Soil Dynamics and Earthquake Engineering 165 (2023) 107654

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

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

Haifa Tebib *, John Douglas, Jennifer J. Roberts

Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

A R T I C L E I N F O

Keywords:
Value of information
Data collection
Bayesian updating
Decision trees
Site-response analysis
Seismic hazard assessment

A B S T R A C T

When funding or conducting a seismic hazard assessment, facility owners and seismic hazard analysts need to estimate the possible added value that could be obtained by collecting additional data to inform the assessment. This added value needs to be balanced against the budget and time available for the collection of the data. In other words, they need to answer the question "Is it worth paying to obtain this information?". Conducting a Value of Information (VoI) analysis before any data collection would help to answer this question and to optimise the data collection process.

In this article, we develop and illustrate a method to assess the VoI of improving estimates of the average shear-wave velocity in the top 30 m within the site-response component of a seismic hazard assessment to decide on the optimal seismic design for a reference building in Greece. The approach is based on decision trees to translate the causal relationships between the input parameters in site-response analysis and Bayesian inference to update the model when new data are considered.

The results show that VoI is highly sensitive to prior probabilities and the accuracy of the data collection method (e.g. geophysical survey). This stresses the importance of defining prior probabilities based on available information as well as only considering data collection methods that are suitable for a project's needs and budget.

1. Introduction

The safety of structures must be guaranteed against external hazards, including earthquakes. As a result, hazard assessments are used to assess existing structures as well as inform the design of new structures. Hazard assessments are particularly important for structures such as nuclear facilities, the failure of which could have great consequences [1,2].

Many types of data are needed to accurately assess the seismic hazard at a given location. The data used to inform such assessments are often associated with considerable uncertainties. Uncertainties are generally categorized in two types: epistemic (due to lack of knowledge of a parameter or a process) and 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 Information (VoI) used as inputs to seismic hazard assessments (SHAs).

SHA is an essential step in defining the appropriate seismic design of a structure and to prevent significant damage and collapse. The higher the seismic design levels, generally the costlier the design is to implement. New information can lower the design requirements and consequently lower the costs of constructing new facilities or retrofitting existing ones. Seismic design must serve 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 [3].

In SHAs, parameters should be known to an accuracy that depends on their importance in the assessment. It has been shown that the sensitivity of the results of SHAs (hazard curves, uniform hazard spectra, and eventually, the seismic design) to different inputs can vary considerably [4]. As such, it is important to know to what extent we should seek 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 ~100 m) 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 (Vs) profiles;

* Corresponding author.
E-mail address: haifa.tebib@strath.ac.uk (H. Tebib).

https://doi.org/10.1016/j.soildyn.2022.107654
Received 27 June 2022; Received in revised form 30 September 2022; Accepted 10 November 2022

depth to bedrock and its Vs; Vs_{30}, average Vs in the first 30 m; and f_0, fundamental resonance frequency of the site. If information to constrain these inputs is not available or not known precisely, uncertainties must be integrated into site-response analyses [4–7]. These uncertainties can have a considerable impact on the overall results of the SHA. Epistemic uncertainties can be reduced by collecting new information through geophysical and geotechnical surveys and/or by installing on-site seismometers. Some surveying techniques (e.g. cross-hole 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, or measure the entire Vs profile, but are non-intrusive, cheaper and quicker to undertake. In this context, VoI assessment could help seismic hazard analysts, investors, insurance providers and facilities owners to prioritise which methods, or combination of methods, should be used to characterise a site or location of interest.

Assessing VoI enables the optimisation of the time and money that one is willing to spend to collect new data. Although past authors may not refer to VoI calculations or concepts, there are examples of previous work in earth sciences and civil engineering which attempt to estimate the benefit of a piece of information or a revised design. Examples include studies to assess risk and reliability for retrofitted structures [8], designing site investigations [9] and making drilling decisions [10].

The VoI usually represents the difference between the posterior and prior values (optimal-expected value in the case of a risk-neutral decision maker) which is then compared to the price, P_e, of a given experiment, analysis, or survey. To estimate the VoI, it is crucial to account for the causal-relationships and the dependencies amongst parameters as well as the probabilities (expressing degrees of belief) assigned to each of them [11]. When collecting more information, the procedure followed should enable these probabilities to be updated based on new evidence. To answer the question “Is this parameter worth investigating further?” the framework used to estimate VoI should include the consequences of decisions, as well as the monetary cost that the decision would imply.

This article presents, for the first time, a method to assess the VoI of key parameters in SHA. Although VoI has been used in other fields of study, the challenge here lies in defining the main requirements of VoI and propagating the uncertainties of a measurable variable through the steps of a SHA in order to assess the estimated VoI. The significance of this work comes from demonstrating its potential usefulness in this field where epistemic uncertainties in inputs might lead to high uncertainties in the final results. SHA teams, facilities owners and insurance companies could benefit from evaluating VoI when uncertainties could be reduced by data collection. Indeed, gap analysis within SHA might lead to long debates between clients, stakeholders and SHA analysts on whether data collection should be performed. The question ‘should we collect more data?’ deserves a more quantifiable answer. Helping answer this question is a key objective of this study.

Firstly, we introduce the principles of VoI, example applications, and definitions used in VoI assessment, as well as outlining our approach, which uses decision trees and Bayesian updating within a framework that uses VoI as a measure to assess the benefits of collecting more information. We then present an application of this method for a relatively simple case, where better site characterisation would be useful in determining the optimal seismic design of a building. Our study focuses on site-response analysis and we consider both discrete and continuous uncertainties to show the scope of this approach, and perform sensitivity analyses to assess the relative importance of different inputs. We then discuss the implications of our findings and present some conclusions.

2. Value of information

Raffai and Schlaifer pioneered the use of VoI and provided much of the mathematical background concerning decision making in an uncertain world [12]. Knowing 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.

VoI can be considered as the amount that someone would be willing to pay to obtain a piece of information. It is the difference between the utility of having the information and the utility without that information. In several fields, the decision is often made based on the information available and, in case of uncertainties, decisions are made 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.

2.1. Applications of VoI

Keisler et al. reviewed VoI applications in 260 peer-reviewed articles published between 1990 and 2011 [13]. They find that VoI assessment is typically used to serve two purposes: to guide decision-makers to focus on the information that has the most impact on a decision and to reduce unwanted consequences; and to increase the robustness of the decision-making process. While VoI is becoming more widespread, currently there are few applications within policy and risk, or in geotechnical and civil engineering [13]. To date, most VoI applications are in fields of medicine [14,15] or economics [16,17]. However, in the last decade, VoI approaches have started to be developed in earth sciences [21], remote sensing [22,23], structural health monitoring [24–26], geotechnical site investigation [27] and by the petroleum industry to help making drilling decisions [18–20]. The increase in application of VoI reflects the need to develop more quantitative objective and rigorous decision-making methods.

In SHA and earthquake engineering there is still a gap when it comes to justifying decisions about data collection. It is currently difficult to estimate whether collecting a particular piece of information will be useful for a SHA, and whether, for example, it will change a structural design or retrofit decision. Although cost-benefit analyses have been carried out for many different applications, including SHA, 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 data. This study presents an approach that can start to fill the knowledge gap in data collection decisions for SHA.

2.2. Expected value of perfect/imperfect information

How VoI is modelled is both field and application specific. A utility function and unit of measurement must be defined based on the stakeholders and the decision maker’s interests. The utility function will either help estimate avoided losses, which are often used for external hazards [8], or estimate the maximization of gains, which is often used in marketing and pharmaceutical applications [28]. The unit of this function can be monetary (i.e. representing profit or revenue) or other value indicators (such as happiness, welfare, reputation or equality). In earthquake engineering, it is more common to work toward minimizing the *Expected Loss* [3]. 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 (e.g. probability values or distributions).

VoI can be calculated by quantifying the *Opportunity Loss*, which represents the cost of being wrong when making a decision. We can define the *Expected Opportunity Loss* (EOL) as:

$$EOL = chances\ of\ being\ wrong \times cost\ of\ being\ wrong \quad (1)$$

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

$$EVI = EOL_{Before\ Info} - EOL_{After\ Info} \quad (2)$$

The EVI represents the reduction in risk after considering extra information. When it comes to perfect information, i.e. complete elimination of uncertainty, the associated EOL will be zero and the EVI will simply be the EOL without that information. This is called the *Expected Value of Perfect Information* (EVPI). 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.

In most fields, and particularly in SHA, 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 value that should be maximized is the *Expected Net Gain of Sampling* (ENGS) that considers the cost of obtaining the information.

2.3. Requirements for assessing the VoI

To assess the VoI, the following three components are required to express the relationships and dependencies between the various variables.

2.3.1. Conditional probabilities

Conditional probabilities describe the probability of a value given a known variable. These are important for VoI analyses as conditioning an observation from information could lead to an improvement in decision making. For conditional probabilities it is essential to express the dependencies amongst the variables.

2.3.2. Graphical models

Graphical models are powerful tools to understand the degree to which variables are linked, connected and influenced by each other. Example graphical models include Bayesian networks (BNs), Bayes nets or belief/decision trees. By using conditional probability density functions within the statistical model, the evidence regarding a parameter is propagated to other nodes.

2.3.3. Priors, likelihood functions and Bayes rule

When collecting a piece of information y on a measure of interest x, y can be "perfect" meaning that it perfectly informs us about x, or, more likely, "imperfect/partial" (e.g., due to noise or because it represents only one variable of a multivariate set).

In SHA, a Vs profile could be the measure of interest x and y could be the dispersion curves obtained from multi-channel analysis of surface waves (MASW). We would like to compute the posterior model for x conditioned on y, i.e. $p(x|y)$. This is done using Bayes rule and requires a prior model for x, $p(x)$, a conditional probability density function on the data y known as the likelihood function $p(y|x)$ and the marginal probability density function on the data y, $p(y)$. Thus, $p(x|y)$ is expressed as follows:

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} \quad (3)$$

The posterior can be viewed as a combination of prior knowledge and information brought by the data. Bayes rule is used to construct posterior distributions that are essential for VoI calculations.

3. Application of VoI for seismic design

This section defines the case study that we use to illustrate an application of VoI to determine whether additional data collection is justified when deciding on an appropriate level of seismic design for a hypothetical 4-storey 3-bay reinforced concrete building in the city of Patras, Greece (Fig. 1). This building was designed for different levels of design peak ground acceleration [3], PGA_d, according to Eurocodes [29,30] and using the Type 1 horizontal design acceleration spectrum of Eurocode 8 (EC8). 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. The application is presented both for discrete and continuous priors. For simplicity, we consider a lack of information about a single parameter, Vs_{30}, which is a commonly used proxy for near-surface site amplification within seismic design codes.

The dilemma here, which we use VoI to inform, is between these two options:

1 Choose a particular seismic design and take the risk of choosing: (a) a higher, and more costly, seismic design than needed; or (b) a lower and less-resistant design but 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, but the tests will have a cost that depends on their type, the company hired to perform them, and other factors such as the price of buying or renting the testing equipment.

3.1. The case of a discrete uncertain parameter

In our first scenario, the uncertain parameter (Vs_{30}) is assumed to have discrete values. This simplification is useful to demonstrate how VoI is calculated and to understand the impact on the results of various inputs. 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. A second scenario (in Section 3.2) considers the more realistic case of continuous distributions.

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

3.1.1. 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 so the more resistant the building needs to be. As such, *design1* is more resistant to seismic loads than *design2*, associated to V_2. 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-optimized 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 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.

3.1.2. After additional data collection

3 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

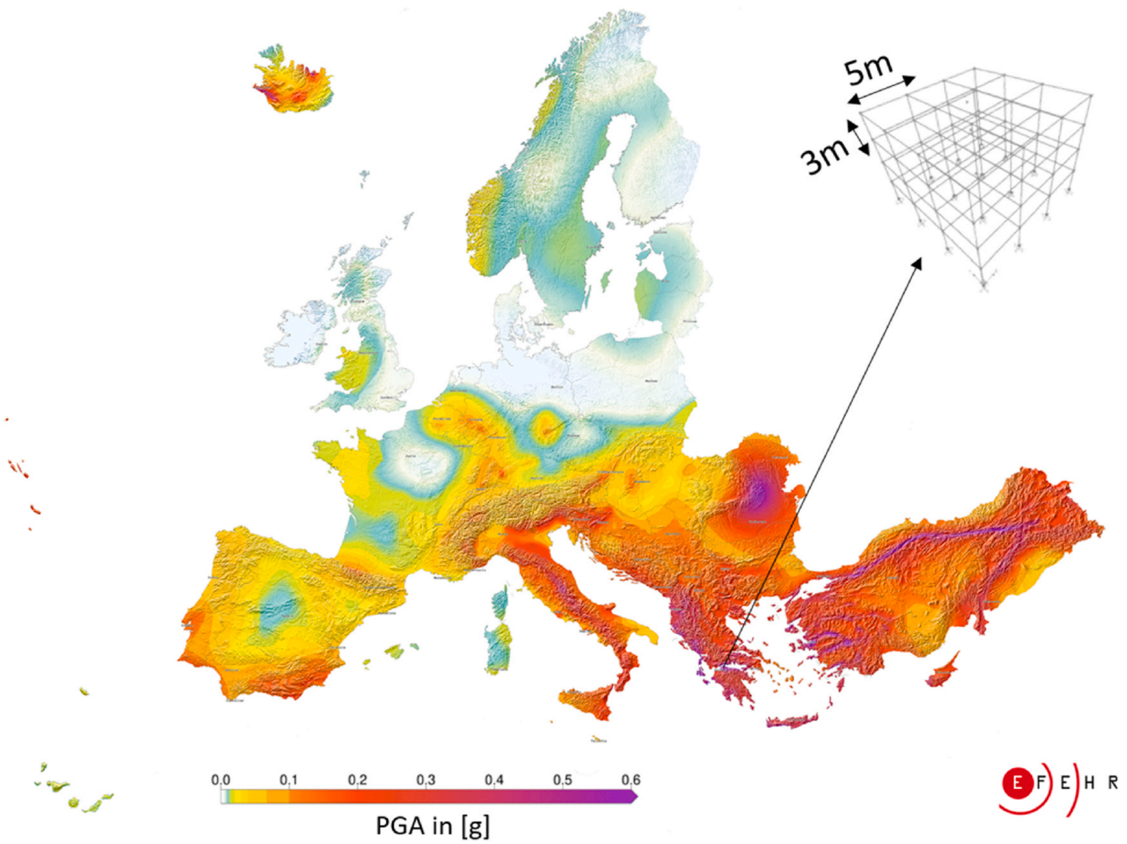

Fig. 1. ESHM20 map of peak ground acceleration [PGA] for 10% probability of exceedance in 50 years (average return period of 475 years) [31]. The 3D building design model is represented and linked to its location in Patras [3].

tests, where a geophone is lowered into a borehole and shots are fired to estimate the Vs within the vertical soil column under the site. An imperfect test would be geophysical survey techniques such as MASW or ambient vibration measurements.

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

3.1.3. *Input and parameters for calculating VoI*

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.

Fig. 2 shows an influence diagram that summarises the parameters that are computed and/or used to estimate the appropriate PGA. The components are:

- Vs_{30}: Average shear-waves velocity in the first 30 m. This is considered here to fully represent the site characterisation.
- Site-amplification factor, Fs. Eq. (4)
- PGA on rock: Peak ground acceleration at a reference rock site
- PGA on soil: Product of PGA on rock and the site-amplification factor
- Design PGA: PGA to which the building is seismically designed
- Expected Losses: Considered as the outcomes for VoI calculations and detailed in this subsection.

The hazard curves 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 fixed to 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 [32].

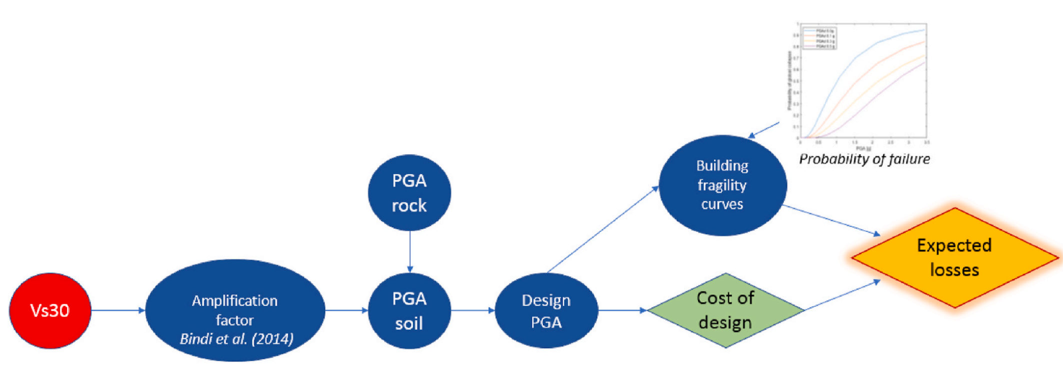

Fig. 2. Framework for site-response analysis and the estimation of the expected costs and losses. Blue circles represent known parameters, the red circle (Vs_{30}) is the uncertain parameter, the lozenge green node is the initial cost of design and the yellow node is the outcome (estimated costs and losses).

In this simple case study, site-response analysis is simplified by neglecting non-linear effects and by assuming Vs_{30} completely controls the near-surface site amplification. Then the frequency-independent amplification factor F_s of the site is assumed to be [33]:

$$Fs = \gamma \log_{10}\left(\frac{Vs_{30}}{V_{ref}}\right) \tag{4}$$

Where V_{ref} is fixed to 800 m/s and γ = -0.3019.

3.1.3.1. Prior probabilities. For this binary case study, the prior probability, *p*, is the probability that V_1 is the true Vs_{30} at the site. Similarly, *1-p* is the probability that V_2 is believed to be the true Vs_{30}.

3.1.3.2. Probability of failure. The fragility curves (derived using Incremental Dynamic Analysis (IDAs) [34]) for the defined building for different PGA_d are obtained from a recent study [3]. The fragility curves, *fc,* indicate the probability of damage for each PGA value.

3.1.3.3. Design PGA. Several approaches can be used to infer the appropriate seismic design, e.g. the uniform-hazard (e.g. Eurocode 8) method or the risk-targeted approach, where the design PGA (PGA_d) is obtained, using an iterative process, by expressing the mean annual frequency of collapse λ_f that will secure the building to an acceptable and controllable risk level [35]. In this case study, we assume that one of these methods is used to infer the design PGA on rock. The design PGA is then simply the retrieved PGA on rock multiplied by the

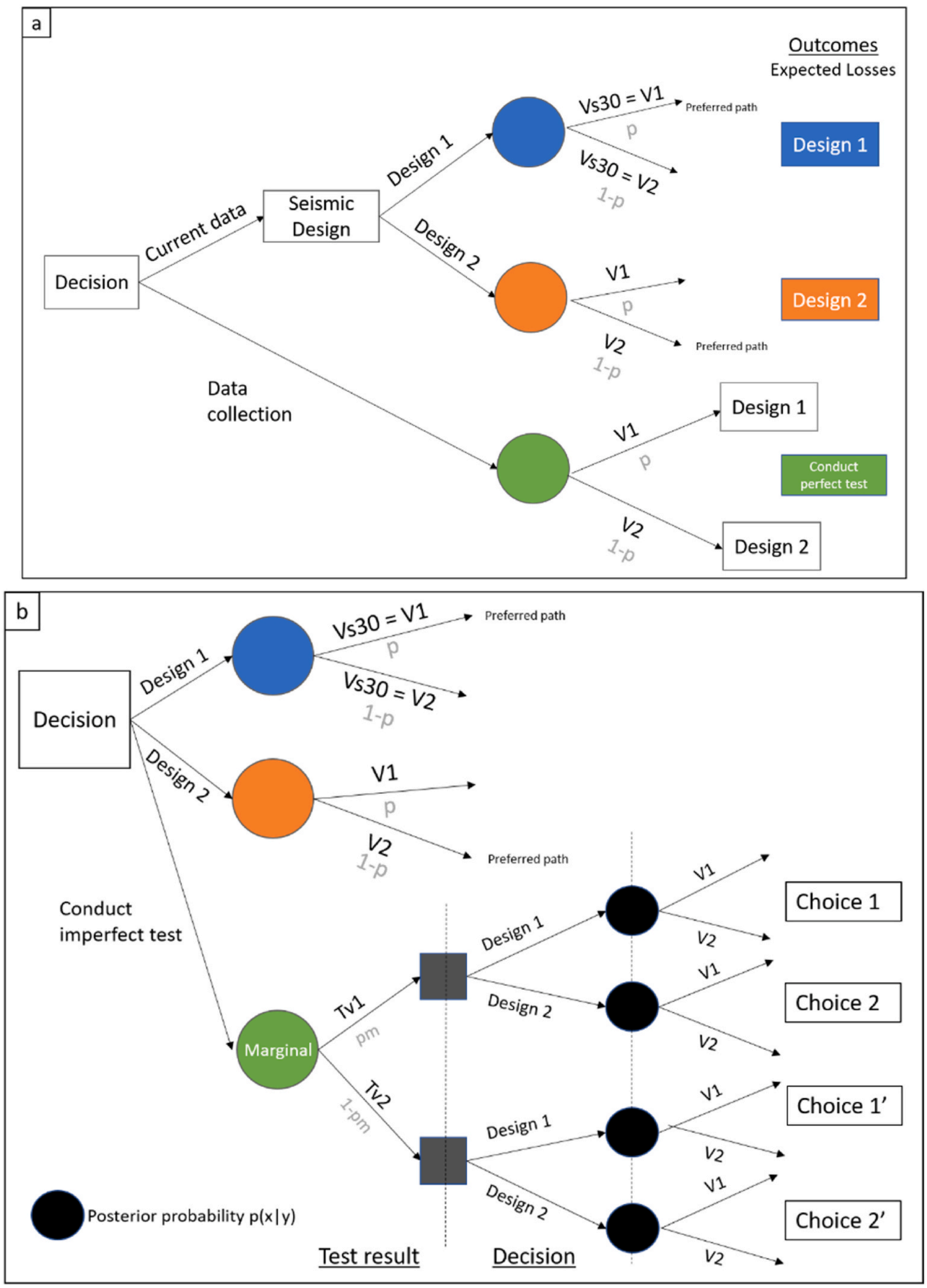

Fig. 3. Decision tree for the computation of EVPI (*a*) and EVII (*b*). Probabilities are displayed in grey with pm as the marginal probability of the test result being V_1.

site-amplification factor (PGA on soil).

3.1.3.4. Outcomes: expected losses. As the VoI here is in monetary units, we need to estimate the economic losses associated to designing the building for a specific PGA_d. The losses due to possible future earthquakes are a function of the seismic 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. 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, $o(x, d)$. E[LCC] includes replacement costs due to total failure as well as additional losses from personal property damage, injuries and fatalities as well as the loss of the function of the building. The inputs and detailed framework for computing the losses are detailed in Ref. [3].

For a decision *d*, the expected outcomes are as follows:

$$E(o(x,d)) = \sum_x o(x,d)p(x) \tag{5}$$

where *o(x,d)* represents the outcomes for a decision *d* if 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 *d* are expressed as follows:

$$o(x,d) = C_d + E[LCC](d).f_c(d,x) \tag{6}$$

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

3.1.4. VoI calculations

As described in section 2.3, decision trees are a useful way of representing the various parameters and their causal dependency. In the decision trees presented in Fig. 3, decisions emerge from the square nodes and probabilities from circle nodes. These two trees depict three possible decisions:

- Applying *Design1* (associated with a $Vs_{30} = V_1$) with available data.
- Applying *Design2* (associated with a $Vs_{30} = V_2$) with available data.
- Conduct a perfect test (a)/imperfect (b) test to obtain more information on Vs_{30}.

3.1.4.1. Prior Value. In the case of *Before information*, two choices are possible, applying *Design1* or *Design2*. The expression of the associated *Prior Value*, *PV*, is as follows:

$$PV = max_{d\in D}\{E(o(x,d))\} = max_{d\in D}\left\{\sum_x o(x,d)p(x)\right\} \tag{7}$$

Where 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*.

3.1.4.2. 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.

$$PoV = \sum_x E(o(x,d_x))p(x) \tag{8}$$

where d_x is the appropriate design decision for state *x*.

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. Since the test is imperfect, uncertainty must be included in the decision tree. An accuracy probability is then assigned to the test. This probability is set by experts from available information about the particular test or by the person applying the test, and it expresses its reliability. If *y* is a value from the imperfect test and *x* is the measure of interest, then the probability of the test being truthful to the real state of *x* is *p(y|x)*, which is called the *likelihood.* This parameter equals 1 in case of perfect information. We define the *posterior model* of *x* conditioned on the data *y*, *p(x|y)*, which is computed using Bayes' rule in Eq. (3) to perform Bayesian updating.

Therefore, the *PoV* is expressed as follows:

$$PoV = \sum_y p(y).\max\left\{\sum_x E(o(x,d))p(x|y)\right\} \tag{9}$$

where *p(y)* is the marginal probability.

3.1.4.3. Value of information. The VoI is then the difference between the *PoV* and the prior value (PV).

$$EVPI \ or \ EVII = PoV - PV \tag{10}$$

Irrespective of its type, PV is always constant. VoI is never negative as adding information always has benefits or no impact in the decision-making process. The VoI is then compared to the cost of the test to decide whether to proceed with data collection.

3.1.5. Scenario 1 results

We acknowledge that the VoI definition and expressions are being used for the first time in this type of application. Thus, several sensitivity analyses are performed in this section and key values computed to validate the method.

The Vs_{30} couple that translates our uncertainty is assumed fixed to [100, 500] m/s and we vary the prior probability. The likelihood probability in case of imperfect information is set to 70%.

Fig. 4-a displays the expected outcomes in euros (losses) combining the construction cost and the expected damage for the three main branches of our decision tree (Fig. 3). In the legend, *Ec1* refers to the expected consequences computed from the branch associated to performing *Design1*, *Ec2* to *Design2* and *Ect perfect/imperfect* are the expected losses after obtaining the perfect/imperfect information and choosing the optimal design. The outcomes for each decision are computed for a range of all possible prior probabilities for V_1 and V_2. Fig. 4-b represents the VoI for several prior probabilities. Considering imperfect information reduces the VoI independently of the prior probabilities. The intersection between *Ec1* and *Ec2* at $p(V_1) \approx 0.4$ in Fig. 4-a is called the *indifference point* [9], where outcomes for both decisions are equal. At this prior probability, the VoI is at its maximum. 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. 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. By changing the construction costs or the probabilities 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. Associated sensitivity analysis showed that the point of indifference is usually in the range of [0.4,0.6].

The likelihood is shown to have a significant impact on the EVII when prior probabilities are fixed. EVII equals zero when the test accuracy is 50%. A test with 50% accuracy is of no help in this binary case as there is 50% chance the test is right and 50% chance it is wrong. On the other side, a test accuracy of 100% is equivalent to a perfect test and EVII is then equal to EVPI. 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 Fig. 5, which shows the EVII for several prior probabilities and different likelihoods and shows that EVII for a

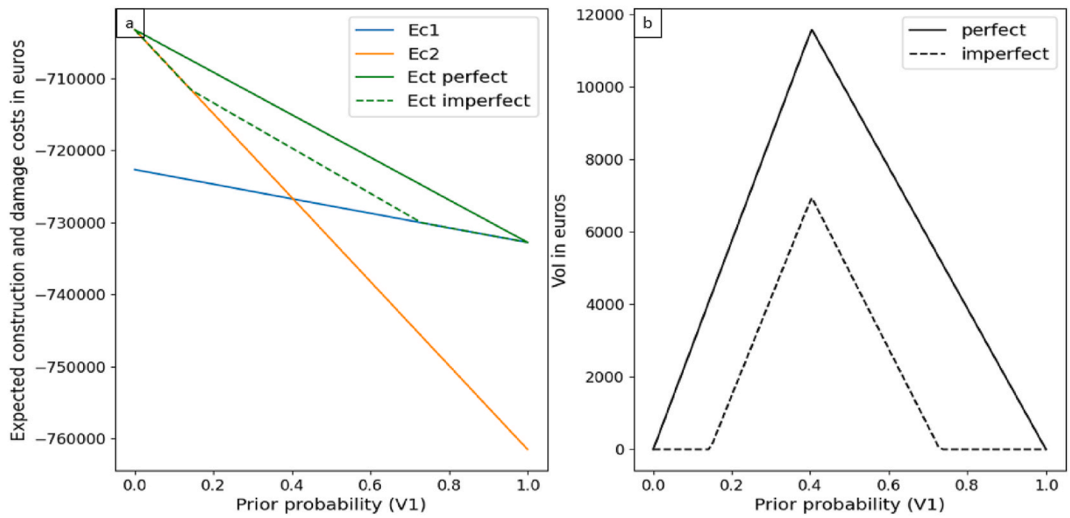

Fig. 4. Sensitivity to prior probabilities for the Expected outcomes of the three main decisions in the decision tree (a) and EVPI (solid line), EVII (dashed line) for the couple [100, 500] m/s (b).

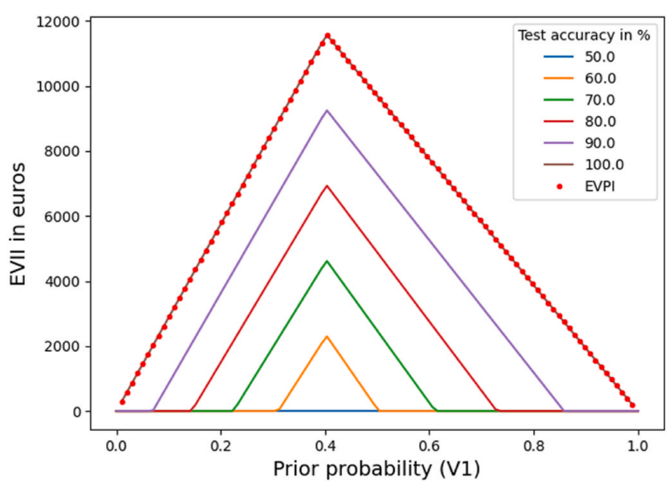

Fig. 5. Sensitivity of EVII to prior probability for different likelihoods and comparison with EVPI.

test accuracy of 100% is equal to the EVPI. This is an additional validation of the method. Moreover, for the additional information to have value, there are conditions for choosing the test when prior probabilities are known. If we take the example of a test that is 90% accurate, we find that beyond a V_1 prior probability of about 85%, there is no value to the additional information. For any likelihood, the value of information is null when the prior probability is equal to or above the likelihood. In other words, it is not worth conducting a test if we are more confident about the value of the measure of interest than the test itself can provide.

3.2. *The case of a continuous uncertain parameter*

The same case study and inputs are considered in this section. In this second scenario, Vs_{30} is not assumed to be discrete but instead continuous. This assumption allows a more realistic definition of uncertainties regarding Vs_{30}. Vs_{30} uncertainties are expressed through probability density functions.

3.2.1. *Input and parameters for calculating VoI*

If some inputs like the PGA on rock, the fragility curves and the definition of expected outcomes remain unchanged, the definition of probabilities and VoI computations can be readily adapted for continuous parameters.

3.2.1.1. Continuous prior probability distribution $p(x)$. Integrating available data and consulting experts are useful in estimating the probability distribution to be used. Available data may suggest a complex distribution, but expert elicitation is more straightforward when using a standard distribution. Here we assume that Vs_{30} uncertainties can expressed by a normal distribution with mean μ and standard deviation σ. It is assumed that the current prior information defines a range of possible Vs_{30} along with extreme possible values, which makes the normal distribution a good assumption. 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.

3.2.1.2. Decisions. The decision-maker must make a choice on applying a particular seismic design to the building based of 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 $D = \{d_1, \ldots, d_i, \ldots d_M\}$ is set *a priori* from the range of possible Vs_{30} based on current information. Indeed, the optimal design PGAs, d_1 and d_n are based on extreme values of Vs_{30}. The range d_1-d_n is then discretised for additional decisions on the seismic design.

3.2.1.3. Outcomes. The outcomes are defined as follow:

$$o(x, d_i) = \begin{cases} C_{d_i} & PGA_{opt}(x) > PGA(d_i) \\ C_{d_i} + E[LCC](d_i).f_c(d_i, x) & PGA_{opt}(x) < PGA(d_i) \end{cases} \tag{11}$$

$PGA_{opt}(x)$ 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 D, only the construction costs are considered. Otherwise, expected losses from total collapse are included in the outcomes. $PGA_{opt}(x)$ is computed by minimizing Eq. (6).

3.2.2. *VoI calculations*

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

$$PriorValue = \max_{d \in D}\left\{ \int o(d_i, x) p(x) dx \right\} \tag{12}$$

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

To approximate the integral in Eq. (12), Monte Carlo simulations [36] are performed to infer samples x_k from the probability distribution $p(x)$. Let n be the size of the samples. Monte Carlo simulation will generate n number of samples according to the chosen probability distribution. This translates into a higher number of samples where the probability is high and vice-versa.

The structure of the decision tree is similar to the discrete case after performing Monte Carlo simulations. 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.

The Monte Carlo integral approximation is used to evaluate the *Prior Value* before information as follows:

$$PriorValue = \max_{D}\left\{\int o(d_i, x)p(x)dx\right\} \sim \max_{D}\left\{\frac{1}{n}\sum_{k=1}^{n} o(d_i, x_k)\right\} \tag{13}$$

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). The *prior value* is then obtained by choosing the decision that will minimize the expected losses. We recall that the *max* operator is used since the losses are expressed as negative values.

EVPI is computed in case of a perfect test where the information is equal to the true value of Vs_{30}. The *PoV* is computed by simply reversing the integral and the *max* operator in Eq. (12):

$$PoV = \int \max_{d_i \in D}\{o(d_i, x)\}p(x)dx \sim \frac{1}{n}\sum_{k=1}^{n}\max_{d_i \in D}\{o(d_i, x_k)\} \tag{14}$$

This calculation requires no additional value calculations if we have computed it for all samples and all alternatives for the prior value approximation in Eq. (12).

The EVPI is then simply calculated by applying Eq. (10). The stability of the results depends on the number n of random samples generated. A low number of samples may result in unstable VoI estimates but a high number of samples may have a high computational cost. To assess the minimum number of samples needed to obtain stable results, EVPI is computed for various numbers of samples and different prior standard deviations. For σ = 120 m/s, EVPI values fluctuate 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>4000. 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, where σ = 120 m/s, the number of samples is fixed to 10 000.

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. The function $e(x)$ represents the *likelihood*.

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

1 Simulation of test observations
2 Construction of observation's probability distribution: Marginal distribution p(y)
3 Monte Carlo sampling of ***N*** observations from p(y): $y = \{y_1, .., y_j, ..y_N\}$
4 Computation of the posterior distribution for an observation y_j using Bayes rule, $p(x|y_j)$
5 Computation of expected outcomes conditioned on each observation y_j and decision d_i
6 Monte Carlo sampling from the posterior distribution $p(x|y_j)$ to infer ***m*** samples of $x = \{x_1, .., x_k, ..x_m\}$ and Monte Carlo integral approximation to compute the Expected Outcomes, *EO*, conditioned on y_j for a decision d_i:

$$EO(d_i|y_j) = \int o(d_i, x)p(x|y_j)dx \sim \frac{1}{m}\sum_{k=0}^{m} o(d_i, x_k) \tag{15}$$

Consequently, the VoI can be computed using this equation:

$$VoI = \int_y p(y).\max_{d_i \in D}\left\{\int_x o(d_i, x)p(x|y)dx\right\}dy - PriorValue$$

$$\sim \frac{1}{N}\sum_{j=0}^{N}\max_{d_i \in D}\left\{\frac{1}{m}\sum_{k=0}^{m} o(d_i, x_k)\right\} - PriorValue \tag{16}$$

This method allows the simulation of a large number of probable observations that include the uncertainties inherent to the test. The expected outcomes for each decision are computed conditioned on the observation and only the decision with the minimum of losses is chosen using the posterior distribution. 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.

3.2.3. Scenario 2 results

EVPI and EVII have been computed for the parameter Vs_{30} with the prior probability distribution defined by $N \sim (\mu = 500, \sigma = 120)\ m/s$. Computation of EVII is more complex and time-consuming than computation of EVPI. For a fixed prior and test error distributions, EVII can be computed in approximately 15 min on a standard PC if N = 1000 observations are simulated from p(y).

The assigned prior distribution has a strong impact on the results. To illustrate this, EVPI is computed using prior distributions of the same mean μ but different standard deviations σ (Fig. 6). EVPI is shown to increase linearly with σ, as a higher prior standard deviation suggest higher uncertainties, i.e. less prior knowledge. The less available information, the higher the chance that a perfect test will be beneficial. The VoI in the case of a continuous variable is slightly smaller than for the discrete case as expected losses are computed including a larger range of alternative possibilities.

3.2.3.1. Posterior distribution behaviour. Next, we assess how posterior distributions for a given observation behave as well as their correlation with the definition of the prior probability distribution *p(x)*. The observation is assumed to derive from a test that has an error function

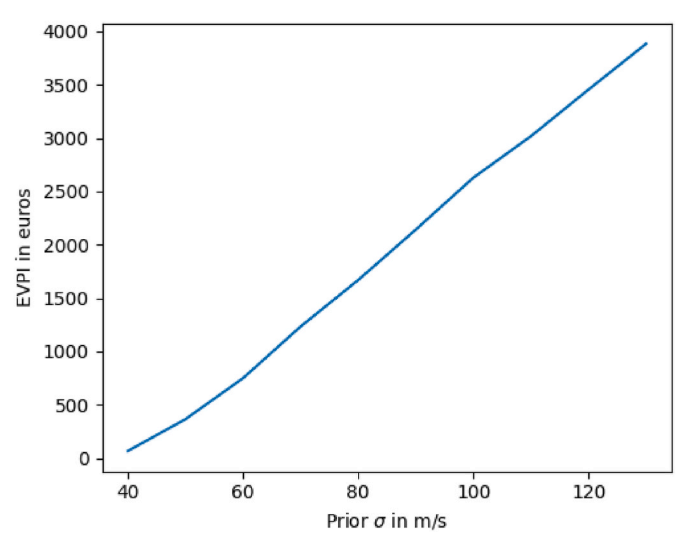

Fig. 6. Sensitivity of EVPI to prior standard deviation.

expressed as $N \sim (\mu_t = 0, \sigma_t = 30)\ m/s$. The prior probability distribution *p(x)* is compared to posterior probability distributions for three different observations y^*, 300, 500 and 700 m/s in Fig. 7. These values are 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)*. The observation y = 500 m/s equals μ.

The posterior of the observation y = 300 m/s has a mean that is slightly higher than the value of the observation (i.e. the result of the test). For y = 700 m/s, the mean is slightly lower. For y = 500 m/s, the mean of the posterior corresponds exactly to the observation. These results show the impact of the prior distribution. For y lower or higher than μ, the associated posterior's *mean* tends to get closer to the highest prior probability meaning that credibility of the observation is not at its highest and suggests closer values to the actual highest prior. This result is confirmed for the case where the observation equals the highest prior probability value 500 m/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, I, chosen for this study is the difference between the posterior distribution mean μ_{post} and the observation. Its value is shown on Fig. 8 for different prior mean μ and observations y_{obs}.

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

$$I = \frac{\mu_{post} - y_{obs}}{\mu}.100 \quad (17)$$

with μ representing the prior mean.

We observe that the further the observation is from the prior mean, the more the posterior's mean is shifted 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.

3.2.3.2. Sensitivity to likelihood. The workflow described above was applied to compute EVII for several likelihood functions. Fig. 9 displays the sensitivity of EVII to different likelihood standard deviations. The result is in accordance with the intuition that a more accurate test (lower error standard deviation) leads to a higher value of information.

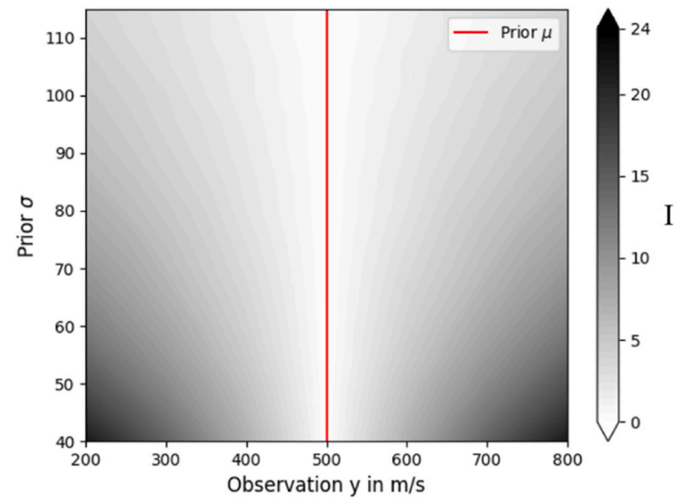

Fig. 8. Impact of prior standard deviation σ on posterior distribution mean μ_{post}

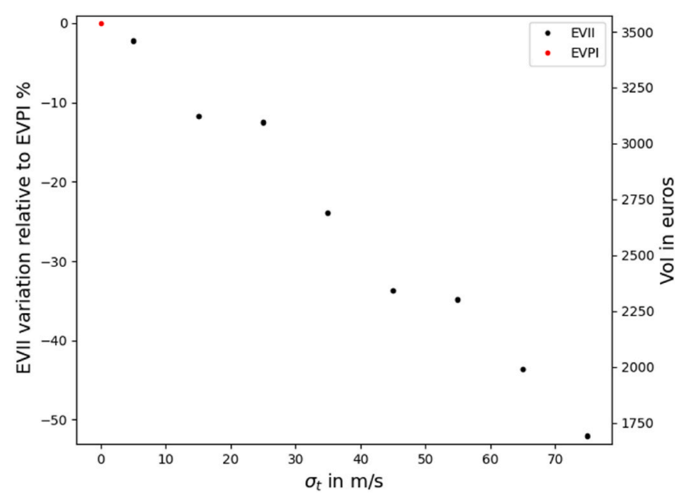

Fig. 9. EVII sensitivity to test's error standard deviation. Variation relative to EVPI (left axis) VoI (right axis).

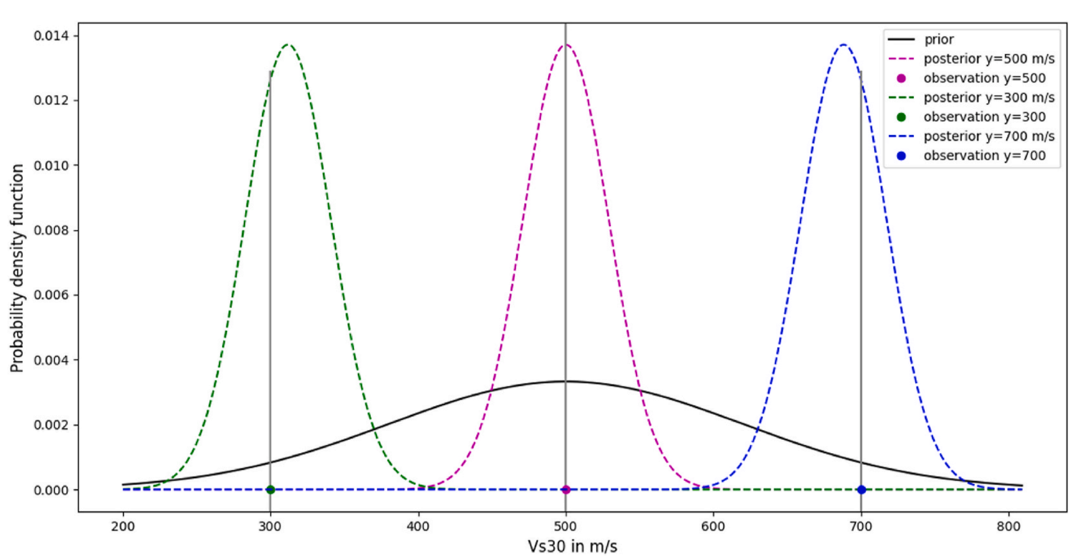

Fig. 7. Prior distribution (black, solid line) and posterior distributions (dashed line) for observations 300 (green), 500 (purple) and 700 m/s (blue).

Moreover, when the test standard deviation tends to zero, the EVII tends to the EVPI.

In a discrete or continuous definition of the uncertain parameter, the likelihood has a strong impact on the EVII. This indicates the need to study and define the reliability of a test based on past experience.

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/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. The EVII should also be compared to the cost of the test. The computation of EVII(test)-cost(test) is an efficient way to choose the type of test (or not to test at all).

3.2.3.3. Other prior distributions. Until now, only normal distributions have been considered when defining priors. 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 used to express uncertainty in Vs_{30} as obtained from surveys [37–40]. Lognormal distributions are skewed to the right (Fig. 10).

To study the sensitivity of VoI for lognormal distributions, we computed EVPI for four different lognormal distributions, with a fixed scale and different shape parameters, $\sigma_{\ln(Vs30)}$ (0.2, 0.3, 0.4 and 0.5). Fig. 7 displays the associated probability density functions and 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.

4. Discussion and conclusions

In this work, we use a case study to develop a method to compute VoI for Vs_{30} within the site-response analysis used in the determination of an optimal seismic design of a particular building. The method was built starting from the simple binary case with perfect information but it was extended to continuous variables and imperfect information, which is more realistic for applications in SHA. The results of the developed method have been shown to follow the expected behaviour. For example, VoI decreases when information is imperfect, i.e. EVII is always lower than EVPI. The level of confidence given to a test has a strong correlation with the prior probabilities when it comes to VoI. The results show that tests or surveys associated with higher uncertainties than the prior probabilities defined by available data and expert knowledge are not worth conducting. Moreover, the more accurate the test the higher the benefits of obtaining the information. It is crucial to bear in mind that more accurate tests and surveys are usually more costly. The VoI approach allows all aspects of a test expenses in time, budget and resources to be considered which will help optimise decision-making. The optimal decision not only represents the collection of information but also the main goals of a project (e.g., choosing an appropriate seismic design, whether to enforce post-earthquake evacuation, or whether to retrofit a building).

The various sensitivity analyses highlight the parameters and inputs that most influence the VoI. These analyses demonstrate that some inputs need to be estimated and chosen carefully to obtain reliable results. The prior probabilities given to the possible values of the uncertain parameter Vs_{30} were shown to have a strong influence on the results. Thus, it is important to use all available information to wisely define the current state of knowledge. These estimates can be guided by expert judgements based on past experience as well as through Empirical Bayes Estimation, which has proven to provide a good estimate of these probabilities in other fields [41–44].

Using continuous variables to compute the VoI is more complex and has higher computational costs than using discrete variables. Nevertheless, assuming the Vs_{30} prior distribution to be continuous is more realistic as it considers a large set of possible values. The definition of the prior distribution and likelihood have a strong impact on the posterior distributions, where the level of confidence in an observation is dependent on the marginal probability density of that same observation. This can be used to define a level of uncertainty below which it is not necessary to descend before making a decision.

This study has demonstrated that calculating the VoI is a helpful step

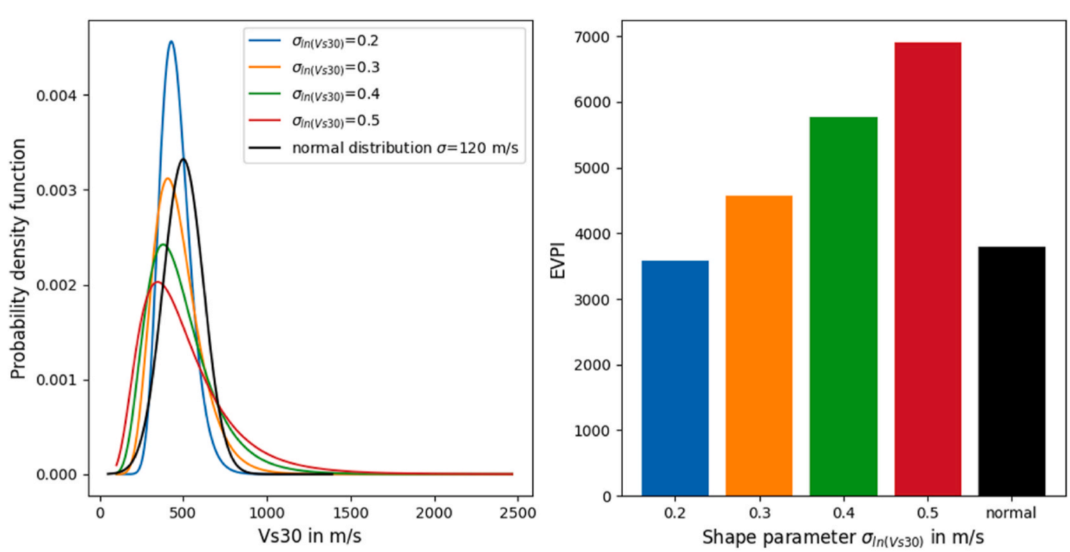

Fig. 10. VoI for lognormal distributions (a) probability density functions (b) EVPI.

in assessing the importance of collecting additional data as well as the maximum investments (in terms of money, time and resources) one should be willing to spend to obtain these data. This guidance would be useful for seismic hazard analysts and facilities owners as budgets are often limited but safety requirements are of utmost importance.

Author statement

Haifa Tebib: Conceptualization, Methodology, Investigation, Software, Writing - Original Draft, Writing - Review & Editing **John Douglas**: Conceptualization, Methodology, Supervision, Writing - Review & Editing **Jennifer J. Roberts**: Conceptualization, Methodology, Supervision, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The first author is undertaking a PhD funded by a University of Strathclyde "Student Excellence Awards" studentship, for which we are grateful. The authors would like to thank Prof. John Quigley for valuable discussions on VoI calculations. The authors are also grateful for helpful discussions with David Hamilton (EDF Energy) and members of the seismic hazard team at Jacobs (Iain Tromans, Angeliki Lessi-Cheimariou and Guillermo Aldama-Bustos). Finally, we thank Dr Milad Kowsari and an anonymous reviewer for their detailed and constructive comments which have strengthened this article.

References

[1] ONR Expert Panel on Natural Hazards. Analysis of seismic hazards for nuclear sites. NS-TAST-GD-013 Annex 1 reference paper. Expert Panel report: GEN-SH-EP-2016-1. 2017.
[2] International Atomic Energy Agency. Seismic hazards in site evaluation for nuclear installations. IAEA safety standards series No. SSG-9 (Rev. 1). 2022. Vienna.
[3] Gkimprixis A, Tubaldi E, Douglas J. Evaluating alternative approaches for the seismic design of structures. Bull Earthq Eng 2020;18(9):4331–61. https://doi.org/10.1007/s10518-020-00858-4.
[4] Aguilar-Meléndez A, Ordaz MG, Puente JD la, Pujades L, Barbat A, Rodríguez-Lozoya HE, Monterrubio-Velasco M, Escalante Martínez JE, Campos-Rios A. Sensitivity analysis of seismic parameters in the probabilistic seismic hazard assessment (PSHA) for barcelona applying the new R-CRISIS. Computación y Sistemas 2018;22:1099–122. https://doi.org/10.13053/cys-22-4-3084.
[5] McGuire RK, Shedlock KM. Statistical uncertainties in seismic hazard evaluations in the United States. Bull Seismol Soc Am 1981;71(4):1287–308. https://doi.org/10.1785/BSSA0710041287.
[6] Ordaz M, Arroyo D. On uncertainties in probabilistic seismic hazard analysis. Earthq Spectra 2016;32(3):1405–18. https://doi.org/10.1193/052015EQS075M.
[7] Abrahamson NA, Bommer JJ. Probability and uncertainty in seismic hazard analysis. Earthq Spectra 2005;21(2):603–7. https://doi.org/10.1193/1.1899158.
[8] Williams RJ, Gardoni P, Bracci JM. Decision analysis for seismic retrofit of structures. Struct Saf 2009;31(2):188–96. https://doi.org/10.1016/j.strusafe.2008.06.017.
[9] Gilbert RB, Habibi M. Assessing the value of information to design site investigation and construction quality assurance programs. In: Phoon KK, Ching J, editors. Risk and reliability in geotechnical engineering. first ed. Boca Raton: CRC Press Taylor and Francis Group; 2015. p. 491–532.
[10] Eidsvik J, Mukerji T, Bhattacharjya D, Dutta G. Value of information analysis of geophysical data for drilling decisions. In: European association of geoscientists & engineers; 2015. https://doi.org/10.3997/2214-4609.201413621 [cited 2020 Nov 25]. p. cp-456-00038.
[11] Raffai H. Decision Analysis: introductory readings of choices under uncertainties. McGraw-Hill; 1997.
[12] Raffai H, Schlaifer R. Applied statistical decision theory. Cambridge MA: M.I.T Press; 1970.
[13] Keisler JM, Collier ZA, Chu E, Sinatra N, Linkov I. Value of information analysis: the state of application, vol. 34. Environment Systems and Decisions; 2014. p. 3–23.
[14] Eckermann S, Willan A. Expected value of information and decision making in HTA. Health Econ 2007;16:195–209. https://doi.org/10.1002/hec.1161.
[15] McFall RM, Treat TA. Quantifying the information value of clinical assessments with signal detection theory. Annu Rev Psychol 1999;50(1):215–41. https://doi.org/10.1146/annurev.psych.50.1.215.
[16] Levitt S, Syverson C. Market distortions when agents are better informed: the value of information in real estate transactions. Rev Econ Stat 2008;90(4):599–611.
[17] Fisher A. Investment under uncertainty and option value in environmental economics. Resour Energy Econ 2000;22:197–204. https://doi.org/10.1016/S0928-7655(00)00025-7.
[18] Eidsvik J, Bhattacharjya D, Mukerji T. Value of information of seismic amplitude and CSEM resistivity. Geophysics 2008;73(4):R59–69. https://doi.org/10.1190/1.2938084.
[19] Bhattacharjya D, Eidsvik J, Mukerji T. The value of information in spatial decision making. Math Geosci 2010;42(2):141–63. https://doi.org/10.1007/s11004-009-9256-y.
[20] Martinelli G, Eidsvik J, Sinding-Larsen R, Rekstad S, Mukerji T. Building Bayesian networks from basin-modelling scenarios for improved geological decision making. Petrol Geosci 2013;19:289–304. https://doi.org/10.1144/petgeo2012-057.
[21] Eidsvik J, Mukerji T, Bhattacharjya D. Value of information in the earth sciences: integrating spatial modeling and decision analysis [Internet]. Cambridge: Cambridge University Press; 2015. https://doi.org/10.1017/CBO9781139628785.
[22] Macauley M. The value of information: measuring the contribution of space-derived earth science data to resource management. Space Pol 2006;22:274–82. https://doi.org/10.1016/j.spacepol.2006.08.003.
[23] Brathwaite J, Saleh JH. Bayesian framework for assessing the value of scientific space systems: value of information approach with application to earth science spacecraft. Acta Astronaut 2013;84:24–35. https://doi.org/10.1016/j.actaastro.2012.10.036.
[24] Cantero-Chinchilla S, Chiachío J, Chiachío M, Chronopoulos D, Jones A. Optimal sensor configuration for ultrasonic guided-wave inspection based on value of information. Mech Syst Signal Process 2020;135:106377. https://doi.org/10.1016/j.ymssp.2019.106377.
[25] Iannacone L, Giordano PF, Gardoni P, Limongelli MP. Quantifying the value of information from inspecting and monitoring engineering systems subject to gradual and shock deterioration. Struct Health Monit 2022;21(1):72–89. https://doi.org/10.1177/1475921720981869.
[26] Giordano PF, Iacovino C, Quqa S, Limongelli MP. The value of seismic structural health monitoring for post-earthquake building evacuation. Bull Earthq Eng 2022. https://doi.org/10.1007/s10518-022-01375-2.
[27] Ching J, Phoon KK. Value of geotechnical site investigation in reliability-based design. In: Advances in structural engineering; 2012. https://doi.org/10.1260/1369-4332.15.11.1935. 1935–45.
[28] Wilson ECF. A practical guide to value of information analysis. Pharmacoeconomics 2015;33:105–21. https://doi.org/10.1007/s40273-014-0219-x.
[29] CEN. Eurocode 2: design of concrete structures—Part 1-1: general rules and rules for buildings. London: British Standard Institution; 2004.
[30] CEN. Eurocode 8: design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings. Brussels: European Committee for Standardization; 2004.
[31] 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, Giardini D. The 2020 update of the European seismic hazard model: model overview. EFEHR Technical Report 001, v1.0.0. 2021.
[32] Tselentis GA, Danciu L. Probabilistic seismic hazard assessment in Greece-Part 1: engineering ground motion parameters. Hazards Earth Syst Sci 2010;10:25–39 [cited 2021 Nov 1].
[33] Bindi D, Massa M, Ameri G, Pacor F, Puglia R, Augliera P. 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. Bull Earthq Eng 2014;12. https://doi.org/10.1007/s10518-013-9525-5.
[34] Vamvatsikos D, Cornell CA. Incremental dynamic analysis. Earthq Eng Struct Dynam 2002;31(3):491–514. https://doi.org/10.1002/eqe.141.
[35] Kennedy RP. Performance-goal based (risk informed) approach for establishing the SSE site specific response spectrum for future nuclear power plants. In: Nuclear engineering and design; 2011. p. 648–56. https://doi.org/10.1016/j.nucengdes.2010.08.001.
[36] Gelman A, Carlin JB, Stern HS, D B. Bayesian data analysis. first ed. Chapman and Hall/CRC; 1995. p. 552. https://doi.org/10.1201/9780429258411.
[37] Mital U, Ahdi S, Herrick J, Iwahashi J, Savvaidis A, Yong A. A probabilistic framework to model distributions of VS30. Bull Seismol Soc Am 2021;111(4):1677–92. https://doi.org/10.1785/0120200281.
[38] Mori F, Mendicelli A, Moscatelli M, Romagnoli G, Peronace E, Naso G. A new Vs30 map for Italy based on the seismic microzonation dataset. Eng Geol 2020;275:105745. https://doi.org/10.1016/j.enggeo.2020.105745.
[39] Seyhan E, Stewart JP, Ancheta TD, Darragh RB, Graves RW. NGA-West2 site database. Earthq Spectra 2014;30(3):1007–24. https://doi.org/10.1193/062913EQS180M.
[40] Boore DM, Thompson EM, Cadet H. Regional correlations of V S30 and velocities averaged over depths less than and greater than 30 meters. Bull Seismol Soc Am 2011;101(6):3046–59. https://doi.org/10.1785/0120110071.

[41] Casella G. An introduction to empirical Bayes data analysis. Am Stat 1985;39(2): 83–7. https://doi.org/10.2307/2682801.

[42] Efron B. Large-scale inference: empirical Bayes methods for estimation, testing, and prediction [Internet]. Cambridge: Cambridge University Press; 2010. https://doi.org/10.1017/CBO9780511761362 [cited 2022 Jun 8].

[43] Robbins H. An empirical Bayes approach to statistics. University of California Press; 1956 Jan.

[44] Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis, vol. 69. London: Chapman and Hall; 1996.