



**D 3.1**

**DELIVERABLE**

**PROJECT INFORMATION**

Project Title: **Harmonized approach to stress tests for critical infrastructures against natural hazards**

Acronym: **STREST**

Project N°: 603389

Call N°: FP7-ENV-2013-two-stage

Project start: 01 October 2013

Duration: 36 months

**DELIVERABLE INFORMATION**

Deliverable Title: **Report on the effects of epistemic uncertainties on the definition of LP-HC events**

Date of issue: 31 March 2015  
updated 01 October 2015

Work Package: WP3 – Integrated low probability-high consequence hazard assessment for critical infrastructures

Editor/Author: Jacopo Selva  
(INGV)

Reviewer: Fabrice Cotton  
(UJF/ GFZ)

REVISION: Version 3



Project Coordinator: Prof. Domenico Giardini  
Institution: ETH Zürich  
e-mail: [gardini@sed.ethz.ch](mailto:gardini@sed.ethz.ch)  
fax: + 41 446331065  
telephone: + 41 446332610



---

# Abstract

Stress Tests (STs) are performed for evaluating the resilience of Critical Infrastructures (CIs) to potential perturbations, in STREST limited to the ones coming from natural hazards. Significant unwanted consequences due to damages in a CI mainly occur because of design deficit, or of the occurrence of “unlikely” events, that is, events with a low probability of occurrence. Given that few data are available on rare events, the analysis of such events requires a full exploration of epistemic uncertainty. Here, we develop a coherent process (EU@STREST, Epistemic Uncertainty in STREST) to ensure a robust management of epistemic uncertainty within a project aimed to perform a ST. EU@STREST follows the state-of-art methodological and procedural guidance obtained (i) from ENSREG (European Nuclear Safety Regulations Group) and IAEA (International Atomic Energy Agency), and (ii) the critical review of existing methods for quantifying epistemic uncertainty. EU@STREST is designed to ensure an improved, standardized and robust treatment of the epistemic uncertainty emerging from hazard and hazardous phenomena selection, alternative models implementation, and exploration of the tails of distributions, taking into account the diverse range of views and opinions of experts, but also the potential limitations in budget for STs for non-nuclear CIs and their potential regulatory impact.

*Keywords: Epistemic uncertainty, multiple-expert integration, EU@STREST process*



---

## **Acknowledgments**

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 603389.



---

## Deliverable Contributors

INGV<sup>†</sup> Jacopo Selva

Sarfraz M. Iqbal

Matteo Taroni

Warner Marzocchi

GFZ/UJF<sup>‡</sup> Fabrice Cotton

TNO<sup>§</sup> Wim Courage

Linda Abspoel-Bukman

Simona Miraglia

ETH Zurich<sup>||</sup> Arnaud Mignan

AUTH<sup>⊥</sup> Kyriazis Pitilakis

Sotirios Argyroudis

Kalliopi Kakderi

Dimitris Pitilakis

Grigoris Tsinidis

Chiara Smerzini

<sup>†</sup> National Institute of Geophysics and Volcanology, Italy.

<sup>‡</sup> Helmholtz Centre Potsdam, Germany/Joseph Fourier University, France

<sup>§</sup> Netherlands Organisation for Applied Scientific Research, the Netherlands.

<sup>||</sup> Swiss Federal Institute of Technology in Zurich, Switzerland.

<sup>⊥</sup> Aristotle University of Thessaloniki, Greece.



---

# Table of Contents

<b>1</b>	<b>Introduction</b> .....	<b>15</b>
1.1	ALEATORY / EPISTEMIC / ONTOLOGICAL UNCERTAINTY.....	15
1.2	GOALS & ORGANIZATION OF THE DELIVERABLE.....	19
<b>2</b>	<b>A critical review of existing methods</b> .....	<b>21</b>
2.1	INTRODUCTION.....	21
2.2	LOGIC TREES ( <i>LT</i> ).....	23
2.3	BAYESIAN/ENSEMBLE APPROACH (BEA).....	30
2.4	CLASSICAL EXPERT ELICITATION ( <i>CEE</i> ) .....	37
2.5	MULTIPLE EXPERT INTEGRATION ( <i>MEI</i> ).....	45
2.6	CRITICAL COMPARISON OF APPROACHES .....	49
<b>3</b>	<b>Selection of hazards and hazardous phenomena</b> .....	<b>54</b>
3.1	INTRODUCTION.....	54
3.2	EXISTING GUIDELINES .....	55
3.3	MULTI-HAZARD AND MULTI-RISK ASSESSMENT METHOD.....	58
3.4	METHODS TO PRIORITIZE NATURAL HAZARDS .....	62
<b>4</b>	<b>Guidelines for treating epistemic uncertainties in Stress Tests (ST) for non-nuclear CIs</b> .....	<b>66</b>
4.1	INTRODUCTION.....	66
4.2	METHOD SELECTION FOR STRESS TESTS.....	67
4.2.1	Methodological requirements.....	68
4.2.2	Exploration of tails in PHA and PRA.....	74
4.3	EU@STREST: MULTIPLE EXPERT INTEGRATION FOR STRESS TESTS OF NON-NUCLEAR CRITICAL INFRASTRUCTURES .....	79
4.3.1	EU@STREST vs SSHAC process .....	80
4.3.2	Core actors.....	81
4.3.3	Key features of the process .....	84
4.3.4	Project design .....	84
4.3.5	Project's PHASE 0: Preparation .....	86
4.3.6	Project's PHASE 1: Initialization .....	87
4.3.7	Project's PHASE 2: Evaluation and Integration .....	89
4.3.8	Project's PHASE 3: Finalization.....	91

---

<b>5</b>	<b>Case studies.....</b>	<b>92</b>
5.1	INTRODUCTION.....	92
5.2	CASE STUDIES.....	92
5.2.1	Sensitivity of induced seismicity ground motion accounting for regional conditions (CI-B2).....	92
5.2.2	Sensitivity of $M_{max}$ accounting for fault interactions (CI-B1).....	95
5.2.3	Sensitivity of performance assessment for urban infrastructure and harbours to spatial correlations in the hazard .....	97

---

## List of Figures

Figure 2-1: Methods for the analysis of the epistemic variability based on multiple-experts opinion. ....	22
Figure 2-2: LT method - Example of a simplified logic tree for three model parameters (from Geist and Parsons 2006). ....	24
Figure 2-3: LT method - Graphical representation of the MECE criterion (from Bommer and Scherbaum, 2008). ....	26
Figure 2-4: SHARE logic tree. Colours depict the branching levels for the earthquake source models (yellow), maximum magnitude models (green) and ground motion models (red). Values below the black lines indicate the weights, for the source model these indicate the weighting scheme for the different return periods. Tectonic regionalization is not branching levels (grey) of the model; however, it defines the GMPEs to be used (from Woessner et al., 2013, Giardini et al. 2013). ....	28
Figure 2-5: Probabilistic Seismic Hazard Analysis (PSHA) results of the project SHARE (Woessner et al., 2013, Giardini et al. 2013). ....	29
Figure 2-6: BEA method - Probability distribution of probability for a simultaneous treatment of aleatory and epistemic uncertainties in probabilistic hazard curves. ....	31
Figure 2-7: Probability of exceedance of 5m in different locations in Southern Italy, as evaluated for a preliminary assessment of PTHA (Selva et al. 2015). ....	32
Figure 2-8: Bayesian/Ensemble Approach (BEA) - Graphical representation of the inference process of BEA. ....	33
Figure 2-9: Fit of the prior model probability in a representative location: (a) the starting dataset, (b) the elaboration of input data, and (c) the obtained prior probabilities (from Selva and Sandri, 2013) ....	35
Figure 2-10: (a) Hazard curve (in 50 years) for the city of Bologna, from prior and posterior models. (b) Same as (a), but for the city of Rimini (from Selva and Sandri, 2013) ....	36
Figure 2-11: (a) Hazard map (10% in 50 years) obtained by the medians of posterior model on sites in g units. (b) The equivalent number of data (inversely proportional to the epistemic variance) , relative to the PGA value $1/0.1g$ , for the posterior model. (c) Same as (a), with past data only for sites with complete historical record (from Selva and Sandri 2013). ....	36
Figure 2-12: illustration on calibration and information for cEE. ....	43
Figure 2-13: An example of testing two hazard models with the same mean and different epistemic uncertainty against an observed frequency of exceedance. ....	53

---

Figure 3-1: Diagram showing the possible scenarios of cascading events among the hazards considered in the MARTIX project (Garcia-Aristizabal et al. n.d.).	59
Figure 3-2: Schematic view of the steps followed in the framework multi-risk assessment (Liu et al., submitted).	60
Figure 3-3: An example of hierarchy with three alternatives and four criteria for a specific goal (from Wikipedia).	63
Figure 4-1: Examples of disaggregation reporting alternative models to describe epistemic uncertainty, in the case of LT (left panel, from Lin and Baker, 2011) and of BSA (right panel).	69
Figure 4-2: Risk migration matrices showing how risk migrates as a function of frequency and aggregated losses when new information is added to the system. Risk increase in a given frequency-loss domain is represented in red and risk decrease in blue. Risk scenarios are represented by points, in white for the first hypothesis and in black for the second, taken from Mignan et al. (2014).	72
Figure 4-3: 200 exceedance probabilities relative to the average PGA for 2% in 50 years for Los Angeles, coming from one logic tree developed in the framework of UCERF3 (Field et al., 2014), as compared with an equivalent Ensemble model (from Marzocchi et al. 2015).	75
Figure 4-4: PSHA obtained by a simple logic tree composed by 5 different seismicity rate models arbitrarily selected from the Italian CSEP experiment (Schorlemmer et al., 2010), and three GMPEs, as compared with ensemble model (from Marzocchi et al., 2015).	76
Figure 4-5: 200 exceedance probabilities relative to the average PGA for 2% in 50 years for Redding, coming from one logic tree developed in the framework of UCERF3 (Field et al., 2014), as compared with two potential Ensemble models (from Marzocchi et al. 2015).	77
Figure 4-6: The basic interactions among the core actors in the process of EU@STREST.	83
Figure 4-7: Graphical representation of the proposed EU@STREST process, showing the different phases and the interactions between the four main actors in each phase.	85
Figure 5-1: Ground motion due to an $M_w = 4$ , $h = 3$ km deep, induced earthquake: a. Logic tree; b. PGV versus epicentral distance	93
Figure 5-2: Length- $M_{max}$ relationships: a. ESHM13; b. Including cascades (modified from Mignan et al., in press)	96
Figure 5-3: Proposed $M_{max}$ logic tree for CI-B1 (Strike-slip faults in Anatolian Peninsula)	97
Figure 5-4: Seismic zones, epicenter of M6.4, 1978 event (red star) and study area (red rectangle).	100
Figure 5-5: MAF curves for performance loss (EPN, WSS, RDN, HBR) with and without considering spatial correlations of intensity measures (1000 runs, one seismic scenario, M=6.5).	104

---

---

Figure 5-6: MAF curves for performance loss (EPN, WSS, RDN, HBR) with and without considering spatial correlations of intensity measures (10000 runs, 5 seismic zones, M=5.5-7.5)..... 105

Figure 5-7: Correlation of components' functionality to system PI with and without considering spatial correlations of intensity measures (up: 1 seismic scenario, down: 5 seismic zones)..... 106

Figure 5-8: Disaggregation to magnitudes and seismic zones for given PI exceedance (5% for TCoH, 20% for ECL), with and without considering spatial correlations of intensity measures..... 107



---

## List of Tables

Table 2-1: Attributes of various SSHAC levels. ....	48
Table 3-1: Matrix of all possible interactions among the hazards considered in MATRIX project (Garcia-Aristizabal et al. n.d.). ....	59
Table 4-1: Analytical description of actions along the Phase 0. ....	86
Table 4-2: Analytical description of actions along the PHASE 1. According to Figure 4.7, the phases of interaction between TI and the IR are highlighted in red, while the ones between TI and the PoE are highlighted in green. ....	88
Table 4-3: Analytical description of actions along the PHASE 2. According to Figure 4-6, the phases of interaction between TI and the IR are highlighted in red, while the ones between TI and the PoE are highlighted in green. ....	90
Table 4-4: Analytical description of actions along the PHASE 3. According to Figure 4-6, the phases of interaction between TI and the IR are highlighted in red. ....	91



# 1 Introduction

## 1.1 ALEATORY / EPISTEMIC / ONTOLOGICAL UNCERTAINTY

Hazard and risk assessment deals with the problem of assessing the probability of exceedance of different levels of intensity for one phenomenon (hazard) and for consequent potential losses (risk) within a time window (exposure time) in a target area. For natural hazards, we deal with systems with many degrees of freedom, often characterized by non-linear dependencies, thus the task of quantitative hazard and risk assessment being particularly challenging. Because of these complexities, many sources of uncertainty exist. To deal with such uncertainties, hazard and risk assessment are typically based on probabilistic formulations. Given that natural systems are not even completely known, different potential implementations of such probabilistic formulations are possible, leading to additional uncertainties.

Hazard practitioners often distinguish uncertainties among different kinds, commonly adopting the terms aleatory and epistemic uncertainties. In particular, they adopt the term aleatory uncertainty to describe the intrinsic natural randomness of the hazard intensity (e.g., peak ground acceleration due to earthquakes, wave height due to tsunamis, loading of tefra deposit due to volcanic eruptions, etc.). This uncertainty is a characteristic of the natural system, and it is irreducible. On the other hand, they use the term epistemic uncertainty to characterize all uncertainties due to our limited knowledge about the true model describing the aleatory variability. This uncertainty depends on the limited knowledge of the modellers on the systems, and it is reducible. There are essentially two main reasons supporting such a separation. On the one hand, it allows isolating the uncertainty that may potentially be reduced by new research activity (the epistemic uncertainty). This potentially allows a better targeting of potential future research (e.g., Marzocchi et al. 2008), since the distinction between epistemic and aleatory uncertainties has the goal of separating the uncertainty due to limitations of scientific knowledge on the studied system from the one due to the intrinsic unpredictability of system itself, which is not reducible. On the other hand, to keep the epistemic uncertainty separated allows communicating to the end users the effective variability of the results, which actually depends on subjective choices of practitioners (e.g., Paté-Cornell 1996, SSHAC 1997), providing results in a more robust format when they are targeted to regulatory concerns.

Even if the distinction between aleatory and epistemic uncertainties is theoretically clear, in common practice it becomes sometimes rather subjective. For example, different practitioners may have different feelings about the variability of one parameter, and some others may think that the parameters' variability is completely natural and it would remain if the system would be perfectly known (for example, the variability of the outcome of one dice, *alea* in Latin), conversely others may think that it is only a matter of knowing better the system. For this reason, SSHAC (1997) proposed a more practical definition of aleatory and epistemic uncertainties. They proposed to call *epistemic* the variability of results that emerges when different scientifically acceptable procedures are adopted in assessing the hazard. In other words, different model proponents may have different sensibility in defining the aleatory uncertainty in their models, and the epistemic variability is defined *ex post* through the variability among such models. Following this definition, the assessment of epistemic uncertainty becomes the quantification of "*the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study*" (SSHAC 1997). This variability may be defined as inter-model variability (Selva et al. 2012) and it is certainly practically convenient, but clearly deviates from the original definition of aleatory and epistemic uncertainties, since while inter-model variability is clearly epistemic in nature, the remaining uncertainty may combine uncertainties related to natural variability and to limited knowledge on the system.

Even if the distinction between aleatory and epistemic uncertainties is both popular and useful in practice, many authors and philosophers argue that this separation is ambiguous, and it does not have a theoretical significance, because, as far as our knowledge of the system may increase, all uncertainties become necessarily epistemic (e.g., NRC, 1997; Bedford and Cooke, 2001; Jaines, 2003). This different interpretation may look academic, but it implies a different treatment and quantification of uncertainty, and it has significant implications in the testability of hazard results. These two different lines of thinking about the distinction between aleatory and epistemic uncertainties directly emerge from the very meaning of probability, which has two main interpretations. For frequentists, a probability is the limiting frequency of a random event or the long-term propensity to produce such a limiting frequency (Popper 1983; Mayo, 1996); for Bayesians, it is a subjective degree of belief that a random event will occur (Lindley 2000). Among the other differences, we mention here the fact that the frequentist interpretation connects the probability definition to a measurable quantity that can be theoretically known by analysing an infinite sequence of outcomes for repeatable event. On the opposite, the subjective interpretation of probability represents a state of knowledge, and it can be defined also for non repeatable events (e.g, the election of one president, or the sun rising tomorrow). Advocates on both sides have

argued that degrees of belief cannot be measured and, by implication, cannot be rejected (de Finetti 1989; Mayo 1996; Jaynes 2003). In the words of one author (Vick 2002), “*the degree of belief probability is not a property of the event (experiment) but rather of the observer. There exists no uniquely true probability, only one’s true belief.*” Within the subjectivist Bayesian framework, one’s true belief can be informed, but not rejected, by experiment. On the opposite, frequentists tend to an objective definition of probabilities, with the goal of rejecting any subjectivity from its assessment.

Marzocchi and Jordan (2014) proposed an *unificationist* approach in which the use of subjective probability, such as expert opinion, poses no problems for ontological testability as long as one experimental concept defines sets of exchangeable data with long-run frequencies determined by the data-generating process. These frequencies, which characterize the aleatory variability, have epistemic uncertainty described by the experts’ distribution. Expert opinion is thus regarded as a measurement system that produces a model that can be tested. A striking example of this is reported in Marzocchi and Jordan (2014). Sir Francis Galton, during a tour of the English countryside, recorded 787 guesses of the weight of an ox made by a farming crowd and found that the average was correct to a single Pound (Galton 1907). The experts’ distribution he tabulated passes the appropriate Student’s test (retrospectively, since the test was invented in 1908). This example shows that the conceptual problem is not posed by the subjectivity of the assessment (the farming crowd made subjective estimations, as experts of the target assessment), but by the possibility of testing the estimations (the weight of the ox is potentially a measurable thing, therefore it is objective). This means that subjective expert opinion poses no problems as long as it exists an experimental concept that defines the data to test the subjective variability. The example also shows that subjectivity can be managed through a distribution describing where the true value may be. This distribution describes the subjectivity on the assessment, that is, the epistemic uncertainty.

This approach is particularly important to assure the scientific testability of hazard/risk results subject to epistemic uncertainty, providing testability of scientific results in the framework of a well-defined experimental concept. In general, the experimental concept defines collections of data, observed and not yet observed, that are judged to be exchangeable (Draper et al., 1993); for instance, when we decide the set of data that we want to describe and that will be used to test the model, we are implicitly defining an experimental concept. In probabilistic hazard assessments, a typical experimental concept consists of collecting data on the exceedance of reference hazard intensity levels during N equivalent time intervals in one specific site or region (Marzocchi et al, 2015). The time intervals are assumed to be

exchangeable and the observation is zero if no exceedance is observed in the time interval and one otherwise. In the experimental concept, the aleatory variability is described by the exceedance frequency of the exchangeable dataset, and the epistemic uncertainty is represented by the lack of knowledge of what is the right model to describe such a frequency, both in terms of functional mathematical formulation and/or of its parameters. Therefore, the definition of aleatory variability is not associated to the real physical process, but to the sequence of data that are judged to be exchangeable.

From this theoretical framework, it derives a clear and univocal taxonomy of uncertainties. The distinction between aleatory and epistemic uncertainties is not only of practical convenience, but it is of primary importance to make any probabilistic assessment a testable, and, consequently, scientific one (see Marzocchi and Jordan, 2014 for a discussion about the issue of testability in their approach and in traditional ones as, e.g., SSHAC, 1997). These uncertainties in any hazard or risk assessment can be organized into a three-level hierarchy, defining a clear taxonomy of the uncertainty involved. Aleatory uncertainty is quantified by the expected (long-run) frequencies of random events within the system model; such frequencies are thus objective probabilities. Hypotheses about aleatory variability can be tested against observations by frequentist (error-statistical) methods. Epistemic uncertainties measure lack of knowledge in the estimation of such frequencies; they are often based on expert opinion and are thus described by subjective probabilities. Ontological error is identified by the rejection of a null hypothesis, called the “ontological hypothesis” (Marzocchi and Jordan, 2014), which states that the true frequencies of the random events are samples from the (joint) probability distribution describing the epistemic uncertainties.

Note that the theoretical framework of Marzocchi and Jordan (2014) is not in contrast with the different interpretations described above, but more it provides a theoretical framework to interpret the differences among them. In particular, the use of frequencies to define aleatory variability is of paramount importance to make hazard models comparable with real data. Being frequencies measurable through the experimental concept, thus they are a property of the system, and not of the practitioners. The subjectivity in assessing such frequencies, due to lack of data and knowledge, is the epistemic uncertainty. As far as our knowledge of the system may increase, such frequencies do not change. What it may change is our capability of assessing the true value of such frequency (epistemic uncertainty reduction). Thus, the epistemic uncertainty again represents “*the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study*” (SSHAC 1997). Of course, it may potentially change also our experimental concept, but this means that it changes also the target of our assessment.

In this document, we make reference to this theoretical framework, and we focus on the quantification of epistemic uncertainty defined as the variability induced by subjective choices that practitioners necessarily do whenever they formulate a model to assess the aleatory uncertainty, both selecting a specific model (e.g., tapered Pareto vs truncated Pareto) and selecting a specific parameter for the selected model ( $b=1$  vs  $b=1.2$ ).

## 1.2 GOALS & ORGANIZATION OF THE DELIVERABLE

The goal of this deliverable is to define a general framework regarding the management and the quantification of epistemic uncertainty in Stress Test (ST) for non-nuclear Critical Infrastructures (CI). Hereinafter, we refer to this method as EU@STREST (Epistemic Uncertainty in STREST). The specific goal of EU@STREST is to define a robust framework for this assessment, in order to increase the reliability of ST results. In particular, we target to make sure that the ST results (for example, acceptance / not acceptance) do not depend on specific subjective choices of the practitioner performing the assessments. In other words, we want to avoid the results of the ST to be completely controlled by *a priori* choices made either by the management of the CI and/or by a restricted group of experts. The development of EU@STREST is here linked to the ST methodology of STREST, developed in WP5. However, EU@STREST is meant to be general and applicable to any ST for non-nuclear critical infrastructure.

Many choices potentially subjective are necessary in any ST, starting from the choice of the hazards to be modelled, going through the choices made to quantitatively assess the hazards, up to the selection of failures modes and performance indicators of the CI, the selection of the components CI and of models of intra- (within the CI) and inter (with other extern systems, like for example the ones that supply services, like electricity, to the CI) dependencies. All these choices may largely influence the results of any ST. For example, if one hazard is *a priori* neglected, its potential impact will be completely neglected. In order to avoid an *a priori* control of the results of the ST, it is required that a minimum level of involvement of multiple experts is guaranteed, both in setting up the methodological framework of the ST and in performing computations. Indeed, the quantification of epistemic uncertainties as defined above (*“the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study”*, SSHAC 1997) makes the assessment theoretically not dependent on the specific analyst, then more *objective*. However, the extended involvement of a large community of experts is very expensive, easily overcoming the limitations on the budget of STs for non-

nuclear CIs. This is the very final challenge of the deliverable: guarantee the minimum level of objectivity of the results of any ST, both sufficiently involving the larger community and accounting for limitations on the budget of STs.

Even if the analysis of epistemic uncertainty is often limited to hazard assessments, the variability on the results due to different and subjective choices related to the vulnerability assessments may potentially introduce non-predictable consequences in loss/risk assessments (e.g., Pate-Cornell 1996, Winkler 1996). However, besides being not widely accepted, this extension typically suffers of severe computational problems. For instance, a seismic hazard logic tree with many branches can be hardly used for risk calculations if we still want to honour the logic tree structure (Field et al., 2005; SYNER-G 2010-2013). Noteworthy, several new approaches to deal with the problem of the assessment and the propagation of epistemic uncertainty have been recently proposed (e.g., Selva and Sandri 2012, Marzocchi et al. 2013, ByMuR 2010-2013), enabling to overcome such computational problems. While the theoretical discussions about methods will be mainly based on the applications in hazard assessments, the extendibility of each procedure to risk assessment should be discussed case by case.

In order to reach these goals, we first review and compare the main methods that exist in literature to treat and quantify epistemic uncertainties (*section 2*). Then, we focus on the problem of selecting the phenomena that a ST should focus on, reviewing existing methods on this topic (*section 3*). In *section 4*, we summarize the results of *sections 2* and *3*, discussing the potential implications of different choices in uncertainty modelling, and formulating the EU@STREST process, to be followed to guarantee the minimum level of uncertainty treatment within a ST for non-nuclear Critical Infrastructures. In particular, EU@STREST process defines the participation and the interaction of different experts in the definition of methods to be used and in the quantification of epistemic uncertainty, making use of a combination of the methodologies discussed in the previous sections. Finally, we discuss several case studies mainly focused on preliminary sensitivity tests (*section 5*). Such kind of studies are of primary importance during any epistemic uncertainty quantification based on the analysis of the variability of different assumptions, in order to help focusing on the parts of the assessment that really matter, as well as allowing a controlled simplification of the methodological framework that does not lose its generality.

## 2 A critical review of existing methods

### 2.1 INTRODUCTION

When dealing with epistemic uncertainties in hazard and risk assessment, it may be distinguished into two main types: the one related to the values of parameters within statistical models; and the one related to the models themselves (e.g., Bommer and Scherbaum 2008, MATRIX 2010-2013; Rougier et al. 2013). While several methods have been proposed to treat parameter uncertainty (e.g., Kijko and Sellevoll, 1992, Straub and Der Kiureghian 2008, Koutsourelakis 2010, Franchin et al. 2013, Keller 2014), typically based on (hierarchical) Bayesian procedures, only very few methods have been proposed to treat both sources simultaneously. In particular, we here focus on the main methods that have been adopted in the assessment of natural hazards: the Logic Tree (*LT*), the Bayesian/Ensemble Approach (*BEA*), the classical Expert Elicitation (*cEE*), and the Multiple Expert Integration (*MEI*).

The main characteristic of all these methods is that they target to the exploration of the variability of the hazard results depending on the different, but scientifically acceptable, choices that different experts may take. In this sense, such methods are tools that enable treating alternative acceptable approaches within a multiple-expert analysis in order to quantify the variability among different technical interpretations, in order to quantify the epistemic variability. They mainly differ in how the different experts' opinions are treated and merged into a single model. In particular, *cEE* tries to extract the target information (e.g., a given probability of occurrence) directly from the experts through structured questionnaires, assuming that the target knowledge is present in (or may be hypothesized by) the pool of experts; *LT* and *BEA* take as input models that different experts have developed, and they merge them in order to provide an aggregated model; *MEI* gathers together the experts, asking them to discuss and formulate an agreed set of alternative models (project's models) that are then integrated based on expert's opinion through an integration process (in literature, typically based on *LT*). These different processes are graphically represented in **Error! Reference source not found.**

In the followings, we will first review such methods, and then we critically compare them, in order to highlight main commonalities and main differences.

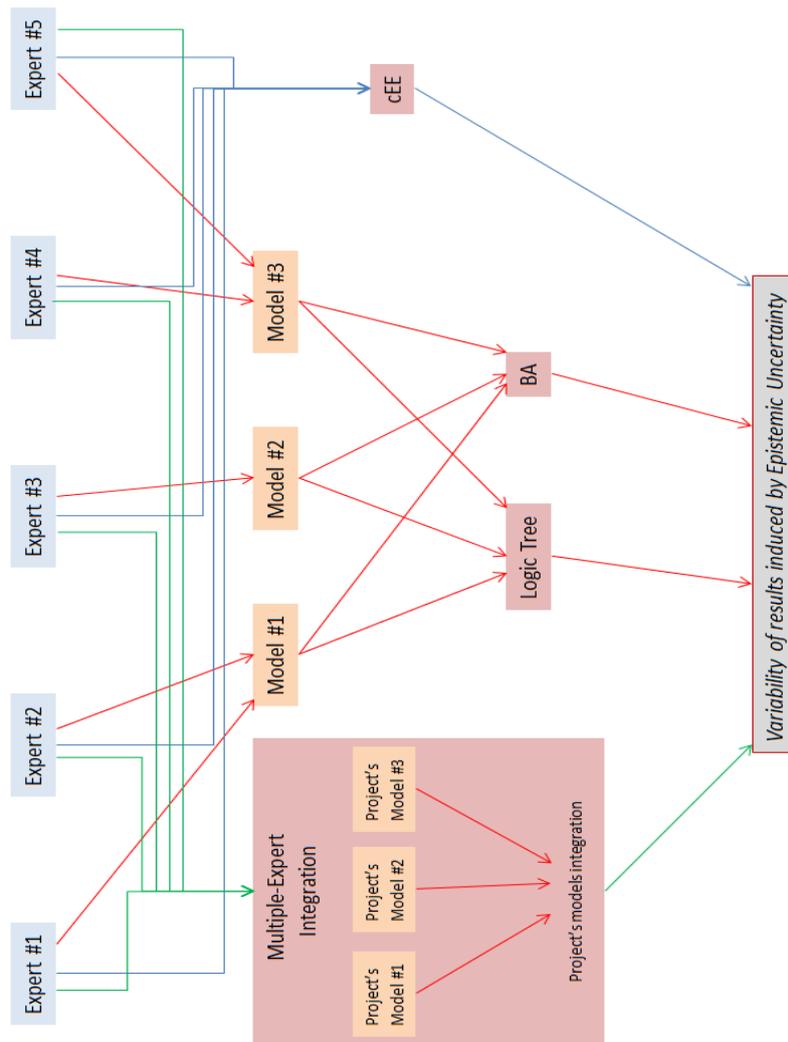
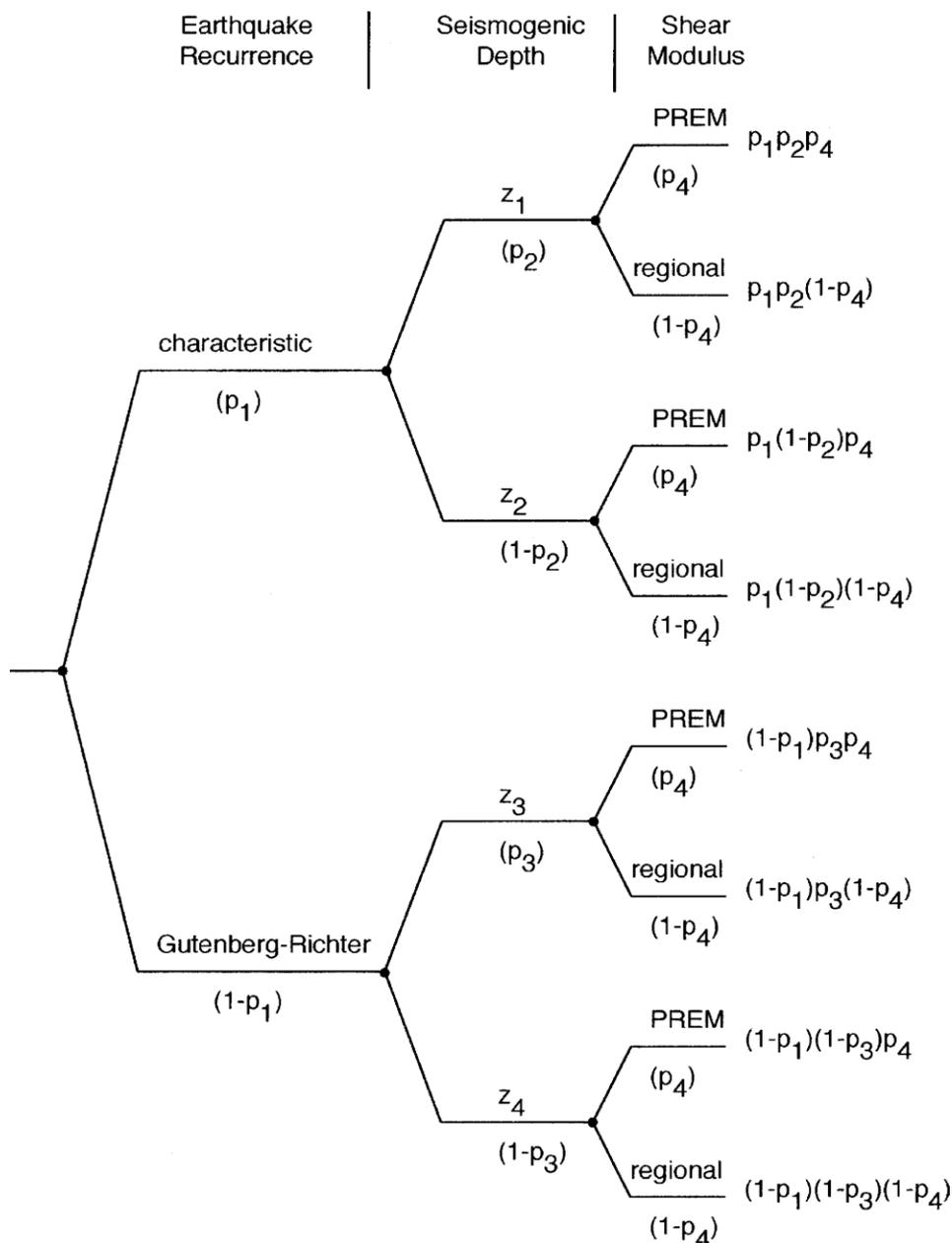


Figure 2-1: Methods for the analysis of the epistemic variability based on multiple-experts opinion.

## 2.2 LOGIC TREES (LT)

The use of the logic-tree framework has almost become standard practice in probabilistic seismic hazard analysis (PSHA), and it has also found its applications in probabilistic tsunami hazard assessments (PTHA), typically as an extension of PSHA for tsunamis of seismic origin, and rarely in the assessment of probabilistic volcanic hazard (PVHA). In PSHA, it is very rare to see a published hazard study or a site-specific hazard analysis of regulatory relevance that does not include a *LT*. Part of the popularity of logic trees is related to the fact that they are technically easy to implement. It is a powerful tool to organize one's thinking in situations where alternative models, in which the analyst has different degrees of confidence, might apply. In Figure 2-2, we report an example of Logic Tree that has been proposed in the framework of PTHA (Geist and Parsons 2006). In this simplified logic tree, the alternative models are considered for three model factors: (1) earthquake recurrence (characteristic, Gutenberg–Richter), (2) seismogenic depths ( $z_1 \dots z_4$ ), and (3) shear modulus: PREM (Dziewonski and Anderson, 1981); regional observations (e.g., Bilek and Lay, 1999). For each alternative formulation, the weighting factor is shown below in parentheses. Joint probabilities for each parameter are shown on right-hand side.

There is an open debate related to the interpretation of the branch weights in logic trees and whether they are probabilities or simply subjective indications of relative merit (Abrahamson and Bommer 2005, McGuire et al. 2005, Musson 2005). However, in their current application (mainly based on the framework envisioned by SSHAC 1997 and following specifications), logic trees are an implementation of probability trees. This means that the weights associated with a particular branch of a logic tree, which express the degree-of-belief of the hazard analyst(s) in the corresponding model, are subsequently treated as (subjective) probabilities to calculate a distribution of hazard curves (Bommer and Scherbaum, 2008).



**Figure 2-2:** LT method - Example of a simplified logic tree for three model parameters (from Geist and Parsons 2006).

The logic tree is then one of the several flavours of graphical probability trees (issue tree, event tree, fault tree, etc.) that aim to dissect a specific problem into its basic components. The structure and the kind of the tree are defined according to its practical use. Yet, we can identify one common feature for almost all the tree structures: the branches emerging from each node must represent a Mutually Exclusive and Collectively Exhaustive (MECE) set of events (e.g., Smith, 1988), in order to meet the Kolmogorov axioms for probabilities. Some relaxations of this assumption have been made, mostly motivated by practical aspects (see,

e.g., Newhall and Hoblitt, 2002; Marzocchi et al., 2010), but these changes involve much more cumbersome calculations. For this reason, almost all probability tree structures, and in particular the logic tree that is used in PSHA (Kulkarni et al., 1984; Bommer and Scherbaum, 2008) are based on MECE assumption. In PSHA, the different branches of the logic tree are meant to describe the epistemic uncertainties related to the different components (nodes) of the hazard model. Referencing from Bommer and Scherbaum (2008), “*the mutual exclusivity requires that the model on one of the branches is applicable but not a combination of models, while the collective exhaustiveness requires that one of the models in the model set considered fully applies*”, as illustrated in Figure 2-2. The need of having a MECE set of events is imposed by how probabilities are combined in the probability tree. For example, consider a logic tree where we are interested in the probability of one specific event  $E$  and the terminal branches of the tree mimic different ways (different models) that can be used to get such a probability. The final assessment is given by the law of total probability that reads

$$\Pr(E) = \sum_{i=1}^N \Pr(E \cap H_i) = \sum_{i=1}^N \Pr(E | H_i) \Pr(H_i). \quad (1)$$

Here  $\Pr(E)$  is the probability of the event of interest,  $\Pr(E|H_i)$  is the conditional probability of the event  $E$  given the model  $H_i$ , and  $\Pr(H_i)$  is the probability that the model  $H_i$  is the true model. As long as equation (1) is used in the probability tree, the models  $H_i$  must represent a set MECE. The most important consequence of the MECE assumption is that one path of the tree must represent the truth. Since no practitioner believes that one of the paths of the logic tree represents the true hazard, the MECE assumption is pragmatically resumed replacing the term “*true* with the *one that should be used*” (Scherbaum and Kuhen, 2011).

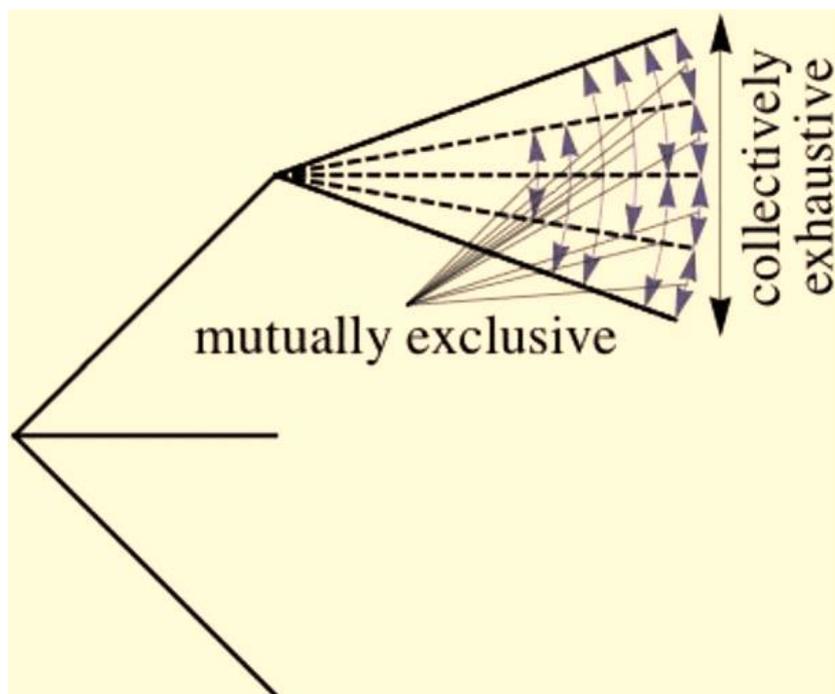


Figure 2-3: LT method - Graphical representation of the MECE criterion (from Bommer and Scherbaum, 2008).

Trying to meet the requirements of the MECE is in general a difficult task. As discussed in Bommer and Scherbaum (2008), the MECE criterion easily may lead practitioners to include many branches (following the reasoning that more branches, i.e., alternative models, increase the chance that the true model is among the options), as well as many sections. On the other hand, more branches also increase the chances of model redundancy, automatically violating the condition of mutual exclusivity (Burnham and Anderson 2002). In addition, the practical achievement of mutual exclusivity is even more complicated because of the subtle nature of the interdependence of models. The clearest case in point is the ground-motion prediction equations since the models applied to a PSHA in a particular region will almost always be derived from datasets that have appreciable overlap in terms of the strong-motion recordings employed. According to Dreyfus and Dreyfus (1986), to understand what is relevant in the main characteristic distinguishing an expert from a novice. Without a sensitivity analysis, it is also impossible to judge the relevance of a particular logic-tree section. Therefore, Bommer and Scherbaum (2008) suggest that setting up a logic tree should always be accompanied by a sensitivity study, in order to reduce the number of branches and of sections of the actual logic tree.

Although the inclusion of a logic tree in PSHA is now *de rigueur*, we note that since the beginning (Kulkarni et al., 1984) the logic tree has been sometimes applied to PSHA in a

peculiar way (Marzocchi et al. 2015). For instance, it has been used to define percentiles of the expected ground shaking probability, reporting the average value (given by equation 1) and a confidence interval defined through the percentiles. In some cases, it has been argued that the hazard should be represented by a percentile instead of the average (Abrahamson and Bommer, 2005). The use of percentiles is questioned by other practitioners (e.g., Musson, 2005) who assert that a logic tree can provide only one number; using their words, the mean hazard is *the* hazard that is obtained over epistemic uncertainties.

This dichotomy in the interpretation of the results of logic trees is very present in the literature of PSHA. In our opinion, the reason for this long-term discussion is that the use of percentiles, even if not easily justifiable from a theoretical point of view, seems reasonable in practice, enabling the analysis and the visualization of epistemic uncertainty. The body and the range of the acceptable technical interpretations cannot indeed be appreciated reporting only the *mean* hazard. In addition, the explicit exploration of such range allows a more effective capability of comparing different hazard results in different areas (where the level of epistemic uncertainty may be substantially different), as well as, of testing them against real data. Here, we simply note that the common practice of representing the results of a LT in terms of percentiles resembles the Ensemble modelling framework that is discussed in *section 2.3*.

A large set of applications of LTs can be found in literature. The use of LT has become almost a standard in PSHA, often in the context of a Multiple Expert Integration (MEI, *section 2.5*). LTs have been proposed also for Probabilistic Tsunami Hazard Analysis (e.g., Geist & Parsons 2006), typically as an extension of PSHA to tsunamis of seismic origin (SPTHA, following Lorito et al. 2015). Several applications exist also for Probabilistic Volcanic Hazard Analysis (Smith et al, 1997), even if in PVHA methods based on BEA (*section 2.3*) and classical Expert Elicitations (*section 2.4*) are prevalent.

In Figure 2-4, we report the example of the LT developed in the European FP7 project SHARE, 2013). Three main levels are indicated with different colours: earthquake source model (yellow), Maximum magnitude (green) and ground motion (red). The tectonic regionalization is not a branching level (grey), but it defines the Ground Motion Prediction Equations to be used. Under each branch, it is reported the weight of each branch, to be interpreted as the probability of the branch being the best among the others. For the source models, different weighting schemes have been adopted for the different average return periods. The results are reported in Figure 2-5, the resulting mean hazard map for a probability level of 10% in 50 years is reported.



**Figure 2-4:** SHARE logic tree. Colours depict the branching levels for the earthquake source models (yellow), maximum magnitude models (green) and ground motion models (red). Values below the black lines indicate the weights, for the source model these indicate the weighting scheme for the different return periods. Tectonic regionalization is not branching levels (grey) of the model; however, it defines the GMPs to be used (from Woessner et al., 2013, Giardini et al. 2013).

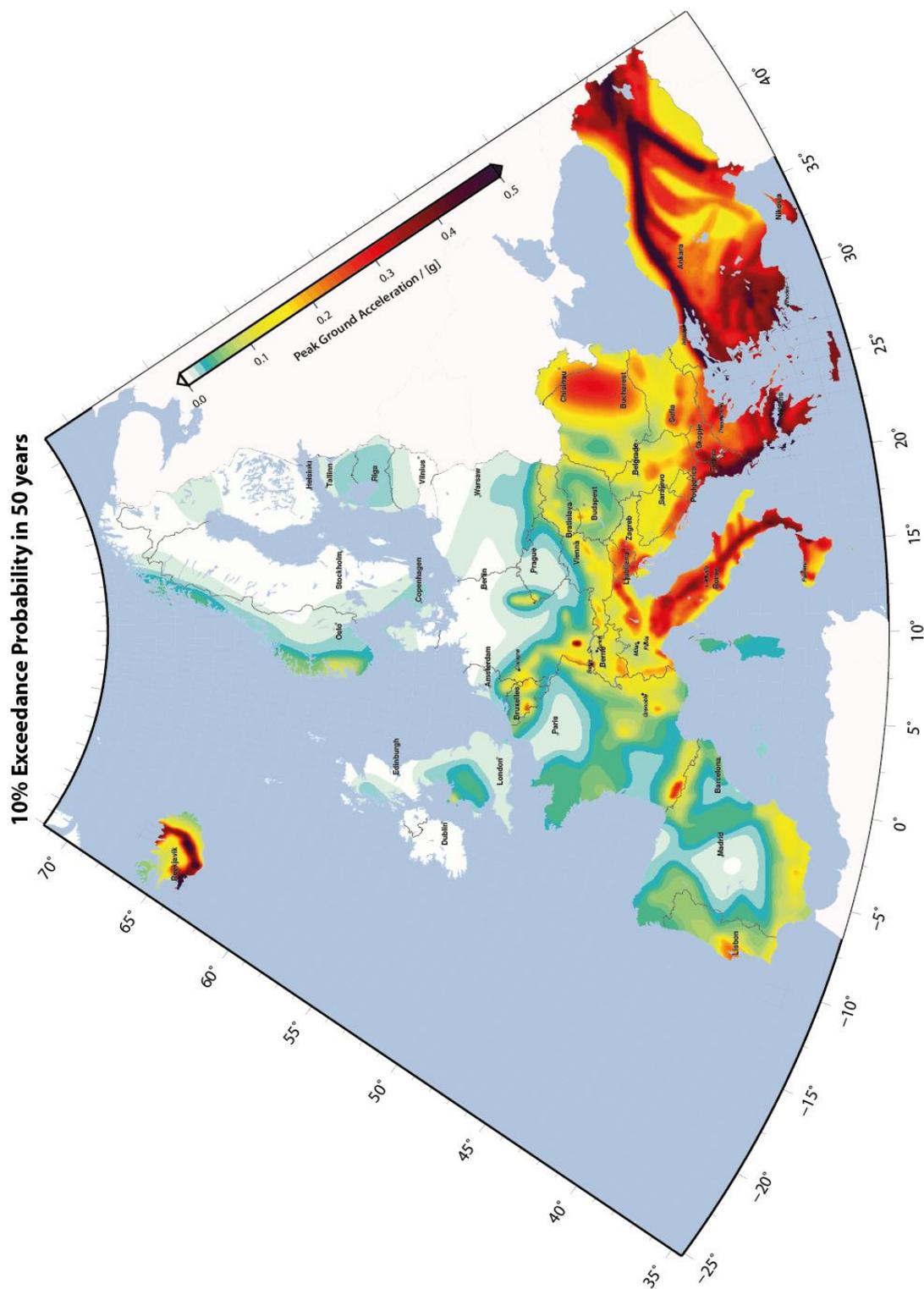


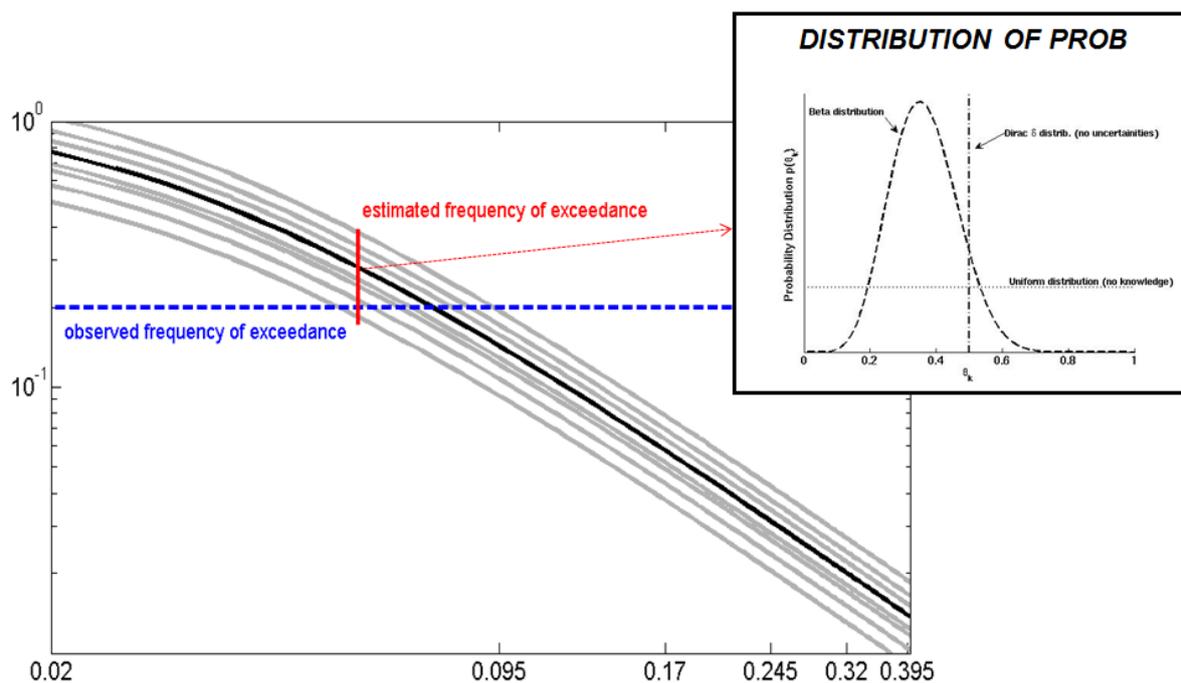
Figure 2-5: Probabilistic Seismic Hazard Analysis (PSHA) results of the project SHARE (Woessner et al., 2013, Giardini et al. 2013).

### 2.3 BAYESIAN/ENSEMBLE APPROACH (BEA)

Many different applications can be made in the general framework of Bayesian inference. In the field of natural hazards, many examples exist, mainly focused on the assessment of one model's parameters (e.g., recent examples for seismic hazard and risk assessments are, among the many others, Franchin et al. 2013, Selva et al. 2013, and Keller et al. 2014). However, here we focus on approaches able to deal with both model and parameter uncertainties. In recent literature, several methods of this kind have been presented, mainly following the Bayesian formulation proposed by Gelman et al. (1995). In few words, the core of the method consists of selecting as target inference variable a probability value  $\theta$  that is updated through the Bayes' rule with the observed frequency of the phenomena in the past. The result is the assessment of a probability distribution relative to the probability  $\theta$ , as reported in Figure 2-6.

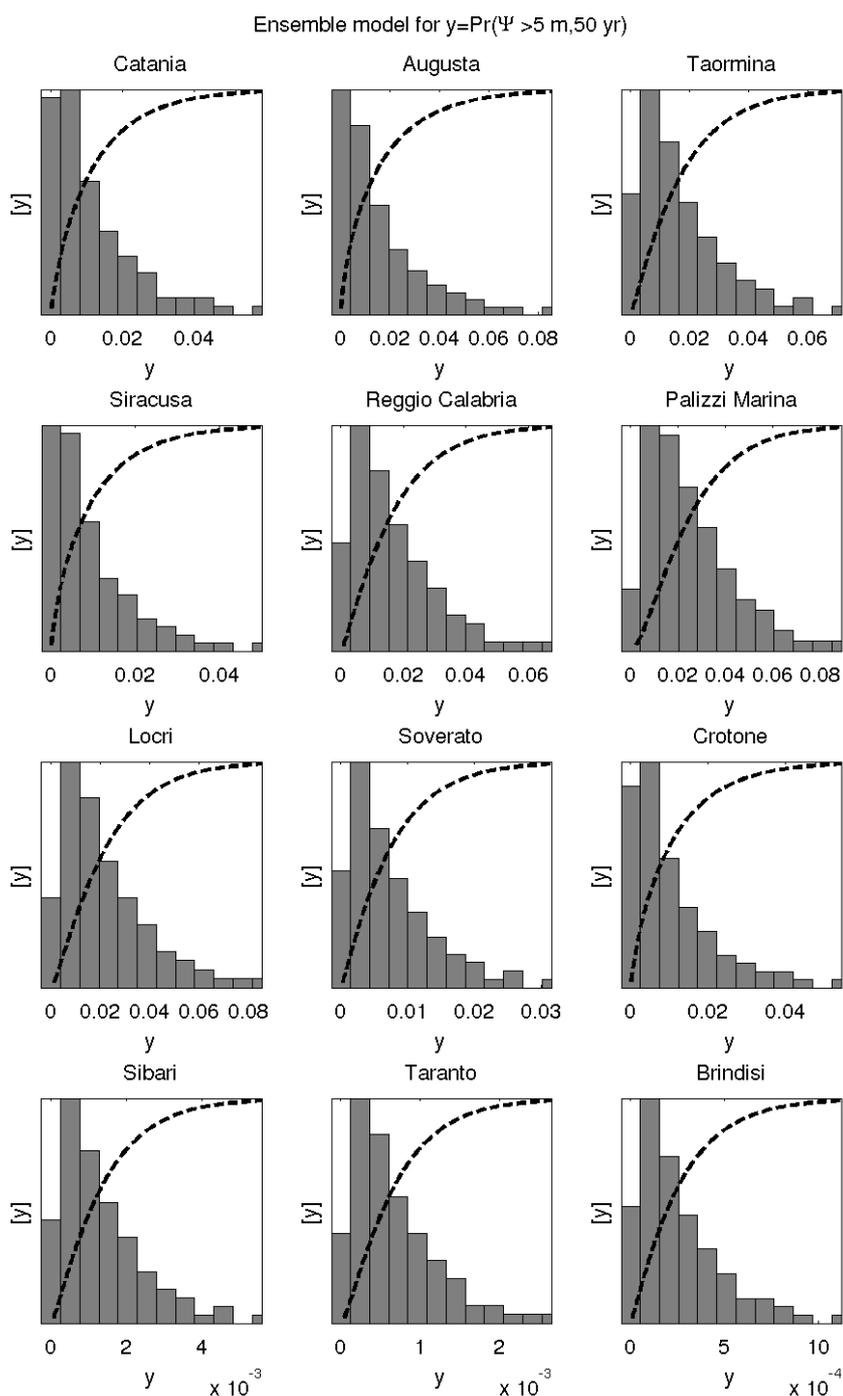
The use of a probability distribution of probability has been a matter of discussion and controversies in statistical literature and among practitioners (e.g. Lindley, 2000; Cox et al., 2008). These controversies have been addressed by Marzocchi and Jordan (2014) who provided a formal and consistent probabilistic framework to interpret such distributions. The central value of this distribution is the best guess of the frequency of an exchangeable dataset (i.e., the aleatory variability), and the dispersion around the central value mimics the epistemic uncertainty (see also Marzocchi et al., 2008). This distribution has an intuitive interpretation, because it tells where the true aleatory variability is expected to be (e.g., SSHAC, 1997).

In this conceptual framework, it is possible to introduce the concept of *Ensemble* modelling. In general, PSHA practitioners have a finite set of different models  $\{\theta_i, \omega_i\}$  ( $i = 1, \dots, N$ ), where  $y_i$  and  $\omega_i$  are the outcome and the weight of the  $i$ -th model in a set of  $N$  different models describing one specific variable of interest  $\theta$ , e.g., the hazard curve for a specific intensity parameter. Ensemble modelling considers  $\{\theta_i, \omega_i\}$  as a sample of an unknown parent distribution  $f(\theta)$  that describes the variable  $\theta$  taking into account simultaneously the aleatory variability and epistemic uncertainty. The sample  $\{\theta_i, \omega_i\}$  can either derive from one or more logic trees, or from a collection of models; the only requirement is that  $\{\theta_i, \omega_i\}$  represents an unbiased sample of the epistemic uncertainty. The weight associated with each model/branch should properly take into account not only the confidence on each model (based on expert opinion and/or on quantitative evaluation the forecasting performances), but also the possible strong correlation among models, as discussed in Bommer and Scherbaum (2008), Marzocchi et al. (2012) and Rhoades et al. (2014).



**Figure 2-6:** BEA method - Probability distribution of probability for a simultaneous treatment of aleatory and epistemic uncertainties in probabilistic hazard curves.

Ensemble modelling can be applied to different variables. For PSHA, it is often convenient to set  $\theta_z = \Pr(Z \geq z)$ , where  $\theta_z$  is the exceedance probability for a specific value  $z$  of the ground shaking intensity parameter, for instance the ground shaking value used for the building code. In other words, the target assessment is the hazard curve itself. In this case, ensemble modelling provides a distribution of the exceedance probability associated to each specific value for the intensity. If a single value is required to characterize the distribution, the mean value has the same legitimacy than any other single statistics, like the mode and the median to represent the distribution. Examples of these distributions are reported in Figure 2-7, where hazard curves for PTHA in the Ionian Sea are used to visualize the probability exceedance of more than 5 m of  $h_{max}$  at various locations (Selva et al. 2015). Note that this theoretical framework provides also a base for interpreting the percentiles of LT results, whose use is indeed controversial within the LT literature, since such percentiles may be simply referred to the parent distribution  $f(\theta_z)$ .



**Figure 2-7:** Probability of exceedance of 5m in different locations in Southern Italy, as evaluated for a preliminary assessment of PTHA (Selva et al. 2015).

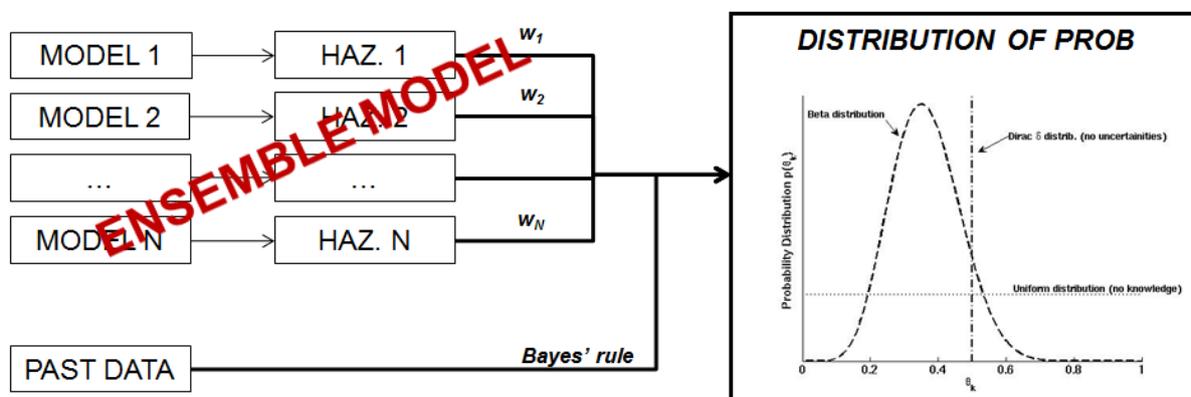
Although ensemble modelling does not impose any specific parametric distribution for  $f(\theta_z)$ , the Beta distribution is commonly used to describe a unimodal random variable bounded between 0 and 1 (Gelman et al., 2003). Different almost equivalent schemes to set the

parameters of the Beta distribution can be selected, ranging from setting the distribution's moments (Marzocchi et al, 2015), weighted likelihood (Selva et al. 2013), minimization of fractiles (Selva and Sandri 2013), or constraining means and adding a subjective measure of the variance based on the subjective credibility of each modelling approach (e.g., Marzocchi et al. 2008; Selva et al. 2012). As discussed in Marzocchi et al. (2015), if a sufficiently large set of alternative models is used to set the prior, also non parametric distribution can be used, without the need of selecting a Beta distribution.

Independent past data, collected in agreement to the conceptual experiment, may also be introduced into the assessment making use of the Bayes' rule, making Ensemble modelling intrinsically compatible with Bayesian inference. In practice, the distribution  $f(\theta_z) = [\theta]$  may be assessed through

$$[\theta] \equiv [\theta|\{y\}] = \frac{[\{y\}|\theta][\theta]_{prior}}{[\{y\}]} \sim [\{y\}|\theta][\theta]_{prior} \quad (2)$$

where  $[\theta]_{prior}$  represents the prior distribution, and it is based on theoretical models and/or prior beliefs,  $\{y\}$  is a vector of past data (recorded frequencies of the target event), and  $[\theta|\{y\}]$  represents the posterior distributions that accounts for both prior information and past data (e.g. Gelman et al., 1995). A visual representation of this process is reported in Figure 2-8.



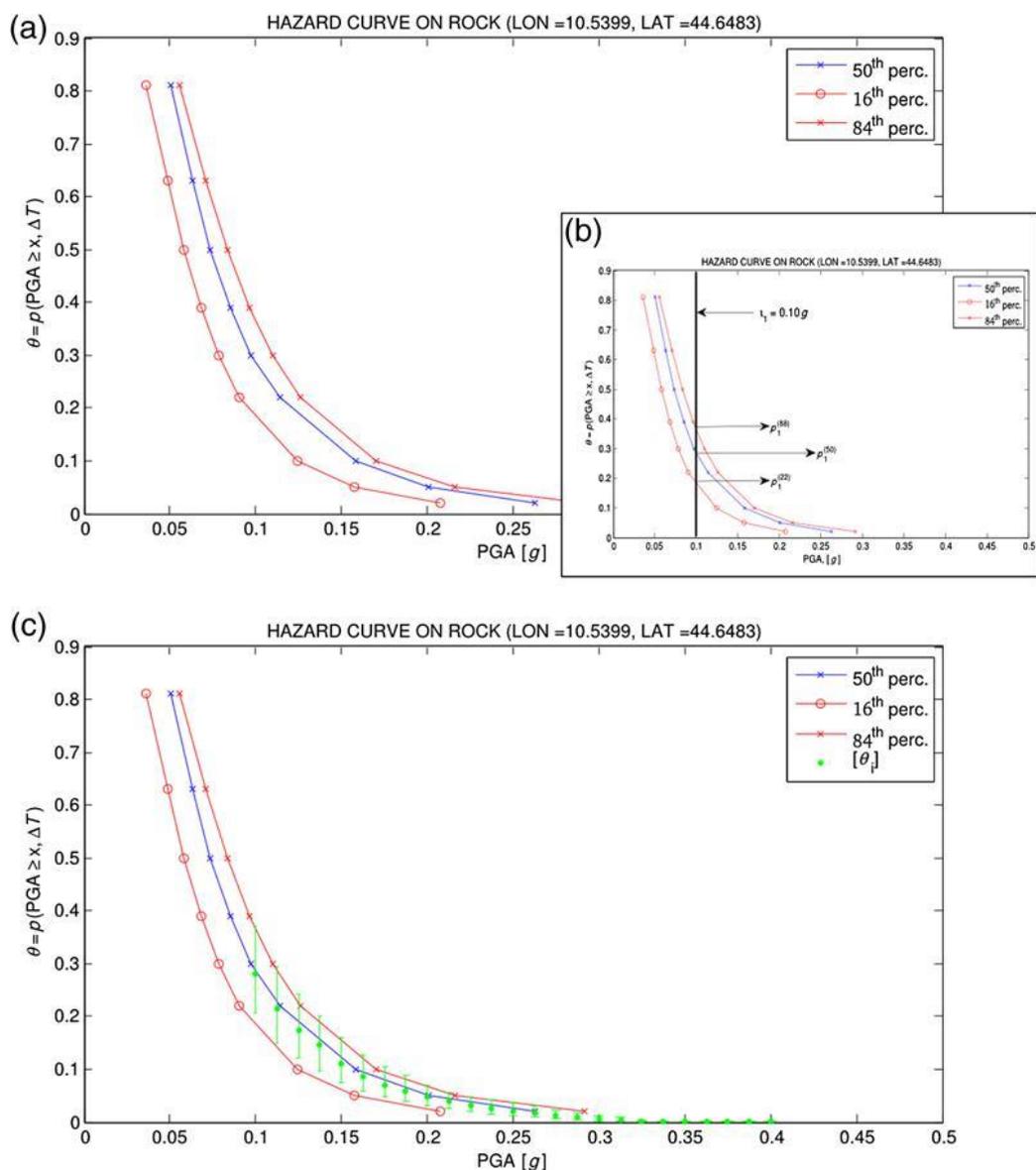
**Figure 2-8:** Bayesian/Ensemble Approach (BEA) - Graphical representation of the inference process of BEA.

In summary, the epistemic and aleatory uncertainty are treated simultaneously through a probability distribution  $f(\theta)=[\theta]$ , where  $\theta$  represents a given probability value. The inference of this distribution can be based on any available source of information, from theoretical models to past data. In this kind of application ensemble modelling offers to natural hazard practitioners some remarkable advantages. First, replacing few probability outcomes with a

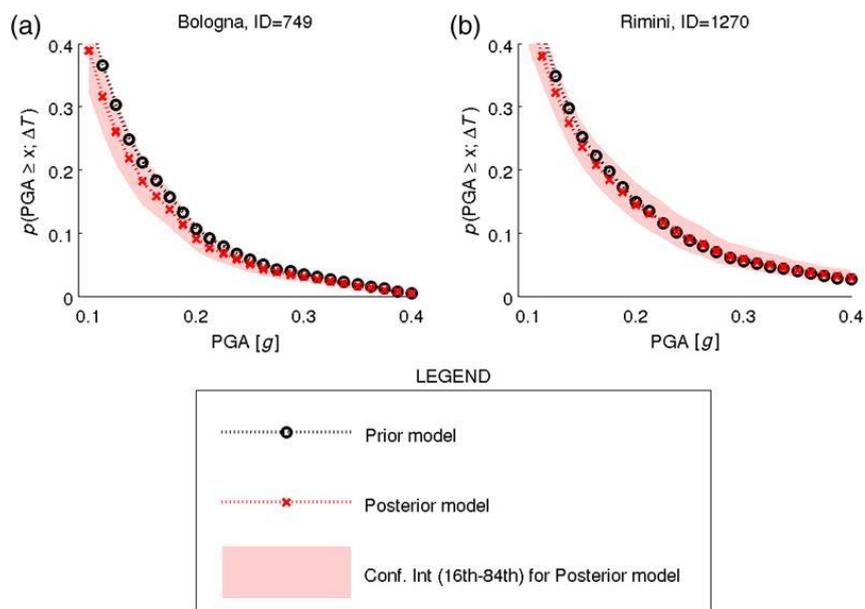
continuous distribution describing the aleatory variability and epistemic uncertainty is crucial for a meaningful test of any probabilistic hazard model (Marzocchi and Jordan, 2014). Moreover, describing the epistemic uncertainty with a distribution allows more meaningful comparisons between hazard in different sites and it shows the sites where the true hazard may be more distant from the mean hazard. For example, two sites with the same mean hazard may have a quite different dispersion of the exceedance probability distribution; this means that, although the mean hazard is the same, the site with the largest dispersion may have a true hazard much lower (or much higher) than the other site. In addition, showing to the stakeholders the dispersion around the mean for each site may indicate the sites where we expect the largest variations of the mean hazard in future hazard evaluations. Similarly, showing the dispersion is also essential to evaluate which are the sites where more monitoring or research are needed in order to reduce uncertainties. Finally, the use of a continuous distribution explicitly shows that the most likely value of the exceedance probability is not necessarily associated with one branch (in contradiction with the MECE assumption), and the probability that the true exceedance probability is lower (larger) than the lowest (largest) value of the logic tree is not negligible.

Several models following this general scheme have been presented in literature for different hazards, ranging from PSHA (e.g., Selva and Sandri 2013, Marzocchi et al., 2015), to PVHA (Marzocchi et al. 2008, 2010; Selva et al. 2010, 2015; Sandri et al. 2014) and PTHA (e.g., Grezio et al. 2012), as well as for fragility models (Selva et al. 2013).

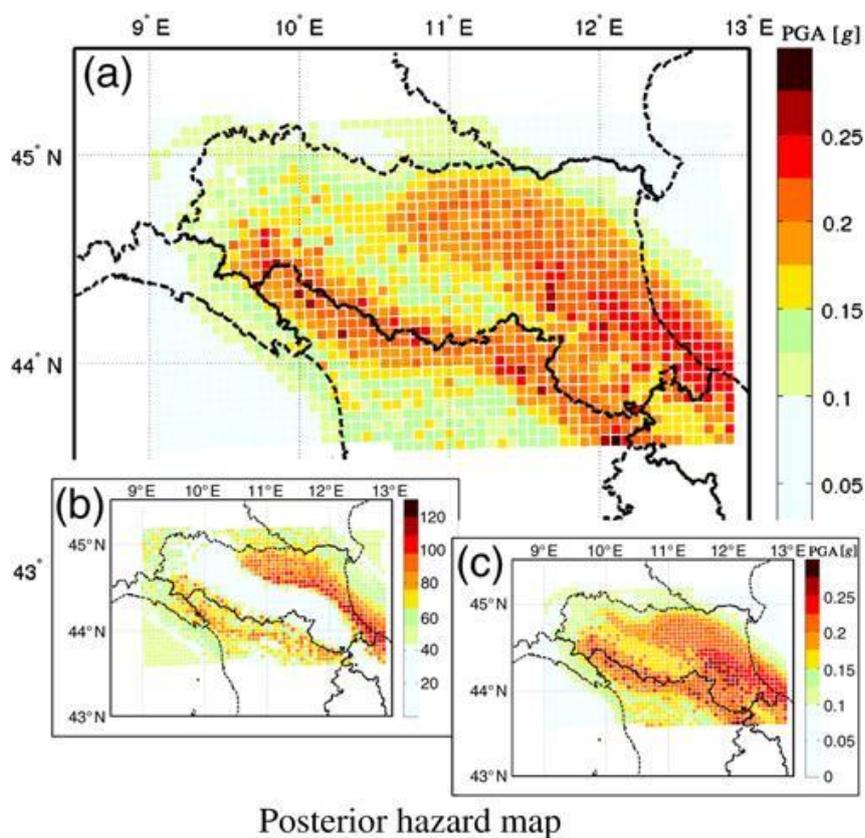
As an example, we report an application for Bayesian PSHA for the Emilia-Romagna region in Northern Italy, published in Selva and Sandri (2014). In this example, the prior distribution has been set making use of the alternative models emerging from the LT adopted for Italian PSHA (GdL MPS, 2004) and corrected for site effects adopting the amplification factors established by the Italian laws (OPCM3431, 2005). In Figure 2-9, we report the fit of the Ensemble model of the percentiles emerging from the alternative models developed for a LT (Stucchi et al. 2011) on hard rock (i.e., before site effects). This prior distribution has been updated based on the historical macro-seismic data (Stucchi et al. 2007), accounting for the uncertainty on the transformation between macroseismic intensity and peak ground acceleration (Faenza and Michelini 2010). The resulting hazard curves, for different sites, and the hazard maps for a probability level of 10% in 50 years (along with the relative epistemic variance for a PGA level of 0.1 g) are reported in Figure 2-10 and Figure 2-11, showing that past data, in specific locations, are able to significantly change prior distributions.



**Figure 2-9:** Fit of the prior model probability in a representative location: (a) the starting dataset, (b) the elaboration of input data, and (c) the obtained prior probabilities (from Selva and Sandri, 2013)



**Figure 2-10:** (a) Hazard curve (in 50 years) for the city of Bologna, from prior and posterior models. (b) Same as (a), but for the city of Rimini (from Selva and Sandri, 2013)



**Figure 2-11:** (a) Hazard map (10% in 50 years) obtained by the medians of posterior model on sites in g units. (b) The equivalent number of data (inversely proportional to the epistemic variance) , relative to the PGA value  $\tau 1 0:1g$ , for the posterior model. (c) Same as (a), with past data only for sites with complete historical record (from Selva and Sandri 2013).

## 2.4 CLASSICAL EXPERT ELICITATION (*cEE*)

In any probabilistic hazard and/or risk analysis, the hazardous events are treated in terms of rate of occurrence, although very seldom large amounts of data are available. Expert Judgement techniques are useful for quantifying models in which situations it has been impossible to make enough observations to quantify the model with 'real data'. This can happen because of cost limitations, technical difficulties or the uniqueness of the situation under study (e.g. frequency of rare events). Data collected through Expert Judgement may also be used to refine estimates from 'real data' when a highly refined model is required or to estimate model parameter uncertainties. Indeed, Expert Judgement has always played a large role in science and engineering since engineers are quite able to provide engineering data which are essentially subjective data driven by engineering models and experience.

A first critical point in the application of Expert Judgement techniques is the choice of the experts because if on one hand it is an advantage to consider a wide range of experts, on the other hand the choice of excluding some experts as belonging to a certain "school of thought" can easily lead to evident bias. Biases can occur at different levels, from the influence of mindset of the expert to overconfidence of the expert, anchoring of the expert to a certain opinion based on some previous estimate or cognitive bias for which occurrence of common events are overestimated while occurrence of rare events underestimated. Biases can be limited only to a certain extent (Bedford et al. 2001).

Protocol and procedure are also key aspects in expert elicitation. Although *informal* expert judgement can be held in case of involvement of *subject-matter experts*, more *formal structured* expert judgement involves multiple experts at different levels. According to NUREG-1150 and NUREG-1163, we can distinguish the experts as:

- *normative* expert, who has training and experience in statistics, decision analysis and probability encoding and who has the duty of structuring the formal elicitation and train the subject-matters experts in probability encoding;
- *generalist* expert, who understands the context in which the data provided by the experts will be used, guides the normative experts in preparing the elicitation to make sure that the data gained are useful, provides documentation to the subject-matter experts and help training them;
- *subject-matter* expert, who provides the subjective judgment that are encoded as probabilities (e.g. probability distribution or a point estimate of an uncertain parameter).

Typically the elicitations are conducted to quantify uncertainties (an uncertain parameter, frequency and likelihood, degree of causality of variables, correlations, and relative merits of alternative conceptual models...), the information regarding uncertainty is then represented by encoding the subjective probabilities from each subject-matter expert. This is common practice rather than give a point estimate, since the data collected represent only a small sample from a highly variable population and probability distributions can be directly related to the likelihood of various future events, frequencies of various events and can be used as direct input to probabilistic performance assessments.

With reference to Goossens et al. (2008), the protocol for a structured expert elicitation consists of the following fifteen steps:

- Preparation for elicitation
  - Definition of case structure
  - Identification of target variables
  - Identification of query variables
  - Identification of performance variables
  - Identification of experts
  - Selection of experts
  - Definition of elicitation format document
  - Dry run exercise<sup>1</sup>
  - Expert training session
- Elicitation
  - Expert elicitation session
- Post-Elicitation
  - Combination of expert assessment
  - Discrepancy and robustness analysis
  - Feedback
  - Post-processing analysis
  - Documentation.

---

<sup>1</sup> The dry run exercise aims at finding out whether the case structure document and the elicitation format document are unambiguously outlined and whether they capture all relevant information and questions. One or two experts, preferably not used in the elicitation session, are given both these documents and asked to provide their comments.

The selection of the expert should be done among those professionals who really stand out for their knowledge and experience in the specific field, but there are no objective criteria for proclaiming a person an expert. Therefore, as reported in Goossens et al.(2008), a first selection of experts by peer designations is considered the best option available, further experts can be indicated by the first selection experts.

The elicitation format depends strongly on the definition of the problem and the identification of target variables. All the parameters to be assessed by the experts are potential target variables. A formal procedure needs to be used to screen and select the final set of target variables among the most important ones, since the experts will be called to answer a limited number of questions. To this aim a ranking procedure or a pairwise-comparison procedure or a structured expert judgment can be used.

Query variables (object of the elicitation) and target variables are not necessary the same. Indeed, target variables may not be appropriate for a direct elicitation and in this case query variables (proxy for the target variables) have to be identified. Query variables should be represented by observable quantities and the questions formulated in the way the expert is used to treat those quantities.

Performance variables are also called *seed* variables. Seed variables are usually known to the analyst and supported by experimental data, while they are unknown to the experts.

The elaboration of the elicitation results depends on the chosen post-processing method in terms of combining the experts' opinions (distributions).

The methods for combining experts' distributions are divided in *Bayesian* and *non-Bayesian* methods. Many Bayesian models have been proposed in the literature but they are not commonly used. The most known is the model of Apostolakis, that solves the problem of estimation of the distribution of an unknown quantity. Given an unknown quantity of interest  $X$ , and a prior density distribution  $p(x)$  over  $X$ , chosen by the decision maker, the problem is to use the Bayes' theorem to update the density  $p(x)$  with the information  $x_1, x_2, \dots, x_n$ , gained from experts' opinions under the hypothesis of independent expert opinions. The hypothesis of independent experts allows to compute the likelihood simply as the product of the conditional probabilities  $p(x|x_i)$ . Among non-Bayesian models to combine experts' opinions, the Delphi method is widely used (Hsu, C. C., and Sandford, B. A. 2007). Fuzzy methods are also used. The Delphi method of Helmer, prescribes to first collect the opinion in an anonymous way, then median responses and interquartile ranges are returned to experts. The experts whose answers were on top and bottom of the quartiles are asked to review their opinion (to defend or revise). By iterating this procedure up to three or four times, the results are then representative of a group census. This procedure resembles in many ways with MEI methods (*section 2.5*), even explicit formulations of models is not required here.

Combination rules of expert opinion are object of strong arguments since some authors strictly consider the expert opinion independent from each other (independence preservation property), while other authors are against this idea.

A performance-based weighting method for combining the expert opinion was proposed by Cooke in 1991 and is described in the following section.

### **Cooke's method for Structured Expert Judgment**

One of the methods to perform Structured Expert Judgment is the so called *Classical model* or *Cooke's method* for structured expert judgment. Since every rational individual has his own subjective probability (and in particular, experts usually do have different opinions) it is necessary to find a way of building consensus (see Bedford et al. 2001).

This method aims to provide a *rational consensus* of the opinions of the consulted experts. This rational consensus is attained through statistical combination of the expert opinions according to a performance weighting principle. This requires that the expert opinions to be gathered conform to a structured method.

The fundamental principle of performance-based weighting is that the weights used to combine expert distributions are chosen according to the performance of the experts on so-called *calibration questions*, questions for which the answers are known to the analyst but not to the experts. Then the experts' answers to the calibration questions are scored with respect to calibration or statistical likelihood, and informativeness. These scores are used to compute weights which satisfy an *asymptotic strictly proper scoring rule* that prescribes that an expert achieves his/her maximal expected weight, in the long run, by and only by stating assessments corresponding to his/her true beliefs.

Using this method, the experts are asked to provide their opinion and express their (un)certainly around this opinion. This is most commonly achieved through asking the experts to provide:

- A quantity which the expert expects it to be close to the actual quantity. This is usually described as the 50<sup>th</sup> percentile, indicating the expert expects there to be a 50% chance the actual quantity will be higher or lower than this quantity.
- A quantity of which the expert does not expect the actual quantity to be lower than this quantity. This is usually described as the 5<sup>th</sup> percentile, indicating the expert expects there to be a very small chance (5%) the actual quantity will be lower than this quantity.
- A quantity of which the expert does not expect the actual quantity to be higher than this quantity. This is usually described as the 95<sup>th</sup> percentile, indicating the expert

expects there to be a very small chance (5%) the actual quantity will be higher than this quantity.

Using this method of questioning, two types of questions can be asked to the experts:

1. *Variables of interest*

The percentiles are assessed for uncertain quantities whose value is not known, not to the analysts or to the experts at the moment of the elicitation. These questions are used to provide the combined expert opinion.

2. *Seed variables or calibration variables*

The percentiles are assessed for quantities whose value is known (or will be known within the time frame of the research) to the analysts but not to the experts at the moment of the elicitation. These questions are used to ensure empirical control of the expert's uncertainty assessments.

With the seed variables two measures of performance are computed: the *calibration* and *information* scores. Roughly, calibration measures the degree to which experts are statistically accurate while information measures the degree to which experts' uncertainty estimates are concentrated relative to a background measure.

**Calibration**

Let us assume that we have answers from  $e = 1, \dots, k$  experts on  $i = 1, \dots, N$  calibration variables. Let us further assume that we assess three quantiles: 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> for each uncertain quantity. For each quantity, each expert divides his/her belief range into four inter quantile intervals for which the corresponding probability of occurrence are:  $p_1 = 0.05$  for a realization value  $\leq 5^{\text{th}}$  percentile,  $p_2 = 0.45$  for a realization value  $\in (5^{\text{th}}, 50^{\text{th}})$  percentile,  $p_3 = 0.45$  for a realization value  $\in (50^{\text{th}}, 95^{\text{th}})$  percentile, and  $p_4 = 0.05$  for a realization value  $\geq 95^{\text{th}}$  percentile. The empirical version of  $\mathbf{p} = (p_1, \dots, p_4)$  for expert  $e$ , is denoted  $\mathbf{s}(e) = (s_1, \dots, s_4)$ . Where

$$s_1(e) = \frac{\# \text{ realizations in seed variables } \leq 5^{\text{th}} \text{ percentiles assessed by expert } e}{N}$$

$$s_2(e) = \frac{\# \text{ realizations in seed variables } \in (5^{\text{th}}, 50^{\text{th}}) \text{ percentiles assessed by expert } e}{N}$$

$$s_3(e) = \frac{\# \text{ realizations in seed variables } \in (50^{\text{th}}, 95^{\text{th}}) \text{ percentiles assessed by expert } e}{N}$$

$$s_4(e) = \frac{\# \text{ realizations in seed variables } \geq 95^{\text{th}} \text{ percentiles assessed by expert } e}{N}$$

One way to measure the difference between  $\mathbf{p}$  and  $\mathbf{s}(e)$  is through relative information or entropy.

$$2 N I(\mathbf{s}(e), \mathbf{p}) = 2 N \sum_{j=1}^4 s_j \ln \left( \frac{s_j}{p_j} \right) \quad (2)$$

Expert assessments are treated as statistical hypotheses. Let us consider for each expert the null hypothesis:

$\mathbf{H}_0$ : the inter quantile interval containing the true value for each variable is drawn independently from the probability vector  $\mathbf{p}$ .

The quantity in equation (1) is asymptotically  $\chi_3^2$  (the degrees of freedom are the number of inter quantile intervals minus 1). This quantity can be used to test  $\mathbf{H}_0$  and it defines the calibration score:

$$CS(e) = P\{2 N I(\mathbf{s}(e), \mathbf{p}) \geq r\}, \quad (3)$$

where  $r$  is the value for expert  $e$  computed as in equation (2). The probability can be evaluated by a  $\chi_3^2$  distribution. The calibration score  $CS(e)$  is the probability that a deviation at least as large as  $r$  could be observed on  $N$  realizations if  $\mathbf{H}_0$  were true. Values of calibration close to zero mean that it is unlikely that the experts probabilities are correct.

### Information

The information score measures the degree to which a distribution is concentrated with respect to a background measure. In the classical model the uniform or log uniform background measures are used. An intrinsic range is calculated for the background measure  $[q_l, q_h]$  where  $q_l = q_5 - k(q_{95} - q_5)$  and  $q_h = q_5 + k(q_{95} - q_5)$ .  $k$  is typically chosen as 0.10 (10% overshoot).  $q_5$  is the lowest 5<sup>th</sup> percentile assessed across experts for a particular item and  $q_{95}$  the largest 95<sup>th</sup> percentile assessed across experts for the same item. The information score is then computed as

$$IS(e) = \frac{1}{N} \sum_{i=1}^N I(f_{e,i}, g_i) \quad (4)$$

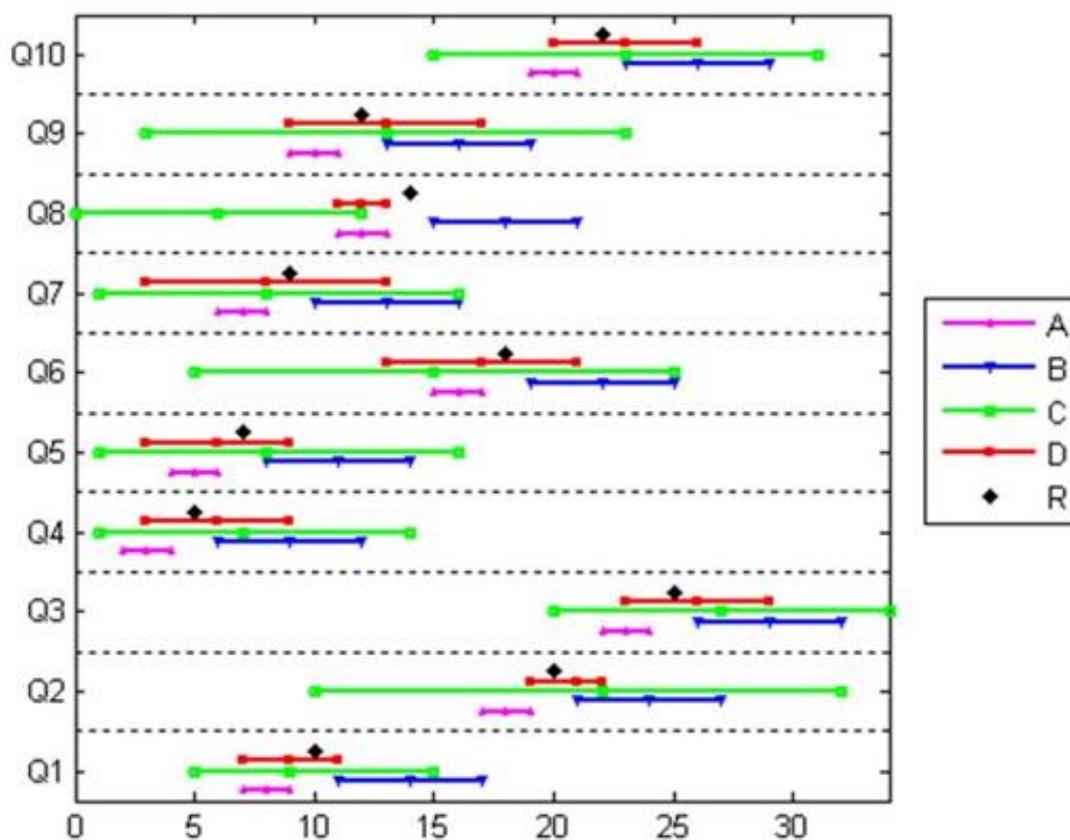
where  $g_i$  is the background density for item  $i$  and  $f_{e,i}$  the density for expert  $e$  on item  $i$ .  $I(f_{e,i}, g_i)$  is the mutual entropy between the densities of interest.

### Illustration on calibration and information

As an illustration of the previous two paragraphs, Figure 2-12 is given. In this figure the uncertainty assessments of 4 experts (A to D) and the actual values (R) are represented. On the vertical axis 10 questions are displayed (Q1 to Q10) and on the horizontal axis the quantities as given (0 to 35).

For example:

- The uncertainty assessment of expert C on question Q1:
  - o 5th percentile = 5
  - o 50th percentile = 9
  - o 95th percentile = 1
- The actual value corresponding to question Q1 = 10



**Figure 2-12: illustration on calibration and information for cEE.**

This figure illustrates the following:

Expert A:

- seems confident, considering the small uncertainty intervals
- always underestimates (not well calibrated)

Expert C:

- Always captures the actual value (well calibrated).

- Expresses large uncertainty, considering the large uncertainty intervals (not too informative)

Expert B or D:

- not too confident, but not too uncertain
- D is calibrated and informative
- B is overestimating

### Combination

In the classical model the combination of experts assessments is called a *Decision Maker* (DM). This is a weighted average of individual estimates. When the weights are determined based on the performance of experts in the seed variables, we speak of performance based DM.

It needs to be clarified that the weight  $w_\alpha(k)$  of expert  $k$  used in the combination of expert opinions does not have the interpretation of 'the probability that expert  $k$  is correct. The fact that the experts' weights sum to 1 does not imply that precisely one expert can be correct. Indeed, the experts' weights represent and may be interpreted purely as scores.

The DM distributions are thus:

$$DM_\alpha(i) = \frac{\sum_{e=1}^k w_\alpha(e) f_{e,i}}{\sum_{e=1}^k w_\alpha(e)} \quad (5)$$

Where the weights  $w_\alpha(e) = 1_{\{CS(e) > \alpha\}} CS(e) IS(e)$ .

Values of  $\alpha < 0.05$  would fail to confer the study the required level of confidence, so that experts with a calibration score under  $\alpha$  are automatically given a score of zero. Notice that the DM can also be evaluated in terms of calibration and information and is hence referred to as virtual expert. Given weights for each expert  $w_\alpha(k)$ , the combined distribution function is now  $DM_\alpha(i)$ .

## 2.5 MULTIPLE EXPERT INTEGRATION (MEI)

Probabilistic hazard and risk assessment inherently contains uncertainties due to limited information related to models, parameters, and experts' opinion. The treatment of uncertainties in selected models, data and experts' opinion can produce highly variable and inconsistent hazard and risk estimates in the assessment for a critical infrastructure in question (Bernreuter D.L. et al. 1989; EPRI 1989). Due to such inconsistencies, a set of guidelines were formulated in 1997 by a committee of seismic hazard practitioners called Senior Seismic Hazard Analysis Committee (SSHAC) resulting in the development of a structured and multi-level assessment process, also known as SSHAC process (SSHAC 1997). The purpose of the SSHAC process is basically to ensure the regulatory matters for critical infrastructures (CIs) in the framework of a probabilistic seismic hazard assessment (PSHA), emphasizing the involvement of multiple experts and focusing on assessing the epistemic uncertainty (Hanks et al. 2009).

The SSHAC guidelines are primarily based on four levels (1 – 4) of complexity spanning from simple and least resource intensive to complex and resource intensive studies. Level 1 is the simplest and least resource intensive, and level 4 is the most complex and resource intensive. The completeness and the level of assurance of each SSHAC level are based on to what extent the uncertainties are treated in the assessment. All levels are a composition of different participants engaging different roles and responsibilities [1] such as Project Sponsor (PS), Project Manager (PM), Participatory Peer Review Panel (PPRP), Technical Integrator (TI), hazard analyst, resource experts, evaluator experts and proponent experts, etc. The roles of some key participants are described below.

- The **Project Sponsor** is the entity that provides financial support for a project and “owns” the results of the study in the sense of property ownership.
- The **Project Leader** is the entity that takes managerial and technical responsibility for organizing and executing the project, oversees all other project participants, and makes decisions regarding the level of study of particular issues.
- The **Technical Integrator** is defined as a single entity—e.g., an individual or team—that is responsible for conducting the evaluation and integration processes.
- A **Resource expert** is a technical expert with specialized knowledge of a particular data set, model, or method of importance to the hazard analysis.

- A **Proponent expert** is a technical expert who advocates a particular hypothesis or technical position.
- An **Evaluator expert** evaluates the models developed by the other experts, assimilating the best aspects of all the models, and derives their own distributions that represent the uncertainty of the “informed technical community”.
- **Hazard analysts** are the PSHA cognoscenti who actually perform the PSHA calculations.
- **Participatory Peer Review Panel** is a panel of experts who are responsible for reviewing both the technical and process aspects of a project under consideration.

A generic summary of important attributes in different levels is presented in **Error! Reference source not found.** that is taken from the original SSHAC report (Committee and Budnitz 1997). **Error! Reference source not found.** shows the contribution of different aspects acting in the SSHAC framework. The most significant differences in most of the study attributes occur between Levels 2 and 3, with much less differences between 1 and 2, or 3 and 4. A brief description of the SSHAC Levels is presented here.

*Level 1)* SSHAC level 1 is the simplest procedure for implementation of probabilistic risk assessment. In this level, the Technical Integrator (TI) is the principal participant who reviews and evaluates the literature, datasets, and models related to the technical issues. The job of TI is to quantify uncertainties and to express her/his view of the Informed Technical Community (ITC) and then to compile documents with respect to all models, parameters and their technical basis. A preliminary document of hazard results is peer reviewed at a late stage. This level of assessment does not differ significantly from what has been discussed in *sections 2.2* (Logic Tree) and *2.3* (Bayesian/Ensemble Approach), since models are selected from literature and thus they do not arise from experts’ discussions, but only from the subjective choice of the TI team.

*Level 2)* SSHAC level 2 is the extension of level 1. This level includes all aspects of level 1 and promotes the communication and connection between different experts and analysts. The TI team contacts members of the ITC regarding applicable databases (published, unpublished, or restricted access) and directly communicates with proponents of alternative viewpoints in order to understand the alternatives and the technical bases behind them. The level 2 study emphasizes the need of meetings in order to resolve key issues and obtain

feedback during the process. Peer review of the work is suggested to be conducted at the late stage.

*Level 3)* SSHAC level 3 is a more complex procedure than the levels 1 and 2 in which the TI team works closely with resource and proponent experts for better understanding of alternative models and methods. Their close collaboration results in finding out the strengths and weaknesses of the various methods, models, and databases pertinent to hazard at a specific location. A series of workshops is advised to yield better communication among experts and peer review is usually participatory.

*Level 4)* SSHAC level 4 is the most complex procedure among all levels in which high level of regulatory assurance is set. The resource and proponents experts play the similar role as they play in level 3, but additionally a new participant is added, the Technical Facilitator Integrator (TFI), to ensure that key technical issues are fully addressed. This addition relies on the use of evaluator or evaluator teams rather than just a single TI team with purpose to obtain more robust estimate of the community distribution. Moreover, all evaluators and experts are equally informed about the available methods, models, and data. Both technical and process reviews are recommended in this Level. The technical review is primarily conducted among the experts and TFI, and the process review of interaction and development of community distribution like other levels is concerned by the participatory Review Panel (RP).

**Table 2-1: Attributes of various SSHAC levels.**

SSHAC Level	Level 1	Level 2	Level 3	Level 4
<b>Number of participants</b>	<ul style="list-style-type: none"> <li>• Project Manager</li> <li>• Small TI team</li> <li>• Peer reviewers</li> <li>• Hazard calculation team</li> </ul>	<ul style="list-style-type: none"> <li>• Project Manager</li> <li>• Small TI team</li> <li>• Peer reviewers</li> <li>• Hazard calculation team</li> <li>• Resource experts</li> </ul>	<ul style="list-style-type: none"> <li>• Project Manager</li> <li>• Project TI</li> <li>• Larger TI team</li> <li>• Peer reviewers</li> <li>• Resource experts</li> </ul>	<ul style="list-style-type: none"> <li>• Project Manager</li> <li>• Project TFI</li> <li>• Small TFI team</li> <li>• Panel(s) of evaluator experts</li> <li>• Peer reviewers</li> </ul>
<b>Interaction</b>	<ul style="list-style-type: none"> <li>• Limited or no contact with proponent and resource experts</li> </ul>	<ul style="list-style-type: none"> <li>• Proponent and resource experts contacted individually</li> </ul>	<ul style="list-style-type: none"> <li>• Proponent and resource experts interact with TI Team in facilitated workshops</li> </ul>	<ul style="list-style-type: none"> <li>• Proponent and resource experts interact with evaluator experts in facilitated workshops</li> </ul>
<b>Peer review</b>	<ul style="list-style-type: none"> <li>• Late stage</li> </ul>	<ul style="list-style-type: none"> <li>• Late stage</li> </ul>	<ul style="list-style-type: none"> <li>• Participatory</li> </ul>	<ul style="list-style-type: none"> <li>• Participatory</li> </ul>
<b>Ownership</b>	<ul style="list-style-type: none"> <li>• TI Team</li> </ul>	<ul style="list-style-type: none"> <li>• TI Team</li> </ul>	<ul style="list-style-type: none"> <li>• TI Team</li> </ul>	<ul style="list-style-type: none"> <li>• TFI team and evaluator experts</li> </ul>
<b>Transparency</b>	<ul style="list-style-type: none"> <li>• Dependent on documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Dependent on documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Interested parties can view interactions at workshops</li> <li>• Participatory peer reviewers observe workshops, participate in Workshop #3</li> <li>• Dependent on documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Interested parties can view interactions at workshops</li> <li>• Participatory peer reviewers observe workshops, participate in Workshop #3</li> <li>• Dependent on documentation</li> </ul>
<b>Regulatory Assurance*</b>	<ul style="list-style-type: none"> <li>• Limited or no interaction with proponent and resource experts reduces confidence</li> <li>• Depends on TI team and degree to which data, models, and methods are readily available</li> </ul>	<ul style="list-style-type: none"> <li>• Individual interaction with proponent and resource experts increases confidence over Level 1</li> <li>• Depends on TI team; degree to which data, models, and methods are readily available; and success in obtaining additional information and understanding from individual interactions</li> </ul>	<ul style="list-style-type: none"> <li>• Interaction among proponent, resource, and evaluator experts in facilitated workshops greatly increases confidence over Level 2</li> <li>• Documentation of evaluation and integration process by TI Team key to high levels of confidence</li> </ul>	<ul style="list-style-type: none"> <li>• Interaction among proponent, resource, and evaluator experts in facilitated workshops greatly increases confidence over Level 2</li> <li>• Documentation of evaluation and integration process by evaluator experts key to high levels of confidence</li> </ul>
<p>*Regulatory Assurance is defined as confidence that views of the larger technical community has been considered and that the centre, body, and range of technically defensible interpretations has been represented.</p>				

Many studies have been carried out in compliance with the SSHAC guidelines since its publication, providing valuable insights to make reliable decisions regarding the protection of critical infrastructures. The focus of the studies in practice has been on the implementation of Level 3 and Level 4 of SSHAC processes (Hanks et al. 2009) and their implementation has shown to be an effective and efficient procedure for the protection of nuclear critical infrastructures (Hanks et al. 2009). The motivation behind the implementation of higher levels is mainly because of higher demand of nuclear regulatory assurance.

The process originally focuses on the theory and practice of conducting PSHA where background knowledge in the assessment is represented in a transparent manner by multiple experts. Examples of applications of this kind are the EPRI GMC study (Level 3) for eastern North America conducted by EPRI (EPRI 2004) for updating applicable ground-motion models in light of significant new information; the CEUS SSC Project (Level 3) to develop a stable and long-lived CEUS SSC and to provide comprehensive seismic-source model for the entire CEUS (EPRI 2008); the Thyspunt site in South Africa (Level 3) for evaluating a potential nuclear power plant location (Bommer et al. 2013); the Yucca Mountain (Level 4) for use in seismic design of preclosure facilities (prior to the closure and sealing of the repository) and evaluations of the performance of the repository system following closure (CRWMS M&O 1998); the PEGASOS project for existing nuclear power industry in Switzerland (Abrahamson N.A. et al. 2002; Coppersmith, Youngs, and Sprecher 2009). The implementation of SSHAC guidelines has also been extended also to as in the Yucca Mountain PVHA (Level 4) employed for Yucca Mountain (Coppersmith K.J. et al. 1996; CRWMS M&O 1996) regarding the designated radioactive-waste repository, first was completed in 1996 and an update was carried out in 2008 Sandia National Laboratories.

## **2.6 CRITICAL COMPARISON OF APPROACHES**

### **Use of expert opinion**

Generally, each expert is specialized in a particular area of research, but performs poorly when a study requires expertise on interdisciplinary or multidisciplinary areas of research. By combining several such experts, the overall performance can be boosted significantly.

MEI and cEE allow the direct involvement of experts in the definition of the problem, with the goal of providing a consensus result. Even if many potential definitions of consensus may be applied (see Bedford et al. 2001), the two methods differ substantially on how to treat the interaction among the experts (SSHAC 2013). MEI requires the explicit formulation of multiple models and ask the experts to discuss about specific assumptions and modelling

choices. On the contrary, cEE typically (but not always) assumes the independence of experts, and thus specific interactions among experts are not allowed in order to avoid cross-contamination. When explaining the object of the elicitation in cEE, there will be for sure some discussion and interaction, but typically the facilitator tries to keep them as minimum as possible. This difference emerges from a different interpretation on the participants' expertise. In cEE, the objective is to obtain answers to well-defined questions from carefully selected experts. In such a process, the answers are assumed to already exist in the minds of the experts and so the goal is skilful extraction of this information; the expert assessments are then combined with calibration of the experts, which allows ascertaining which answers constitute the most reliable data. In other words, the experts are treated as repositories of knowledge, and it is assumed that the knowledge only needs to be elicited from them. This sits in marked contrast to the MEI processes that expressly encourage and foster structured interactions among experts during the assessment process (SSHAC 2013). In MEI, the evaluators, chosen in part because of their appropriate technical backgrounds, actually become experts in the course of the project as they evaluate project-specific data and learn from the interactive process. Subject matter experts are asked to participate in an interactive process of ongoing data evaluation, learning, model building, and, ultimately, quantification of uncertainty (SSHAC 2013). Note however that atypical cEE, targeting to have interactions in a controlled environments. Examples are the Delphi-like methods, based on cyclical elicitation procedures (Selva et al. 2012).

Another important difference between MEI and cEE is that MEI targets to the explicit definition of multiple acceptable models, letting them to emerge from the discussion among the experts. This process is typically not required in cEE, where more often the final target values (e.g., probabilities, parameters) for a predefined model are elicited (e.g., Neri et al 2008; Selva et al. 2012). This may limit somehow the potential use of cEE results. Indeed, several important methodologies for managing and analysing hazard and risk results require the explicit definition of multiple models, as for example the disaggregation process to trace back a potential consequence to its potential causing sources. However, it is worth noting that this difference has also important implications in the budget required to perform either MEI or cEE. Indeed, a cEE process relies mainly in carefully preparing the questionnaires to be filled by experts. This potentially long process is made by a restricted group, while the whole group of experts is involved only at the end. This clearly may help in limiting the process's budget. Therefore, cEE has the clear advantage of reducing the costs of the analysis. This advantage may be particularly important when experts are asked to define issues for which they may have experience on, in which case the discussion process that

MEI requires may be redundant. In these cases, a cEE may be the more effective solution to obtain the very same results.

The explicit formulation of models emerging from the group of experts that MEI considers makes MEI formally dependent from an 'integration' model allowing combining the multiple-models. LT and BEA represent two potential candidates for dealing with this process. In other words, while LT and BEA do not necessarily consider as input a set of models arising from the discussion among experts within a MEI process, they do represent a methodological solution required for MEI. Indeed, MEI can be seen as the process of defining the set of alternative models, and relative weights, to be input in LT or BEA.

### **Integrating alternative models**

The integration of alternative models is made implicitly in cEE and explicitly with LT or BEA. Indeed, the experts participating to a cEE do know potential alternative models, since they are expert of the target assessment, and they clearly consider them when they answer the questionnaires. Conversely, the explicit definition (either through MEI or simply from literature) of models is the starting point for LT and BEA.

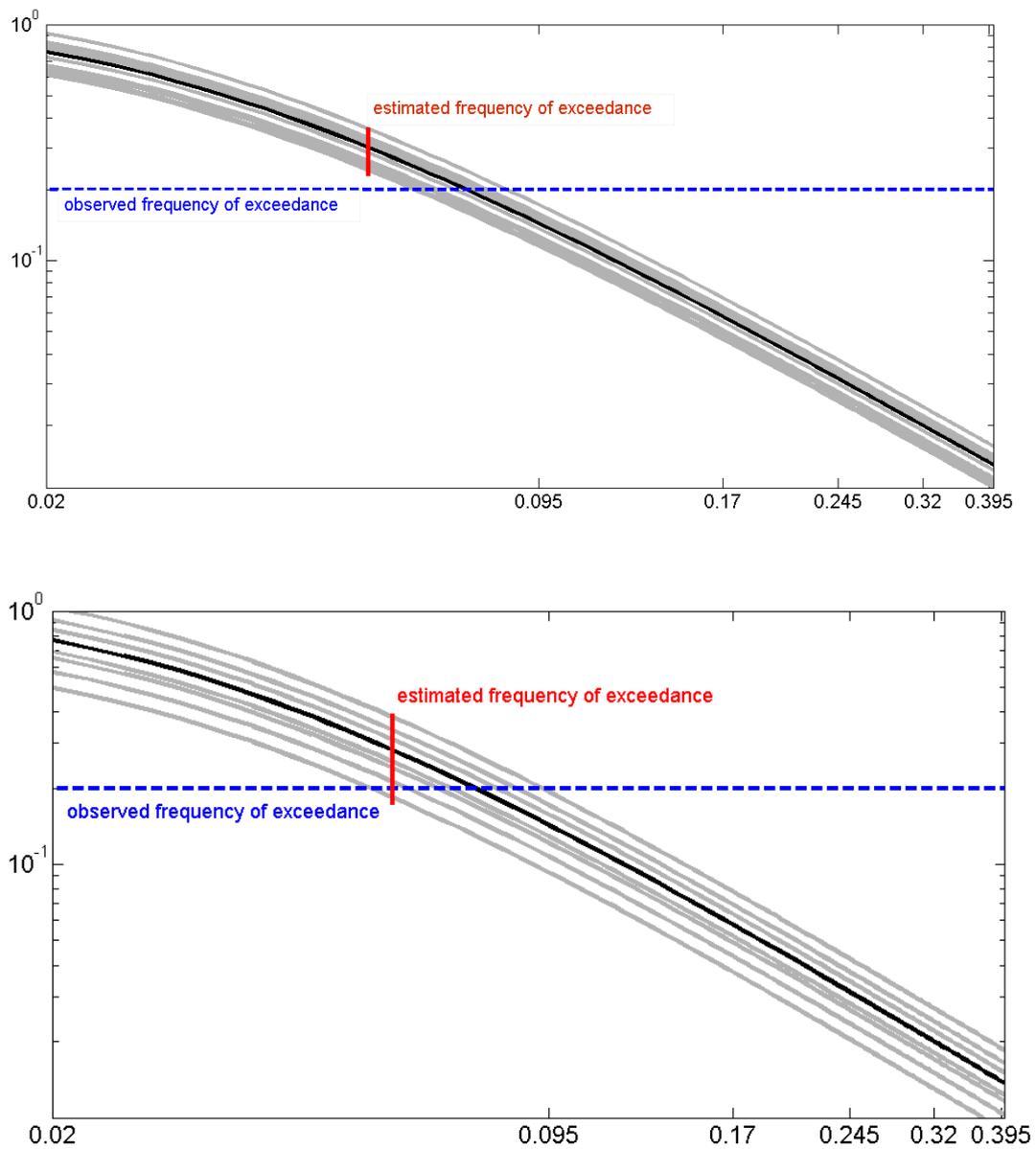
Both LT and BEA get as input a set of alternative model formulations for the assessment. In particular, BEA provides a continuous distribution assuming that input models are an unbiased sample of the epistemic uncertainty; this sample is then used to infer the parent distribution, whose central value represents the best guess of the aleatory variability, and the dispersion around the central value mimics the epistemic uncertainty (Marzocchi et al. 2015). Ensemble modeling assumes that models are independent or that the weights associated to each model account for possible correlation between models (Bommer and Scherbaum, 2008; Marzocchi et al., 2012). Conversely, LT assumes that the set of models represents a mutually exclusive and collectively exhaustive (MECE) set of events. This means that the models must be independent and that weights have a strict probabilistic interpretation. This also means that the only statistically meaningful result of a LT is the mean value, that is, the mean of the alternative models with respect to the epistemic uncertainty (Musson, 2005; 2012).

Many practitioners interpreted the results of LTs in terms of percentiles. In practice, they give more emphasis to the full discrete distribution of the final branches outcome using the percentiles (e.g., Abrahamson and Bommer, 2005; Stucchi et al., 2011; Field et al., 2014). This peculiar way of interpreting the outcome of a LT is in contrast with its probabilistic framework (Musson, 2005, 2012), but it can be theoretically justified in the framework of

BEA. Indeed, Ensemble modelling does not replace necessarily the logic tree, but it solves the largest part of the misconceptions and discussion about its use in PSHA, and it provides a framework to interpret the logic tree structure and outcomes. In this framework the logic tree does not have the probabilistic structure of classical probability tree and it is just a convenient tool to produce alternative hazard models which are meant to sample the epistemic uncertainty. Conversely, the ensemble modeling is much more flexible than the logic tree structure, because it also considers models that do not fit the hierarchical structure of the logic tree (Marzocchi et al., 2015).

The use of distributions instead of single mean values (coming from LTs used as alternative trees, or directly from BEA) has also important consequences in the testability of the results. Marzocchi and Jordan (2014) show that the practice of using only the mean to test hazard models (e.g. McGuire and Barnhard, 1981; Stirling and Petersen, 2006; Albarello and D'Amico, 2008; Stirling and Gerstenberger, 2010) leads often to the rejection of good models. An example of this is reported in Figure 2-13. In this example, two models with the same mean are compared with an observed frequency of exceedance. If only the mean model curve is considered in the testing phase, both models should be rejected. Conversely, if we look also at the distribution for the epistemic uncertainty, the first model should be rejected, while the second model not. In other words, testing the distribution allows to test the accuracy of the epistemic uncertainty distribution, while the use of the mean is equivalent to assume an infinity precision of the assessment.

Other important consequences of the use of distributions are the possibility of a more aware comparison and interpretation of the results, and of showing to the stakeholders the sites with the highest epistemic uncertainty, i.e., the sites where future large variations of the mean hazard are more likely (e.g., Paté-Cornell 1996; Marzocchi et al. 2015).



**Figure 2-13:** An example of testing two hazard models with the same mean and different epistemic uncertainty against an observed frequency of exceedance.

## 3 Selection of hazards and hazardous phenomena

### 3.1 INTRODUCTION

Different types of natural hazards and hazardous phenomena exist. Most of them occur individually, but many of them are interrelated in which one hazard or phenomenon can cause the direct and indirect occurrence of another (Gill and Malamud 2014; Smith 2013), called multi-type natural hazards. The occurrence of an individual or of multi-type natural hazards poses serious threat to both nuclear and non-nuclear Critical Infrastructures (CIs).

In order to evaluate the safety and resilience of CIs against different types of natural hazards, a transparent and comprehensive hazard/risk assessment methodology (Stress Tests) is adopted (Alzbutas et al. 2005; Pfister et al. 2012; Zio and Ferrario 2013; Zwicky, Fajfar, and Giardini 2014). Generally, the risk assessments are carried out for individual natural hazards to a specific CI, but the attention has been diverted to the assessments of potential multi-type natural hazards, since the incident of *Fukushima Daiichi Nuclear Power Plant* (NPP) happened in 2011. Moreover, the methodologies of assessing risk of single hazards are well established and found in literature, but methodologies and tools to tackle multi-type hazard and risk are being developed focusing the safety of different CIs (Komendantova et al. n.d.).

In the nuclear sector, IAEA (International Atomic Energy Agency) and ENSREG (European Nuclear Safety Regulations Group) provide useful guidelines for performing the hazard and risk assessment (Stress Tests)<sup>2</sup> of a variety of hazards for the construction and safety of nuclear CIs highlighted within STREST project (Zwicky et al. 2014). If we look up the guidelines of ENSREG and IAEA, we encounter a diverse list of natural hazards for assessments. Of course, the list varies from one specific site to another. IAEA suggests taking into account almost all natural hazards and ENSREG suggests most of the EU-concerned hazards. Regulatory concerns and available sources in the nuclear sector might allow considering the treatment of most of the hazards in terms of assessing hazard and risk, but it might not be possible in the non-nuclear sector compared to nuclear sector due to low regulatory concerns and limited available funding, Given that the resources are limited, it

---

<sup>2</sup> Note the notion Stress Tests (STs) is adopted by ENGREG for the transparent and comprehensive risk assessment.

is imperative to prioritize the natural hazards/phenomena of interest for the assessments for any non-nuclear CI in question that could be an existing one or a site evaluation for building a new one.

Stress Tests are essentially required to construct and protect the non-nuclear CIs. IAEA and ENSREG guidelines provide useful information to perform the tests; however, they do not provide sufficient information to select the most hazardous single and multi-type hazards for assessment as well as assessment procedure to conduct multi-type risk assessment of corresponding hazards. Therefore, two basic questions need to be answered before STs are performed challenging the safety of non-nuclear CIs which are:

- i) How to prioritize single and multi-type natural hazards for risk assessment? and
- ii) How to include multi-type hazards in their assessments of risk?

In order to answer the abovementioned questions, we explore what types of natural hazards are being identified and selected in the nuclear section (see *section 3.2*). Since the STREST project builds upon some of the ideas and methods developed in other European projects, therefore, we can benefit from the developed methodologies in the MATRIX project (Mignan et al., 2014). So, we look up the guidelines of an EU project MATRIX for the selection and assessment of multi-type natural hazards (see *sections 3.2 & 3.3*). In particular, in *section 3.3*, we describe the step-wise approach to conduct the hazard and risk assessment of multi-type hazards designed in the MATRIX and in *section 3.4* we propose two methods to prioritize the single and multiple-type hazards. These methods are Paired Comparison method (Bradley-Terry model) denoted by PC and Analytic Hierarchy Process denoted by AHP, which are based on expert judgments. Numerous applications of these methods are found in literature (Huang, Lin, and Weng 2004; Matthews and Morris 1995; Zahedi 1986). We foresee the potential applications of the PC and AHP methods and MATRIX approach in carrying out the EU@STREST for stress tests (see *section 4.3*) for non-nuclear CIs focused in STREST project.

## **3.2 EXISTING GUIDELINES**

### **IAEA Guidelines**

The International Atomic Energy Agency (IAEA) published a series of safety standards concerning the assessment of natural hazards for nuclear critical infrastructures (IAEA NS-

G-1.5 2003). The published series particularly emphasized on the natural events both external and human-induced for the protection of existing nuclear plant in terms of safety re-assessment (IAEA NS-G-2.10 2003) and for the site evaluation to install a new nuclear power plant (IAEA NS-G-1.2 2001). With regard to external events, no particular exhaustive list exists for the comprehensive risk assessments. It is emphasized to preliminary take into account all the significant hazards with identification of their potential hazardous phenomena and consequences for a particular site and plant. More specifically, all such hazards should be evaluated in accordance with specific requirements for safety requirements established by each EU member State. By looking at the guidelines of IAEA's safety standards, we find the following significant important hazards to be considered for the comprehensive risk assessments (IAEA NS-G-1.5 2003) which are:

- Earthquake;
- Extreme meteorological conditions (of temperature, snow, hail, frost, subsurface freezing and drought);
- Floods (due to tides, tsunamis, seiches, storm surges, precipitation, waterspouts, dam forming and dam failures, snow melt, landslides into water bodies, channel changes and work in the channel);
- Cyclones (hurricanes, tornadoes and tropical typhoons) and straight winds;
- Abrasive dust and sand storms;
- Lightning;
- Volcanism;
- Biological phenomena;
- Collision of floating debris (ice, logs, etc.).

The general description of abovementioned significant hazards, their related occurring phenomena and guidance to assess the hazard and/or risk against specific natural hazard, either for the site specific review or for the protection of nuclear CI, are given in different reports of IAEA (see: <http://www-ns.iaea.org/standards/>). For instance, Reference (IAEA NS-G-3.5 2003) gives guidance for a site specific review of the potential risk of flooding of a site due to diverse initiating causes and scenarios (and relevant potential combinations); Reference (IAEA NS-G-3.4 2003) gives guidance for a site specific review of extreme meteorological events as well as the potential risk of tropical (typhoons and hurricanes) and extra-tropical cyclones, both generated on the ground (tornadoes) and on seas or large water bodies (waterspouts); Reference (IAEA SSG-21 2012) gives the guidance on volcanic

activity that may affect the site. The evaluation of seismic risk in relation to earthquake and ground shaking scenarios is published in (IAEA NS-G-1.6 2003).

### **ENSREG Guidelines**

Similar to IAEA, there is no exhaustive list of natural hazards to be considered for the STs in ENSREG. However, their focus has been to deal with extraordinary triggering hazards. More specifically, three categories of the extraordinary triggering natural hazards are identified in the guidelines of ENSREG, i.e.

- Earthquake;
- Flooding; and
- Extreme weather conditions.

This identification is made by collected national reports under the National Action Plan (ENSREG 2013). Every member state per se is responsible under the National Action Plan to conduct the stress tests for the selected natural hazards and produce guidelines related to European nuclear critical infrastructures. The results of STs are shared with other EU members' states to seek harmonization in the applications of the tests for the better safety and protection of CIs.

### **MATRIX Guidelines**

Since the MATRIX in an EU based project, therefore it deals with single and multi-type natural hazards which are relevant to EU territory for the purpose of hazard and risk assessment. The MATRIX project guidelines provide the consideration of following hazards and their interactions:

- Earthquakes;
- Volcanic Eruptions;
- Tsunamis;
- Wild fires;
- Landslides;
- Meteorological events;
- Storms; and

— Flooding.

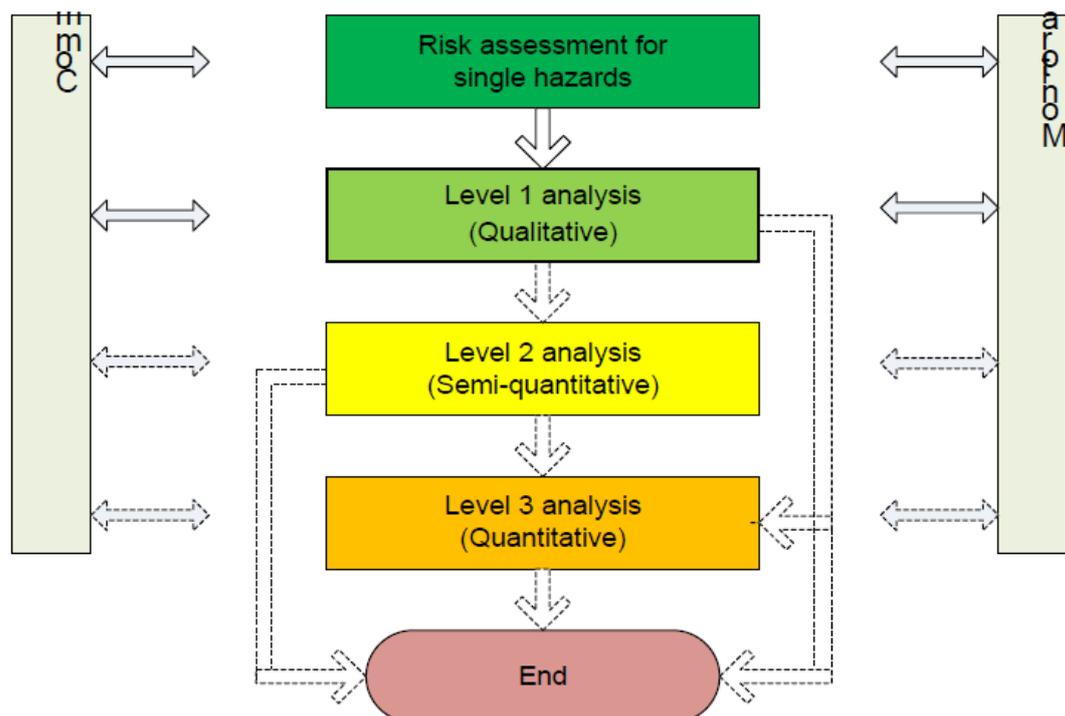
The interactions were classified into triggering events and cascading effects. The triggering events with possible direct interactions are shown in the matrix form shown in Table 3-1 and possible scenarios of cascading effects are shown in Figure 3-1. In the **Error! Reference source not found.**, the different hazards considered in the MATRIX are classified as triggering (running in the x-axis) against the 'triggered' (running in the y-axis) events. The figure also allows us to better understand the existing relationships between the different kinds of hazardous events and, their relative level in the chain. In this way, the occurrence of different phenomena may be considered from the possible triggering factors. In the MATRIX project, a framework for multi-hazard and multi-risk assessment methods for Europe has been developed and implemented to assess the hazard and risk of multi-type hazards. The short description of this framework is given in the section below.

### 3.3 MULTI-HAZARD AND MULTI-RISK ASSESSMENT METHOD

MATRIX project proposed various frameworks for multi-type hazard and risk assessment accounting for the possible spatial and temporal interactions (Mignan et al., 2014; Liu et al., submitted) that allows multi-type hazards and risks to be assessed in terms of their probabilities and the expected total losses. The general overview of the MATRIX is described below.



The overall framework is composed of four sophistication stages including three levels which are (I) risk assessment for single hazards, (II) Level 1: qualitative multi-risk analysis, (III) Level 2: semi-quantitative multi-risk analysis, and (IV) Level 3: quantitative multi-risk analysis. The schematic view of the each stage is depicted in the Figure 3-2.



**Figure 3-2:** Schematic view of the steps followed in the framework multi-risk assessment (Liu et al., submitted).

The stage 1 of the framework in the Figure 3-2 assumes the classical approach for assessing the hazards and risks of single hazard. In the other stages, the three levels of analyses are employed for the assessment varying from qualitative to quantitative analysis. The choice of employing any level depends on the context of study and level of treatment of epistemic uncertainty. The level of treatment of uncertainty means that epistemic uncertainty is most likely to be treated in a qualitative way at the level 1 and in a quantitative way at the level 3. Once the hazard or risk of single hazard is assessed, the assessor embarks on the three level processes. The description of these levels is described below.

*Level 1 analysis*) Level 1 analysis is essentially qualitative in nature comprising on comprehensive hazard identification to demonstrate that the activity does not pose a significant off-site risk. It consists of a flow type of questions that provides the end-user as to whether or not a multi-type assessment approach is required. These questions explicitly account for cascading hazards and dynamic vulnerability within the context of conjoint or successive hazards. An exhaustive list of answers is supplied to each question to be chosen by the user. Level 1 suggests whether higher levels of analyses are required or not. The user is suggested to skip the Level 2 analysis if the cascading phenomenon is potentially a serious concern.

*Level 2 analysis*) Level 2 supplements the qualitative analysis using semi-quantitative methods in which the interactions among hazards and dynamic vulnerability are assessed approximately. It is a matrix-based approach in which hazard interactions and time-dependent vulnerability are considered. A matrix is developed by means of the choice of a pair of hazards (considered as the basic components of the system followed by a clockwise scheme of interaction with the description of the mutual influence between different hazards. Once the description is inserted in the matrix, their degree of intensity of interaction is expressed by whole numbers (0 to 3) representing no interaction to strong interaction. Moreover, it is possible to verify the degree of the impact of each hazard on the others and the effect from other hazards. The hazard interaction index, which is the sum of the codes for all the off-diagonal terms, is evaluated and compared to a given threshold value in order to avoid the excessive weighting of a single hazard.

*Level 3 analysis*) Level 3 is a fully quantitative analysis to assess the interactions among hazards and dynamic vulnerability. A new quantitative multi-risk assessment model based on Bayesian networks (Liu et al., submitted) is introduced to both estimate the probability of a triggering/cascade effect and to model the time-dependent vulnerability of a system exposed to multi-hazard. A conceptual Bayesian network multi-risk model may be built based on Bayes' Theorem (Marzocchi et al. 2012). To determine the whole risk from several threats, the network takes into account possible hazards and vulnerability interactions. Another generic approach to treat multi-hazard assessment is a simulation framework based on the Monte Carlo method (Mignan et al., 2014). The former approach requires the definition of specific multi-risk scenarios while the simulation approach is more flexible; by realizing numerous time series and quantifying all hazard interactions via a hazard correlation matrix, the emergence of "surprise" cascading effects (i.e., not defined in advance explicitly) becomes possible. The simulation approach of Mignan et al. (2014) is tested in STREST

deliverable 3.5 for earthquake-earthquake interactions and cascades at dams (multiple interactions between natural hazards and operational hazards).

In the next section we briefly describe two methods to prioritize the natural hazards to a specific CI as a part of guidelines to perform STs. Moreover, these methods can be used to prioritize the components of a system (critical infrastructure) in risk assessment.

### 3.4 METHODS TO PRIORITIZE NATURAL HAZARDS

#### The Paired Comparison method

There are various paired comparison method (PC), but we choose the Bradley-Terry model that was introduced by Bradley and Terry (Bradley and Terry 1952). It is a statistical model that is applied in situations where the strength of an object needs to be evaluated with respect to another in pairs. The evaluation of strength of objects can be determined in terms of rankings and ratings, and can be described in the form of probabilities ( $Pr$ ) of possible outcomes.

If there are  $T$  objects to be ranked. For any two elements  $i$  and  $j$  of the set ( $T$ ), the Bradley and Terry model suggests that

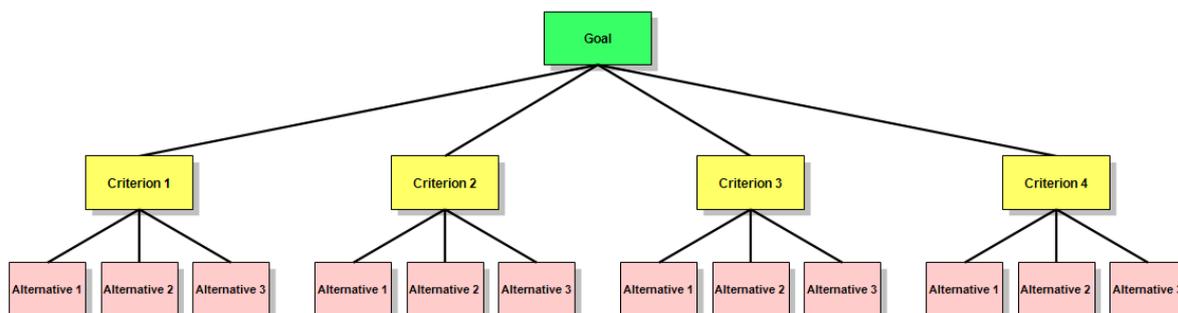
$$\Pr(i \text{ beats } j) = \frac{\lambda_i}{\lambda_i + \lambda_j} \quad (6)$$

where  $\lambda_k > 0$  is a parameter associated to element  $k \in \{1, \dots, T\}$  that represents the strengths of the objects. Here the objects that can be thought of are single or/and multi-type natural hazards as well as the components of a CI. In order to prioritize them, we ask experts to give their opinions on all possible paired of hazards or system's components what is required for the model. The experts' opinions are then processed using the Bradley and Terry model to get the probabilities of ranked outcome. Certainly, classical expert judgements (see *section 2.4*) can play an important role when the comparisons are made.

#### Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) developed originally by Thomas L. Saaty (Saaty 1990) is a multi-criteria decision-making method that is useful to make decisions under complex problems. The hierarchy process breaks down the complex decisions into a series of pairwise comparisons, synthesizes the results, and then helps to take into account both subjective and objective aspects of the decision. Additionally, the process incorporates a useful technique for checking the consistency of the expert judgements, thus reducing the bias in the process of decision making.

The process works by decomposing the decision-making problem into a hierarchy of evaluation criteria and alternative options among which the best decision is to be made. The best decision refers to the goal of the analysis. The structure of the hierarchy is shown in the Figure 3-3.



**Figure 3-3:** An example of hierarchy with three alternatives and four criteria for a specific goal (from Wikipedia).

The structure of modelling the decision in Figure 3-3 consists of an overall goal, a group of options or alternatives for reaching the goal, and a group of factors or criteria that relate the alternatives to the goal. It is important to note here that the criteria can be further broken down into subcriteria, sub-subcriteria, and so on, in as many levels as the problem requires.

The AHP can be implemented in three simple steps. First, a weight is generated for each evaluation criterion according to the experts' pairwise comparisons of the criteria. The importance in each criterion is judged by worth of the weight ( $w$ ). The higher the weight, the more important the criterion is. In the next step, for a fixed criterion, a score ( $s$ ) is calculated to each alternative according the expert's pairwise comparisons of the alternatives with respect to the criterion under consideration. The relative importance of one criterion/alternative over the other is usually expressed with numeric rating from one (equal important) to nine (extreme important) (Saaty 1990) and their pairwise comparisons are reported in the form of matrices to calculate the weights ( $w$ ) and scores ( $s$ ). Additionally, the weights and scores are assigned based on computing eigenvectors (Saaty and Hu 1998). At the final stage, the process combines the weights and scores of respective criteria and alternatives, thus determining a global score for each alternative, and a consequent desired ranking. The calculated global score for each alternative is a weighted sum of the scores obtained with respect to all the criteria.

Some inconsistencies may arise when many pairwise comparisons are performed (Harker and Vargas 1987), which is typically measured by Consistency Ratio (CR) (Saaty 1990). A perfectly consistent judgement by experts should always be zero i.e.  $CR = 0$ , however, inconsistencies are tolerated if  $CR < 0.1$  (Saaty 1990). Being the most critical issue in the application of AHP, the consistency ratio less than a cut off value of 0.1 is seen as a strict rule (Goepel 2013). It is therefore suggested to relax this cutoff value up to 0.3 depending on the number of criteria and the kind of project (Goepel 2013).

As an illustrative example, we demonstrate the mechanism of the AHP with the objective to prioritize natural hazards to a critical infrastructure. Five alternatives (a group of relevant options) and five criteria (a group of the most relevant factors) are hypothetically chosen listed here.

**Five alternatives:** Earthquake (H1); Tsunami (H2); Floods (H3); Extreme weather conditions (H4); and Volcanic eruption (H5).

**Five criteria:** The frequency of occurrence of natural hazards (C1); Intensity of natural hazards (C2); Geographical location of natural hazards (C3); Impact of natural hazards (C4); Critical infrastructure resilience against natural hazards (C5).

The matrix of experts' provided pairwise comparisons for five criteria is built based on the expert numeric values and written as

$$A = \begin{pmatrix} 1 & 7 & 3 & 1 & 1 \\ 1/7 & 1 & 1/7 & 1/5 & 1/5 \\ 1/3 & 7 & 1 & 1 & 1 \\ 1 & 5 & 1 & 1 & 3 \\ 1 & 5 & 1 & 1/3 & 1 \end{pmatrix}.$$

Note that the values in the upper triangular matrix of A are directly obtained from the expert and the lower triangular matrix of A is the reciprocal values of the corresponding values of upper triangular matrix. The corresponding weight vector, a computed eigenvector of the matrix A is  $w = (0.31 \ 0.04 \ 0.20 \ 0.22 \ 0.22)^T$ . Then based on expert judgments for five alternatives for a fixed criterion, the corresponding matrices are built and written as

$$B^1 = \begin{pmatrix} 1 & 1/4 & 4 & 1/6 & 1 \\ 4 & 1 & 1 & 1/3 & 5 \\ 1/4 & 1 & 1 & 3 & 1 \\ 6 & 3 & 1/3 & 1 & 1 \\ 1 & 1/5 & 1 & 1 & 1 \end{pmatrix}, B^2 = \begin{pmatrix} 1 & 4 & 1 & 1/4 & 2 \\ 1/4 & 1 & 3 & 1/3 & 5 \\ 1 & 1/3 & 1 & 1 & 3 \\ 1/4 & 3 & 1 & 1 & 1 \\ 1/2 & 1/5 & 1/3 & 1 & 1 \end{pmatrix},$$

$$B^3 = \begin{pmatrix} 1 & 1 & 3 & 3 & 1/4 \\ 1 & 1 & 5 & 1 & 1/3 \\ 1/3 & 1/5 & 1 & 2 & 1 \\ 1/3 & 1 & 2 & 1 & 2 \\ 4 & 3 & 1 & 1/2 & 1 \end{pmatrix}, B^4 = \begin{pmatrix} 1 & 1/2 & 3 & 2 & 1 \\ 2 & 1 & 7 & 1 & 1/5 \\ 1/3 & 1/7 & 1 & 2 & 1 \\ 1/2 & 1 & 1/2 & 1 & 1/3 \\ 1 & 5 & 1 & 3 & 1 \end{pmatrix} \text{ and}$$

$$B^5 = \begin{pmatrix} 1 & 1 & 3 & 1/4 & 1/7 \\ 1 & 1 & 3 & 2 & 1 \\ 1/3 & 1/3 & 1 & 1/5 & 3 \\ 4 & 1/2 & 5 & 1 & 2 \\ 7 & 1 & 1/3 & 1/2 & 1 \end{pmatrix}.$$

The corresponding score vectors the computed eigenvectors of  $B^1$ ,  $B^2$ ,  $B^3$ ,  $B^4$ , and  $B^5$  are  $s^1 = (0.16 \ 0.24 \ 0.20 \ 0.29 \ 0.11)^T$ ,  $s^2 = (0.20 \ 0.23 \ 0.14 \ 0.33 \ 0.10)^T$ ,  $s^3 = (0.21 \ 0.20 \ 0.13 \ 0.16 \ 0.29)^T$ ,  $s^4 = (0.19 \ 0.25 \ 0.12 \ 0.10 \ 0.34)^T$ ,  $s^5 = (0.19 \ 0.22 \ 0.16 \ 0.23 \ 0.20)^T$ . Hence the score matrix is

$$S = (s^1 \ s^2 \ s^3 \ s^4 \ s^5) = \begin{pmatrix} 0.16 & 0.20 & 0.21 & 0.19 & 0.19 \\ 0.24 & 0.23 & 0.20 & 0.25 & 0.22 \\ 0.20 & 0.14 & 0.13 & 0.12 & 0.16 \\ 0.29 & 0.33 & 0.16 & 0.10 & 0.23 \\ 0.11 & 0.10 & 0.29 & 0.34 & 0.20 \end{pmatrix}$$

and the global score vector defined as  $v = S.w$  is  $(0.18 \ 0.23 \ 0.16 \ 0.21 \ 0.12)^T$ . This shows that the second alternative, tsunami (H2), is of high importance, and extreme weather conditions (H4), earthquake (H1), floods (H3) and volcanic eruption (H5) receive the second, third, fourth and the fifth priorities respectively.

One important issue in AHP is to aggregate the judgements if many experts are involved in the process. Different approaches are employed to aggregate their individual or group opinions, however, weighted geometric mean is the most common aggregation method considering the equal or subjective weights on experts (Goepel 2013; Zio 1996) and will be employed in order to aggregate judgements for the prioritization of hazards. One challenge still remains concerning the assignation of more realistic weights to experts in the aggregation method under consideration. One possible way is to employ the AHP in parallel to derive weights on different experts with regards to the most relevant aspects of the judgemental process and performance (Zio 1996). The relevant aspects can be categorized such as experts' performance measure, experts' source of information, experts' past information etc.

## **4 Guidelines for treating epistemic uncertainties in Stress Tests (ST) for non-nuclear CIs**

### **4.1 INTRODUCTION**

Moving toward a safer and more resilient society requires improved and standardized tools for hazard and risk assessment of low probability-high consequence (LP-HC) events, and their systematic application to whole set of classes of CIs, targeting integrated risk mitigation strategies. The consistent design of Stress Tests (STs) and their application to specific infrastructures, to classes of infrastructures as well as to whole systems of interconnected infrastructures, is a basic step required to verify the safety and resilience of individual components as well as of whole systems.

Within the project STREST, a specific ST methodology is being designed in WP5, for testing the vulnerability and resilience of individual European non-nuclear CIs and infrastructure systems. The methodology is mainly based on single and/or multi-risk Probabilistic Risk Analysis (PRA), disaggregation, and/or Scenario-Based Risk Analysis (SBRA) for potential events that cannot be included in PRA. This methodology considers several ST-levels, foreseeing different levels for the management of epistemic uncertainty. In particular, the following levels are foreseen:

- ST-level 1 “Design basis check”: structural design check
- ST-level 2a “Re-evaluation of design bases events I”: single-hazard PRA
- ST-level 2b “Re-evaluation of design bases events II”: single-hazard PRA + epistemic uncertainty
- ST-level 3a “Evaluation of beyond design basis events I”: single-hazard PRA + epistemic uncertainty
- ST-level 3b “Evaluation of beyond design basis events II”: single- and multi-hazard PRA + epistemic uncertainty
- ST-level 3c “Evaluation of beyond design basis events III”: SBRA for selected thinkable scenarios (based on epistemic uncertainty)

Different analyses related to epistemic uncertainty are required at the different ST-levels. In particular, even if all hazard/risk analyses are affected by a large epistemic uncertainty, their

explicit quantification is not required at ST-levels 1 and 2a. At ST-levels 2b, 3a, and 3b, epistemic uncertainty quantification in the context of PRA is foreseen, involving the analysis of events in the tails of hazard and risk curves (in particular for ST-levels 3a and 3b). At ST-level 3c, epistemic uncertainty regards essentially the selection of “beyond design basis events” that cannot be explicitly included into PRA, but may represent an input for SBRA.

In this section, we present a harmonized and consistent approach to treat epistemic uncertainty within the framework of the STREST ST methodology. However, the approach can be employed also in other contexts, where a moderate level of public concern is required (as for non-nuclear CI, target of STREST). This coherent approach, hereinafter referred to as EU@STREST (Epistemic Uncertainty in STREST) process, follows the state-of-art methodological, philosophical and procedural guidance obtained i) from ENSREG and IAEA described in Deliverable 2.2, and ii) the review of existing methods for quantifying epistemic uncertainty described above. The EU@STREST process is designed to ensure an improved, standardized and robust treatment of epistemic uncertainty within STs of critical infrastructures.

In order to introduce EU@STREST, we first analyse (*section 4.2*) the potential use of the methods discussed in *chapter 2* and *chapter 3*, with reference to the ST-levels defined in the STREST methodology. Then, we discuss the EU@STREST process (*section 4.3*) that proposes the procedure to be followed for allowing a robust quantification of epistemic uncertainty within a hypothetical project aimed to perform a ST. Here, we develop the general framework, to be documented when it applies to one critical infrastructure in WP6.

## **4.2 METHOD SELECTION FOR STRESS TESTS**

Stress Tests are performed for evaluating the resilience of CIs to potential perturbations, in STREST limited to the ones coming from natural hazards. Significant unwanted consequences due to damages in a CI mainly occur for two reasons. They may be related to design deficit with respect to the effective hazards, or they may be connected to the occurrence of “unlikely” events, that is, events with a low probability of occurrence. In the ST-levels described above, the epistemic uncertainty quantification is mainly devoted to treat such unlikely events.

In this section, we discuss the use of the methods discussed in *chapter 2* and *chapter 3* in the framework of the ST methodology adopted in STREST and its ST-levels. First, we discuss them in terms of their applicability within the methodology (*section 4.2.1*). Then, we

specifically focus on the problem of analysing low probability and/or high consequence events, that is, of exploring the tails of hazard and risk curves (*section 4.2.2*).

#### 4.2.1 Methodological requirements

At all ST-levels, STs may eventually be linked to specific policies to make CIs more resilient to natural hazards, so STs are of *regulatory concern*. In addition, ST-levels 2b, 3a and 3b are based on Probabilistic Risk Assessment (PRA) and hazard/risk disaggregation, and the epistemic uncertainty analysis should allow for those procedures. Then, at all ST-levels, a certain degree of subjectivity is present in *selecting the phenomena to be modelled*. Finally, at all ST-levels, but especially at ST-level 3c, we must face the problem of missing models, the models which are theoretically required, but their inclusion is made impossible by substantial lack of procedures or of basic knowledge.

#### Regulatory concern

MEI (Multiple-Expert Integration, *section 2.5*) has been developed to define the processes to be followed to provide scientific quantifications of interest for regulatory concerns, within the context of nuclear facilities. The basic idea of MEI is that the involvement of multiple experts must increase as the regulatory concern of the assessment increases, and that model's variability should emerge from the discussions among the experts.

ST-levels 1 and 2a essentially require the use of available assessment, not specifically implemented for the target CI, in order to perform testing with low budgets. In this case, the regulatory concern is essentially managed through the selection of the *reference* assessments and the use of their *best guess* assessment (central value). This selection is then the key for managing the epistemic uncertainties that, even if they always exist, are not quantified at these levels.

For ST-levels 2b, 3a, 3b and 3c, the explicit quantification of epistemic uncertainty is critical, with explicit reference to a target CI. Also in this case potential reference assessments may exist, but they are mostly used for comparison than as input of the analysis. In this case, the primary input for the ST comes from a pool of experts, which have experience in the required hazard/risk assessments and/or in the CI. The key for managing uncertain choices is the involvement of a pool of experts in taking these choices letting alternative models to emerge from the discussion among experts. In other words, such ST-levels require a MEI method that takes into account the potential limitation in budget for ST of non-nuclear CI. This MEI

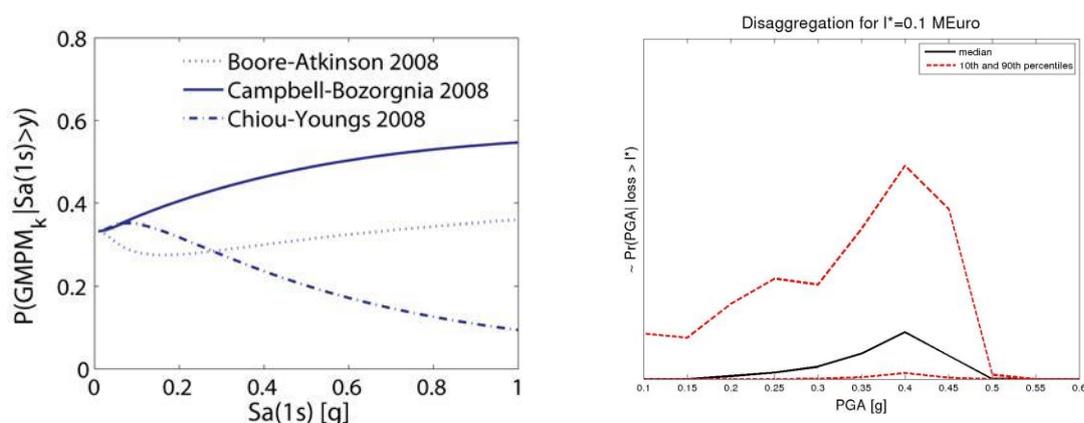
process can be seen as a coherent framework in which the use of the specific techniques discussed in *section 2* and *section 3* is regulated.

In *section 4.3*, we propose the EU@STREST process that is a MEI process specifically developed for non-nuclear CI and that enables to enforce ST at all the ST-levels.

### Hazard/risk disaggregation and uncertainty propagation

Hazard/risk disaggregation is required at ST-levels 2a, 2b, 3a and 3b. Hazard/risk disaggregation is the process of identifying the most probable causes for a selected event of interest, defined in terms of hazard intensity at site (hazard disaggregation) or loss level in the CI (risk disaggregation). This process may help in focusing risk mitigation actions, like targeted monitoring or retrofitting plans. Hazard and risk disaggregation is based on the inversion of probabilistic hazard and risk equations (e.g., Bazzurro and Cornell 1999) and it requires the explicit formulation of hazard and risk models.

Considering the methods discussed in *chapter 2*, only LT (Logic Tree) and BEA (Bayesian/Ensemble Approach) can enable for hazard/risk disaggregation considering epistemic uncertainty (see discussion in *section 2.6*), as required at ST-levels 2b, 3a and 3b. Indeed, both LT and BEA are based on the explicit formulation of models that enable to track back a selected consequence to its possible causative events. In both cases, the disaggregation may be performed on the single model alternatives adopting the procedure presented in Bazzurro and Cornell (1999), and epistemic uncertainty in disaggregation is quantified by analysing the variability of the results that emerge from the set of alternatives. In Figure 4-1, we report several examples of this.



**Figure 4-1:** Examples of disaggregation reporting alternative models to describe epistemic uncertainty, in the case of LT (left panel, from Lin and Baker, 2011) and of BSA (right panel).

The propagation of epistemic uncertainty to risk results is theoretically possible for all the methods discussed in *section 3*. However, the effective propagation is affected by severe computational problems (Field et al., 2005). BEA theoretically overcomes this problem by allowing the use of limited samples of alternatives (Marzocchi et al., 2015). An equivalent process is possible for LT, by trimming the branches of the trees based on the results sensitivity tests. In either case, the most important problem is represented by the general lack of model alternatives for many potential hazards (e.g., spatial correlation models) as well as for many of the potential components of CIs (e.g., SYNER-G 2009-2013). We will specifically discuss this problem in paragraph *Missing Models*.

### **Selection of phenomena of interest**

A necessary step to implement any ST is the selection of the phenomena to be included in the assessments regarding both hazards and the CI modelling. Indeed, at all ST-levels, a certain degree of subjectivity is present in selecting the phenomena to be modelled.

In general, the hazard selection may consist of the selection of the potential hazardous phenomena to be considered (e.g., ground shaking, ground failure, tsunami waves, wind, etc.), their potential sources (e.g., different seismic zones, maximum magnitudes) and their potential combination and interaction (e.g., leading to phenomena acting simultaneously or in short-time intervals). At ST-levels 2b and 3a (single-hazard/risk analysis), the main target is only the analysis of the completeness of sources. On the contrary, the selection of hazards and their combination is critical for ST-level 3b (multi-hazard/risk analysis).

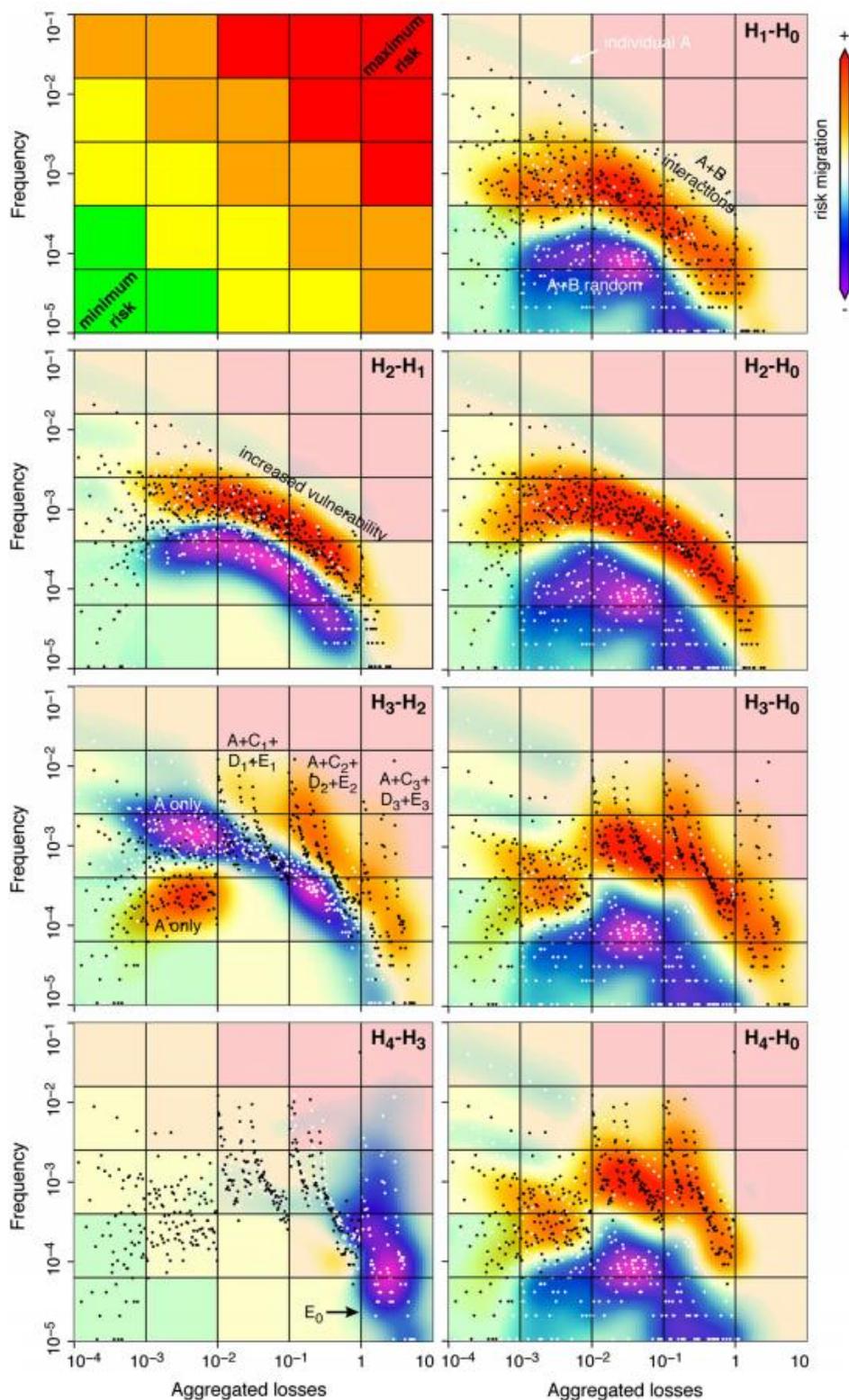
The selection of the phenomena to be investigated for the CI is focused on selecting the components to be modelled, their interaction, their failure modes, and the performance indicators that better describe the potential consequences due to damages on the CI. These choices are fundamental for any ST, at all the ST-levels.

The main goal of formalizing these choices is to avoid unwanted “dragon kings” (Sornette and Ouillon, 2012), that is, possible events in the tail of distributions due to a physical mechanism not represented in the distribution tail considered. This effect is clearly shown in Figure 4-2, where Mignan et al. (2014) discuss the changes in risk curves whenever several critical interactions among risks are included into the analysis.

This problem must be theoretically managed by a careful evaluation of all the possible paths to damages in a multiple-expert environment. In order to avoid unwanted biases induced by insufficiently scrutinized choices, it is necessary to include in any ST formal documented feedbacks from the larger community, with the goal of adopting participatory definitions of

the phenomena to be included in STs. The main assumption is that the pool of experts should be able i) to prepare a sufficiently large set of alternative models to fully explore the effective epistemic variability, and ii) to find and fill eventual lacks in the modelling procedures. Within the EU@STREST process, these fundamental feedbacks can be formally managed through a prioritization technique, like the PC (Paired Comparison) or the AHP (Analytic Hierarchy Process), discussed in *section 3.4*. These techniques are specific formulation within the general class of cEE (*section 2.4*), structuring the feedbacks from the experts' community and keeping the cost of ST at reasonable levels. Indeed, AHP enables the distillation of experts' opinion through formal *ad hoc* questionnaires, concentrating the main efforts into a restricted team preparing the questionnaires, and reducing by far the temporal involvement of the entire pool of experts (see *section 2.4*). At the same time, the participation of the larger community in critical choices regarding any ST is granted, enabling the production of participatory definitions that reduce the possibility of unwanted biases and make the results more defensible for *ex post* criticisms.

Of course, the feedback of experts should be linked also to sound science. In particular, the prioritization of hazards, of components and of modelling procedures for CIs should be based on the extensive use of *sensitivity tests*, as suggested by Bommer and Scherbaum (2008). One example of sensitivity test to be considered in order to better constrain the selection of the modelling approach is reported in *section 5.2.2*, where the impact on risk curves of modelling spatial correlations for a single hazard for a spatially extended CI is presented. Indeed, the results of these tests enable to specifically quantify how strongly one assumption can influence the ST results, increasing the awareness of the pool of experts in formulating their opinions.



**Figure 4-2:** Risk migration matrices showing how risk migrates as a function of frequency and aggregated losses when new information is added to the system. Risk increase in a given frequency-loss domain is represented in red and risk decrease in blue. Risk scenarios are represented by points, in white for the first hypothesis and in black for the second, taken from Mignan et al. (2014).

### Missing models

Any ST requires the formulation of a potentially large number of models, from hazard to fragility and serviceability for all the CI components, to models simulating the CI system (and system of systems) as a whole. In a multi-risk prospective, models simulating the interaction effects are also required (e.g., Selva et al., 2013; Mignan et al., 2014), in order to consider the impact of multiple-hazards as well as the effect of hazards on pre-damaged components. In order to complete ST when missing models exist, it is also required to set a procedure to fill these gaps.

When single models are required to complete a PRA analysis (ST-levels 2b, 3a and 3b), it is possible to set such models through cEE, following for example the Cook's method (see *section 2.4*). For example, if a single fragility model for one of the components of the CI is missing, either the parameters of a fragility curves with a pre-defined distribution function (e.g., log-normal for seismic fragility), or the probability of occurrence of damages at different levels of the hazard intensity can be elicited directly from the pool of experts.

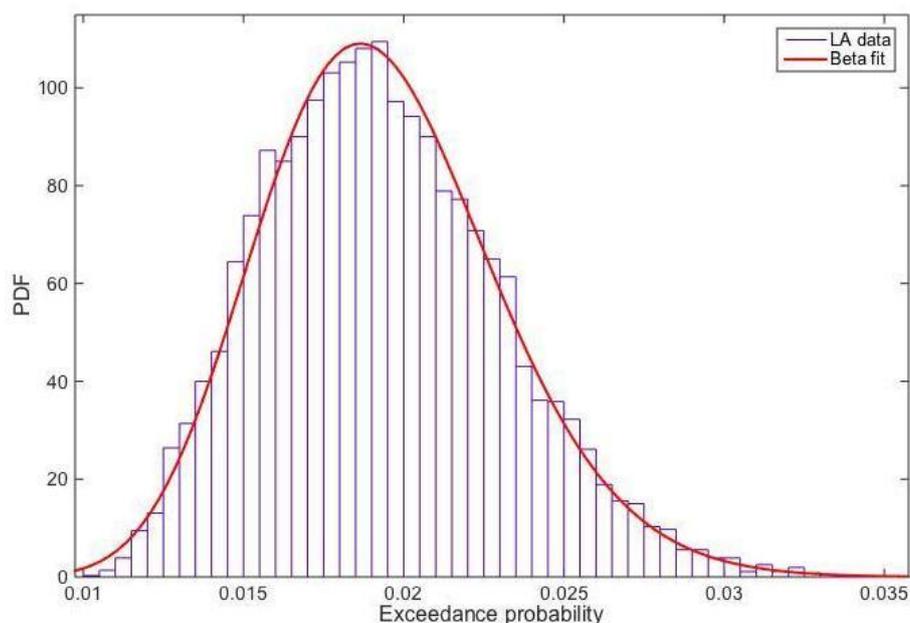
There are cases in which the formulation of models through cEE is quite complicated, or the number of required models is too large to be effectively implemented. For example, a few models exist for simulating the spatial correlations of seismic intensity, or for constraining the probability of occurrence of off-shore landslides that may cause potentially large tsunamis at the target site. In these cases, a complete exploration of epistemic uncertainty is not possible, given that alternative modelling approaches do not exist. If also a structured formalization of this problem within a cEE context result too complicated (because of the possible too large number of degrees of freedoms that should be elicited), the inclusion of such damage mechanisms is not possible within a reliable PRA that quantifies the epistemic uncertainty, that is, ST-levels 2b, 3a and 3b cannot be implemented. However, the simple fact of neglecting them in a ST creates the possibility of "grey swans" (events almost not expectable) or "dragon kings" (mechanisms not considered in the tails of the adopted models). A straightforward procedure to quantitatively include such events in ST is the one discussed for ST-level 3c: define specific "thinkable" scenarios to produce the input for SBRA. This means that experts are asked to define possible scenarios that, for whatever reason, cannot be included into PRA. Also this selection can be performed by applying the Cook's method of cEE (*section 2.4*), asking for the definition of the source parameters for a given specific event, or the probability of overcoming pre-defined intensity thresholds at site, or the quantification of an expected intensity at site for a pre-defined probability level (or mean return period). The practical implementation of a cEE process of this kind strongly

depends on the actual target assessment. However, the principle is that such definition(s) should arise from a pool of experts representing the larger community within a ST project.

#### 4.2.2 Exploration of tails in PHA and PRA

Generally speaking, “unlikely” events are events in the tails of distributions. Using the words of Mignan et al. (2014), “*anomalous events, or outliers, have been repackaged into fancier animals, some of which made their way into popular culture*”. In particular, Taleb (2007) called “black swan” to describe rare events, which in principle cannot be anticipated. Since an event that cannot be anticipated surely cannot enter to any quantitative hazard/risk assessment, we may re-coin this definition with “grey swans”, meaning those events that can somehow be anticipated, but that cannot enter a quantitative PRA. These events enter in the general category of “missing models”, and they have been already discussed in *section 4.1*. Another popular term is “perfect storm”, which refers to an event resulting of a rare combination of circumstances (Paté-Cornell, 2012). The analysis of “perfect storms” should be based on a careful exploration of the tails of distributions, which is discussed in this section.

Given that few data may be available on rare events, the analysis of tails should be based on a full exploration of epistemic uncertainty based on model alternatives that are well sampled (that is, they form an unbiased sample) from the larger community of experts, and the use of the testing procedures to analyse the compatibility of such models with data corresponding to the longest mean return periods (very low annual frequencies). With reference to the methods discussed in *chapter 2*, MEI is important to guarantee a comprehensive definition of the set of alternative models to explore epistemic uncertainty. Typical cEEs have the disadvantage of not defining explicitly the alternative models, making the disaggregation analysis impossible. In addition, most of cEE (with the exception of Delphi-like methods) typically require that the knowledge is already present in the expert, an assumption that is difficult to overcome when rare events are the target of the assessment. Therefore, the LT and BEA are the most likely candidates to explore the tails of distribution, when applied in the context of a MEI. Note that, hereinafter, we assume that the LT is adopted exploring the full distribution of alternative branches, and not only the mean value (see discussions in *section 2.2* and *section 2.6*).

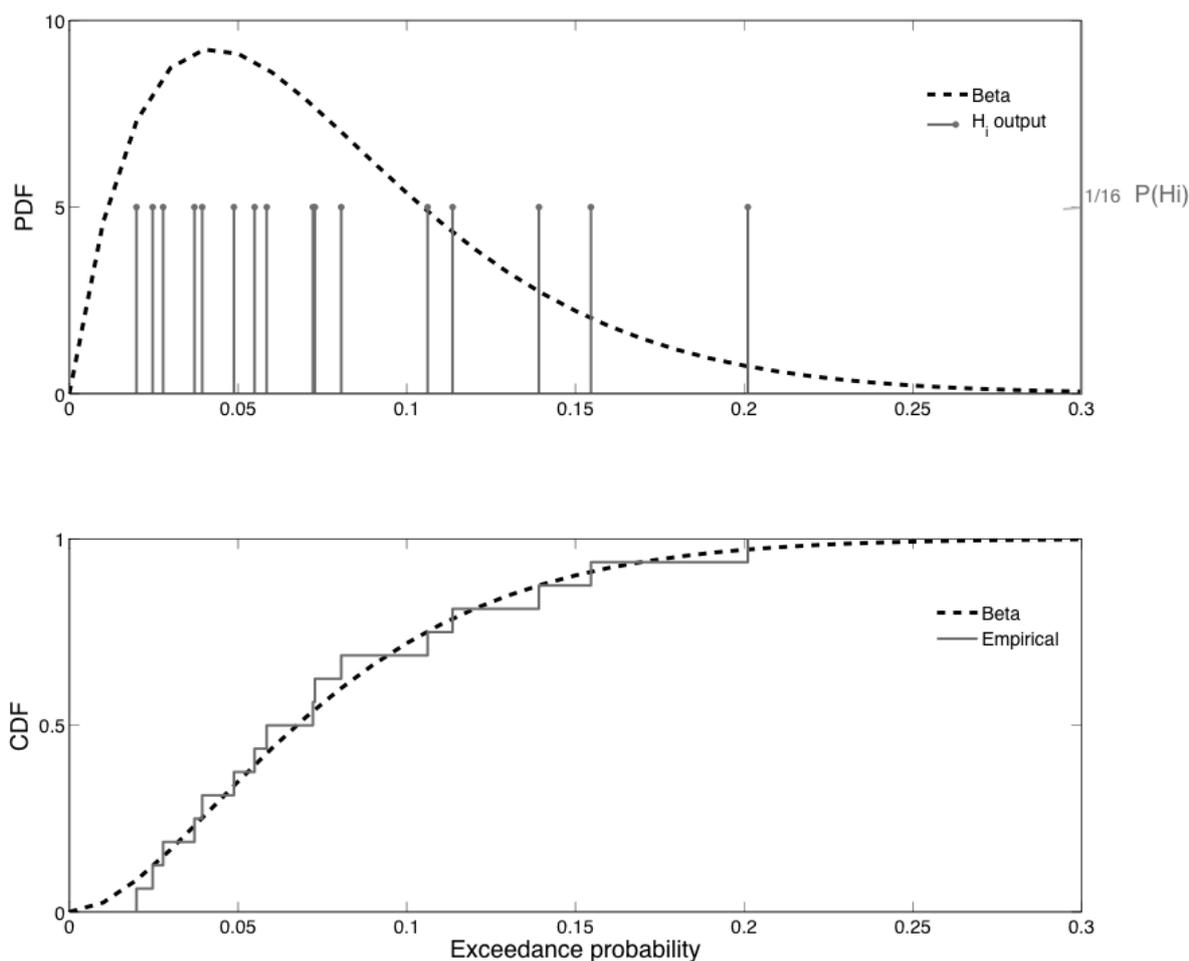


**Figure 4-3:** 200 exceedance probabilities relative to the average PGA for 2% in 50 years for Los Angeles, coming from one logic tree developed in the framework of UCERF3 (Field et al., 2014), as compared with an equivalent Ensemble model (from Marzocchi et al. 2015).

If many alternative models are considered, LTs and BEAs should theoretically perform equally. When many alternative models have been selected within a MEI process, the number of considered alternative models may be considered large enough to completely describe the epistemic variability. In this case, the distribution of the alternative models coming from a LT can be considered equivalent to the target distribution of BEA. In Figure 4-3, we report an example of this, where the histogram of LT alternatives is compared to an Ensemble model based on a Beta distribution that fits the same model alternatives with the same weights (Marzocchi et al. 2015).

However, the use of a large number of alternatives, besides being criticized by several authors (e.g., Bommer and Scherbaum 2008 and discussion therein), increases the computational problems posed by the propagation of epistemic uncertainty to risk assessments (e.g., Field et al 2003), as discussed in *section 4.2*. Therefore, in order to enable the propagation of epistemic uncertainty, the total number of alternatives should be somehow reduced. However, reducing this number, the LT alternatives only will poorly describe the epistemic variability. In this case, the ensemble model enables a better exploration of the variability. An example of this is reported in Figure 4-4, in which it is evident that the continuous distribution of the Ensemble model shows that most likely value of the exceedance probability is almost never associated with one branch (in contradiction

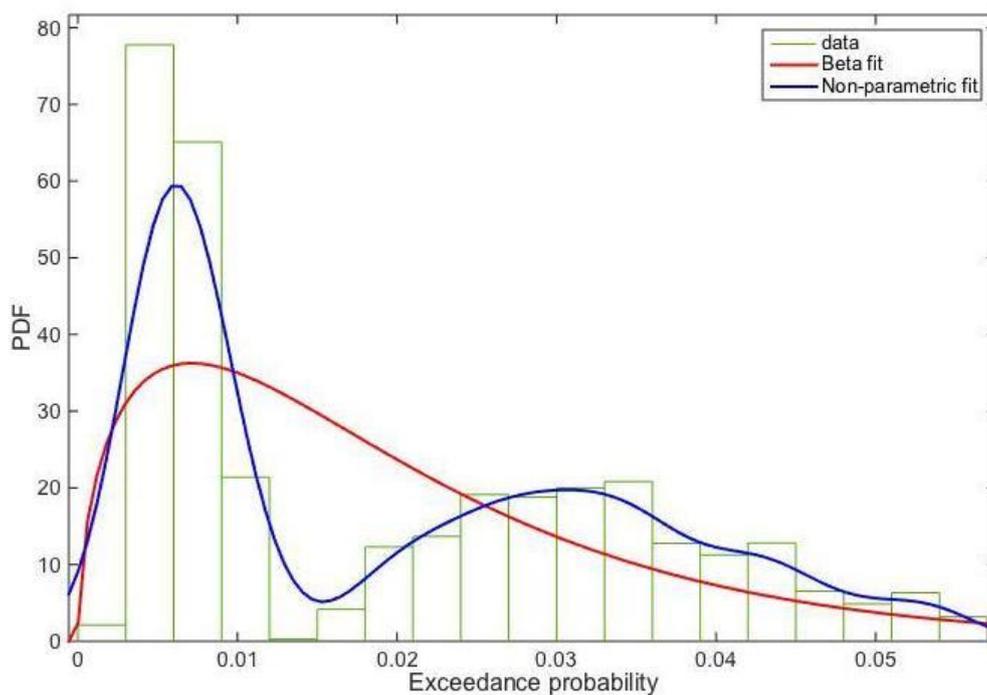
with the MECE assumption), and the probability that the true exceedance probability is lower (larger) than the lowest (largest) value of the LT is not negligible.



**Figure 4-4:** PSHA obtained by a simple logic tree composed by 5 different seismicity rate models arbitrarily selected from the Italian CSEP experiment (Schorlemmer et al., 2010), and three GMPEs, as compared with ensemble model (from Marzocchi et al., 2015)

A further complication may arise when a marked binomiality is found in the distribution of model alternatives. In this case, ensemble distribution may be assumed either unimodal or bimodal, as reported in Figure 4-5. The choice between the two possible ensemble models should be discussed depending on the cause of the binomiality. For example, if the two modes emerge from two end-cases and the values between these two cases are judged possible, a unimodal distribution may be more appropriate. On the contrary, if the two modes emerge from two alternatives that cannot consider intermediate cases, a bimodal distribution is more appropriate. Note that this possibility, which naturally emerges when BEA forces us to select the ensemble distribution, is typically neglected with LTs, where either only the

mean is used (so the problem is completely neglected), or the use of percentiles can be considered similar to the hypothesis of unimodality.



**Figure 4-5:** 200 exceedance probabilities relative to the average PGA for 2% in 50 years for Redding, coming from one logic tree developed in the framework of UCERF3 (Field et al., 2014), as compared with two potential Ensemble models (from Marzocchi et al. 2015).

If data relative to rare events at site are available, they should be considered either to set the model (e.g., Selva and Sandri 2013) or, at least, to test it (e.g., Albarello and D’Amico, 2008). A meaningful comparison between such rare data and the adopted (hazard) models can be performed only when the output results are reported in terms of distributions (Marzocchi and Jordan 2014), as discussed in *section 2.6*. Indeed, large epistemic uncertainties are involved in the assessment of the tails of distributions, and to test only the mean values may rule out models that are actually well sampling the epistemic uncertainty. In other words, models that do not explore the tails of epistemic uncertainty may be favoured, if such models actually over-fit the data without fully exploring the epistemic uncertainty. Indeed, in this case, the results may be precise but possibly not accurate. On the other hand, the goal of testing must be to test the whole distribution, and not only its mean, and the eventual rejection of one model must consider also the epistemic variability. This fact has been specifically discussed in *section 2.6* (see, for example, Figure 2-13). In practice, this means that, if a LT procedure is selected, the whole set of percentiles should be tested. However, due to the fact that few alternative models may have been considered in

order to make feasible the propagation to risk of epistemic uncertainty (see *section 4.2.1*), the definition of such percentile may be weak. This fact should be accounted for when hazard models are tested.

To conclude, both LT and BEA are feasible methods to evaluate the epistemic uncertainty on the tail of the distributions and to enable testing of models regarding rare events, once the set of alternative models and relative weights (input for both methods) are well defined. BEA presents several practical advantages that make it preferable. However, as discussed in *section 2.6*, we note that Ensemble modelling does not replace necessarily the logic tree, but it solves the largest part of the misconceptions and discussion about its use in PSHA, with particular reference to the analysis of the variability of LT outcomes through percentiles. So, the main difference is on how the results of an alternative tree are interpreted, and then BEA does not radically change the picture that a well-defined LT provides, but simply re-evaluates its actual outcomes.

### 4.3 EU@STREST: MULTIPLE EXPERT INTEGRATION FOR STRESS TESTS OF NON-NUCLEAR CRITICAL INFRASTRUCTURES

Multiple Expert Integration methods (MEI, *section 2.5*) become vital when the regulatory concerns are relatively high, as in conducting appropriate hazard and risk assessment for the robustness and resilience of nuclear and non-nuclear critical infrastructures (see discussion in *section 4.2*). In conducting such assessments, namely Stress Tests (STs), the consideration of the diverse range of views and opinions of experts, their active involvements in the assessments process, and their formal feedbacks can play a significant role in providing robust results on which to base solid regulatory decisions. Here we are concerned with the safety of European non-nuclear Critical Infrastructures (CIs) and develop a harmonized and systematic process to perform STs named EU@STREST, which stands for Epistemic Uncertainty at STREST. This process is based on a MEI method (see *section 2.5*) in which the epistemic uncertainties are explicitly addressed, quantified and communicated by means of a structured interaction of multiple experts.

The goal of the EU@STREST process is to produce guidelines for performing robust Stress Tests given that 1) relatively limited funding is typically available for any single CI, for both testing an existing one or designing a new one, and 2) important regulatory concerns may be involved. EU@STREST is focused to guarantee the minimum level of multiple-experts participation, within the given resource limitation, to deal with the treatment of the unavoidable uncertainties. Following the EU@STREST process, the output of a ST is expected to provide stable and robust results in terms of *“the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study”* (SSHAC 1997, 2013) for all the relevant threats concerning the CI. In other words, after a careful screening of all potential natural hazard initiators, the results should cover all the relevant threats and they should include best estimate (as supported by current data and models) together with an explicit estimate of the full range of uncertainty (resulting from the limitations of the currently available data and models) (see, *chapter 1*). In EU@STREST, particular emphasis is also given to transparency, independence and responsibility of the different actors participating in the process.

The process EU@STREST adapts to the different levels of Stress Tests (ST-levels), as defined in the ST methodology developed in STREST (WP5). In particular, EU@STREST process can be adopted for all such ST-levels, proposing specific technical solutions for the management of uncertainty quantification as required by all ST-levels. Of course, the

amount and quality of data to be considered, nature of models to be used and alternative methods to be employed in the analysis may vary from one ST-level to other ST-level.

The level of involvement of experts representing the informed technical community (SSHAC 1997, 2013) may largely vary within EU@STREST. Indeed, the requirement of keeping the budget for the analysis low forces to minimize (when necessary) the heavy involvement of large groups of experts. Thus, EU@STREST considers a certain degree of flexibility in how to include external expertise, depending on i) ST-level, ii) needs and requests of the CI authorities and/or of European or local regulations, and iii) the quality and quantity of disagreements among the experts on important issues regarding the ST.

#### **4.3.1 EU@STREST vs SSHAC process**

With reference to SSHAC (1997, 2013) levels, EU@STREST is roughly positioned between SSHAC-levels 2 and 3.

If this is not explicitly required by regulation or by special CI authorities' needs, external experts in EU@STREST (representatives of the informed technical community) are asked to provide only formalized feedbacks (*interviewing processes*) during the whole process, without active interactions among them. These interviewing processes are formalized through classical Expert Elicitation (cEE, for a discussion on the similarities and differences between SSHAC-MEI and cEE, see SSHAC 2013). In this case, the external experts form a pool that provides only inputs to the process. With reference to SSHAC (1997, 2013) levels, this configures EU@STREST roughly at SSHAC-level 2. However, single experts of the pool may also act as proponent and advocate a particular hypothesis or technical position (both regarding resources and methods) through individual communications at intermediate stages of the process. In addition, EU@STREST considers participatory review throughout the course of the project. Both these things are typical of SSHAC-level 3. Moreover, if during the first *interviewing processes* significant contrasts in experts' opinions emerge, formal and transparent interaction among the experts can be planned in order to seek consensus, transforming the pool in a panel. This significantly pushes EU@STREST toward SSHAC-level 3.

It is though worth noting that significant differences exist between EU@STREST and all SSHAC levels. Indeed, EU@STREST focuses on STs based on risk assessments (not only hazard) in a multiple hazard environment, including potential interactions and interconnectivities at the hazard and the CI levels. In addition, EU@STREST makes use of classical Expert Elicitation (cEE) tools. In this, EU@STREST and SSHAC processes differ

significantly, since cEE techniques are usually not considered in SSHAC. Consequently, they significantly differ in the selection of experts and in the process of integration of experts' opinion. In SSHAC processes the integration is mainly based on meetings among all the experts, who become 'experts on the topic' with such discussions that take place during the project (SSHAC 2013). In EU@STREST, the integration process is conducted mainly by the process scientific leader (technical integrator, see next section), based on the results of the aggregation of experts' opinions performed through cEE. This integration process is then independently reviewed throughout the project. In this sense, the technical integrator is the only actor of the process becoming 'expert of the topic' during the project. However, if significant and irreconcilable divergences among experts emerge from the cEE aggregation processes, the technical integrator may promote discussions among experts (closer to what SSHAC processes foresee), in order to assure a fair integration of such divergent opinions.

#### 4.3.2 Core actors

The EU@STREST process involves several types of actors which play an important role for the successful completion of the project. The core actors in the project are the Project Manager (PM), the Technical Integrator (TI), the Evaluation Team (ET), a Pool of Experts (PoE), and an Internal Reviewers (IR). The ET and IR may involve several participants, with different background knowledge, but in specific cases may be reduced to individuals. The size of such groups depends on the purpose and given resources of the study.

**Project Manager (PM):** Project manager is a stakeholder who owns the problem and is responsible and accountable for the successful development of the project. It is the responsibility of the PM that his/her decisions appear rational and fair to the authorities and public. The PM specifically defines all the questions that the ST should answer.

**Technical Integrator (TI):** The technical integrator is an analyst responsible and accountable for the scientific management of the project. The TI is responsible for capturing the views of the informed technical community in the form of a community distribution, to be implemented in the hazard and risk calculations. Thus, the TI explicitly manages the integration process. The TI should have i) expertise on managing classical Expert Elicitation (cEE), preparing questionnaires and analysing the results in order to manage the interviews to extract from the larger community feedbacks regarding critical choices/issues that any test involves (e.g., the selection of appropriate scientifically acceptable models); ii) experience in hazard and risk calculations; iii) experience in expert integration techniques, in order to

manage the quantification and the propagation of epistemic uncertainty out of acceptable models.

**Evaluation Team (ET):** The Evaluation Team is a group of analysts that actually perform the hazard and risk assessments required by the ST, under the guidelines provided by the TI. The team is selected by consensus of the TI and PM, and it may be formed by internal resources and/or external experts. The ET represents also the interface between the project and the CI authorities, guaranteeing the successful and reciprocal acknowledgement of choices and results.

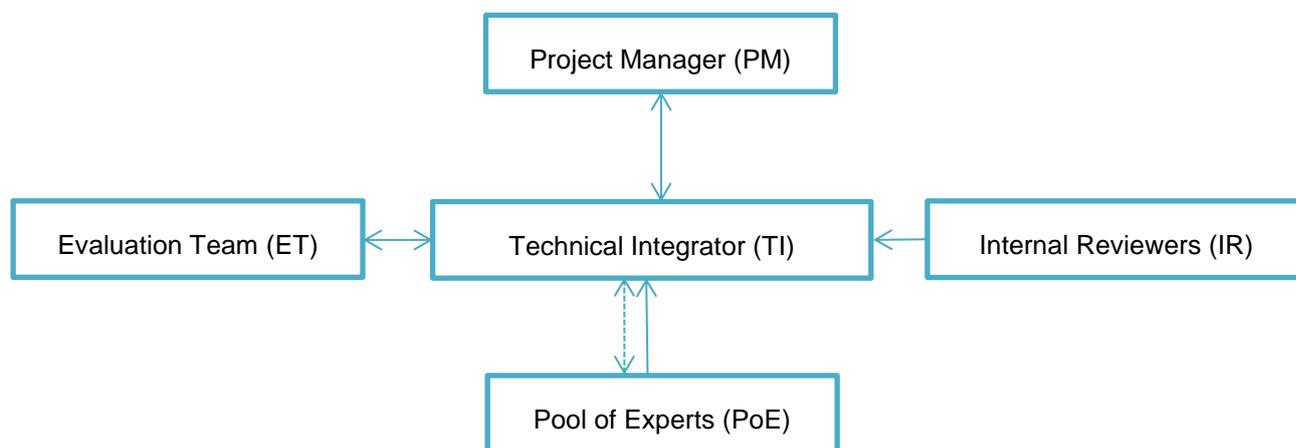
**Pool of Experts (PoE):** This pool has the goal of representing the larger technical community within the process. Two sub-pools are foreseen, which can partially overlap: PoE-H (a pool of hazard analysts) and PoE-V (a pool of vulnerability and risk analysts). PoE-H should have either site-specific knowledge (e.g., hazards in the area) and/or expertise on a particular methodology and/or procedure useful to the TI and the ET team in developing the community distribution regarding hazard assessments. The PoE-V should have expertise on the specific CI and/or on the typology of CI and/or on a particular methodology and/or procedure useful to the TI and the ET team in developing the community distribution regarding fragility and vulnerability assessments. Individual experts of the pool may also act as proponent and advocate a particular hypothesis or technical position, in individual communications with the TI (referring to SSHAC documents, the PoE includes both resource and proponent experts). They participate to the *interviewing processes* (either in remote or through specific meetings) lead by the TI as pool of experts, providing the TI for quantitative opinions on critical choices/issues. If requested by the CI authorities or if irreconcilable disagreements among the experts of the pool emerge during the interviewing processes (in both Phase 1 or Phase 2, see next section), TI and PM may decide to organize meetings with the PoE (or parts of it), in order to openly discuss about controversial issues. In this case, the pool acts as a panel, and TI is responsible for moderating the discussion. TI certifies that there is no conflict of interest with the PoE and that is representative of the technical informed community.

**Internal Reviewers (IR):** One expert or a group of experts on subject matter under review that independently peer reviews and evaluates the work done by the TI and the ET. This group provides constructive comments and recommendations during the implementation of the project. In particular, IR reviews the coherence between TI choices and PM requests, the TI selection of the PoE in terms expertise coverage and scientific independence, the fairness of TI integration of PoE feedbacks, and the coherence between TI requests and ET

implementations. In particular, IP reviews the project both in terms of technical and procedural aspects of the project (actor's independency, transparency, consistency with the project plan). The IR makes sure that the TI has captured the centre, body and range of technically defensible interpretations.

The CI authorities select the PM and the ET. The PM selects the TI and IR. After this selection, the PM interacts only with the TI, taking care of all political choices (e.g., selection of the ST level, definition of acceptable risks, etc.). The TI implements the scientific analysis based on PM choices, providing to the ET all the essential scientific choices that enable the actual implementation of the ST. The TI interacts with the PoE, and it integrates PoE feedbacks into the analysis. The ET implements the ST, following the TI choices. The IR reviews the work of TI and the ET, in order to maximize the reliability of ST results and increase their robustness. The interactions among the core actors are shown in the Figure 4-6.

Note that the IR actively plays an important role during the project and thus is part of the project. If regulators foresee an external review of the ST results, this second review is performed independently and after the end of the project. In EU@STREST, the internal review by the IR is considered essential also in this case, in order to increase the likelihood of a successful external *ex post* review.



**Figure 4-6:** The basic interactions among the core actors in the process of EU@STREST.

### 4.3.3 Key features of the process

The stability and robustness of the results of the process depends on three key features, namely:

**Transparency:** data, models and methods choices are documented;

**Independence:** The Pool of Experts (PoE), Evaluation Team (ET), Technical Integrator (TI) and Internal Reviewers (IR) are independent;

**Responsibility:** PM holds the responsibility of the project and about all “political choices” adopted in it (e.g., selection of ST level, of target perils, etc.). TI holds the intellectual ownership of the process and is responsible for all “scientific choices” in the project (e.g., selection of scientific acceptable models, treatment of uncertainty) and for the results. ET is responsible for performing the ST following the TI requests. IR is responsible for the conformance between the scientific development of the project, the Stress Test level rules, and the EU@STREST guidelines.

### 4.3.4 Project design

A four-phase systematic process is proposed in order to perform the appropriate Stress Test (ST) smoothly. These phases are named as preparation, initialization, integration and evaluation, and finalization. A brief description of these phases can be found below.

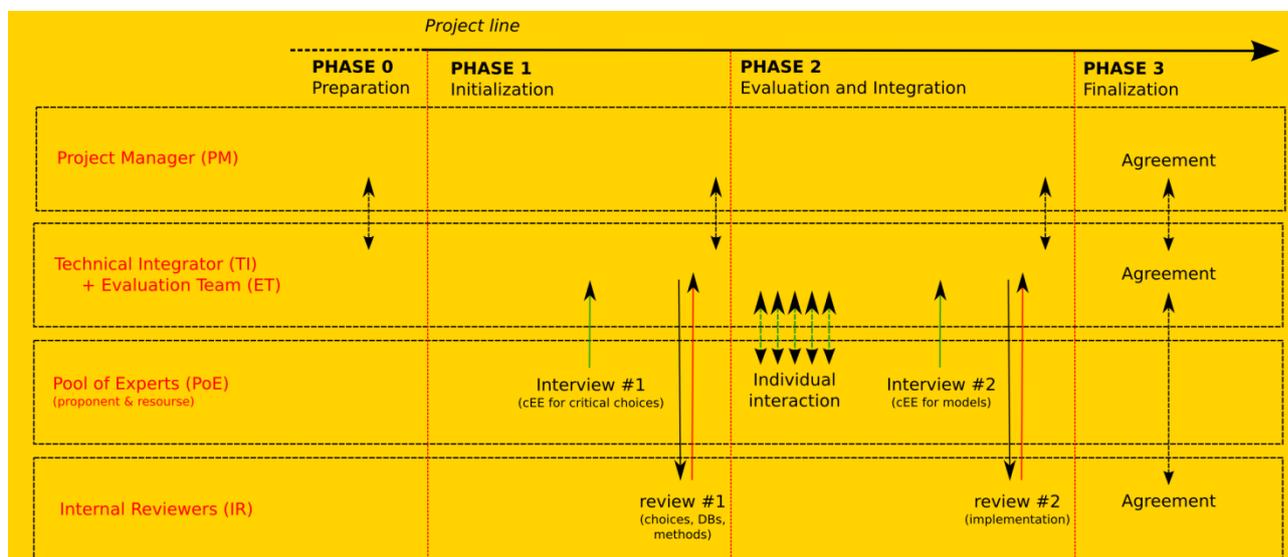
**Phase 0:** The *preparation* phase defines the target ST level and its main goals, and it shapes the project plan for its implementation to the target Critical Infrastructure (CI).

**Phase 1:** The *initialization* consists of 1) the review of all the available data, methods and models for hazard and risk assessment proposed by the larger technical community and of interest for the ST, and 2) the first proposal for the hazard and risk methodologies, potentially identifying methodological gaps (e.g., missing models).

**Phase 2:** The *integration and evaluation phase* builds representation(s) (e.g., in the form of alternative or logic tree) of technically-defensible methods for all the hazards and risks (and interactions) of interest, and it implements them to perform a first ST through an integration method (logic tree or ensemble modeling).

**Phase 3:** The *finalization phase* consists of the revisions and finalization of the ST.

In **Figure 4-7**, we report the graphical representation of the different phases, involvement of main actors and their phase-wise interactions. The main actions in each phase are described with more details in the next sections.



**Figure 4-7:** Graphical representation of the proposed EU@STREST process, showing the different phases and the interactions between the four main actors in each phase.

### 4.3.5 Project's PHASE 0: Preparation

The preparation phase is an organizational phase of the in EU@STREST in which appropriate ST-level is selected and the project plan is developed, outlining the overall organization of the project. More specifically, CI authorities select a PM. Then, the PM selects the TI and IR; PM and TI jointly select the ET; TI selects the PoE, having the different expertise required for PHASE I (hazard, CI components and interaction selection and prioritization). Key roles and responsibilities are assigned to each participant. PM and TI agree the time limits within the project.

PM selects the ST-level and specify the target peril (e.g., injuries/deaths, pollution, etc.), depending on the goal of the analysis and the required results. The PM defines also levels of acceptable risk and performance target (e.g., in terms of number of injured/deaths and/or economical losses and/or minimum performance). In other words, the PM selects all the questions that the ST has to answer to. Depending on this, one or more of the actions in PHASES 1 to 3 may be not required.

All the choices, in this as well as in all the following phases, are summarized in specific documents (fact-sheets), in order to guarantee the transparency.

All the actions required in PHASE 0, their temporal sequence and the required documentation are reported in **Table 3**.

**Table 4-1:** Analytical description of actions along the Phase 0.

<i>ACTIVITY</i>	<i>ACTOR</i>	<i>ACTION</i>	<i>DOCUMENTATION</i>
Preliminary choices	PM	- Definition of the main requirements of the ST (i.e., single/multi hazard, target perils, etc.) - Selection of ST level	ST fact-sheet
Selections of participants	PM	Selection of Technical Integrator (TI) Internal Reviewers (IR), and Evaluation Team (ET)	PoE selection fact-sheet, including Invitation letter
	TI	Selection of the preliminary Pool of Experts (PoE) for PHASE I	
Plan timing	PM+TI	PM + TI agree the total timing, and the one for each of next phase	Project plan

#### 4.3.6 Project's PHASE 1: Initialization

This phase concerns the selection of phenomena of interest, and the primarily development of project's database, through a technical evaluation of the available data, models, and methods from the larger technical community. All the actors play important roles in this phase.

During this phase, it is made the selection of the phenomena of interest (hazards and interactions among hazards, if required by the ST-level, together with the selection of CI components typologies, inter-connections among components and with external systems). To this goal, the PoE should include expertise in the target area (local hazards) and on the typology of the CI. Three steps are foreseen for this first interview process of PoE: preselection, screening and prioritization. In the pre-selection phase, an exhaustive list of hazards of interest and of CI components (both internal and external to the core of the CI) is prepared by the ET. This starting list must be based on existing documents (e.g., NUREG guidelines) and their experience on the CI. In the screening phase, the TI prepares simple forms to be sent to the PoE to screen such lists. The possibility to add potential missing "possibilities" must be allowed, in order to enable potential active feedback from the PoE. The ET prepares also specific supporting materials (e.g., CI general description, general overview of the area, etc.) to be provided to the PoE. In the prioritization phase, with the support of the ET, the TI prepares specific questionnaires for the cEE, finalized to the prioritization of hazards, hazard interactions, CI components and interactions. Based on PoE answers to these questionnaires, a decision method based on classical Expert Elicitation (cEE) is used to prioritize them.

Based on the PoE formal feedback through the cEE, the TI identifies specific requirements on hazard and risk assessments (e.g., spatial correlations). If specific inconsistencies are found among the experts, the TI may decide to convene the PoE (or parts of it) to openly discuss about such issues.

After PoE formal feedbacks, the TI and the ET review the literature for available models and methods for both hazard and risk analyses. At this stage, a revision of the target ST-level is possible, like in the case in which PRA (Probabilistic Risk Assessment, as requested by ST-level 3b) is judged not to be feasible because of the lack of too many necessary factors for the analysis that cannot be filled in the next project's phases (e.g, Probabilistic Hazard is not possible, many fragility models are not available, etc.). For example, the TI may suggest switching to ST-level 3c that foresees the use of SBRA (scenario based risk assessment), which is actually of easier implementation. After this review, TI is ready to set up a preliminary list of approaches to be potentially adopted in the hazard and risk assessments

for the ST. A list of alternative models / approaches should be defined at this stage, in order to allow the management of epistemic uncertainty in the next phases. These choices may be also supported by several sensitivity tests performed by the ET under request of the TI.

The PoE is then updated by the TI with new experts, covering the expertise required to implement the selected hazards, system analysis and interactions (for hazards and CI components).

The IR reviews the consistency of TI choices with PoE feedbacks and PM requirements, the effectiveness of the PoE update and the existence of important gaps in the review.

All the actions required in PHASE 1, their temporal sequence and the required documentation are reported in **Table 4**.

**Table 4-2:** Analytical description of actions along the PHASE 1. According to Figure 4.7, the phases of interaction between TI and the IR are highlighted in red, while the ones between TI and the PoE are highlighted in green.

ACTIVITY	ACTOR	ACTION	DOCUMENTATION
Preparation for pre-selection	TI	ET prepares the complete list of hazards and of typologies of CI components, based on existing guidelines and CI experience. All possible interactions are considered in this first step. This includes the definition of a preliminary list of potential: single hazards of interest (for <i>ST-level 3b and 3c</i> ); interaction among hazards of interest (for <i>ST-level 3b and 3c</i> ); main components of the CI of interest; main interaction among components of interest (including the external systems). With the support of the ET, the TI prepares simple forms in order to screen all these possibilities. The possibility to add potential missing possibilities must be allowed.  ET and TI also prepare some accompanying material, which includes a general description of CI, of its component, and a general overview on its location.	CI fact-sheet
Pre-selection	TI + PoE	TI asks the PoE to fill simple forms, to pre-select a restricted list of hazards, interaction among hazards, components and interaction among components.	Pre-selection fact-sheet
Preparation of the first elicitation process	TI	Preparation of questionnaires for PoE. The questionnaires are focused to prioritize only the phenomena pre-selected in the "Pre-selection" phase.	Documentation for cEE
First elicitation process	TI + PoE	<i>Formal feedback on PHASE 1 through cEE (decision theory):</i>  <i>- Selections CI components &amp; interactions through a decision method (for all ST-levels)</i>  <i>- Selection of hazards and interaction through a decision method (for ST-levels 2b, 3a, 3b, 3c)</i>	Feedback document #1
Develop project database (DB)	TI + ET	- Review of potential DB of interest  - Review of system analysis requirements and methods available  - Review of potential regional (SHARE, ASTARTE, etc.), sub-regional (National Hazard Maps), and local (Previous studies	DB fact-sheet

## Methods for the analysis of epistemic uncertainties

		for the CI area) hazard assessments - Review of fragility for CI components	
Develop preliminary model(s) and hazard/risk Input Document (HRID)	TI + ET	- Selection of models / approaches for implementing the ST (for ST-levels in which epistemic uncertainty quantification is not required) - Selection of potential alternative models / approaches for implementing the ST (for ST-levels in which epistemic uncertainty quantification is required) - Identification of potential lacks in modelling, and potential review of ST-level (e.g., from ST-level 3b to 3c, since PRA judged not feasible) - This phase also includes the production by the ET of sensitivity analyses potentially supporting TI decisions.	Hazard fact-sheet Fragility fact-sheet Interdependencies fact-sheet Sensitivity analyses
Addition of further expertise (if required)	TI	Review of the Pool of Experts (PoE) in order to include expertise related to all the processes prioritized and selected in PHASE 1	PoE integration fact-sheet
<i>First review</i>	<i>IR</i>	<i>Formal revision of PHASE 1</i>	<i>Review #1</i>

### 4.3.7 Project's PHASE 2: Evaluation and Integration

The purpose of this phase is to complete the development of all the methods required for a complete implementation of the ST, based on the choices and the review of PHASE 1, in order to carry out the preliminary hazard and risk assessments. In this phase, also the integration method (e.g., Logic Tree or Ensemble modelling, see Section 2) should be selected. All the actors play important roles in this phase.

The ET develops all the required models based on TI guidelines, also through individual interactions with PoE participants. Then, the TI prepares questionnaires and supporting material for the second interview phase. In this phase, different requirements exist for the different ST-levels and the choices made at the previous phases. If modelling gaps have been found, a cEE is required to fill those gaps. If a SBRA (ST-level 3c) has been selected, the PoE should quantify the scenarios to be modelled. If a full quantification of epistemic uncertainty has to be performed, the potential alternative methods / approaches for hazard and risk calculations should be selected/scored/ranked, depending on the selected method for models' integration: the possibility to add potential missing methodologies must be allowed. Supporting material should include all possible available information, depending on the goals of the second interview phase, e.g., about missing models (if any) and/or about selected methods (if quantification of epistemic uncertainty is required) and/or about the historical record of the target hazard (for SBRA). After this second interview phase, made through cEE, the TI must be able to fully implement a preliminary ST. As in PHASE 1, if

specific inconsistencies are found among the experts, the TI may decide to convene the PoE (or parts of it) to openly discuss about such issues.

Then, ET finalizes the ST analysis, based on TI requests and guidelines. Finally, the IR reviews the consistency of TI choices with PoE feedbacks and PM requirements.

All the actions required in PHASE 2, their temporal sequence and the required documentation are reported in **Table 5**.

**Table 4-3:** Analytical description of actions along the PHASE 2. According to Figure 4-6, the phases of interaction between TI and the IR are highlighted in red, while the ones between TI and the PoE are highlighted in green.

ACTIVITY	ACTOR	ACTION	DOCUMENTATION
Finalization of project's DB, models and methods for hazard and risk analyses	TI + ET + PoE	<i>Individual interaction regarding the update of project's DBs, methodological aspects, enabling the practical implementation of alternative methodologies (e.g., branches of an Alternative or a Logic Tree)</i>	Methods fact-sheet
Preparation of the second interview process	TI	Production of new questionnaires & supporting material for the PoE, depending on ST-level.  Supporting material should contain information about missing models (if any) and/or about selected methods (if quantification of epistemic uncertainty is required) and/or about the historical record of the target hazard (for SBRA).	Documentation for cEE, including supporting material
<i>Second elicitation process</i>	<i>TI + PoE</i>	<i>- cEE for producing experts' based models, filling potential gaps emerged during phase 2 (e.g., through Cook's method, see section 2.4)</i>  <i>- for ST-level 3c: Selection of hazard scenarios (e.g., through a decision method or Cook's method, see Section 2.4)</i>  <i>- for ST-levels 2b, 3a and 3b: cEE about selecting/scoring/ranking alternative methods for quantifying epistemic uncertainty (e.g., through AHP, see Section 3.4)</i>	<i>Feedback document #2</i>
Preliminary hazard and risk calculations	ET + TI	Preliminary Stress Test	Draft project report(s)
<i>Participatory peer review</i>	<i>IR</i>	<i>Formal revision of PHASE 2</i>	<i>Review #2</i>

### 4.3.8 Project’s PHASE 3: Finalization

The PHASE 3 is the conclusive phase in which the EU@STREST process is finalized, including all feedbacks in PHASE 2 (if any). TI, ET and IR play a decisive role in this phase.

TI and ET revise the final models considering the feedback of reviewers and perform the final computations. PM, IR and TI arrange the final meeting to reach an agreement.

All the actions required in PHASE 3, their temporal sequence and the required documentation are reported in Table 6.

**Table 4-4:** Analytical description of actions along the PHASE 3. According to Figure 4-6, the phases of interaction between TI and the IR are highlighted in red.

<b>ACTIVITY</b>	<b>ACTOR</b>	<b>ACTION</b>	<b>DOCUMENTATION</b>
Finalize models	TI	In light of Feedback document #2 (if any)	
Perform final hazard calculations and sensitivity analyses.	ET + TI	Final Stress Test (if needed)	Final project report(s)
Final meeting	TI+IR	Discussion on final results and inclusion of all feedbacks	Report on all project documents (including feedbacks)
<i>Agreement</i>	<i>PM+IT+IR</i>		<i>Agreement statement</i>

## 5 Case studies

### 5.1 INTRODUCTION

Sensitivity tests are required to enable a more robust selection of methods with any assessment of epistemic uncertainty (e.g., Bommer and Scherbaum 2008). The output of these tests represents also a fundamental input for the pool of experts, in order to support them in providing their structured feedbacks within the EU@STREST process. In this chapter, we present several sensitivity tests that have been performed in the project. Such applications, along with others, will be used in the case studies to guide and support potential controversial decisions.

### 5.2 CASE STUDIES

#### 5.2.1 Sensitivity of induced seismicity ground motion accounting for regional conditions (CI-B2)

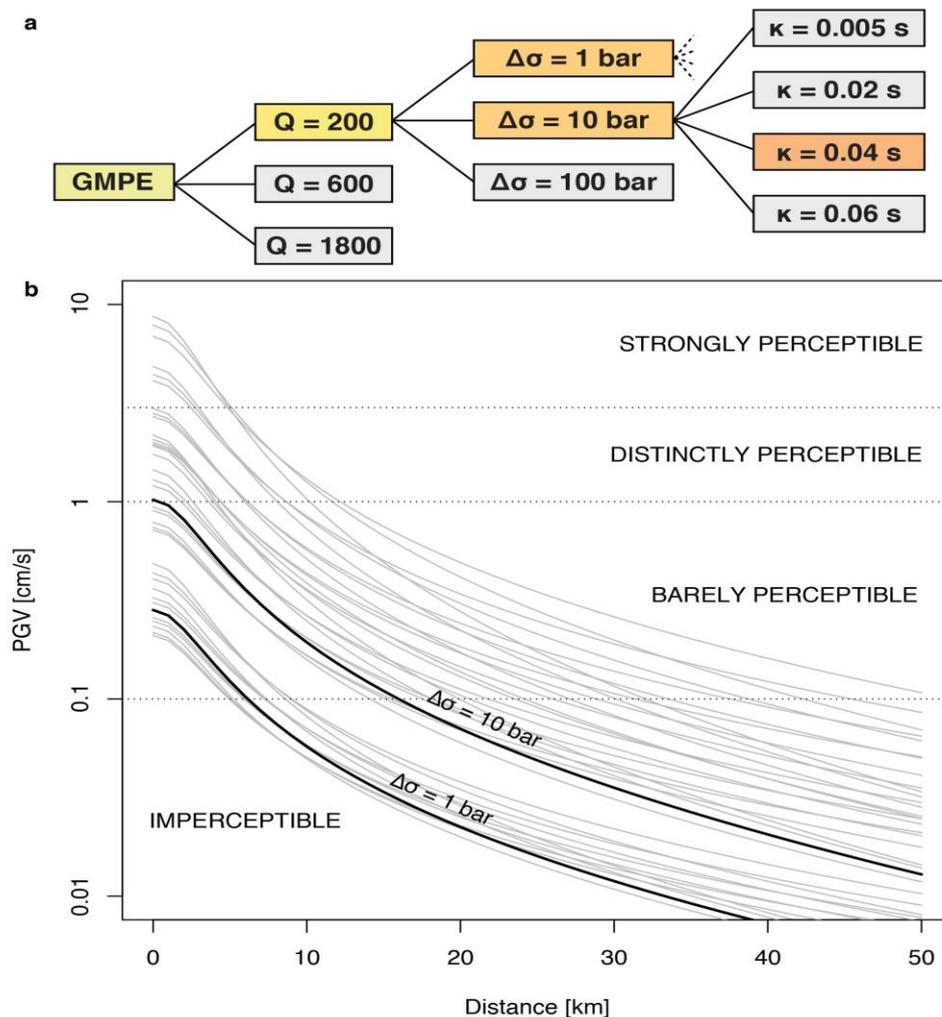
This section investigates how ground motion due to small-to-moderate induced earthquakes varies due to epistemic uncertainties and how taking into account regional physical conditions can reduce these uncertainties. We here use the ground motion prediction equations (GMPEs) of Douglas et al. (2013) and the logic tree proposed by Mignan et al. (2015) for induced seismicity risk assessment. We apply the method to the Groningen gas field (where STREST CI-B2 [Gasunie gas network] is located).

To account for epistemic uncertainties in induced seismicity-based GMPEs, Douglas et al. (2013) developed 36 stochastic models representing different stress drops  $\Delta\sigma$  (1, 10 and 100 bars) and attenuation parameters  $Q$  (200, 600 and 1800) and  $\kappa$  (0.005, 0.02, 0.04 and 0.06 s). Mignan et al. (2015) included these models in a logic tree for the risk assessment of induced seismicity in Basel, Switzerland. For that region, 8 GMPE models were selected in agreement with regional conditions (i.e.  $\Delta\sigma = 3-10$  bars,  $Q = 1200$ ,  $\kappa = 0.016$  s).

Here, we use the same approach for the Groningen gas field. We obtain the necessary parameter estimates from Kraaijpoel and Dost (2013), with  $\Delta\sigma = 3-17$  bars,  $Q = 20$ ,  $\kappa = 0.04$  s. Figure 5-1a shows the 2 GMPE models selected in the logic tree (colored) based on the parameter set for the Groningen field, i.e.  $Q = 200$  (minimum available),  $\kappa = 0.04$  s and  $\Delta\sigma =$

1 or 10 bars. Figure 5-1b shows the expected ground motion variability for an  $M_w = 4$ ,  $h = 3$  km deep induced earthquake, for all the models (grey curves) and for the 2 selected models (black curves). While the peak ground velocity (PGV) varies within about two orders of magnitude when considering all the models, it varies within less than one order of magnitude when considering parameters in agreement with the conditions at the Groningen gas field. We note also that PGV in Groningen are on the lower range of possible values due to small to moderate stress drops and an intermediary  $\kappa$ . For the example shown in Figure 5-1b, we see that such earthquake would be barely perceptible in average (adapting from Bommer et al., 2006). PGV values are however expected to be higher in 50% of cases (median values shown in Figure 5-1b) due to aleatory uncertainties (not shown).

**Figure 5-1:** Ground motion due to an  $M_w = 4$ ,  $h = 3$  km deep, induced earthquake: a. Logic tree; b. PGV versus epicentral distance





### 5.2.2 Sensitivity of $M_{max}$ accounting for fault interactions (CI-B1)

This section investigates how estimates of the maximum magnitude  $M_{max}$  of strike-slip earthquakes change once knowledge about fault interactions is added. The new approach proposed to estimate  $M_{max}$  – based on concepts of dynamic stress – is summarized below and applied to faults in the Anatolian Peninsula (where STREST CI-B1 [hydrocarbon pipelines] is located). For the full description of the approach and of the results, the reader should refer to STREST Deliverable 3.5 and to Mignan et al. (in press). This section of Deliverable 3.1 focuses on the  $M_{max}$  sensitivity.

Earthquakes are known to potentially propagate over several segments (e.g., 2002  $M_w = 7.9$  Denali, Alaska, earthquake; 2011  $M_w = 9.0$  Tohoku, Japan, earthquake). Yet, fault segments are still modeled as individual faults in most regional seismic hazard models. Mignan et al. (in press) (see also STREST D3.5) defined a set of simple criteria to assess cascades from individual fault segments by using dynamic stress modeling assumptions (e.g., Kame et al., 2003; Harris et al., 2002) and field observations (e.g., Wesnousky, 2006; Barka and Kadinsky-Cade, 1988): These criteria consider bending, branching and jumping between fault segments, as follows:

1. Compatibility of segments: Segments involved in a cascade must have the same mechanism (left- or right-lateral) and the same dip direction (i.e. not antithetic).
2. Maximum distance: The minimum distance  $\Delta$  between two sources must be lower than 5 km.
3. Maximum strike difference: The relationship  $\psi - \delta \leq \varphi \leq \psi + \delta$  must be respected with  $\varphi$  the angle between two segments (i.e. strike difference),  $\psi = \gamma(45 - \Psi - 180 \text{ atan}(\mu_d)/2\pi)$  the optimal angle for rupture,  $\Psi > 0$  the angle between the first segment and the direction of maximum compressive stress  $S_{max}$ ,  $\mu_d$  the dynamic friction coefficient,  $\gamma = 1$  for right-lateral and  $\gamma = -1$  for left-lateral, and  $\delta = 30^\circ$  the range of preferred orientation.
4. Relative position of segments: The rupture can propagate from all or part of a segment to all or part of another segment if the angle between the two subsequent segments remains obtuse (i.e., no backward branching/bending allowed).

These criteria were applied to a subset of strike-slip faults in the Anatolian Peninsula, as defined within the 2013 released *European Seismic Hazard Model* (ESHM13) (Basili et al., 2013; Giardini et al., 2013).

The  $M_{max}$  parameter in ESHM13 is obtained by the weighting of various empirical magnitude scaling relationships (Wells and Coppersmith, 1994; Mai and Beroza, 2000; Hanks and

Bakun, 2002; Leonard, 2010), as illustrated in Figure 5-2a. Following recommendations from Stirling et al. (2013), we additionally consider the relationship proposed by Wesnousky (2008). In ESHM13,  $M_{max}$  is linked to the effective length  $L_{eff}$  instead of the total length  $L_{tot}$  of fault segments to take into account lower slip rates. For strike-slip faults in the Anatolian Peninsula,  $6.5 \leq M_{max}$  (ESHM13)  $\leq 8.1$  (with  $17 \leq L_{tot} \leq 360$  km).

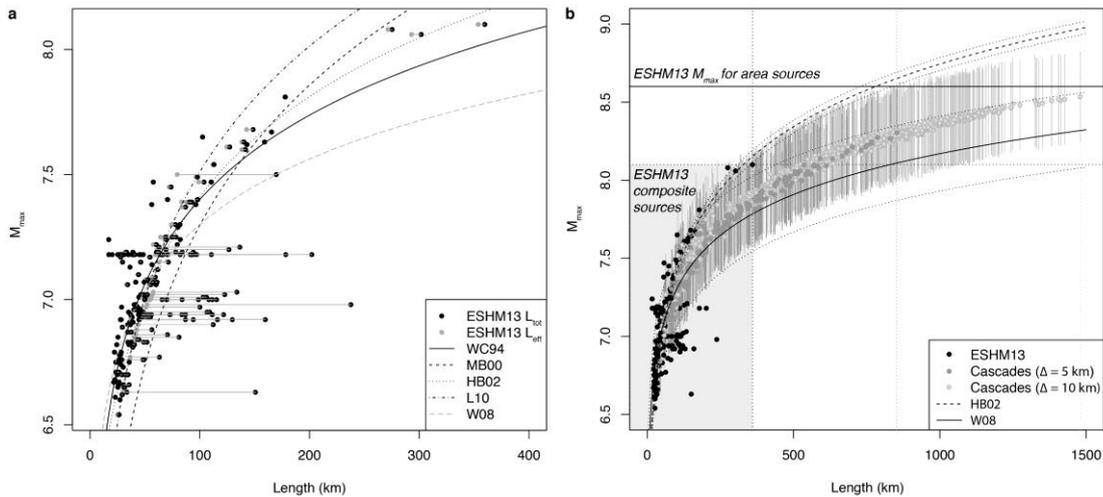


Figure 5-2: Length- $M_{max}$  relationships: a. ESHM13; b. Including cascades (modified from Mignan et al., in press)

Figure 5-2b shows estimates of  $M_{max}$  obtained for the fault length distribution of cascading ruptures generated by applying the criteria defined above (with  $33 \leq L_{casc}(\Delta = 5 \text{ km}) \leq 853$  km and  $33 \leq L_{casc}(\Delta = 10 \text{ km}) \leq 1480$  km). The range of rupture cascade lengths exceeds the validity range of the magnitude scaling relationships, especially in the case of Mai and Beroza (2000) ( $L_{max} = 180$  km) and Leonard (2010) ( $L_{max} = 50$  km for strike-slip ruptures), which are here rejected. Therefore only the relationships of Hanks and Bakun (2002) and of Wesnousky (2008) ( $L_{max} \sim 430$  km) are shown in Figure 5-2b. Since these scaling relationships are not well constrained at very high length values, physical constraints were also considered by Mignan et al. (in press) by taking into account the role of the slip rate on  $M_{max}$  estimation using the equation of Anderson et al. (1996) (points on Figure 5-2b). The plot of Figure 5-2b shows that the Hanks and Bakun (2002) estimates may represent an upper bound for  $M_{max}$ , the Wesnousky (2008) estimates a lower bound and the Anderson et al. (1996) estimates intermediary values. For  $L_{max}(\Delta = 5 \text{ km}) = 853$  km,  $M_{max}(\text{HB02}) = 8.65 \pm 0.05$ ,  $M_{max}(\text{W08}) = 8.11 \pm 0.24$  and  $M_{max}(\text{A96}) = 8.30 \pm 0.29$ . For  $L_{max}(\Delta = 10 \text{ km}) = 1480$  km,  $M_{max}(\text{HB02}) = 8.97 \pm 0.04$ ,  $M_{max}(\text{W08}) = 8.32 \pm 0.24$  and  $M_{max}(\text{A96}) = 8.53 \pm 0.29$ .

Figure 5-3 shows an  $M_{max}$  logic tree based on these results. It should be noted that  $\Delta = 5$  km is considered as standard in the criteria for multi-segment rupture. However, uncertainties on fault traces may require changes in  $\Delta$ . Choice of a higher  $\Delta$ , here  $\Delta = 10$  km, was made by Mignan et al. (in press) to reproduce the plate boundaries (e.g. the North Anatolian Fault) as continuous structures. Let's finally add that strike-slip ruptures are unlikely to reach M9 values due to their scaling with  $L$  instead of  $L^2$  for dip-slip events (Romanowicz and Ruff, 2002). The 2012 equatorial Indian Ocean earthquakes of  $M_w$  8.6 and 8.2 were two of the largest strike-slip earthquakes ever recorded (Duputel et al., 2012). As a consequence, the branch "HIGH – Hanks & Bakun (2002)" should have the lowest weight.

The proposed logic tree could be used in the design of new stress tests (WP5/6) where the impact of cascade-based  $M_{max}$  what-if scenarios would be investigated. For CI-B1, the maximum fault displacement can be inferred from  $M_{max}$  to then be able to estimate the performance of pipelines to extreme cascading ruptures (WP4/5/6).

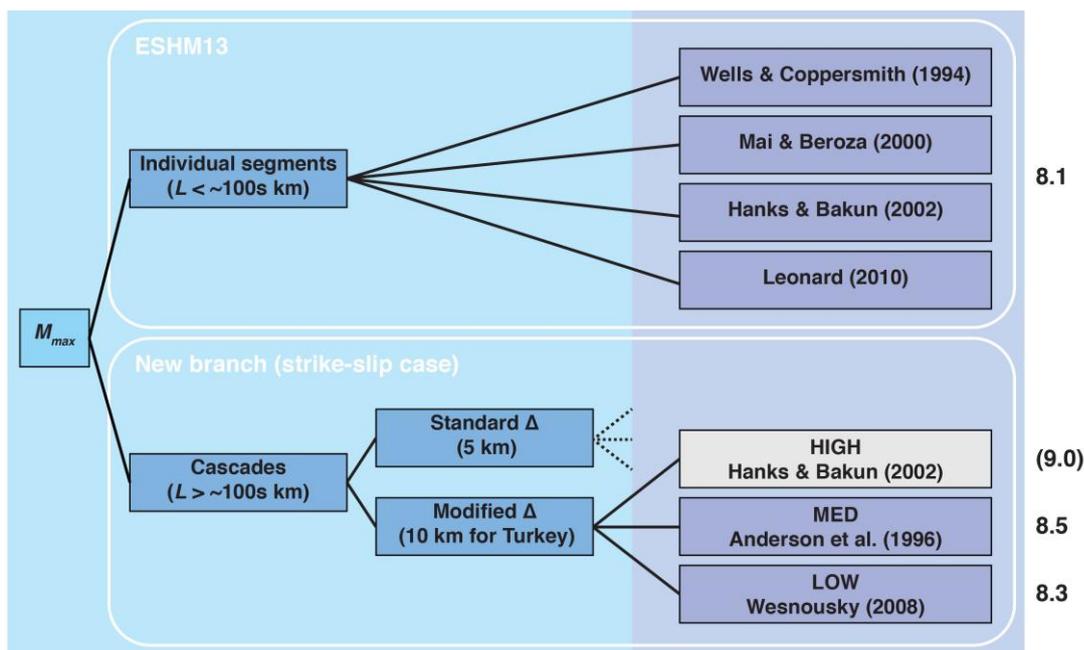


Figure 5-3: Proposed  $M_{max}$  logic tree for CI-B1 (Strike-slip faults in Anatolian Peninsula)

### 5.2.3 Sensitivity of performance assessment for urban infrastructure and harbours to spatial correlations in the hazard

The objective of the analysis is to evaluate probabilities or mean annual frequency of events defined in terms of loss in performance of networks. The analysis is based on an object-

oriented paradigm where systems are described through a set of classes, characterized in terms of attributes and methods, interacting with each other. The physical model for each network starts from the SYNER-G taxonomy (Pitilakis et al. 2014) and requires: a) for each system within the taxonomy, a description of the functioning of the system (intra-system dependencies) under both undisturbed and disturbed conditions (i.e., in the damaged state following an earthquake); b) a model for the physical and functional (seismic) damageability of each component within each system (fragility functions); c) identification of all dependencies between the systems (inter-dependencies); and d) definition of adequate Performance Indicators (PIs) for components and systems, and for the infrastructure as a whole (Franchin 2014).

The computational modules include the following main models: a) seismic hazard class modeling earthquake events and corresponding seismic intensity parameters, b) network class modeling physical damages of networks' components and the overall system's performance, c) interdependencies models simulating specific interactions between systems. The hazard model provides the means for: 1) sampling events in terms of location (epicentre), magnitude and faulting style according to the seismicity of the study region and 2) maps of sampled correlated seismic intensities at the sites of the vulnerable components in the infrastructure ('shakefields' method, Weatherill et al. 2014). When more vulnerable components exist at the same location and their fragility is expressed with different intensity measures, the model assesses them consistently. Probabilistic evaluation of the performance of networks is carried out by means of Monte Carlo simulations. For simplicity, the methodology is focused on performance without reparations (emergency phase). The final goal is to assess the exceedance probability of different levels of performance loss for each system under the effect of any possible seismic input and physical damages. This output, represents the performance curve, and is the equivalent of risk curves for non-systemic probabilistic assessments in single (e.g., PEER formula, Cornell and Krawinkler, 2000) and/or multi-risk (e.g., Selva, 2013) analysis. Further specifications for each physical system are given in Modaressi et al. (2014), Argyroudis et al. (2014).

## Seismic Hazard

For the seismic hazard input of the present application, five seismic zones with  $M_{\min}=5.5$  and  $M_{\max}=7.5$  are selected based on the results of SHARE European research project (Giardini et al. 2013). In addition, a seismic source representing the destructive earthquake ( $M=6.5$ , 20 June 1978) that caused extensive damage in the study area is adopted for a specific scenario analysis Figure 5.4. A Monte Carlo simulation (MCS) is carried out sampling seismic events for the considered zones following the ‘shakefield’ approach (Weatherill et al. 2014). In particular, earthquakes are sampled from the seismic zones in terms of localization and magnitude and local intensity values at the sites of vulnerable components are evaluated. First, a scalar random field of a so-called primary Intensity Measure (IM) on rock is sampled (i.e. peak ground acceleration, PGA), on a regular grid covering the study region, as a function of the sampled magnitude and epicenter location, employing the ground motion prediction equation (GMPE) by Akkar and Bommer (2010). The spatial variability for PGA is modelled using the correlation models provided by Jayaram and Baker (2009) as adapted for European events consistently with the selected GMPE (Esposito and Iervolino, 2011). For each site of the grid, the averages of primary IM from the specified GMPE are calculated, and the residuals are sampled from a random field of spatially correlated Gaussian variables. Then, the primary IM is interpolated to all vulnerable sites and the secondary IMs are sampled from their distribution conditional on the primary IM value. All values are amplified on the basis of local soil conditions using the amplification factors proposed in EC8 (2004). The geotechnical hazard model proposed by HAZUS is used to sample permanent ground deformations (PGD) due to liquefaction for components whose fragility model requires one (e.g. pipelines, harbor cranes). In case of the specific scenario analysis, only shakefields are sampled for the selected event.

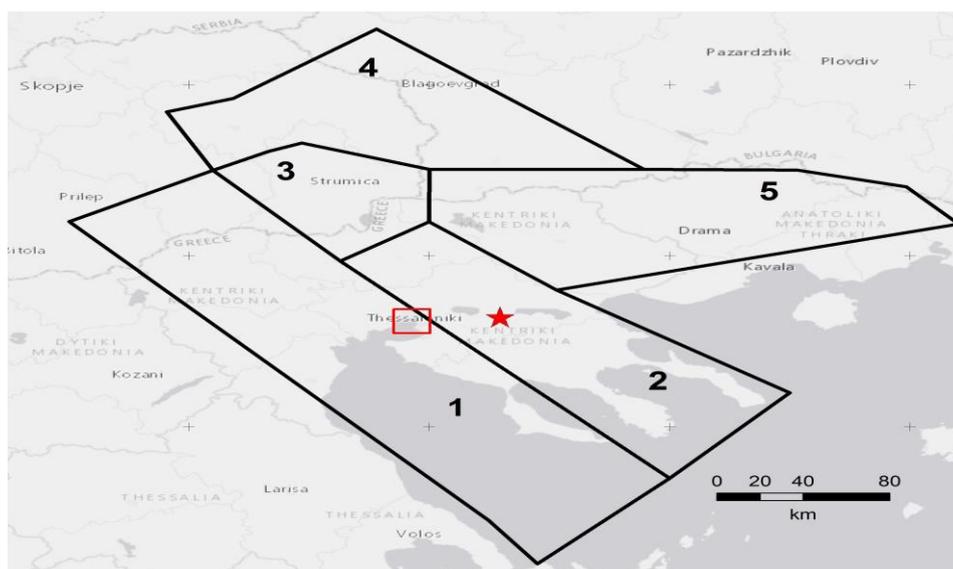


Figure 5-4: Seismic zones, epicenter of M6.4, 1978 event (red star) and study area (red rectangle).

### System specifications

The main features of the networks under study are given in the following. For more details the reader is referred to Argyroudis et al. (2014).

#### *Electric Power Network (EPN)*

The EPN is modelled as a directed graph, i.e., a graph in which all edges have a travelling direction, from node  $i$  to node  $j$ . The main network for the case study is composed of 30 nodes and 29 edges. Nodes consists of 1 generator (over-high voltage 400-150 kV station), 8 high-voltage substations (150 kV /20 kV) and 21 demand points located at WSS pumping stations. In this way the interaction with the WSS is simulated through the connection of pumping stations with the reference EPN load bus (here substation). Only transmission substations are considered as vulnerable components, while edges are non-vulnerable transmission lines (underground and overhead) connecting the generator with the transmission substations and the transmission substations with the demand points. The fragility curves for transmission substations are classified in 3 classes (open, mixed and closed-type) and given in terms of peak ground acceleration (PGA). During each simulation the nodes that are non-functional are removed from the system and residual connectivity is reassessed. The adopted PI is the Electric power Connectivity Loss (ECL), which relates the number of connected nodes in seismic and non-seismic conditions.

### *Water Supply System (WSS)*

The WSS for the case study is comprised of 477 nodes and 601 edges with total length of about 280 km. The nodes consist of demand nodes, pumping stations and tanks; the latter considered as water sources for the system. The simulated network includes 445 demand nodes, 21 pumping stations and 11 tanks. The WSS is modelled as a directed graph, as EPN. Pipelines have 24 different diameter values (between 500-3,000 mm); their construction materials include asbestos cement, cast iron, PVC and welded steel. The fragility functions of ALA (2001) that correlate the repair rate (RR) with peak ground velocity (PGV) and permanent ground deformation (PGD) are used. The considered PI is the Water Connectivity Loss (WCL), which relates the number of connected nodes in seismic and non-seismic conditions.

### *Roadway Network (RDN)*

The RDN for the case study is composed of 594 nodes and 674 edges. The nodes consist of 15 external nodes, 127 Traffic Analysis Zone (TAZ) centroids and 452 simple intersections. The RDN is modelled as a directed graph, 495 edges are two-ways and 179 are one-way roads. Specific fragility curves that were constructed for bridges of the study area are applied as a function of PGA for two damage grades, namely yielding and ultimate. The fragility curves provided in HAZUS (FEMA 2003) are used for the road segments exposed to PGD due to liquefaction. The effect of collapsed overpasses and buildings to the road functionality is also considered based on the results of the fragility analysis of these structures and other models that relate the induced debris to road blockage. Two performance indicators are used, the Simple Connectivity Loss (SCL) and the Weighted Connectivity Loss (WCL). SCL measures the average reduction in the ability of sinks to receive flow from sources. WCL upgrades the simple connectivity loss by weighting the number of sources connected to each sink, in the seismically damaged network and in non-seismic conditions.

### *Harbor (HBR)*

HBR components are 72 nodes and 17 sides. The nodes consist of pier-edges (non-vulnerable) and (non-anchored) cranes (vulnerable). Edges include only (gravity type) waterfronts. The fragility models used for cranes and waterfronts are expressed in terms of PGD and PGA. In addition, we considered the EPN sub-network within the HBR, consisting of 17 distribution substations, 74 edges and 48 demand nodes (cranes), to model EPN-HBR interaction. This sub-network is supplied by the EPN of the city and is comprised of

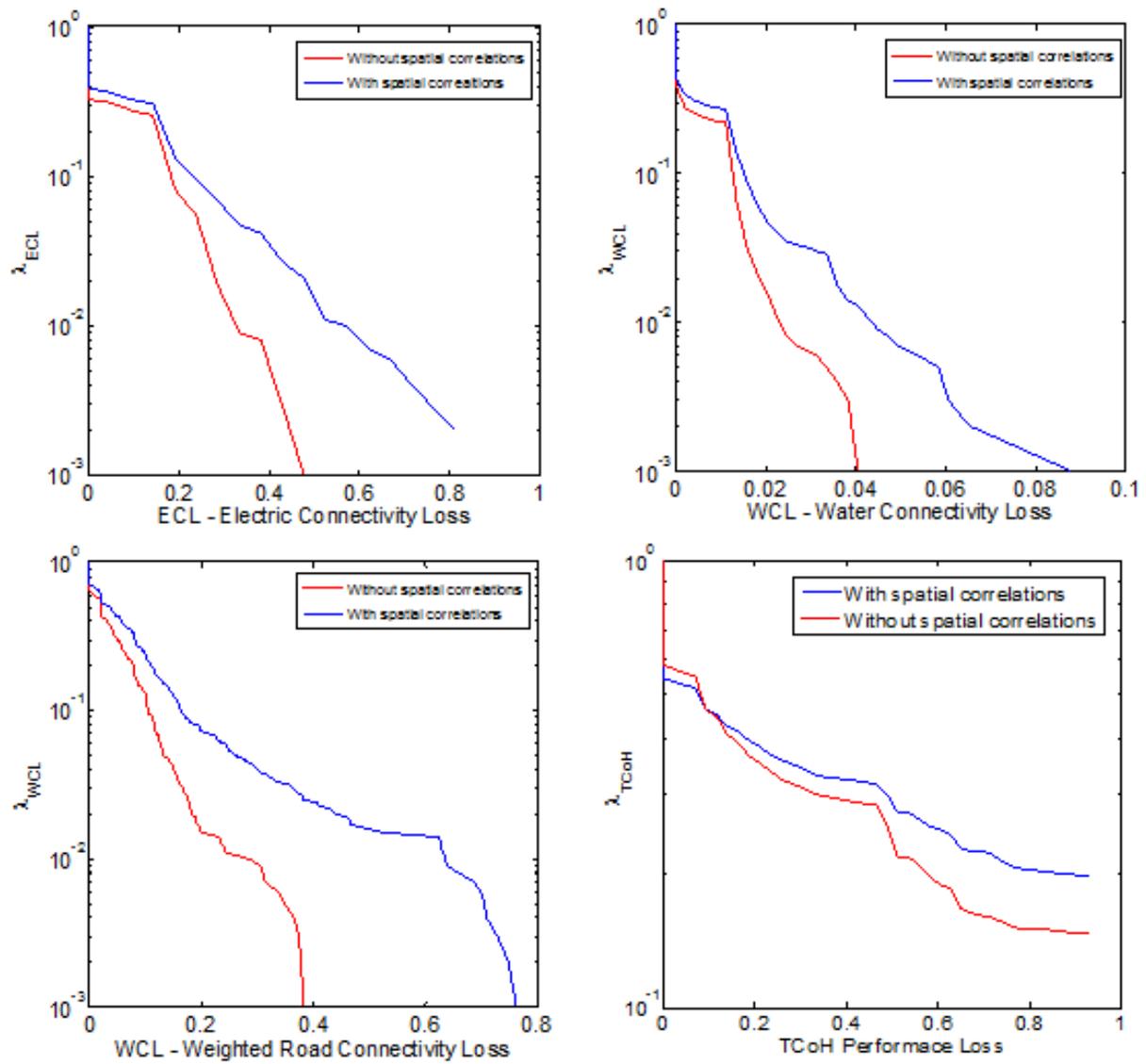
vulnerable components (distribution substations). The interdependency between the cargo handling equipment and the electric power supply to cranes is considered. The PI used is the total containers handled (TCoH) per day (more details in Kakderi et al. 2014).

### **Application and discussion**

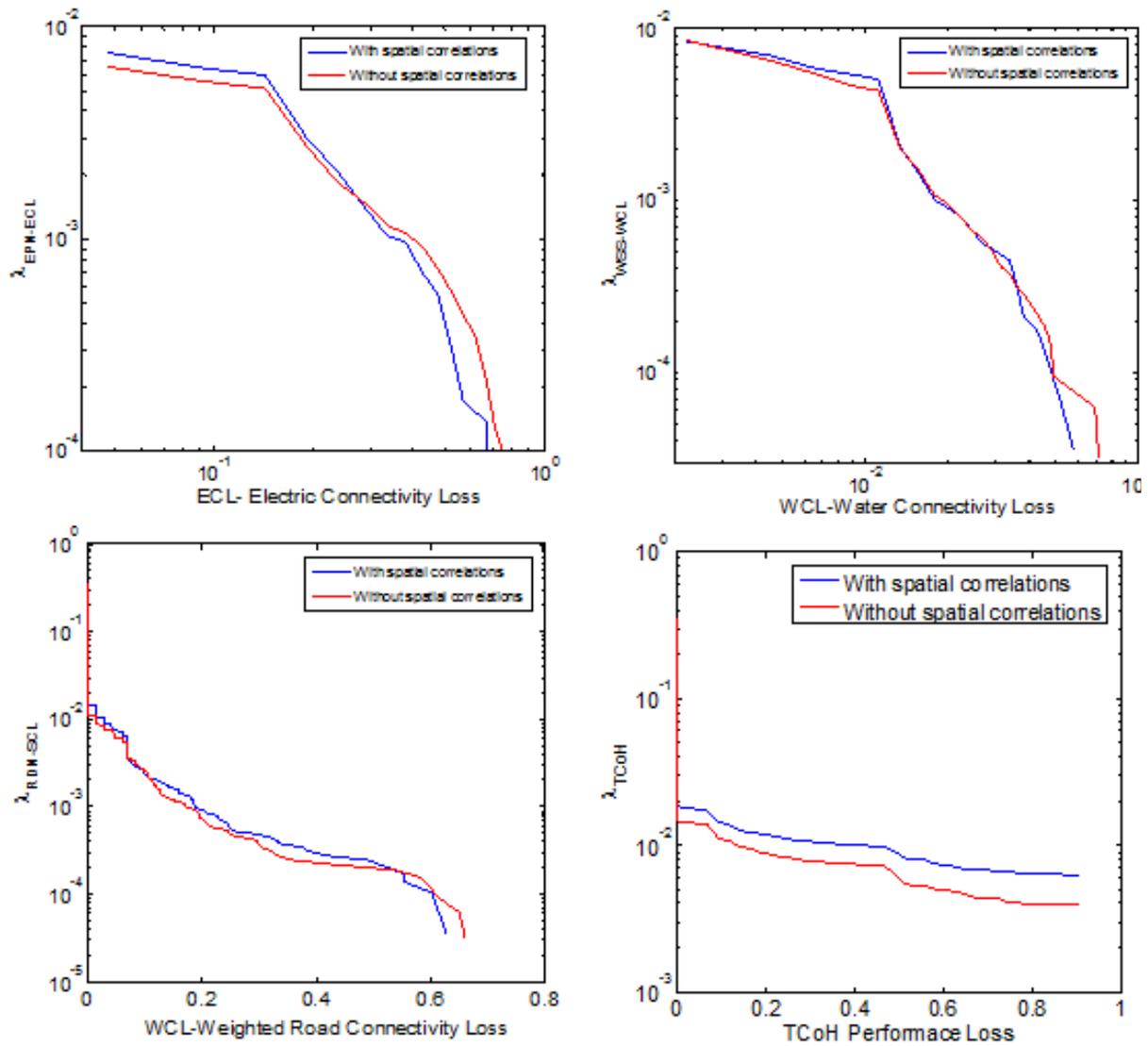
In Figure 5-5 and Figure 5-6 we show the differences in the estimated mean annual losses (MAF) for EPN, WSS and RDN connectivity loss and HBR performance loss, with and without considering spatial correlations of the intensity measures. The results are based on 1,000 runs for a specific seismic scenario ( $M=6.5$ ) and on 10,000 runs (MCS) for the five seismic zones ( $M=5.5-7.5$ ). In Figure 5-7 we show the correlation of non-functional components (water pumping stations and electric power substations) to the corresponding system's PI, with and without considering spatial correlations of the intensity measures. In this way the most correlated components to the system's PI can be defined for all possible or for a specific seismic scenario, that is, the components that mostly control the performance of the system. In Figure 5-8 we show the contribution of various magnitudes and seismic zones to the probability of exceeding a specific threshold, with and without considering spatial correlations of the intensity measures.

In case of one seismic scenario it is observed that when spatial correlations of IMs are not considered, the expected losses can be underestimated (Figure 5.5). In particular, neglecting spatial correlations results the considerably lower losses at lower annual probabilities of exceedance. In this case random IMs are computed (based on the GMPE), which are not correlated for adjacent cells/components, and consequently the estimated damages and losses are also not correlated. For example, two adjacent components can have completely different seismic demand (IM), and thus completely different damages, which is rather not realistic, resulting to bias of the expected losses. It is also observed (Figure 5-7 @ up) that, in both cases (with/without spatial correlations) the most correlated (e.g. the four most correlated) components are identical. However, the 'critical' components seem to emerge better when no spatial correlations are considered. This is maybe related to the fact that correlated damages for spatially proximate components may induce a 'false' correlation between the PI and a component that is not relevant for the system performance, but it is spatially close to another component that is actually relevant for the system. In case of all possible seismic scenarios (Figure 5-6) the differences in MAF curves are not as clear as before for the EPN, WSS and RDN which are systems distributed in a large area (up to 180 sq. km) with the exception of the HBR that is extended in a smaller area (3 sq. km). This may be due to the fact that the most important sources for the hazard are the local ones

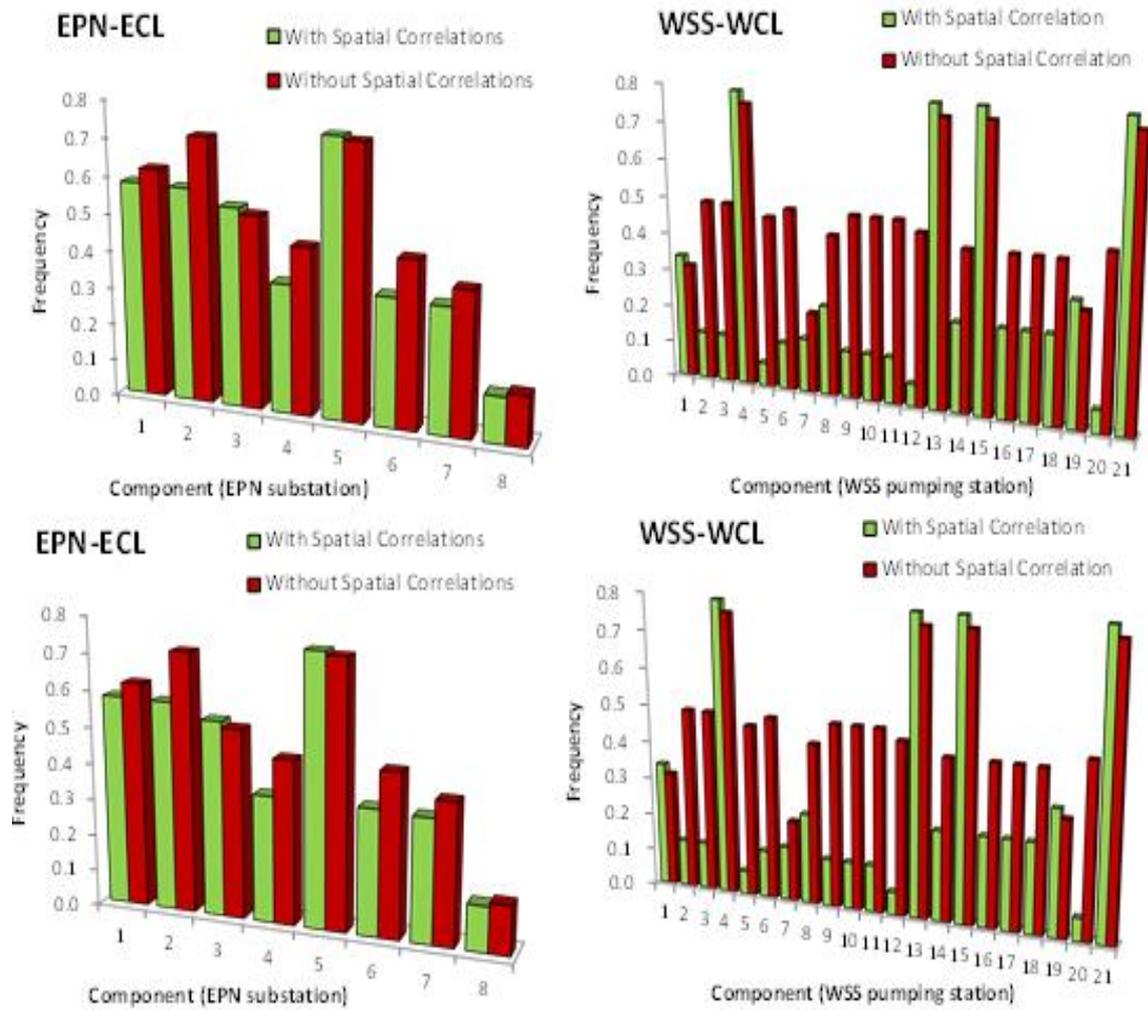
(Zone 1 & 2, Figure 5-4 and Figure 5-8), and most of components of the larger systems are quite far away to each other and spread all over the city. Thus, earthquakes occurring within the city area may dominate the performance loss, and systems' components are spread around the epicenters. On the opposite, HBR is localized, with components very close to each other, and earthquakes have a small chance of being localized within such a limited area. Similar observations on the reduced impact of spatial correlations for larger spatial 'footprint' of an exposure portfolio have been made by Weatherill et al. (2015). The same trends for the most correlated components are derived from Figure 5-7 (down) as for the case of the single scenario. The disaggregation also shows Figure 5-8 that for both cases (with and without spatial correlations) smaller magnitudes contribute more to the loss (for the considered low thresholds), since those events are more probable than larger ones. However, the likelihood of seismic scenarios depends on the considered loss threshold, the proximity of networks to the seismic zones and the seismicity of the zones. On the other hand, different magnitudes and seismic sources can have similar contribution to a given loss exceedance; therefore a single scenario cannot be determined as the most probable one.



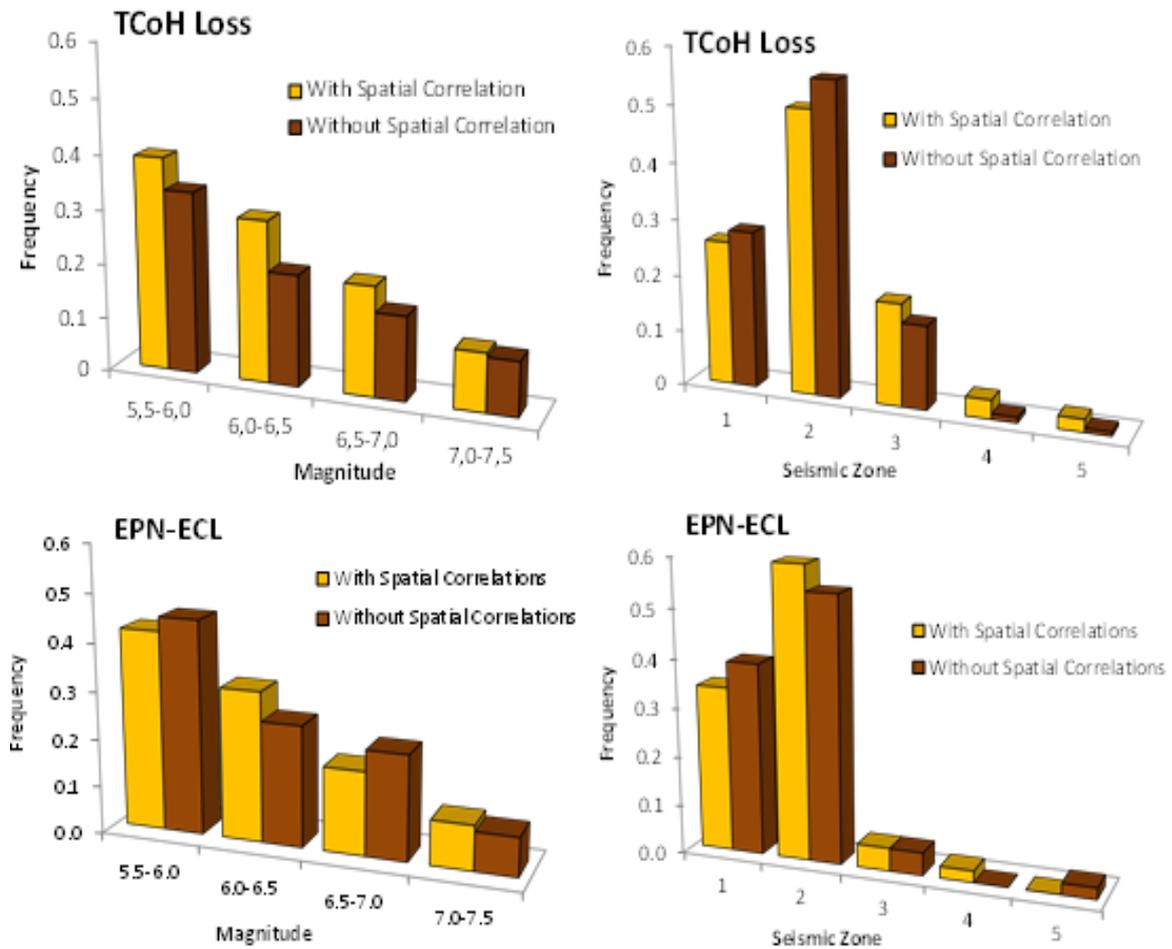
**Figure 5-5:** MAF curves for performance loss (EPN, WSS, RDN, HBR) with and without considering spatial correlations of intensity measures (1000 runs, one seismic scenario,  $M=6.5$ ).



**Figure 5-6:** MAF curves for performance loss (EPN, WSS, RDN, HBR) with and without considering spatial correlations of intensity measures (10000 runs, 5 seismic zones, M=5.5-7.5).



**Figure 5-7:** Correlation of components' functionality to system PI with and without considering spatial correlations of intensity measures (up: 1 seismic scenario, down: 5 seismic zones).



**Figure 5-8:** Disaggregation to magnitudes and seismic zones for given PI exceedance (5% for TCoH, 20% for ECL), with and without considering spatial correlations of intensity measures.



---

## References

- Abrahamson, N. A., and J. J. Bommer (2005). Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra* 21, 603-607.
- Anderson, J. G., S. G. Wesnousky and M. W. Stirling (1996), Earthquake Size as a Function of Fault Slip Rate, *Bull. Seismol. Soc. Am.*, 86, 683-690
- Akkar S, Bommer JJ. Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. *Seismological Research Letters* 2010; 812:195-206.
- Argyroudis S, Selva J, Kakderi K, Pitilakis K. Application to the city of Thessaloniki, in: K. Pitilakis et al. (eds). *SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*. Springer, Netherlands, 2014.
- ALA. Seismic fragility formulations for water systems. American Lifeline Alliance, ASCE, 2001.
- Abrahamson N.A. et al. 2002. "PEGASOS—A Comprehensive Probabilistic Seismic Hazard Assessment for Nuclear Power Plants in Switzerland: 12th European Conference on Earthquake Engineering, London." Paper No.6.
- Alzbutas, Robertas et al. 2005. "External Events Analysis and Probabilistic Risk Assessment Application for Iris Plant Design." in *The 13th international conference on nuclear engineering abstracts*.
- Barka, A. A. and K. Kadinsky-Cade (1988), Strike-slip fault geometry in Turkey and its influence on earthquake activity, *Tectonophysics*, 7, 663-684
- Basili R., et al. (2013), The European Database of Seismogenic Faults (EDSF) compiled in the framework of the Project SHARE, doi:10.6092/INGV.IT-SHARE-EDSF, available at: <http://diss.rm.ingv.it/share-edsf/> (last assessed September 2014)
- Bedford, T., and R. Cooke. 2001. *Probabilistic risk analysis: foundations and methods*, Cambridge University Press.
- Bilek, S. L. and Lay, T.: 2000, Depth dependent rupture properties in circum-Pacific subduction zones, In: J. B. Rundle, D. L. Turcotte and W. Klein (eds), *GeoComplexity and the Physics of Earthquakes*, American Geophysical Union, pp. 165–186.
- Bommer, J. J., and Scherbaum, F., 2008. The use and misuse of logic-trees in probabilistic seismic hazard analysis, *Earthquake Spectra* 24, 997–1009.
- Bommer, J. J., S. Oates, J. M. Cepeda, C. Lindholm, J. Bird, R. Torres, G. Marroquin and J.

- Rivas (2006), Control of hazard due to seismicity induced by a hot fractured rock geothermal project, *Engineering Geology*, 83, 287-306, doi: 10.1016/j.enggeo.2005.11.002.
- Burnham, K. P., and Anderson, D. R., 2002. *The Model Selection and Multimodel Inference: A practical information-theoretic approach*, 2nd edition, Springer, New York, NY, 488 pp.
- Bernreuter D.L. et al. 1989. "Seismic Hazard Characterization of 69 Nuclear Plant Sites East of the Rocky Mountains, Vols. 1-8."
- Bommer, Julian J. et al. 2013. "A SSHAC Level 3 Probabilistic Seismic Hazard Analysis for a New-Build Nuclear Site in South Africa." *Earthquake Spectra*.
- Bradley, Ralph Allan and Milton E. Terry. 1952. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika* 324–45.
- Cornell C, Krawinkler H. Progress and challenges in seismic performance assessment. *PEER Center News* 2000; 3(2). Douglas, J., B. Edwards, V. Convertito, N. Sharma, A. Tramelli, D. Kraaijpoel, B. Mena Cabrera, N. Maercklin and C. Troise (2013), Predicting Ground Motion from Induced Earthquakes in Geothermal Areas, *Bull. Seismol. Soc. Am.*, 103, 1875-1897, doi: 10.1785/0120120197
- Crowley H, Bommer J. Modelling seismic hazard in earthquake loss models with spatially distributed exposure. *Bulletin of Earthquake Engineering* 2006; 4(3):249–273. Dreyfus, H. L., and Dreyfus, S. E., 1986. *Mind over Machine*. The Free Press, New York, 252 pp.
- Committee, Senior Seismic Hazard Analysis and Robert J. Budnitz. 1997. *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*. US Nuclear Regulatory Commission Washington, DC.
- Coppersmith K.J. et al. 1996. Lessons learned—The Use of Formal Expert Assessment in Probabilistic Seismic and Volcanic Hazard Analysis, Chap. 20 of Connor, C., and Connor, L., Eds., *Volcanism, Tectonism, and Siting of Nuclear Facilities*. edited by Cambridge University and Press.
- Coppersmith, Kevin J., Robert R. Youngs, and Christian Sprecher. 2009. "Methodology and Main Results of Seismic Source Characterization for the PEGASOS Project, Switzerland." *Swiss Journal of Geosciences* 102(1):91–105.
- CRWMS M&O. 1996. *Probabilistic Volcanic Hazard Analysis for Yucca Mountain, Nevada: Las Vegas, Nevada*.
- CRWMS M&O. 1998. *Probabilistic Seismic Hazard Analyses for Fault Displacement and Vibratory Ground Motion at Yucca Mountain, Nevada*.
- Draper D., J. Hodges, C. Mallows and D. Pregibon (1993). Exchangeability and data analysis (with discussion). *J. Roy. Statist. Soc, Ser. A* 156, 9-37
- Dreyfus, H. L., and Dreyfus, S. E., 1986. *Mind over Machine*. The Free Press, New York, 252 pp.
- Douglas, J., B. Edwards, V. Convertito, N. Sharma, A. Tramelli, D. Kraaijpoel, B. Mena Cabrera, N. Maercklin and C. Troise (2013), Predicting Ground Motion from Induced Earthquakes in Geothermal Areas, *Bull. Seismol. Soc. Am.*, 103, 1875-1897, doi: 10.1785/0120120197

## References

---

- Duputel, Z., H. Kanamori, V. C. Tsai, L. Rivera, L. Meng, J. P. Ampuero and J. M. Stock (2012), The 2012 Sumatra great earthquake sequence, *Earth Planet. Sci. Lett.*, 351, 247-257, doi: 10.1016/j.epsl.2012.07.017
- Dziewonski, A. M. and Anderson, D. L.: 1981, Preliminary reference Earth model, *Phys. Earth Planet. Interiors* 25, 297–356.
- Esposito S, Iervolino, I. PGA and PGV spatial correlation models based on European multi-event datasets. *Bulletin of the Seismological Society of America* 2011; 101(5):2532-2541.
- Eurocode 8, EC8. Design of structures for earthquake resistance. European Committee for Standardisation. The European Standard EN 1998-1, 2008.
- ENSREG. 2013. National Action Plan Workshop, Summary Report.
- EPRI. 1989. Probabilistic Seismic Hazard Evaluations at Nuclear Power Sites in the Central and Eastern United States—Resolution of the Charleston Earthquake Issue: Palo Alto, Calif.
- EPRI. 2004. CEUS Ground Motion Project: Palo Alto, Calif.
- EPRI. 2008. Project Plan: Central and Eastern United States Seismic Source Characterization for Nuclear Facilities: Palo Alto, Calif.
- Faenza, L., and A. Michelini (2010). Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap, *Geophys. J. Int.* 180, 1138–1152.
- FEMA. HAZUS-MH Technical Manual. Federal Emergency Management Agency, Washington DC, USA, 2003.
- Field, E.H., N. Gupta, V. Gupta, M. Blanpied, P. Maechling, and T. H. Jordan (2005). Hazard calculations for the WGCEP-2002 earthquake forecast using open-SHA and distributed object technologies. *Seismol. Res. Lett.* 76, 161-167.
- Franchin P. A computational framework for systemic seismic risk analysis of civil infrastructural systems, in: K. Pitilakis et al. (eds). *SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications.* Springer, Netherlands, 2014.
- Franchin, P. & Cavalieri, P. (2013), Seismic vulnerability of a complex interconnected infrastructure. In: Tesfamariam, S. & Goda, K. (eds) *Handbook of seismic risk analysis and management of civil infrastructure systems*, Woodhead Publishing Ltd, Cambridge, UK.
- Galton F (1907) *Vox populi*. *Nature* 75(1949):450-451. Gelman A, Carlin J, Stern H, Rubin D (1995) *Bayesian data analysis*. Chapman and Hall/CRC, Boca Raton
- Geist, E. & Parsons, T., 2006. Probabilistic analysis of tsunami hazards, *Nat. Hazards*, 37, 277–314.
- Giardini D et al. Seismic Hazard Harmonization in Europe (SHARE). Online Data Resource, <http://portal.share-eu.org:8080/jetspeed/portal/>, doi: 10.12686/SED-00000001-SHARE, 2013.

- Grezio, A., Sandri, L., Marzocchi, W., Argnani, A., Gasparini, P. & Selva, J., 2012. Probabilistic tsunami hazard assessment for Messina Strait Area (Sicily, Italy), *Nat. Hazards*, 64, 329–358.
- Garcia-Aristizabal, Alexander et al. n.d. “Identifying and Structuring Scenarios of Cascade Events in the MATRIX Project.” *New Multi-Hazard and Multi-Risk Assessment Methods for Europe*.
- Gill, Joel C. and Bruce D. Malamud. 2014. “Reviewing and Visualizing the Interactions of Natural Hazards.” *Reviews of Geophysics* 52(4):2013RG000445. Retrieved (<http://dx.doi.org/10.1002/2013RG000445>).
- Goepel, Klaus D. 2013. “Implementing the Analytic Hierarchy Process as a Standard Method for Multi-Criteria Decision Making in Corporate Enterprises—a New AHP Excel Template with Multiple Inputs.” in *Proceedings of the international symposium on the analytic hierarchy process, Kuala Lumpur, Malaysia*.
- Goossens L.H.J., Cooke R.M., Hale A.R., Rodic-Wiersma L.J. (2008) *Fifteen Years of Expert Judgment at TU Delft*. Safety Science, vol 46. Elsevier.
- Hanks, T. C. and W. H. Bakun (2002), A Bilinear Source-Scaling Model for  $M$ -log  $A$  Observations of Continental Earthquakes, *Bull. Seismol. Soc. Am.*, 92, 1841-1846
- Harris, R. A., J. F. Dolan, R. Hartleb and S. M. Day (2002), The 1999 Izmit, Turkey, Earthquake: A 3D Dynamic Stress Transfer Model of Intraearthquake Triggering, *Bull. Seismol. Soc. Am.*, 92, 245-255
- Hanks, Thomas C., Norm A. Abrahamson, David M. Boore, Kevin J. Coppersmith, and Nichole E. Knepprath. 2009. “Implementation of the SSHAC Guidelines for Level 3 and 4 PSHAs—experience Gained from Actual Applications.” *US Geological Survey Open-File Report 1093:66*.
- Harker, Patrick T. and Luis G. Vargas. 1987. “The Theory of Ratio Scale Estimation: Saaty’s Analytic Hierarchy Process.” *Management science* 33(11):1383–1403.
- Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research & evaluation*, 12(10), 1-8.
- Huang, Tzu-Kuo, Chih-Jen Lin, and Ruby C. Weng. 2004. “A Generalized Bradley-Terry Model: From Group Competition to Individual Skill.” Pp. 601–8 in *Advances in Neural Information Processing Systems*.
- IAEA NS-G-1.2. 2001. “Safety Assessment and Verification for Nuclear Power Plants.”
- IAEA NS-G-1.5. 2003. “External Events Excluding Earthquakes in the Design of Nuclear Power Plants Safety Guide.”
- IAEA NS-G-1.6. 2003. “Seismic Design and Qualification for Nuclear Power Plants.”
- IAEA NS-G-2.10. 2003. “Periodic Safety Review of Nuclear Power Plants.”
- IAEA NS-G-3.4. 2003. “Meteorological Events in Site Evaluation for Nuclear Power Plants.”
- IAEA NS-G-3.5. 2003. “Flood Hazards for Nuclear Power Plants on Coastal and River Sites.”
- IAEA SSG-21. 2012. “Volcanic Hazards in Site Evaluation for Nuclear Installations.”

- Jayaram N, Baker J. Correlation model of spatially distributed ground motion intensities. *Earthquake Engineering and Structural Dynamics* 2009; 38:1687-1708.
- Jayaram N, Baker JW. Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering & Structural Dynamics* 2010; 39(10): 1109-1131.
- Jaynes, E.T. 2003. *Probability theory: the logic of Science*, Cambridge Univ. Press, New York.
- Kame, N., J. R. Rice and R. Dmowska (2003), Effects of prestress state and rupture velocity on dynamic fault branching, *J. Geophys. Res.*, 108, 2265
- Kakderi K, Selva J, Pitilakis K. Application in the harbor of Thessaloniki, in K. Pitilakis et al. (eds). *SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*. Springer, Netherlands, 2014.
- Keller M, Pasanisi A, Marcihac M, Yalamas T, Secanell R, Senfaute G, (2014) A Bayesian Methodology Applied to the Estimation of Earthquake Recurrence Parameters for Seismic Hazard Assessment, *Qual. Reliab. Engng. Int.*, 30 921–933.
- Kijko , A., and M. A. Sellevoll (1992). Estimation of earthquake hazard parameters from incomplete data files, Part II, Incorporation of magnitude heterogeneity, *Bull. Seismol. Soc. Am* 82, 120-134
- Kraaijpoel, D. and B. Dost (2013), Implications of salt-related propagation and mode conversion effects on the analysis of induced seismicity, *J. Seismol.*, 17, 95-107, doi: 10.1007/s10950-012-9309-4
- Koutsourelakis P (2010) Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: a Bayesian framework. *Probab Eng Mech* 25:49–60.
- Kulkarni, R. B., R. R. Youngs, and K. J. Coppersmith (1984). Assessment of confidence intervals for results of seismic hazard analysis, *Proceedings of the Eighth World Conference on Earthquake Engineering*, San Francisco, 1.
- Komendantova, Nadejda et al. n.d. "New Multi-Hazard and Multi-Risk Assessment Methods for Europe."
- Lindley, D.V. (2000). The philosophy of statistics. *Statistician* 49, 293-337.
- Liu, Z., F. Nadim, A. Garcia-Aristizabal, A. Mignan and K. Fleming, A three-level framework for multi-risk assessment, submitted to *Georisk*.
- Leonard, M. (2010), Earthquake Fault Scaling: Self-Consistent Relating of Rupture Length, Width, Average Displacement, and Moment Release, *Bull. Seismol. Soc. Am.*, 100, 1971-1988, doi: 10.1785/0120090189
- Mai, P. M. and G. C. Beroza (2000), Source Scaling Properties from Finite-Fault-Rupture Models, *Bull. Seismol. Soc. Am.*, 90, 604-615
- Marzocchi, W., L. Sandri, and J. Selva (2008). BET\_EF: A probabilistic tool for long-and short-term eruption forecasting, *Bull. Volcanol.* 70, 623–632.

- Marzocchi W., L. Sandri, J. Selva (2010). BET VH: a probabilistic tool for long-term volcanic hazard assessment. *Bull. Volcanol.* 72, 705-716.
- Marzocchi W., and T. H. Jordan (2014). Testing for Ontological Errors in Probabilistic Forecasting Models of Natural Systems. *Proc. Nat. Acad. Sci.* 85, 955-959.
- Marzocchi W., Taroni M., Selva J. (2015). Accounting for Epistemic Uncertainty in PSHA: Logic Tree and Ensemble Modeling, *Bulletin of the Seismological Society of America*, 105 (4), doi: 10.1785/0120140131.
- Mayo DG (1996) *Error and the Growth of Experimental Knowledge* (Univ of Chicago Press, Chicago).
- McGuire, R.K., C.A. Cornell, (Abrahamson, 2005) and G.R. Toro (2005). The case for using mean seismic hazard, *Earthquake Spectra* 21, 879-886.
- Mignan, A., D. Landtwing, P. Kästli, B. Mena and S. Wiemer (2015), Induced seismicity risk analysis of the 2006 Basel, Switzerland, Enhanced Geothermal System project: Influence of uncertainties on risk mitigation, *Geothermics*, 53, 133-146, doi: 10.1016/j.geothermics.2014.05.007.
- Mignan, A., Wiemer S. and Giardini D. (2014), The Quantification of Low-Probability–High-Consequences Events: Part I. A Generic Multi-Risk Approach, *Natural Hazards*, 73, 1999-2022, doi: 10.1007/s11069-014-1178-4
- Mignan, A., L. Danciu and D. Giardini, Reassessment of the maximum fault rupture length of strike-slip earthquakes and inference on  $M_{max}$  in the Anatolian Peninsula, Turkey, *Seismol. Res. Lett.*, in press
- Modaressi H, Desramaut N, Gehl P. Specification of the vulnerability of physical systems, in: K. Pitilakis et al. (eds). *SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*, Springer, Netherlands, 2014
- Musson, R. M. W. (2005). Against fractiles, *Earthquake Spectra* 21, 887-891.
- Marzocchi, Warner, Alexander Garcia-Aristizabal, Paolo Gasparini, Maria Laura Mastellone, and Angela Di Ruocco. 2012. "Basic Principles of Multi-Risk Assessment: A Case Study in Italy." *Natural hazards* 62(2):551–73.
- Morales-Napoles O., Worm D., Abspoel-Bukman L., Huijbregtse E., Courage W. (2015). Trends and uncertainties regarding rain intensities in the Netherlands. TNO-Report of ETP Project GAMES2R.
- Matthews, J. N. S. and K. P. Morris. 1995. "An Application of Bradley-Terry-Type Models to the Measurement of Pain." *Applied Statistics* 243–55.
- Newhall, C., and R. Hoblitt (2002). Constructing event trees for volcanic crises. *Bull. Volcanol.* 64, 3-20.
- NRC, National Research Council Panel on Seismic Hazard Evaluation (1997). *Review of Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, National Academy of Sciences, Washington, D.C., 84 pp., , ISBN:0-309-56207-4.
- Neri, A., Aspinall WP, Cioni R, Bertagnini A, Baxter PJ, Zuccaro G., Andronico D., Barsotti S., Cole PD, Esposti Ongaro T, Hinks TK, Macedonio G, Papale P, Rosi M, Santacroce

## References

---

- R, Woo G., (2008) Developing an Event Tree for probabilistic hazard and risk assessment at Vesuvius, *J Volcanol and Geotherm Res* 178(3), 397–415.
- Pitilakis K, Crowley E, Kaynia A (eds). SYNER-G: Typology definition and fragility functions for physical elements at seismic risk. *Geotechnical, Geological and Earthquake Engineering*, 27, Springer Netherlands, 2014.
- Popper KR (1983) *Realism and the Aim of Science* (Hutchinson, London).
- Pfister, Andrew, C. Goossen, K. Coogler, and J. Gorgemans. 2012. "AP1000 R Design Robustness against Extreme External Events-Seismic, Flooding, and Aircraft Crash." in *Proceedings of the 2012 International Congress on Advances in Nuclear Power Plants-ICAPP'12*.
- Romanowicz, B. and L. J. Ruff (2002), On moment-length scaling of large strike slip earthquakes and the strength of faults, *Geophys. Res. Lett.*, 29, 1604, doi: 10.1029/2001GL014479
- Rougier, J., Sparks, S., Hill, L.J. (eds.) (2013), *Risk and Uncertainty Assessment for Natural Hazards*, Cambridge University Press, Cambridge (UK), ISBN: 9781107006195.
- Scherbaum, F., and N. M. Kuehn (2011). Logic tree branch weights and probabilities: Summing up to one is not enough, *Earthquake Spectra* 27, 1237-1251.
- SSHAC (Senior Seismic Hazard Analysis Committee) (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute, NUREG/CR-6372, UCRL-ID-122160, Vols. 1/2.
- Selva J, Marzocchi W, Papale P, Sandri L (2012), Operational eruption forecasting at high-risk volcanoes: the case of Campi Flegrei, Naples, *J. Applied Volcanology*, 1:5, DOI:10.1186/2191-5040-1-5
- Selva, J., Argyroudis, S., and K. Pitilakis (2013). Impact on loss/risk assessments of inter-model variability in vulnerability analysis, *Nat Hazards*, 67 (2), 723-746, DOI 10.1007/s11069-013-0616-z
- Selva, J., and L. Sandri (2013) Probabilistic Seismic Hazard Assessment: Combining Cornell-Like Approaches and Data at Sites through Bayesian Inference, *Bulletin of the Seismological Society of America*, 103 (3), 1709–1722, doi: 10.1785/0120120091
- Selva J. Long-term multi-risk assessment: statistical treatment of interaction among risks, *Natural Hazards* 2013; 67(2): 701-722. SSHAC, Senior Seismic Hazard Analysis Committee (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute; NUREG/CR-6372, UCRL-ID-122160, vol. 1-2.
- Selva J, Tonini R, Molinari I, Tiberti MM, Romano F, Grezio A, Melini D, Piatanesi A, Basili R, Lorito S (2015), Exploring seismic source variability in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA), in prep.
- Smith, J.Q. (1988). *Decision analysis: a Bayesian approach*, Chapman and Hall, London.
- Stergiou E, Kiremidjian AS. *Treatment of uncertainties in seismic risk analysis of transportation systems*. Engineer Thesis, Stanford University, 2006.

- Straub D, Der Kiureghian A (2008) Improved seismic fragility modeling from empirical data. *J Struct Saf* 30:320–336
- Stirling, M., T. Goned, K. Berryman and N. Litchfield (2013), Selection of Earthquake Scaling Relationships for Seismic-Hazard Analysis, *Bull. Seismol. Soc. Am.*, 103, 2993-3011, doi: 10.1785/0120130052
- Stucchi, M., R. Camassi, A. Rovida, M. Locati, E. Ercolani, C. Meletti, P. Migliavacca, F. Bernardini, and R. Azzaro (Editors) (2007). DBMI04 il database delle osservazioni macrosismiche dei terremoti italiani utilizzate per la compilazione del catalogo parametrico CPTI04, *Quaderni di Geofisica* 49., 38 pp., <http://emidius.mi.ingv.it/DBMI04/> (last accessed April 2013) (in Italian).
- Stucchi, M., C. Meletti, V. Montaldo, H. Crowley, G. M. Calvi, and E. Boschi (2011). Seismic hazard assessment (2003–2009) for the Italian building code, *Bull. Seismol. Soc. Am.* 101, 1885–1911.
- SYNER-G (2013), Systemic seismic vulnerability and risk analysis for buildings, lifeline networks and infrastructures safety gain. European Collaborative research project (FP7/2007-2013). <http://www.syner-g.eu>.
- Saaty, Thomas L. 1990. "How to Make a Decision: The Analytic Hierarchy Process." *European journal of operational research* 48(1):9–26.
- Saaty, Thomas L. and G. Hu. 1998. "Ranking by Eigenvector versus Other Methods in the Analytic Hierarchy Process." *Applied Mathematics Letters* 11(4):121–25.
- Smith, Keith. 2013. *Environmental Hazards: Assessing Risk and Reducing Disaster*. Routledge.
- Vick SG (2002) *Degrees of Belief: Subjective Probability and Engineering Judgment* (ASCE Press, Reston, VA).
- Vrbanić, Ivan and Ivica Bašić. 2011. "PSA in the Context of Stress Tests." in *Nordic PSA Conference-Castle Meeting 2011*.
- Wells, D. L. and K. J. Coppersmith (1994), New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement, *Bull. Seismol. Soc. Am.*, 84, 974-1002
- Wesnousky, S. G. (2006), Predicting the endpoints of earthquake ruptures, *Nature*, 444, 358-360
- Wesnousky, S. G. (2008), Displacement and Geometrical Characteristics of Earthquake Surface Ruptures: Issues and Implications for Seismic-Hazard Analysis and the Process of Earthquake Rupture, *Bull. Seismol. Soc. Am.*, 98, 1609-1632, doi: 10.1785/0120070111
- Weatherill G, Esposito S, Iervolino I, Franchin P, Cavalieri F. Framework for seismic hazard analysis of spatially distributed systems, in: K. Pitilakis et al. (eds). *SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*. Springer, Netherlands, 2014.
- Winkler RL (1996) Uncertainty in probabilistic risk assessment. *Reliab Eng Syst Saf* 54:127–132.

## References

---

- Woessner J., Danciu L., Giardini D., Crowley H., Cotton F., Grünthal G., Valensise G., Arvidsson R., Basili R., Demircioglu M., Hiemer S., Meletti C., Musson R., Rovida A., Sesetyan K., Stucchi M. and the SHARE consortium 2013. The 2013 European Seismic Hazard Model. Submitted.
- Zahedi, Fatemeh. 1986. "The Analytic Hierarchy Process-a Survey of the Method and Its Applications." *interfaces* 16(4):96–108.
- Zio, E. 1996. "On the Use of the Analytic Hierarchy Process in the Aggregation of Expert Judgments." *Reliability Engineering & System Safety* 53(2):127–38.
- Zio, E. and E. Ferrario. 2013. "A Framework for the System-of-Systems Analysis of the Risk for a Safety-Critical Plant Exposed to External Events." *Reliability Engineering & System Safety* 114:114–25. Retrieved March 12, 2015 (<http://www.sciencedirect.com/science/article/pii/S0951832013000161>).
- Zwicky, Peter, Peter Fajfar, and Domenico Giardini. 2014. "Harmonized Approach to Stress Tests for Critical Infrastructures against Natural Hazards."