International Journal of Disaster Risk Reduction 110 (2024) 104641

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

A novel multiple-expert protocol to manage uncertainty and subjective choices in probabilistic single and multi-hazard risk analyses

J. Selva[a,*], S. Argyroudis[b], F. Cotton[c], S. Esposito[d], S.M. Iqbal[e], S. Lorito[e], B. Stojadinovic[f], R. Basili[e], A. Hoechner[c,g], A. Mignan[f,h], K. Pitilakis[i], H.K. Thio[j], D. Giardini[f]

[a] Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse (DiSTAR), Università degli Studi di Napoli 'Federico II', Naples, Italy
[b] Brunel University London, Kingston Lane, UB8 3PH, Uxbridge, United Kingdom
[c] Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany
[d] Swiss Re, Zurich, Switzerland
[e] Istituto Nazionale di Geofisica e Vulcanologia, Italy
[f] ETH, Zurich, Switzerland
[g] Gempa GmbH, Potsdam, Germany
[h] Mignan Risk Analytics GmbH, Greater Zurich, Switzerland
[i] Aristotle University, Faculty of Engineering, Department of Civil Engineering, Thessaloniki, Greece
[j] AECOM, Los Angeles, CA, USA

ARTICLE INFO

Keywords:
Expert elicitation
Epistemic uncertainty
Stress test
Critical infrastructure
Multi-hazard risk
Natural hazards

ABSTRACT

Integrating diverse expert opinions in hazard and risk projects is essential to managing subjective decisions and quantifying uncertainty to produce stable and trustworthy results. A structured procedure is necessary to organize the gathering of experts' opinions while ensuring transparency, accountability, and independence in judgements. We propose a novel Multiple-Expert management Protocol (MEP) to address this challenge, providing procedural guidelines for conducting single to multi-hazard risk analyses. MEP establishes a workflow to manage subjectivity rooted in (i) moderated and staged group interactions, (ii) trackable blind advice through written elicitations with mathematical aggregation, (iii) participatory independent review, (iv) close cooperation between scientific and managerial coordination, and (v) proper and comprehensive documentation. Originally developed for stress testing critical infrastructure, MEP is designed as a single, flexible, technology-neutral procedural workflow applicable to various sectors. Moreover, its scalability allows it to adapt from high to low-budget projects and from complex probabilistic multi-hazard risk assessments to standard single-hazard analyses, with different experts' degree and type of involvement depending on available funding and emerging controversies. We present two compelling case studies to showcase MEP's practical applicability: a multi-hazard risk analysis for a port infrastructure and a single-hazard regional tsunami hazard assessment.

* Corresponding author. University of Naples, Federico II, Dept. of Earth, Environmental, and Resources Sciences, Complesso di Monte Sant'Angelo, Via Vicinale Cupa Cintia, 21, 80126, Naples, Italy.
E-mail address: jacopo.selva@unina.it (J. Selva).

https://doi.org/10.1016/j.ijdrr.2024.104641
Received 8 October 2023; Received in revised form 10 May 2024; Accepted 24 June 2024
Available online 26 June 2024

1. Introduction

Natural hazard and consequent risk modelling are characterized by many degrees of freedom, often with non-linear dependencies and sometimes by a relative lack of observations (e.g., for rare events). Hazard and risk quantifications should provide the probability of exceeding different levels of intensity for one or more phenomena within a time window (exposure time) in a target area [1–5].

Natural phenomena are never perfectly understood or predictable, and past data are often insufficient to discriminate among existing modelling alternatives to quantify the consequent hazard and risk. Consequently, different and potentially diverging implementations of probabilistic analyses are possible, and alternative formulations may produce significantly different and sometimes inconsistent estimates [6–10]. Since the results of such analyses may be relevant for planning and policy-making, hence ultimately affecting public safety, it is essential to estimate the emerging uncertainty [11–14].

The uncertainty related to different opinions, interpretations and/or different implementations may be quantified and communicated in the form of a probability distribution, describing "*the center, the body, and the range of technical interpretations that the larger informed technical community would have if they were to conduct the study*" [12,15]. This distribution is referred to in the literature with different expressions, like the technical community distribution [16] or the extended experts' distribution [13]. The uncertainty that this distribution quantifies is often referred to as "epistemic uncertainty" or "inter-model variability" [5,9–13,15,17,18]. From now on, we will univocally refer to this distribution as epistemic uncertainty or community distribution.

The quantification of this community distribution is still rather inhomogeneous across natural hazards [11]. In seismic hazard, epistemic uncertainty is usually quantified by implementing multiple scientifically acceptable (not rejected by data) methods, including both alternative parametrizations of individual models and the adoption of alternative models [13,19,20]. This is in common with many other fields, like for example for weather forecasts, flood hazards or climate models [21–27]: all such methods foreseen the development of ensembles of alternative acceptable models, which are then aggregated. Aggregation methods largely differ in the different fields. For example, in seismic hazard models, alternatives are aggregated through a so-called logic tree [28], a method that has been widely debated and is still an object of active research [7,16,19,29]. In other fields, like volcanology, Bayesian approaches are often preferred [18,30–33].

The involvement of multiple experts becomes increasingly important as the relevance and the regulatory concerns of a project grow (e.g., Ref. [12,15,25,34,35]). Engaging multiple experts enables better constraining of community distribution. This allows the technical community to consider all existing alternative views better, thus better matching the effective variability. It also helps facilitate the management of subjective choices that lead to uncertainty and minimising possible a posteriori criticism if a forecast model fails [12,36–40]. To this end, there is an increased need to organize group interaction, define roles and responsibilities, and build clear and transparent documentation of the interaction process [12,15].

Several protocols have then been defined in the past to manage disagreements and consensus in setting up hazard and risk models, defining interaction within a multiple-expert environment to build community distributions. In the field of natural hazards, and in particular in seismic hazards, which represents one of the oldest and most advanced fields for natural hazards [3,41], the most applied and referenced protocol to manage multi-expert interaction is the one developed by the Senior Seismic Hazard Analysis Commission (SSHAC) in 1997 [12], with its subsequent specifications and integrations [15,17,42], hereinafter referred to as SSHAC guidelines. While defining their guidelines, SSHAC analysed several previous seismic hazard analyses, concluding that differences in the results are often due to procedural rather than technical differences, highlighting the need of establishing appropriate procedures to define the community distribution through the interaction of multiple experts. SSHAC guidelines proposed four distinct levels for conducting seismic hazard analyses, ranging from a single expert summarizing the literature variability (Level 1), to multiple groups organizing discussions among different groups of experts, to provide alternative integration processes for hazard results (Level 4). SSHAC guidelines have the paramount merit of having introduced standardized procedures for Probabilistic Seismic Hazard Analysis - PSHA, remaining the primary reference in this topic for over two decades.

In the meantime, quite different approaches have been developed to deal with similar problems in other domains. For example, in the context of volcanic hazard assessment, the input of multiple experts to hazard models is usually dealt with structured elicitations of small groups of experts [37,43–49], with typical panels of 10–20 experts [38,43,50,51]. Elicitation results are then used to initialize one specific (usually Bayesian) hazard model [18,30–33]. This approach is radically different from what is foreseen in SSHAC. First, in such methods, alternatives remain hidden within expert opinions and are not made explicit. Second, in classical elicitations, the experts act independently [36,40,48,52,53], assuming that the required knowledge already exists within individual experts. On the contrary, SSHAC foresees a structured interaction of multiple actors that concur with the definition of "integrated" models. The development of such integrated models is based on a multi-staged interaction among the experts with strong non-formalized group interactions and face-to-face meetings between twhe experts to integrate their opinions through the project [17]. This approach, which is typical also of other methods like, for example, the Delphi method [38,54], is instead criticized by the practitioners of classical elicitations for the potential biases that such interaction may introduce, e.g., due to the undesirable effects of personalities and reputations that may unavoidably bias the group quantification (e.g., Ref. [43,44,55–58]).

A homogenous management of epistemic uncertainty in a multiple-expert environment would be a fundamental step toward an effective multi-hazard risk, as required by modern holistic risk reduction programs (e.g., Ref. [121]). However, achieving the effective applicability of existing protocols beyond their primary scope poses significant technical and methodological challenges. Ensuring scalability, methodological flexibility, and technological neutrality are essential considerations when adapting these methods to different fields. These challenges are particularly evident in situations requiring interaction between different communities, such as in multi-hazard risk analyses for stress testing critical infrastructure systems [59–62].

Existing protocols are typically not scalable. They pre-select the level of complexity and the number of potential participants based on the available resources and the regulatory concern [42], adopting field-specific choices. For nuclear facilities, seismic hazard is usually estimated through large-scale projects based on the SSHAC protocol, with tens to hundreds of participants [16,17,42,63,64] and with a process of feedback, interaction, and learning among the experts [15]. Instead, classical structured expert elicitations typically involve a few tens of participants who act independently and meet only to execute the elicitation (see Ref. [43,48,51]). In either case, the scale of the project is usually defined in advance. This is quite standard when a scientific project has to be proposed. Still, in other situations, such as implementing stress tests for critical infrastructures, a more flexible approach may be preferable, allowing for dynamic scalability that reflects the needs emerging during the project's development [60].

Existing protocols for multiple-expert interaction are also defined in a single-hazard environment, with field-specific methods and limited methodological flexibility and neutrality. For example, developing numerous alternative models, prescribed by SSHAC protocols to quantify epistemic uncertainty, is possible within the seismological community but may not be feasible in other fields that lack a sufficient number of alternative models. For instance, probabilistic tsunami hazard assessment is a relatively less-established methodology [4,65]. Few or no models exist for dealing with problems like fragility for multiple hazards (e.g., Ref. [66,67]) or for pre-damaged structures (e.g., Ref. [68]), or restoration for assessing infrastructure resilience (e.g., Ref. [69]). In addition, in fields like multi-hazard risk and resilience assessment (e.g., Ref. [67,70–78,122]) or systemic risk and resilience analyses (e.g., Ref. [79–81,123]), the selection of the sources, hazards and/or system's components and interactions, pose critical challenges typically not encountered in single hazard analyses. While an explicit formulation of alternatives should be favoured to increase transparency and readability of the results, procedural decision-making steps to deal with defining priorities and managing the absence of relevant models are still required in multi-hazard risk models to deal with complex, real-world scenarios effectively.

Similarly, existing protocols often lack technological neutrality, as they not only define the multi-expert interaction but also prescribe the scientific tools to be used (e.g., Ref. [13,15,30,31,45,48,49,51,55,82,120]). For example, the epistemic uncertainty quantification in SSHAC guidelines for seismic hazard relies on logic trees, while other fields adopt different strategies. Most of the elicitation protocols predefine a specific elicitation technique. This is again linked to the fact that most protocols have been developed in a single-hazard environment. Instead, more neutrality may be achieved by defining the required steps toward the hazard/risk definition, as in existing multi-hazard protocols [34,70,72,74,78], which however typically do not detail the management of epistemic uncertainty within a multiple expert environment.

Existing protocols in the setup of hazard/risk assessments are also not stringent in the management of subjective decisions during the development of the analysis, leaving this part entirely under the responsibility of the project coordination. However, this may be critical in projects with relevant societal implications and regulatory concerns. In hazard and risk assessments, a number of subjective choices are always made. The consequences of such choices are often, if not always, not quantified. The effect of not-considered sources and/or interactions may sometimes be unpredictable [67,74–76,83,84]. Therefore, a transparent decision-making process should be introduced. There is a vast literature on methods for decision-making in multiple-expert environments, ranging from classical elicitation methods to Multi-Criteria Decision Analysis - MCDA [61,62,85,86]. These methodologies assign weights to criteria and assess diverse options, thereby enabling comparisons across multiple criteria and helping quantify uncertainty's impact on rankings and decisions. These approaches provide the means to scrutinize alternative scenarios and solutions, facilitating an understanding of the impact of different choices and exploring the implications of different choices to assess stakeholder preferences and priorities better. Thus, such methods allow for transparent decision-making processes based on evaluating alternatives considering multiple criteria. They may also provide fundamental hints for the decisions that are ultimately necessary at the stage of setting up hazard and risk analyses, increasing the transparency and readability of such assessments.

This paper introduces a Multiple-Expert management Protocol (MEP), which defines a transparent workflow of tasks that allows for generalizations and significant flexibility. MEP incorporates and merges many of the key elements of existing protocols, generating an innovative checks-and-balances interaction between project coordination and panels of experts and reviewers to manage the process toward consensus, aiming to achieve the comprehensiveness and accuracy required in hazard/risk analyses to support decision-making for increased safety and reliability of complex technological systems. MEP was developed in synergy with the development of the STREST stress test framework [60] and the NEAM Tsunami Hazard Model 2018 - NEAMTHM18 [87]. While MEP was primarily developed within the geophysical community, it can be readily extended to other natural hazards and domains.

2. Multiple-expert management protocol (MEP): goals and key features

MEP is targeted to guide the development of scientific projects aimed at producing single-to-multi-hazard and risk assessments, especially when their results may imply potential regulatory concerns, for which the consistent evaluation of existing uncertainty in a multi-expert environment is particularly relevant. The main goal of MEP is the production of robust results that fairly account for alternative hypotheses in the community, collected through a traceable and transparent process that treats the different levels/steps of the assessment homogeneously. MEP pursues the integration of different expert opinions based on the following.

- a clear definition of the roles and responsibilities of the different experts;
- a staged development and revision workflow of tasks toward procedural consensus of the multiple experts participating in the project, organized in three phases (pre-assessment, assessment, finalization) with specific participatory review and technical documentation; to pursue technical neutrality, all stages define specific goals and milestones, without defining specific techniques to obtain them;

- a revisable and homogeneous decision-making process for managing all critical choices and for the quantification of the epistemic uncertainty, based on trackable advice from a restricted group of experts that acts independently from project management, collected through decision-making methods like written classical elicitations and MCDA in the different phases;
- a dynamic management of project scope and complexity, allowing scalability (complexity should be scaled with the scopes) and flexibility (emerging complexities may be dealt with by changing the scope or scaling up the project);
- a solid quantification of the community distribution that represents the potential range of the results based on scientifically acceptable opinions.

These key features are discussed in more detail in the following paragraphs. To avoid unnecessary confusion, we adopt, when possible, naming and terminology used in existing guidelines, mainly from SSHAC guidelines and classical elicitation literature.

2.1. Definition of roles and responsibilities

MEP defines five main actors/groups with different and pre-defined roles, revising and integrating the group definitions foreseen in SSHAC guidelines. The **Project Manager (**PM**)** is responsible for managing the project in the name of the project's funder. The PM defines all the questions the project should answer, defining timing and funding based on existing tasks and potential consequences. The **Technical Integrator (TI)** leads the scientific process for answering the project's questions. The PM and TI collectively manage the project, representing the funder and scientific components. The TI coordinates an **Implementation Team (IT)** to implement the planned evaluations. PM, TI, and IT form the core of the project, with the PM having a managerial role while TI and IT have a more scientific role. Unless alternative agreements are found, TI and IT collectively hold its intellectual property. Two other groups support the project's activity with feedback that is out of the direct control of PM and TI. A **Pool of Experts (PoE)** participates in formal written expert elicitations to provide the TI with independent judgements or quantifications over specific subjective choices. An **internal Participatory Peer Review Panel (iPPRP)** reviews the whole process to maximise the results' reliability and increase their robustness. This review process takes place throughout the development process. The main interactions among the core actors are illustrated in Fig. 1.

The PM is responsible and accountable for the successful development and execution of the project. The main stakeholder (infrastructure owner or funding agency) generally appoints the PM. Hence, it is the PM's responsibility to make rational and fair decisions for stakeholders, authorities, and the public. Specifically, the PM (i) selects the target level of detail and complexity of the analysis, (ii) oversees, along with the TI, the participant selection budget and guides the project's sophistication to balance the required funding and the requested level of details in results; (iii) is the sponsor of the elicitation experiments, without playing any formal role in such experiments. The PM is equivalent to SSHAC's "project leader", but in MEP, we prefer to change this name to reflect the fact that the effective leadership of the project is shared (with different roles) with the TI.

The TI is an individual responsible and accountable for the scientific management of the project. To achieve this effectively, the TI captures the views of the informed technical community to be implemented in the hazard and risk calculations in the form of trackable opinions and community distributions. More specifically, the TI, within the budgetary constraints set by the PM, (i) carefully selects experts for various roles, guaranteeing a balancing mix of competences, and proposes to the PM potential integrations when required; (ii) moderates meetings of experts and defines elicitation methods, while by coordinating elicitation experiments successfully extract opinions from the PoE; (iii) oversees the hazard and risk calculations, defining the techniques to manage the quantification of the community distributions and uncertainty propagation; (iv) coordinates the methodological review process based on feedback from the iPPRP. Notably, as per SSHAC guidelines, the TI coordinates the process of evaluating and integrating alternative datasets and models. However, in MEP, the TI's role is not that of an evaluator or integrator; instead, the TI acts as a moderator during the evaluation/integration process based on the formal elicitations of the PoE. To excel in this multifaceted role, the TI may consider forming a small support team with diverse expertise to cover all necessary skills adequately.

The Implementation Team (IT) is a group of analysts who actually implement, under the coordination of the TI, the hazard and risk assessments and all other complementary activities (e.g., sensitivity analyses and pre- and post-processing of results). The IT does not participate in managing the project or expert elicitations, which is entirely managed by the TI (or the TI team). Consequently, as

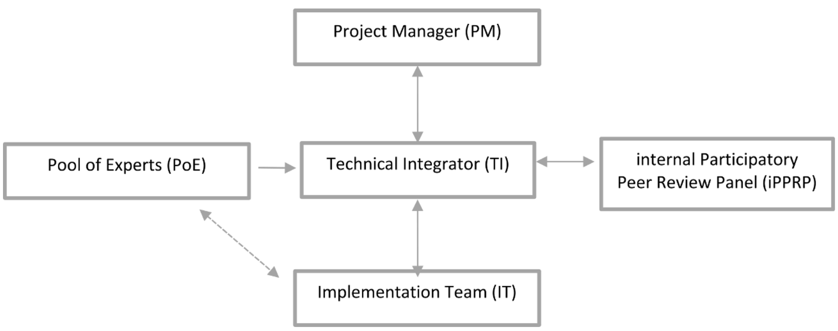


Fig. 1. Primary interactions among the main groups of MEP. To guarantee fairness, intellectual independence among Project Manager (PM), Technical Integrator (TI), and internal Participatory Peer Review Panel (iPPRP) must be guaranteed. In addition, project management (TI and PM) members cannot participate in the Pool of Experts (PoE).

further discussed below, individuals from the IT may participate in elicitations, creating an overlap between the two groups. The IT team is selected jointly by the TI and the PM. If the stakeholders have a scientific team in their personnel, such experts may join the IT. In this sense, the IT may also help improve the interaction with stakeholders, for example, help to define a common language and facilitate the reciprocal comprehension, interpretation and communication of the results to the stakeholders. Notably, the IT should include both resource and proponents experts, as defined in SSHAC guidelines.

The Pool of Experts (PoE) is intended to represent the larger technical community, with members from the project and the external community. Its main goal is to provide blind input to the TI with respect to critical subjective choices. This group is not present in SSHAC guidelines. In MEP, the PoE is instead central in the process of evaluation and integration of alternative models, for which MEP does not charge any individual but relies on a group response of the PoE. To manage potential sources of cognitive biases - including (but not limited to) the impact of “over-opinioned” experts or experts with strong personalities or with scientific or political power - the PoE opinions should be collected through structured and blind elicitations, with written questionnaires collected anonymously, with mathematical aggregation. The procedure may include the definition of weights for the different experts [43] and the adoption of multiple criteria for evaluating alternative decisions [85]. To favour the integration of experts' knowledge during the project, more than a mere aggregation of opinions pre-existing to the project [17], MEP allows for the inclusion of some experts from the IT in the PoE. This allows for including in the elicitations experts that fully participated in the project's development and actively participated in implementing methods. To preserve the independence of roles, the PoE cannot include TI, PM or review panel members (iPPRP). Since this interaction among experts may also generate “fake” consensus due to a levelling or averaging of opinions during discussions [43,56,88], MEP still requires a substantial presence of completely external experts (not included in any other groups; e.g., > 50 %), allowing to verify the consistency of results and bringing new expertise and potentially different views.

This pool's size and selection process are discussed in various papers [36,40,48,52,53] and references therein). Typically, PoE has 7 to 15 members (e.g., Ref. [43]); however, larger panels may be appropriate for some projects [89]. In the case of complex assessments, sub-pools on specific topics may be foreseen. Experts may have site-specific (e.g., hazards in the area) and/or target-specific (e.g., a critical infrastructure) knowledge or expertise on a hazard/risk methodology and/or on the typology of infrastructure (e.g., nuclear or non-nuclear such as transport, energy or communication networks). In the elicitation, the experts of the PoE must act independently, unless interaction is requested and moderated by the TI.

The iPPRP comprises one expert or a group of experts who independently peer-review the project developments. This group corresponds to the review panel of SSHAC, sharing equivalent goals and responsibilities. The terms “internal” and “Participatory” are specifically used to emphasize that the iPPRP's involvement occurs throughout the project, and no post-project reviews can serve as a substitute. The iPPRP engages in a participatory review process, providing written reviews at predetermined stages during project implementation. These reviews encompass technical aspects, such as evaluating selected models and covering interdependencies and procedural aspects, like assessing actor's independence, expertise coverage, transparency, and consistency with MEP. The iPPRP must possess knowledge about all relevant issues, and if necessary, topical meetings may be organized with IT members. The requested expertise may be reached collectively; thus, the iPPRP size may vary depending on the project's complexity. It is important to note that MEP follows a review process akin to scientific papers, wherein a member of the iPPRP acts as a coordinator (editor), while other members of the iPPRP conduct their assessments independently, at least in an initial phase. To maintain objectivity, the iPPRP members cannot have affiliations or shared membership with any other group.

2.2. Project's workflow toward consensus

The procedural consensus among the key actors guarantees the fairness of the results in MEP. Procedural consensus does not imply an effective consensus of divergent opinions but rather a fair representation of existing divergences within the overall procedure. This consensus is reached with an explicit agreement among the PM, TI and iPPRP on the final documentation, each holding specific responsibilities as outlined above. Consensus should be reached on both procedural and technical aspects. Procedural aspects address the completeness of the documentation, the conformance to guidelines and rules, and the overall fairness of the process. Technical aspects address the scientific method and its implementation, drawing from PoE inputs, iPPRP reviews, and PM/TI consequent choices that also account for temporal and budgetary constraints and any specific requests from stakeholders. Note that the results cannot be the object of the agreement since they can actually be checked only by the TI and the IT.

MEP aims at favouring the process toward consensus with a workflow of actions organized in two rounds of assessment and review, both including the contribution of all actors, and followed by a final phase of outreach (Fig. 2). To favour this procedural convergence, all the necessary subjective decisions should be taken adopting solid decision-making methods, and their ground is reviewed during the project. To guarantee transparency and enable future scrutiny, all choices, data, models and methods must also be documented throughout all phases, with a structured documentation plan that should become available to all actors at the end of each phase. Potentially sensitive information may be blurred or hidden for security reasons. Still, this action should be explicitly motivated and should not hinder the ex-post scrutiny of the overall process of hazard and risk quantification process. In the end, all documents and reports are collected in a container that grants the reproducibility of the results and a long-term curation and maintenance.

Critical subjective choices include, at least, the selection and the evaluation of available data and models; the quantification of their credibility weights; the selection of the hazards or hazard source areas; the definition of model components, such as discretization or approximations; the selection of hazard interactions modes for multi-hazard risk assessments; the selection of infrastructure assets and inter-dependencies for systemic risk; the definition of performance indicators; the management of theoretical gaps, including missing models; the design of sensitivity analyses or sanity checks.

In the decision-making phase, MEP aims at balancing the opposite requirements of large interacting communities, as in SSHAC guidelines, and of small groups of independent experts, as in classical elicitation procedures, by initializing all decisions with formal

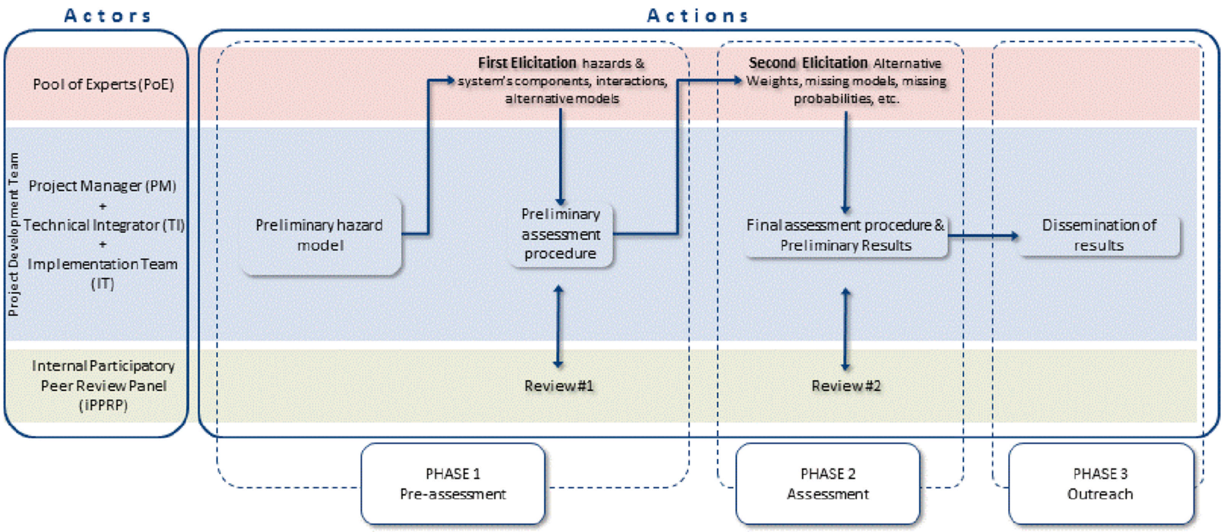


Fig. 2. MEP's project workflow (modified from Ref. [87]).

methods like classical elicitation and/or multi-criteria analysis methods with a limited group, the PoE. This should be based on written questionnaires to ensure reproducibility. Then, the project management (PM and TI) proposes solutions based on such input, potentially fostering new discussions and/or needing new expertise. Finally, decisions are made explicitly, discussed and reviewed by an unimpeded and independent review panel to make the decision-making process solid, that is, trackable, independently reviewed, and open to ex-post scrutiny after the end of the project.

Notably, this process has the goal of explicating all decisions. This means that the PoE formal feedback does not necessarily translate into actions but rather leads to a dynamic re-evaluation of the project scopes and means, eventually simplifying the process or redefining the final objectives. The TI and PM may also consider process sophistications or additions, if possible. For example, potential controversies may suggest topical meetings among experts; new significant topics may emerge, suggesting the extension of expertise within the PoE and the IT or the planning of further analyses (e.g., sensitivity tests). Multiple elicitations of the updated PoE may be foreseen, but this keeps the project manageable as the PoE is a relatively small group. In any case, such possible iterations should occur within each of the defined stages until PM, and TI judge the overall assessment procedure sufficiently strong to be sent to the iPPRP for the internal review.

2.3. Comparison between MEP and existing protocols

MEP incorporates several elements of the existing protocols, producing an innovative combination of requirements and tackling some fundamental additional key issues. In particular, from SSHAC guidelines, MEP takes the definition of the main roles and the existence of a participatory review, adding the Pool of Experts and revising the specific assignments of all the other groups. As in typical multi-hazard risk protocols, MEP defines a list of goals and targets of the analysis, assigning specific targets to each defined role. However, MEP includes such tasks in the multiple-expert environment, which aligns with the SSHAC protocol. In addition, specific attention is given in MEP to managing subjective decisions during the assessment, which must be based on solid, revisable and explicit decision-making methods. Thus, MEP incorporates the methods typical of decision-making, like classical elicitations or MAPE. In addition, in line with protocols like SSHAC, MEP requires a transparent, participatory review and an extended documentation process to further open all the elements of hazard and risk models for future scrutiny. Still, unlike SSHAC, this also includes all the decision-making points made explicit through the elicitations.

SSHAC guidelines identify four different study levels, primarily considering the different levels of complexity of the study. In very short, SSHAC level 1, the simplest in the scale, considers a TI that reviews the literature, formulates the implementation plan, evaluates the results, and passes it through a peer-review process; SSHAC level 2 adds the interaction with representatives from the larger technical community, the potential organization of topical meetings, and potential interaction with reviewers at earlier stages; in SSHAC level 3, the TI formally organizes workshops to discuss data and model alternatives, while the review is participatory; in SSHAC level 4, alternative TI teams are considered, and a technical facilitator is in charge of producing the final community distribution. In this framework, MEP ranges between SSHAC levels 2 and 3. Indeed, in MEP, experts from the informed technical community participate mainly through formalized feedback (expert elicitations) during the whole process, without active interactions among them (SSHAC's workshops). This configures MEP at SSHAC level 2. However, single experts of the pool may also act as proponents and advocate a particular hypothesis or technical position (regarding both resources and methods) at intermediate stages of the process (during Phases 1 and 2) through individual communications with the TI. In addition, MEP considers participatory review throughout the course of the project. Both these things are elements of SSHAC level 3. Moreover, if significant contrasts in experts' opinions emerge during the elicitations, formal and transparent interaction among the experts can be planned, transforming the pool into a panel and further pushing MEP toward SSHAC level 3.

A substantial difference exists between MEP and SSHAC processes regarding the integration of alternative models to quantify the community distribution. MEP uses conventional expert elicitation tools (for a discussion on the similarities and differences between SSHAC elicitations and classical expert elicitations [17]). In particular, MEP uses such elicitations to establish a formal decision-making process for managing all subjective matters, including the management of model evaluation and integration. In MEP, no individuals are expected to act as evaluators/integrators of the different models. Still, there is a process of integration initiated by elicitations and finalized by the project coordination with the assistance of a review panel. The technical integrator acts as a moderator of the integration process, not as an integrator/evaluator. This allows for making feedback traceable down to a single expert and avoiding the need for very large projects. This also allows for avoiding a complete dependence on the supposed (hard-to-reach) objectiveness or independence of evaluators/integrators from resource/proponent experts and dealing with the potential problems related to group interaction [36,40,43,57]. Consequently, MEP and SSHAC differ significantly in the process of integrating experts' opinions and reaching consensus. While SSHAC is mainly based on meetings among all the experts, MEP is based on the combination of written elicitations (whose members act independently and interact only if significant and irreconcilable divergences emerge) and reviews. This means that MEP potentially generates fewer meetings/travels, making it eco-friendly and easier to manage.

MEP differs significantly also from standard elicitation protocols in which experts are mainly asked to act independently. In MEP, all experts are involved in the project activity, favouring an integration of different opinions. In this sense, the experts become experts of the specific topic during the projects, as discussed in the SSHAC protocol [17]. Formal elicitations are used to initialize a decision-making process that involves different actors (project coordination and reviewers). This process is meant to guarantee integration throughout the project of expertise but also to fully document the reasoning behind subjective choices. This allows future scrutiny of such matters to build a factual path for improving hazard/risk quantifications in the subsequent generations based on documented controversies, drawbacks and criticisms.

3. MEP: detailed project workflow

MEP foresees that the project workflow is articulated in 3 phases. Each phase consists of several sequential stages with specific tasks that are logically or sequentially ordered. In Fig. 3, we report the graphical representation of the different phases and stages, the involvement of the main actors and their phase-wise interactions. The main actions in each phase are described in more detail in the next sections.

3.1. PHASE 1: pre-assessment

PHASE 1 is devoted to defining a preliminary implementation plan that enables the target assessment. PHASE 1 starts with selecting the core actors and elaborating a reliable plan for the results. It includes a first elicitation experiment and ends with the first formal review. PHASE 1 is divided into 5 stages, each producing specific documentation. All the actions required at each stage, their temporal sequence and the required documentation are reported in Table 1.

3.1.1. PHASE 1 - STAGE 1: project definition & kick-off

The pre-assessment phase starts with the project's organisational part (timing and total maximum budget) and the selection of all the main actors. This phase ends with the organization of a kick-off meeting involving PM-TI-IT.

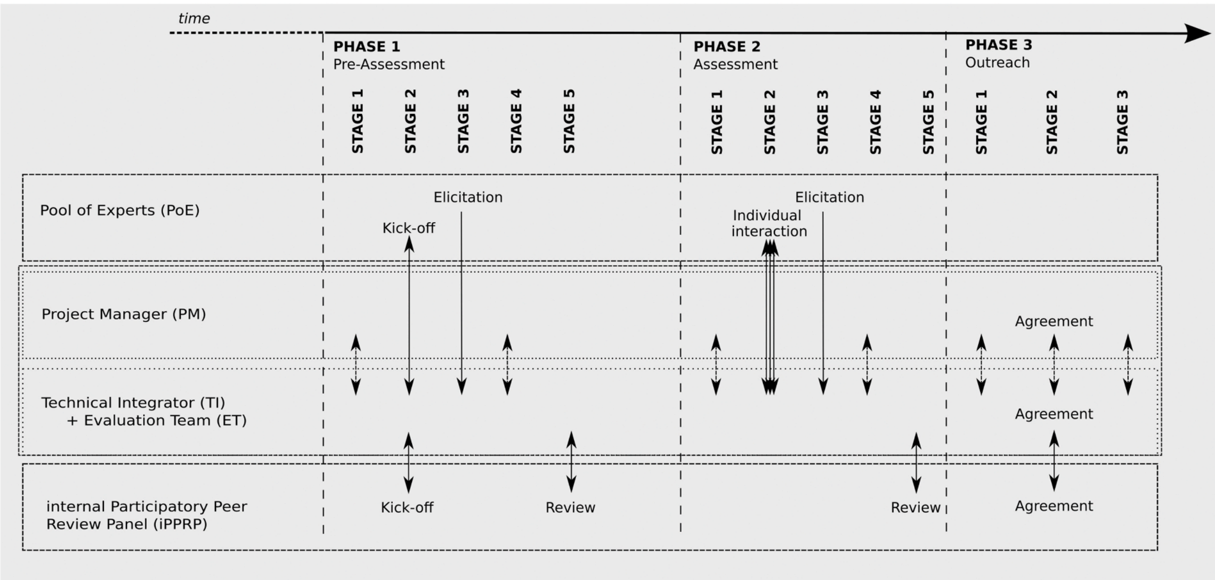


Fig. 3. Graphical representation of MEP, showing the different phases and the interactions between the main actors in each phase.

Table 1
Description of actors, actions and required documentation in the stages of PHASE 1.

ACTIVITY	ACTORS	ACTIONS	DOCUMENTATION
PHASE 1 - STAGE 1: Project definition & kick-off	PM, TI	Selection of the core actors (PM, TI, and IT) and of the main characteristics of the analysis, including budget and timing.	Report: Project summary
PHASE 1 - STAGE 2: PoE and iPPRP kick-off	PM, TI, PoE	Selection of the iPPRP and of the PoE. The TI organizes the PoE kick-off meeting. If weights to experts are foreseen, they are assigned at this stage.	Report: Kickoff agenda
PHASE 1 – STAGE 3: First elicitation experiment	TI + IT, PoE	PoE elicitation for prioritization, to define the framework of the analysis (e.g., selected hazards, sources, components, interaction) and to manage existing alternatives. TI and IT review the literature, prepare the elicitations, and elaborate PoE answers. Elicitation results may trigger topical meetings and/or subsequent analyses, if required.	Report: Results of the first elicitation round
PHASE 1 - STAGE 4: Plan finalization	PM, TI + IT	The TI defines an implementation plan, agreed with the PM. The existence of methodological gaps potentially impacting the results is also checked, in order to avoid an unwanted under-exploration of potential paths to risk.	Report: Preliminary implementation plan
PHASE 1 - STAGE 5: Review & Finalization	IR, TI, PM	The iPPRP review PHASE 1 documents, triggering a peer-review process to converge toward an agreed implementation plan.	Report: First review

More specifically, stakeholders select a PM. The PM selects the TI, and jointly, the PM and TI select the starting nucleus of the IT. The PM (with the technical assistance of TI and IT) selects the target level of deepening of the assessment (e.g., regional vs site-specific analysis, literature vs ad hoc models) and specifies its regulatory requirements (e.g., exposure time window, limits in the mean return periods) and its target (e.g., list of intensity measures for different hazards, list of risk metrics, such as injuries/deaths, economic losses, pollution), as well as any other definition that completely define the target analysis (e.g., levels of acceptable risk, target performance, see Ref. [60]). This also includes the rules to be followed regarding intellectual properties, disclosure of the documentation, and dissemination. In other words, the PM selects all the questions the project has to answer. All these definitions must be tuned in collaboration with the stakeholders to agree on the project's budget and time limits, and should be agreed upon with the TI. In the case of projects in competitive calls, all these selections are implicit in the development of the proposal, and the funding agencies accept them as the project is funded.

At this stage, an approximate size of participants for the different multi-expert groups (IT, PoE, and iPPRP) is defined. MEP is fully scalable from large to small groups, with a minimum of 5–10 experts in the PoE and a few individuals in both iPPRP and IT. The actual size of groups should reflect the need to cover the expertise range, but may also be tuned to favour the general community's acceptance of results. This strategic decision should be agreed upon between PM and TI, based on the needs of stakeholders, also accounting for the fact that it might become necessary to update some of the groups. This may happen at the beginning of phase 2 (Section 3.2) and may be particularly important when innovative analyses with little background are foreseen.

Documentation includes a summary of all the choices discussed and finalized in a kick-off meeting organized by the PM and involving TI, IT and representatives of stakeholders.

3.1.2. PHASE 1 - STAGE 2: PoE and iPPRP kick-off

Once the project framework is defined, the TI selects PoE and iPRRP. A kick-off meeting for both is organized, in which all fact-sheets of PHASE 1 are shared and discussed. The kickoff may include training sections, and if a weighting scheme is foreseen for the experts of the PoE, it may be initialized through seed questionnaires at this point. This enables to perform the subsequent elicitations also remotely. Documentation includes programs, presentations, and seed questionnaires.

3.1.3. PHASE 1 – STAGE 3: first elicitation round

The first round of elicitation is focused on strengthening the definition of the general framework of the assessment. This may involve multiple tasks, depending on the application. To this end, two major milestones are (i) the definition of the main requirements of hazard/risk modelling, and (ii) the enumeration of the available alternative models.

Defining the main requirements of hazard/risk modelling is not always straightforward. For example, it may include the selection of the hazards to be considered, the selection of the sources, or the main interactions to be considered. This is critical in all multi-hazard or systemic risk quantifications, but it may also be important in a single hazard, such as tsunamis that may have multiple sources [4]. The TI and the IT should extensively review the scientific and technical literature about existing guidelines and previous experiences, and organize an ad hoc prioritization elicitation whenever a sufficient ground for decisions does not exist.

Once requirements are set up, the TI and the IT review the literature for available data, methods and models. The review should include potential databases of interest, system analysis requirements, and available methods at regional, sub-regional/national, and local levels. In this, individual interactions with PoE members or other experts may play a fundamental role, in which all experts may act as resource and proponent experts (as defined in SSHAC guidelines). After this review, the TI can set up a preliminary list of approaches to be potentially adopted, and existing alternatives should be screened. Indeed, excluding models can drastically modify the community distribution, and, at the same time, including all the available models is useless and sometimes dangerous [19]. Hence, considering the cost of implementing all alternatives, organizing an elicitation to prioritize actions is fundamental.

The elicitation experiments may be in person or remote, and different techniques may be adopted or specifically developed. Many prioritization techniques based on expert elicitation exist, like the PC (Pairwise Comparison, e.g., Ref. [90]), the AHP (Analytic Hierarchy Process, [108]), BBN (Bayesian Belief Networks, e.g., Ref. [91]), other MCDA (Multi-Criteria Decision Analysis) [61,62,85,86], or many other methods. As in all elicitations, the PoE should be supported by significant supporting material, including the description of methods, alternative models and/or potential sensitivity tests. If specific inconsistencies emerge from the elicitation, the TI may decide to convene the PoE (or parts of it) to openly discuss such issues and eventually define new sensitivity tests to be performed, allowing for final decisions.

Documentation for this stage includes preparatory material, questionnaires, and analysis of the first round of elicitation results.

3.1.4. PHASE 1 - STAGE 4: plan finalization

Based on the results of the elicitation and the literature review, the TI defines a concrete implementation plan agreed upon with the PM. The implementation plan should discuss the method planned for aggregating the selected alternatives and the data that will be used for testing the results.

The existence of methodological gaps potentially impacting the results is also checked at this point to avoid an unwanted under-exploration of potential paths to risk [83]. Such gaps may be managed by planning further elicitations (for PHASE 2) or supplementary analyses (for example, specific scenarios as in Ref. [60]) and/or sensitivity tests to verify the practicability of simplified modelling approaches (for example, defining simplified end-members to explore their impact on results).

Documentation includes a detailed implementation plan, a description of databases, models and implementations, and a summary of the whole decision process, indicating limitations due to project constraints.

3.1.5. PHASE 1 - STAGE 5: review

This stage concludes PHASE 1 with the iPPRP review. The iPPRP reviews the detailed document produced in PHASE 1 – STAGE 4 as well as all the other documents produced in PHASE 1 (Table 1). iPPRP should pay attention to both the scientific plan and the decision process applied by the TI and PM, evaluating i) the consistency of the TI choices with the PoE feedback and PM requirements, ii) the completeness of the collection of data and methods, and iii) the effectiveness of the defined overall method. The iPPRP members should put themselves in the role of TI and PM and evaluate how the actual PM and TI managed the decision-making process and how the final results reflect the project's general goals.

This early-stage formal review plays a pivotal role within the iPPRP task, as its outcomes can significantly impact the project at its initial phase. The TI can facilitate the review process by preparing specific questionnaires highlighting crucial points for evaluation. To favour the convergence, the review process may incorporate written and iterative interactions between the TI and iPPRP, similar to standard peer-review processes, coordinated by an iPPRP coordinator nominated by the panel. In cases where the TI and PM cannot fully implement iPPRP suggestions based on scientific, technical or managerial (funding/timing) grounds, they may reshape the project's goals by, for example, diminishing the target level, explicitly agreeing upon such changes with the IPPRP, and/or explicitly exposing existing criticisms.

In a first phase, each iPPRP member must perform the review independently, possibly in an anonymous form. Subsequent panel discussions are possible. Then, an iPPRP coordinator manages the review process by interacting with the TI. Documentation includes the review of each member of the iPPRP and the potential interactions with TI.

3.2. Project's PHASE 2: assessment

The purpose of PHASE 2 is to produce the preliminary results. PHASE 2 starts with defining the implementation plan based on the iPPRP review, includes a second elicitation, and ends with a second review. PHASE 2 is logically organized in 5 stages, as defined in Table 2.

3.2.1. PHASE 2 - STAGE 1: implementation plan & group revision

At the beginning of PHASE 2, TI and PM finalized and documented the final implementation plan, which emerged from the iPPRP review. If this implies new analyses that require expertise not sufficiently covered by existing PoE or IT, the TI and the PM may have to discuss potential updates of these groups and organize individual meetings with the newly engaged project members.

Table 2
Description of actors, actions and documents in the stages of PHASE 2.

ACTIVITY	ACTORS	ACTIONS	DOCUMENTATION
PHASE 2 – STAGE 1: Finalization of the implementation plan	PM, TI, IT	The TI and the PM agrees over the methodology to be implemented. If needed, TI and PM update the groups of experts and the objectives of the project.	Report: Implementation plan
PHASE 2 – STAGE 2: Single model implementation	TI, IT, PoE	TI and the IT update the project's databases and implement all models and tests, also through individual interactions with PoE members. Codes and databases are properly documented.	Repository: Data and codes
PHASE 2 – STAGE 3: Second elicitation round	TI, PoE	The TI organizes the second elicitation to quantify the weights of alternatives, to manage missing models, etc.	Report: Second elicitation round
PHASE 2 – STAGE 4: preliminary results	PM, TI, IT	The implementation of the analysis produces the preliminary results.	Report: Preliminary results
PHASE 2 – STAGE 5: Review & finalization	IR, TI	The iPPRP review PHASE 2 documents, triggering a peer-review process to converge toward final results.	Report: Second review

The final report on the implementation plan should explicitly discuss the updates from PHASE 1 based on the interaction with the iPPRP. This document and the review document should be immediately submitted to the iPPRP to enable an agreed-upon finalization of the review process of PHASE 1.

3.2.2. PHASE 2 - STAGE 2: individual models implementation

In Stage 2, the TI and IT collaborate to update the project's databases and implement all required models and tests. In this stage, individual interactions with members of the PoE or with other experts are essential, particularly when specific members are considered reference experts for certain methods or data. During implementation, the TI also finalizes a list of consistency and/or sanity checks to be implemented.

Computationally extensive tasks, like, for example, building a tsunami unit sources database (see Section 4.2) [92], might have started earlier, especially if they influence decisions made in the previous stage (PHASE 2 – Stage 1). However, their effective use is defined at this point.

A thorough technical implementation documentation is required to ensure reproducibility and transparency. Results should be linked to published datasets and models, and any new databases, data processing tools, or simulation codes must be fully documented. This documentation should be consolidated into one document, providing detailed workflows, data, and codes whenever possible, to enable reproducibility.

3.2.3. PHASE 2 - STAGE 3: second elicitation round

Stage 3 may run in parallel with Stage 2, providing input to models through a second elicitation. This includes quantifying the weights of alternatives to integrate them to produce the community distribution; this may also include other elicitations to fill existing gaps, provide missing input, or define the details for additional analyses (e.g., test scenarios).

The quantification of the weights of alternative models is an essential part of integrating alternatives, probably representing the most subjective part of the quantification of the community distribution. To avoid complete control of the TI over such weights, MEP foresees that at least a factor of such weights should be based on the elicitation of the PoE, which represents the larger technical community [87,93]. These weights may be eventually combined with other weights, like the ones based on statistical testing [94,95]. The PoE should be supported by an exhaustive description of the selected methodologies and the implemented tests, as well as the results of consistency and/or sanity checks (e.g., Ref. [19,96,115]). In practice, the quantification of weights may be based on prioritization techniques (as in Section 3.1.3).

Ignoring modelling gaps may lead to sever risk underestimation [83]. To avoid this, substantial gaps must be explicitely managed. Expert elicitation may be adopted to evaluate missing probabilities that cannot be set based on past records [18,31,97] or to quantify expert-based missing models (e.g., missing fragility model for one component [98]), or even to define complementary scenarios to evaluate the potential of existing gaps [60]. This allows for basing these potentially critical choices with the blind support of the PoE, rather than leaving them under the full control of TI and PM.

Documentation of this second round of elicitations includes preparatory material, questionnaires, supporting material, and analysis of the results.

3.2.4. PHASE 2 - STAGE 4: alternatives aggregation and preliminary results

Based on Stage 2 implementations and Stage 3 input, TI and IT can produce the preliminary results.

At this stage, alternative models are aggregated. As discussed above, different methods exist to perform aggregation. An important issue to tackle is to check if marked multi-modality is found in the distribution of model alternatives, as in the case of UCERF 3 seismic hazard model for the site of Redding, California [7,99], or in the case of tephra hazard in Naples [29]. In these cases, groups of alternative models diverge in their forecast, and the cause of such diversion should be explicitly discussed to understand whether intermediate values should be expected (and preferred).

The results of consistency and/or sanity checks and other post-processing analyses like disaggregation must be reported to increase the results' readability. Consistency checks should show that the results statistically agree with existing evidence (i.e., they are not rejected), and this should be tested explicitly through standard hypothesis testing (e.g., Ref. [64,100–102]). As data used for testing are usually already considered in formulating the hazard/risk model, such tests are not proper statistical tests but simply sanity checks. Disaggregation results are also fundamental to check the consistency of results and the potential implications of several choices, as well as to balance the contribution of the different hazard/risk factors (source, propagation, fragility, etc.) to the final results [103].

Documentation is prepared by the TI and the IT as a final report to present the results and the post-processing analyses. The eventual lack of testing should be explicitly discussed by the TI and reviewed by the iPPRP.

3.2.5. PHASE 2 - STAGE 5: second review

This stage concludes PHASE 2 with the iPPRP review of the final implementation and the preliminary results. The iPPRP reviews the detailed documents produced in PHASE 2 – STAGE 1 and PHASE 2 – STAGE 4 as well as all the other documents produced in PHASE 2 (Table 2). As in the first review round (PHASE 1 – STAGE 5; ***Section 3.1.5***), the iPPRP should evaluate technical and procedural aspects. The specific foci of this second review are: i) the consistency of the decision-making process based on iPPRP and PoE feedback for defining the final implementation plan and its overall scientific quality, ii) the coherence of results with implementation, and iii) the adequacy of the presented sanity checks and other post-processing analyses. The TI should facilitate iPPRP's task by preparing specific questionnaires containing the main points to be evaluated by the reviewers, leaving a free field to accommodate any general comment of the iPPRP. Also in this second round, to favour the convergence toward final results, the review process may foresee a written interaction between TI and reviewers under the coordination of IPPRP, with replies to comments and reviewers' re-evaluation, as in ordinary peer-review processes. This process is similar to what is described in PHASE 1 – STAGE 5, and specific documentation is produced.

3.3. Project's PHASE 3: finalization & dissemination

PHASE 3 is the conclusive phase in which the project's results are finalized, an agreement among the main actors on the project's methods is reached, and final hazard/risk assessment results are disclosed. PHASE 3 is logically organized in 3 stages, which are not necessarily sequential in time, but most likely may proceed in parallel. All the actions required in PHASE 3 and the required documentation are reported in Table 3.

3.3.1. PHASE 3 - STAGE 1: finalization of the model and documentation

The TI and the IT jointly revise the final models and results considering the feedback from the iPPRP and perform the final computations. If significant changes were agreed with the iPPRP in PHASE 2, this process may imply communicating with the iPPRP until finalization.

At this stage, the TI produces a document that synthetically presents the final results and tests. Complementary to this document, an extended abstract of the project is produced that synthesizes the whole process and describes all the activities carried out, the main decisions made, the potential limitations, and suggestions for potential future improvements. To this end, iPPRP suggestions and critics should be discussed, especially whenever suggestions cannot be fully implemented for financial or technical reasons. The TI and IT also update previous documents regarding the implementation plan, producing a second version of this document that accounts for possible final changes required to accommodate the second round of review.

3.3.2. PHASE 3 - STAGE 2: agreement

At this stage, the PM, TI and iPPRP should agree on the project's procedure, methods, and documentation. To this end, the TI interacts with the PM and iPPRP to reach an agreement on the final implementation plan and results, as reported in a specific document. The TI can favour the convergence by arranging a final meeting with the PM and iPPRP. The agreement may be formalized through a formal signed statement and/or by co-authoring a common document describing the overall procedure, results, and documentation. Once the agreement is reached, all documents produced in all phases and stages are curated, for example, through the assignment of a Digital Object Identifier (DOI), and made available to the stakeholders and, if possible, to the extended technical community and the general public, following the rules agreed at the beginning of the project.

The documentation of Stage 3 includes meeting notes, the signed agreement statement and/or the jointly authored document.

3.3.3. PHASE 3 - STAGE 3: dissemination

The PM and the TI organize the dissemination phase. The dissemination should be designed to reach, at least, the extended technical community. A project website should be prepared, in which at least the extended abstract and the documents regarding the agreement are made openly available to the public. MEP strongly encourages that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people. At the same time, ethics and standards should guide exceptions for not sharing data. The iPPRP is involved one last time to check that the key project's results are reproducible.

Scientific publications (if pertinent) should be organized at this stage by the TI, in agreement with PM and stakeholders, and stored in a project repository. The stakeholders should store and maintain all the documents and the website with the support of the PM and the TI.

4. Applications of MEP in real case studies

To showcase the flexibility and scalability of MEP, we present two distinct applications: a multi-hazard risk analysis for stress testing a critical infrastructure (section 4.1) and a regional single hazard analysis (section 4.2). In both cases, the involvement of multiple experts through MEP has proven pivotal, allowing a robust decision-making process that effectively manages subjective choices and selections. For multi-hazard applications, MEP enables the selection of hazards through collaborative input from various experts. Additionally, it facilitates the definition of the modelling approach for systemic risk assessment, including identifying infrastructure components and their interactions. In single hazard analyses at the regional level, MEP allows alternative approaches to address epis-

Table 3
Description of actors, actions and documents in stages of PHASE 3.

ACTIVITY	ACTORS	ACTIONS	DOCUMENTATION
PHASE 3 – STAGE 1: Finalization of the model and of the documentation	PM, TI	The TI and the PM fully document the methodology and the results, and prepare final documentation. This includes also the preparation of high-level summary documents to increase the readability of results for end-users and/or the general public. All documents are crystallized, to prevent future changes, and made fully available to stakeholders.	Report: Extended abstract Report and Repository: Results Report: Updated implementation plan
PHASE 3 – STAGE 2: Agreement	PM, TI, iPPRP	A final meeting between PM, TI and iPPRP is held, to reach an agreement on project's procedure, methods, and documentation.	Report or publication
PHASE 3 – STAGE 3: Dissemination	PM, TI, iPPRP	Project's data, codes, results and methods are disseminated outside the project. Dissemination tools are maintained by the stakeholder. PM and TI organize scientific publications. The iPPRP check that the results are reproducible.	Public Repository and Website with documentation

temic uncertainty quantification, significantly improving the overall model's validity and reliability. In both cases, this approach enhances objectivity, accountability, and results acknowledgement, garnering support from end-users and the broader technical community.

4.1. Stress test of the port of hessaloniki, Greece

The MEP approach was applied to assess the safety of critical infrastructure at the port of Thessaloniki (PoTh) in Greece. This study was part of the European FP7 STREST research project [59] and utilized a risk-based multi-level stress test methodology [60]. The port occupies a total space of 1.5 million sq. m, covering a length of 3.5 km. Its infrastructure comprises six piers for cargo and passenger traffic, extending along a 6200 m quay with a sea depth down to 12 m and storage areas of 600,000 sq. m. The MEP implementation was tailored for this specific case study within a research project, and some stages were simplified accordingly. As this was a demonstration case study rather than a formal application, some modifications were also made to ensure feasibility and practicality, as discussed below.

4.1.1. Phase 1: Pre-assessment

The stress test involves a probabilistic risk analysis for the spatially extended port infrastructure, considering its critical interdependencies, such as energy supply, for multiple hazards, with quantification of epistemic uncertainties. Single and multiple hazards and interdependent infrastructure are not predefined; hence, their selection and classification are central to the project and are addressed in Stage 1.

In Stage 1, the roles and responsibilities of PM, TI and IT are defined in Table 4. The iPPRP was formed with reviewers defined within STREST for relevant project deliverables. The TI and the PM identified PoE members. A formal invitation letter was sent to 21 experts with diverse backgrounds, including the port authority, industry, and academia. The PoE comprises experts in hazard analysis (e.g., physicists, seismologists) and infrastructure design and assessment (e.g., civil and infrastructure engineers). A kick-off meeting (Stage 2) was organized on February 25, 2016 in Thessaloniki, Greece, where the MEP process and the role of experts were presented; the meeting was attended by PM, TI, IT and 15 PoE members. To quantify the potential impact of a weighting scheme on the results, all the elicitations were processed using three weighting schemes: Equal Weighting (EW), Acknowledgement-based Weighting (AW) and Performance-based Weighting (PW). PW and AW weights were assigned to experts through seed questionnaires. Subsequently, two rounds of specific elicitation (Stage 3) took place to screen and prioritize natural hazards and port infrastructure relevant to the PoTh. Round A involves screening single hazards and port infrastructure, while round B focuses on prioritizing hazards and infrastructure.

The screening process of hazards and their possible interactions was based on a list of 21 natural hazards [71], including geophysical, hydrological, shallow earth processes, atmospheric, biophysical and space/celestial hazards. For single hazards, screening criteria were defined based on their relevance to the site and port infrastructure, e.g., pertinence in the region, negligible impact in the region, redundancy, lack of relevance to PoTh, and negligible influence on PoTh. For multiple hazards, the screening was based on interaction criteria, such as one hazard triggering another, simultaneous occurrence of two hazards, increased infrastructure vulnerability due to a hazard, or increased exposure to a second hazard.

The screening of single port infrastructure was based on performance and vulnerability criteria. This entailed evaluating whether an infrastructure influences the target performance indicators of the port and assessing its vulnerability to all relevant hazards. The screening of infrastructure interdependencies was based on criteria relevant to physical/functional, informational, geographical, restoration, substitute, sequential, logical, societal, and general types of interdependencies [104].

Based on the elicitation results provided by ten experts for all the weighting schemes, the most relevant screened hazards included the following **single hazards**: earthquake (t_1), tsunami (t_2), flood (t_6), ground collapse (t_9), local soil subsidence (t_{10}), storm (t_{12}), and **multiple hazards**: earthquake and tsunami ($t_{1,2}$), earthquake and ground collapse ($t_{1,9}$), storm and flood ($t_{12,6}$), earthquake, tsunami and ground collapse ($t_{1,2,9}$). Screening a **single infrastructure** indicated that most of it was deemed relevant to the PoTh. The AW and EW had similar results, identifying most infrastructures except for wastewater (HBR6) and water systems (HBR7). The PW results

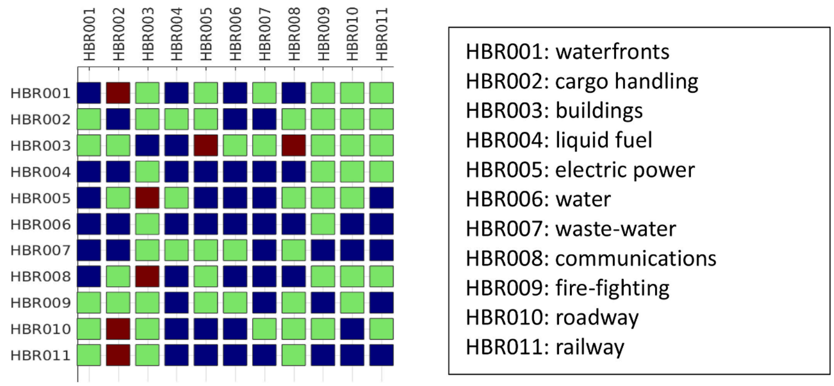
Table 4
Roles and responsibilities in MEP applications of Sections 3.1 and 3.2.

	Multi-hazard risk at Thessaloniki port, project STREST (Section 3.1)	NEAM Tsunami Hazard Model (NEAMTHM18), project TSUMAPS-NEAM (Section 3.2)
Project Manager (PM)	K. Pitilakis	R. Basili, supported by the TSUMAPS-NEAM Advisory Board
Technical Integrator (TI)	J. Selva, supported by S. Iqbal and S. Argyroudis	J. Selva, supported by S. Iqbal, S. Lorito, A. Hoechner, H·K. Thio
Implementation Team (IT)	S. Iqbal and S. Argyroudis	Project's Task 1 (Hazard Assessment) members
internal Participatory Peer Review Panel (iPPRP)	Independent peer review organized during the Project	J. Behrens, M. Dolce, D. Di Bucci, T. Parsons, E. Geist, M. Pagani, A. Amato, A. Zaytsev, M. Gonzalez, J. M. Gonzalez Vida
Pool of Experts	*Experts on hazards:* A. Kiratzi, G. Papathanasiou Z. Roumelioti, C. Papaioannou, E. Sokos, N. Theodoulidis *Experts on infrastructure:* M. Alexoudi, S. Christopoulos, S. Fotopoulou, K. Kakderi, S. Karafagka, I. Krestenitis, K. Makra, D. Pitilakis, E. Rovithis	*Internal to the project*: F. Romano, R. Omira, F. Lovholt, A. Babeyko, A. Yalciner, G. Papadopoulos, M. Canals, A. Bouallegue *External*: A. Armigliato, M. Sorensen, C. Ozer, G. Davies, W. Power, J. Polet, C. Meletti

tend to confirm the same trend; however, PW provided an intermediate score (still larger than the waste-water and water systems, but smaller than the others) for three infrastructures, namely, the fire-fighting (HBR09), the communication (HBR08) and liquid fuel systems (HBR04). Furthermore, based on the assessment of the experts' judgment for EW, the relevancy of screened **interdependencies** was categorized in three classes from non-relevant to relevant (Fig. 4A). The most relevant interdependencies included the following: waterfronts (HBR001) and cargo handling (HBR002) - unidirectional; buildings (HBR003) and electric power (HBR005) - bidirectional; buildings (HBR003) and communications (HBR008) - bidirectional; roadway (HBR010) and cargo handling (HBR002) – unidirectional; and railway (HBR011) and cargo handling (HBR002) - unidirectional.

To prioritize hazards and infrastructure during round B, experts were asked to provide judgments using a numerical scale from 1 to 7 derived from a verbal scale, where numerical values were linked to the verbal expressions typically adopted to describe the relative frequency of the events [24]. The elements assessed included (i) the relative likelihood of significant intensities of screened hazards; (ii) the likelihood of significant damages to the port infrastructures given significant intensities of single or combined hazards; (iii) the non-functionality of the screened port infrastructures, given significant damages; (iv) the impact on the port performance. The latter was defined in terms of two performance indicators: the total cargo/containers handled (p1) and the total cargo/containers handled and delivered to the port's gate (p2). Overall, no single hazard got the same priority under the three weighting schemes. Hazards $t_{1,9}$ and t_6 were identified as the most and least threatening under EW and AW, while t12 and $t_{1,2}$ were the most and least threatening under PW. HBR005 was consistently identified as the most vulnerable among port infrastructures, regardless of the weighting scheme. Conversely, HBR009 was found to be the least vulnerable under EW and AW for both p1 and p2. The prioritization of interdependencies was not performed since all selected interdependencies were implemented (Fig. 4A).

A)



Port infrastructures	Infrastructure code	Directivity of the link
Waterfronts and Cargo Handling	HBR01 and HBR02	Unidirectional
Buildings and Electric Power	HBR03 and HBR05	Bidirectional
Buildings and Communications	HBR03 and HBR08	Bidirectional
Roadway and Cargo Handling	HBR10 and HBR02	Unidirectional
Railway and Cargo Handling	HBR11 and HBR02	Unidirectional

B)

Prioritization of Steps

No.	Model code	Description
1	Step-1	Definition of the seismic source variability and quantification of the long-run frequencies of all the seismic sources
2	Step-2	Tsunami generation and offshore propagation
3	Step-3	Near-shore tsunami propagation and inundation
4	Step-4	Computation of the weights of the alternative models developed in Steps 1 to 3 to measure their credibility, and construction of the "ensemble" model

Fig. 4. A) Screening (above) and selection (below) of port infrastructure interdependencies and categorization to relevant (brown), intermediate (green), and non-relevant (blue), based on judgments provided by ten experts. B) Results of the elicitation to rank the need to explore modelling alternatives in the four steps of tsunami hazard quantification: green, yellow and red indicate low, medium and high priority, respectively [87].

After considering the elicitation results, the risk assessment model was finalized (Stage 4). At this stage, it was decided to include only the relevant hazards, infrastructure, and interdependencies in the model, considering the availability of data, models (e.g., fragility function), and resources. The risk assessment model was then internally reviewed within the research project (Stage 5).

4.1.2. Phase 2: Assessment

Following the Phase 1 revision, the risk assessment plan was finalized (Stage 1) and subsequently implemented (Stage 2) to perform the stress test [105]. Due to temporal constraints and the limited availability of full-scale risk analysis models for spatially distributed infrastructures, the analysis considered only one model without any alternative implementations. On the contrary, alternative implementations were available in the literature for seismic and tsunami hazards [5,106]. PM and TI deemed further elicitations unnecessary for the seismic model in Stage 3, as official hazard models already provided the required weights. However, an ad hoc elicitation was instead made to quantify potential alternative implementations of the source model of an ad hoc tsunami hazard analysis produced following the procedure defined in Ref. [107]. The weights of alternative models indicate each model's (subjective) credibility in the technical community; thus, the experts must compare pairs of alternative modelling implementations and assign a relative degree of preference. This involved an Analytical Hierarchy Process (AHP) based elicitation [108], a multi-criteria prioritization method involving pairwise comparisons of alternative options to assign numerical scores ranging from 0 to 1. The estimated numerical scores were normalized and used as weights of the alternatives in an ensemble model [106]. The results of the application [109] were produced as part of the stress-testing methodology [60], providing not only the results of the risk quantification but also offering suggestions for potential system strengthening (Stage 4). The STREST project deliverables documented the final results and underwent a standard review process (Stage 5).

4.1.3. Phase 3: Finalization and dissemination

The final model was implemented and documented, producing the final results of Stage 1 [105]. As this application served as a demonstration case study, a formal agreement (Stage 2) was not required, which would have been necessary in a real stress test scenario. The results of this case study were effectively disseminated (Stage 3) during the final meeting of the STREST project [110] as well as in project reports and publications discussing the stress testing approach to the PoTh [59,105,109,111].

4.2. Tsunami hazard: NEAMTHM18

The MEP workflow has been adopted in producing the reference regional hazard model for tsunamis originating from seismic events in the NEAM (North East Atlantic, Mediterranean and connected seas), named NEAM Tsunami Hazard Model 2018 – NEAMTHM18 [87,112,113]. This probabilistic hazard model has been produced in the framework of the project TSUMAPS-NEAM (http://www.tsumaps-neam.eu/). NEAMTHM18 has been adopted by the IOC (Intergovernmental Oceanographic Commission) of UNESCO in the Intergovernmental Coordination Group for the NEAM Tsunami Warning System (ICG NEAMTWS) 2030 Strategy document [114]. The full documentation of the project is reported in Ref. [112].

Here, we discuss MEP implementation for NEAMTHM18, highlighting the main procedural choices and the decision-making strategy. Being NEAMTHM18 a single-hazard model, many of MEP steps and stages are simplified or omitted, demonstrating MEP's overall scalability and flexibility.

4.2.1. Phase 1: Pre-assessment

Stage 1 started with the elaboration of the proposal coined "Probabilistic TSUnami Hazard MAPS for the NEAM Region - TSUMAPS-NEAM", submitted to the European Civil Protection and Humanitarian Aid Operations (DG-ECHO, https://ec.europa.eu/echo/index_en), which played the role of end-user. Representatives of all project partners (nine institutions from Italy, Norway, Portugal, Germany, Turkey, Spain, Greece, Morocco and Tunisia) agreed with DG-ECHO regarding the effort's main goals, timing and budget. The TSUMAPS-NEAM's kick-off was held in Rome on 10–11 February 2016. During the kick-off meeting, the main characteristics of MEP (in its early development stage) were discussed, and all partners agreed to follow MEP and shape the project's workflow accordingly. PM, TI and IT members were established (Table 4). Both PM and TI were assigned support teams to represent end-users and better organize expert elicitations, respectively.

For Phase 1 – Stage 2, The PoE (Stage 2) was formed by mixing representatives of the project partnership (one representative per partner) and external experts (reported in Table 4 as "Internal' and 'External', respectively), covering expertise on probabilistic hazard quantification (specifically, for seismic and/or tsunami hazards), seismic source uncertainty management including knowledge of the regional sources, tsunami and seismic hazard, tsunami numerical modelling, and knowledge of the historical tsunamis. The iPPRP comprised only external experts, considering expertise similar to the one described for the PoE, with additional expertise on the end-use of tsunami hazards (civil protection representatives). PM and TI and their respective supporting teams made the selection of external members (iPPRP and part of the PoE). At the same time, each institution proposed the PoE's internal components, which were agreed upon by the TSUMAPS-NEAM Advisory Board (PM support team).

All members of PoE and iPPRP were formally invited by the PM, and a joint iPPRP-PoE kick-off meeting was organized along with a project's technical meeting in Athens (Greece) from 29 June to July 1, 2016. The members of iPPRP and PoE were invited to participate in the entire meeting, during which a training session was held to familiarise them with the TSUMAPS-NEAM project approach, goals, and partners. The general scheme of the hazard assessment in the TSUMAPS-NEAM project was presented and discussed in the technical meeting. In the kick-off session, the TI presented the general scheme of MEP, the specific roles of the PoE, and the iPPRP, discussing the interaction process among the different actors and its scope. The TI also introduced expert elicitation methods, discussing their rationale, existing methods, and the rationale for weighing the experts in elicitations. It was agreed to implement multiple weighting schemes to test their impact on the results and let the PoE score the preferred one during the elicitation. Three weighting schemes have been adopted: equal, performance-based [37], and acknowledgement-based [51]. The fifteen members of the PoE were invited to answer a seed questionnaire and to acknowledge two other experts of their own choice within the PoE. The alternative weighting schemes were scored in the second elicitation round, resulting in a slight preference toward performance-based weights (0.57 against 0.43) [112]. These activities were propaedeutic to quantify the performance-based and acknowledgement-based weights, respectively.

In Stage 3, the TI organized the first elicitation experiment. The main goal has been to trim the potential alternative models for the epistemic uncertainty quantification. This was fundamental as many models exist, but the project's constraints (especially timing) and common practice [19,115] suggest limiting the total number of implemented models. The elicitation was held online, sending the questionnaires via email and asking for written replies (questionnaires and results can be found in Ref. [112]). The trimming was performed by prioritizing the steps (and the substeps) of hazard implementation regarding the need to explore alternatives. To this end, an AHP method [108] was implemented. To produce aggregate results, we implemented an aggregation of individual judgments technique considering both weighted arithmetic and geometric means to evaluate group central tendency [116–118], using all three weighting schemes. Since all weighting schemes provided coherent results, no further effort was made to aggregate them. The results clearly indicated the need for significant exploration of alternatives in modelling the frequency of high-magnitude events (e.g., annual frequencies, slip distributions), with relatively less effort to be implemented in other parts of the assessment, such as the implementation of alternative tsunami modelling for source-to-site propagation in deep sea [87,112].

In Stage 4, PM, TI and IT defined the preliminary implementation and prepared a detailed document to describe the details of hazard quantification. This document and all the documents of Phase 1 have then been submitted to the iPPRP for review (Stage 5). The review was done by asking each reviewer to express opinions regarding the project processes and methods described in the implementation plan, appendices, and other documents. Reviewers were asked to submit a numerical rating from 0 to 5 (5 being the highest rating) on 6 questions regarding the clarity of the documentation, the capability of the proposed methodology to fulfil the needs of end-users, the opinions expressed by the PoE, and to perform a satisfactory quantification of the hazard and of the community distribution. Reviewers were allowed to add specific suggestions and comments to accompany their responses and give unlimited general comments.

Phase 1 ended with the analysis of iPPRP reviews, which were summarized in a specific document [112]. Some concern was expressed about the preliminary documentation (too dispersive) and some specific modelling choices (e.g., treatment of the tails of the frequency-size distribution of earthquakes, the shape of unit sources for tsunamis, and the use of topo-bathymetric data). Suggestions were made also about potential sanity and consistency checks to be performed. The PM and the TI defined a strategy to address iPPRP concerns, agreeing to develop and implement an updated implementation plan in Phase 2.

4.2.2. Phase 2: Assessment

In Stage 1, TI and PM updated the preliminary plan to fulfil the review. The implementation plan was updated to produce a final version, with a clear definition of Phases 2 and 3 documentation and a plan for post-processing results and producing sanity checks. No need to update the PoE was identified. Then, the IT implemented all the selected models (Stage 2), and prepared a full documentation for both input data and developed codes. For publicly available data, the documentation provided the links to the original sources and relevant documentation. For new data, the documentation included a technical description of the data files with enough details to guarantee potential reuse. All files and codes were stored in a private repository, making them available upon request. The data-processing codes were also preserved in a computational platform that can be accessed upon request and subject to obtaining the necessary credentials.

The second PoE elicitation (Stage 3) focused on quantifying the weights of the implemented alternative models. A document with further details on implemented models was produced to support the PoE and provide a report on the implementation plan. AHP was again selected as the method for the elicitation because of the experience gained by the PoE participating in the first elicitation round. To merge the three weighting schemes into one weight, the preference of PoE to either performance-based or acknowledgement-based weighting schemes was asked through the AHP, and the score emerging from the elicitation, evaluated with equal weights for the experts, provided the weights to produce weighted means of performance-based and acknowledgement-based results. Following Goepel [117] and Forman and Peniwati [116], the weighted geometric means were considered to estimate the group's central tendency. The final scores have been adopted as weights for the alternative models.

The aggregation of alternative models (Stage 4) was performed using an ensemble modelling [7,106], providing, as a result, a family of hazard curves at each target point and their statistics (mean model, different percentiles). A short and synthetic document illustrating the preliminary results and the results of the main tests and sanity checks performed was then prepared.

In the second revision (Stage 5), iPPRP members were asked to express their opinions on the final project implementation, results, and assessment of the sensitivity of results to inputs, tools, and documentation. Opinions were collected through individual documents and summarized in a single document. As in Phase 1, iPPRP opinions were collected as a quantitative rating of the different aspects (method, integration of PoE suggestions, documentation, etc.), and reviewers were allowed to add specific suggestions and comments. Phase 2 ended with the analysis of iPPRP reviews [112], which showed that the scores improved with respect to the first round, with a general agreement about the implemented method and coherence of the results. Several suggestions were made about improving the uncertainty treatment (e.g., uncertainty on the size of finite faults, uncertainty in tsunami modelling parameters, uncertainty in inundation estimation, potential non-linearity in some target sites) and the representation of the results. Interesting suggestions were made about potential issues to be accounted for in the next-generation hazard quantifications, which were triggered by the project results. At this point, a final implementation plan was agreed.

A specific meeting was organized to discuss the final results with the iPPRP, held in Tunis, Tunisia, on 11–12 September 2017. PM, TI, IT, iPPRP, and end-user representatives participated in the meeting. The meeting included sessions dedicated to the presentation of the final results, discussion of the answer to reviewers (from the second round), an open discussion session with the participation of the iPPRP, PoE representatives and end-user representatives, and a session organized by the iPPRP dedicated to present the final iPPRP recommendations. Notably, PM, TI, and iPPRP agreed to have a further open review step based on access to the final results of the iPPRP panel and of the extended technical community, to collect potential feedback. To this end, the final results were disclosed and advertised to the larger technical community in December 2017. Results remained open for a three-month moratorium, during which feedback was collected. After this, the method and results were crystallized.

4.2.3. Phase 3: Finalization & dissemination

The documentation of methods was updated, and final results were produced (Stage 1). An agreement was finally reached, and all the participants in the entire process completed a common publication. This goal was reached in 2021 [87].

The preparation of the final documentation (Stage 2) and the dissemination to the larger technical community (Stage 3) partially overlapped in time. The final documentation included, in addition to technical documents on model implementation, a Layman's Report, guidelines for end-users, and other dissemination material (e.g., brochure, videos, etc.). The finalized documentation was published on September 12, 2019 [112], with an assigned Digital Object Identifier (DOI).

Dissemination (Phase 3 – Stage 3) started at the very beginning of the project with the publication of several preliminary, intermediate, and final results using the project website (http://www.tsumaps-neam.eu/). This was also done by publishing NEAMTHM18 on the EPOS ICS-C (https://www.ics-c.epos-eu.org/), along with other services provided by the newly formed candidate TCS Tsunamis [119], as discussed below. These websites fully illustrate the project development and provide links to access the project documentation. The most relevant web pages are.

- https://www.tsumaps-neam.eu/documentation/ provides access to project progression, technical reports, other dissemination material, and all the PHASEs 1 & 2 documents.
- https://www.tsumaps-neam.eu/publications/ lists the relevant articles published in peer-reviewed journals.
- https://www.tsumaps-neam.eu/neamthm18/ is the landing page associated with the Digital Object Identifier minted by INGV through DataCite on September 12, 2019 (https://datacite.org/). This page includes 1) links to the interactive mapper where the hazard and probability maps can be navigated, interrogated, and tsunami hazard data downloaded; 2) links to hazard data distributed through the Open Geospatial Consortium standard protocols via the INGV platform https://tsumaps-neam.eu/web-services/; 3) link to the relevant metadata associated with the dataset (https://data.ingv.it/en/dataset/309) and stored in the INGV Open Data Portal; 4) the license terms of use (CC BY 4.0); 5) citation and abstract. All these elements guarantee that the project results remain persistently findable, accessible, interoperable, and reusable (FAIR).

Full access to the hazard model results and output data was obtained through a dedicated Interactive Hazard Curve Tool advertised to the larger tsunami community and the public.

To guarantee reproducibility, codes, input, and intermediate data were stored in a repository for long-term maintenance and made available upon request. In the meantime, the model was also presented and discussed at several international Congresses.

5. Final remarks

The MEP framework presented in this paper provides a systematic and transparent workflow for multiple-expert interaction in managing uncertainty and subjectivity within single to multiple hazard and risk projects, with a primary focus on natural hazards. We successfully applied this protocol and demonstrated its versatility of MEP through two case studies: (i) a multi-hazard risk analysis for the port of Thessaloniki in the FP7 STREST project [59,60] and (ii) the NEAM seismic probabilistic tsunami hazard model, in the DG-ECHO TSUMAPS-NEAM project [87].

In a nutshell, MEP facilitates a formal decision-making process through consensus among multiple experts, ensuring transparency and independence and clearly establishing responsibility, accountability and ownership of results. MEP's key actors include a Project Manager (PM) managing the project and represent the end-users, a Technical Integrator (TI) who coordinates the scientific aspects of the project, an Implementation Team (IT) supporting the scientific implementations, a Pool of Experts (PoE) representing the larger technical community to support decision-making on subjective matters, and an internal Participatory Peer Review Panel (iPPRP) providing a participatory review. MEP consists of three consecutive phases: the pre-assessment phase (Phase 1), in which the method is defined; the assessment phase (Phase 2), in which the preliminary results are produced; and the finalization phase (Phase 3), in which a formal agreement is reached and results are disseminated.

The two case studies highlight the overall flexibility of MEP and its capability to adapt to multiple tasks and deal with all the main issues that may emerge during such projects. For example, in the Thessaloniki case study, the role of PoE was fundamental in guiding and making objective the selection of the hazards to be considered in the multi-hazard analysis and of the main elements and interactions to be included in the systemic vulnerability model. In practice, there was an infinite spectrum of possible choices. The fact of starting the implementation from an external and formalized opinion, independent from the project management (PM and TI), guaranteed that their choices were not based on the existing capability of the implementation group. Indeed, without this external input, PM and TI more or less consciously may select those hazards and those components whose competences are already covered in the project, rather than on their actual importance. This process also guaranteed that the adopted simplifications were explicitly discussed, favouring future scrutiny for improving subsequent applications.

MEP allows for dynamic management of the complexity that may vary depending on the complexity of the assessment, as well as the available budget, the deepening level requested by the end-users, and the level of disagreements among experts. This aspect was strikingly important in the NEAMTHM18 development. During the iPPRP and PoE kick-off meeting held in Athens, two important requirements emerged from the iPPRP: the need to model the background seismicity (earthquakes outside subduction zones, typically not considered in tsunami hazard, but ultimately fundamental in small seas like the Mediterranean) and the requirement of considering alternatives to quantify the rates of subduction zones. These requirements were difficult to reach within the budget and temporal constraints of the project. This forced the discussion toward practical proposals. In particular, for the background seismicity, a prioritization action was introduced to define the extension of the background seismicity areas, allowing for a strong reduction of the computational effort required for modelling the potential tsunamis. For the rates, several PoE members proposed the introduction in the preliminary NEAMTHM18 model (the one proposed at the end of PHASE 1) of an existing model previously developed for another global hazard model [94]. This model was introduced within the range of the proposed alternative hazard models to be implemented in NEAMTHM18 and ultimately resulted as one of the most weighted in the ensemble [87,112]. All these implementations were initially unforeseen in the project kickoff meeting: they helped the process of convergence toward an overall procedural consensus, also increasing the awareness of end-users on the limits of the developed model, as they actively witnessed the entire revision/decision process.

These applications also demonstrated that one of the added values of the full traceability of the entire decision-making process over subjective matters is that, by increasesing the awareness of end-users, makes MEP budget-conscious, in that the targets and the limitations for evaluation are explicit and established based on the effective needs of the entity funding the hazard/risk assessment.

Finally, it is worth noting that MEP's technological neutrality made it applicable across various hazard, risks and scales, since it provides a flexible workflow rather than rigid prescriptions. This allows for a uniform process for diverse 'clients', such as industries and communities, to develop comparable hazard and risk assessments. This is crucial for developing transparent regulations and public policy decisions across large geographic regions or globally. In this sense, MEP contributes towards safe and reliable complex technological systems and effective regional and global risk governance, aligning with the goals of the Sendai framework for disaster risk reduction of the United Nations Office for Disaster Risk Reduction (UNDRR) [121].

CRediT authorship contribution statement

J. Selva: Conceptualization, Formal analysis, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. **S. Argyroudis:** Conceptualization, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **F. Cotton:** Conceptualization, Methodology, Project administration, Writing – review & editing, Funding acquisition. **S. Esposito:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **S.M. Iqbal:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **S. Lorito:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Validation, Writing – review & editing. **B. Stojadinovic:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Writing – review & editing. **R. Basili:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Validation, Writing – review & editing. **A. Hoechner:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **A. Mignan:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Writing – review & editing. **K. Pitilakis:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Validation, Writing – review & editing. **H.K. Thio:** Conceptualization, Formal analysis, Methodology, Writing – review & editing. **D. Giardini:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors thank the anonymous reviewers for their constructive comments and criticism, which helped improve the paper. The work presented in this paper was conducted within the project "STREST: Harmonized approach to stress tests for civil infrastructures against natural hazards", funded by the European Community's Seventh Framework Programme under Grant Agreement No. 603389, and the project The NEAMTHM18 was prepared in the framework of the European Project TSUMAPS-NEAM (http://www.tsumaps-neam.eu/) funded by the mechanism of the European Civil Protection and Humanitarian Aid Operations with grant no. ECHO/SUB/2015/718568/PREV26 (https://ec.europa.eu/echo/funding-evaluations/financing-civil-protection-europe/selected-projects/probabilistic-tsunami-hazard_en).

References

[1] C.A. Cornell, H. Krawinkle, Progress and challanges in seismic performance assessment, PEER Cent. News 3 (2000).
[2] A. Der Kiureghian, Non-ergodicity and PEER's framework formula, Earthq. Eng. Struct. Dynam. 34 (2005) 1643–1652, https://doi.org/10.1002/eqe.504.
[3] M.C. Gerstenberger, W. Marzocchi, T. Allen, M. Pagani, J. Adams, L. Danciu, E.H. Field, H. Fujiwara, N. Luco, K.F. Ma, C. Meletti, M.D. Petersen, Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges, Rev. Geophys. 58 (2020), https://doi.org/10.1029/2019RG000653.
[4] A. Grezio, A. Babeyko, M.A. Baptista, J. Behrens, A. Costa, G. Davies, E.L. Geist, S. Glimsdal, F.I. González, J. Griffin, C.B. Harbitz, R.J. LeVeque, S. Lorito, F. Løvholt, R. Omira, C. Mueller, R. Paris, T. Parsons, J. Polet, W. Power, J. Selva, M.B. Sørensen, H.K. Thio, Probabilistic tsunami hazard analysis: multiple sources and global applications, Rev. Geophys. 55 (2017) 1158–1198, https://doi.org/10.1002/2017RG000579.
[5] J. Woessner, D. Laurentiu, D. Giardini, H. Crowley, F. Cotton, G. Grünthal, G. Valensise, R. Arvidsson, R. Basili, M.B. Demircioglu, S. Hiemer, C. Meletti, R.W. Musson, A.N. Rovida, K. Sesetyan, M. Stucchi, The SHARE Consortium, in: The 2013 European Seismic Hazard Model: Key Components and Results. Bull. Earthq. Eng., 13, 2015, pp. 3553–3596, https://doi.org/10.1007/s10518-015-9795-1.
[6] D.L. Bernreuter, J.B. Savy, R.W. Mensing, J.C. Chen, Seismic Hazard Characterization of 69 Nuclear Plant Sites East of the Rocky Mountains: Results and Discussion for the Batch 4 Sites (No. NUREG/CR--5250-VOL.5), Nuclear Regulatory Commission, United States, 1989.
[7] W. Marzocchi, M. Taroni, J. Selva, Accounting for epistemic uncertainty in PSHA: logic tree and ensemble modeling, Bull. Seismol. Soc. Am. 105 (2015) 2151–2159, https://doi.org/10.1785/0120140131.
[8] T.N. Palmer, G.J. Shutts, R. Hagedorn, F.J. Doblas-Reyes, T. Jung, M. Leutbecher, Representing model uncertainty in weather and climate prediction, Annu. Rev. Earth Planet Sci. 33 (2005) 163–193, https://doi.org/10.1146/annurev.earth.33.092203.122552.
[9] M.E. Paté-Cornell, Uncertainties in risk analysis: six levels of treatment, Reliab. Eng. Syst. Saf. 54 (1996) 95–111, https://doi.org/10.1016/S0951-8320(96)00067-1.
[10] J. Selva, S. Argyroudis, K. Pitilakis, Impact on loss/risk assessments of inter-model variability in vulnerability analysis, Nat. Hazards 67 (2013) 723–746, https://doi.org/10.1007/s11069-013-0616-z.
[11] K.J. Beven, S. Almeida, W.P. Aspinall, P.D. Bates, S. Blazkova, E. Borgomeo, J. Freer, K. Goda, J.W. Hall, J.C. Phillips, M. Simpson, P.J. Smith, D.B. Stephenson, T. Wagener, M. Watson, K.L. Wilkins, Epistemic uncertainties and natural hazard risk assessment – Part 1: a review of different natural hazard areas, Nat. Hazards Earth Syst. Sci. 18 (2018) 2741–2768, https://doi.org/10.5194/nhess-18-2741-2018.
[12] R. Budnitz, G. Apostolakis, D. Boore, L. Cluff, K. Coppersmith, C.A. Cornell, P. Morris, Senior Seismic Hazard Analysis Committee (SSHAC): Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts: Main Report, 1997 No. NUREG/CR-6372), NUREG/CR-6372.
[13] W. Marzocchi, T.H. Jordan, Testing for ontological errors in probabilistic forecasting models of natural systems, Proc. Natl. Acad. Sci. USA 111 (2014) 11973–11978, https://doi.org/10.1073/pnas.1410183111.
[14] P. Van Der Keur, C. Van Bers, H.J. Henriksen, H.K. Nibanupudi, S. Yadav, R. Wijaya, A. Subiyono, N. Mukerjee, H.-J. Hausmann, M. Hare, C.T. Van Scheltinga, G. Pearn, F. Jaspers, Identification and analysis of uncertainty in disaster risk reduction and climate change adaptation in South and Southeast Asia, Int. J. Disaster Risk Reduc. 16 (2016) 208–214, https://doi.org/10.1016/j.ijdrr.2016.03.002.
[15] J.P. Ake, C. Munson, J. Stamatakos, M. Juckett, K. Coppersmith, J. Bommer, Updated Implementation Guidelines for SSHAC Hazard Studies (No. NUREG-2213), U.S. Nuclear Regulatory Commission, 2018.
[16] J.J. Bommer, Challenges of building logic trees for probabilistic seismic hazard analysis, Earthq. Spectra 28 (2012) 1723–1735, https://doi.org/10.1193/1.4000079.
[17] A.M. Kammerer, J.P. Ake, Practical Implementation Guidelines for SSHAC Level 3 and 4 Hazard Studies (Office of Nuclear Regulatory Research No. NUREG-2117), 2012.
[18] A. Neri, A. Bevilacqua, T.E. Ongaro, R. Isaia, W.P. Aspinall, M. Bisson, F. Flandoli, P.J. Baxter, A. Bertagnini, E. Iannuzzi, S. Orsucci, M. Pistolesi, M. Rosi, S. Vitale, Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: 2. Pyroclastic density current invasion maps, J. Geophys. Res. Solid Earth 120 (2015) 2330–2349, https://doi.org/10.1002/2014JB011776.
[19] J.J. Bommer, F. Scherbaum, The use and misuse of logic trees in probabilistic seismic hazard analysis, Earthq. Spectra 24 (2008) 997–1009, https://doi.org/10.1193/1.2977755.
[20] J. Rougier, S. Sparks, L. Hill (Eds.), Risk and Uncertainty Assessment for Natural Hazards, Cambridge University Press, Cambridge, 2013, https://doi.org/10.1017/CBO9781139047562.
[21] M. Araujo, M. New, Ensemble forecasting of species distributions, Trends Ecol. Evol. 22 (2007) 42–47, https://doi.org/10.1016/j.tree.2006.09.010.
[22] H.L. Cloke, F. Pappenberger, Ensemble flood forecasting: a review, J. Hydrol. 375 (2009) 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005.
[23] T. Gneiting, A.E. Raftery, Weather forecasting with ensemble methods, Science 310 (2005) 248–249, https://doi.org/10.1126/science.1115255.
[24] Intergovernmental Panel On Climate Change, in: Climate Change 2013 – the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, first ed., Cambridge University Press, 2014, https://doi.org/10.1017/CBO9781107415324.
[25] W.M.O. IPCC (Ed.), Climate Change: the 1990 and 1992 IPCC Assessments, IPCC First Assessment Report Overview and Policymaker Summaries and 1992 IPPC Supplement. IPCC, Geneve, 1992.
[26] F. Molteni, R. Buizza, T.N. Palmer, T. Petroliagis, The ECMWF ensemble prediction system - methodology and validation, https://doi.org/10.21957/JL1B71OQJ, 1994.
[27] Z. Toth, E. Kalnay, Ensemble forecasting at NMC: the generation of perturbations, Bull. Am. Meteorol. Soc. 74 (1993) 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.
[28] R. Kulkarni, R. Youngs, K. Coppersmith, Assessment of confidence intervals for results of seismic hazard analysis, in: Proceedings of the Eighth World Conference on Earthquake Engineering. Presented at the Eighth World Conference on Earthquake Engineering, 1984, pp. 263–270 San Francisco, California.
[29] J. Selva, A. Costa, G. De Natale, M.A. Di Vito, R. Isaia, G. Macedonio, Sensitivity test and ensemble hazard assessment for tephra fallout at Campi Flegrei, Italy, J. Volcanol. Geoth. Res. 351 (2018) 1–28, https://doi.org/10.1016/j.jvolgeores.2017.11.024.
[30] W.P. Aspinall, G. Woo, B. Voight, P.J. Baxter, Evidence-based volcanology: application to eruption crises, J. Volcanol. Geoth. Res. 128 (2003) 273–285, https://doi.org/10.1016/S0377-0273(03)00260-9.
[31] T.K. Hincks, J.-C. Komorowski, S.R. Sparks, W.P. Aspinall, Retrospective analysis of uncertain eruption precursors at La Soufrière volcano, Guadeloupe, 1975–77: volcanic hazard assessment using a Bayesian Belief Network approach, J. Appl. Volcanol. 3 (2014) 3, https://doi.org/10.1186/2191-5040-3-3.
[32] W. Marzocchi, L. Sandri, J. Selva, BET_EF: a probabilistic tool for long- and short-term eruption forecasting, Bull. Volcanol. 70 (2008) 623–632, https://doi.org/10.1007/s00445-007-0157-y.
[33] W. Marzocchi, L. Sandri, P. Gasparini, C. Newhall, E. Boschi, Quantifying probabilities of volcanic events: the example of volcanic hazard at Mount Vesuvius: quantifying volcanic hazard, J. Geophys. Res. Solid Earth 109 (2004), https://doi.org/10.1029/2004JB003155.
[34] K.J. Beven, W.P. Aspinall, P.D. Bates, E. Borgomeo, K. Goda, J.W. Hall, T. Page, J.C. Phillips, M. Simpson, P.J. Smith, T. Wagener, M. Watson, Epistemic uncertainties and natural hazard risk assessment – Part 2: what should constitute good practice? Nat. Hazards Earth Syst. Sci. 18 (2018) 2769–2783, https://doi.org/10.5194/nhess-18-2769-2018.
[35] E.E.H. Doyle, D.M. Johnston, R. Smith, D. Paton, Communicating model uncertainty for natural hazards: a qualitative systematic thematic review, Int. J. Disaster Risk Reduc. 33 (2019) 449–476, https://doi.org/10.1016/j.ijdrr.2018.10.023.
[36] W. Aspinall, A route to more tractable expert advice, Nature 463 (2010) 294–295, https://doi.org/10.1038/463294a.
[37] R.M. Cooke, Experts in uncertainty: opinion and subjective probability in science. Environmental Ethics and Science Policy Series, Oxford university press, New York Oxford, 1991.
[38] H.A. Linstone, M. Turoff (Eds.), The Delphi Method: Techniques and Applications, Addison-Wesley Pub. Co., Reading, Mass, 1975 Advanced Book Program.
[39] W. Marzocchi, J.D. Zechar, Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model, Seismol Res. Lett. 82 (2011) 442–448, https://doi.org/10.1785/gssrl.82.3.442.

[40] M.G. Morgan, Use (and abuse) of expert elicitation in support of decision making for public policy, Proc. Natl. Acad. Sci. USA 111 (2014) 7176–7184, https://doi.org/10.1073/pnas.1319946111.
[41] C.A. Cornell, Engineering seismic risk analysis, Bull. Seismol. Soc. Am. 58 (1968) 1583–1606.
[42] T.C. Hanks, N.A. Abrahamson, D.M. Boore, K.J. Coppersmith, N.E. Knepprath, Implementation of the SSHAC Guidelines for Level 3 and 4 PSHAs--Experience Gained from Actual Applications (Open-File Report No. 2009–1093), U.S. Geological Survey Open-File Report, 2009.
[43] W.P. Aspinall, R.M. Cooke, Quantifying scientific uncertainty from expert judgement elicitation, in: J. Rougier, S. Sparks, L. Hill (Eds.), Risk and Uncertainty Assessment for Natural Hazards, Cambridge University Press, Cambridge, 2013, pp. 64–99, https://doi.org/10.1017/CBO9781139047562.005.
[44] T. Bedford, R. Cooke, Probabilistic Risk Analysis: Foundations and Methods, first ed., Cambridge University Press, 2001, https://doi.org/10.1017/CBO9780511813597.
[45] R. Cooke, Experts in uncertainty: opinion and subjective probability, in: Nachdr science (Ed.), Environmental Ethics and Science Policy Series, Oxford Univ. Press, New York, 2011.
[46] R.M. Cooke, L.J.H. Goossens, Procedures Guide for Structured Expert Judgment: Final Report, EUR Nuclear Science and Technology, Off. for Off. Publ. of the Europ. Communities, Luxembourg, 2000.
[47] S.C. Hora, Expert Judgment in Risk Analysis, 2009.
[48] S.C. Hora, Eliciting probabilities from experts, in: W. Edwards, R.F.Jr. Miles, D. von Winterfeldt (Eds.), Advances in Decision Analysis, Cambridge University Press, Cambridge, 2007, pp. 129–153, https://doi.org/10.1017/CBO9780511611308.009.
[49] A.B. Knol, P. Slottje, J.P. van der Sluijs, E. Lebret, The use of expert elicitation in environmental health impact assessment: a seven step procedure, Environ. Health 9 (2010) 19, https://doi.org/10.1186/1476-069X-9-19.
[50] W.P. Aspinall, Structured elicitation of expert judgement for probabilistic hazard and risk assessment in volcanic eruptions, in: H.M. Mader, S.G. Coles, C.B. Connor, L.J. Connor (Eds.), Statistics in Volcanology. The Geological Society of London on Behalf of the International Association of Volcanology and Chemistry of the Earth's Interior, 2006, pp. 15–30, https://doi.org/10.1144/IAVCEI001.2.
[51] J. Selva, W. Marzocchi, P. Papale, L. Sandri, Operational eruption forecasting at high-risk volcanoes: the case of Campi Flegrei, Naples. J. Appl. Volcanol. 1 (2012) 5, https://doi.org/10.1186/2191-5040-1-5.
[52] E.L. Melnick, B. Everitt (Eds.), Encyclopedia of Quantitative Risk Analysis and Assessment, John Wiley, Chichester, West Sussex, England ; Hoboken, N.J, 2008.
[53] R.L. Winkler, R.T. Clemen, Multiple experts vs. Multiple methods: combining correlation assessments, Decis. Anal. 1 (2004) 167–176, https://doi.org/10.1287/deca.1030.0008.
[54] V. Mahajan, H.A. Linstone, M. Turoff, The Delphi method: techniques and applications, J. Mark. Res. 13 (1976) 317, https://doi.org/10.2307/3150755.
[55] A. Babuscia, K.-M. Cheung, An approach to perform expert elicitation for engineering design risk analysis: methodology and experimental results, J. R. Stat. Soc. Ser. A Stat. Soc. 177 (2014) 475–497, https://doi.org/10.1111/rssa.12028.
[56] C.M. Goodman, The Delphi technique: a critique, J. Adv. Nurs. 12 (1987) 729–734, https://doi.org/10.1111/j.1365-2648.1987.tb01376.x.
[57] J.-U. Klügel, Problems in the application of the SSHAC probability method for assessing earthquake hazards at Swiss nuclear power plants, Eng. Geol. 78 (2005) 285–307, https://doi.org/10.1016/j.enggeo.2005.01.007.
[58] F. Woudenberg, An evaluation of Delphi, Technol. Forecast. Soc. Change 40 (1991) 131–150, https://doi.org/10.1016/0040-1625(91)90002-W.
[59] S.A. Argyroudis, S. Fotopoulou, S. Karafagka, K. Pitilakis, J. Selva, E. Salzano, A. Basco, H. Crowley, D. Rodrigues, J.P. Matos, A.J. Schleiss, W. Courage, J. Reinders, Y. Cheng, S. Akkar, E. Uçkan, M. Erdik, D. Giardini, A. Mignan, A risk-based multi-level stress test methodology: application to six critical non-nuclear infrastructures in Europe, Nat. Hazards 100 (2020) 595–633, https://doi.org/10.1007/s11069-019-03828-5.
[60] S. Esposito, B. Stojadinović, A. Babič, M. Dolšek, S. Iqbal, J. Selva, M. Broccardo, A. Mignan, D. Giardini, Risk-based multilevel methodology to stress test critical infrastructure systems, J. Infrastruct. Syst. 26 (2020) 04019035, https://doi.org/10.1061/(ASCE)IS.1943-555X.0000520.
[61] I. Linkov, B. Trump, J. Trump, G. Pescaroli, A. Mavrodieva, A. Panda, Stress-test the resilience of critical infrastructure, Nature 603 (2022) https://doi.org/10.1038/d41586-022-00784-2, 578–578.
[62] I. Linkov, B.D. Trump, J. Trump, G. Pescaroli, W. Hynes, A. Mavrodieva, A. Panda, Resilience stress testing for critical infrastructure, Int. J. Disaster Risk Reduc. 82 (2022) 103323, https://doi.org/10.1016/j.ijdrr.2022.103323.
[63] N.A. Abrahamson, K.J. Coppersmith, M. Koller, P. Roth, C. Sprecher, G.R. Toro, R. Youngs, Probabilistic Seismic Hazard Analysis for Swiss Nuclear Power Plant Sites (PEGASOS Project) (No. Final Report), NAGRA, 2004.
[64] C. Meletti, W. Marzocchi, V. D'Amico, G. Lanzano, L. Luzi, F. Martinelli, B. Pace, A. Rovida, M. Taroni, F. Visini, M.W. Group, The new Italian seismic hazard model (MPS19), Ann. Geophys. 64 (2021) 6, https://doi.org/10.4401/ag-8579.
[65] J. Behrens, F. Løvholt, F. Jalayer, S. Lorito, M.A. Salgado-Galvez, M.B. Sørensen, S. Abadie, I. Aguirre-Ayerbe, A.-Q. Iñigo, A. Babeyko, M. Baiguera, R. Basili, S. Belliazzi, A. Grezio, K. Johnson, S. Murphy, R. Paris, I. Rafliana, R. De Risi, T. Rossetto, J. Selva, M. Taroni, M. Del Zoppo, A. Armigliato, V. Bureš, P. Cech, C. Cecioni, P. Christodoulides, G. Davies, F. Dias, H.B. Bayraktar, M. González, M. Gritsevich, S. Guillas, C. Harbitz, U. Kânoğlu, J. Macias Sanchez, G.A. Papadopoulos, J. Polet, F. Romano, A. Salamon, A. Scala, M. Stepinac, D.R. Tappin, H.K. Thio, R. Tonini, I. Triantafyllou, T. Ulrich, E. Varini, M. Volpe, E. Vyhmeister, Probabilistic tsunami hazard and risk analysis – a review of research gaps, Front. Earth Sci. 9 (2021) 628772.
[66] S.A. Argyroudis, S.A. Mitoulis, M.G. Winter, A.M. Kaynia, Fragility of transport assets exposed to multiple hazards: state-of-the-art review toward infrastructural resilience, Reliab. Eng. Syst. Saf. 191 (2019) 106567, https://doi.org/10.1016/j.ress.2019.106567.
[67] A. Mignan, S. Wiemer, D. Giardini, The quantification of low-probability–high-consequences events: part I. A generic multi-risk approach, Nat. Hazards 73 (2014) 1999–2022, https://doi.org/10.1007/s11069-014-1178-4.
[68] I. Iervolino, M. Giorgio, E. Chioccarelli, Markovian modeling of seismic damage accumulation: markovian modeling of seismic damage accumulation, Earthq. Eng. Struct. Dynam. 45 (2016) 441–461, https://doi.org/10.1002/eqe.2668.
[69] S.A. Mitoulis, S.A. Argyroudis, M. Loli, B. Imam, Restoration models for quantifying flood resilience of bridges, Eng. Struct. 238 (2021) 112180, https://doi.org/10.1016/j.engstruct.2021.112180.
[70] S. De Angeli, B.D. Malamud, L. Rossi, F.E. Taylor, E. Trasforini, R. Rudari, A multi-hazard framework for spatial-temporal impact analysis, Int. J. Disaster Risk Reduc. 73 (2022) 102829, https://doi.org/10.1016/j.ijdrr.2022.102829.
[71] J.C. Gill, B.D. Malamud, Reviewing and visualizing the interactions of natural hazards: interactions of Natural Hazards, Rev. Geophys. 52 (2014) 680–722, https://doi.org/10.1002/2013RG000445.
[72] S. Hochrainer-Stigler, R. Šakić Trogrlić, K. Reiter, P.J. Ward, M.C. De Ruiter, M.J. Duncan, S. Torresan, R. Ciurean, J. Mysiak, D. Stuparu, S. Gottardo, Toward a framework for systemic multi-hazard and multi-risk assessment and management, iScience 26 (2023) 106736, https://doi.org/10.1016/j.isci.2023.106736.
[73] M.S. Kappes, M. Keiler, K. Von Elverfeldt, T. Glade, Challenges of analyzing multi-hazard risk: a review, Nat. Hazards 64 (2012) 1925–1958, https://doi.org/10.1007/s11069-012-0294-2.
[74] Z. Liu, F. Nadim, A. Garcia-Aristizabal, A. Mignan, K. Fleming, B.Q. Luna, A three-level framework for multi-risk assessment, Georisk 9 (2015) 59–74, https://doi.org/10.1080/17499518.2015.1041989.
[75] W. Marzocchi, A. Garcia-Aristizabal, P. Gasparini, M.L. Mastellone, A. Di Ruocco, Basic principles of multi-risk assessment: a case study in Italy, Nat. Hazards 62 (2012) 551–573, https://doi.org/10.1007/s11069-012-0092-x.
[76] J. Selva, Long-term multi-risk assessment: statistical treatment of interaction among risks, Nat. Hazards 67 (2013) 701–722, https://doi.org/10.1007/s11069-013-0599-9.
[77] A. Tilloy, B.D. Malamud, H. Winter, A. Joly-Laugel, A review of quantification methodologies for multi-hazard interrelationships, Earth Sci. Rev. 196 (2019) 102881, https://doi.org/10.1016/j.earscirev.2019.102881.
[78] J. Wang, Z. He, W. Weng, A review of the research into the relations between hazards in multi-hazard risk analysis, Nat. Hazards 104 (2020) 2003–2026, https://doi.org/10.1007/s11069-020-04259-3.
[79] S.A. Argyroudis, J. Selva, P. Gehl, K. Pitilakis, Systemic seismic risk assessment of road Networks considering interactions with the built environment: systemic seismic risk assessment of road networks, Comput. Aided Civ. Infrastruct. Eng. 30 (2015) 524–540, https://doi.org/10.1111/mice.12136.
[80] F. Cavalieri, P. Franchin, P. Gehl, B. Khazai, Quantitative assessment of social losses based on physical damage and interaction with infrastructural systems:

quantitative assessment of social losses, Earthq. Eng. Struct. Dynam. 41 (2012) 1569–1589, https://doi.org/10.1002/eqe.2220.
[81] K. Pitilakis, P. Franchin, B. Khazai, H. Wenzel (Eds.), SYNER-G: Systemic Seismic Vulnerability and Risk Assessment of Complex Urban, Utility, Lifeline Systems and Critical Facilities: Methodology and Applications, Geotechnical, Geological and Earthquake Engineering, Springer, Netherlands, Dordrecht, 2014, https://doi.org/10.1007/978-94-017-8835-9.
[82] R.J. Budnitz, G. Apostolakis, D.M. Boore, L.S. Cluff, K.J. Coppersmith, C.A. Cornell, P.A. Morris, Use of technical expert panels: applications to probabilistic seismic hazard analysis, Risk Anal. 18 (1998) 463–469, https://doi.org/10.1111/j.1539-6924.1998.tb00361.x.
[83] E. Paté-Cornell, On "black swans" and "perfect storms": risk analysis and management when statistics are not enough: **on black swans and perfect storms**, Risk Anal. 32 (2012) 1823–1833, https://doi.org/10.1111/j.1539-6924.2011.01787.x.
[84] D. Sornette, G. Ouillon, Dragon-kings: mechanisms, statistical methods and empirical evidence, Eur. Phys. J. Spec. Top. 205 (2012) 1–26, https://doi.org/10.1140/epjst/e2012-01559-5.
[85] I. Linkov, Multi-criteria decision analysis: case studies in engineering and the environment. Environmental Assessment and Management, second ed., CRC Press, Taylor & Francis Group, Boca Raton, FL, 2021.
[86] I. Linkov, E. Moberg, Multi-Criteria Decision Analysis, 0 ed., CRC Press, 2011, https://doi.org/10.1201/b11471.
[87] R. Basili, B. Brizuela, A. Herrero, S. Iqbal, S. Lorito, F.E. Maesano, S. Murphy, P. Perfetti, F. Romano, A. Scala, J. Selva, M. Taroni, M.M. Tiberti, H.K. Thio, R. Tonini, M. Volpe, S. Glimsdal, C.B. Harbitz, F. Løvholt, M.A. Baptista, F. Carrilho, L.M. Matias, R. Omira, A. Babeyko, A. Hoechner, M. Gurbuz, O. Pekcan, A.C. Yalciner, M. Canals, G. Lastras, A. Agalos, G.A. Papadopoulos, I. Triantafyllou, S. Benchekroun, H.A. Jaouadi, S.B. Abdallah, A. Bouallegue, H. Hamdi, F. Oueslati, A. Amato, A. Armigliato, J. Behrens, G. Davies, D. Di Bucci, M. Dolce, E.L. Geist, J.M. Gonzàlez Vida, M. González, J. Macias Sanchez, C. Meletti, C.O. Sozdinler, M. Pagani, T. Parsons, J. Polet, W. Power, M.B. Sørensen, A. Zaytsev, The making of the NEAM tsunami hazard model 2018 (NEAMTHM18), Front. Earth Sci. 8 (2021) 616594, https://doi.org/10.3389/feart.2020.616594.
[88] S. Keeney, F. Hasson, H.P. McKenna, A critical review of the Delphi technique as a research methodology for nursing, Int. J. Nurs. Stud. 38 (2001) 195–200, https://doi.org/10.1016/S0020-7489(00)00044-4.
[89] S. Hoffmann, P. Fischbeck, A. Krupnick, M. McWilliams, Elicitation from large, heterogeneous expert panels: using multiple uncertainty measures to characterize information quality for decision analysis, Decis. Anal. 4 (2007) 91–109, https://doi.org/10.1287/deca.1070.0090.
[90] M. Maida, K. Maier, N. Obwegeser, Pairwise comparison techniques for pr wise comparison techniques for preference elicitation: using ence elicitation: using test-retest reliability as a quality indicator, in: CONF-IRM 2012 PROCEEDINGS 65, 2012.
[91] M.E. Bayraktar, M. Hastak, Bayesian belief network model for decision making in highway maintenance: case studies, J. Construct. Eng. Manag. 135 (2009) 1357–1369, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000111.
[92] I. Molinari, R. Tonini, S. Lorito, A. Piatanesi, F. Romano, D. Melini, A. Hoechner, J.M. Gonzàlez Vida, J. Maciás, M.J. Castro, M. de la Asunción, Fast evaluation of tsunami scenarios: uncertainty assessment for a Mediterranean Sea database, Nat. Hazards Earth Syst. Sci. 16 (2016) 2593–2602, https://doi.org/10.5194/nhess-16-2593-2016.
[93] J. Griffin, M. Gerstenberger, T. Allen, D. Clark, R. Cuthbertson, V.-A. Dimas, G. Gibson, H. Ghasemi, R. Hoult, N. Lam, M. Leonard, T. Mote, M. Quigley, P. Somerville, C. Sinadinovski, M. Stirling, S. Venkatesan, Expert Elicitation of Model Parameters for the 2018 National Seismic Hazard Assessment: Summary of Workshop, Methodology and Outcomes, Geoscience Australia, 2018, https://doi.org/10.11636/Record.2018.028.
[94] G. Davies, J. Griffin, F. Løvholt, S. Glimsdal, C. Harbitz, H.K. Thio, S. Lorito, R. Basili, J. Selva, E. Geist, M.A. Baptista, A global probabilistic tsunami hazard assessment from earthquake sources, Geol. Soc. Lond. Spec. Publ. 456 (5) (2018) 219–244, https://doi.org/10.1144/SP456.
[95] F. Visini, B. Pace, C. Meletti, W. Marzocchi, A. Akinci, R. Azzaro, S. Barani, G. Barberi, G. Barreca, R. Basili, P. Bird, M. Bonini, P. Burrato, M. Busetti, M. Carafa, O. Cocina, R. Console, G. Corti, N. D'Agostino, S. D'Amico, V. D'Amico, M. Dal Cin, G. Falcone, U. Fracassi, R. Gee, V. Kastelic, C. Lai, H. Langer, F. Maesano, A. Marchesini, L. Martelli, C. Monaco, M. Murru, L. Peruzza, M. Poli, S. Pondrelli, A. Rebez, R. Rotondi, A. Rovida, F. Sani, M. Santulin, D. Scafidi, J. Selva, D. Slejko, D. Spallarossa, A. Tamaro, G. Tarabusi, M. Taroni, M. Tiberti, G. Tusa, T. Tuvè, G. Valensise, P. Vannoli, E. Varini, A. Zanferrari, E. Zuccolo, Earthquake rupture forecasts for the MPS19 seismic hazard model of Italy, Ann. Geophys. 64 (2021) 3, https://doi.org/10.4401/ag-8608.
[96] J.J. Bommer, J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, D. Fah, On the selection of ground-motion prediction equations for seismic hazard analysis, Seismol Res. Lett. 81 (2010) 783–793, https://doi.org/10.1785/gssrl.81.5.783.
[97] W.P. Aspinall, G. Woo, Santorini unrest 2011–2012: an immediate Bayesian belief network analysis of eruption scenario probabilities for urgent decision support under uncertainty, J. Appl. Volcanol. 3 (2014) 12, https://doi.org/10.1186/s13617-014-0012-8.
[98] K. Jaiswal, D. Wald, D. Perkins, W. Aspinall, A. Kiremidjian, Estimating structural collapse fragility of generic building typologies using expert judgment, in: G. Deodatis, B. Ellingwood, D. Frangopol (Eds.), Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, CRC Press, 2014, pp. 879–886, https://doi.org/10.1201/b16387-130.
[99] E.H. Field, R.J. Arrowsmith, G.P. Biasi, P. Bird, T.E. Dawson, K.R. Felzer, D.D. Jackson, K.M. Johnson, T.H. Jordan, C. Madden, A.J. Michael, K.R. Milner, M.T. Page, T. Parsons, P.M. Powers, B.E. Shaw, W.R. Thatcher, R.J. Weldon, Y. Zeng, Uniform California earthquake rupture forecast, version 3 (UCERF3)--The time-independent model, Bull. Seismol. Soc. Am. 104 (2014) 1122–1180, https://doi.org/10.1785/0120130164.
[100] D. Albarello, V. D'Amico, Testing probabilistic seismic hazard estimates by comparison with observations: an example in Italy, Geophys. J. Int. 175 (2008) 1088–1094, https://doi.org/10.1111/j.1365-246X.2008.03928.x.
[101] I. Iervolino, E. Chioccarelli, P. Cito, Testing three seismic hazard models for Italy via multi-site observations, PLoS One 18 (2023) e0284909, https://doi.org/10.1371/journal.pone.0284909.
[102] M. Taroni, W. Marzocchi, D. Schorlemmer, M.J. Werner, S. Wiemer, J.D. Zechar, L. Heiniger, F. Euchner, Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy, Seismol Res. Lett. 89 (2018) 1251–1261, https://doi.org/10.1785/0220180031.
[103] P. Bazzurro, C.A. Cornell, Disaggregation of seismic hazard, Bull. Seismol. Soc. Am. 89 (1999) 501–520.
[104] S.M. Rinaldi, J.P. Peerenboom, T.K. Kelly, Identifying, understanding, and analyzing critical infrastructure interdependencies, IEEE Control Syst. Mag. 21 (2001) 11–25, https://doi.org/10.1109/37.969131.
[105] K. Pitilakis, S. Akkar, S. Argyroudis, A. Babic, A. Basco, R. Basili, B. Brizuela, Y. Cheng, W. Courage, H. Crowley, M. Dolšek, M. Erdik, S. Esposito, S. Fotopoulou, D. Giardini, S. Iqbal, S. Karafagka, S. Lorito, J.P. Matos, A. Mignan, A. Piatanesi, J. Reinders, D. Rodrigues, F. Romano, E. Salzano, A.J. Schleiss, B. Stojadinović, R. Tonini, G. Tsionis, E. Uckan, M. Volpe, Guidelines for Stress-Test Design for Non-nuclear Critical Infrastructures and Systems, applications. Publications, Office, LU, 2016.
[106] J. Selva, R. Tonini, I. Molinari, M.M. Tiberti, F. Romano, A. Grezio, D. Melini, A. Piatanesi, R. Basili, S. Lorito, Quantification of source uncertainties in seismic probabilistic tsunami hazard analysis (SPTHA), Geophys. J. Int. 205 (2016) 1780–1803, https://doi.org/10.1093/gji/ggw107.
[107] J. Selva, S. Iqbal, M. Taroni, W. Marzocchi, F. Cotton, W. Courage, L. Abspoel-Bukman, S. Miraglia, A. Mignan, K. Pitilakis, S.A. Argyroudis, K. Kakderi, D. Pitilakis, G. Tsinidis, C. Smerzini, Report on the Effects of Epistemic Uncertainties on the Definition of LP-HC Events, Deliverable 3.1, Project STREST: Harmonized approach to stress tests for critical infrastructures against natural hazards, 2015, p. 119.
[108] R.W. Saaty, The analytic hierarchy process—what it is and how it is used, Math. Model. 9 (1987) 161–176, https://doi.org/10.1016/0270-0255(87)90473-8.
[109] K. Pitilakis, S. Argyroudis, S. Fotopoulou, A. Anastasiadis, D. Pitilakis, D. Raptakis, E. Riga, A. Tsinaris, K. Mara, J. Selva, S. Iqbal, M. Volpe, R. Tonini, F. Romano, B. Brizuela, A. Piatanesi, R. Basili, S. Lorito, E. Salzano, A. Basco, A.J. Schleiss, J.P. Matos, S. Akkar, Y. Cheng, E. Uckan, M. Erdik, W. Courage, J. Reinders, H. Crowley, D. Rodrigues, Integrated report detailing analyses, results and proposed hierarchical set of stress tests for the six CIs covered in STREST, Deliverable 6.1 of project STREST: Harmonized approach to stress tests for critical infrastructures against natural hazards, 2016, p. 298.
[110] G. Tsionis, Final Workshop Report with Conclusions and Recommendations, 2016.
[111] K. Pitilakis, S. Argyroudis, S. Fotopoulou, S. Karafagka, K. Kakderi, J. Selva, Application of stress test concepts for port infrastructures against natural hazards. The case of Thessaloniki port in Greece, Reliab. Eng. Syst. Saf. 184 (2019) 240–257, https://doi.org/10.1016/j.ress.2018.07.005.
[112] R. Basili, B. Brizuela, A. Herrero, S. Iqbal, S. Lorito, F.E. Maesano, S. Murphy, P. Perfetti, F. Romano, A. Scala, J. Selva, M. Taroni, M.M. Tiberti, H.K. Thio, R. Tonini, M. Volpe, S. Glimsdal, C.B. Harbitz, F. Løvholt, M.A. Baptista, F. Carrilho, L.M. Matias, R. Omira, A.Y. Babeyko, A. Hoechner, M. Gurbuz, O. Pekcan, A.C. Yalçıner, M. Canals, G. Lastras, A. Agalos, G.A. Papadopoulos, I. Triantafyllou, S. Benchekroun, H.A. Jaouadi, K. Attafi, S. Ben Abdallah, A. Bouallegue, H. Hamdi, F. Oueslati, NEAMTHM18 documentation: the making of the TSUMAPS-NEAM tsunami hazard model 2018, Zenodo (2019), https://doi.org/10.5281/

zenodo.3406625.
[113] R. Basili, B. Brizuela, A. Herrero, S. Iqbal, S. Lorito, F.E. Maesano, S. Murphy, P. Perfetti, F. Romano, A. Scala, J. Selva, M. Taroni, M.M. Tiberti, H.K. Thio, R. Tonini, M. Volpe, S. Glimsdal, C.B. Harbitz, F. Løvholt, M.A. Baptista, F. Carrilho, L.M. Matias, R. Omira, A.Y. Babeyko, A. Hoechner, M. Gurbuz, O. Pekcan, A.C. Yalçıner, M. Canals, G. Lastras, A. Agalos, G.A. Papadopoulos, I. Triantafyllou, S. Benchekroun, H.A. Jaouadi, K. Attafi, S. Ben Abdallah, A. Bouallegue, H. Hamdi, F. Oueslati, NEAM Tsunami Hazard Model 2018 (NEAMTHM18): Online Data of the Probabilistic Tsunami Hazard Model for the NEAM Region from the TSUMAPS-NEAM Project, Istituto Nazionale di Geofisica e Vulcanologia (INGV), 2018.
[114] IOC/UNESCO, {Intergovernmental Oceanographic Commission, North-Eastern atlantic, the mediterranean and connected seas tsunami early warning and mitigation system. NEAMTWS): 2030 Strategy (No. IOC/2022/TS/171), UNESCO, Paris, France, 2023.
[115] F. Cotton, F. Scherbaum, J.J. Bommer, H. Bungum, Criteria for selecting and adjusting ground-motion models for specific target regions: application to central europe and rock sites, J. Seismol. 10 (2006) 137–156, https://doi.org/10.1007/s10950-005-9006-7.
[116] E. Forman, K. Peniwati, Aggregating individual judgments and priorities with the analytic hierarchy process, Eur. J. Oper. Res. 108 (1998) 165–169, https://doi.org/10.1016/S0377-2217(97)00244-0.
[117] K.D. Goepel, Implementing the analytic hierarchy process as a standard method for multi-criteria decision making in corporate enterprises – a new AHP excel template with multiple inputs. Presented at the the International Symposium on the Analytic Hierarchy Process, 2013, https://doi.org/10.13033/isahp.y2013.047.
[118] E. Zio, On the use of the analytic hierarchy process in the aggregation of expert judgments, Reliab. Eng. Syst. Saf. 53 (1996) 127–138, https://doi.org/10.1016/0951-8320(96)00060-9.
[119] A. Babeyko, S. Lorito, F. Hernandez, J. Lauterjung, F. Løvholt, A. Rudloff, M. Sørensen, A. Androsov, I. Aniel-Quiroga, A. Armigliato, M.A. Baptista, E. Baglione, R. Basili, J. Behrens, B. Brizuela, S. Bruni, D. Cambaz, J. Cantavella Nadal, F. Carillho, I. Chandler, D. Chang-Seng, M. Charalampakis, L. Cugliari, C. Denamiel, G.G. Doğan, G. Festa, D. Fuhrman, A.-A. Gabriel, P. Galea, S. Gibbons, M. González, L. Graziani, M.-A. Gutscher, S. Harig, H. Hebert, C. Ionescu, F. Jalayer, N. Kalligeris, U. Kânoğlu, P. Lanucara, J. Macias Sánchez, S. Murphy, Ö. Necmioğlu, R. Omira, G. Papadopoulos, R. Paris, F. Romano, T. Rossetto, J. Selva, A. Scala, R. Tonini, K. Trevlopoulos, I. Triantafyllou, R. Urgeles, R. Vallone, I. Vilibić, M. Volpe, A. Yalciner, Towards the new thematic core service tsunami within the EPOS research infrastructure, Ann. Geophys. 65 (2022) DM215, https://doi.org/10.4401/ag-8762.
[120] R.L. Boring, D. Gertman, J. Joe, J. Marble, W. Galyean, L. Blackwood, H. Blackman, Simplified expert elicitation procedure for risk assessment of operating events, 2005 https://doi.org/10.2172/911228, No. INL/EXT-05-00433, 911228).
[121] Sendai Famework for Disaster Risk Reduction 2015-2030, UNISDR - United Nations Office for Disaster Risk Reduction, No. UNISDR/GE/2015-ICLUX EN5000), 2015 (Geneve).
[122] O.M. Nofal, K. Amini, J.E. Padgett, J.W. van de Lindt, N. Rosenheim, Y.M. Darestani, L. Duenas-Osorio, Multi-hazard socio-physical resilience assessment of hurricane-induced hazards on coastal communities, Resilient Cities Struct. 2 (2) (2023) 67–81.
[123] A. Tafur, J.E. Padgett, A flow-based commodity-independent port capacity model for resilience assessment of intermodal freight networks subjected to coastal hazards, Reliab. Eng. Syst. Saf. (2024) 110280.