more accurately than others?
Dr. Aldama-Bustos who is a specialist in bedrock-related geotechnical/geophysical parameters and GMPEs stated that what controls and leads to major uncertainties in the assessed hazard are the GMPEs selection. Site-specific parameters like κ_0 (related to the attenuation of high frequencies in the shallow crust) and the Q attenuation factor, which are sometimes either implicitly or explicitly included in the GMPEs, are also important. The standard deviation associated to the median prediction of the ground motion (from the GMPEs) controls the uncertainties in the PSHA as well. Other factors are of a middle level of influence like the definition and characterisation of the seismic sources in seismicity models and uncertainties related to the activity rates of the sources.

Dr. Lessi-Cheimariou shared her experience in site-response analysis and believes that one of the critical parameters is the shear-wave velocity (V_s) and the impedance contrast between each layer, which are translated into the stiffness of the site. Depending on the stiffness of the site, the second set of critical parameters are those related to the nonlinear properties of the near-surface materials, including the shear modulus degradation with shear strain and the increase in damping as the shear strain increases.

She also stated that as part of site-response analysis for PSHA purposes, the variability of V_s across the site and at depth are important to determine. These parameters define the standard deviation of the natural logarithm of V_s, i.e. σ_{lnV_s} which is also important for the site-response analysis.

Dr. Tromans added that complex geology is also a factor. In the case where there are multiple zones within the same site with different geological features, the requirements for site characterisation would mean having to obtain information about the potential V_s profiles in each of these different areas of the site.

All of these parameters can be full of uncertainties even with rigorous ground investigations and geophysical techniques. How do you manage including these uncertainties within the calculations?

Dr. Lessi-Cheimariou stated that for the site-response analysis, one of the ways of incorporating aleatory and epistemic is through Monte Carlo simulations. The variability is included as an input in the σ_{lnV_s} parameter explained earlier, which can also take account of epistemic uncertainty. As an output, uncertainty is expressed in terms of standard deviation of the natural logarithm of amplification factors obtained through a convolution approach. Another way is to use logic trees when there are different geologies or different best estimates of the velocity profiles.

Dr. Aldama-Bustos added that it is sometimes difficult to obtain a standard deviation associated with the median estimate of any parameter. In the case of GMPEs, tests are performed to assess if models can explain observed strong-motion data from the region of interest. According to the results of comparison between GMPE predictions and observations, a level of confidence is estimated for each GMPE, which will be translated into weights assigned to the different branches of the logic tree. The κ_0 parameter mentioned earlier is often expressed as a model giving an average value and a standard deviation because this parameter is often difficult to determine precisely. Like for the GMPEs, weights will be assigned to the different values characterising its distribution in the logic tree.

Data collection

Gap analysis: Before performing any calculations, it is common to have an overview of the available data and proceed to a gap analysis, which is a review of the existing database and the data that need to be gathered. When and how does this procedure take place?

Gap analysis constitutes the first stage of the PSHA within Jacobs approach. New nuclear sites in the UK are generally adjacent to existing nuclear sites. For that reason, there will always be an available database of the site properties. In this context, a couple of examples were given:

Example 1: Nuclear power plant 1A is being decommissioned. Nuclear power plant 1B is still operating. Nuclear power plant 1C is the proposed site. The data for 1A are from the late 1950s to early 1960s, which is quite common for nuclear sites in the UK. Although useful, these data are of limited quality due to the techniques that were used at that time. 1B data were collected in the 1980s and 1990s and represent a huge step forward in terms of quantity and quality. Nevertheless, they would not always meet the requirements of modern practice for ground investigations (GIs).

This first stage (gap analysis) usually lasts for 3 to 4 months. Gap analysis often continues during the whole duration of the project as the database is continually being updated.

Example 2: In one of the sites studied by Jacobs, there was the possibility of drilling incline boreholes. This possibility was identified as being potentially useful during the gap analysis at the beginning of the project. Whether these costly boreholes would be drilled or not had to be decided at a later stage. When more data was gathered following the original gap analysis, these boreholes were seen to be not required. This type of decision making is an ongoing process during the entire length of the project.

What are the usual timelines for data collection?

GIs can take time. The duration depends on the project as it is related to, for example, the resources available. The PSHA starts when most of the GI campaigns have been carried out. GI is important, not only, for PSHA purposes but also other design calculations as the GI provides geotechnical parameters critical for foundation design and so forth. The exact duration of data collection depends on the available dataset

(quality and amount). All new-build nuclear sites within the UK are adjacent to existing facilities, which means data collection often takes less time than it would if completely new sites were being investigated.

Example: Project Site 2

Jacobs has been involved in this project since the beginning. There was at least 6 months of preparation before the GI started.

The four phases of GI, their duration and purpose were the following:

- Phase 1 and 2 (6-9 months): Surface geophysics and H/V studies
- Phase 3 (9 months): Intrusive GI
- Phase 4 (2-3 months): Offshore GI

Do requirements for data collection depend on the type of facility and the level of seismic hazard?

The type of GI is closely related to the type of project. There are fewer requirements when it comes to a residential development project than for nuclear power plants, for example. For nuclear projects in the UK, which represents a region of low seismicity, there are very onerous requirements of GI. From all projects in the UK, those related to nuclear facilities have the most detailed GI. For two nuclear projects based in areas of different seismic hazard, site-response analysis would require consideration of nonlinear site response (high seismicity regions like California) and in the others, such as the UK, an equivalent linear approach. This type of consideration helps define which GI needs to be performed.

Site effects

What are the site effects that you consider most in your PSHA studies?

The type of site effects included in PSHA studies depend on the characteristics of the site and whether there is a 1-dimensional or 2-dimensional site response. Near-surface effects are taken into account within the site-response analysis, which considers features from the ground level down to what is defined as the bedrock. These effects include the de-amplification and amplification due to the presence of softer sediments. Associated critical parameters are the V_s and the nonlinear response of these materials. Liquefaction, another type of site effect, is a separate analysis from PSHA and includes

different parameters, including the water level. It is also important to look at the geomorphology of the site, e.g., dipping strata, basements and specific impedance contrasts.

Which data related to site effects are usually available and which need to be gathered?
The information used in the analysis should always be site specific when it comes to assessing site effects. There is usually no database related to site-specific properties existing unlike earthquake catalogues, for example. Some data may be available like laboratory data or dynamic parameters, which can be used within empirical relationships to establish V_s profiles. However, these are rarely representative of soil properties. Hence, GIs are essential to capture the soil features and site effect parameters.

Estimation of the value of information

Gathering more data can reduce epistemic uncertainties but it also costs money and takes time. What are the common strategies used for prioritising data and determining which data should be collected?

Examples of useful guidance include the following:

- TAG13 by the ONR: data requirements and use of existing data.
- The International National Atomic Agency (IAEA) guidelines: site characterisation and site selection.
- United States Nuclear Regulatory Commission (USNRC) reports
- Reports of the INTERPACIFIC project: Combining intrusive and non-intrusive investigation data

Are there methods usually used to determine the cost and benefits of collecting data?
No specific method is currently used for the purpose of determining the benefit in terms of reduced costs at Jacobs but progress is being made towards developing such a method. There is not always a clear cost-benefit from gathering data. Most of the time data collection is related to regulatory assurance. It is not always about knowing the right answer but more related to knowing the average hazard with a good level of confidence. When performing PSHA for a critical site or facilities, it is important to capture the range of epistemic and aleatory uncertainties. It is sometimes hard to quantify that in terms of cost-benefit because generally, regulatory assurance is the primary concern.

Example: Project Site 1

The client asked the company about the impact on the PSHA if certain datasets were not available. The seismic team developed a risk matrix where various datasets were listed, the associated effects of not including them in the PSHA and the knock-on impact on the capital cost of the project. However, these values are still difficult to define.

In some cases, instrumentation is installed on the site to measure a certain feature, e.g., downhole array seismometers to characterise the κ_0 value. It is often known in advance that the initial model of the target parameter would not change by much (e.g., κ_0 in this case) but this initiative is important for the client to prove to the owner or regulator that he is willing to go the extra mile. However, putting a cost and benefit on such effort is quite challenging.

Whilst the client might not benefit directly from an investment on certain instrumentation in the short term, benefits can become valuable in the long term. Because of the 10 years cycle of periodic reviews that the ONR requires, the clients are setting up contracts with the British Geological Survey (BGS) to have ongoing data saved on the BGS servers. These data could be useful when it is necessary to update the PSHA study in 10-years' time, for example. These data would help to update specific models and should significantly reduce the uncertainty bounds.

A GI technique can also be beneficial for different purposes. The client is keen to benefit from it for PSHA but also for the geotechnical design of the project. Data gathered during GI phases can be used to inform preliminary design stages for the geotechnical design requirements of the foundations, for example.

What are the arguments you make to convince owners to invest money and time in collecting information?

As a part of the gap analysis, data that are believed to be needed in the project are presented to the client. The arguments made to encourage data collection include regulatory requirements even though they can be vague and make reference to best practice. Jacobs then need to discuss with the client what is relevant good practice for site investigation and PSHA. The INTERPACIFIC project outputs, industry standard

approaches, state of the art approaches and other guidelines are used to assert the importance of combining different data sets.