This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603839

# ASTARTE

## Assessment, STrategy And Risk Reduction for Tsunamis in Europe

**Grant Agreement no:** 603839
**Organisation name of lead contractor:** IPMA
**Coordinator:** Maria Ana Baptista

## Deliverable 8.39

## Quantifying and managing uncertainty in tsunami risk assessment methods: application to the ASTARTE test sites

| | |
|---|---|
| **Due date of deliverable:** | M38 |
| **Actual submission date to PC:** | M40 |
| **Start date of the project:** | 01/11/2013 |
| **Duration:** | 42 months |

| | |
|---|---|
| **Work Package:** | WP8 "From Hazard To Risk Assessment" |
| **Lead beneficiary of this deliverable:** | INGV |
| Author(s): | **INGV**: J. Selva, S. Lorito, S. Orefice, B. Brizuela, S. Iqbal, F. Romano, M. Volpe, R. Tonini, R. Basili, M.M. Tiberti; **NGI**: F. Løvholt, S. Glimsdal, C. Harbitz; **GFZ**: A. Hoechner, A. Babeyko; **UNIBO**: A. Armigliato, G. Pagnoni, S. Tinti; **IPMA**: R. Omira; **METU**: S. Acar, A. Yalciner, B. Aytore, N. Dogulu, O. Cabuk, U. Kanoglu, B. Yalciner; **KOERI**: O. Necmioglu, C. O. Sozdinler, M Ozel; **SRB-RAS**: A. Zaytsev |
| **Version:** | Final |

| **Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)** | |
|---|---|
| **Dissemination Level** | |
| **PU** Public | x |
| **PP** Restricted to other programme participants (including the Commission Services) | |
| **RE** Restricted to a group specified by the consortium (including the Commission Services) | |
| **CO** Confidential, only for members of the consortium (including the Commission Services) | |

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

ASTARTE project has dealt with uncertainty in tsunami hazard and risk assessment in several deliverables. To provide a homogeneous description of uncertainty treatment throughout the project, the main concepts are summarised into a single final document. This deliverable is meant to serve as initial set of guidelines for future assessments, to be likely updated during ongoing and future efforts (e.g. the Global Tsunami Model - GTM, www.globaltsunamimodel.org).

The earliest ASTARTE document about uncertainty treatment was the deliverable D4.13, where some concepts were developed already. Those concepts are further elaborated and extended to a more general framework here. We also deepen into the practical implementation of uncertainty quantification in tsunami hazard and risk assessment, while attempting to preserve a multi-risk perspective.

The general framework of uncertainty treatment described in D4.13 is first discussed and updated. Then, best practices and state-of-the-art in treating and quantifying uncertainty for natural hazards (with focus on seismic and volcanic applications) and highlighting some peculiarities for tsunamis is discussed also from a multi-hazard perspective. Several different methods adopted in the tsunami community, their rationale, pros and cons, and methods for comparing different results from different approaches are discussed also by means of examples of uncertainty treatment adopted in some of the hazard and/or risk analyses at the ASTARTE test sites, previously reported in WP8 deliverables (D8.8, D8.14, D8.33).

Finally, we introduce a novel general framework for tsunami hazard quantification at the regional scale, enabling a full exploration of uncertainty for tsunamis of seismic origin. The method also establishes a standardized protocol to manage subjectivity within a multiple-expert environment. The method is being applied in the TSUMAPS-NEAM project (http://www.tsumaps-neam.eu/), which will provide the first community-based regional Seismic Probabilistic Hazard Analysis for the NEAM region. TSUMAPS-NEAM was made possible by the work done within ASTARTE for the definition of good practices for the source treatment (D3.12, D3.40) and of methods for the regional-scale hazard analysis (e.g. D8.8). A multiple-expert approach for the management of critical choices was instead mainly developed within the STREST project (http://www.strest-eu.org/); its application in TSUMAPS-NEAM involves several ASTARTE partners. An example is drawn and presented from this regional-scale analysis for hazard calculated at offshore points. Finally, a proposed methodology for treating the uncertainty in approximated methods for the calculation of maximum inundation heights, where the uncertainty in inundation quantities is derived from numerical simulations at ASTARTE test sites, is presented.

# DOCUMENT INFORMATION

| | | | |
|---|---|---|---|
| **Project Number** | FP7 – 603839 | **Acronym** | ASTARTE |
| **Full Title** | Assessment, STrategy And Risk Reduction for Tsunamis in Europe | | |
| **Project URL** | http://www.astarte-project.eu/ | | |
| **Document URL** | | | |
| **EU Project Officer** | Denis Peter | | |

| | | | | |
|---|---|---|---|---|
| **Deliverable** | **Number** | D8.39 | **Title** | Quantifying and managing uncertainty in tsunami risk assessment methods: application to the ASTARTE test sites |
| **Work Package** | **Number** | WP8 | **Title** | From Hazard to Risk Assessment |

| | | | | |
|---|---|---|---|---|
| **Date of Delivery** | **Contractual** | M38 | **Actual** | M40 |
| **Status** | version | | final X | |
| **Nature** | prototype □ report X dissemination □ | | | |
| **Dissemination level** | public X consortium □ | | | |

| | | | | |
|---|---|---|---|---|
| **Authors (Partner)** | INGV | | | |
| **Responsible Author** | **Name** | Jacopo Selva | **E-mail** | jacopo.selva@ingv.it |
| | **Partner** | INGV | **Phone** | |

| **Version Log** | | | |
|---|---|---|---|
| **Issue Date** | **Rev. No.** | **Author** | **Change** |
| 0.1.0 | 06/09/2016 | INGV | Table of contents |
| 0.2.0 | 01/03/2017 | All | First draft |
| 0.3.0 | 28/03/2017 | All | Final draft |
| 1.0.0 | | INGV | Submitted |
| | | | |
| | | | |
| | | | |

# List of Figures

## List of Tables

# 1. Introduction

Aleatory uncertainty is quantified by the expected long-run frequencies of random events within the model of the system. Such frequencies are objective probabilities $\vartheta$, and they can be potentially measured through a well-defined experimental concept. The experimental concept defines collections of data - observed and not observed yet – that are judged to be exchangeable. The long-run frequencies are determined by this data-generating process (Marzocchi & Jordan, 2014). The quantification of the aleatory uncertainty is the main goal of any probabilistic model for hazard or risk. For tsunamis, we will indicate these analyses as Probabilistic Tsunami Hazard Analysis (PTHA) and Probabilistic Tsunami Risk Analysis (PTRA); we will also refer to PTHA for tsunamis of seismic origin as seismic-PTHA (S-PTHA).

Epistemic uncertainties measure the lack of knowledge in the estimation of such frequencies. Models assessing $\vartheta$ are often based on expert opinions, thus epistemic uncertainty is described by subjective probabilities. Bayesian methods are appropriate for reducing epistemic uncertainties as new knowledge is gained through observations. The epistemic uncertainty arising from this framework describes '*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, 2012; Marzocchi et al., 2015).

Different approaches have been used in literature of natural hazards to develop alternative models of the aleatory uncertainty, and merge them into a single model simultaneously quantifying both aleatory and epistemic uncertainty. The most common approaches are based on expert elicitations (e.g., Cook, 1991; Bedford and Cooke, 2001; Neri et al., 2008; Morgan, 2014), multiple-expert processes (SSHAC, 1997; Selva, et al. 2015), Bayesian methods (e.g., Gelman, et al. 1995; Selva and Sandri, 2013; Marzocchi et al., 2010; Grezio et al., 2010), Logic Trees (from Kulkarni et al., 1984; Geist & Parsons, 2006), and ensemble modelling approaches (e.g., Toth and Kalnay, 1993; Araujo and New 2007; Cloke and Pappenberger, 2009, Marzocchi et al., 2015; Selva et al., 2016).

The quantification of epistemic uncertainty is often limited to hazard assessments only, and more seldom considered in subsequent risk assessments; it has been recognized though that the variability of the results due to possible different and subjective choices related to the vulnerability may potentially introduce unforeseen consequences in loss/risk (e.g., Pate-Cornell,,z 1996, Selva et al., 2013).

Ontological error is identified by the rejection of a null hypothesis which states that the true (long-run) frequencies of the random events are samples from the (joint) probability distribution describing the epistemic uncertainties. In other words, if this null hypothesis is rejected, it means that our model describing the uncertainty on the definition of the true frequency is 'ontologically' rejected, possibly indicating that the selected framework is not capable of satisfactory describing uncertainty (Marzocchi and Jordan, 2014). The possibility

of formally testing the full uncertainty quantification against real (and independent) data is one of the main advantages of the described framework (Marzocchi et al., 2015).

In the following sections, we discuss how the quantification of uncertainty is generally treated in the field of natural hazards (with focus to earthquakes and volcanoes, which are also sources of tsunamis), and we compare the main methods adopted within ASTARTE, which well represent the most used methods adopted in the common practice of tsunami hazard and risk assessments.

Quite advanced techniques in uncertainty quantification have been historically developed in the field of seismic hazard and risk assessment (SSHAC, 1997, 2012; Cornell and Krawinkle, 2000; Der Kiureghian, 2005). Generally non homogeneous levels of quantification of these types of uncertainty are instead adopted in the scientific literature when dealing with tsunami hazard and risk analysis.

Some of the techniques used in different fields may be extended to tsunamis (e.g., Geist and Parsons, 2006). However, many technical and practical limitations do exist, mainly due to several peculiarities of tsunamis (e.g., Geist and Lynett, 2014; Lorito et al., 2015; Selva et al., 2016; Davies et al., 2017). To facilitate this discussion, some examples of applications to the ASTARTE test sites of Gulluk bay (Turkey), Sines (Portugal), and Siracusa (Italy) are presented, together with the comparison and critical review of complementary and/or alternative scenario-based and probabilistic approaches to hazard and risk analysis.

By limiting ourselves to seismic sources only, we finally discuss and present an example which illustrates the methodology for uncertainty quantification in a regional S-PTHA project. Clearly, all methods as well as any specific assessment may involve rather subjective choices at all levels, which may be even hidden in the complexity of the process. This very much critical aspect of hazard and risk analysis is a challenge for scientists and practitioners responsible for the analysis. Here, we then also discuss, in the frame of the above-mentioned regional assessment, a possible structured process for adopting working simplifications; for performing critical choices between possible alternative models; and for their relative weighting.

Overall, this approach represents a refinement of the method described in deliverable D8.8 of ASTARTE and published in Lorito et al. (2015) and Selva et al. (2016); it nevertheless benefitted of the huge amount of work made within ASTARTE by many partners; and also by the efforts developed in the STREST project. TSUMAPS-NEAM is in a sense a spin-off of the ASTARTE project; however, it clearly marks the transition between hazard analyses and studies with uncertainty treatment on one hand, towards an operational assessment motivated by regulatory concerns on the other hand, and commissioned by the DG-ECHO for Civil Protection purposes.

Since the computational feasibility of site-specific / high-resolution probabilistic inundation maps is a challenge *per-se*, an example is also presented here concerning a possible two-step approximate method for dealing with tsunami inundation quantities. Rather crude approximate methods based on local amplification factors of the offshore wave amplitude probabilities are first used. Then, detailed numerical simulations at the ASTARTE test sites are used for assessing a probability density function, which accounts for uncertainty propagation due to the crude inundation approximation (see also Davies et al., 2017). The approach developed here is being as well applied in the TSUMAPS-NEAM assessment.

Ideally, the present deliverable will be a first step of a longer-term process initiated within ASTARTE, as it is foreseeable that these guidelines will be further developed and updated in the future, for instance by the Global Tsunami Model (GTM, www.globaltsunamimodel.org), a network of coordinated action for tsunami hazard and risk assessment worldwide. This GTM has been already endorsed by GFDRR and by UN-ISDR, with the auspice of contributing to the implementation of the 2015-2030 Sendai framework for disaster risk reduction (http://www.unisdr.org/we/coordinate/sendai-framework).

# 2. Treatment of uncertainty in hazard and risk, in a multi-risk perspective

In any given place, different phenomena may occur that can represent a hazard for an asset or a person located in that place. However, it is clear that the specific interest of any individual or stake holder or decision maker is to prevent the potential odd consequences independently from the hazard originating the consequence. For instance, a municipality may be interested to know in which direction and field the mitigation efforts have to be implemented (which hazard should we try to mitigate first?), considering that available economical and human resources are always limited.

Even if this type of questions is of fundamental and evident interest for the society, it is not a simple task so far to answer quantitatively and exhaustively. The quantification of risks is indeed usually done through the independent analysis of single hazardous phenomena, developed in different fields of science. Additionally, this implies the inappropriate in many cases assumption of independence among the different hazards. Therefore, the quantification of the risks caused by different hazardous phenomena is based on different definitions and/or approaches and/or assumptions, making them substantially not comparable (e.g., Marzocchi et al., 2009).

Different studies have shown the importance of making compatible assessments, allowing for a quantitative comparison of the different risks (e.g., FEMA, 1997; Grünthal et al., 2006; Douglas, 2007), as well as, for the treatment of the important interaction that may occur along the risk chains due to different hazard initiators (e.g., Bovolo et al., 2009; Marzocchi et al., 2009, 2012; Selva, 2013; Mignan et al., 2014, 2016).

The development of parallel and homogeneous and coordinated management of risks (multi-hazard risk approach) requires also a correct treatment of uncertainties related to each individual risk assessment and in the “aggregated” results. Indeed, a common set of boundary conditions for the risk quantification (such as the exposure time window and/or the definition of the considered type of damage) are only a pre-requisite for an effective risk comparison. Very different levels of knowledge exist for the different hazards. For example, while seismic hazard and risk analyses have nowadays rather well defined standards adopted worldwide (e.g., SSHAC, 1997; USNRC, 2012; Woessner et al., 2015), at least as far as the management of the critical choices and the communication of uncertainties are concerned, risks due to other hazards do not have accepted standardizations of the assessment process itself (e.g., the ones linked to volcanic sources, see IAEA-PVHA, 2016; Loughlin et al., 2015; or for tsunamis, hence the GTM, www.globaltsunamimodel.org).

Very different levels of knowledge exist even for the different factors of the same risk quantification. For example, the different choices on vulnerability modelling in risk assessments have been seen to result in significant discrepancies between the seismic risk assessments made by different authorities for the same location, structure type and seismicity (e.g., Rossetto and Elnashai, 2005; Selva et al., 2013). Also, for some structures many different alternative vulnerability models may be available (e.g., RC buildings), while for other structures only few models (if any) may be available (Pitilakis and Crowley, 2015), leading to completely different levels of confidence on the resulting risk models. If all the sources of uncertainty are not correctly propagated into the risk assessment, the differences between detailed multiple-expert analyses and simplified single-expert quantifications may completely disappear or be overlooked, leading to potentially misleading comparisons (see discussions in Selva et al., 2013).

In the next subsections, we first briefly discuss the state of the art for risk assessments within a multi-hazard environment in ***Section 2.1***. In ***Section 2.2***, we briefly recap the main tsunami hazard and risk assessment methods, discussing the main specificities that tsunami hazard/risk quantifications have with respect to other hazards. Then, in ***Section 2.3***, we compare the different methods, providing examples from the ASTARTE case studies and delineating some guidelines for the comparison of the results obtained when different methods are adopted.

## 2.1 State of the art for natural hazards

As discussed in *Section 1*, the general framework for the quantification of uncertainties in risk analyses has been anticipated in ASTARTE in deliverable D4.13, in order to provide a general homogenized reference framework for the assessments at an early stage of the project. Here, this framework is first discussed in more general terms, without specifically referring to tsunamis.

The quantification of risk induced by natural phenomena is commonly based on the combination of three main risk factors: the probability that a certain event will occur, the value of the assets that may be affected by the event and the extent of damages that the event may induce in these assets. Thus, it is often quantified by

$$R = H \times L \times V \tag{2.1}$$

where H is the hazard, V the vulnerability, and L the value at risk (e.g., EN 1050, 1997; Marzocchi et al., 2012a). Equation 2.1 is obviously a simplification since hazard, vulnerability and value cannot be satisfactory described by single values. This equation is rather a conceptual non quantitative description of the problem. It should be also clear that risk is a non-normalized probability, with a zero lower limit (no appreciable possibility of an adverse event or the event cannot cause damages), but not necessarily an upper limit (unless Equation 2.1 is provided as percentages of the area's total value, i.e., normalized risk, Marzocchi et al., 2012a).

In the next subsections, we first discuss the basic principles allowing for the quantification of uncertainties (*Section 2.1.1*) in hazard and risk assessments. Then, we discuss the general formulation (*Section 2.1.2*) that is required to quantify such uncertainty, discussing in details the most common factorizations (*Section 2.1.3*) of hazard and vulnerability assessment. This quantification framework leads to a rather complex computational framework that, in many cases, requires specific simplification strategies to be adopted in real case study. In *Section 2.1.4*, we briefly discuss the most common simplification strategies adopted.

### 2.1.1 Quantification of uncertainties in hazard and risk assessments

A natural event (e.g., an earthquake) may generate a phenomenon (e.g., seismic waves) that reaches a place and may cause damages. Different events may cause similar phenomena (e.g., different earthquakes may cause similar ground shaking at a given place). Also, the same phenomenon may cause different damages (e.g., similar ground shaking may cause different level and type of damage to one asset). This is due to the natural variability of the phenomenon that is usually referred to as aleatory uncertainty.

Hence, since unpredictable in the strict sense, one would desire to know the probability that a certain type of damage will occur in the future, considering this natural variability. To

answer to this apparently simple question, a probabilistic formulation of hazard and risk is required.

The basic principle of the probabilistic formulation is that – ideally – all the potential sources and the potential consequences should be taken into account, in order to quantify the probability that the odd consequence (e.g., the collapse of one building) happens within a time window (exposure time). To do this, the total hazard or risk can be computed by applying the theorem of total probability, which expresses the total probability of an outcome in terms of the conditional probabilities of the events through which it can be realized. This means that the potential consequences that all the possible sources may cause should be quantified and combined, by weighting them by their probability of occurrence. In practice, this means that any hazard or risk quantification should take into account all the potential sources that may cause the phenomenon (e.g., variability in earthquake generation), the uncertainties related to the propagation from source to target of the hazardous phenomenon (e.g., for seismic waves propagation), and of the generation of damages (e.g., variability in the response of the asset).

In probabilistic hazard and risk assessments, all aleatory uncertainties are formally included, as it will be discussed in *Section 2.1.2*. However, alternative formulations of the probabilistic framework, also concerning the very physics and modelling of the problem at hand, are usually possible. Due to our limitations in the knowledge of the studied system and the paucity of past data, we often cannot quantitatively select among scientifically acceptable (not falsifiable) alternative models. This uncertainty is usually referred to as epistemic uncertainty. Note that the definition of aleatory and epistemic may be source of debate among practitioners. In the followings, we refer to the definitions and discussion already addressed in *Section 1* of this deliverable and in the previous D4.13, while adopting a quite pragmatic approach in the present discussion.

The quantification of epistemic uncertainty derives from alternative applications of the probabilistic formulation, adopting different modelling choices for one or more of the risk terms. The application of these alternatives, along with the quantification of the representativeness of such alternatives within the technical community, is used to quantify the so-called community (or ensemble) distribution (e.g., Bommer, 2012; Marzocchi et al., 2015) describing the epistemic uncertainty (e.g., SSHAC, 1997).

Alternative formulations can be separated into two main broad groups, exploring the variability of the values of parameters within statistical models or, on the other side, exploring the one related to alternative models and parameterizations (e.g., Bommer and Scherbaum, 2008; Rougier et al., 2013). Many methods have been proposed to treat parameter uncertainty (e.g., Kijko and Sellevoll, 1992; Straub and Der Kiureghian, 2008; Koutsourelakis, 2010; Franchin and Cavalieri, 2013; Keller, 2014). Most of them are based on hierarchical Bayesian procedures, in which statistical distributions on the parameters values are adopted to model their potential variability. Conversely, few methods have been

proposed in geosciences to treat both sources simultaneously. In the scientific literature dealing with natural hazards, different techniques have been adopted in different contexts, such as logic trees (from Kulkarni et al, 1984), ensemble modelling (Toth and Kalnay, 1993; Araujo and New 2007; Cloke and Pappenberger, 2009; Marzocchi et al., 2015; Selva et al. 2016) and Bayesian modelling (Gelman et al., 1995; Straub and Der Kiureghian, 2008; Selva and Sandri, 2013; Selva et al., 2013). Other methods are more strictly related to extract the relevant information directly from experts, like structured expert elicitation methods based on mathematical aggregation (e.g., Cook, 1991; Bedford & Cook, 2001; Neri et al., 2008; Morgan, 2014), or on behavioral aggregation based on complex group interaction procedures (e.g., Delphi methods as in Linstone and Turoff, 1975; multiple-expert integration as in SSHAC 1997, Budnitz et al., 1998, USNRC 2012).

The development of alternatives is often produced in a multiple-expert context (SSHAC 1997 and following specifications), in order to standardize all the procedural aspects of expert interaction, and produce more robust results. This specific topic is discussed in more details in *Section 3*.

### 2.1.2 Basic formulation of probabilistic hazard and risk analyses

A more comprehensive formulation of risk than that in the generic definition of Equation 2.1 is based on the quantification of aleatory uncertainty by means of risk curves. Risk curves represent the probability of exceeding different loss levels due to the potential occurrence of a class of events (e.g., earthquakes, volcanic eruptions, tsunamis, floods, etc).

More specifically, the risk curve due to a generic event *E* for a given area in a given exposure time $\Delta$T represents the probability that a given loss value *l* is overcome in a target area and in the exposure time $\Delta$T. In deliverable D4.13, we discussed a formulation representing a simplification of the Pacific Earthquake Engineering Research (PEER) formula (Cornell and Krawinkle, 2000; Der Kiureghian, 2005), that has been proposed as general formulation for any kind of natural risks in several papers (Marzocchi et al., 2012a; Selva et al., 2013, Mignan et al., 2014) and European projects (Marzocchi et al., 2009; MATRIX: Liu et al., 2015; STREST: Esposito et al., 2017). This formulation reads:

$$\boldsymbol{\lambda(L \geq l; \Delta T) = \int_d \int_h G(l|d) \cdot dG(d|h) \cdot d\lambda(h)} \quad (2.2)$$

where

- *l* represents a loss measure in a specific metric (e.g., economic loss, casualties, dead); *d* represents a given damage level (ranging from no damages to complete loss); *h* a given hazard intensity measure (typically, but not necessarily a scalar). In different contexts, they are given different names. For example, in seismic engineering, h is often referred to as Intensity Measure (*IM*), *d* as Damage Measure (*DM*) and *l* as decision variable (DV) (e.g., Yeo and Cornell, 2005).
- $\lambda(x)$ represents the mean annual rate of exceedance of a generic variable *x.*
- $G(y|x)$ represents a conditional survivor function, that is, $G(y|x)$ represents the probability of exceedance of *y*, conditioned to the value *x*.

The hazard intensity measure *h* is usually (but not necessarily) selected as a scalar variable that well correlate with damages (e.g., Pinto, 2007). In common practice, all the variables *h*, *d* and *l* are evaluated at a discrete number of levels, each of them effectively representing an entire interval of possible values (e.g., NIBS, 2004). A rougher discretization is typically adopted for the damage measure, which is often defined over a relatively small number of pre-defined limit states (e.g., Minor, Moderate, Extensive and Complete damages).

The curve $\lambda(h)$ represents the mean annual rate of exceedance of a given value for the hazard intensity *h*, and it is typically referred to as *hazard curve* (e.g., SSHAC, 1997). The hazard curves are usually quantified adopting a computational approach, that is, by modelling the occurrence of all possible sources and the propagation of the hazardous

phenomenon from the source to the target, and then aggregating the results into the hazard curves as absolute or exceedance probabilities or rates of a set of pre-defined intensity thresholds.

An alternative approach is the empirical one, in which data at the target are used to fit a probabilistic model. Usually the computational approach is preferred over the empirical one, since (especially for rare events like earthquakes, tsunamis, volcanic eruptions, etc.) the past data rarely may completely cover the full variability of possible hazard intensities.

The distribution *G(d|h)* represents the fragility function of a given target asset, that is, the probability that the damage level *d* is overcome due to a hazard intensity h (e.g., ALA, 2001; NIBS, 2004). In their original form, fragility curves have been introduced in seismic risk analysis and they quantify the probability of a structure being damaged beyond a specific damage state for various levels of ground shaking (e.g., ALA, 2001; NIBS, 2004). Since then, fragility curves have been widely adopted in seismic loss and risk assessments (e.g., NIBS, 2004; Cornell and Krawinkler, 2000; Pitilakis et al., 2006). Indeed, they not only provide a quantitative base for risk quantification accounting for uncertainties, they also provide contributes in the retrofitting decisions, emergency response planning and estimation of direct and indirect losses of built environments as well as lifeline systems (Pitilakis et al., 2006; Kappos et al., 2008; Azevedo et al., 2010). More recently, fragility curves have been adopted for other hazards (e.g., Spence et al., 2005; Suppasri et al., 2013), and have been proposed as the general framework for vulnerability assessment for all natural risks (e.g., Douglas, 2007; Schmidt et al., 2011). Many methods are used to quantify fragility models, such as the empirical method based on fitting a statistical model to past recorded damages (e.g., Basöz et al., 1999; Maruyama et al., 2010; Rossetto and Elnashai, 2003; Suppasri et al., 2013), analytical modeling of the structures (e.g., Akkar et al., 2005; Moschonas et al., 2009), expert judgment (e.g., ATC-13, 1985; ATC-25, 1991) or on a combination of such methods (hybrid methods, e.g., Kappos et al., 2006; for buildings).

The distribution G*(l|d)* is the probability that a given loss level *l* is overcome, given the damage level *d*. In practice, the expected cost to repair a generic asset is often written as fraction of the total replacement cost and a cost damage factor that defines the fraction of replacement cost necessary to repair a five damage state *d* (Stergiou and Kiremidjian, 2006).

The results of the convolution of *Equation (2.1.2)* is $\lambda(l)$ that represents the cumulative mean annual rate of exceedance of a given value of loss *l.* The curve $\lambda(l)$ is typically referred to as *risk curve*. In order to allow a simpler comparison among the same risk in different areas, as well as, different risks in the same area, a single-risk index is often considered, instead of using the whole risk curve (e.g., Marzocchi et al. 2009). As risk index, the average (mean) of losses in the target area in the exposure time is often considered. In the case of earthquakes, by normalizing the average per 1 year, this risk index is referred to as Average Annual Earthquake Losses (AEL in NIBS, 2004; FEMA, 2008).

All the terms of this formulation are assumed valid and stationary within the exposure time $\Delta$T. Starting from $\lambda(I)$, it usually quantifies the probability of observing at least one exceedance episode within the exposure time $\Delta$T. To do that, it is required to define a generation process, which is usually assumed to be Poisson (starting from Cornell, 1968). In this case, the probability of observing at least one exceedance episode in $\Delta$T is computed as 1 minus the probability of observing exactly 0 exceedances, that is:

$$\mathbf{Pr}(\boldsymbol{X} \geq \boldsymbol{x}; \boldsymbol{\Delta T}) = \mathbf{1} - \mathbf{exp}\,(-\boldsymbol{\lambda}(\boldsymbol{x}) \cdot \boldsymbol{\Delta T}) \tag{2.3}$$

In several cases, the Poisson process hypothesis does not really hold. For example, earthquakes are typically clustered in space and time and, to adopt the Poisson assumption, the analysis is referred to the so-called mainshocks (main events in each cluster), individuated through a declustering procedure of the original catalogue of events (e.g. Gardner and Knopoff, 1974; Reasenberg, 1985). Analogous reasoning can be made for tsunamis (e.g., Geist and Parsons, 2008).

Epistemic uncertainty arising from the adoption of alternative modelling procedures can be quantified by modelling the different probability values obtained, for example, from Equation (2.3). This variability can be interpreted in the framework of ensemble modelling, as introduced by Marzocchi et al. (2015) and, specifically for tsunamis of seismic origin, by Selva et al. (2016). In this framework, as previously discussed, aleatory uncertainty is quantified by objective probability θ that represent the expected long-run frequencies of random events within the model of the system. Epistemic uncertainties measure the lack of knowledge in the estimation of θ. This uncertainty is quantified through a probability distribution, described by subjective probabilities, that describes '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; USNRC, 2012; Bommer, 2012; Marzocchi et al., 2015; Selva et al., 2016).

To quantify such community distribution, a finite set of different models $\{\vartheta_i , \omega_i\}$ ($i$ = 1, . . . , N) can be developed, where $\vartheta_i$ and $\omega_i$ are the outcome and the weight of the $i$-th model. The N different models describe the variable of interest $\vartheta$, that is, in this case, the hazard or risk curve of *Equation (2.3)*. Each alternative model is considered as a sample of the unknown parent distribution $f(\vartheta)$, the community distribution. In order to produce a consistent ensemble model, the set of samples and weights $\{\vartheta_i , \omega_i\}$ should represent an unbiased sample of the epistemic uncertainty (see discussion in Marzocchi et al., 2015), that is, they should consistently represent the variability of models and opinions within the reference community. The development of alternatives and the quantification of weights should therefore emerge from a multiple-expert process (e.g., SSHAC, 1997), as discussed in *Section 3*.

### 2.1.3 Factorization of risk factors

Further factorizations are typically introduced in each of the factors in *Equations 2.2*, for both the hazard and the fragility factors.

As discussed in the deliverable D4.13 for PTHA, while empirical approaches target to assess directly the hazard curves from past data, in computationally based hazard assessments the hazard factor is often factorized by source and propagation contributions:

$$\lambda(h) = \int_{\Sigma} G(h|\sigma) d\lambda(\sigma) \tag{2.4}$$

where, again, *h* represents the hazard intensity at the target, and $\lambda$ and *G* represent mean annual rates and a probabilistic survivor function, respectively, while $\sigma$ represents a single possible source in the set $\Sigma$ that includes all the possible sources that may cause an intensity *x* at the target. This factorization allows to separate the analysis of sources σ, their aleatory variability Σ and their mean annual frequencies λ(σ), for any σ ∈ Σ, from the analysis of the propagation of the tsunami from any source σ to the site of interest.

The source factor $\lambda$(x) is usually further factorized in at least two parts (from Cornell, 1968), expressing the size of the event (e.g., magnitude for an earthquake, volcano explosivity index or intensity for volcanic hazard, volume for landslides) and a spatial factor (e.g., hypocentres, vent position and landslide position, for earthquakes, eruptions and landslides, respectively). Further factors do depend on the specific hazard under consideration. These subfactors are usually quantified using an empirical approach, that is, a statistical model fitted to the empirical data (e.g., smoothed seismicity fitted to a seismic catalog, e.g., Hiemer et al., 2014).

The propagation factor *G(h|σ)* quantifies the probability that a given hazard intensity is overcome at the target, given the occurrence of the source *σ,* that is, *G(h|σ) = Pr(H ≥ h|σ)*. Depending on the specific problem, this probability distribution may include different processes, like boundary conditions for the propagation like wind for tephra fallout of volcanic hazard (e.g., Costa et al., 2009), or intra-event variability of ground motion prediction equations (e.g. Cotton et al., 2008). It can also include some variability not included in the source modelling, like the intra-event variability of ground motion prediction equations, or earthquake slip distribution for tsunami hazard (e.g., Davies et al., 2017). For some hazards (e.g., seismic hazard or regional tsunami hazard), this probability is factorized in propagation and site amplification (e.g., Field et al., 2000; Løvholt et al, 2012). The quantification of this propagation factors may be based on empirical data (e.g., ground motion prediction equations of seismic hazard), or on explicit (time-domain) numerical modelling (e.g., tsunami or volcanic hazard).

Also the fragility factor may be further factorized, depending on the analysis. For example, in seismic risk analysis (e.g., Yeo and Cornell, 2005) the fragility factor is sometimes factorized by considering the engineering demand parameters (EDPs) as an intermediate factor between hazard intensity *h* and damage *d*, that is

$$\mathrm{G}(d|h) = \int G(d|edp) \cdot dG(edp|h) \tag{2.5}$$

In practice, this means that the hazard intensity is first translated in what the structure should do to resist to the hazard sollicitation (demand), and the probability that a given structure can sustain this demand. The quantification of these factors is usually based on direct modelling of the structures, adopting a variety of empirical hazard inputs (e.g., ground motion time history caused by a given earthquake).

### 2.1.4 Common simplifications of the general formulation

As discussed in ***Section 2.1.2*** and ***Sections 2.1.3***, the assessment of the factors of ***Equation 2.1*** and the quantification of the epistemic uncertainty through the community distribution are rather demanding, in terms of scientific and computational effort. There is indeed a clear trade-off between the computational feasibility and the exploration of uncertainty (e.g., for tsunamis, and relatively to the computational cost of high-resolution inundation simulations, Lorito et al., 2015). For this reason, several simplification approaches are often adopted.

The level of simplification should in theory depend on the regulatory concern of the assessment (e.g., nuclear power plant vs small infrastructures) and/or the type (and the cost) of decisions that will be based on the assessment (e.g., SSHAC, 1997; Marzocchi and Woo, 2009; Esposito et al., 2017). For example, if there is no regulatory concern or the decision to be taken does not imply significant societal and/or economical consequences, large simplifications may be acceptable. Other important examples of the dependency of the level of the simplification on the goal of the assessment include the scale of the problem (if the analysis is required at a regional scale, many site-specific details and modelling sophistications can be avoided), the sensitivity of the results on model sophistications (specific requirements are only sometimes required; for example, spatial correlations can be in many case neglected, since they are often required only if there are inter-dependencies among spatially distributed components, e.g., Argyroudis et al., 2015), or the amount and the quality of the available data (there is no need of sophistications if the epistemic uncertainty will hide their impact and/or the data cannot support the sophistications).

Many different simplifications strategies have been selected in literature, based mainly on reducing i) the quantification of the epistemic uncertainty, ii) the quantification of the aleatory uncertainty, and/or iii) the total computational cost.

The simplification in the quantification of the epistemic uncertainty is typically based on reducing the number of alternatives to be implemented, even to the limit of one model implementation (no epistemic uncertainty). In other cases, only the mean or the modal models are considered. Another potential simplification arises in how alternative models are selected and weighted (e.g., see definitions of levels in SSHAC, 1997).

The simplification in the quantification of the aleatory uncertainty is usually based on the reduction of the source variability. This is reducing the number of discrete scenarios to be accounted for, by adopting coarser and coarser grids or simplification strategies like Event Trees (for tsunamis of seismic origin: Lorito et al., 2015; Selva et al., 2016). This reduction is often pushed toward the adoption of very few scenarios, like for example the use of reference scenarios. In these cases, the most critical point becomes the selection of such reference scenarios.

The reduction of the total computational cost is usually obtained by reducing the resolution and the number of times the propagation models are to be used. Propagation models are indeed typically highly sophisticated and computationally challenging. Depending on the cases, one possible and often adopted strategy to reduce the computational cost of hazard and risk assessments by replacing the direct propagation modelling with analytic or very simplified models (e.g., IAEA-Volcanoes, 2012), as well as, fully empirical (statistical) laws (as for earthquake Ground Motion Prediction Equations, e.g., Cotton et al., 2008).

Another drastic simplification of the computational cost is to build the analysis only upon past observations, as for the empirical probabilistic hazard assessments (e.g., for tsunamis: Geist & Parsons 2006). However, computationally based analyses are usually preferred, due to the scarcity of past observations. Indeed, such data cannot fully represent the effective variability of potential phenomena at source and at target (e.g., for tsunamis: Geist & Lynett, 2014), especially in the tails of distributions where events are very rare (e.g., Geist & Parsons, 2014).

## 2.2 State of the art for tsunamis

This section focusses on comparing the quantification frameworks and simplifications usually adopted for tsunami hazard and risk assessments with the general assessment framework discussed in *Paragraph 2.1*. Some examples concerning the modelling of the tsunami sources and the simplifications that can be made in view of tsunami hazard analysis have been already discussed in D3.40, which is then in this respect complementary to this deliverable.

Here, in particular, in *Section 2.2.1* we highlight the main peculiarities that tsunamis have in comparison with other natural hazards. In *Section 2.2.2*, we recap from deliverables D8.8, D8.14 and D8.33, and we describe the main methodologies adopted in ASTARTE, also in comparison with these peculiarities.

### 2.2.1 Comparison between tsunamis and other natural hazards

With respect to other natural hazards, tsunamis have several peculiarities that influenced the development of methods for the quantification of uncertainty. It is possible that these peculiarities led, in time, to many simplifications that are reflected in the methods adopted for tsunami risk and hazard quantification methods.

The attention to quantitative uncertainty assessment was only lately taken into progressively more serious consideration, mainly after the 2004 Indian Ocean tsunami. One reason for this is that both the 2004 Indian Ocean and the 2011 Tohoku tsunamis were caused by earthquakes of a magnitude which was unanticipated at that locations. Some features of those earthquake ruptures had probably never testified before, such as the rupture length on one hand for the Sumatra-Andaman earthquake, and the amount and characteristics of the slip distribution (e.g. the huge slip close to the trench); taken as a whole, these two events document the possibility of a large variability around the mean values of empirical earthquake scaling laws. Several other earthquakes and tsunamis in the last decade or so have shown somehow surprising features as well (see e.g. Lorito et al., 2016). It is moreover well known that the Tohoku earthquake also raised serious questions on how to improve tsunami warning systems as well as to increase the resilience of coastal communities.

Indeed, fully probabilistic methods that quantify (epistemic) uncertainties have been only recently developed for tsunami hazard but mainly focussing in the treatment of seismic sources (e.g. Selva et al., 2016; Davies et al., 2017); and sometimes only for some components of the tsunami risk assessment (for example only the PTHA: Chock et al., 2016).

Some peculiar features of tsunamis (not necessarily exclusive) which influence the process of hazard and risk assessment are briefly discussed in the next pages. This is however not meant to be an exhaustive discussion, but a starting point for future discussions (e.g. within the GTM); the purpose is only to highlight with some examples the need for proper uncertainty quantification.

*Tsunamis are low frequency high impact natural disasters*. A direct implication of this is that the past tsunami observations are relatively scarce, and that computationally-based hazard assessments are complex and characterised by an inherent high uncertainty; but the challenge is worth the effort given the potentially disastrous consequences of tsunamis. Appropriate methods need to be used for taking into account the uncertainty associated to rare events (e.g. Geist and Parsons, 2014; Davies et al., 2017), for example in the tails of the earthquake magnitude frequency distributions, which are relatively unconstrained by data (too short catalogues), and cause the largest tsunamis, as tsunami impact doesn't tend to saturate for high magnitudes as earthquake impact may do. The rarity of the events in turn also means that in many areas of the world there is necessarily a relatively poor awareness of tsunamis, even in areas where well-documented and very large historical tsunamis happened in relatively recent times (e.g., Lisbon in Portugal, Messina in Italy).

*Tsunamis are multi-source hazardous phenomena*. They can be generated by different sources (e.g., Geist and Lynett, 2014), including earthquakes, landslides, volcanoes, and even meteorological and astronomical events. These sources are usually studied by different and non-overlapping communities, in which different methods have been developed through time. In addition, in the different fields, very different levels of knowledge and different standards exist, also due to the differences in source complexity and tsunami generation mechanism, or to the variable extent to which is possible to constrain the source recurrence. See D3.40 for a discussion of some of these issues, particularly as regards the complexity of tsunami source modeling. From an hazard and risk assessment perspective, instead, we note that, in the seismological community, the probabilistic treatment of sources is pretty well established (since Cornell, 1968); this treatment has been often inherited in PTHA studies from PSHA, as discussed also below. In other fields, like in volcanology, tsunamis are well known to happen, but they are rarely included in volcanic hazard documents (Paris et al. 2014). In these fields, tsunamis have been studied only by analysing specific case studies (e.g., Tinti et al., 1999). In many cases, these difficulties probably led either to focus mainly on the propagation and shoaling of the tsunamis, limiting the source variability to a collection of few scenarios (Tonini et al., 2011), or to focus mainly on a homogenous treatment of sources, limiting the quality of propagation modelling (e.g., Grezio et al., 2012). Still, these are quite remarkable attempts of multi-source tsunami hazard modelling.

*The complexity of the tsunami propagation and impact is high.* Systematic analyses of the source uncertainty and its consequences in tsunami hazard and risk quantification have

been developed by limiting the source variability, that is, by modelling only one of types the potential sources for tsunamis. For example, in seismology, the treatment of sources is well established. This treatment can be naturally extended to tsunamis (e.g., Geist and Parsons 2006). In these cases, tsunamis are seen as purely secondary events of earthquakes, to be propagated from source to target as seismic waves. However, in the seismological community, it is well established the use of empirical laws to model this propagation, the so-called Ground Motion Prediction Equations (e.g., Cotton et al, 2008). On the contrary, tsunamis force to propagate explicitly waves from source to target (e.g., Sørensen et al. 2012; see also Geist, 2009, for a review). This introduced a trade-off between the exploration of the variability of sources and the accuracy in modelling propagation (Lorito et al., 2015). This trade-off lead to a dichotomy between regional studies (with propagation on deep water only and simplified amplification, like Sørensen et al., 2012; Davies et al., 2016) and local studies (with propagation modelled in nested grids, but conversely simplified source variability model, like González et al., 2009). Only recently, specific techniques have been developed to try and solve this dichotomy, based on the selection of "representative scenarios" that statistically approximate the full source variability (Burbidge et al., 2008; Thio et al., 2010; Mitsoudis et al., 2012; Lorito et al., 2015).

*Tsunamis are sensitive to source features*. Some source characteristics (see also D3.40) may have limited or less important effects on "primary" hazards like earthquakes themselves. For example, Ground Motion Prediction Equations (GMPE) usually depend only on few specific characteristic of sources (usually, magnitude, distance and, sometimes, tectonic regime, e.g., Bazzurro and Cornell, 1999; Cotton et al., 2008); even if near-source effects may be important also in this context and there is recent research devoted to that (e.g., STREST Project). Other source parameters are included into the statistical treatment of the empirical models. For example, fault geometry and dimensions are included into the computation of the source-target distance for GMPEs. Such approaches are not satisfactorily y enough in PHTA, where defining individual scenarios is essential for tsunami modelling (e.g., Sørensen et al., 2012), such as the dip and the slip distribution which determine the features of the initial sea level elevation, tsunami energy focussing in the broad-side of the source direction, and the following complex interaction with bathymetry (see e.g. Geist, 2002; Geist, 2009; Geist, 2012). More in general, tsunamis may be sensible to specific characteristics of the sources that may not be relevant in the original hazard (e.g., the seismic hazard), and that instead require a special treatment of uncertainty (Selva et al., 2016).

*The strategies to tsunami risk mitigation are diverse*. Another important peculiarity of tsunamis is that the wave propagation is slow compared, for example, to seismic waves. This makes the early warning strategy very effective, especially in large seas like oceans for distant tsunamis, while near-field tsunami warning presents some issues and must be complemented by a very good preparedness to respond to natural signs of a strong earthquake and tsunami possibly coming (intense and/or very long ground shaking, sea

withdrawal, roars, etc.). In addition to this, tsunami hazard may be mitigated by the design of specific coastal structures, such as coastal walls and vertical evacuation structures; and by a wise land-use planning. Both these possibilities are partly or essentially dependent of the development of good propagation models, as well as, in efficient near-real-time source inversion, like for example implemented by US NOAA PMEL, or for the GiTEWS, or under development for other regions in the world. Nevertheless, a strong attention is needed in this context too to progressively improve the effectiveness of risk-reduction measures.

*Tsunamis have multiple paths to damages*. Tsunamis pose both a direct and an indirect threat to people, that is, may kill both directly when a tsunami hit people and indirectly when a tsunami hits a building that collapses, potentially killing people inside. In addition, tsunamis are usually generated by another dangerous event (e.g., an earthquake) that may have already pre-damaged (e.g., Iervolino et al., 2016) one structure, or simply altered the distribution of people (e.g., Selva, 2013). Also, cascade non-natural phenomena like in the case of the Fukushima nuclear disaster may be triggered which determine an huge worsening of the tsunami consequences. The study of this type of paths to risk has increased the difficulty of risk assessments and often lead to the development of either simplified risk models or the selection of few scenarios, neglecting uncertainties. More recently, several technical solutions to deal with this problem have been developed in the frame of single and multiple-hazard risk assessments (e.g., Selva, 2013; Mignan et al., 2014; Liu et al., 2015; Chock et al., 2016).

### 2.2.2 Hazard and risk quantification methods adopted in ASTARTE

Following the general definition of risk expressed by Equation 2.1, the ASTARTE Consortium has tackled separately the hazard, vulnerability and risk assessment problems, drawing up three distinct documents (D8.8, D8.14 and D8.33). Each of them contains a methodological introductory part, followed by a more extensive description of the application of the methodologies to the ASTARTE test sites. We provide here a short recap of methodological approaches for the hazard and vulnerability quantification, detailed respectively in D8.8 and D8.14. We will refer neither to any specific result nor to any specific partner, keeping the discussion at a methodological level.

***Tsunami Hazard***

Regarding tsunami hazard assessment, two main general approaches were followed:

- Probabilistic Tsunami Hazard Analysis (PTHA)
- (credible worst-case) Scenario-Based approach (SBTHA)

In line with what was already recalled in the previous sections for probabilistic hazard analyses, the main goal of PTHA is to determine the probability of exceedance of a given tsunami metric (flow depth, run-up, current speed, etc.) over a given exposure time for a selected coastal area/site, quantifying both aleatory and epistemic uncertainty. To reach the goal, two different philosophies were adopted in D8.8. One was computationally-based PTHA that target to a full treatment of uncertainty as described in Section 2.1. In ASTARTE, the PTHA analyses have been limited to seismic sources only, and thus they are herein referred to as Seismic PTHA (S-PTHA; from Lorito et al., 2015). S-PTHA is a computationally-based method that takes into account all potential (tectonic) tsunamigenic sources, irrespective of whether it is known that they generated or not tsunamis in the past. The potential tectonic sources are fully characterised in terms of magnitude ranges, geometry, focal parameters and activity rates. For each of the resulting configurations, tsunami generation, propagation and inundation at the selected coastal area/site are simulated numerically.

The other general approach was referred to as Empirical PTHA: it takes as reference the effects of tsunamis recorded in the past, typically maximum run-ups for events occurred in the pre-instrumental era (the largest part in the NEAM region) and tide-gauge and deep ocean sensor records for more recent events. These are treated statistically to retrieve tsunami activity rates and height-frequency relationships.

Typical products of the PTHA are hazard curves, providing the expected maximum value of a given tsunami metric (e.g. tsunami height) as a function of return period, and inundation maps, or maps of maximum water height along a pre-selected offshore isobath, for different return periods. The results are usually provided as aggregate products, but de-aggregation is always possible when the effect of a specific source has to be analysed.

Similarly to computationally-based PTHA, SBTHA heavily relies upon a correct choice of the source areas of interest for the selected target coastal area/site: this is carried out by means of a thorough analysis of catalogues (when available) of historical occurrences of a given tsunamigenic source type, of available databases/studies of potential/known tsunamigenic sources, of the geomorphology and geodynamics of the relevant geographical domain. This initial step (choice of the source areas) can be long and challenging: in addition to earthquakes, in SBTHA also landslides and volcanoes can be taken into account, even if volcanoes have not been tackled in ASTARTE. This enlarges the spectrum of published studies and data to be analysed. Once the source areas have been identified, SBTHA foresees that for each type of tsunamigenic source identified in each area (earthquake, landslide, volcanic eruption) the “credible worst case” is selected. This is the most delicate and critical point of the entire SBTHA approach. In the case of seismogenic faults, defining the “credible worst case” means choosing the maximum value for the magnitude of the earthquakes that a given seismogenic fault might produce. The choice can be based on past earthquakes, on the geometry of the seismogenic area, on its background seismicity and/or strain accumulation rates. Whatever the criterion or combination of criteria adopted, the maximum chosen magnitude allows to establish the geometry of the fault and the slip based on scaling laws. In the “credible worst case” perspective, the focal mechanism (fault strike, dip and rake) is chosen as a trade-off between the constraints posed by local tectonics/geomorphology, and the need to maximize the tsunamigenic potential and the tsunami energy focussing towards the target area. In the case of landslides, the tsunamigenic potential results from a combination of the sliding mass volume, of the position from which the landslide detaches (the depth in the case of submarine landslides), of the expected slide dynamics (translational or rotational), partly on the run-out distance. Geomorphological and geotechnical analyses are of fundamental importance in defining at least some of the above factors, especially the maximum volume of mass that can be mobilised along a selected slope. Unfortunately, it is not rare that this type of information is insufficient or even missing, so that pure speculation must be used. For each of the selected sources and related credible-worst-case scenarios, tsunami generation, propagation and inundation at the selected coastal area/site are simulated numerically, usually on nested grid configurations allowing fine resolutions (down to few tens of meters of even few meters) in specific target areas.

Typical products of the SBTHA are aggregated fields of maximum water height (offshore and inland), flow depth onshore, particle velocities (offshore and onshore), maximum momentum flux (onshore).

The previous short excursus on the two main approaches for the hazard quantification adopted in ASTARTE (as emerged from D8.8) highlights some important points of interest in light of the scope of the present deliverable.

One of the most evident differences emerging from the ASTARTE applications is that SBTHA more easily deals with different tsunamigenic sources: for example by analysing jointly and comparing the effects of several tsunami scenarios generated by both earthquakes and landslides. Empirical PTHA intrinsically deals with all sources, since local data are not filtered based on the source. Conversely, computationally-based PTHA has been applied here only to earthquake sources, and thus named S-PTHA. This limits for now the applicability of S-PTHA to those test areas where the main threat in terms of tsunami is related to tectonic sources. However, some first cases of applications of computationally-based PTHA to non-seismic sources are being developed (see e.g. discussion in Geist and Lynett, 2014), despite of some challenges, including the difficulty in constraining the source recurrence (see e.g. discussion in D3.40).

Limiting to seismic sources only, it is evident that the computational effort involved by the two approaches is completely different. In S-PTHA, all potential seismogenic sources must be included and weighted according to their relative probability of occurrence. In principle, for each source a multi-dimensional space of parameters should be explored, resulting from the combination of the discrete values that each relevant geometrical and focal property of the fault can assume. The often prohibitive number of simulations to be run makes it feasible only to run linear simulations over coarse or medium-resolution grids, limiting the computation of the relevant tsunami metrics along a selected offshore isobath. If inundation mapping on fine-resolution grids is needed, then a filtering procedure must be adopted at the source level, an example of which is described in D8.8. Even in this case, however, the number of simulations to be computed is very much larger than that involved in SBTHA, where at most 10-15 high-resolution models are produced.

Another difference regards the “return period” concept, which is a key feature of PTHA methodology and results, while it is not apparent in the SBTHA approach. This concept is intrinsically related to the possibility of quantifying the uncertainty of hazard quantification and, as discussed in section 2.4, represents the basic concept to make hazard quantification comparable and scientifically testable. Here, we stress that in the “credible-worst-case” perspective the return period is not absent, and even if in some cases it is not explicitly addressed, it can be thought to be on the order of the known or hypothesised recurrence time of the largest-magnitude earthquakes that can be generated by a given fault, hence it is dependent on the specific case and it would be recommendable to discuss it case by case when performing SBTHA.

Finally, as explained more in detail in the next sections, while the sources of uncertainty are very similar in PTHA and SBTHA, the way they are treated is totally different in the two approaches. Limiting to the results developed in ASTARTE, uncertainties were taken into account only in PTHA.

**Tsunami Vulnerability & Risk**

The first part of Deliverable D8.14 provides a complete review of the main topics related to tsunami vulnerability definition and assessment. We briefly summarise here what we think are the key concepts used in ASTARTE.

First of all, it is recognized within the tsunami community that a clear definition of vulnerability related to tsunamis need to be established: in ASTARTE, we define tsunami vulnerability everything that regards the characteristics and circumstances of a community, system or asset that make it susceptible to the damaging effects of tsunami hazard. Exposure is another concept that is tightly linked to that of vulnerability, consisting in the presence of people, property, systems or other elements in tsunami hazard zones and hence subject to potential losses. Finally, we define tsunami damage the loss or harm caused by a destructive tsunami.

Like for many other geo-hazards, vulnerability related to tsunamis can be treated through qualitative, semi-quantitative and quantitative approaches. Given a set of elements whose vulnerability is to be assessed (e.g. buildings or people), a qualitative approach assigns attributes to the elements, scores to each element attribute on the basis of some (often subjective) criteria and groups them in classes of vulnerability, based on suitable combination of the scores. In a quantitative approach, as discussed in **Section 2.1.2**, the vulnerability is defined through curves that relate physical indicators (like fragility/damage for buildings) or social indicators (like mortality in the case of people) to relevant tsunami metrics, such as the flow depth and/or the flow velocity. It is worth noting that uncertainties are quantified only in the case of quantitative vulnerability quantifications.

The approaches used in ASTARTE D8.14 focussed almost exclusively on buildings and involved both a qualitative (PTVA-3) and a semi-quantitative (SCHEMA) approach. For comparative reasons, in the present document (in **Section 2.3.3**), a quantitative method will also be introduced based on fragility curves, i.e., probabilistic vulnerability models that describe the conditional probability that a damage state will be reached or exceeded for a given hazard level (see **Section 2.1.2** for more details).

Full details on the PTVA-3 and the SCHEMA approaches can be found in D8.14, together with the adaptations made for the different test sites. Here we summarize only the relevant characteristics.

The building classification scheme proposed by the SCHEMA (SCenarios for tsunami Hazard-induced Emergencies Management) project was set up according to intrinsic properties of

the buildings, construction material and structure, including foundation. It was built upon the analysis of the damage to constructions caused in Banda Aceh, northern Sumatra, by the 2004 Indian Ocean tsunami (Leone et al., 2010), with modifications to adapt it to European building standards (Valencia et al., 2011). The tsunami hazard metric (in **Section 2.1**, called hazard intensity measure) chosen as the one influencing most the damage is the flow depth. Based on the vulnerability classification and the experimental data on tsunami flow depth, for each building class a damage function was built, providing the maximum level of damage that a building belonging to that class may incur corresponding to a given maximum flow depth. The resulting damage discretization in classes leads to a discretization also of the damage functions: the result is a damage matrix, with rows and columns coinciding respectively with the damage and vulnerability classes and with the elements being the discrete intervals of the flow depth. The damage matrix can be adapted to a certain extent to cope with the specific characteristic at any given test-site.

PTVA is acronym standing for "Papathoma Tsunami Vulnerability Assessment": it was first devised by the Greek researcher Papathoma who applied it to the Cretan town of Heraklion (Papathoma et al. 2003). PTVA-3 is "version 3" of PTVA and has been introduced by Dall'Osso et al. (2009). In PTVA-3 the main vulnerability index is the "Relative Vulnerability Index" (RVI), which is defined as follows. The first step consists in defining the index of vulnerability of a building (Bv): this is an integer ranging from 1 to 5. In order to compute it, PTVA-3 introduces first an intermediate index, Bv', that is computed by taking into account seven structural parameters (attributes), each of which is graded by a score selected by the evaluator among a list of possibilities. The index Bv' is the weighted average of these scores with weights that are numerical constants calibrated in the field (all details in Dall'Osso et al. 2009). By construction, the index Bv' can take values within the interval [-1, +1] and is linearly mapped and discretized in the integer index Bv with five levels from 1 to 5. The second step requires the calculation of three more variables related to the water column (Ex and WV) and to the level of protection (Prot). The parameter Ex depends directly from the water column; WV is the ratio between the number of flooded stories and the total number of stories; Prot expresses the level of protection against the tsunami according to the type of obstacles that can mitigate the effects of the tsunami itself (houses, stone walls, seawalls, trees or other natural protections). "Natively", the different parameters may attain values in different intervals: anyway, in the PTVA-3 procedure they are all rescaled to a [+1, +5] interval. Finally, RVI is computed as:

$$RVI=2/3\cdot SV+1/3\cdot WV \tag{2.6}$$

where

$$SV = (BV) \times (Ex) \times (Prot) \tag{2.7}$$

SV can vary by definition between 1 and 125, but is then rescaled to the interval [+1, +5]: it is this scaled value that enters the definition of RVI. The same applies to WV. As a result, RVI

ranges from 1 to 5 but does not necessarily attain an integer value. As a consequence, this interval has been further divided into 5 equal sub-intervals of 0.8 extension, corresponding to the relative level of the expected damage of the building: minor, moderate, average, high and very high.

It is worth stressing that there are still several problems that make the generalized application of such approaches difficult. The most important issue is that the two approaches are not computational. In particular, the SCHEMA approach is empirical: the original fragility and damage curves proposed by the methods are based on empirical data from Banda Aceh after the 2004 Indian Ocean tsunami, and not analytical and based on direct modelling of the target structures (as briefly discussed in Section 2.1.2). This implies that the structures used to build the curves are typically dissimilar to those found in other places, like the ASTARTE test sites, requiring an adaptation tailored to each site. This is one problem that would exist also in the development of quantitative tsunami vulnerability and risk based on fragility (quantitative method), since for tsunamis only very few examples of analytical models exist. This may potentially introduce substantial epistemic uncertainty that can be quantified by adopting alternative models (see discussion in Section 2.3.3).

On the other hand, PTVA-3 is a qualitative approach providing a vulnerability classification and damage assessment in absence of fragility curves. It is a quite flexible approach in which, anyway, the Relative Vulnerability Index defines only a comparative ranking of damage among buildings but cannot define an absolute state of damage.

## 2.3 Comparing the results of different levels of simplifications in treating uncertainty: examples from ASTARTE Test sites

In this section, we present some examples from the hazard and risk quantifications in some of the test sites of ASTARTE, with focus in the comparison between different levels of simplifications used in these assessments.

### 2.3.1 Güllük Bay: Probabilistic VS Deterministic tsunami hazard

Deterministic and probabilistic tsunami hazard analyses are performed for Güllük bay.

Accurate bathymetry/topography data and rupture parameters of tsunami source are the required inputs for convenient numerical modelling in order to perform tsunami hazard analysis. Bathymetry/topography data are obtained from GEBCO© with 900m resolution in the sea, and from ASTER satellite data with 30m resolution at land. In order to improve the quality of the data in the shallow zone and data of shoreline additional digitizations from navigational charts are performed. Gridded data sets are developed from the available bathymetry/topography data and also digitized data for both nested and single domain simulations with different grid sizes.

In the deterministic approach, probable worst case seismic tsunami sources (s18-z22) with their estimated rupture parameters and non-seismic tsunami sources (caldera motion of Santorini and Columbus volcanos) with their dimensional parameters are determined and simulated.

In the probabilistic approach, the data based on seismic monitoring between 1950-2014 years is used and the earthquake magnitudes which may occur in 100, 500, 1000 years return periods are computed statistically by extreme value analysis. The rupture parameters are determined with the help of measured rupture parameters from past earthquakes in the region (Kalafat, 2009).

The tsunami numerical tool NAMI DANCE is used as the computational tool in numerical modelling. Simulations are performed for each selected tsunami scenario according to 240min simulation duration (for seismic cases) and 300min simulation duration (for non-seismic cases). Simulation durations are determined according to distance between tsunami source and the study region to properly compute and visualize generation, propagation and coastal amplification of tsunami. In the deterministic approach, nested domain gridded data are used (Domain B with 270m grid size, Domain C with 90m grid size, Domain D with 30m grid size). Single domain gridded data are used (Domain B with 30m grid size) in the probabilistic approach. Nested and single domain simulation results are very close to each other, hence the results can be compared to each other confidently.

Some of the main critical regions under tsunami attack in the Güllük bay are selected as numerical gauge points. Among those "Didim Tekağaç" is selected as one of the numerical gauge location because of the findings of the traces of historical tsunamis (Santorini eruption approximately in 1600 BC) from the excavations (Papadopoulos et al., 2012) in the region. Other vulnerable locations in the region are "Fish Farms". Those are selected due to their vulnerability not only against storm waves but also long waves. Another numerical gauge point is selected at Gulluk village called "Güllük" where the Milas Bodrum Airport is located on the low land area at the end of Gulluk Bay. The commercial port near Güllük village is another critical structure in Gulluk Bay. "Yalıkavak" is selected as another numerical gauge point to monitor the probable coastal amplifications around Yalıkavak marina, nearby touristic coastal utilities and residential areas. The time histories of water level fluctuations and computed near shore tsunami parameters for those numerical gauge points are obtained to provide data for the tsunami hazard analysis for Güllük bay.

Among the scenarios in probabilistic analysis, the tsunami due to the 1000 years mean return period seismic source (s18-z22) is found to cause higher tsunami parameter values in Güllük bay in comparison to other sources (Acar, 2015). In addition to this source, probable effects of seismic source (s18-z22) with 500 years return period should be considered as important tsunami sources for probabilistic tsunami hazard analysis. Maximum wave amplitudes are computed as 3.7m for “Didim Tekağaç”, 3.57m for “Fish Farms”, 4.51m for “Güllük” and 3.57m for “Yalıkavak”. In addition, maximum current velocities are computed as 2.18 m/sec for “Didim Tekağaç”, 1.55 m/sec for “Fish Farms”, 0.47 m/sec for “Güllük” and 0.33 m/sec for “Yalıkavak”. Current velocity near fish farms is greater than 1 m/sec (1.17 m/sec according to 500 years return period simulation, 1.55 m/sec according to 1000 years return period simulation).

Probabilistic analysis gives important and reliable results comparing to deterministic seismic source (s18-z22) results. Extreme value analysis gives earthquake magnitudes on the same order of that selected for the deterministic approach. Rupture parameters (dimensions of the fault, surface displacement) are estimated by the help of available and valid empirical equations. In the light of probabilistic analysis results, deterministic seismic source (s18-z22) has the probability to occur with less than a 500 years return period.

It does not seem that devastating tsunami effects and resulting damages may not be foreseeable in the near future in Güllük bay according to the resolution of the data used for hazard analysis here. However, marinas, commercial port, aquaculture plants (fish farms), airport, nearby touristic coastal utilities and residential areas located in Güllük bay are still critical locations even if a small size tsunami enters Güllük bay. These critical regions would be more extended under the tsunami with 500 and 1000 years return periods. Inundation zones are formed along Güllük coasts especially in low elevated coastal areas.

### 2.3.2 Sines: Probabilistic VS Deterministic tsunami hazard

Tsunami hazard assessment is conducted for Sines test-site using both deterministic (DTHA) and probabilistic (PTHA) approaches. Only tsunamis of tectonic origin were considered in both analyses. The tsunami hazard was assessed through deriving high-resolution inundation maps. Inundation maps depict the spatial distribution of the maximum flow depths for DTHA, while for PTHA, they show probabilities of exceedance considering different exposure times.

Bathymetric/topographic dataset was built to serve accurate assessment of tsunami hazard in both deterministic and probabilistic approaches. Another common tool that was employed in both analyses is the non-linear shallow water equations code that allowed simulating the tsunami from the source up to the coast, including inundation. The tsunamigenic source models, on the other hand, were defined in different ways. In the DTHA, particular source scenarios (maximum credible) were considered. In the PTHA, where the contribution of small and large events is included, a broad range of near- and far-field potential sources was used.

A set of bathymetric/topographic data grids is generated to cover the areas from the tsunami source area to the studied test-site. The grids are nested for consecutive calculations of tsunami generation, propagation and inland inundation. For the near-field source zone (SW Iberia Margin), four nested grid layers of increasing spatial resolutions (640 m, 160 m, 40 m, and 10 m) have been employed. The far-field tsunami simulations (Gloria and Caribbean source zones) were conducted using a system of five nested grids (resolutions from 2560 m to 10 m). Bathymetric Chart of the Oceans 30 arc-sec data (GEBCO), together with the SW Iberian margin bathymetry compilation (Zitellini et al., 2009) were used to build the coarse grids. To generate the digital elevation model for the Sines test-site, different types of data sets were used: LIDAR data (2 m resolution), completed by both bathymetric model of the Hydrographical Institute (100 m resolution) and data from the nautical charts of the zone. The compiled digital elevation model of Sines was validated through field survey GPS-RTK data.

In both DTHA and PTHA, the numerical code NSWING (Nonlinear Shallow water WIth Nested Grids) was used to simulate the tsunami propagation and inundation. Tsunami simulations were conducted to quantify the tsunami metrics (wave height, flow depth, velocity etc) that were used to assess the tsunami hazard.

Tsunamigenic sources were defined in different ways for DTHA and PTHA. In the deterministic approach, a set of maximum credible earthquake scenarios were considered. These include a limited number of events: four source models in the SWIM zone of magnitudes ranging from 8.25 to 8.75 and one 8.3 magnitude scenario in Gloria fault. Unlike the scenario-based approach, the probabilistic analysis for Sines test-site involved the use of a broad range of near- and far-field potential sources (Mw7.5 up to Mw9.0) with their

annual recurrence rates. Three main source zones were distinguished (SWIM, Gloria and Caribbean) and for each one of them the maximum possible earthquake magnitudes were assigned considering instrumental and historical catalogues in addition to apply a seismic probability model allowing to derive the recurrence rates of individual earthquake scenarios. Also, aleatory uncertainty into the source location was treated by letting the earthquake fault, representing the tsunami event, float with a regular interval along the fault trace allowing to cover possible rupture zones.

Another source of uncertainty that was considered in the PTHA at Sines test-site regards the tidal stage. This is important for Sines test-site, located in the NE Atlantic coast, because the tide variation is significant and the peak-to-peak tidal amplitude can be as high as 4–5 m in this region. To incorporate the tide uncertainty in the PTHA the following steps were adopted (further details can be found in Omira et al. 2016): i) Run the numerical code for a number of 'static' tide stages (i.e. MLLW, MSL, and MHHW), and compute η (hazard case: wave height/flow depth) at each point of the simulation grid; ii) Construct the curve of evolution of η (vertical axis) in function of the tidal stage (horizontal axis) for each grid point using a piecewise linear interpolation; iii) Compute, from long time-series recorded by the tide gauge (or predicted if records are not available) the tide probability density function (PDF), which expresses the cumulative probability (vertical axis) of exceedance of tide levels (vertical axis); iv) Selecting the exceedance threshold and then a horizontal cut of the curve of evolution of η in function of the tidal stage gives the tidal stage (Ts) that leads to exceed the threshold; v) Using this Ts value, a vertical cut of the tide PDF curve gives the cumulative probability of exceeding the Ts.

Hazard curves, as a key result of any PTHA, were computed at each point of the grid and presented for some PoIs off- and on-shore. A vertical cut of these hazard curves, considering a selected exceedance threshold, led to derive the probability exceedance maps for Sines test-site considering exposure times of 100, 500, and 1000 years.

Considering the contribution of large and small tsunami sources in PTHA led to derive probabilistic maps that provide a more complete presentation of tsunami hazard than the DTHA. The fact that there is no single accepted way of determining the worst-case scenario in DTHA constitutes the main limitation of the scenario-based approach in comparison with the probabilistic analysis. This is clearly highlighted in the obtained tsunami hazard results for the Sines test-site showing that the calculated inundation area using the probabilistic approach (~4 km2) is much larger than those simulated for four among the five considered scenarios in the DTHA. Moreover, probabilistic tsunami hazard results for Sines test-site show that most coastal zones, including the critical installations, are prone to tsunami inundation with probabilities reaching 80% for 500-year exposure time. On the other hand, the same zones were assumed as "safe" when considering some maximum tsunamigenic scenarios impacting the studied test-site.

### 2.3.3 Siracusa: Probabilistic VS Deterministic tsunami risk

In previous WP8 deliverables, and particularly in D8.8 and D8.14, the name "Siracusa" was used to describe a wide geographic area ranging from the town of Augusta to the north down to Siracusa and its inner gulf to the south. Here the focus will be only on the town of Siracusa: the reader interested in its history, including the accounts and evidences of the impact of past tsunamis, and in the results of the tsunami hazard assessment is referred to deliverable D8.8. Here we extract some of the results presented in D8.14 concerning the vulnerability classification and damage assessment performed by applying the PTVA-3 and SCHEMA-project methodologies to selected zones of Siracusa. These results, expressed in terms of number of expected damaged buildings for a "bathtub scenario" of 5 m, are compared to those computed for the same scenario by means of a purely probabilistic method based on fragility curves. The outcomes of the comparison are discussed.

A set of 1:2000 scale numerical maps provided by the "Servizi Informatici Territoriali e Cartografia, Nodo Regionale S.I.T.R." of the "Regione Siciliana, Assessorato Territorio e Ambiente, Dipartimento Urbanistica, Area 2 Interdipartimentale" allowed to extract the buildings interested by a scenario "bathtub" inundation level of 5 m: the resulting number of extracted buildings is 2446.

***SCHEMA and PTVA-3 approaches***

As described in D8.14 and in Pagnoni and Tinti (2016), the SCHEMA classification adapted to Siracusa involves building vulnerability classes from A (light constructions) to E (reinforced concrete buildings). The damage to buildings is discretized in 5 levels going from D1 (light damage) to D5 (total collapse). The damage functions themselves are discretized in a different way with the respect to the original SCHEMA formulation, especially as regards vulnerability classes A and B (Pagnoni and Tinti, 2016). In light of the comparison that is presented in this document, we may further classify the building typologies in "Brick" buildings (covering SCHEMA classes A to C) and "Reinforced Concrete" (RC) buildings (covering SCHEMA classes D and E). The resulting vulnerability classification for the 2446 buildings in Siracusa counts 92% of "Brick" buildings (the largest part in SCHEMA class C – individual buildings, villas) and only 8% in "RC". For the considered 5-m inundation scenario, only about 10% of the buildings would suffer a structural damage leading to partial or total collapse (damage classes D4 and D5). These larger damages are expected in correspondence with the harbour areas, the commercial/industrial areas and the shantytown (see D8.14).

The application of PTVA-3 to Siracusa results in different aspects. It was evident that the definition of at least two of the attributes that enter the definition of BV ("foundation strength" and "shape and orientation") are very difficult to define and risk to depend too much on the subjectiveness of the evaluator. This said, it resulted that no buildings fall into class 1, which is the most resistant class. Moreover, 91% of the buildings fall into the least

resistant classes (4 and 5), 75% being in class 4: these buildings are found mainly in the commercial and industrial area and in the shantytown in the north-western part of the bay.

Regarding the damage level, where about 21% of the damage is classified from HIGH to VERY HIGH (classes 4 and 5), and only 9% as VERY HIGH.

The results obtained through SCHEMA and PTVA-3 for Siracusa were compared both at the level of vulnerability classes and of SCHEMA damage level vs PTVA-3 RVI. The details are given in D8.14. We think it is worth to recall here that, notwithstanding the implicit assumption that the two damage scales are comparable (but they differ since RVI is a relative damage scale, while the SCHEMA scale is an absolute one), we obtained that in 45% of cases the method PTVA-3 results in a damage level larger than SCHEMA and in only 3% it gives a smaller damage (66 buildings). In other words, SCHEMA tends to underestimate the damage level with respect to PTVA-3.

***Probabilistic approach based on Fragility functions***

For the probabilistic method, we consider the fragility functions adopted from Suppasri et al. (2013). These fragility curves were obtained by fitting the survey data from 2011 the Great East Japan tsunami for different building typologies. Suppasri et al. (2013) define 6 damage states:

1. Minor damage: there is no significant structural or non-structural damage, possibly only minor flooding. Possible to be used immediately after minor floor and wall clean up.
2. Moderate damage: Slight damages to non-structural components. Possible to be used after moderate reparation
3. Major damage: Heavy damages to some walls but no damages in columns. Possible to be use after major reparations
4. Complete damage: Heavy damages to several walls and some columns. Possible to be used after a complete reparation and retrofitting
5. Collapsed: Destructive damage to walls (more than half of wall density) and several columns (bent or destroyed). Loss of functionality (system collapse). Non-repairable or great cost for retrofitting
6. Washed away: Washed away, only foundation remained, total overturned. Non-repairable, requires total reconstruction.

Here, we consider 2 building typologies: Brick and RC buildings. For these typologies of buildings, the fragility curves are assumed log-normal. The fragility curves are reported in *Figure 1*. As expected, masonry buildings are more fragile than RC buildings. For example, for a flow depth of 10 m, the probability of being washed away is around 0.80 for Brick buildings and 0.20 for RC buildings.

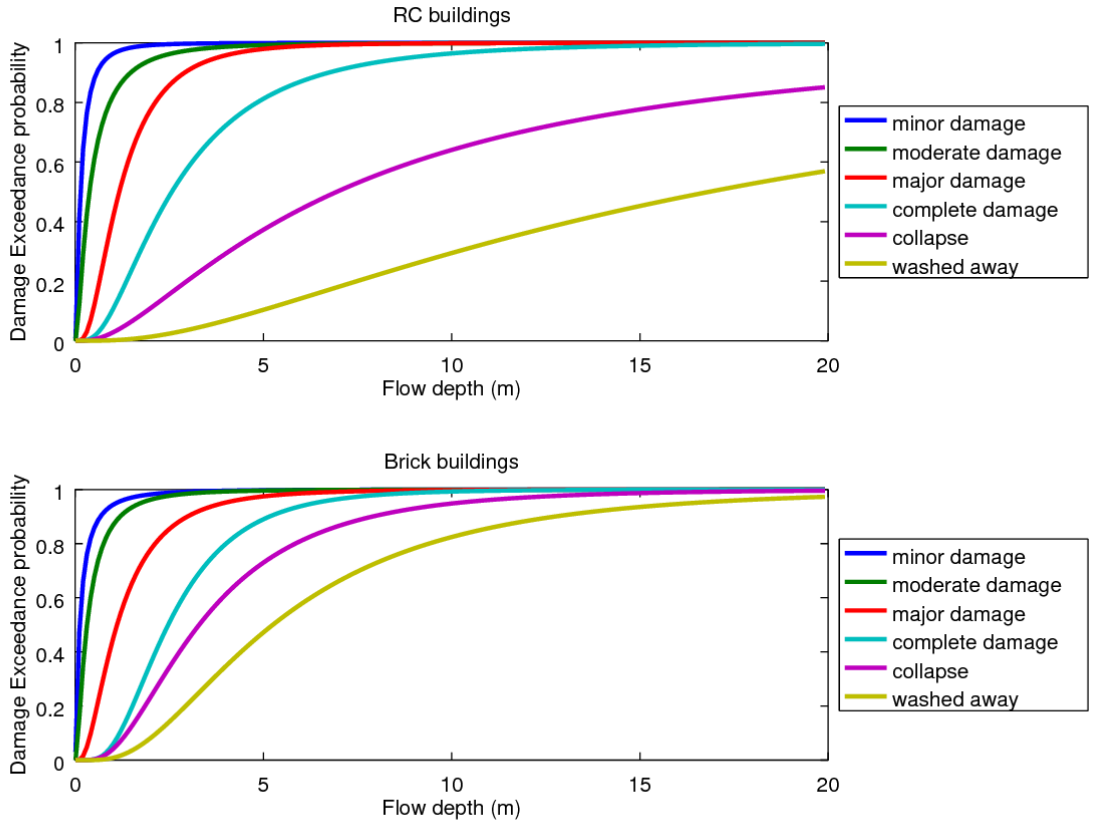



***Figure 1**: Fragility curves for RC (above) and Brick (below) buildings, for all damage states (from Suppasri et al. 2013)*

To propagate the (aleatory) uncertainty quantified by the fragility curve, we adopt a Monte Carlo method. In particular, we first sample 1000 times the damage state for each building, by comparing one randomly generated number between 0 and 1 and the probability of exceeding each damage state for the intensity at the place of the building given by the selected scenario. Then, we count the number of buildings with the same damage state in each sampled configuration of damages. We report the results in the histograms of ***Figure 2***, for all the buildings and for the two classes of buildings separately. The effect of propagating the aleatory uncertainty is that we have a distribution of the number of buildings in each damage state, and not a single number. In this way, it is explicitly shown the range of variability that one should expect, given one pre-selected event. It is worth to note that, even considering this uncertainty, the results are quite informative. For example, the most represented damage state seems to be 'collapse', with a number of buildings ranging between 300 and 350. The less represented damage state is instead 'washed away', with a number of collapses, with a number of buildings ranging between 225 and 275. Instead, if we look at RC buildings only, the number of collapsed and washed away buildings is significantly lower than all the other damage states.

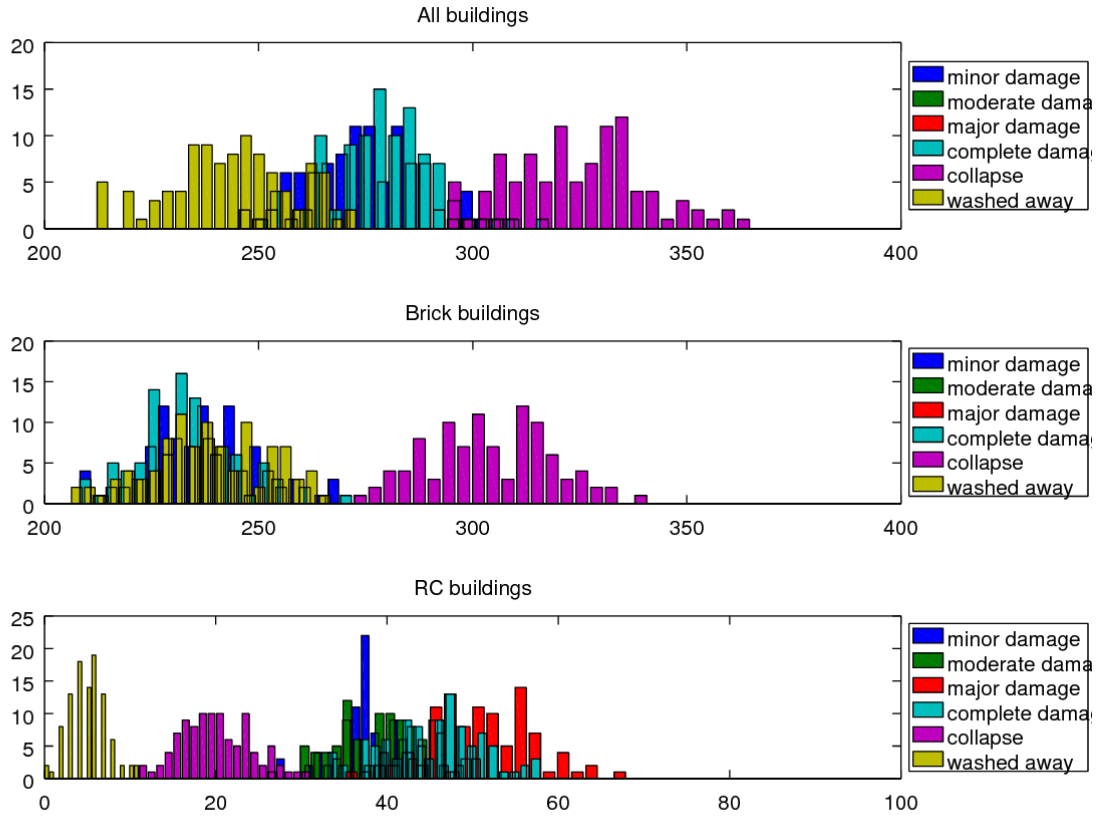



***Figure 2**: Number of buildings in each damage state, for all buildings (top), Masonry buildings (middle) and RC buildings (bottom), adopting the probabilistic method based on Suppasri et al. (2013).*

***Comparison***

To compare the results of the probabilistic method with the ones of SCHEMA and PTVA-3, we consider the fragility curve corresponding to collapse and indicating the probability of exceeding collapse, that is, collapse and washed away. By adopting the same MC procedure described above, and summing the number of buildings with damages equal or larger than collapse, we obtain a sample of 1000 total number of heavily damaged buildings. These results are reported in ***Figure 3***, and compared with the equivalent SCHEMA and PTVA-3 results. These results clearly show that probabilistic methods are much more informative and they allow for make judgement that would have been impossible in the case of more qualitative methods. Indeed, the plots of ***Figure 3*** clearly show that SCHEMA and PTVA-3 results are incompatible with the probabilistic results based on Suppasri et al. (2013) results. This conclusion is possible only because the probabilistic methods quantify the full probabilistic distribution. In this case, in ***Figure 3*** we can note that both SCHEMA and PTVA-3 results are well outside the uncertainty bounds of the probabilistic method. In particular, they foresee much less heavily damaged buildings. This may be a symptom of a potentially severe underestimation of the (epistemic) uncertainty in the assessments.

It is fair to argue that this is the result of two main factors that make the comparison illustrated here only indicative. The first factor, which applies to the SCHEMA results, is that SCHEMA is based on a vulnerability analysis (Valencia et al. 2011) relying on data and statistical processing that are significantly different (and alternative to) those of Suppasri et al. (2013). This is confirmed by the analysis presented in Tarbotton et al. (2015): the plot they show in their Figure 13b reveals the huge difference in the fragility curves by Suppasri et al. (2013) and Valencia et al. (2011), clearly explaining why the SCHEMA results are so largely underestimated even with respect to the lower bound of the probabilistic distribution presented here. Secondly, we proposed a comparison based on a classification in "Brick" and "Reinforced concrete" buildings which is very hard to carry out for the town of Siracusa, especially when compared with the Japanese situation. Finally, the comparison was made by assuming that the damage levels in SCHEMA, the RVI of PTVA-3 and the damage classes of Suppasri are somehow superimposable, an assumption which deserves at least a careful validation.

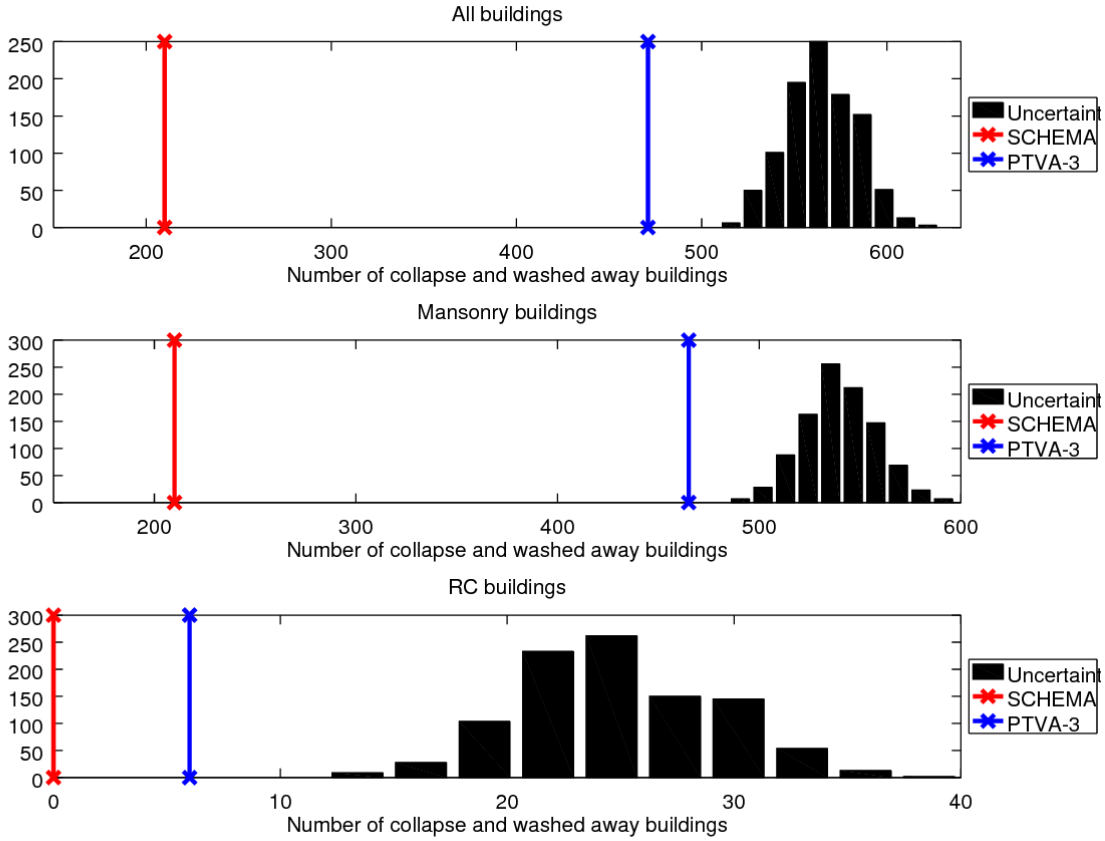



***Figure 3**: Number of buildings with damage states greater or equal to collapse (D5 and D6 of Suppasri et al., 2013) for all buildings (top), Masonry buildings (middle) and RC buildings (bottom), adopting the probabilistic method based on Suppasri et al. (2013) fragility curves (histogram) and the qualitative methods SCHEMA and PVHA-3.*

## 2.4 Comparison between scenario-based and probabilistic tsunami hazard and risk assessment: guidelines for future tsunami hazard and risk assessments

In general, we can state that probabilistic and scenario-based assessments differ in some respects and share some common features, at least in the formulation presented in ASTARTE (Section 2.2.2 and reference therein). In light of the scope of this document, we are not going to discuss whether one approach is better than the other in general. Indeed, as discussed below, they may answer to different questions that may be of interest for stakeholders and local planners. Here, we focus on comparing probabilistic and scenario-based approaches in terms of their capability to account for and quantify uncertainty, to compare to each-other, and to be tested against real data.

As shown in ***Section 2.3***, when available, probabilistic hazard, vulnerability, and risk quantifications are more informative than scenario-based and/or qualitative assessments. In particular, they allow for quantifying the uncertainty and, in this respect, for understanding the effective meaning and strength of scenario-based results. For example, in Sections 2.3.1 and 2.3.2, only the probabilistic tsunami hazard results allow quantifying the representativeness of the scenario-based hazard, by quantifying its mean return period. It is also shown that only the probabilistic results allow for the comparison among different results. For example, in Section 2.3.3, it is shown that the different models adopted for risk quantification are incompatible to each other, meaning that a significant epistemic uncertainty exists over these risk results. This means that either only one of these models is assumed true and the other false, or all models may be scientifically acceptable. In most of cases, given that tsunamis are rare, the paucity of data prevents to falsify most of the models (e.g., Marzocchi and Jordan, 2014), and robust results can be obtained only by quantifying the variability of among scientifically acceptable models in the form of community distribution (e.g., SSHAC, 1997).

Consequently, hazard and risk quantifications that account for uncertainty (both aleatory and epistemic) not only provide more solid results on which to base decision making, but also allow for a deeper understanding of their meaning and their potential limitations. Of course, this derives from the fact that all the probabilistic methods are built by the computation of a large number of scenarios, weighted by their probability of occurrence in the exposure time, and combined by considering alternative and scientifically acceptable models. This complexity includes more information that, if well communicated, allows increasing awareness for decision making. However, the communication is complicated by the complexity of the results, increasing the need of outreach efforts.

The increasing complexity of hazard and risk quantifications that account for all the potential sources of uncertainty is also the main limit of probabilistic analyses. Indeed, they are difficult to be implemented. Probably for this reason, in ***Section 2.3***, we could report only 3 case studies. More in general, looking at all ASTARTE case studies, we can note that

probabilistic methods are essentially limited only to hazard (and not extended to risk) and to seismic sources. In other words, most of the hazard and risk computations made in ASTARTE test sites are based on scenario-based assessments. The main reason for this is that fully probabilistic hazard assessments are more complicated and recent in literature. However, we must note that scenario-based assessments may be selected also because i) the state of knowledge about some of the possible tsunami sources is not sufficient for full probabilistic assessments (e.g., submarine landslides), and ii) the computational power required by fully probabilistic methods is actually difficult to be reached.

The relative simplicity of the analysis of single scenarios allows focusing on the analysis of some details that should be neglected otherwise. For example, since there is not the need of simplifying the propagation, it may provide very important clues for design purposes, as well as for investigating physical processes. Also, they provide a base for investigating cascade effects, or multiple hazards caused by the same event. All these types of studies can be described as “what if” analyses. The use of scenarios for hazard and risk quantification is instead more critical, since it enables producing results at much lower costs, but it is entirely based on the “selection” of the scenario(s). This selection and the ground to make it is both theoretically and practically complicated.

Some authors based the selection of this scenario on the search for the maximum credible (MC) event, often claiming the all people and structures should be protected against this maximum, no matter how infrequent it may be (e.g., Wyss et al., 2012; Peresan and Panza, 2012). However, the definition of this maximum is in practice very challenging, and it has been challenged in literature for its very physical ground. For example, Allen (1995; and many after this paper) described MCE as the event “which is only a shade smaller than the minimum incredible earthquake”. This tongue-in-cheek highlights the basic difficulty in rigorously defining the MCE concept, that is, define what it is actually possible and impossible (even if absolutely rare). Events like the Boxing day earthquake in Indonesia, Tohoku in Japan, or even Christchurch in New Zealand, clearly demonstrated that things apparently incredible may indeed happen, causing destructive events.

Another possible interpretation of MC event may be based on the definition of a time window (Allen, 1995). In a row, this time window may be associated to a “return period” of classical probabilistic hazard assessments (e.g., Hanks and Cornell, 1994). However, sometimes, this reference time window (or mean return period) is not defined explicitly. This is very dangerous and it poses serious ethical issues (e.g. Marzocchi, 2013), since the selection of the scenario implicitly attributes the role of defining the acceptable risk to the scientists, while this choice has not a scientific ground (e.g., Jordan et al., 2011; Marzocchi et al., 2012; Geller et al., 2013). If instead this selection is made explicitly, it may indeed provide a physical ground for the selection of the scenarios (e.g., Løvholt et al., 2012). The main difficulty is the fact that it is sometimes difficult to do this without a full probabilistic analysis, since it may be possible to define such scenarios depending on their sources (e.g.,

their magnitude), but it is more complicated to do it depending on their impact (e.g., run-up at site). Indeed, the complexity of the tsunami energy path from the source to the site and the strongly non linear character of the tsunami inundation process cause a strong heterogeneity of the tsunami impact. One possible solution is recurring to simplifying assumptions for exploring the dependence of impact variability on the source variability, that may be then challenged *ex post* (e.g., following a method similar to Lorito et al., 2015). Assuming that it is possible to define an objective ground to make this selection *ex ante* (before producing a fully probabilistic assessment), this definition may help producing simplified tsunami hazard and risk assessments that may result useful whenever the cost of potential risk mitigation actions is so low that do not require highly detailed analyses (see discussion in Marzocchi et al., 2012).

The selection of the scenarios based on a reference mean return period may provide also the ground to compare probabilistic and scenario-based results to each other and to test them against the available data. One example of this comparison is made in **Figure 4**. The probabilistic results are reported in terms of a family of hazard curves reporting the mean annual frequency of exceedance of different values of hazard intensity in one defined location (**Figure 4A**). Epistemic uncertainty is reported in grey as a confidence interval around a best guess value.

A scenario-based hazard quantification consists of one point in the same graph for the location, reporting hazard intensity value and the reference mean return period (blue dot in **Figure 4B1** and **4B2**). The intercept between the hazard curves and the scenario hazard intensity defines an interval of annual frequency of exceedance for that intensity value (black dashed box in **Figure 4B1** and **4B2**). This range should be compared to the reference mean return period of the scenario-based analysis (blue dot) and, in particular, this reference values should in theory be within the range, as in **Figure 4B2**. If the reference value is outside the range, as it is in the example of **Figure 4B1**, the two results are incompatible and should be rechecked. In this specific case, it seems that either the probability of occurrence estimated for that scenario is too large in the probabilistic analysis, or the scenario selected is too small (in terms of impact) with respect to the target return period. Since these results may have been produced independently by two different groups without any interaction and by possibly different models of the earthquake (scaling laws, slip distribution, numerical modelling strategy), of the tsunami generation and propagation (modelling assumptions such as hydrostaticity, the adopted numerical scheme), this may be a symptom that epistemic uncertainty have not been fully addressed in the probabilistic analysis, which then is not presently fully representing the community distribution, or just that “sanity-checks” of both models are necessary. This type of problems is not new at all. The main motivation for establishing SSHAC and the SSHAC process was the existence of incompatible hazard estimates provided by different yet highly respected groups, which had put the US Nuclear Regulatory Committee in a stall situation in view of their regulatory concerns. In the next sections, we’ll briefly touch on these issues, by

presenting a feasible scheme for the decision-making process applicable to tsunami hazard and risk analysis; this process was initially discussed and applied within the STREST project, and now is being applied for the TSUMAPS-NEAM PTHA.

With a similar logic, the results can be tested against real data. By counting the number of known events which exceed a given threshold, the observational record can be translated into observed frequency of exceedance for any given hazard intensity threshold. Actually, provided the observational would be complete, which is likely inherently impossible for low probability / high consequences events like tsunamis, this would be enough for assessing tsunami hazard; in many cases, frequencies estimated from tsunami data need to be considered as a lower bound not to be exceeded by hazard estimates. In **Figure 4C1 and 4C2**, we report for explanatory reasons two potential observation relative to the same intensity defined for the scenario-based hazard is represented as a horizontal green point. For what it concerns the probabilistic hazard assessment, this observation represents the ground for a quantitative test against ontological errors in probabilistic hazard assessments, as discussed in Marzocchi et al. (2015). Indicatively, this green point should be inside the range of possible mean annual frequencies defined in the black dashed box (**Figure 4C2**). If instead the point is outside the range (**Figure 4C1**), the probabilistic hazard model can be likely rejected, either because it is not test (as discussed above) or because it includes an ontological error (see discussion in Section 1). A similar comparison (but only qualitative) can be made also with the scenario-based analysis: if the observed frequency (green dot) is distant from the target mean return period adopted for selecting the scenario (blue dot), a new selection should be considered.

This discussion had the simplifying assumption that the uncertainty associated to the observation point is negligible, but it can be easily extended to account also for measure uncertainty by means of standard hypothesis testing theory (Marzocchi and Jordan, 2014). In particular, the measure uncertainty would result with both horizontal and vertical error bars in **Figure 4C1 and 4C2**, expressing the measurement error and the uncertainty in the frequency estimation. Note that this might be useful in the specific case of tsunamis, given that they are be generated by different sources, and a reason for inconsistency might also be that the hazard assessment can be focused on only one (or few) source, while the tsunami catalogues may mix different sources. This indeed should be taken into account, and it can be possibly quantified in terms of measure uncertainty.

All the described comparisons can be done also for risk results. The only difference is that in the abscissa there will be a loss measure, instead of the hazard intensity.

To summarize:

- Probabilistic analyses allow for a deeper understanding of results, potentially increasing the awareness of decisions based on hazard and risk results; the analysis

of results is more complicated, and a closer interaction between hazard/risk analysts and end-users (decision makers) is required.

- The analysis of scenarios may provide important clues in all “what if” studies, allowing for some specific modelling otherwise not feasible; for the consideration into hazard/risk analysis, the selection of scenarios should be linked to a reference “mean return period”. In this case, a scenario-based analysis can be seen as the maximum simplification of the hazard/risk analysis, that is, the definition of one point in the intensity/probability graph.
- Probabilistic hazard and risk analysis is a complex and resource-demanding process. Sanity checks of the data, methods and results are necessary. Critical choices during each of the analysis steps must be made in a transparent and reproducible way, and by defining a structured process on purpose, for example by means of expert elicitations, participatory and independent reviews.
- While the PTHA for earthquake-generated tsunamis is probably mature enough for operational applications, this is not yet the case for other tsunami sources such as landslides and volcanoes, which at least locally may seriously contribute to tsunami hazard. For example because the source frequency is presently more difficult to constrain.
- When available, probabilistic scenario-based analyses are comparable only if scenarios are defined in terms of a reference mean return period. In this case, both results can also be compared, in principle, and at least qualitatively, with past data. However, for tsunamis, catalogues are likely to be under-representing the largest low-frequency events, and it is difficult to discriminate the causative tsunami source.

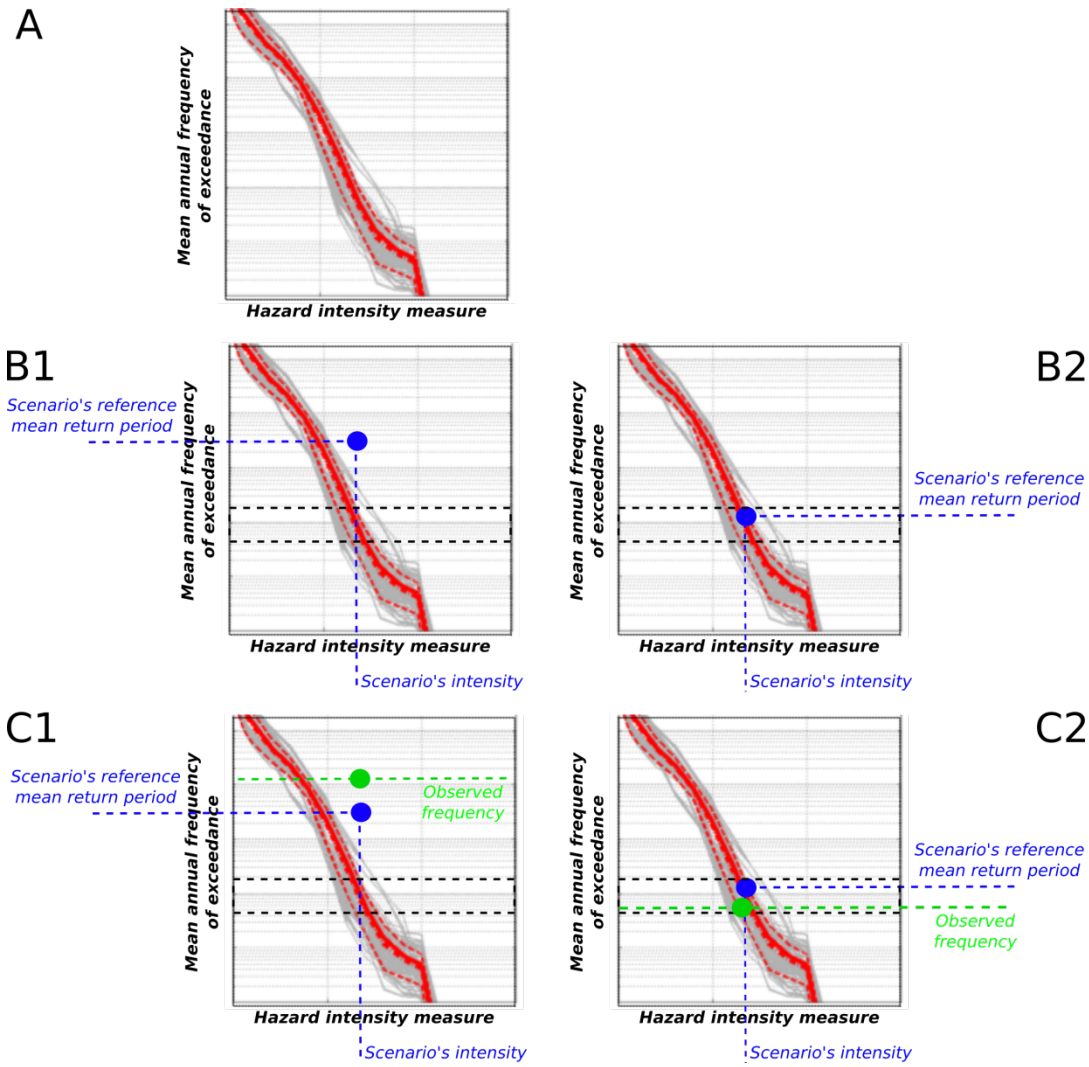



***Figure 4**: Comparison between Probabilistic Hazard Analysis (hazard curve in red and grey), scenario-based intensity (blue), and past data (green), modified from Selva et al. (2016).*

# 3. Proposal for a homogenised treatment of uncertainty for tsunami hazard and risk, with examples

For natural hazards and consequent risks, we deal with systems with many degrees of freedom, often characterized by non-linear dependencies, and sometimes by a relative lack of observations. So, and particularly when potential regulatory concerns are relatively high, the management of scientific and technical controversies within a multiple-expert environment is critical for any hazard/risk assessment project.

As discussed already in the Introduction and in some details in **Section 2**, there might be, and indeed there are usually, different opinions within a group of scientists concerning some aspects of the hazard/risk assessment chain; it is then mandatory not to overlook or underrate such controversies, as different opinions and approaches typically lead to very different estimates of hazard and risk.

Alternative probabilistic formulations may produce and they have in fact produced in the past highly variable and in some cases inconsistent hazard and risk estimates (e.g., Paté-Cornell, 1996; Bernreuter et al., 1989; Selva et al, 2013; Marzocchi et al., 2015). Since the results of such analyses may have consequences of public relevance, it is fundamental to manage the emerging uncertainty (e.g., SSHAC, 1997; USNRC, 2012; IPCC, 2013).

The integration of different expert opinions is then needed for managing subjective decisions and in quantifying the epistemic uncertainty in the form of "technical community distributions". To produce robust and stable results, the experts' diverse range of views and opinions, their active involvement, and their formal feedbacks need to be organized into a structured process, ideally granting transparency, accountability and independency.

In the field of natural hazards, and in particular in seismic hazard and risk quantification, the most applied and referenced protocol is the one developed by the Senior Seismic Hazard Analysis Commission (SSHAC) in 1997, and its subsequent specifications and integrations (e.g., Hanks et al., 2009; USNRC, 2012). One of the most relevant conclusions in SSHAC guidelines about uncertainty management and use of experts is that differences in PSHA assessment are often due to procedural rather than technical differences, and so a great effort has been put in establishing appropriate procedures.

However, when the regulatory concern is not the highest possible (e.g. for Nuclear CIs), and/or in small/medium size projects, applicability of the full SHAAC procedure may not be feasible. Furthermore, the group interaction and face-to-face meetings between the experts can be challenging for the undesirable effects of personalities and reputations that unavoidably may bias the group quantification (e.g., Bedford & Cooke, 2001; Aspinall and Cooke, 2013). Lastly, some "technical neutrality" is desired, since what really matters is the standardization of the process, not the specific tool (e.g. a logic tree or an ensemble) to be used.

A more flexible management of subjectivity in probabilistic single/multi-hazard/risk assessments has been addressed within a structured Multiple-Expert Management Protocol within the STREST project (EU@STREST in Selva et al., 2015, STREST D3.1). A development of this approach is being now applied specifically to seismic probabilistic tsunami hazard assessment (S-PTHA) within the TSUMAPS-NEAM project.

The TSUMAPS-NEAM project benefits of some data analyses and - more importantly with respect to the topics discussed herein - of some several methodological advancements addressed within ASTARTE. They regard for example the seismic source definition and treatment (see e.g. D3.12 and D3.40), the general approach to uncertainty quantification (as discussed within this deliverable and in D4.13), and several others, including very preliminary regional hazard results presented in D8.8.

A further specific advancement achieved within ASTARTE, and presented in this deliverable, concerns the development of the uncertainty treatment for approximated Maximum Inundation Height (MIH) estimation through local amplification factors, applied to each single scenario contributing to PTHA before probabilistic aggregation.

The next step following these technical achievements, as well as several discussions with the ASTARTE partners, was taken in TSUMAPS-NEAM as regards the application of a method for the general management of uncertainty, including those emerging from possible alternative formulations of the same problem.

In what follows, we then briefly describe the TSUMAPS-NEAM project as a whole and the PTHA methodology applied therein, with specific focus on the Multi-Expert Uncertainty Management Protocol. We will then report on the methodology for the treatment of uncertainty for MIH, developed within ASTARTE and which will be applied in TSUMAPS-NEAM.

## 3.1 Regional Seismic - Probabilistic Tsunami Hazard Analysis (SPTHA) in the NEAM Region, under development within TSUMAPS-NEAM Project

As said in the Introduction and earlier in this Section, TSUMAPS-NEAM has its roots within the ASTARTE Project. Also thanks to application, further development, and integration of several studies performed within the ASTARTE project, TSUMAPS-NEAM will develop the first homogeneous long-term PTHA for earthquake-induced tsunamis (S-PTHA), which is presently unavailable for the coastlines of the NEAM region (NE Atlantic, the Mediterranean, and connected seas).

The expected results are: 1) long-term PTHA (complete *hazard curves*), i.e. exceedance probability for maximum offshore amplitude at a 50-m depth, and for Maximum Inundation

Height (MIH), at points distributed along the NEAM coasts; 2) regional hazard maps (hazard intensity with given exceedance probability; 3) probability maps (probability corresponding to given intensities).

This regional S-PTHA is meant to be the term of comparison or a basis for existing or future national and site-specific S-PTHA and Risk assessment efforts, for example as a preliminary screening for the prioritization of subregions or smaller areas for site-specific studies. Common risk assessment, long-term risk mitigation and planning, at the national and regional levels, and several specific applications (e.g. land-use and evacuation plans, identification of Critical Infrastructures (CIs) at risk) can clearly benefit from having a region-wide S-PTHA as input and reference.

Therefore, this S-PTHA will rely on common understanding of the best viable practices and complies with EU scientific and policy standards for hazard and risk assessment (e.g., seismic hazard for building codes). Moreover, the development of standardized PTHA products (hazard and probability curves, maps, documentation, web-tools for their analysis) is the first step for multi-source tsunami hazard assessment, i.e. the complete PTHA (e.g. including landslide and volcano sources), tsunami risk assessment, and also to include tsunamis in multi-hazard risk assessments.

A quantitative definition of the tsunami hazard in the region is also crucial for better addressing future risk reduction political choices in the context of the NEAM TWS (Tsunami Warning System), and a critical tool for a thorough effort for awareness raising in the region.

### 3.1.1 TSUMAPS-NEAM Methodology for Regional-scale S-PTHA

The workflow for S-PTHA adopted in TSUMAPS-NEAM is organized into the following 4 STEPS:

- STEP 1: PROBABILISTIC EARTHQUAKE MODEL
- STEP 2: TSUNAMI GENERATION & MODELING IN DEEP WATER
- STEP 3: SHOALING AND INUNDATION
- STEP 4: HAZARD AGGREGATION & UNCERTAINTY QUANTIFICATION

Each of these steps is further divided into several Levels, which will be omitted in this brief overview. These Levels describe the finer grain of the analysis workflow within each STEP, and the definition of the databases on which the analyses rely on.

In general, at each of the Levels within each STEP several alternative models (as discussed in the introduction of this section) might be used for the same problem: a different magnitude frequency distribution (MFD) for the earthquakes, two different corner magnitudes for the

same MFD, a shallow water or a Boussinesq approach to tsunami modeling, two different bathymetric models, and many others. Among all the alternatives considered technically sound, only a selection of them is generally worth to be implemented in a real assessment, since only a sub-set of them usually control the largest part of the epistemic uncertainty. The role of the panel of experts (PoE) through a structured process, as described further below (in Section 3.1.2), is to provide guidance on this selection ("trimming" of the alternatives); then, to provide guidance on the relative weights of the different alternatives (usually within a logic tree; here within a more general "alternative" tree, within an ensemble modeling scheme; Selva et al., 2016).

The TSUMAPS-NEAM STEPS as a whole are very synthetically described in what follows, that is without entering into the details of the Levels within each STEP. Then, the different roles within the project are described, in particular that of the PoE and their role in the decision-making process.

**STEP 1** - PROBABILISTIC EARTHQUAKE MODEL

The goals of STEP 1 are the definition of the set of seismic sources, their frequency, their parameterization, and the probabilities associated to different configurations of the source parameters.

This analysis, i.e. the treatment of the aleatory (natural) variability of the seismic sources, is conducted through the definition of an Event Tree (ET). An ET is a branching graph representation of events in which individual branches are possible alternative steps from a general prior event, state, or condition, and which evolve into increasingly specific subsequent events. Examples of the use of an ET for S-PTHA can be found in Lorito et al. (2015) and in Selva et al. (2016); see also ASTARTE D8.8. This type of approach helps decreasing the overall computational coast of S-PTHA, and it is an alternative to more classical approaches for the discretization of the total probability in S-PTHA (e.g. Geist and Parsons, 2006).

STEP 1 defines the ET used in TSUMAPS-NEAM for the treatment of the seismic sources aleatory variability. All Levels in STEP 1 except for Level 0 coincide with the nodes of this ET. At each Level (i.e. at the ET nodes), discrete probabilities are evaluated for the parameters under investigation. The Levels are organized in a logical sequence, and probabilities at each Level are conditioned to the branches ("events") at the previous Level.

Without entering in any detail, we just mention that the Earthquake model at STEP 1 is composed by two types of seismic sources, which are: the "Predominant Seismicity (PS)" (already called "Interface Seismicity (IS)" in D8.8, and in Selva et al., 2016), and the "Background Seismicity (BS)".

PS is used when well-known fault structures (e.g. geometrically well-constrained) may be of particular relevance for tsunami generation (typically, subduction interfaces). With

reference to terminology adopted for the types of seismic sources defined in the proposed standardization by ASTARTE D3.40, these are the “Earthquake fault sources”.

BS seismicity is used instead since no one can exclude that earthquakes may occur outside well-known faults, and we cannot exclude that faults are not mapped well enough everywhere (which is likely impossible, particularly offshore). Hence, when dealing with PS we use complete fault models for the earthquakes, while we allow earthquakes to occur everywhere in a given volume with a given variability of the faulting mechanism. In D3.40, they were termed “Earthquake Grids”. The presence of known faults (not those already treated as PS), historical seismicity (focal mechanisms) and dominant stress regime can be used to constrain the faulting mechanism probability.

In each region defined by the regionalization (see D3.40), three different situations may occur: 1) a region is treated as a mix of PS and BS (e.g. a subduction zone and the crustal earthquakes above it); 2) a region is treated as pure BS (similarly to some PSHA approaches, in the present case were no really major structures that are mapped well enough are present); 3) a pure PS region (e.g. subduction zones located very far away from the target coast, for which modelling the largest earthquakes occurring on the known subduction interface is enough, such as the Caribbean subduction with respect to the target NEAM coastlines.

**STEP 2** - TSUNAMI GENERATION & MODELING IN DEEP WATER

The goals of STEP 2 are the numerical simulation of the sea floor displacement, and the numerical simulation of the tsunami generation and propagation from the source up to a given bathymetric depth offshore of the target area. The tsunami propagation is simulated up to a chosen isobath offshore of the target coast, up to which the propagation is assumed linear.

This assumption allows the formulation of the problem as a linear combination of pre-calculated tsunami Green’s functions produced by unit sources (‘unitary’ Gaussian-shaped sea level elevation; Molinari et al., 2016; see also Selva et al., 2016; ASTARTE D8.8). This limits the computational burden and makes the analysis computationally feasible. Further commonly adopted simplifications (cf. ASTARTE D3.40) are also employed at all levels of this STEP (e.g. Okada-like faults in homogeneous half space).

However, the Levels inside STEP 2 (coseismic displacement, tsunami generation, propagation in deep waters) are separated in a way that makes the simulation ‘modular’ allowing in principle to adopt more complex approaches for each Level (see e.g. D3.40, D4.19). The uncertainty introduced by these simplifications is treated jointly with propagation in shallow water, shoaling and inundation, at STEP 3.

**STEP 3** - SHOALING AND INUNDATION.

The goals of STEP 3 are the simulation of the last phases of the tsunami impact, including combination with tides, and the quantification of the associated uncertainty (using also the methodology developed within ASTARTE and presented in Section 3.2).

This STEP is kept separated from STEP 2 in TSUMAPS-NEAM as the linearity assumption doesn't hold anymore during the later stages of tsunami evolution in very shallow waters and beyond (shoaling, inundation). Moreover, the typical discrete length/time-scales for the numerical simulation of these stages need to be much finer than for modelling the propagation in the deep ocean. As a consequence, the computational cost strongly increases, and full inundation calculations for each considered scenario are generally not feasible for a regional hazard quantification (e.g., Davies et al., 2017). Hence, these last phases need to be treated with specific and generally quite crude approximations replacing direct numerical simulations (e.g. Løvholt et al., 2012); as a consequence, a specific uncertainty treatment (see Section 3.2 and also Davies et al., 2017) is needed and the applicability of the results is nevertheless quite limited as far as site-specific studies are concerned. Again, the modularity of Levels leaves the possibility open for different approaches for inundation calculations if enough computational resources and detailed enough topo-bathymetric models are available; or the re-use of the regional results for subsequent local studies (e.g. Lorito et al., 2015), disaggregation or sensitivity analyses.

**STEP 4** - HAZARD AGGREGATION & UNCERTAINTY QUANTIFICATION

The goals of STEP 4 is the quantification of the hazard curves at the target sites, that is of the exceedance probability in a given time window for the chosen hazard metric at the points of interest including treatment of epistemic uncertainty. It merges the results of STEP 1, namely the probability associated to each considered scenario, with the tsunami impact due to each scenario (STEP 2 + STEP 3) to calculate the hazard curves at chosen points of interest. At sites where this is possible because of data availability, this STEP includes comparison of the results of computationally based S-PTHA with observations (historical tsunamis, paleotsunamis).

For quantifying the uncertainty on hazard curves, the workflow of STEPS 1-3 is repeated for each alternative model (or a sample of it, see Selva et al., 2016) that is adopted at any Level within these STEPS. The process of alternative selection and weighting through expert elicitations is discussed in the next subsection 3.1.2. Note that the elicitation process influences the alternatives considered at all levels in the previous steps, as well as, at this STEP, for example by 'trimming' alternative models that are judged unnecessary.

The treatment of the epistemic uncertainty, expressed by alternative hazard curves, is finally treated by ensemble modeling (see previous discussions in this deliverable; D8.8; Selva et al., 2016).

In Figure 5, we report some preliminary results of the actual implementation of the computational chain described so far, restricted to the Mediterranean see. For sake of examples, we report the hazard maps (intensity corresponding to a given Mean Return Period - MRP) for the mean of the ensemble models (epistemic uncertainty), and different MRP.

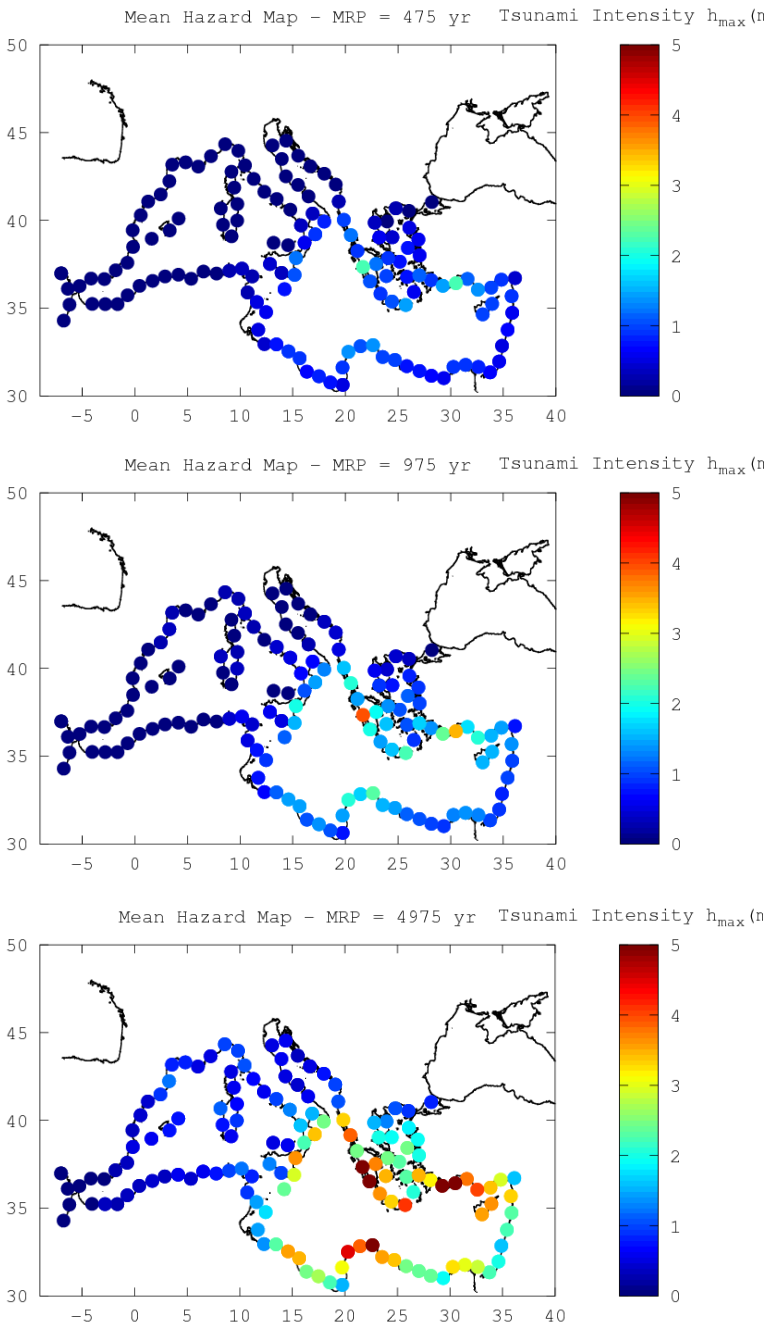



**Figure 5:** Preliminary Regional Hazard Maps (covering the whole Mediterranean) relative to an Average Return Period of 475 yr (above), 975 yr (center), and 4975 yr (below), corresponding to probabilities of 0.10, 0.05 and 0.01 in 50 yr within a Poisson process, respectively. The mean of the epistemic uncertainty is reported.

### 3.1.2 TSUMAPS-NEAM Multiple-expert process for uncertainty quantification

The procedural for the management of subjective choices and uncertainty quantification of TSUMAPS-NEAM is rooted in a clear definition of roles and interaction among the different experts, structured elicitations based on mathematical aggregation of a pool of experts, and in a participatory independent review. This is achieved with coordination of a technical integrator and a project manager, within a formalised process. Several groups are defined in the framework of this process. The roles of the participants to each group will be briefly described below.

Among the different parts of the process and the different groups, the multi-step / multi-expert process with the panel of experts (PoE), for addressing subjective choices, will be described in slightly greater detail, since it is the core process for addressing subjective choices.

Note that TSUMAPS-NEAM included several experts that are also ASTARTE partners in this process.

In a nutshell, the purpose of the protocol is:

1. To establish roles and responsibility, in order to guarantee transparency, independency of roles, accountability and achievement of procedural consensus;
2. To homogenize the management of decision making for subjective choices, guaranteeing documented and trackable decision making;
3. To establish homogeneous principles for the management of alternatives, that is, alternative and scientifically acceptable implementations for quantifying the community distribution.

Roles in the TSUMAPS-NEAM process have been defined as follows:

- the Project Manager (PM);
- the Technical Integrator (TI);
- the Evaluation Team (ET);
- the Pool of Experts (PoE);
- the Internal Reviewers (IR).

The PM coordinates the development of the whole project, and he is typically the Project Coordinator appointed by the main stakeholder / representative of the funding agency. The PM has the responsibility for decisions to be rational and fair to stakeholders, authorities and to the public whenever this is the case. The PM don't usually have a formal role in structured expert elicitations. In TSUMAPS-NEAM, the PM role is played by the Project Coordinator, with the support of one or two Steering Committee members.

The TI is project scientific manager; his main responsibility is to try and capture, and to integrate, the views of the "informed technical community" to be implemented in the hazard and risk calculations. The final results are to be expressed in the form of "community distributions" (SSHAC, 1997), and the opinions expressed along the process should be trackable. The TI has a role that merge the one of the standard elicitator of structured elicitations (e.g., Aspinall and Cooke 2013) and the one of technical integration SSHAC (1997; and followings). It is desirable to have a small team rather than a single person acting as TI; in TSUMAPS-NEAM this is performed by several Task Leaders and some project partners.

The ET is the group who actually perform the hazard assessment; this group is led by the TI. This team is selected jointly by the TI and the PM and, in TSUMAPS-NEAM, it is constituted by the project consortium.

The PoE is meant to be representative of the broader technical community. It supports the ET for critical subjective choices trough the structured elicitation experiments. The PoE may partially overlap with the ET, and it actually does in TSUMAPS-NEAM, but it should contain a substantial presence of external experts too, with no overlap with other groups. Experts 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 in developing the community distribution regarding hazard assessments. The experts of the pool act independently and PoE meetings are moderated by the TI. Typical sizes of panels are in the order of 7-15 components (e.g., Aspinall and Cooke, 2013), even if sometimes larger panels may be appropriate in some circumstances (Hoffmann et al., 2007). The PoE of TSUMAPS-NEAM is made of 15 experts, selected among experts of earthquakes, tsunamis, both in general and with specific knowledge of the NEAM region, and of probabilistic analyses. It is composed by 8 experts selected among the partners, and by 7 invited external experts belonging to the international scientific community. In particular, 3 of them are from European Institutions, 4 from Institutions outside Europe.

The IR a group of experts who are meant to provide a peer-review of the process, the methods, and of the results. Comments and recommendations can be provided during the implementation of the project (not only at the end); this participatory review process happens at pre-defined discrete stages. In particular, IR reviews the project both in terms of technical (scientific models, testing procedures, etc.) and procedural (actor's independency, transparency, consistency with the project plan, etc.) aspects (SSHAC, 1997). The IR group

size may on the complexity of the project. Note that through the participatory review scheme, the IR plays an active role during the project and thus it is part of the project itself. If regulators or other external authorities foresee an external review of the project results, this further review is performed independently and after the end of the project. In TSUMAPS-NEAM, the IR is composed by 5 experts, including end users belonging to Civil Protection authorities.

We won't enter here into the details of the many activities and the complex interactions among all these groups which occurred during several meetings, remote interactions and day by day work. We will limit ourselves to describe, schematically, the main steps of the tasks performed or to be performed with the involvement of the PoE; these tasks are the core of a solid decision-making process concerning the subjective choices among possible alternative implementations and models, alternative weighting, and uncertainty quantification.

The starting point for decisions is constituted by two formal feedbacks requested to the PoE, occurring during two distinct TSUMAPS-NEAM project phases.

The first one is the pre-assessment, when a prioritization of the STEPS and Levels within which the analysis of the alternatives should be deeper than for other STEPS and Levels, as they are expected to influence the epistemic uncertainty; hence, during this pre-assessment phase, the full set of models to be actually implemented is defined, and consequently other possible models are excluded ("trimming" of the alternatives). When doing this, the PoE should be provided with the necessary supporting material, including the description of methods, alternative models and/or potential sensitivity tests.

This issue is critical, since the exclusion of some models can drastically modify the body and the range of the community distribution; in other works, it can significantly alter the results of the assessment. Conversely, including all the available models is virtually impossible, probably useless, and sometimes dangerous (Bommer and Scherbaum, 2008).

The second phase is the assessment; the assessment includes the selection of the weights for the alternatives included in the ensemble model of S-PTHA uncertainty.

Both selection and weighting of alternative models may be based again on prioritization techniques, like the PC (Pairwise Comparison, e.g., Maida et al., 2012), the AHP (Analytic Hierarchy Process, from Saaty, 1980), and BBN (Bayesian Belief Networks, e.g., Bayraktar et al., 2009).

This process is conducted in TSUMAPS-NEAM as follows.

The pre-assessment phase starts with assessing the organizational part of the project within the given constraints (e.g. timing & total budget); the selection of all the main actors, the pre-selection of the structure of the assessment (STEPS and Levels presented above), the

pre-selection of a vast range of possible alternative implementation. These activities end with the organization of a Project kick-off meeting, which results in full documentation of all these aspects.

Then a second meeting, the PoE kick-off meeting, is organized, where the experts are presented with all the necessary elements and documentation. During this PoE kick-off meeting, the assessment workflow and the possible alternatives are discussed in great detail and initial feedback is also requested about potential gaps the PoE may identify. Moreover, the experts are presented with the workflow of the foreseen decision-making process in which they will be involved. This enabled an effective discussion with experts to clarify the scope of the tasks of expert elicitations and their specific role.

The formal feedbacks of the PoE are based on the use of structured expert elicitations of the PoE (for a review of decision making based on expert elicitations, see Morgan 2014), that is through a trackable process. Note that the use of expert judgements is well established in many fields, such economics, politics, medicines, climate studies, volcanology and seismology, and in general in decision making under uncertainty (e.g., Aspinall and Cooke, 2013; Morgan, 2014).

The experts themselves are assigned weights for quantifying the "credibility" of their feedbacks. Weighting of experts can be done in different ways and it is a fundamental part of the elicitation process, even though often equal-weighted procedures are still considered (e.g., Cooke, 1991; Budnitz et al., 1998; Aspinall and Cooke, 2013). In any case, weighted and not weighted (i.e. equal-weighted) results are always cross-checked for consistency.

To achieve this, during the PoE kick-off meeting, the 15 experts of TSUMAPS-NEAM answered to a seed questionnaire, prepared by the TI, which consisted of two parts.

The first part of the questionnaire concerned topics related to S-PTHA, such as earthquake and tsunamis in general, probabilistic methods, previous assessments in the NEAM region. The experts were asked to express their best guess and confidence intervals ($5^{th}$, $50^{th}$, and $95^{th}$ percentiles) to each question to quantify their own uncertainty on each subject. Their assessments were used to estimate weights using Excalibur, a software package for structured expert judgement elicitation using the Classical Model (also called Cooke's Model; Cooke, 1991).

During the second part, each expert was asked to acknowledge two other experts of her/his choice within the PoE.

These activities are propaedeutic to the application of different weighting schemes. Three alternative weighting schemes are indeed taken into account in the elicitation experiments of TSUMAPS-NEAM. The alternatives schemes considered are Performance-based Weighting (PW; aiming to quantify the "credibility" of experts' answers in terms of their capability in constraining their subjective uncertainty; Cooke, 1991) and Acknowledge-based Weighting

(AW; aiming to quantify the "credibility" of experts' answers in terms of their representativeness within the community; Selva et al., 2012). Weights are assessed through the questionnaire (Figure 6). Equal Weighting (EW) is also considered as a baseline. The results obtained in elicitation experiments under alternative weighting schemes will be compared, in order to check consistency and sensitivity of the results, and all will provide the input for final decision-making.

Once all is set during the PoE kick-off meeting, the formal interviews (i.e. the two elicitation steps, one for prioritization and trimming, and one for weighting) can be performed remotely, which decreases their cost.

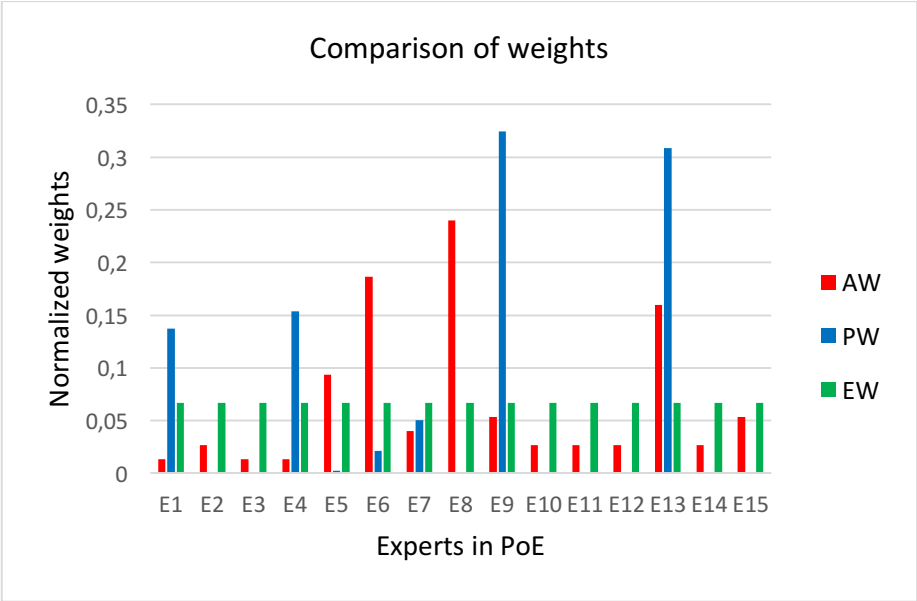



***Figure 6*** *Comparison of weights assigned to experts based on alternative weighting schemes.*

A first elicitation round then occurs in the period of time following the PoE meeting. This elicitation round is meant to design the final scheme for the analysis that concludes the pre-assessment phase.

In this first elicitation the PoE is asked to make detailed pairwise comparison (PC, e.g., Maida et al., 2012) between each pair of STEPS, and then between each pair of Levels within each STEP, according to **Table 1**. The analysis of the results is conducted with an Analytic Hierarchy Process (AHP, Saaty, 1980).

***Table 1:*** *Fundamental scale of absolute numbers*

| **Intensity of Importance** | **Definition** | **Explanation** | **Weights of models** | ***Standard AHP weights*** |
|---|---|---|---|---|
| 1 | Equal importance | Two steps/levels/sublevels contribute equally to the objective | 0.5-0.5 | *0.5-0.5 (x1)* |
| 3 | Moderate preference | Experience and judgment slightly favor one step/level/sublevel over another | 0.6-0.4 (x1.5) | *0.75-0.25 (x3)* |
| 5 | Strong preference | Experience and judgment strongly favor one step/level/sublevel over another | 0.75-0.25 (x3) | *0.83-0.17 (x5)* |
| 7 | Very strong preference | A step/level/sublevel is favored very strongly over another; its dominance demonstrated in practice | 0.95-0.05 (x19) | *0.86-0.14 (x7)* |
| 9 | Extreme preference | Overwhelming evidence favoring one step/level/sublevel over another | 0.99-0.01 (x99) | *0.90-0.10 (x9)* |

As an example, we report only the result of the prioritization between the 4 STEPS presented before (Table 2). Analogous Tables exist for the Levels within each STEP. We point out once again that prioritization here means what STEP needs to be addressed by implementing a larger number of alternative models since it is characterized by larger epistemic uncertainty; epistemic uncertainty needs, in turn, to be described through these alternatives.

***Table 2**: Results for the prioritization of STEPS*

| No. | Model code | Description |
|---|---|---|
| 1 | STEP1 | Definition of the seismic source variability and quantification of the long-run frequencies of all the seismic sources |
| 2 | STEP2 | Tsunami generation and off-shore propagation |
| 3 | STEP3 | Near-shore tsunami propagation and inundation |
| 4 | STEP4 | Computation of the weights of the alternative models developed in STEPs 1 to 3 to measure their credibility, and construction of the “ensemble” model |

In particular, the prioritizations obtained by the different weighting schemes are compared both in terms of central values and of inter-expert distributions. Based on this comparison, the steps and levels of the SPTHA are ranked in three groups:

- **High priority (red):** steps/levels with clear high priority in all weighting schemes. For these steps/levels, alternative implementations are strongly recommended by the PoE. In this case, the alternatives should be carefully selected to represent a range of models that cover the full range of scientifically acceptable modeling alternatives (following SSHAC 2012, "the center, body, and range of technically defensible interpretations").

- **Medium priority (orange):** steps/levels with either high priority in one (but not all) the weighting schemes, or intermediate priority in all weighting schemes. For these steps/levels, an evaluation of the potential consequence of alternative implementations is recommended by the PoE. In this case, some alternative implementations should be considered and/or some sensitivity test should be planned.

- **Low priority (green):** steps/levels with low priority in all weighting schemes. For these steps/levels, the PoE suggests a relatively low potential impact of epistemic uncertainty, so that one preferred implementation can be considered.

In **Figure 7**, we also report an example of the AHP results, obtained from the answers of the PoE during this first elicitation. The plots contain:

- The empirical CDF of the scores of the proposed alternatives, obtained by considering the prioritization of the different experts as weighted samples;

- The parametric variability of the scores of the proposed alternatives, considering arithmetic and geometric means and percentiles $16^{th}$, $50^{th}$ (median) and $84^{th}$;

- The Consistency Ratio (CR) (Saaty, 1990) of all the experts, compared with thresholds of 0.1 and 0.3;

- The weights of the experts, adopting the different weighting schemes.

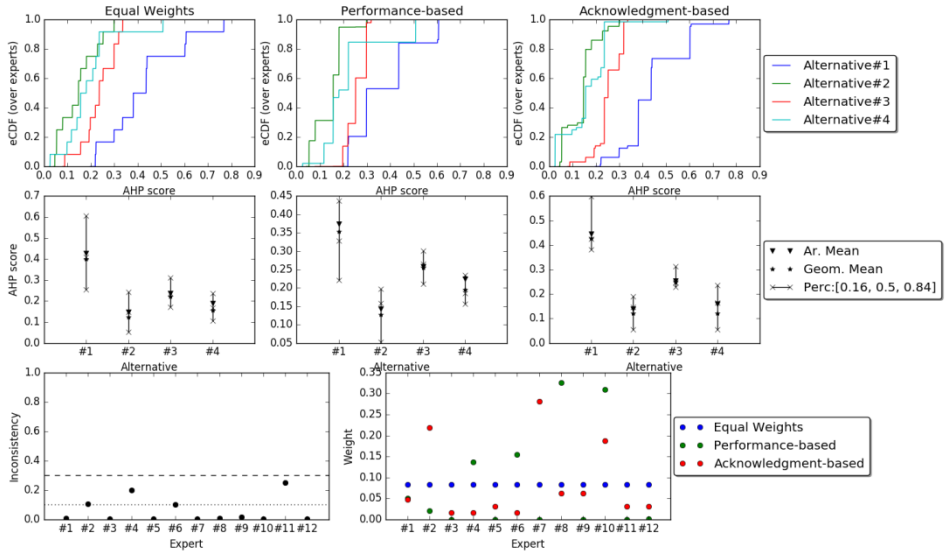



***Figure 7:*** *AHP results for the prioritization of STEPS, as defined in Table 3.*

After this elicitation the method could be drafted, with the selection of the alternative implementation. Based on the above results, a larger number of alternative models have been developed to carefully explore and quantify the epistemic uncertainty for STEPS 1 and 3 than for STEPS 2 and 4 (and within them for the Levels that were judged more important). Of course, alternative models are either implemented, if this is feasible within the resources allocated to the project; or the need for their implementation in a future assessment is to be clearly reported. This drafted method is then sent to reviewers (IR). After this review, the method can be finalized and the final assessment phase could be started.

In answering to the first elicitation, or during the revision, it may emerge the existence of substantial gaps that prevent the quantification of the results under specific conditions. If these gaps cannot be filled within the project, these types of exclusions can lead to a “grey swan” (an event which is foreseeable, but cannot be considered in the analysis; Paté-Cornell 2012). To quantify the effects of these gaps, missing models and the selection of scenarios might be, performed, for example through the so called “Classical model” or “Cooke’s method” for structured expert judgment (e.g., Cook, 1991; Neri et al., 2008). In TSUMAPS-NEAM, one critical aspect that emerged during the PoE kick-off meeting, was the immediate need for alternatives at STEP3 and more in general the uncertainty quantification related to this STEP. For this reason, a specific study has been developed within ASTARTE (describe in the next Section 3.2), to devise an uncertainty quantification method to be then applied to the TSUMAPS-NEAM S-PTHA. This is presented in the next Section.

At this stage of TSUMAPS-NEAM, a full preliminary S-PTHA has been conducted, so far only for the Mediterranean (as already reported in Figure 5). Preliminary weights have been assigned by now to the alternative models. The second elicitation will concern the

assessment through the PoE of the final weights, with similar methods to those illustrated for the first elicitation.

To quantify the weights of alternative models, this subjective credibility may be also mixed with more objective criteria (e.g., results of sanity checks; results of model's correlations; likelihood score for independent data), based on the performance of the models on independent data (e.g., Marzocchi et al., 2012; Davies et al., 2017) and/or on sanity checks on recent data.

Once all methods are implemented and weighted, the different results should be aggregated to express the community distribution. This is done in TSUMAPS-NEAM by means of ensemble modelling (e.g. Selva et al., 2015).

## 3.2 Quantification of tsunami run-up uncertainty using approximate amplification factor methods

The standard way of estimating the tsunami run-up and inundation maps is to apply depth averaged nonlinear shallow water (NLSW) models that include drying-wetting schemes. However, these detailed numerical inundation simulations are too intensive computation wise when (i) we need to estimate tsunami run-up height for many (often thousands and millions) of scenarios (such as for PTHA) and (ii) we need to cover large stretches of coastlines with complex geometries, for instance covering scales over one or more countries.

While intermediate methods limiting the number of scenarios in PTHA exists (e.g. the pioneering application of Gonzalez et al., 2009 and the formal procedures developed by Lorito et al., 2015), we need simpler methods for determining tsunami inundation heights in PTHA and over large regions. To this end, a faster procedure, with a higher degree of approximation, is to relate the near-shore surface elevations at the hazard points to the maximum shoreline water levels. The surface elevation at the shoreline then acts as an approximation for the maximum inundation height or run-up height along the shoreline. The method is described by Løvholt et al. (2012) and Løvholt et al. (2015). Combined with results from offshore tsunami simulations, it can be used to estimate the mean or median tsunami run-up or maximum inundation height at a coastal location.

A new version of the amplification factor method, takes into account the local bathymetry. The methodology has been developed for production of tsunami hazard maps for the TSUMAPS project (http://www.tsumaps-neam.eu/). The new method, as briefly described below, uses a set of different local transects normal to the coastline as a basis for estimating the amplification factor to be applied to the incoming tsunami as modeled at a single point of interest in front of the stretch of coast under examination. The local amplification factor method is expected to replicate the median tsunami inundation height more accurately

than the previous method using idealized profiles (Løvholt et al., 2012). However, as the method is intrinsically deterministic, we need to compare the method with local inundation models to quantify its associated bias and uncertainty. Quantifying this uncertainty is also necessary for the general uncertainty quantification of tsunami hazards in TSUMAPS.

Previous attempts to measure the bias of the amplification factor method, as well as quantifying tsunami inundation height uncertainty, is limited to a single study (Davies et al., 2017). This study estimated the amplification factor uncertainty and bias by comparing tsunami simulations combined with amplification factor analysis with observed run-up from field observations for four large tsunamis (Chile, 1960; Alaska, 1964; Indian Ocean, 2004; Tohoku, 2011). We also note that the analysis of Davies et al. (2017) used a version of the amplification factors that was based on idealized and much simplified bathymetric transects.

Davies et al. (2017) found that tsunami simulations combined with the amplification factor gave a relatively small bias compared to field observations (the amplification factor overestimated the maximum inundation height), however, a large lognormal variance of ($\sigma^2 \approx 1$) was found. However, in these results the uncertainty was mixed with the source uncertainty. Since Davies et al. (2017) considered real cases where field data are compared to full simulations of tsunamis from the generation to the target site, this comparison implicitly combines the variability due to heterogeneous slip, other simplifications in source and tsunami generation and propagation modeling, in addition to the inherent variability in the run-up process itself. It is therefore desirable to conduct a dedicated study that only address the uncertainty of the run-up process, by directly comparing detailed high-resolution inundation simulations to those using the amplification factors, but starting by a common source and open ocean propagation model.

In the present section, we carry out a set of local inundation simulations for different earthquake source magnitudes, and compare the distribution of the maximum inundation heights (MIH) with results from the local amplification factors. The work is a collaborative effort including partners from NGI, INGV, GFZ Potsdam, and IPMA. In addition, NOA and UB have contributed with local bathymetric and topographic data. The simulations are carried out at three ASTARTE test sites (Heraklion, Crete, Greece; Colonia Sant Jordi, Mallorca, Spain; and Sines, Portugal). For each simulation, we estimate the bias of the amplification factor method, by comparing the results with median local MIH values derived from NLSW models. Furthermore, we derive the lognormal σ of the MIH from the NLSW models. We also discuss the overall uncertainty for all the simulations combined. We use the variable MIH from the NLSW models to include uncertainty to the MIH amplification factor method. In addition, we conduct basic statistics of the amplification factor bias. This statistics is an additional source of epistemic uncertainty (in addition to the local MIH variability).

### 3.2.1 New set of amplification factors based on local bathymetric transects

Here, we briefly explain an improved method of amplification factors, based on the original work of Løvholt et al. (2012). Where Løvholt et al. (2012) used a limited set of idealized bilinear profiles to represent the bathymetry, we here use local bathymetric profiles. The new method is based on the following steps:

- ✓ The amplification factors are computed for evenly spaced hazard points located every 20 km along the shoreline. This is partly to avoid the impression that the amplification factors can be used for local analysis, as they are only to be used for rough regional hazard estimation. In this way, we also obtain the desired slowly varying run-up as a function of the alongshore coastal coordinate.
- ✓ At each hazard point, about 40 depth profiles are extracted more or less normal to the shoreline, each with a distance of 1 km apart (see Figure 8). In some areas the profiles are extracted manually, but most of the profiles are produced automatically from linear interpolation over SRTM bathymetric maps in ArcGIS, see Figure 9.
- ✓ The amplification factors are computed along seven arbitrarily selected profiles (from the 40 profiles) by using a 1HD linear shallow water wave model (LSW model). An initial wave of 1 m height in deep water, shaped as a single period sinusoidal shaped wave (N-wave) is fed over the boundary of the model.
- ✓ For each profile, the following parameters are varied:
  - The polarity of the incident wave, i.e. either a leading trough or leading peak.
  - The wave periods are 120, 300, 600, 1000, 1800, 3600 seconds respectively.
- ✓ For each simulation the maximal surface elevations at 50 m depth and at the shoreline (0 m depth) are extracted. The ratio of the height between the latter and the first is the amplification factor for the present profile.
- ✓ The median value of the seven amplification factors are extracted for each combination of wave period, polarity, and hazard point, see example in Figure 10. All combinations are stored in lookup-tables. As an example, amplification factors for a leading trough polarity and wave period of 600 s for the Mediterranean and Black Sea are depicted in Figure 11.
- ✓ Some hazard points that were accidentally left out without factors in the automatic selection procedure (for various reasons) are set equal to closest neighbouring factor.
- ✓ Two versions of the factors were produced, one set using the raw data values, and one set of factors smoothed along the shoreline with a median filter, to avoid artificially short amplification fluctuations along the shoreline, see example for the Black Sea in Figure 12.

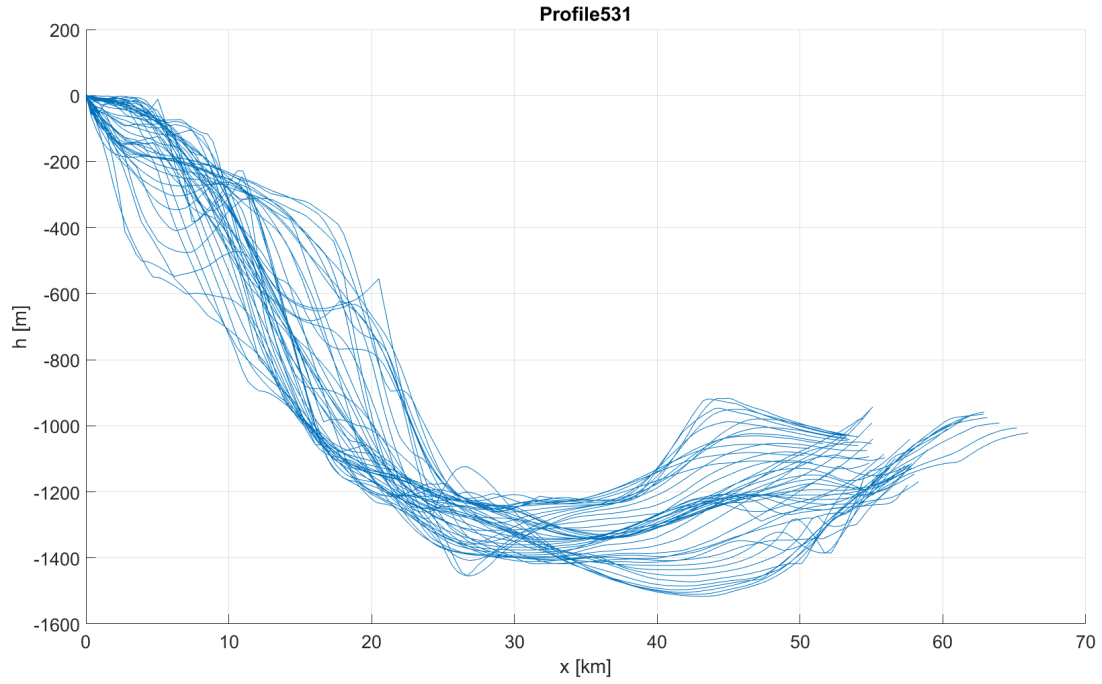



***Figure 8**: Example of a set of 40 depth profiles representing a hazard point.*

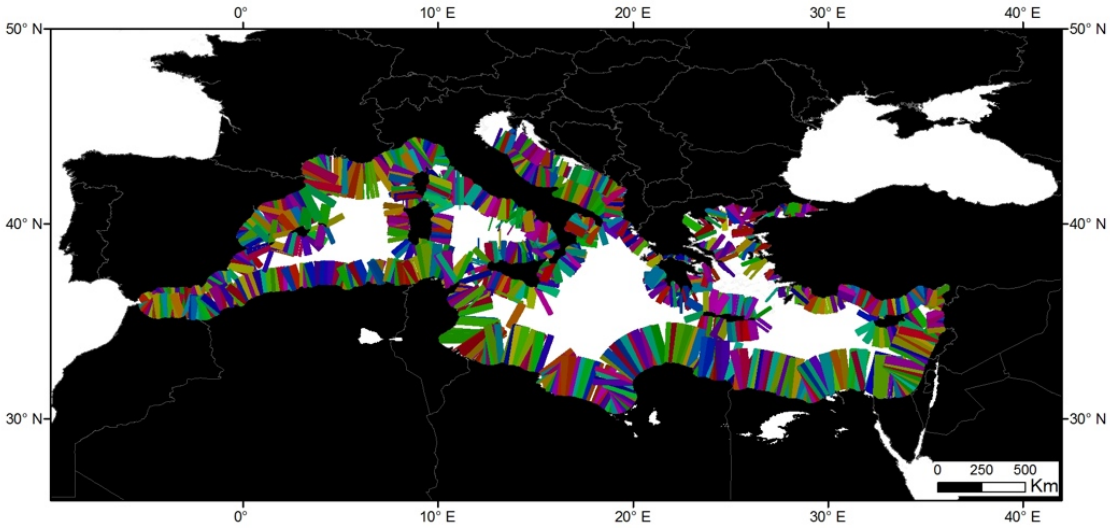



***Figure 9**: Example of a subset of profiles for the Mediterranean.*

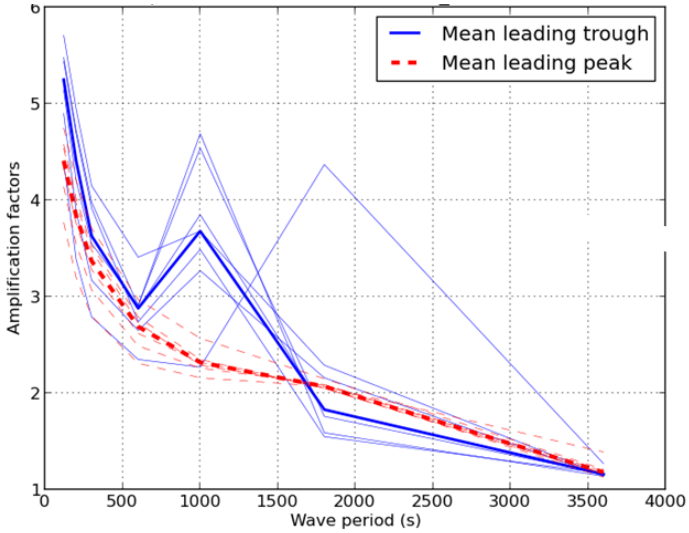



***Figure 10**: The amplification factors for the seven depth profiles related to a given location in the Mediterranean versus the wave period (seconds). The red and blue curves are factors for leading peak and leading trough, respectively. The thin curves are the factors for the seven local profiles, while the thick lines are the median values for leading peak and leading trough at this location.*

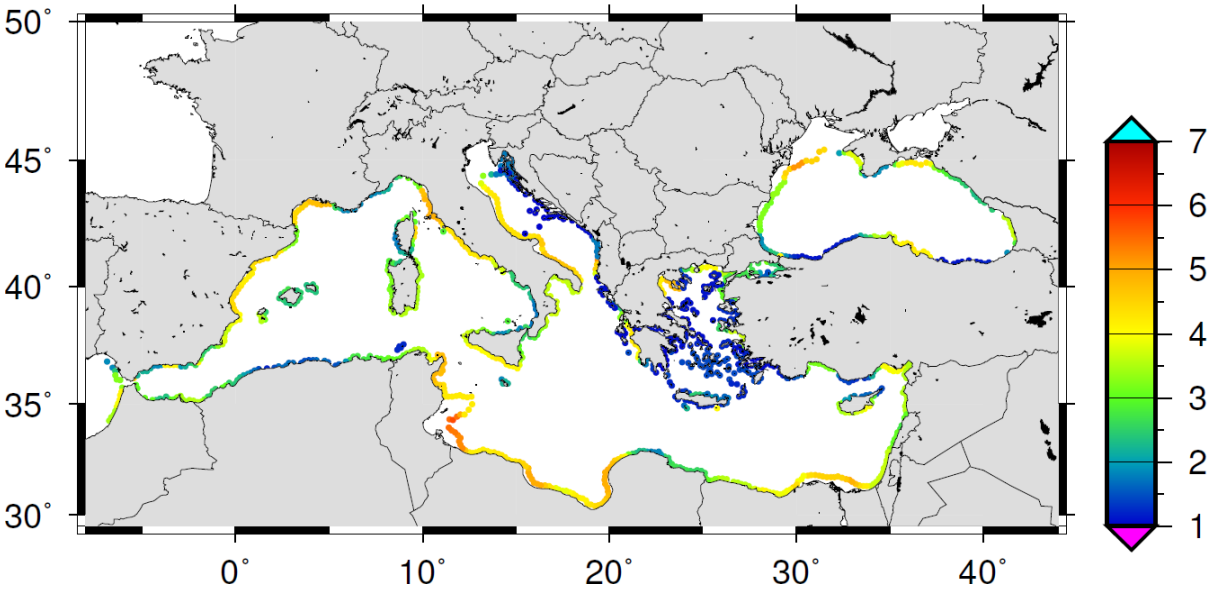



***Figure 11:** The amplification factors for the Mediterranean and Black Sea for the case with a leading trough and a wave period of 600 s. The factors are filtered along the shoreline with a median filter as explained in the text.*

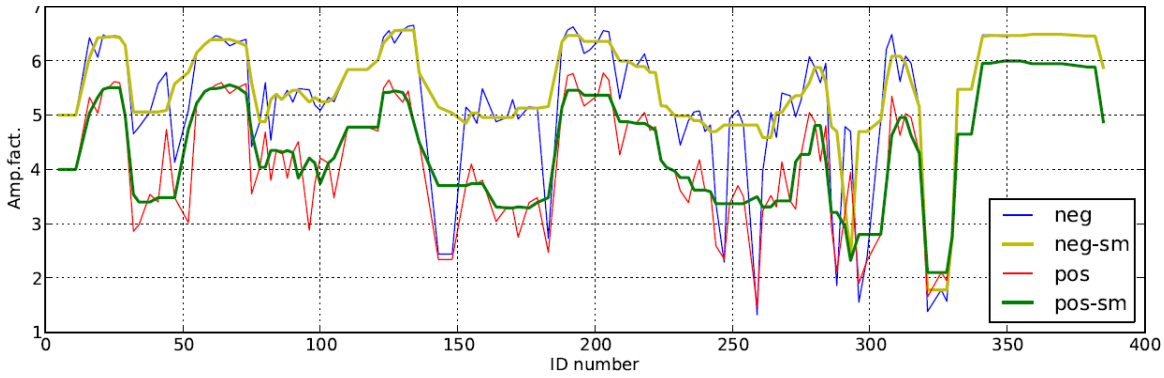



***Figure 12:*** *The amplification factors for Black Sea as function of the hazard points lying along the shoreline (ID numbers). The figure shows the effect of filtering the factors along the shoreline. The labels "neg" and "pos" relate to leading trough and leading peak, respectively. The tag "-sm" means the filtered values (median filter).*

### 3.2.2 Set up of local inundation simulations

The comparisons between the amplification factors and the NLSW models are undertaken at the three ASTARTE test site locations Sines, Colonia Sant Jordi, and Heraklion. For each location, we use 3-4 earthquake sources with varying magnitude. We have employed a total of 11 different sources, 9 with magnitudes $M_w$7.0, $M_w$7.5, and $M_w$8.0 (earthquakes with this source magnitude is employed for all locations). In addition we have applied two sources with larger magnitudes, a $M_w$8.5 earthquake at Sines and a $M_w$9.0 at the Hellenic Arc (impacting Heraklion), respectively.

Different inundation models are employed for the different locations. The tsunami propagation and local inundation simulations for Sines run-up are modelled with the NSWING model (e.g. Wronna et al., 2015). For Colonia Sant Jordi we use both HySEA (see e.g. de la Asunción, 2013) as well as a model combination of GloBouss and ComMIT (see Pedersen and Løvholt, 2008; Titov and Gonzales, 1997). The combination of GloBouss and ComMIT is also used for Heraklion. In both cases, GloBouss is used for the propagation stage and to produce the input to the inundation model ComMIT (propagation files), see Løvholt et al. (2010). All offshore tsunami simulations are conducted on regular grids with a resolution of 30 arcsec. The NLSW models used nested grids for simulating the local inundation. In the present investigations, the resolution of the finest grid is about 10 m at all locations. The Manning-friction (n) is in all simulations set to n=0.03.

For each test site and source scenario, we estimate the deterministic MIH by using the new amplification factor method as described above (Section 3.2.1). The wave characteristics needed to determine the amplification factors are extracted from the simulated offshore surface elevations at the hazard point placed on the 50 m contour outside each test site location. Surface elevations arriving more than 2 hours after the first wave arrival were

neglected. The MIH is then quantified by multiplying the maximum surface elevation at the hazard point by the amplification factors from the lookup-tables. We interpret this value as the median tsunami MIH.

The characteristic wave period used to derive the amplification factor is extracted automatically by a low-pass filter method. The filtering was conducted to remove spurious short oscillations that could lead to too small wave period estimates. The details of the low pass filter method are not provided here. Instead, we provide an example of its output. In Figure 13, we show an example of the simulated (black curve) and a filtered (red curve) time series of a tsunami offshore Colonia Sant Jordi emerging from a $M_w$ 8.0 earthquake. The simulated time series is produced by the GloBouss model. From the filtered time series, the location of the highest peak within the time frame of 2 hours are determined, see vertical magenta line. The maximum surface elevation value is taken at the unfiltered maximum location (marked with a magenta bullet). The wave period is the duration between the troughs ahead and behind the leading peak measured on the filtered time series, and the wave period interval is indicated by cyan coloured bars. We stress that the filtering algorithm is only used to derive wave periods, and not the simulated offshore wave amplitudes.

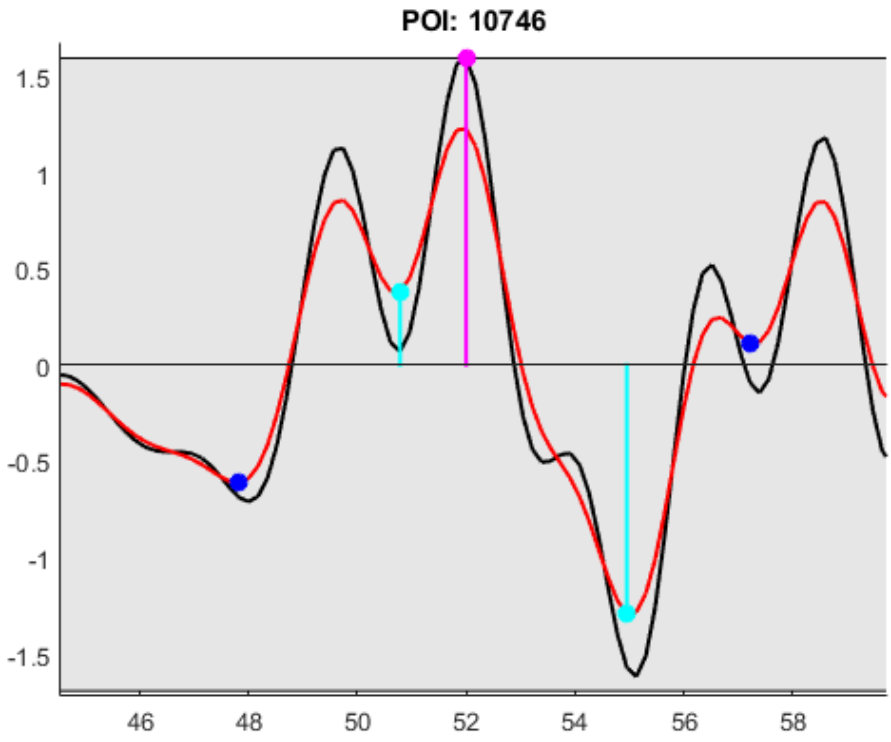



***Figure 13:*** *Example of extracting the wave characteristics from the hazard point on the 30 arcsec simulations (black line) at a hazard point with ID 10746. The vertical axis is the surface elevation in meters, while the horizontal axis is the time in minutes. The maximum*

*value is marked with a magenta bullet, while the wave period is measured between the two cyan vertical lines.*

### 3.2.3 Estimating the maximum inundation height uncertainty

For each scenario, we extract the maximum inundation height from the local inundation simulations for the finest grid. We then make MIH distributions by extracting maxima either in North-South direction or East-West direction line by line. I.e., if the shoreline is mostly East-West, we search along lines oriented in the North-South direction, while if the shoreline is predominantly oriented more North-South we search along lines oriented in the East-West direction. In some special cases the height of the terrain landside of the shoreline is too high and too steep to be inundated. In these situations, we include values of maximum surface elevation for a small distance seaside from the shoreline (at least one cell away).

From this, we obtain a MIH probability density for each simulation (i.e. for each inundated location). Each such MIH probability density is then fitted to a lognormal probability density function (PDF) using standard PDF fitting procedures in Matlab. We note that the lognormal PDF implies a normal distribution of the natural logarithm of the random variable with variance $\sigma^2$, in this case the MIH. The lognormal distribution of the MIH reads:

$$p(MIH) = \frac{1}{\sqrt{2\pi} \cdot MIH \cdot \sigma} e^{-\frac{(ln(MIH)-\mu)^2}{2\sigma^2}}$$

We note that the MIH median and mean values for the local probability distributions are $e^{\mu}$ and $e^{\mu+\sigma^2/2}$, respectively. Both the fitted mean and median MIH's are compared with the results from the amplification factors, in order to quantify biases for the individual simulations. To this end, two different expressions for estimating the biases are used; first a direct comparison between the logarithms of the MIH from the amplification factor with the normal mean from the fitted distribution

$$\epsilon_1 = \ln(MIH) - \mu$$

and second a normalized error measure

$$\epsilon_2 = \frac{e^{\mu} - MIH}{e^{\mu}}$$

### 3.2.4 Results

Various examples of the MIH distributions from NLSW simulations are shown in Figure 14-Figure 19. We note that in our present estimates of the overall bias, we do not take include the effect of the change in the topography due the earthquake displacement. The effect of change in topography affects the largest two events ($M_w$8.5 Sines and $M_w$9.0 Heraklion). For these two events, we show results both with and without the topographic correction.

Without conducting a formal test for quantifying the suitability of the lognormal fit, visual inspection of the different individual simulations reveals that more than 50% of the NLSW simulations exhibit lognormal characteristics. This is particularly the case for the higher magnitude simulations for Heraklion and Colonia Sant Jordi. The $M_w$7.0 events have a tendency to be more irregular. Moreover, the MIH's for the Sines test site seem to separate into different bins, being more complex than a simple lognormal. This may be due to the particular characteristics of this test site combined with the source.

For the Heraklion test site, the amplification factor method seems to overestimate the run-up in most cases, and except for the atypical subduction zone event, we find a logarithmic bias of $\epsilon_1$~0.2. The uncertainty σ seems to be largest for the smallest magnitudes (0.36-0.47), being clearly smaller for the largest ones (0.17). For the subduction zone event, we see that not taking into account the topography changes introduces a large bias (Figure 15).

For the Colonia Sant Jordi test site we find the largest biases compared to the ComMIT model, whereas HySEA provides MIH's close to or slightly smaller than the amplification factor method. For the largest magnitudes, the biases are relatively small, $\epsilon_1$ ranging from -0.09 to 0.23. Uncertainties typically range from 0.2 to 0.3. We also see that HySea provides smaller MIH's than ComMIT. We remark that both of these codes are thoroughly benchmarked towards standard benchmark tests (Liu et al., 2008).

The Sines test site differs from the others by its more complicated distributions, and the fact that the amplification factors seem to systematically underestimate the MIH. The negative bias shows the biggest deviations for the smallest moment magnitude, which is again supposedly due to the characteristics of the site. The MIH uncertainty is otherwise quite comparable to the other test sites. For the largest ($M_w$8.5) event close to Sines, we see that not taking into account the topography changes introduces a large bias (Figure 15).

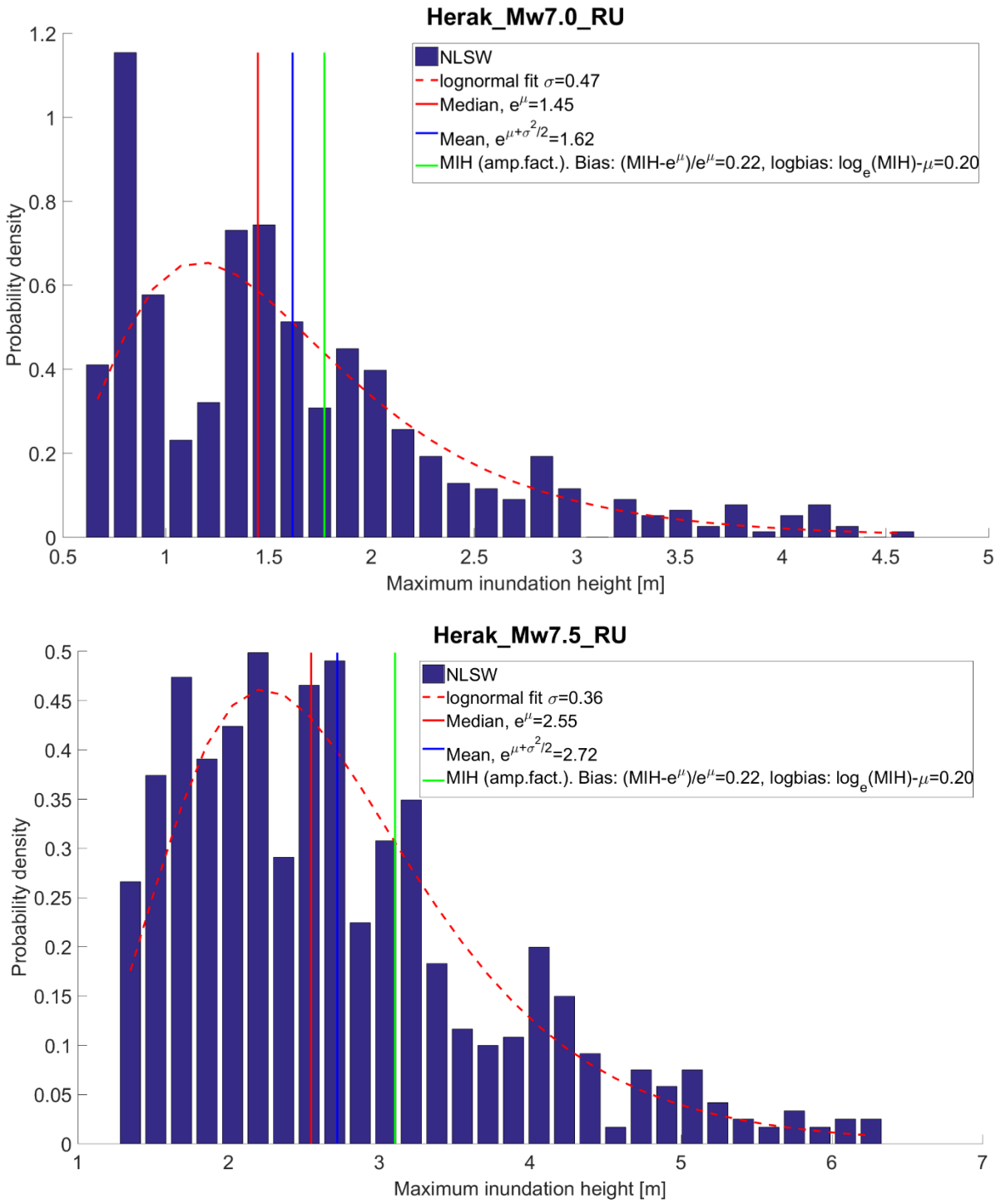



***Figure 14**: Distribution of MIH and fitted lognormal distribution for the Heraklion test site using the ComMIT model. Upper panel, $M_w$7 scenario; lower panel, $M_w$7.5 scenario.*

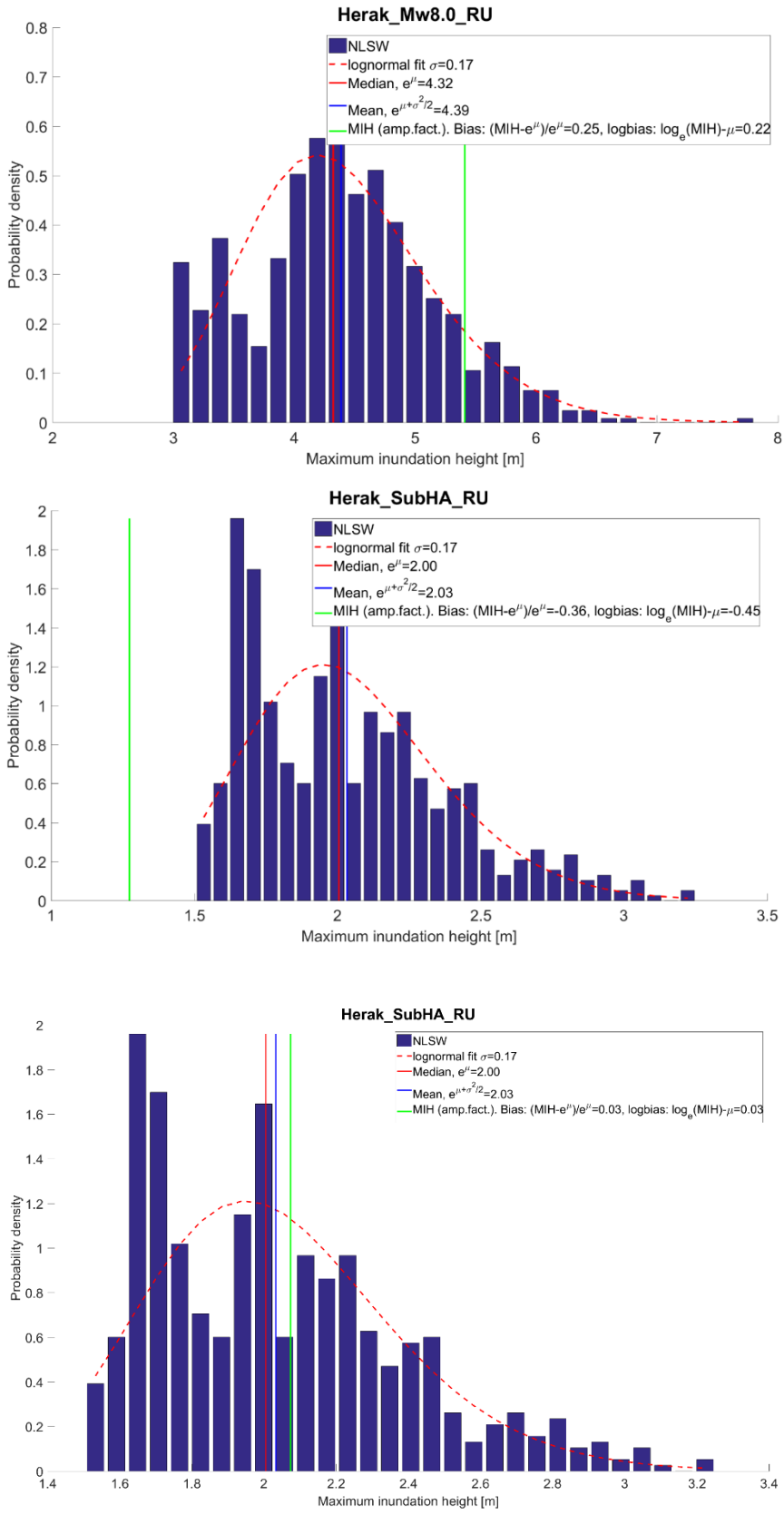
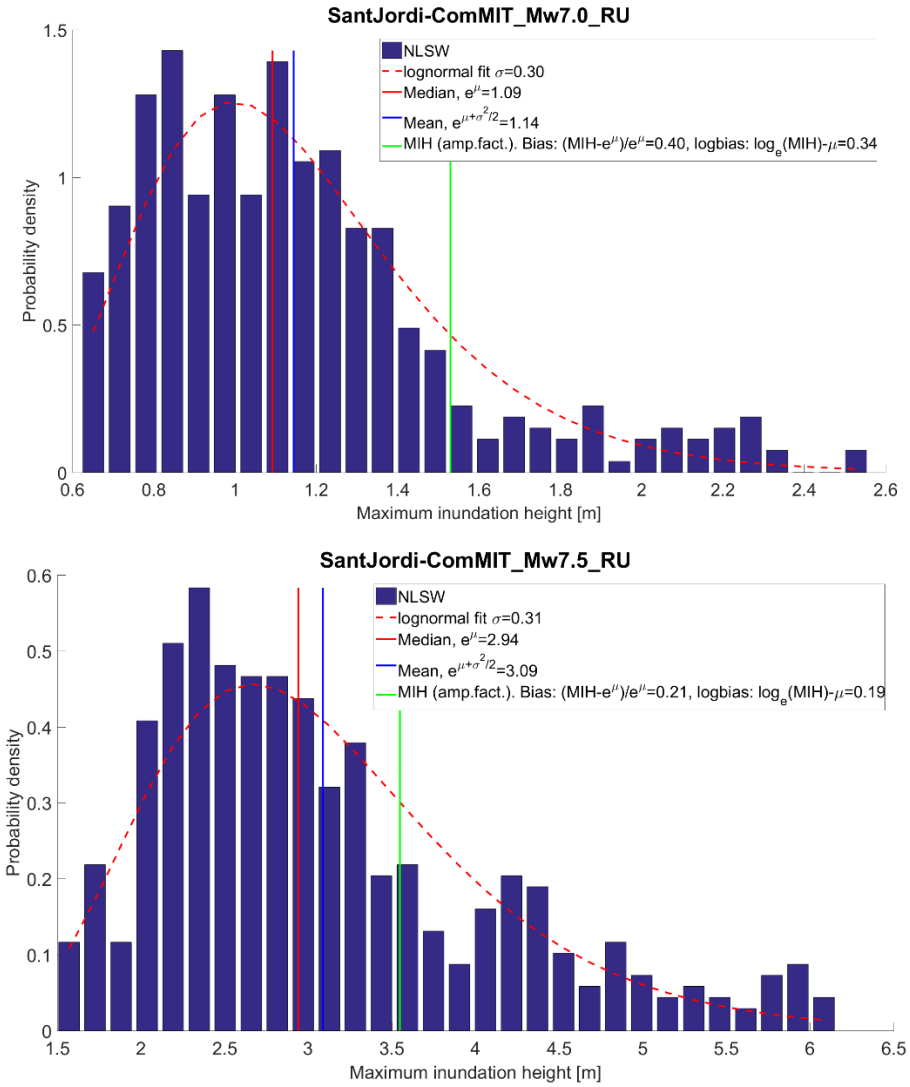



***Figure 15**: Distribution of MIH and fitted lognormal distribution for the Heraklion test site using the ComMIT model. Upper panel, $M_w$8 scenario; mid panel, megathrust subduction*

*scenario from the Hellenic Arc (not corrected for topographic change), lower panel, megathrust subduction scenario from the Hellenic Arc (corrected for topographic change).*


SantJordi-ComMIT_Mw7.0_RU
NLSW
lognormal fit σ=0.30
Median, e^μ=1.09
Mean, e^(μ+σ^2/2)=1.14
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=0.40, logbias: log_e(MIH)-μ=0.34
Probability density
Maximum inundation height [m]
SantJordi-ComMIT_Mw7.5_RU
NLSW
lognormal fit σ=0.31
Median, e^μ=2.94
Mean, e^(μ+σ^2/2)=3.09
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=0.21, logbias: log_e(MIH)-μ=0.19
Probability density
Maximum inundation height [m]

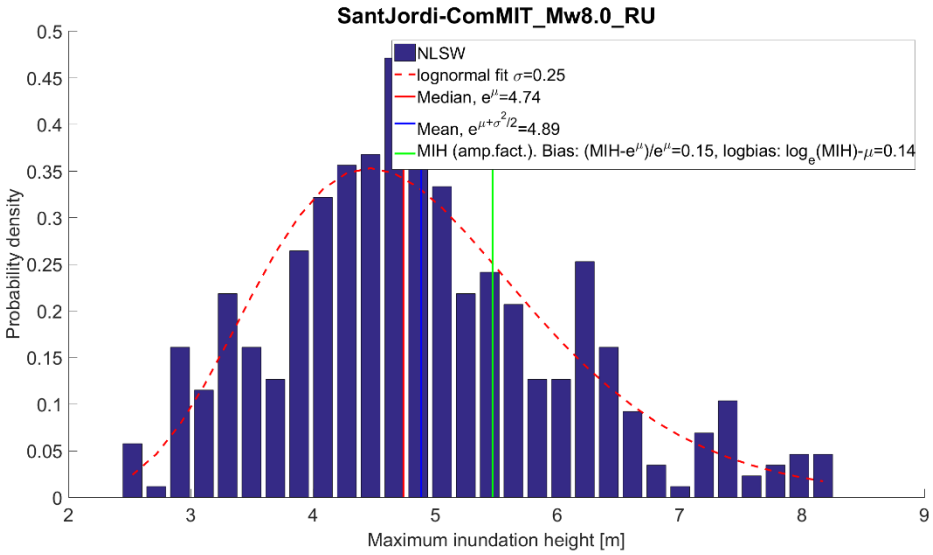
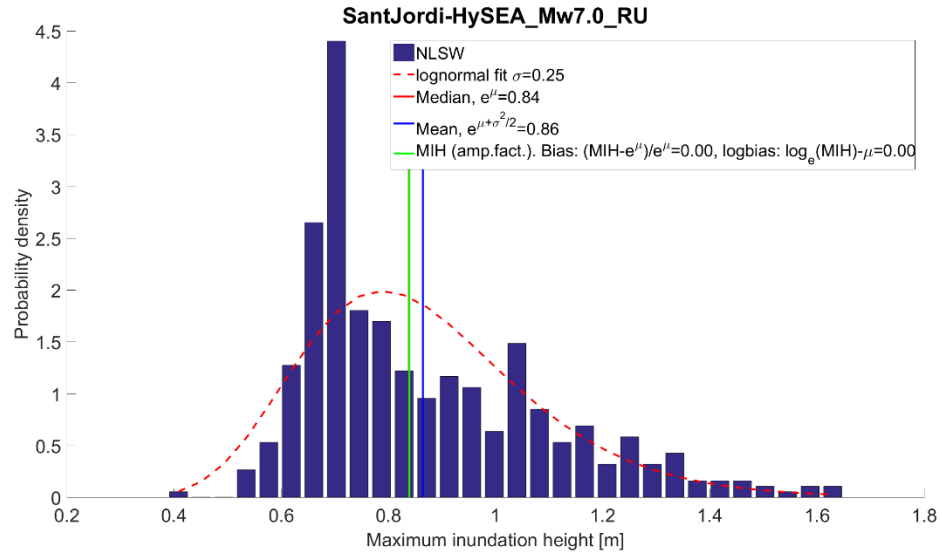
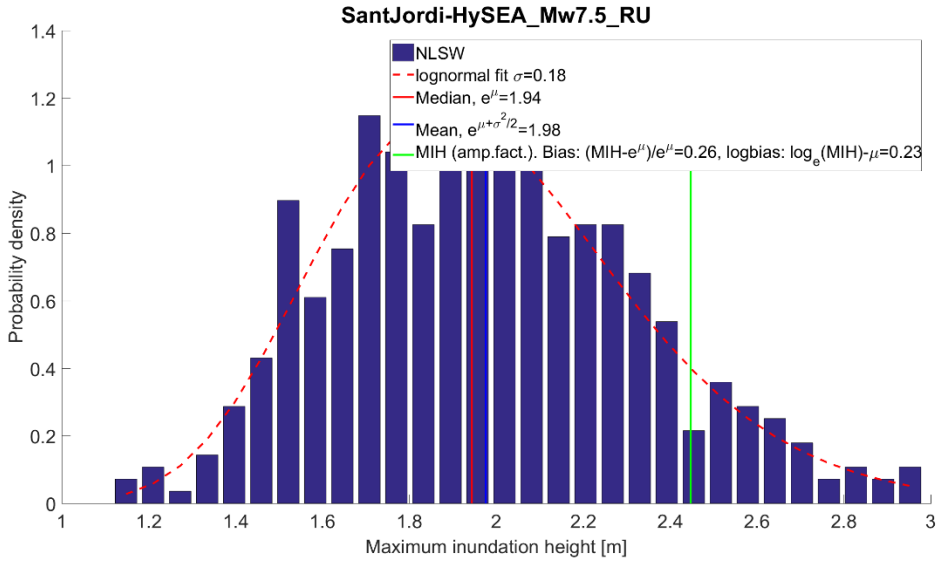



***Figure 16**: Distribution of MIH and fitted lognormal distribution for the Colonia Sant Jordi test site using the ComMIT model. Upper panel, $M_w$7 scenario; mid panel, $M_w$7.5 scenario; lower panel, $M_w$8.0 scenario.*


SantJordi-HySEA_Mw7.0_RU
NLSW
lognormal fit σ=0.25
Median, e^μ=0.84
Mean, e^(μ+σ^2/2)=0.86
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=0.00, logbias: log_e(MIH)-μ=0.00
Probability density
Maximum inundation height [m]


SantJordi-HySEA_Mw7.5_RU
NLSW
lognormal fit σ=0.18
Median, e^μ=1.94
Mean, e^(μ+σ^2/2)=1.98
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=0.26, logbias: log_e(MIH)-μ=0.23
Probability density
Maximum inundation height [m]

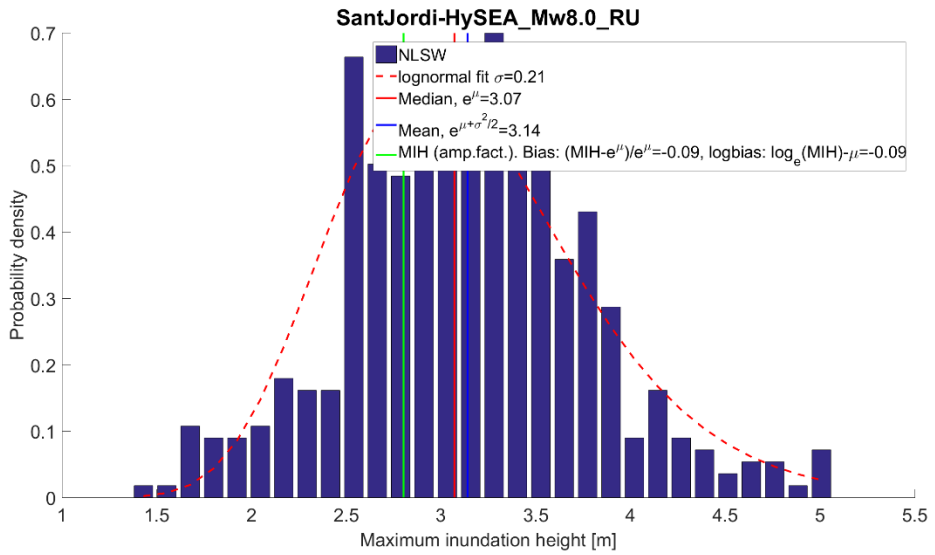
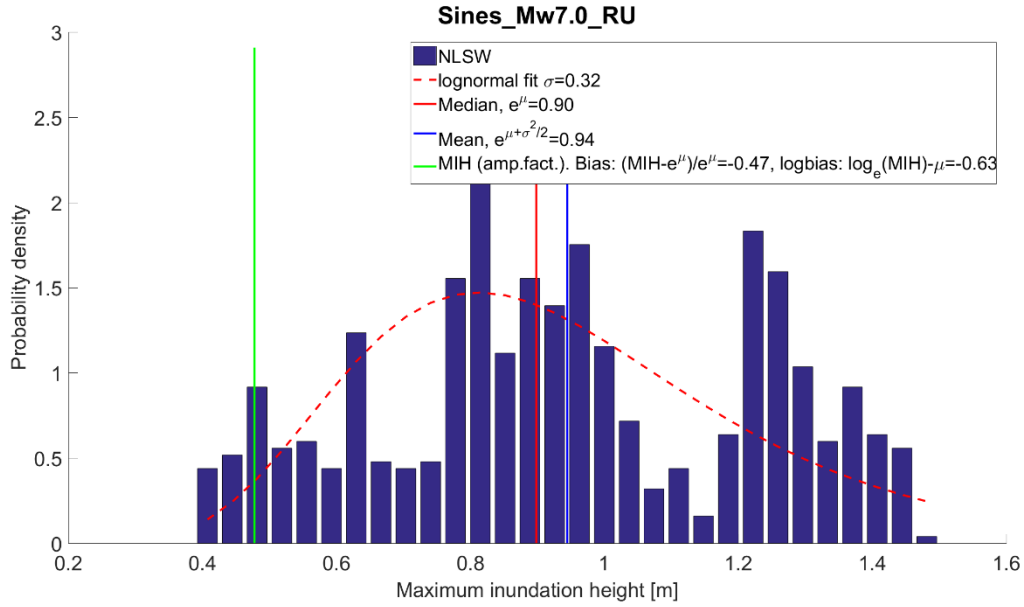

SantJordi-HySEA_Mw8.0_RU
NLSW
lognormal fit σ=0.21
Median, e^μ=3.07
Mean, e^(μ+σ^2/2)=3.14
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=-0.09, logbias: log_e(MIH)-μ=-0.09
Probability density
Maximum inundation height [m]

***Figure 17**: Distribution of MIH and fitted lognormal distribution for the Colonia Sant Jordi test site using the HySEA model. Upper panel, $M_w$7 scenario; mid panel, $M_w$7.5 scenario; lower panel, $M_w$8.0 scenario.*



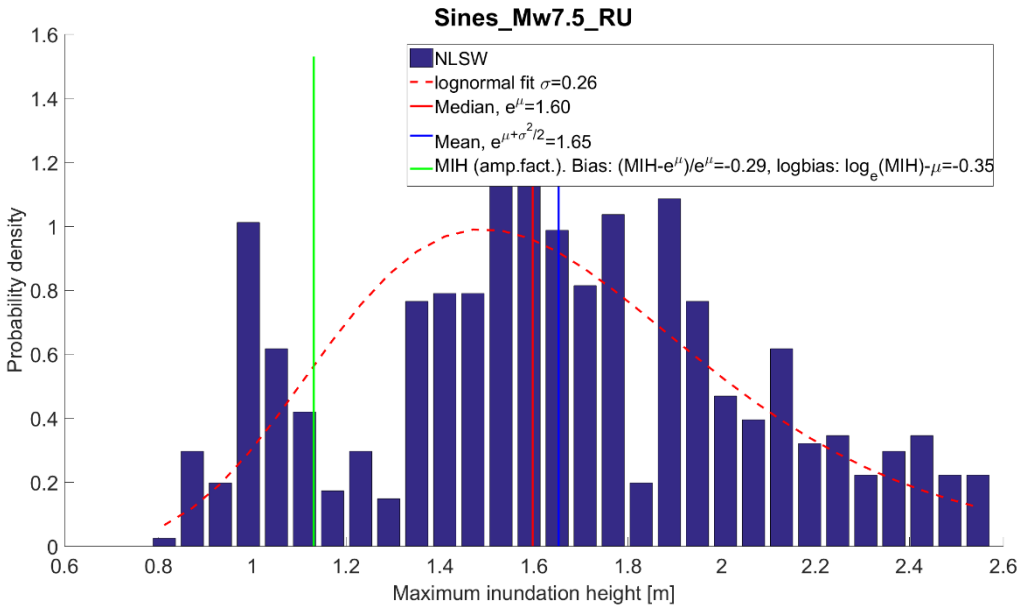
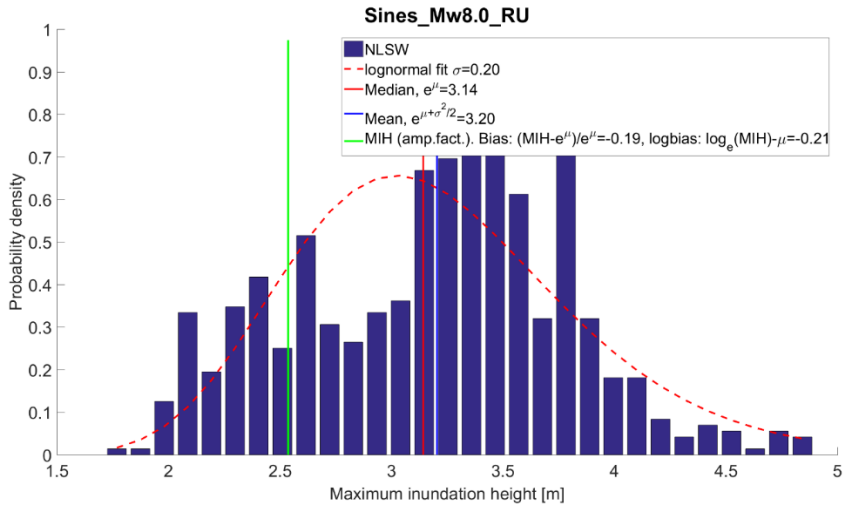
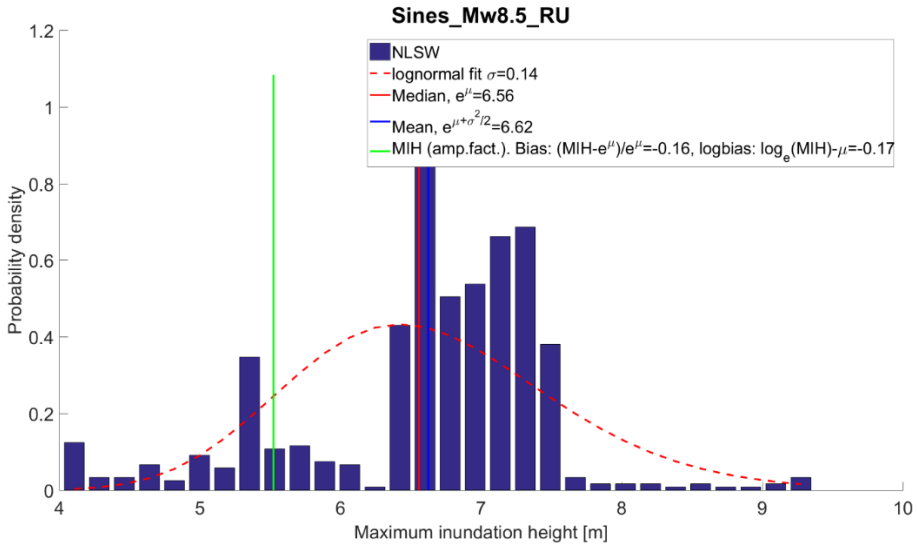




***Figure 18**: Distribution of MIH and fitted lognormal distribution for the Sines test site using the Comcot model. Upper panel, $M_w$7 scenario; lower panel, $M_w$7.5 scenario.*


Sines_Mw8.0_RU
NLSW
lognormal fit σ=0.20
Median, e^μ=3.14
Mean, e^(μ+σ^2/2)=3.20
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=-0.19, logbias: log_e(MIH)-μ=-0.21
Probability density
Maximum inundation height [m]


Sines_Mw8.5_RU
NLSW
lognormal fit σ=0.14
Median, e^μ=6.56
Mean, e^(μ+σ^2/2)=6.62
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=-0.16, logbias: log_e(MIH)-μ=-0.17
Probability density
Maximum inundation height [m]

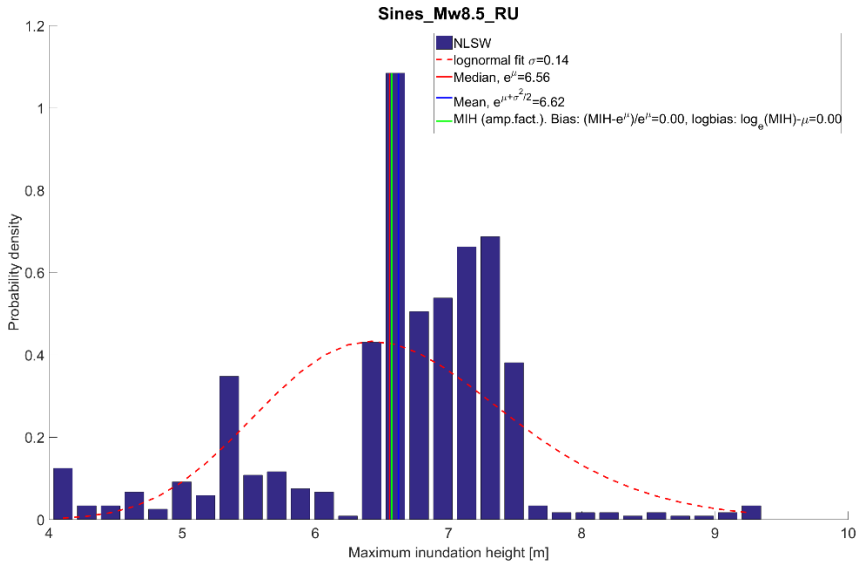

Sines_Mw8.5_RU
NLSW
lognormal fit σ=0.14
Median, e^μ=6.56
Mean, e^(μ+σ^2/2)=6.62
MIH (amp.fact.). Bias: (MIH-e^μ)/e^μ=0.00, logbias: log_e(MIH)-μ=0.00
Probability density
Maximum inundation height [m]

*Figure 19: Distribution of MIH and fitted lognormal distribution for the Sines test site using the Comcot model. Upper panel, $M_w$8.0 scenario; mid panel, $M_w$8.5 scenario (not corrected for topographic change), lower panel, $M_w$8.5 scenario (corrected for topographic change).*

### 3.2.5 Summary and discussion

Both positive and negative amplification factor biases are found for the various simulations. The mean bias is however close to zero (Figure 20). It is noted that the investigation is based on just a few locations, and ideally, many more sites should be investigated to provide a broader set of data for a more reliable distribution, particularly for the bias. On one hand, the present investigation does not hold evidence for the need for correcting the amplification factor method to a large extent. On the other hand, there is a need take into account uncertainty of the median value of the MIH due to the amplification factor method (reflected by the bias, $\sigma_{bias}$=0.3). The mean lognormal σ for all the simulations is 0.26 (Figure 21). Altogether, this local variability (lognormal σ) needs to be added to the variability of the mean value for each simulation. If we assume that the two are independent, a simple strategy can be to add their variances, giving a total uncertainty $\sigma_{tot}$~0.4. This uncertainty should be added to the source uncertainty due to variable slip, fault orientation, variability of source rigidity, etc. For example, Geist (2002) reports of coefficients of variations in the range of 0.25-0.35 for variable slip uncertainty. A total uncertainty would need to incorporate this in addition to the uncertainties found here for the inundation alone.

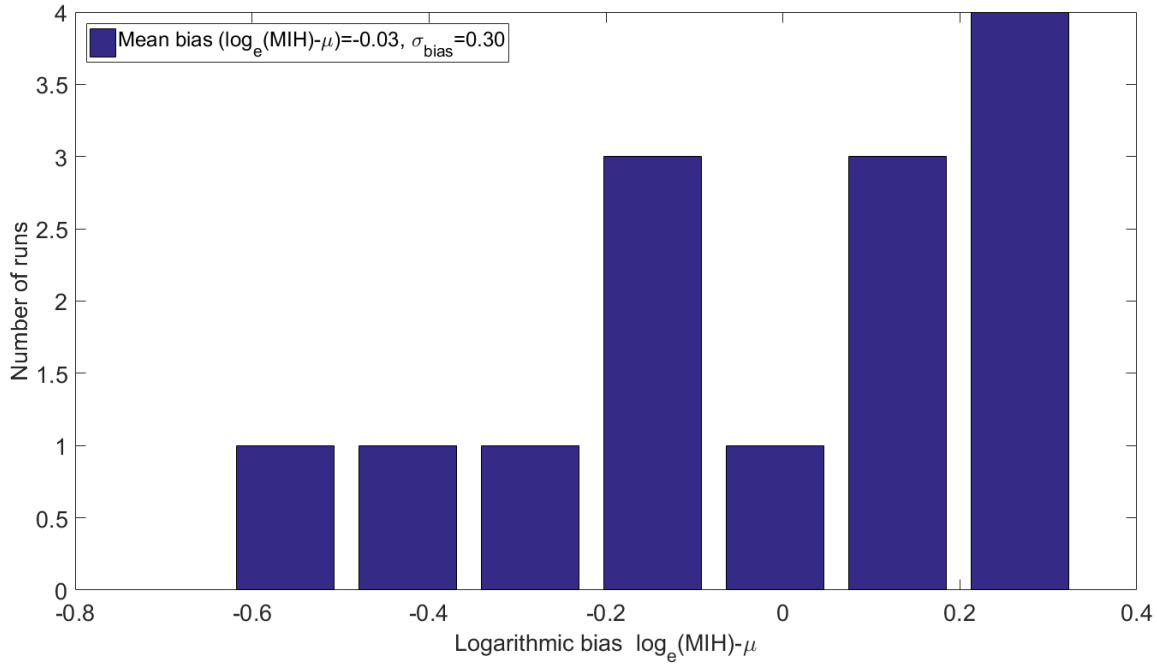



***Figure 20**: Histogram showing the lognormal bias for the different simulations investigated. The mean bias is close to zero. Here, the topography change is not taken into account. If we take into account the topography change, the bias will increase from -0.03 to +0.02.*

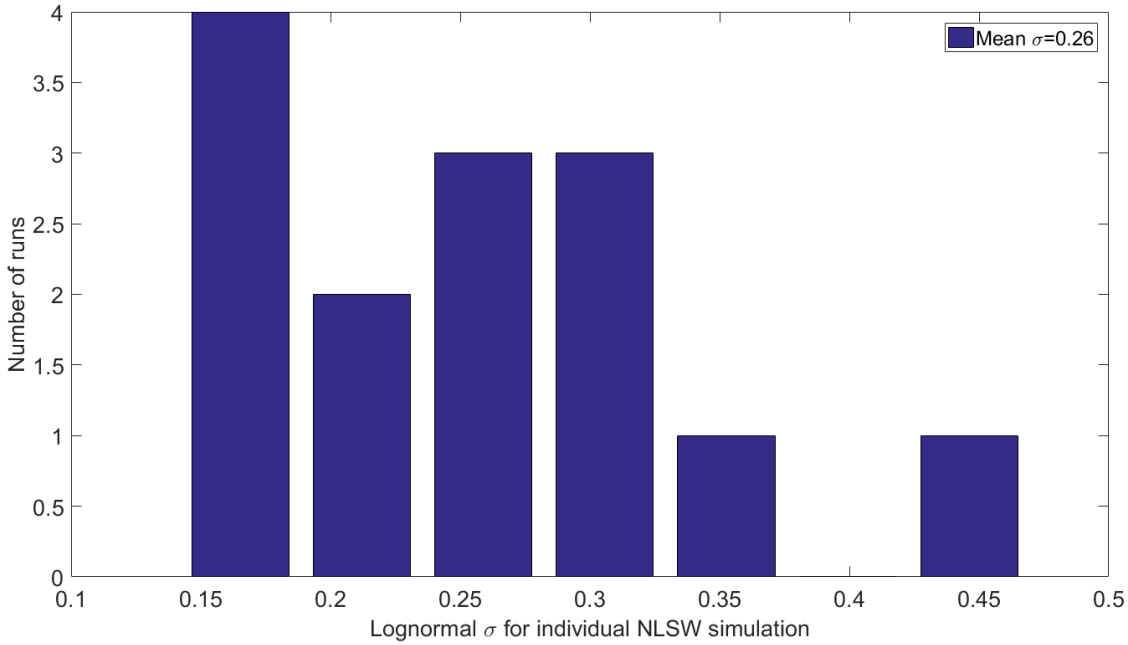



***Figure 21**: Histogram showing the uncertainty distribution for the different simulations investigated.*

## 4. Conclusions and outlook

Uncertainty quantification is the key for comparing hazard and risk analyses with past data, and to scientifically test and eventually falsify them (*Section 1*). This comparison is the only possible strategy to identify ontological errors.

Among the available methods for hazard and risk quantification, only probabilistic models are focused to the quantification of uncertainty (*Section 2*). Scenario-based approaches do not quantify uncertainty, but they may provide important clues to feed decision making as well as probabilistic models. The definition of reference return periods for scenario-based analyses provides a tool to compare scenario-based and probabilistic hazard and risk quantifications. By means of this comparison, scenario-based and probabilistic methods may feed to each other.

Probabilistic approaches target to quantify all sources of uncertainty (*Section 2*). On the one side, they should explore the effective variability of any potential tsunamis source, irrespective of the occurrence of tsunamis from that source in the (known) past. On the other side, they test the impact on the results of alternative and scientifically acceptable computational methodologies. This leads to increase the methodological complexity and the costs of the efforts toward the quantification of hazard and risk. In particular, it requires the establishment of standards for the analysis of source variability and of protocols for the management of unavoidable subjectivity.

Within the tsunami community, fully probabilistic approaches are still in development. In particular, probabilistic tsunami hazard approaches are under development mainly for seismic sources (S-PTHA). Few probabilistic models are available for vulnerability and risk quantification, mainly based on empirical approaches. An S-PTHA model is being produced for the NEAM region (http://www.tsumaps-neam.eu/). A global effort is ongoing for extending the available methodologies and defining common standards and good practices, in the framework of the Global Tsunami Model (GTM, www.globaltsunamimodel.org). In this respect, this document collected and organized several ongoing efforts to establish a first step in this direction (*Section 3*). This effort leads also toward a standardization of tsunami practices in line with the best practices adopted for other (natural) hazards, in multi-hazard risk perspective.

Hazard and risk assessments based on probabilistic approaches produce a large amount of information that can increase the awareness of decision making over these results, for local planners, stake holders and even for the general public. However, the complexity of these results may be an obstacle to this process. Therefore, scientists should make an effort to improve the communication of uncertainty, by organizing appropriate outreach actions and by establishing tools and procedures for assessing the results and distilling the important information included in them.

## References

Acar S (2015). Tsunami Hazard Analysis for Gulluk Bay, MS Thesis in METU Department of Civil Engineering, Ocean Engineering Research Center, Ankara, Turkey, (132 pages)

Akkar S, Sucuoglu H, Yakut A (2005). Displacement-based fragility functions for low- and mid-rise ordinary concrete buildings. Earthquake Spectra, 21(4), 901–927.

ALA - American Lifeline Alliance (2001). Seismic fragility formulations for water system, American Society of Civil Engineers (ASCE) and Federal Emergency Management Agency (FEMA). Available at http://www.americanlifelinesalliance.com/pdf/Part_1_Guideline.pdf. Accessed Mar 2017

Allen CR (1995). Earthquake hazard assessment: Has our approach been modified in the light of recent earthquakes?, Earthquake Spectra, 11, 357–366.

Araújo MB, New M (2007). Ensemble forecasting of species distributions, Trends in Ecology & Evolution, Volume 22, Issue 1, January 2007, Pages 42-47, ISSN 0169-5347, http://dx.doi.org/10.1016/j.tree.2006.09.010.

Argyroudis S, Selva J, Gehl P, Pitilakis K (2015). Systemic seismic risk assessment of road networks in urban areas, Computer-Aided Civil and Infrastructure Engineering, 30, 524-540, DOI: 10.1111/mice.12136

Aspinall WP, Cooke RM (2013). Quantifying scientific uncertainty from expert judgement elicitation, in "Risk and Uncertainty Assessment for Natural Hazards" (Eds J. Rougier, L. Hill, R.S.J. Sparks), Cambridge University Press, Cambridge, UK, ISBN 978-1-107-00619-5

ATC-13, Applied Technology Council Report (1985) Earthquake damage evaluation data for California. Palo Alto, California

ATC-25 - Applied Technology Council Report (1991) Seismic vulnerability and impact of disruption of lifelines in the conterminous United States. Redwood City, California

Azevedo J, Guerreiro L, Bento R, Lopes M, Proença J (2010). Seismic vulnerability of lifelines in the greater Lisbon area. Bull Earthq Eng 8:157–180

Basöz NI, Kiremidjian AS, King SA, Law KH (1999). Statistical analysis of bridge damage data from the 1994 Northridge, CA, earthquake. Earthq Spectr 15(1):25–54

Bayraktar M, Hastak M (2009). Bayesian Belief Network Model for Decision Making in Highway Maintenance: Case Studies, J. Constr. Eng. Manage., 10.1061/(ASCE)CO.1943-7862.0000111, 1357-1369

Bazzurro P, Cornell C (1999). Disaggregation of seismic hazard, Bull. Seismol. Soc. Am. 89, no. 2, 501–520.

Bedford T, Cooke R (2001). Probabilistic risk analysis: foundations and methods, Cambridge University Press.

Bernreuter DL, Savy JB, Mensing RW, Chen JC (1989). Seismic Hazard Characterization of 69 Nuclear Plant Sites East of the Rocky Mountains, Vols. 1-8, prepared for Division of Engineering and System Technology, Office of Nuclear Reactor Regulation, U.S. Nuclear Regulatory Commission, Washington, NUREG/CR-5250 UCID-21517.

Bommer JJ (2012). Challenges of Building Logic Trees for Probabilistic Seismic Hazard Analysis, EARTHQUAKE SPECTRA, Vol: 28, Pages: 1723-1735, ISSN: 8755-2930

Bommer JJ, Scherbaum F (2008). The Use and Misuse of Logic Trees in Probabilistic Seismic Hazard Analysis, EARTHQUAKE SPECTRA, Vol: 24, Pages: 997-1009, ISSN: 8755-2930

Bovolo A, Abele S, Bathurst J, Caballero D, Ciglan M, Eftichidi G, Simo B (2009). A distributed framework for multi-risk assessment of natural hazards used to model the effects of forest fire on hydrology and sediment yield. Comput Geosci 35:924–945. doi:10.1016/j.cageo.2007.10.010

Budnitz RJ, Apostolakis G, Boore DM, Cluff LS, Coppersmith KJ, Cornell CA, Morris PA (1998). Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis, Risk Analysis, Vol. 18 (4), 463-469.

Burbidge, D., Cummins, P., Mleczko, R. & Thio, H., 2008. A Probabilistic tsunami hazard assessment for Western Australia, Pure appl. Geophys., 165, 2059–2088.

Chock G, Yu G, Thio HK, Lynett P (2016). Target Structural Reliability Analysis for Tsunami Hydrodynamic Loads of the ASCE 7 Standard, Journal of Structural Engineering, 142, DOI: http://dx.doi.org/10.1061/(ASCE)ST.1943-541X.0001499

Cloke HL, Pappenberger F (2009). Ensemble flood forecasting: A review, Journal of Hydrology, Volume 375, Issues 3–4, 15 September 2009, Pages 613-626, ISSN 0022-1694, http://dx.doi.org/10.1016/j.jhydrol.2009.06.005.

Cooke R (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford Univ. Press, New York

Cornell C (1968). Engineering seismic risk analysis. Bull Seismol Soc Am 58:1583–1606

Cornell C, Krawinkle H (2000). Progress and challanges in seismic performance assessment. PEER Center News 3(2). http://peer.berkeley.edu/news/2000spring/index.html

Costa A, Dell'Erba F, Di Vito MA, Isaia R, Macedonio G, Orsi G, Pfeiffer T (2009). Tephra fallout hazard assessment at the Campi Flegrei caldera (Italy), Bull. Volcanol., 71, 259–273, doi:10.1007/s00445-008-0220-3.

Cotton F, Pousse G, Bonilla F, Scherbaum F (2008). On the discrepancy of recent European ground-motion observations and predictions from empirical models: Analysis of KiK-net accelerometric data and point-sources stochastic simulations. Bulletin of the Seismological Society of America, 98(5):22442261, Oct 2008. doi: 10.1785/0120060084

Dall'Osso F, Gonella M, Gabbianelli G, Withycombe G, Dominey-howes D (2009). A revised (PTVA) model for assessing the vulnerability of buildings to tsunami damage. Nat. Hazards Earth Syst. Sci., 9, 1557–1565.

Davies G, Griffin J, Løvholt F, Glymsdal S, Harbitz C, Thio HK, Lorito S, Basili R, Selva J, Geist E, Baptista MA (2017). A global probabilistic tsunami hazard assessment from earthquake sources. From: Scourse, E. M., Chapman, N. A., Tappin, D. R. & Wallis, S. R. (eds) Tsunamis: Geology, Hazards and Risks. Geological Society, London, Special Publications, 456, https://doi.org/10.1144/SP456.5, updated version.

de la Asunción M, Castro MJ, Fernández-Nieto ED, Mantas JM, Ortega S, González-Vida JM (2013). Efficient GPU implementation of a two waves TVD-WAF method for the two-dimensional one layer shallow water system on structured meshes. Computers & Fluids, 80:441-452.

Der Kiureghian A (2005). Non-ergodicity and PEER's framework formula. Earthq Eng Struct Dyn 34:1643–1652. doi:10.1002/eqe.504

Douglas J (2007). Physical vulnerability modelling in natural hazard risk assessment. Nat Hazards Earth Syst Sci 7:283–288

EN 1050 (1997). Safety of machinery. Principles for risk assessment

Esposito S, Stojadinovic B, Babič A, Dolšek M, Iqbal S, Selva J (2017). Engineering Risk-Based Methodology For Stress Testing Of Critical Non-Nuclear Infrastructures (STREST Project), Paper N° 4247, 16 th World Conference on Earthquake, 16WCEE 2017.

FEMA - Federal Emergency Management Agency (2008). Estimated annualized earthquake losses for the United States. FEMA 366

Field EH, SCEC Phase III - Working Group (2000). Accounting for site effects in probabilistic seismic hazard analyses of southern California: Overview of the SCEC PhaseIII report, Bull. Seismol. Soc. Am. 90, no. 6B, S1–S31.

Franchin P, Cavalieri F (2013). Seismic vulnerability of a complex interconnected infrastructure, in S. Tesfamariam and K. Goda (eds.), Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems, Wood-head Publishing Ltd., Cambridge, UK, pp. 465–513.

Gardner JK, Knopoff L (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, Bull. Seis. Soc. Am., 64 (5), 1363–1367

Geist EL (2002). Complex earthquake rupture and local tsunamis, J. geo- phys. Res., 107(B5), doi:10.1029/2000JB000139

Geist EL (2009). Phenomenology of tsunamis: statistical properties from generation to runup. Adv Geophys 51:107–169

Geist EL (2012). Near-field tsunami edge waves and complex earthquake rupture. Pure Appl Geophys. doi:10.1007/s00024-012-0491-7

Geist E, Parsons T (2006). Probabilistic analysis of tsunami hazards, Nat. Hazards, 37, 277–314

Geist EL, Parsons T (2008). Distribution of tsunami interevent times, Geophys. Res. Lett., 35, L02612, doi:10.1029/2007GL032690

Geist EL, Lynett PJ (2014). Source processes for the probabilistic assessment of tsunami hazards, Oceanography, 27(2), 86–93

Geist EL, Parsons T (2014). Undersampling power-law size distributions: effect on the assessment of extreme natural hazards, Nat. Hazards, 72, 565–595.

Gelman A, Carlin J, Stern H, Rubin D (1995). Bayesian data analysis. Chapman and Hall/CRC, Boca Raton

Gonzalez FI, Geist EL, Jaffe B, Kanoglu U, Mofjeld H, Synolakis CE, Titov VV, Arcas D, Bellomo D, Carlton D, Horning T, Johnson J, Newman J, Parsons T, Peters R, Peterson C, Priest G, Venturato A, Weber J, Wong F, Yalciner A (2009). Probabilistic tsunami hazard assessment at Seaside, Oregon, for near- and far-field seismic sources. J.Geoph. Res. – Oceans, 114, C11023

Grezio A, Marzocchi W, Sandri L, Gasparini P (2010). A Bayesian procedure for Probabilistic Tsunami Hazard Assessment, Nat. Hazards 53, 159–174.

Grezio A, Sandri L, Marzocchi W, Argnani A, Gasparini P, Selva J (2012). Probabilistic Tsunami Hazard Assessment for Messina Strait Area (Sicily - Italy), Nat. Haz. 64:329-358, DOI: 10.1007/s11069-012-0246-x

Grüntal G, Thieken AH, Shwarz J, Radtke KS, Smolka A, Merz B (2006). Comparative risk assessments for the city of Cologne—storms, foods, earthquakes. Nat Hazards 28:21–44

Hanks TC, Abrahamson NA, Boore DM, Coppersmith KJ, Knepprath NE (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.

Hiemer S, Woessner J, Basili R, Danciu L, Giardini D, Wiemer S (2014). A smoothed stochastic earthquake rate model considering seismicity and fault moment release for Europe, Geophys. J. Int., 198, 1159–1172.

Hoffmann S, Fischbeck P, Krupnick A, McWilliams M (2007). Elicitation from Large, Heterogeneous Expert Panels: Using Multiple Uncertainty Measures to Characterize Information Quality for Decision Analysis, Decision Analysis 4(2), 91 – 109.

IAEA – Volcanoes (2012). Volcanic Hazards in Site Evaluation for Nuclear Installations Specific Safety Guide, IAEA Safety Standards Series No. SSG-21

IAEA – Volcanoes (2016). Volcanic Hazard Assessments for Nuclear Installations: Methods and Examples in Site Evaluation, TECDOC-1795.

Iervolino I, Giorgio M, Chioccarelli E (2016). Markovian modeling of seismic damage accumulation. Earthquake Engineering and Structural Dynamics. 45(3):441–461.

IPCC - Intergovernmental Panel on Climate Change (1998). PRINCIPLES GOVERNING IPCC WORK, Approved at the Fourteenth Session (Vienna, 1-3 October 1998) on 1 October 1998, amended at the Twenty-First Session (Vienna, 3 and 6-7 November 2003), the Twenty-Fifth Session (Mauritius, 26-28 April 2006), the Thirty-Fifth Session (Geneva, 6-9 June 2012) and the Thirty-Seventh Session (Batumi,

14-18 October 2013), available online at http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles.pdf (last checked, 10 Oct. 2016).

Jordan TH, Chen YT, Gasparini P, Madariaga R, Main I, Marzocchi W, Papadopoulos G, Sobolev G, Yamaoka K, Zschau J (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization, Ann. Geophys., 54, 315–391

Kalafat D, Kekovali K, Gunes Y, Yilmazer M, Kara M, Deniz P, Berberoglu M (2009). "A catalogue of source parameters of moderate and strong earthquakes for Turkey and its surrounding area (1938-2008)", KOERI, January 2009

Kappos A, Panagopoulos G, Panagiotopoulos Ch, Penelis G (2006). A hybrid method for the vulnerability assessment of R/C and URM buildings. Bull Earthq Eng 4:391–413

Kappos AJ, Panagopoulos G, Penelis G (2008). Development of a seismic damage and loss scenario for contemporary and historical buildings in Thessaloniki, Greece. Soil Dyn Earthq Eng 28(10–11): 836–850

Keller M, Pasanisi A, Marcilhac 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. Eng. Int., 30, 921–933.

Kijko A, Sellevoll MA (1992). Estimation of earthquake hazard parameters from incomplete data files. Part II. Incorporation of magnitude heterogeneity, Bulletin of the Seismological Society of America 82 (1), 120-134

Koutsourelakis P (2010). Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: a Bayesian framework. Probab Eng Mech 25:49–60. doi:10.1016/j.probengmech.2009.05.005

Kulkarni RB, Youngs RR, Coppersmith KJ (1984). Assessment of confidence intervals for results of seismic hazard analysis, Proceedings of the Eighth World Conference on Earthquake Engineering, San Francisco, California, Vol. 1.

Leone F, Lavigne F, Paris R, Denain JC, Vinet F (2010). A spatial analysis of the December 26th, 2004 tsunami induced damages: lessons learned for a better risk assessment integrating buildings vulnerability. Appl. Geogr., 31, 363–375.

Linstone HA, Turoff M (1975). The Delphi Method: Techniques and Applications. Addison-Wesley Publishing Company Advanced Book Program, Reading

Liu PLF, Yeh H, Synolakis CE eds (2008). Advanced numerical models for simulating tsunami waves and runup, in Advances in Coastal and Ocean Engineering Vol 10, World Scientific.

Loughlin S, Sparks S, Brown S, Jerkins S, Vye-Brown C Eds. (2015). Global Volcanic Hazards and Risk, Cambridge University Press, Cambridge

Lorito S, Selva J, Basili R, Romano F, Tiberti MM, Piatanesi A (2015). Probabilistic hazard for seismically induced tsunamis: accuracy and feasibility of inundation maps, Geophys. J. Int., 200, 574–588.

Lorito S, Romano F, Lay T (2016). Tsunamigenic earthquakes (2004 – 13): source processes from data inversion. In: Meyers, R. (ed.) Encyclopedia of Complexity and Systems Science. Springer Science + Business Media, New York, https://doi.org/10. 1007/978-3-642-27737-5_641-1

Løvholt F, Glimsdal S, Harbitz CB, Nadim F, Zamora N, Peduzzi P, Dao HI, Smebye H (2012). Tsunami hazard and exposure on the global scale, Earth-Science Reviews, Volume 110, Issues 1–4, Pages 58-73, ISSN 0012-8252, 10.1016/j.earscirev.2011.10.002.

Løvholt F, Pedersen G, Glimsdal S (2010). Coupling of dispersive tsunami propagation and shallow water coastal response, Open Oceanography Journal, Caribbean Waves Special Issue, 4, 71-82, doi: 10.2174/1874252101004020071, available online from http://www.bentham.org/open/tooceaj/openaccess2.htm

Løvholt F, Griffin J, Salgado-Galvez M (2015). Tsunami Hazard and Risk Assessment on the Global Scale, R.A. Meyers (ed.), Encyclopedia of Complexity and Systems Science, Springer, DOI 10.1007/978-3-642-27737-5_642-1

Maida M, Maier K, Obwegeser N (2012). Pairwise comparison techniques for preference elicitation: using testretest reliability as a quality indicator, CONF-IRM 2012 Proceedings. Paper 65. http://aisel.aisnet.org/confirm2012/65

Maruyama Y, Yamazaki F, Mizuno K, Tsuchiya Y, Yogai H (2010). Fragility curves for expressway embankments based on damage datasets after recent earthquakes in Japan. Soil Dyn Earthq Eng 30: 1158–1167

Marzocchi W (2013). Seismic Hazard and Public Safety, Eos, Vol. 94, No. 27, 2 July 2013

Marzocchi W, Mastellone ML, Di Ruocco A, Novelli P, Romeo E, Gasparini P (2009). Principles of multi-risk assessment: interaction amongst natural and man-induced risks (Project report). Office for Official Publications of the European Communities, Luxembourg

Marzocchi W, Woo G (2009). Principles of volcanic risk metrics: Theory and the case study of Mount Vesuvius and Campi Flegrei, Italy. J Geophys Res 114:B03213, doi:10.1029/2008JB005908

Marzocchi W, Sandri L, Selva J (2010). BET_VH: a probabilistic tool for long-term volcanic hazard assessment, Bull. Volcanol., 72(6):717–733.

Marzocchi W, Garcia-Aristizabal A, Gasparini P, Mastellone M, Di Ruocco A (2012a). Basic principles of multi-risk assessment: a case study in Italy. Nat Hazards, pp 1–23. 10.1007/s11069-012-0092-x

Marzocchi W, Newhall CG, Woo G (2012b). The scientific management of volcanic crises, J. Volcanol. Geotherm. Res., 247–248, 181–189.

Marzocchi W, Jordan TH (2014). Testing for ontological errors in probabilistic forecasting models of natural systems, Proc. Natl. Acad. Sci. Unit. States Am. 85, 955–959.

Marzocchi W, Taroni M, Selva J (2015). Accounting for epistemic uncertainty in PSHA: Logic Tree and ensemble modeling, Bull. seism. Soc. Am., 105(4), 2151–2159.

Mignan A, Wiemer S, 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, Scolobig A, Sauron A (2016). Using reasoned imagination to learn about cascading hazards: a pilot study, Disaster Prevention and Management 25 (3), 329-344, doi: 10.1108/DPM-06-2015-0137

Mitsoudis DA, Flouri ET, Chrysoulakis N, Kamarianakis Y, Okal E, Synolakis CE (2012). Tsunami hazard in the southeast Aegean Sea, Coast. Eng., 60, 136–148.

Molinari I, Tonini R, Lorito S, Piatanesi A, Romano F, Melini D, Hoechner A, Gonzàlez Vida JM, Maciás J, Castro MJ, de la Asunción M (2016). Fast evaluation of tsunami scenarios: uncertainty assessment for a Mediterranean Sea database, Nat. Hazards Earth Syst. Sci., 16, 2593-2602, doi:10.5194/nhess-16-2593-2016

Morgan MG (2014). Use (and abuse) of expert elicitation in support of decision making for public policy, Proc. Nat. Acad. Sci. 111(20); 7176-7184, DOI:10.1073/pnas.1319946111

Moschonas I, Kappos A, Panetsos P, Papadopoulos V, Makarios T, Thanopoulos P (2009). Seismic fragility curves for Greek bridges: methodology and case studies. Bull Earthq Eng 7(2):439–468

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 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.

NIBS (2004) HAZUS, Technical manuals, vol. Earthquake loss estimation methodology, federal emergency management agency edn. Washington, D.C.

Omira R, Matias L, Baptista MA (2016). Developing an Event-Tree Probabilistic Tsunami Inundation Model for NE Atlantic Coasts: Application to a Case Study. Pure and Applied Geophysics, 173(12), 3775-3794.

Pagnoni G, Tinti S (2016). Application and comparison of tsunami vulnerability and damage models for the town of Siracusa, Sicily, Italy. Pure Appl. Geophys., 173, 3795-3822.

Papathoma M, Dominey-howes D, Zong Y, Smith D (2003). Assessing tsunami vulnerability, an example from Herakleio, Crete. Nat. Hazards Earth Syst. Sci., 3, 377–389.

Paris R, Belousova M, Belousov A, Ontowirjo B, Whelley PL (2013). Volcanic tsunami: a review of source mechanisms, past events and hazards in Southeast Asia (Indonesia, Philippines, Papua New Guinea), Nat Hazards (2014) 70:447–470, DOI 10.1007/s11069-013-0822-8

Paté-Cornell ME (1996). Uncertainties in risk analysis: six levels of treatment, Reliab. Eng. Syst. Saf., 54, 95–111.

Paté-Cornell ME (2012). On "black swans" and "perfect storms": risk analysis and management when statistics are not enough, Risk Anal; 32(11):1823-33, doi: 10.1111/j.1539-6924.2011.01787.x

Pedersen G, Løvholt F (2008). Documentation of a global Boussinesq solver, Preprint Series in Applied Mathematics 1, Dept. of Mathematics, University of Oslo, Norway, URL: http://www.math.uio.no/eprint/appl_math/2008/appl_2008.html

Pinto PE Ed. (2007). Probabilistic methods for seismic assessment of existing structures, vol. LESSLOSS Report No. 2007/06. IUSS Press, Pavia

Pitilakis K, Alexoudi A, Argyroudis S, Monge O, Martin C (2006). Chapter 9: vulnerability and risk assessment of lifelines. In: Goula X, Oliveira CS, Roca A (eds) Assessing and managing earthquake risk, geo-scientific and engineering knowledge for earthquake risk mitigation: developments, tools, techniques. Springer, ISBN 1-4020-3524-1, pp 185–211

Reasenberg P (1985). Second-order moment of central California seismicity, 1969-82, J. Geophys. Res., 90, 5479–5495

Rossetto T, Elnashai A (2003). Derivation of vulnerability functions for European-type RC structures based on observational data. Eng Struct 25:1241–1263

Rossetto T, Elnashai A (2005). A new analytical procedure for the derivation of displacement-based vulnerability curves for populations of RC structures. Eng Struct 27(3):397–409

Rougier J, Sparks RSJ, Hill LJ Eds. (2013). Risk and Uncertainty Assessment for Natural Hazards (eds. Cambridge University Press, isbn: 9781107006195.

Saaty TL (1980). The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, ISBN 0-07-054371-2, McGraw-Hill

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 (2013). Long-term multi-risk assessment: statistical treatment of interaction among risks , Natural Hazards, DOI: 10.1007/s11069-013-0599-9

Selva J, Sandri L (2013). Probabilistic Seismic Hazard Assessment: Combining Cornell-like approaches and data at sites through Bayesian inference , Bull. Seism. Soc. Am., 103(3):1709-1722, DOI: 10.1785/0120120091

Selva J, Argyroudis S, Pitilakis K (2013), Impact on loss/risk assessments of inter-model variability in vulnerability analysis , Natural Hazards, DOI: 10.1007/s11069-013-0616-z

Selva J, Iqbal S, Taroni M, Marzocchi W, Cotton F, Courage W, Abspoel-Bukman L, Miraglia S, Mignan A, Pitilakis K, Argyroudis S, Kakderi K, Pitilakis D, Tsinidis G, Smerzini C (2015). Deliverable D3.1: Report on the effects of epistemic uncertainties on the definition of LP-HC events. STREST project: Harmonized approach to stress tests for critical infrastructures against natural hazards.

Selva J, Tonini R, Molinari I, Tiberti MM, Romano F, Grezio A, Melini D, Piatanesi A, Basili R, Lorito S (2016). Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA) , Geophys. J. Int., 205: 1780-1803, DOI: 10.1093/gji/ggw107

Schmidt J, Matcham I, Reese S, King A, Bell R, Henderson R, Smart G, Cousins J, Smith W, Heron D (2011). Quantitative multi-risk analysis for natural hazards: a framework for multi-risk modelling, Nat. Hazards, 58 (3): 1169–1192

Sørensen MB, Spada M, Babeyko A, Wiemer S, Grünthal G (2012). Probabilistic tsunami hazard in the Mediterranean Sea, J. Geophys. Res., 117, B01305, doi:10.1029/2010JB008169.

Spence RJS, Kelman I, Baxter PJ, Zuccaro G, Petrazzuoli S (2005). Residential building and occupant vulnerability to tephra fall. Nat Hazards Earth Syst Sci 5(4):477–494

SSHAC - Senior Seismic Hazard Analysis Committe (1997). Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts, U.S. Nuclear Regulatory Commission Report NUREG/CR-6372.

Suppasri A, Mas E, Charvet I, Gunasekera R, Imai K, Fukutani Y, Abe Y, Imamura F (2013). Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami, Nat Hazards (2013) 66:319–341, DOI 10.1007/s11069-012-0487-8

Stergiou E, Kiremidjian AS (2006). Treatment of uncertainties in seismic risk analysis of transportation systems. Technical report no. 154, John A. Blume Earthquake Engineering Center, Civil Engineering Department, Stanford University, Stanford, CA

Straub D, Der Kiureghian A (2008). Improved seismic fragility modeling from empirical data. J Struct Saf 30:320–336

Tarbotton C, Dall'Osso F, Dominey-Howes D, Goff J (2015). The use of empirical vulnerability functions to assess the response of buildings to tsunami impact: Comparative review and summary of best practice. Earth Sci. Rev., 142, 120-134.

Thio HK, Somerville P, Polet J (2010). Probabilistic tsunami hazard in California. Pacific Earthquake Engineering Research Center, PEER Report 2010/108, University of California, Berkeley.

Tinti S, Bortolucci E, Armigliato A (1999). Numerical simulation of the landslide-induced tsunami of 1988 in Vulcano island, Italy. Bull. Volcanol., 61, 121-137.

Titov VV, Gonzalez FI (1997). Implementation and testing of the Method of Splitting Tsunami (MOST) model. NOAA. Technical Memorandum ERL PMEL-112, 11 pp.

Tonini R, Armigliato A, Pagnoni G, Zaniboni F, Tinti S (2011). Tsunami hazard for the city of Catania, eastern Sicily, Italy, assessed by means of Worst-case Credible Tsunami Scenario Analysis (WCTSA), Nat. Hazards Earth Syst. Sci., 11, 1217–1232.

Toth Z, Kalnay E (1993). Ensemble Forecasting at NMC: The Generation of Perturbations. Bull. Amer. Meteor. Soc., 74, 2317–2330, doi: 10.1175/1520-0477(1993)074-2.0.CO;2.

USNRC - U.S. Nuclear Regulatory Commission (2012). Practical Implementation Guidelines for SSHAC Level 3 and 4 Hazard Studies, Prepared by AM Kammerer & JP Ake, NRC Project Manager: R Rivera-Lugo, NUREG-2117.

Valencia N, Gardi A, Gauraz A, Leone F, Guillande R (2011). New tsunami damage functions developed in the framework of SCHEMA project: application to European-Mediterranean coasts, Nat. Hazards Earth Syst. Sci., 11, 2835-2846, doi:10.5194/nhess-11-2835-2011

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 RW, Rovida A, Sesetyan K, Stucchi M, the SHARE consortium (2015). The 2013 European Seismic Hazard Model - Key Components and Results. Bulletin of Earthquake Engineering, doi: 10.1007/s10518-015-9795-1.

Wronna M, Omira R, Baptista MA (2015). Deterministic approach for multiple-source tsunami hazard assessment for Sines, Portugal, Nat. Hazards Earth Syst. Sci., 15, 2557-2568, doi:10.5194/nhess-15-2557-2015

Yeo GL, Cornell CA (2005). Stochastic characterization and decision bases under time-dependent aftershock risk in performance-based earthquake engineering, Pacific Earthquake Engineering Research Center

Zitellini N, Gracia E, Matias L, Terrinha P, Abreu MA, DeAlteriis G, *et al.* (2009). The quest for the Africa-Eurasia plate boundary west of the Strait of Gibraltar. Earth and Planetary Science Letters, 280(1), 13–50.