Assessing the Value of Information in Site-Response Analysis

Haifa Tebib - University of Strathclyde, Glasgow, United Kingdom, email: haifa.tebib@strath.ac.uk

John Douglas –University of Strathclyde, Glasgow, United Kingdom, email: john.douglas@strath.ac.uk

Jennifer Roberts –University of Strathclyde, Glasgow, United Kingdom, email: jen.roberts@strath.ac.uk

Abstract: In seismic hazard assessments the importance of knowing different input parameters accurately depends on their weights 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 estimate the possible added value brought by acquiring new data along with 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.
In this article, we propose a method that combines available data and expert elicitation to facilitate the decision-making process within the site-response component of seismic hazard assessment. The approach combines Bayesian networks and decision trees to translate the causal-relationships between the input parameters in site-response analysis and Bayesian inference to update the model when new evidence is considered.
Here, we assess VoI for a hypothetical site. Our analysis shows that VoI is highly sensitive to prior probabilities and the accuracy of the test to be performed. This highlights the importance of defining those from available information as well as considering only tests that are suitable for our needs and budget.
Keywords: VoI, data collection, seismic hazard assessment, Bayesian updating, decision trees, site-response analysis

1.Introduction

The safety of infrastructure including nuclear facilities must be guaranteed against external hazards, including earthquakes. There is a need for various types of data to assess the seismic hazard at a particular location accurately in order to design structures to withstand that level of hazard. Collection of new data and methods for the calculation of ground motion can be associated with considerable uncertainties but the collection of such data can decrease this uncertainty. New data collection, however, can be costly as well as time consuming.

There is a pressing need to estimate the value of a piece of information (VoI) when it comes to considering the collection of data to enhance a seismic hazard assessment (SHA), in particular. In SHAs, multiple parameters need to be known to an accuracy that depends on their importance in the assessment. Assessments can be less sensitive to one parameter and very sensitive to others. In other words, uncertainties in a variable might lead to high variation in the results (hazard curves, uniform hazard spectra, and eventually the seismic design). Thus, it is important to know to what extent we should seek new information (data) to include in the seismic hazard analysis as it can sometimes lower the design requirements and hence, lower the construction costs of new facilities or the retrofit of existing ones.

To estimate the VoI, it is crucial to develop a methodology that takes into account the causal-relationships and the dependencies between parameters as well as the probabilities (beliefs of belief) assigned to each of them. When collecting more information, the method undertaken should be able to allow the updating of these probabilities according to new evidence. Moreover, the method should be able to answer the question "Is this parameter worth investigating?”. This means that the consequences of decisions should be included to the framework as well as the monetary (and time) cost that the decision would imply.

The overall aim of this study is to develop a method to assess the value of information of important parameters in SHA using various case studies and to suggest a framework that uses VoI as a measure to assess the feasibility and benefits of collecting more information. This article presents a simple application of this method, specifically for the case of a simple site-response analysis.

VoI is a helpful tool to assess the importance of a parameter or a geophysical/geotechnical test as well as the maximum investments (in terms of money, time and resources) one should be willing to spend to obtain a piece of information. This guidance could be critical for seismic hazard analysts and/or facilities owners (clients) where budgets are balanced with safety requirements.

2. Definition and applications of VoI

2.1 Definition

The book of Raffai & Schlaifer (1970) pioneered the use of VoI and provides an introduction to the mathematical analysis of decision making in a world of uncertainty. Knowing that, generally, more information leads to a reduction in uncertainty, the question here is whether a decision should be made on current information or whether to invest in additional information by considering its potential impact on the payoff that, as a result, can lead to revisiting the decision. The VoI is a key tool to improving research prioritization and to the collection of more information to reduce uncertainty and make more accurate and less uncertain decisions.

VoI can be considered the amount that you 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 of research, the decision is often made from the information and measures available and, in case of uncertainties, the decision making relies on a subjective approach that constitutes a consensus among experts, decision makers and even stakeholders. VoI is used to tackle this uncertainty as it helps to illuminate the importance of understanding the uncertainty and taking it into account to substantially decrease the overall uncertainty.

2.2 Types of VoI measurements

How we define a VoI model is different in each field and each application. A utility function has to be defined as well as its unit. This depends on the stakeholders or the decision makers interests. This function will either help compute avoided losses, which is often used in the healthcare field and the evaluation of losses due to external hazards (Williams, Gardoni, & Bracci, 2009) or compute a maximization of gains, which is mainly considered in marketing and economics. The unit of this function can be monetary (dollars, euros etc.), representing profit or revenue, or it can be more abstract like welfare or happiness. Many more characteristics are important for VoI calculations. Some of these are the number of alternative decisions, the type and number of parameters considered and their type of uncertainty, e.g. probability values or functions.

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 simply represents the reduction in risk after considering extra information. When it comes to a perfect information that eliminates completely the 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). If EVPI has a value that is less than the cost of the obtained information, obtaining that information should be rejected because even when the information completely eliminates the uncertainty and leads to making the less risky choice, it is not worthwhile in terms of the unit considered (e.g. financial cost).

Acknowledging that in this field of research and especially in SHA, perfect information does not exist and that whatever the measurements, uncertainties will always remain, we will likely not consider EVPI but instead the Expected Value of Imperfect Information (EVII).

3. Assessing the value of information: The must-have

Assessing VoI has advantages in the optimisation of time and money that one is willing to spend. In the field of earth sciences and civil engineering, even if VoI is not always referred to as VoI, we find some attempts in estimating the benefit of a piece of information or a design. VoI has been used to assess risk and reliability of retrofitting structures (Williams et al., 2009), designing site investigations (Gilbert & Habibi, 2015) and making drilling decisions (Eidsvik et al., 2015).

In order to consistently assess the VoI, there are the following three must-have to express the relationships and dependencies between the various variables.

- Conditional probabilities

Conditional probabilities are important for VoI analysis as conditioning an observation from information could lead to improvement in decision making. It is essential, within the framework to be built, to express the dependencies between the variables.

- Graphical models

To understand the degree to which variables are linked, connected and influenced by each other, graphical models are a powerful tool to serve this purpose. Such graphical models are referred to as Bayesian networks (BNs), Bayes nets or belief/decision trees. By using conditional probability density functions within the statistical model, the evidence regarding a parameter will be propagated to other nodes.

- 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 "imperfect/partial" mainly because of noise or because it represents only one variable of a multivariate set. For example, in seismic hazard assessment, 1D shearwaves velocity (Vs) profile could be the measure of interest x.

We would like to compute the posterior model for x conditioned on y $p(x|y)$. 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:

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

The latter can be viewed as a trade-off between prior knowledge and information brought by the data. Bayes' rule is used to construct posterior distributions that are essential for VoI calculations.

4. Application of VoI for a hypothetical case study

Seismic hazard assessment is an essential step to define an appropriate seismic design for structures. Seismic design is mandatory in earthquake-prone regions to ensure sufficient levels of safety and to prevent significant damage and collapse. The higher the seismic design, the costlier it is to implement. Seismic design must serve two potentially conflicting purposes: safety and economics. Thus, there should be a trade-off between constructions costs and the acceptable target levels of safety. Risk-targeted and minimum-cost design procedures has proved to be effective in satisfying this compromise (Gkimprixis et al., 2020).

In this section, we define a case study to build a methodology for VoI and proceed to simple calculations to evaluate the VoI of a parameter used in site-response analysis and hence to infer an appropriate seismic design.

4.1 Methodology

We aim to determine an appropriate seismic design for a reinforced concrete four-storey building. We face a lack of information about one specific parameter, Vs_{30}, i.e. the time-averaged shear-wave velocity in the top 30m. The dilemma here is between two major options:

1- Choose a particular seismic design and take the risk of choosing a higher, and more costly, seismic design than needed, or choose a lower and less-resistant seismic design where the seismic hazard and the soil response at the site of the building can result in damage or even total collapse.
2- Conduct geophysical/geotechnical tests in order to decrease the uncertainties on Vs_{30}. This will reduce the risk of choosing an inappropriate seismic design. Nevertheless, the test will have a cost that depends on the test type, the company hired to perform the test and the price of renting the equipment.

In this specific case, Vs_{30} is likely to have one of two possible values: V_1 or V_2. Available data and experts' 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. More realistically, a probability distribution covering a range of Vs_{30} should be used - this is the object of ongoing work.

4.1.1 Inputs and parameters

The purpose of the 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 amplification factor of the site. This PGA on soil is used to choose the appropriate seismic design.

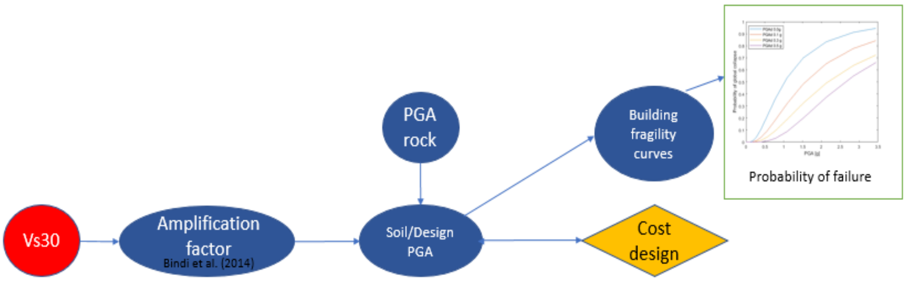

Fig.1 - Framework for site-response analysis and the estimation of the expected costs and losses. Blue circles represent non-uncertain parameters and the red circle (Vs30) is the uncertain parameter. The lozenge yellow node is the outcome (estimated cost and losses)

Fig.1 shows an influence diagram that summarises the parameters that are computed and/or used to infer the appropriate PGA to seismically design the building.

In this simple case study, site-response analysis is simplified by neglecting the non-linear site amplification and by computing the frequency-independent amplification factor F_s of the site using Bindi *et al.*, (2014) for PGA as follow:

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

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

Prior probabilities

The prior probability, *p*, is the probability that V_1 is believed to be the Vs_{30} of the site. Similarly, *1-p* is the probability that V_2 is believed to be Vs_{30}. These probabilities are often inferred by the seismic analyst from available data or expert knowledge.

Probability of failure

The work of Gkimprixis *et al.*, (2020) provides the fragility curves for the defined building for different PGA_d (PGA to which the building is designed). These fragility curves were derived using Incremental Dynamic Analysis (IDAs) (Vamvatsikos & Cornell, 2002). These fragility curves, *fc,* indicate the probability of total collapse for each PGA value.

Outcomes: Expected Losses

As the VoI here is in monetary units, we need to estimate the losses associated to seismically designing the building for a specific PGA_d. The losses due to possible earthquakes are a function of the seismic hazard in the site location and the vulnerability of the structural and non-structural components. The decision's outcomes are the expected consequences for each of the possible decisions. These expected consequences are defined to be the Expected Life-Cycle Losses, E[LCC] in case of total collapse.

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

$$E\big(o(x,d)\big) = C_0 + \sum_x o(x,d)p(x) \tag{5}$$

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:

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

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

4.1.2 Calculations

Decision trees are an excellent way of representing the various important parameters and their causal dependency.

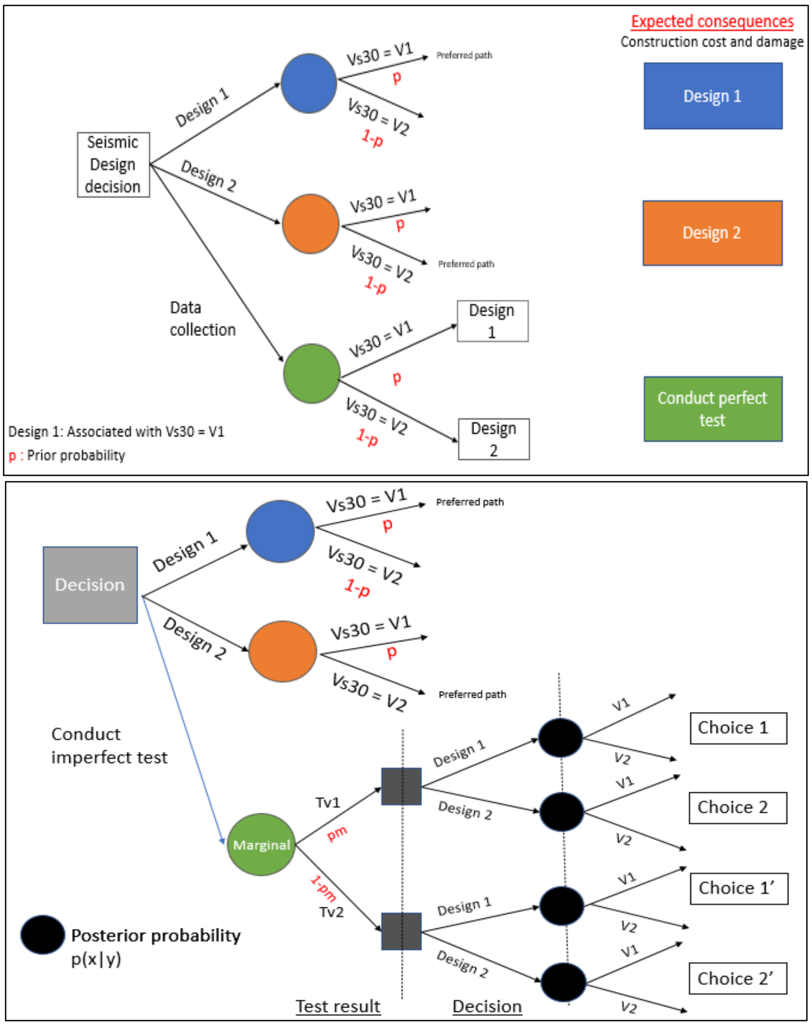

Fig.2 - Decision tree for the computation of EVPI (*up*) and EVII (*bottom*). Probabilities are displayed in red. pm is the marginal probability of test result being V_1.

In the decision trees presented in Fig.2, decisions emerge from the square nodes and probabilities from circle nodes. These two decisions trees depict three possible decisions:

1. Applying seismic design 1 (associated with a $Vs_{30}=V_1$) with available data.
2. Applying seismic design 2 (associated with a $Vs_{30}=V_2$) with available data.

3. Conduct a perfect test (left)/imperfect (right) test in order to have information about Vs_{30}.

Prior Value

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

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

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

EVPI

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

$$PoV = \sum_x E(o(x,d_x))p(x) \quad (8)$$

Where d_x is the appropriate design decision for state x.

EVII

Realistically, geotechnical or geophysical information cannot be completely free of uncertainties, i.e. perfectly accurate. Because most surveys would need analysis and interpretation to infer the measure of interest, results are likely to have dispersion, characterised, for example, by a normal distribution with a given standard deviation. The test being imperfect, there is the need to include this uncertainty 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 expresses its reliability. If *y* is data 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 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 (3) to perform Bayesian updating.

Therefore, the posterior value (PoV) is expressed as follows:

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

Where p(y) is the marginal probability.

Value of Information

The value of information is simply the difference between the *posterior value* and the prior value.

$$EVPI/EVII = PoV - PV \quad (10)$$

Irrespective of the type of value of information to compute, 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.

4.2 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 need to be performed and key values computed to validate the method.

The studied building is assumed to be located in the city of Patras, Greece. The hazard curves associated with this location were used to retrieve the expected losses and the PGA on a reference rock (PGAr). The PGAr is fixed to 0.43g as well as the Vs_{30} couple that translates our uncertainty and we proceed to the variation of the prior probability. V_1 is fixed to 100 m/s and V_2 to 500 m/s in Fig.3. The likelihood probability in case of imperfect information is set to 70%.

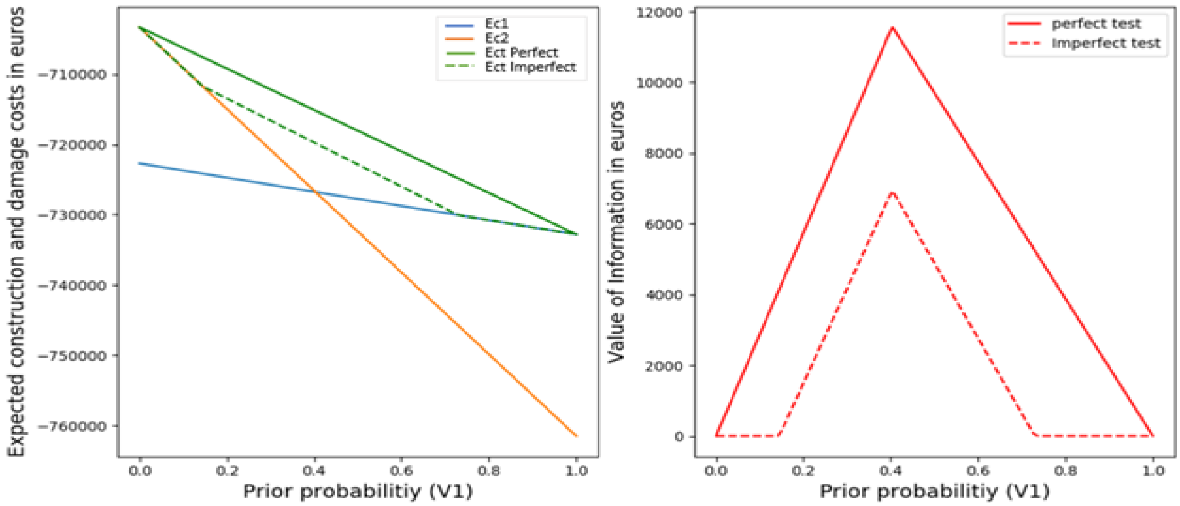

Fig.3 - Sensitivity to prior probabilities for the Expected outcomes of the three main decisions in the decision tree (left) and EVPI (solid line), EVII (dashed line) for the couple [100,500] m/s (right)

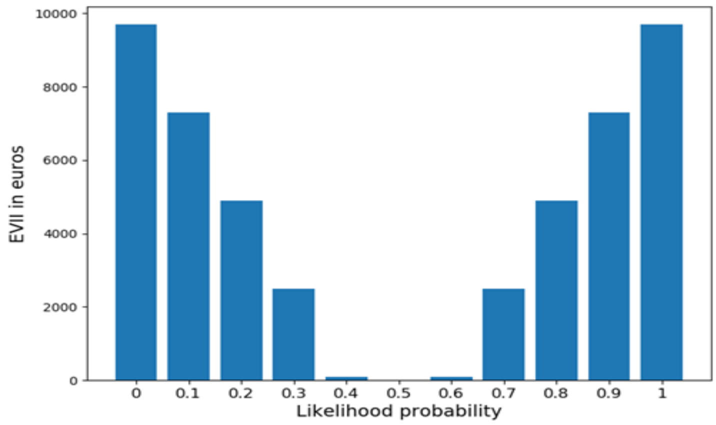

Fig.4 - Sensitivity of EVII to likelihood probability for fixed prior probability (0.5)

The left-hand graph in Fig.3 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.2. 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. The right-hand graph represents the VoI for several prior probabilities. Having an imperfect information reduces the VoI independently of the prior probabilities.

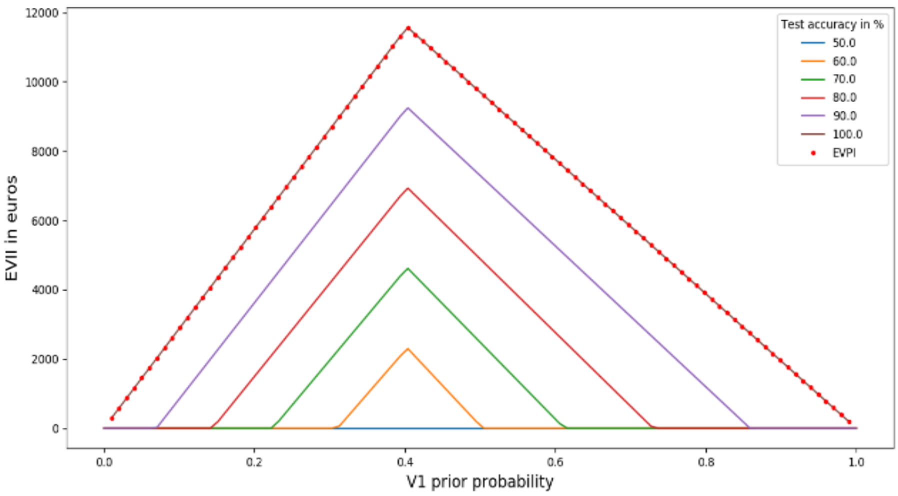

Fig.5- Sensitivity of EVII to prior probability for different likelihoods and comparison with EVPI

Fig.4 displays the EVII for fixed prior probabilities and different likelihoods. There is no EVII when the test accuracy is 50%. A test with 50% accuracy is of no help as there is 50% chance the test being right (or 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 seems to have a linear behaviour. This linearity is further confirmed when visualising, in Fig.5, the EVII for several prior probabilities and different likelihoods where we can also verify that EVII for a test accuracy of 100% is equal to the EVPI. This is an additional validation of the method. Moreover, the less accurate the test, the smaller the range of prior probabilities given to V_1 and V_2. 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 likelihoods, there is no value of information when the prior probability is equal 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.

5.Conclusion

In SHA and earth sciences in general, the EVII is more likely to be used as it expresses the uncertainties regarding the tests conducted to infer information. The developed method has been 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 available data and experts' knowledge is not worth conducting. Moreover, the more accurate the test the higher the benefits of obtaining the information. Another important finding is that for site rock conditions, uncertainties regarding the Vs_{30} have little influence on the EVPI and it might not be worthwhile to conduct tests to obtain more information. Calculations using extreme and specific values have enabled a validation of the presented method.

These sensitivity analyses highlight the parameters and inputs that influence most the VoI. This stresses the fact that some inputs should be estimated and chosen carefully in order to have a high reliability in the VoI estimation. The prior probability given to the possible values of the uncertain parameter Vs_{30} was shown to highly influence the EVI. Thus, it is important to wisely and thoroughly use available information to infer the prior probabilities. The estimation can be given, for example, through experts from past experiences as well as through Empirical Bayes Estimation, which has proven to provide a good estimate of these probabilities in other fields (Carlin & Louis, 1996; Casella, 1985; Efron, 2010; Robbins, 1956).

We recall that the binary uncertainty assumed for the distinction of interest Vs_{30} is not realistic, especially when V_1 and V_2 have a large gap. Uncertainties for a parameter like Vs_{30} should be expressed continuously and ideally through probability distributions. Methods in the literature have proved to infer probability distributions from available data (e.g., Empirical Bayes Theory). An ongoing stage of the research aims at developing the method to express uncertainties with probability distributions. Nevertheless, this method is applicable to other case studies where choices or decisions can be binary. For example, this simplified method can be used when there is uncertainty about a fault being active or inactive (i.e. it does not generate large earthquakes). In this case, this method can be a great tool in justifying in-situ or satellite remote sensing data collection as well as estimating the maximum worthwhile investment in time and money.

References

- Bindi, D., Massa, M., Ameri, G., Pacor, F., Puglia, R., & 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*. https://doi.org/10.1007/s10518-013-9525-5
- Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis* (Vol. 69). Chapman and Hall.
- Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. In *Source: The American Statistician* (Vol. 39, Issue 2).
- 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*. https://doi.org/10.1017/CBO9780511761362
- Eidsvik*, J., Mukerji, T., Bhattacharjya, D., & Dutta, G. (2015). *Value of Information Analysis of Geophysical Data for Drilling Decisions*. cp-456-00038. https://doi.org/10.3997/2214-4609.201413621
- Gilbert, R. B., & Habibi, M. (2015). Assessing the value of information to design site investigation and construction quality assurance programs. In K.-K. Phoon & J. Ching (Eds.), *Risk and Reliability in Geotechnical Engineering* (1st Editio, pp. 491–532). CRC Press Taylor and Francis Group.
- Gkimprixis, A., Tubaldi, E., & Douglas, J. (2020). Evaluating alternative approaches for the seismic design of structures. *Bulletin of Earthquake Engineering*, *18*(9), 4331–4361. https://doi.org/10.1007/s10518-020-00858-4
- Raffai, H., & Schlaifer, R. (1970). *Applied Statistical Decision Theory*. M.I.T Press.
- Robbins, H. (1956). AN EMPIRICAL BAYES APPROACH TO STATISTICS. In *https://doi.org/*. University of California Press.
- Vamvatsikos, D., & Cornell, C. A. (2002). Incremental dynamic analysis. *Earthquake Engineering & Structural Dynamics*, *31*(3), 491–514. https://doi.org/https://doi.org/10.1002/eqe.141
- Williams, R. J., Gardoni, P., & Bracci, J. M. (2009). Decision analysis for seismic retrofit of structures. *Structural Safety*, *31*(2), 188–196. https://doi.org/10.1016/j.strusafe.2008.06.017