A risk-based multi-level methodology to stress test critical infrastructure systems

Simona Esposito[1] (corresponding author)
Institute of Structural Engineering, Swiss Federal Institute of Technology (ETH) Zürich, Stefano-Franscini-Platz 5, 8093, Zürich, Switzerland, e-mail address: simona.esposito@ibk.baug.ethz.ch

Božidar Stojadinović
Institute of Structural Engineering, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland

Anže Babič
Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia

Matjaž Dolšek
Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia

Sarfraz Iqbal[2]

---

[1] Present affiliation: Swiss Re Management Ltd, Mythenquai 50/60, 8022 Zürich, Switzerland

[2] Present affiliation: Faculty of Technology, Linnaeus University, Växjö, Sweden.

Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy

Jacopo Selva
Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy

Marco Broccardo
Swiss Competence Centers for Energy Research, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland

Arnaud Mignan
Institute of Geophysics, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland

Domenico Giardini
Institute of Geophysics, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland

Abstract

Making communities safer requires better tools to identify, quantify and manage the risks. Among the most important tools are stress tests, originally designed to test the risk posed by nuclear power plants. A complementary harmonized multi-level stress test for non-nuclear civil infrastructure systems against natural hazards is proposed. Each stress test level is characterized by a different scope and by a different level of risk analysis complexity to suite different civil infrastructure systems, different hazards and different risks. The stress test comprises the following phases. First, the goals and the methods for the risk analysis are defined. The test is then performed at the component and the system levels, followed by a verification of the findings. A penalty system is defined to adjust the output of the risk assessment according to the limitations of the risk analysis methods used. The adjusted risk assessment results are then passed to a grading system to determine the outcome of the stress test. Finally, the risk assessment results are reported and the stress test outcomes are communicated to stakeholders and authorities.

Keywords: critical infrastructure systems, stress test, natural hazards, multi-level, penalty system, grading system

INTRODUCTION

Critical infrastructure systems (CIs) are crucial for a modern society: they provide the essential functions of public safety and enable, through their services, the higher-level functions of a community. Natural hazard events can interrupt services, cause damage, or even destroy CIs, triggering disruption of vital socio-economic activities, extensive property damage, human injuries or loss of lives. Consequently, the need to understand and model the risks posed by and the resilience of CIs is increasing. The European Program for Critical Infrastructure Protection (EPCIP) was established in 2006 to ensure a high degree of protection of European infrastructure systems. More recently, the EPCIP (2013) clearly declared the need to develop a tool – a stress test - for critical infrastructure systems as a way to verify the risk and increase the resilience of European CIs in the near future.

To date, the stress test tool has been applied in the nuclear and the banking sectors. In the banking sector, stress testing is used to test scenarios and evaluate how certain factors will affect a company, an industry or a specific portfolio (Quagliariello, 2009). This tool became ubiquitous after the 2007 financial crisis to assess the stability of large institutions or complex financial instruments that may have an impact on the national or global economy. Whereas in the nuclear sector, the European Council in the aftermath of the 2011 Fukushima accident mandated the European Commission to review the safety of all nuclear plants in EU countries and to provide a transparent and comprehensive risk assessment, a *de facto* stress tests, (EUCO 10/1/11, 2011, ENSREG, 2011, Cotton et al., 2016).

In this context, a multi-level method to stress test critical civil and industrial infrastructure systems is proposed herein with the goal to expand the scope and applicability of this important risk management tool from the nuclear and financial sectors to a broader range of infrastructure systems that pose human, environmental or societal risks. Since the risks posed by nuclear and financial systems are quite different in geographic extent and duration compared to non-nuclear, but still critical, infrastructure systems, the proposed stress tests methodology is focused on non-nuclear CIs. The aims of a stress test are to identify and quantify the risk posed by the CIs and their individual components on the community they serve due to possible natural hazard events, to compare the quantified risks to acceptable risk exposure levels, and to guide the affected community to effectively and rationally manage these risks over time. The proposed stress test methodology is based on the best-possible characterization of the extreme and other damaging scenarios and consequences (Cornell and Krawinkler, 2000), including multiple hazards (Selva 2013; Liu et al. 2015; Mignan et al., 2014; 2016; 2017, 2018) and systemic amplification (i.e. cascading) effects (Esposito et al. 2015; Argyroudis et al. 2015). The main aspects of the proposed methodology (Esposito et al. 2016, 2017) were developed in the contest of the European Community FP7 project "STREST: Harmonized approach to stress tests for civil infrastructures against natural hazards" (Tsionis et al. 2016). STREST project reports, including example applications, are available at the project web site (STREST 2018, http://www.strest-eu.org/opencms/opencms/results/). A general overview of the proposed methodology is presented first, followed by a discussion of these key common stress test features:

- The use of a multiple expert management protocol to guarantee the robustness of stress test results by managing the subjectivity and quantifying the epistemic uncertainty.
- The workflow of the stress test process.
- The multi-level framework to provide different levels of risk analysis complexity that may be used to stress test different CI system.
- The penalty system to acknowledge the limitation of the methods and models used to assess the performance of the CI system and, consequently, penalize the output of the risk assessment if needed.
- The grading system to quantify the outcome of the stress test, to plan then next periodic stress test, and to prescribe the degree of safety improvement required before the next planned stress test.

Applications of the proposed stress test stress test method to six different CIs, namely the single-site CIs (an oil refinery, a high-seas port, an alpine tall dam), the geographically extended CIs (an oil pipeline and a natural gas field), and the distributed multi-site CIs (a regional industrial district) and different natural hazards are briefly described with references provided for further detailed examination.

STRESS TEST METHODOLOGY

Considering that CIs differ greatly in the types of natural hazards they are vulnerable to and in the potential consequences, as well as in the available financial and technical resources for conducting a stress tests, a multi-level stress test methodology is proposed. Each Stress Test

Level (ST-L) is characterized by a different scope (component or system) and by a different level of risk analysis complexity (starting from straight-forward design code checks and ending with state-of-the-art single and/or multi-hazard probabilistic risk analyses). The selection of the appropriate ST-Ls may depend on the capabilities of the stakeholders, i.e. the available human and financial resources to perform the stress test and on the regulatory requirements that specify acceptable risk exposure levels based on the different characteristics and importance of the CIs (Esposito et al., 2016). This multi-level structure provides the necessary flexibility to apply the proposed stress test to a broad range of CIs. The proposed stress test follows a workflow consisting of four phases (Figure 1):

- Phase 1: Pre-Assessment, during which the data available on the CI (risk context) and on the phenomena of interest (hazard context) is collected. Then, the goal and objectives, the time frame, the most appropriate ST-L, and the total costs of the stress test are defined.
- Phase 2: Assessment, during which the stress test at the component and the system scope is performed.
- Phase 3: Decision, during which the results of the stress test are analyzed according to the goal and objectives defined in Pre-Assessment phase. Then, critical events (i.e. events that most likely cause the exceedance of a given level of loss) and risk mitigation strategies are identified.
- Phase 4: Report, during which the stress test outcome and risk mitigation guidelines based on the findings established in the Decision Phase are formulated first and then presented to the stakeholders.

Multiple-Expert Interaction and Integration Protocol (MEI^3)

Engaging multiple experts is critical in a risk assessment when potential controversies exist and the regulatory concerns are relatively high (SSHAC, 1997; Woessner et al., 2013; Field et al., 2017). In order to produce robust and stable results, the integration of expert inputs plays a fundamental role in managing subjective decisions and in quantifying the epistemic uncertainty as “the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study” (SSHAC, 1997; Marzocchi et al., 2015). To this end, the diverse range of views and opinions of the experts, their active involvement, and their formal feedbacks need to be organized into a structured process ensuring transparency, accountability and independency. To this end, a formalized multiple expert interaction and integration (MEI^3) protocol has been developed by Selva et al., 2015, Esposito et al., 2016, and Stojadinović et al., 2016, and integrated into the proposed stress test methodology (Figure 1). This protocol is designed to clearly document and manage the data, models and methods adopted for the risk assessment and the associated uncertainty quantification, tracking potential subjective choices and controversies.

In general, several groups of experts may be involved in a stress test, each with different background knowledge and tasks, and a different pre-assigned role in the stress test. The size of such groups depends on the selected ST-L, keeping in mind that the number of experts needed, as well as the cost to engage them, increases with the increase of ST-L. The stress test actors are: Project Manager (PM, a stakeholder responsible and accountable for the successful implementation of the stress test), Technical Integrator (TI, a lead analyst responsible and accountable for the scientific management of the stress test), Evaluation Team (ET, a group of analysts that perform the risk assessment), Pool of Experts (PoE, a group of independent experts tasked with providing blind quantitative input to the TI and PM for managing key critical choices/issues), and Internal Reviewer(s) (IR, an expert or a group of experts selected internally by PM and TI and tasked with an independent participatory peer reviews of the stress test).

The proposed MEI^3 protocol defines the role of these actors and structures their interactions during a stress test to ensure transparency and accountability (Figure 1). PM interacts mainly with the TI and defines the questions that a stress test should answer, considering the technical and societal aspects, selects the appropriate ST Level, and defines the appropriate hazard levels and risk acceptance criteria. The TI manages the scientific process, coordinates the ET in the design and implementation of the scientific and engineering analyses, and has the responsibility for making key decisions. The ET includes a number of experts from different fields who interact to combine their expertise to cover the entire scope of the stress test (e.g. natural hazards, infrastructure system operation, consequences, risk evaluation, including resource, proponent and evaluator experts of SSHAC). If needed (depending on the ST level), the TI also organizes the PoE, whose blind inputs are used to strengthen decision making on controversial, but critical, issues. To structure this interaction, the TI organizes formal elicitation experiments with the PoE and aggregates their opinions through conventional aggregation techniques (e.g. Morgan 2014), providing trackable blind inputs to the ET. The IR reviews the entire stress test process starting from its earliest stages in order to guarantee its transparency, ensure its fairness, and maximize the reliability of the stress test results. Thus, under the coordination of the TI, the ET integrate PoE input and IR feedback to form the scientific and engineering basis of the stress test.

To ensure independence and set the stress test team hierarchy, the CI authorities (representatives of the private or the public owners of the CI, e.g. chief operating officer or a public official in charge of ensuring CI services for the community) select the PM. The PM selects the TI and IR. The members of the ET and of the PoE are consensually selected by TI and PM. PM and TI are, in principle, individuals. The ET and IR may involve several participants, with different background knowledge, but in specific cases may be reduced to individuals. The PoE is, by definition, a group of experts formed *ad hoc* to cover the fields of expertise required for meaningful expert elicitation experiments. The PoE may be also organized in thematic sub-groups (e.g., hazard and infrastructure system experts), if required. In all cases, the size of groups depends on the complexity of the stress test and the resources available for it (Esposito et al. 2016, Stojadinović et al., 2016).

Participation of the different actors significantly changes in different phases of a stress test. The PM and TI/ET are the most active participants in the workflow. The PM and the TI participate in all of the stress test steps. The TI is supported by the PM and assisted by the ET, but the level of assistance depends on the ST-L. The PoE is present only at the most complex (highest) stress test levels. If present, the PoE participates in the Assessment and Decision phases through structured expert elicitation experiments and individual interaction with the TI. The IR performs a peer review at the end of the Pre-Assessment and Decision phases. The final conclusions of the stress test are agreed between PM and TI and must include a discussion based on IR feedback. More details on the interaction among the experts and the rationale behind the adopted process are presented in Selva et al. (2015), Esposito et al. (2016) and Stojadinović et al. (2016).

Stress Test Workflow

The workflow is structured in four phases: 1) Pre-Assessment, 2) Assessment, 3) Decision, and 4) Report phases (Figure 1). Each phase is subdivided into a number of steps, with a total of nine steps in a stress test.

Phase 1: Pre-Assessment

STEP 1 - Data collection. First, the PM TI, ET and IR are selected following the MEI[3] protocol described above. Then, TI and the ET collect the data available on the CI and on the phenomena of interest, as well as the data from relevant stress tests performed in the past (on the same CI, on similar CIs in different locations, and on other CIs at the same location).

STEP 2 - Risk Measures and Acceptance Criteria. The PM defines one or more risk metrics (e.g. number of fatalities, magnitude of economic losses), measures (e.g. expected loss over a period of time, annual exceedance probability of an event, etc.) and acceptance criteria (e.g. frequency of a number of casualties, threshold level of loss), based on the regulatory requirements, the technical and societal considerations, and the outcomes of previous stress tests.

STEP 3 - Setup of Stress Test. The time frame, the total costs of the stress test, and the complexity of the stress tests are defined. The date to present of the outcome of the stress test (in the Report phase) is also set: this date is intended to remain fixed to bound the cost and effort of the stress test. Then, based on the regulatory requirements, the PM selects the appropriate ST-L to define the technical implementation of following stress test phases, as well as the Level of Detail with which to conduct the modeling and analysis elements of the stress test, and the associated risk analysis Penalty Factors. All these steps are finally set after a first written revision of the IR.

Note that *STEP 3* may take some time, and may differ substantially depending on the selected ST-L. If the selected ST-Level foresees the presence of the PoE, the PM and TI agree the initial costs and time frame for the assessments to be performed in *STEP 3*. The PoE work is initialized through a kick-off meeting, followed by the first structured expert elicitation to provide input to the TI for selecting the target single and/or multiple hazards, evaluation the relevance of CI components and the importance of system-level risk and scenario analysis with respect to interdependencies of CI components (functional, co-location, etc.), and for identifying the main sources of epistemic uncertainty. If significant disagreements emerge from the first PoE elicitation, the TI may promote further topical discussions. The PM and TI make a first decision based on the PoE feedbacks, and plan the next tasks (e.g. further PoE elicitations, complementary scenario-based assessments). If, however, the selected ST-Level does not foresee the presence of the PoE, the process becomes simpler (even if less transparent and robust) since all critical decisions are taken directly by the TI, who selects the target hazards and the relevant CI components and systems. In either case, the TI coordinates the ET in collecting the applicable models and data needed for the stress test. Based on this, the TI and PM jointly identify the Level of Detail in modeling and analyses conducted the Assessment phase. Additional sensitivity analyses may be performed to better support this decision.

Throughout, the activities of the Pre-Assessment phase are documented by the ET. The IR reviews these documents and provides feedback that is included in the final version of the stress test plan. Based on this, the final costs and the exact timing for the Assessment and Decision phases are established. Note that, based on the IR review, the PM and TI may add new experts to the ET, as well as re-evaluate the Level of Detail of modeling and analysis the Assessment phase of the stress test in order to avoid penalties that may be suggested by the IR. In fact, if the Level of Detail used in the stress test is lower than minimum required, a Penalty System is applied to final stress test risk assessments in the Decision phase to modulate the stress test outcome (see *STEP 6*).

Phase 2: Assessment

STEP 4 - Component Level Assessment. Performance of each component of the CI is checked by the ET using a hazard-based assessment, a design-based assessment or a risk-based assessment approach, as defined by the selected ST-L.

STEP 5 - System Level Assessment. First, ET implements the CI risk assessment models specified in *STEP 3*. If PoE is in place, the TI organizes a second round of PoE structured expert elicitations to fill potential methodological gaps, to identify additional scenarios to examine, and/or to rank/score/weight different models to quantify the epistemic uncertainty. If PoE is not in place but the treatment of epistemic uncertainty is required, the TI and the ET directly assigns scores/ranks/weights to the selected models. Finally, the ET performs the required risk analyses and risk assessments.

Phase 3: Decision

STEP 6 - Risk Acceptance Check and Grading. Results of risk assessments obtained in *STEP 4* and *STEP 5* are compared to the risk acceptance criteria defined in *STEP 2* by the TI and the ET. The analysis and its results are preliminarily reviewed by the IR. Depending on the type of risk metrics, measures and acceptance criteria, as well as on the ST-L and the risk analysis methods used, comparison of the assessed risks to the acceptance criteria may differ. It may range from safety factor or load and resistance factor demand-capacity comparison used in conventional design codes to evaluation of the distance between the acceptance criteria and the computed risk curve. The CI stress test grade is determined using the stress test penalty (Esposito et al. 2016, Stojadinović et al., 2016) and grading systems (Babič and Dolšek, 2016, 2019) presented later.

STEP 7 - Disaggregation/Sensitivity Analysis. Critical events, components and system features (i.e. most likely causes of exceedance of the considered risk acceptance criteria) are identified based on the result of component (*STEP* 4) and system (*STEP* 5) assessment. Sensitive components and systems may be identified using the experience gained during the operation of the CI, as well as the experience of the TI and ET from previous stress tests or from PoE feedbacks. The critical events may be rigorously identified through a disaggregation analysis (Iervolino, 2016; Esposito et al., 2015; Bazurro and Cornell, 1999). For example, the loss may be disaggregated with respect to system response to identify the component whose damage most likely causes the exceedance of the loss value of interest. Scenario-based analyses are another way to rigorously identify critical events, components and system features using the PoE in an elicitation exercise. As in *STEP 5*, if technical problems emerge during the risk disaggregation and sensitivity analyses, the TI may solve them through individual interactions with the PoE (if present) and ET. It is recommended to identify the critical events, components and system features, especially if the CI does not pass the stress test.

STEP 8 – Risk Mitigation Guidelines. Risk mitigation strategies and guidelines are formulated based on the result of component (*STEP* 4) and system (*STEP* 5) assessments and identification of critical events, components and systems features in *STEP 7*. The guidelines are prepared by the TI and ET and preliminarily reviewed by the IR.

Phase 4: Report

STEP 9 - Results Presentation. The work in the Assessment and Decision phases of the stress test is documented, compared to the stress test plan developed in the Pre-Assessment phase, and a preliminary stress test report is prepared by the ET. The IR formally review this report and the compiled data and analyses results, providing a written feedback to the TI. The ET, coordinated by the TI, explicitly lists and considers this feedback, accordingly updates the preliminary stress test report, and completes the supporting documentation. Based on these documents, the PM and the TI determine the stress test outcome. The stress test outcome, in terms of the stress test grade, and the stress test reports are publicly presented to CI Authorities, regulators and the community by the PM and the TI.

Stress test levels

Non-nuclear CIs, such as transportation, manufacturing, petro-chemical, power-generation, energy transfer and storage, communication or water storage and supply are very diverse: potential range of consequences of a failure of these CIs and the types of hazards they are vulnerable to vary greatly, as do the capabilities and the available resources for conducting a stress test. Therefore, it is not possible or reasonable to require the most complex form of a stress test in all situations. To facilitate conducting stress tests across a broad range of non-nuclear CIs, three stress test levels, characterized by a different scope and by a different complexity of the involved hazard and risk analyses (Figure 2) are proposed:

- Stress Test Level 1 (ST-L1): single-hazard CI component-only check;
- Stress Test Level 2 (ST-L2): single-hazard CI system-wide risk assessment; and
- Stress Test Level 3 (ST-L3): multi-hazard CI system-wide risk assessment.

The aim of ST-L1 is to check each component of a CI system independently to determine if the component passes or fails the minimum requirements for its performance, which are usually defined by the CI design codes and operation and maintenance guidelines current at the time of the stress test. ST-L1 is compulsory and, by default, a part of system-wide stress tests (ST-L2 and ST-L3). This stress test level is compulsory because design of most CI components is regulated by design codes or operation and maintenance guidelines, and the data and the expertise to perform theses component-level checks are readily available. However, system-wide stress tests are not compulsory because such tests may require extensive knowledge of both components and systems (that may be difficult to obtain for older and extensively modified CIs) and usually necessitate significant resources (e.g. expert staff, high costs). Nevertheless, system-wide single-hazard (ST-L2) or multi-hazard (ST-L3) risk assessment is highly recommended to reveal the systemic mechanisms that may lead to potentially catastrophic consequences.

Component-level assessment

Component-level assessment in ST-L1 can be done using hazard-based, design-based or the risk-based assessment procedures. These procedures differ in complexity and in required data. Hazard-based component-level assessment is performed by comparing the value of hazard intensity used to design the component (e.g. a building, a pipe, a storage tank) at the time of CI design, to the value of the hazard intensity prescribed in the current regulatory documents (e.g. building code), or to the value of current best-estimate hazard intensity corresponding to a design level exceedance probability. In a design-based assessment, the expert compares the demand, as defined in the current design codes or best knowledge, with the capacity of the component computed using the current design codes or best knowledge. The demands and the capacities can be expressed in terms of forces, stresses, deformations or displacements. Design-based assessment can be done by factoring the data from the existing design documentation or by performing assessment of the component according to current state of practice. Finally, risk-based assessment is performed by convolving the hazard curve at the location of the component and the fragility function of the component, thus obtaining the probability of exceedance of a designated limit state in a period of time (P_{LS}). The outcomes of the hazard-based and the design-based assessments are qualitative and indicate if the component complies with the current design requirements or not. If such design requirements do not exist or cannot be verified, the component is assessed as non-compliant. The outcome of the risk-based component assessment is, however, quantitative and expressed using P_{LS}. A more detailed description of the component-level assessment can be found elsewhere (Esposito et al., 2016, Stojadinović et al., 2016).

System-level assessment

System-level assessment may be performed considering a single hazard (ST-L2) or multiple hazards (ST-L3). Regardless of the hazard selected, the aim is to evaluate the performance of the CI as a system, considering the interactions and co-location of components and their function in the CI as a system. The intent is to promote probabilistic risk analysis (PRA), a systematic methodology to evaluate risks associated with every life-cycle aspect of a complex engineered system (Bedford and Cooke, 2001). The final results of a PRA are a risk curve and the associated uncertainties (aleatory and epistemic) representing the frequency of exceeding a given consequence (e.g. a loss value). The specific quantitative PRA method to use depends upon the context in which the risk is placed and upon the system under consideration. A list of possible methods that may be applied to assess the performance and the risk of the CI can be found in Salzano et al., (2016), Kakderi et al., (2015) and Crowley et al., (2015). It should be noted that there are no standardized approaches for multi-risk system-level assessment (for ST-L3) even though different methods have been proposed recently, such as Liu et al. (2015), Marzocchi et al (2012), Selva (2013) and Mignan et al. (2014; 2016; 2017, 2018).

Consideration of single or multiple hazards, quantification of epistemic uncertainty, sensitivity and disaggregation analysis, and use of scenario-based analysis to enhance the PRA differentiate ST-L2 and ST-L3 sub-levels (Figure 2). Namely, ST-L2 involves single-hazard risk assessment, while ST-L3 involves multi-hazard risk assessment. Quantification of epistemic uncertainty, foreseen in stress test sub-levels b, c and d, aims to assess “the probability distribution representing the epistemic uncertainty within the (expert) community” (Bommer 2012). This can be achieved by selecting a number of appropriate different, yet scientifically acceptable, models and weighting them according to their credibility. Logic (e.g., Bommer and Scherbaum, 2008) or alternative trees methods (Marzocchi et al., 2015), where the risk analysis is divided into a number of consecutive steps and alternative models are defined at each step, is one possible way to organize a systematic epistemic uncertainty quantification. While sensitivity analysis is expected in all ST-Ls, a rigorous risk disaggregation is possible only in stress test sub-levels b, c and d. Scenario-based PRA enhancements are foreseen only in ST-L2d and ST-L3d. More complex risk analyses require more experts in the ET, PoE and IR to successfully complete a stress test.

Selection of stress test level

Component-level check (ST-L1) is mandatory in any stress test, whereas a system-level risk assessment is optional and must be planned in the Pre-Assessment phase based on the following considerations (Figure 2):

ST-L1a: This stress test level is a component-level check that requires knowledge and resources that do not exceed those required to design, operate and maintain the CI. The stress test may require the involvement of PM, TI, ET (consisting of a few individuals), and single-expert IR, a total of 5 to 6 people (3 plus 2/3 experts in the ET). The TI, with the support of the ET, are the only experts making critical scientific or technical decisions. They select the most important hazard to consider in a component-level hazard-based, design-based or risk-based analysis. If more than one hazard is considered to be critical for the CI under study, more than one ST-L1a check should be performed, one for each hazard. Epistemic uncertainties are not considered.

ST-L2a: Even though this is a system-level risk assessment, no quantification of epistemic uncertainties is required, resulting in a stress test team comprising 5 to 6 people similar to the team required for ST-L1a. Clearly, a system-level analysis builds on component-level checks already conducted by the same team. The intent behind allowing stress tests at ST-L2a is to strongly encourage system-level risk assessment.

ST-L2b: This stress test involves a more complex system-level risk assessment that includes quantification of epistemic uncertainties. Up to ten experts may be required to assist the TI, reflecting the increased effort and the wider scope of this stress test. Typically, the ET would involve a few experts internal to the CI and a few external experts. Similarly, the IR would include more than one expert to provide a competent internal peer review. Epistemic uncertainties are treated by the TI and the ET, who select the models based on a literature review and assigns appropriate weights to each one of them.

ST-L2c: Stress tests conducted at this level involve a PoE (consisting of eight to twelve experts to enable stable elicitations; Aspinall and Cooke, 2013) to assist the TI and ET in scientific and technical decisions. Thus, the stress test may require between 10 and 20 experts (PM, TI, ET formed by few individuals internal to the CI and a few external experts, PoE, and an IR with more than one expert). Compared to ST-L2b, a more robust procedure is foreseen to quantify epistemic uncertainties (Selva et al. 2015). In the Pre-Assessment phase, a preliminary list of models is prepared by the TI and the ET, screened by the PoE, and reviewed by the IR. Then, at the beginning of the Assessment phase, an expert elicitation of the PoE is organized by the TI to assign the weights to the models. Finally, the ET implements models and aggregates the weighted outcomes to quantify epistemic uncertainties using methodologies such as Logic Tree (e.g., Bommer and Scherbaum, 2008) or Ensemble Modelling (Marzocchi et al. 2015).

ST-L3c: As opposed to ST-L1 and ST-L2 single-hazard stress tests, this stress test involves a multi-hazard risk assessment. The same structure of experts as in ST-L2c is required, and the treatment of epistemic uncertainties is similar. However, the size of the ET, IR and PoE may increase. Selection of the considered hazards and identification of possible hazard sequences and co-location interactions occurs in the Pre-Assessment phase, based on the results of the PoE expert elicitation.

ST-L2d, ST-L3d: These two stress test levels are complementary to ST-L2c and ST-L3c levels, respectively. Scenario-based analyses should be considered only if, for technical or other reasons, one or more important phenomena cannot be included into a conventional PRA. These additional scenarios, not included in the PRA, are meant to further investigate the epistemic uncertainty by including otherwise neglected events. The choice of the scenarios should be based on *ad hoc* expert elicitation experiments of the PoE (e.g. Marzocchi et al., 2015). The choice of performing such scenario-based assessments (instead of including this into the PRA) should be justified by the TI, documented by the ET, and reviewed by the IR.

STRESS TEST PENALTY SYSTEM

There is a broad range of methods and models to assess the natural hazard risks to CIs. Generally, models reflect a strategy of bounded rationality and therefore are necessarily a simplification of reality (Fischhoff, 2015). However, how accurately a model reproduces the behavior of a prototype may vary greatly.

In the proposed stress test methodology, the notion of the Level of Detail is used describe the accuracy of the models and analysis methods adopted for component and system-level risk assessments. It is as a measure of “the trueness and precision, and the repeatability and reproducibility of the results of the risk assessment” (Bommer, 2012). While the Level of Detail implies the degree of reliability of the results obtained in the stress test Assessment phase, selecting the Level of Detail in the Pre-Assessment phase is challenging since a higher Level of Detail requires more experts with deeper knowledge of the spectrum of the available models and methods, making the stress test more demanding and costly. The state-of-the-practice methods, models and data are expected to have the trueness, precision, repeatability and reproducibility that can be achieved within the established state of knowledge and within a reasonable engineering and analysis effort. The experts involved in a stress test need to characterize the trueness and precision of the state-of-practice methods, models and data, while simultaneously promoting more advanced methods and discouraging less advanced methods. To facilitate this, penalty (Esposito et al. 2016, Stojadinović et al., 2016) and grading systems (Babič and Dolšek, 2016, 2019) are developed for the proposed stress test methodology.

Level of Detail

During the Pre-Assessment Phase (*STEP 3* in Figure 1) the TI and PM select the most appropriate ST-L (Figure 2) to perform a stress test for the given CI and hazard(s). As each ST-L entails a different level of complexity of the hazard and risk analysis, a different minimum Level of Detail is implied by the state of knowledge of the expert community and the state of practice in component and system-level risk assessment. A Target Level of Detail (*TL*) characterizes the minimum accuracy of ST-L. However, since the models, methods and data to perform each step of the stress test are identified by the TI based on scientific and technical grounds, but also on the practical considerations such as the duration of and the resources for the stress test, the Level of Detail of the selected models, methods and data may be different than the *TL*. Therefore, the choice of the models and methods is made and documented jointly by the TI and the PM and evaluated by the IR to establish the Effective Level of Detail (*EL)* of the stress test. The *EL* should be at least as high as the *TL*: otherwise, a penalty system is applied to modulate the outcome of the stress test. PM and TI may change the Level of Detail of the selected methods and models to avoid penalties while considering the resources and time needed to conduct a stress test.

The Level of Detail of the selected methods, models and data in may be described using either a qualitative or a quantitative scale. Three categories are defined on the qualitative scale:

- Advanced: making use of detailed data and advanced methods and models consistent with the best knowledge and the scientific and technical state of the art;
- High: making use of models, methods consistent with the current state of practice and data consistent with the current state of knowledge;
- Moderate: making use of methods, models and data consistent with the applicable design codes and regulations and the available CI design data.

A qualitative *TL* is associated to each ST-L in Table 1. Note that component-level assessment in the compulsory ST-L1a can be hazard-, design- or risk-based with Moderate, High and Advanced *TL*, respectively. ST-L2b,c,d and ST-L3c,d are associated with Advanced *TL* because state-of-the-art methods are required to quantify the epistemic uncertainties.

A Level of Detail is quantified using a *TL* interval [TL_{lb}, TL_{ub}], with the lower and upper bound set by the stress test team for the ST-L at which the stress test is conducted in the Pre-Assessment phase. The *EL* identified for the adopted stress test hazard and risk analysis methods, models and data should be at least equal to the lower *TL* bound TL_{lb}. Quantification of Level of Detail can be refined further for system-level stress tests by evaluating the aggregate *EL* based on the *EL* of each part of the conducted hazard and risk analysis. For example, in a single-hazard risk analysis (ST-L2) comprises three principal steps *i* (hazard, vulnerability and risk), with each one of the steps is characterized by *j* different layers (e.g. source model, ground motion model, elastic, static inelastic, dynamic inelastic response model, direct or indirect loss model). The resulting aggregate *EL* of the risk analysis is a function of the *EL* of each step *i* and layer *j*. Using a weighted average approach, the aggregate *EL* may be computed as:

$$EL = W_1 \sum_{j=1}^{n} w_{1,j} EL_{1,j} + W_2 \sum_{j=1}^{m} w_{2,j} EL_{2,j} + W_3 \sum_{j=1}^{p} w_{3,j} EL_{3,j} \quad , \qquad (1)$$

where *n*, *m* and *p* are the number of layers in each step *i* (hazard, vulnerability, risk), W_i represent the weight of each step *i* of the risk analysis, $w_{i,j}$ the weight of each layer *j* (of each step *i*), and $EL_{i,j}$ is the effective Level of Detail of the layer *j* in the specific step *i*. If all layers (of each step) are considered equally important, then $w_{1,1} = w_{1,2} = ... = w_{1,n} = 1/n$, $w_{2,1} = w_{2,2} = ... = w_{2,m} = 1/m$, $w_{3,1} = w_{3,2} = ... = w_{3,p} = 1/p$. If all steps are considered equally important, then $W_1 = W_2 = W_3 = 1/3$. In case of a multi-hazard risk analysis (ST-L3), aggregate *EL* may be obtained as:

$$EL = \sum_{s=1}^{S} W^{(H_s)} EL^{(H_s)} \tag{2}$$

where $W^{(H_s)}$ represents the weight of hazard *s* assigned by experts. Thus, a multi-hazard risk analysis *EL* corresponds to the weighted mean of the Level of Detail evaluated for each hazard separately. If all hazards are considered equally important, then $W^{(H_1)} = W^{(H_2)} = ... = W^{(H_s)} = 1/s$. The weights and *EL* values are assigned by the TI and reviewed by the IR in the Pre-Assessment and Assessment phases. When epistemic uncertainty analysis is of concern, a hierarchy of additional layers can be added in the same fashion.

Penalty Factor system and Penalized Loss

The objective of the proposed stress test penalty system (Esposito et al. 2016, Stojadinović et al., 2016) is to encourage using as advanced methods, models and data to conduct stress test risk analyses as can be afforded with the given resources. Conversely, the aim is to penalize simplistic, no necessarily conservative, hazard and risk assessment approaches.

A penalty factor is derived by extending the context of probabilistic seismic risk assessment (Cornel and Krawinkler, 2000, Broccardo et al. 2015). The result of a risk assessment at the system level is expressed by the annual exceedance rate of loss metric (*L*), $\lambda(l)$ as:

$$\lambda(l) = \int_d \int_{edp} \int_{im} G(l \mid d) \left| dG(d \mid edp) \right| \left| dG(edp \mid im) \right| \left| d\lambda(im) \right| \qquad (3)$$

where *im* is a hazard intensity measure (e.g., peak ground acceleration, spectral pseudo acceleration, peak or sustained wind speed), *edp* is an engineering demand parameter (e.g., interstory drift), *d* is a damage measure (e.g., minor, medium or extensive,), *l* is the loss variable (e.g., direct monetary losses, downtime), and $G(\mathrm{y} \mid \mathrm{x})$ are the conditional complementary cumulative distribution functions (CCDFs). To penalize an simplistic analysis, a random variable that measures additional uncertainty associated with such methods, models or data, named Uncertainty Penalty ε_P, is introduced. Then, a new loss metric, named Penalized Loss L_P, is defined on a logarithmic scale as:

$$\log(L_P) = \log(L) + \varepsilon_P \qquad (4)$$

The loss variable is assumed to be normalized within a [0,1] interval (e.g. with respect to the present cost to build a new facility) in order to make the assessment independent of currency or inflation considerations. Note that Uncertainty Penalty ε_P acts as a model error: it amplifies the uncertainties associated with simplistic analysis models, methods and data. A probability distribution of ε_P is the Normal distribution, i.e., $\varepsilon_P \sim N(0, \sigma(l))$, where the standard deviation $\sigma(l)$ of a loss variable is defined as:

$$\sigma(l) = \left| PF \cdot \log(l) \right|, l > 0 \qquad (5)$$

where the Penalty Factor (*PF*) is introduced to quantify the uncertainty added by simplistic methods, models and data used in risk analyses. Observe that *PF* acts as a coefficient of variation, in the sense that it increases the uncertainty present when state-of-the-art (advanced) methods, models and data are used. Further, in order to focus on the tails of the loss curve, no model error is added to *PF* when $l = 0$. Note that $\sigma(l)$ is proportional to the loss, as is customarily done in seismic risk assessment. Consequently, the tails of the computed loss curves are penalized both by the presence of an extra uncertainty due to simplistic analysis methods, models and data, as well as by a larger standard deviation $\sigma(l)$ of the loss variable *l*. Further, Penalized Loss L_P is random variable defined conditionally with respect to the loss value *l* of the loss metric *L* obtained from the probabilistic risk analysis. Then, the conditional cumulative complementary distribution of L_P can be written as:

$$G(l_P \mid l) = 1 - F(l_P \mid l) = P(L_P > l_P \mid L = l) \qquad (6)$$

and the

and the annual exceedance rate of L_P can be written, with reference to Eq. 3, as:

$$\lambda(l_P) = \int_l G(l_P \mid l)\left|d\lambda(l)\right| \qquad (7)$$

Given this formulation, the Penalty Factor *PF* can be quantified as the difference between the Expected Level of Detail *EL* and the Target Level of Detail *TL* of the chosen methods, models and data for the risk analyses at he selected stress test level. If the Level of Detail is expressed using a quantitative scale, *PF* is defined as the difference between the *EL* and the lower bound of the *TL* for the selected (TL_{lb}) as follows:

$$\begin{cases} PF = (TL_{lb} - EL) & \text{if} \quad EL < TL_{lb}^{ST} \\ 0 \quad otherwise \end{cases} \qquad (8)$$

Note that the penalty system could be also applied considering the upper bound of the *TL* interval, making it possible to penalize the added uncertainties more rigorously. Clearly, no penalty is applied if the Level of Detail of the selected methods, models and data meets or exceeds the Target Level of Detail. Conversely, the values of the obtained Penalty Factor *PF* should be in the [0,1] interval, with typical values ranging between 0.1 and 0.6, reflecting common estimates of the standard deviation of loss obtained from probabilistic natural hazard CI risk analyses. If a qualitative Level of Detail scale (i.e. Moderate, High, Advanced) is considered, Penalty Factors need to be defined for the three possible cases: i) *TL*=High, *EL* =Moderate, ii) *TL* =Advanced, *EL* = High, iii) *TL*= Advanced, *EL* =Moderate. For example: i)

$PF_{H\text{-}M}$=0.2, ii) $PF_{A\text{-}H}$=0.2, iii) $PF_{A\text{-}M}$=0.4. The actual EL, TL and PF values are set in *STEP 3* of the stress test procedure.

An example of using the proposed penalty system to account for the uncertainty added by simplistic methods, models and data used in risk analyses is provided in Figure 3, where the annual loss exceedance curve of a hypothetical CI has been penalized using different PF values. The solid curve corresponds to PF=0, i.e. the original annual loss exceedance rate L (Eq. 5), where the other two curves represent the annual exceedance rate of the Penalized Loss L_P (Eq. 7) for two values of the Penalty Factor. The Penalized Loss curve is intended for use in the stress test grading system.

STRESS TEST GRADING SYSTEM

The principal outcome of a CI stress test is a grade determined in *STEP 6* (Figure 2). The stress test grade is based on the comparison of the results of the risk assessment to the risk acceptance criteria defined in *STEP 2* of the stress test. The grade is obtained using a grading system developed and integrated into the proposed stress test methodology. Additional insights into the grading system are presented elsewhere (Babič and Dolšek, 2016, 2019).

The grading system (Figure 4) has three outcomes: Pass, Partly Pass, and Fail. A stress-tested CI passes the stress test if it attains grades AA or A. Grade AA corresponds to negligible risk and is expected to be the risk acceptance criterion for new CIs. Grade A corresponds to risk being as low as reasonably practicable (ALARP) (Helm, 1996; Jonkman et al., 2003), and is expected to be the risk acceptance criterion for existing CIs. Grade B corresponds to the existence of possibly unjustifiable risk; in this case, the CI partly passes the stress test. Grade C corresponds to the existence of intolerable risk; in this case, the CI fails the stress test.

Stress test grade boundaries

The PM defines the boundaries between grades (i.e. the risk acceptance criteria) by following the requirements of the regulators and societally acceptable risk norms in *STEP 2* of a stress test. The grade boundaries depend on the type the risk measure used to characterize risk. They can be expressed using point estimates (Figure 5, top row) or continuous functions (Figure 3 and Figure 5, bottom row). Examples of point estimates include the annual probability of risk measure exceedance (e.g. loss of life) and the expected value of the risk measure (e.g. expected number of fatalities per year), whereas continuous function examples include a frequency loss λ-L curve (e.g. Figure 3). Regulatory boundaries may differ between countries and industries. Harmonizing the risk objectives (and risk measures) across a range of critical non-nuclear CIs remains a challenge. This is a task for both the regulatory bodies and the industry associations, who should reconcile the societal and industry interest and develop mutually acceptable risk objectives and the stress test grade boundaries.

Evolution of the stress test grading system in time

The performance of a CI changes over time, and so does its natural hazard risk exposure. This is due to evolution of the understanding of risks and hazards through new findings, as well as changes of the CI due to use, ageing and long-term degradation processes, effects of previous hazard events, and change in the CI-induced community risk exposure (e.g. population increase or decrease). Variation of CI performances may lead to an increase of the probability of failure, loss of functionality, or exacerbate the consequences of failure during the CI lifetime. This means that a CI that passed a stress test at some point in time may not pass the stress test later on. More important, a CI that partially passed a stress test should be incentivized to reduce the risk it poses within a set time period. Crucially, a CI that failed a stress test must be compelled to design and implement retrofits to urgently reduce the risk it poses or manage these risks using other means, such as reduction or relocation of operations, or transfer of risk using financial measures such as insurance.

For these reasons, a stress test is designed to be periodic. Instead of a single stress test, the regulator should prescribe a sequence of stress tests designed to ensure continuous reduction of risk posed by the CI. The period between two consecutive stress tests is determined based on equity of the cumulative risk exposure posed by different CIs (Babič and Dolšek, 2016). The proposed stress test grading system is extended to facilitate such risk-driven periodic stress test organization. If the CI passes a stress test (obtains a grade AA or A), the risk objectives for the next stress test do not change. The time to the next stress test is set to the longest possible period between two stress tests established by the regulator. Such period may be as long as the expected lifetime of the CI system, or as short as dictated by the changes in the estimates of hazard or changed in the design codes. Some CIs may obtain grade B (the risk they pose is possibly unjustifiable) or C (the risk they pose is intolerable). In such cases, the grading system stimulates the stakeholders to act to reduce the risk in the following ways: i) by making the grade boundaries for the next stress test more stringent; and/or ii) by reducing the time between the successive stress tests. Both ways to reduce the risk are based on the equity of risk above the ALARP region over two stress test cycles (Babič and Dolšek, 2016). In particular, if a CI obtains grade B, the boundary between grades B and C is shifted to the left, toward the boundary between grade A and B (Figure 5b). Furthermore, if a CI obtains grade C, the boundary between grades B and C is moved to the boundary between grades A and B, and the period until the next stress test is reduced (Figure 5c). This incentivizes the CI stakeholders to adequately mitigate the risks posed by the CI in as few stress test cycles as possible. It follows that the CI will be upgraded to pass the stress test, or that the regulator will require that the CI ceases operation.

Grading of CI components

A stress test, regardless of the selected ST-L, comprises the compulsory component-level assessment (Figure 2) that can be hazard-based, design-based and risk-based. Each component is assessed by at least one method using an algorithm shown in Figure 6.

The acceptance criteria for a hazard-based assessment and a design-based assessment are obtained directly from the CI design codes or operation and maintenance guidelines, whereas the acceptance criteria for the risk-based assessment need to be defined in *Step 2* of a stress test. Similar to the system-level grades, three component-level grade boundaries need to be defined (between grades AA and A, between grades A and B and between grades B and C, as shown in Figure 5). If a hazard-based or a design-based component assessment results in lack of acceptance, or if the acceptance criteria are unknown, a more advanced method may be used (Figure 6). The proposed progressive component-level assessment procedure guides the transition to more accurate, but more demanding, assessments. If, in the end, a component fails the assessment (obtains grade C), risk mitigation actions must be taken. The time in which the grade needs to be improved depends on the type of assessment. If a hazard-based or a design-based assessment are used, risk mitigation has to be implemented immediately, as the component is not in compliance with the design or regulatory requirements. If a risk-based assessment is used, the time to improve the assessment grade is determined on the basis of the amount of risk corresponding to the component reaching the designated limit state in the time period considered (Babič and Dolšek, 2016), paralleling the evolution of the grading system for system-level CI assessment (Figure 5).

APPLICATION, DISCUSSION AND CONCLUSION

In order to demonstrate the proposed stress test methodology, six different CIs representing three major CI geographic distribution classes, were stress-tested to different hazards: 1) an oil refinery and petrochemical plant in Milazzo (Italy), mainly exposed to earthquake and tsunami hazard; 2) a conceptual alpine earthfill dam in Switzerland under multi-hazard effects (such as earthquakes, floods, internal erosion and electromechanical malfunctions); 3) the Baku-Tbilisi-Ceyhan pipeline in Turkey, focusing on seismic threats at pipe-fault crossing locations; 4) a part of the Gasunie national gas storage and distribution network in the Netherlands, exposed to earthquake and liquefaction effects; 5) the port infrastructure of Thessaloniki, Greece, subjected to seismic, tsunami and liquefaction hazards; and 6) an industrial district in the region of Tuscany, Italy, exposed to seismic hazard. The results are presented in detail in Pitilakis et al., (2018) Pitilakis et al. (2016), and in Rodrigues et al. (2018).

These six stress tests were conducted successfully, showing that the proposed stress test methodology works. However, the teams conducting the stress tests encountered some difficulties. In particular, a common difficulty was the formulation of the risk acceptance criteria (in Pre-Assessment Phase). In most of the cases, risk acceptance criteria have been set based on expert judgment. In practice, selecting risk measures and acceptance criteria is challenging that strongly depends on design, operation and maintenance regulatory requirements, technical, financial and societal considerations, and political context. Therefore, the risk acceptance boundaries are best defined in collaboration of the CI stakeholders: owners, operators, users and the neighboring communities. Other aspects of the stress test methodology that need to be further discussed and enhanced in future studies are summarized below.

A stress test can be conducted at three principal levels (Figure 2). This choice depends on regulatory requirements that should account for the importance/criticality of the CI. Risk ranking of CIs is a challenging task due to their diverse nature, very different potential consequences of failure, different types of hazards, vulnerabilities, etc. Nevertheless, an assessment of infrastructure criticality and importance, aimed at identifying and ranking CIs on a national or regional scale, will support the choice of the appropriate stress test level and the subsequent stress test resource allocation. Some key factors that may be considered to define the criticality and importance of a CI and a possible methodology to rank CIs are presented in Esposito et al. 2016 and Stojadinović et al., 2016.

The proposed penalty system requires that the accuracy of the models, methods and data used in risk analyses (i.e. level of detail) is evaluated and quantified by the stress test team. Therefore, the involved experts must have a clear idea about models, methods and data available in the scientific literature and their practical applicability to perform each step of the required risk analyses for a particular CI. This may not be feasible for all perils that have to be considered in a stress test, nor could it be possible for older and extensively modified CI. Further, this evaluation should change in each stress test, reflecting the progress of the scientific research and new knowledge and insights, as well as the acquisition of new data during operation and maintenance of the CI that is stress tested or other similar CIs. The Target and Effective Levels of Detail should, therefore, evolve over time, in a fashion similar to the proposed time evolution of the stress test grading system, to encourage timely and effective risk mitigation using state-of-the-art methods and best practices.

The proposed grading system is based on the mean annual rate of exceedance of the loss risk measure. However, other options are possible and should be discussed. For example, the grading system may be based on risk measure quantiles, as determined by the stress test PM. Second, it is yet to be determined how grades of single components should affect the outcome of a stress test. For example, if the CI is assigned grade B in an ST-L2 assessment, the stress test outcome is a Partial Pass. However, one or several components of that CI may receive grade C in the component-level assessment. In this case, one conservative option would be to change the outcome of stress test to Fail to compel the CI stakeholders mitigate the risk posed by these components. Another option would be to introduce a complementary outcome of stress test, which would address only single components and would be independent of the outcome obtained based on systemic level assessment. In this case, risk mitigation strategies and guidelines should be defined separately for individual components.

The outcomes of the stress test are intended to support decision makers in the evaluation and management of the risks CIs pose to communities they serve. However, the instantaneous losses of the community services provided by CIs by themselves do not reveal how a community served by the CIs responds to and functions after a natural disaster. The time dimension of the recovery process is key: the evolution of community needs and the ability of the CIs to fulfill these needs (e.g. water, gas, and electricity) is best represented and modelled using the concept of disaster resilience rather than that of natural hazard risk. The proposed stress test methodology was also designed to serve as a basis for development of a new resilience-based stress test concept with the goal to support decision makers in the evaluation of strategies and actions that not only decrease the risks posed by CIs, but to also enhance the resilience of CIs against natural hazards.

Acknowledgements

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. The authors gratefully acknowledge this funding. The authors acknowledge the contributions of the STREST Work Package leaders, Mr. Peter Zwicky, Prof. Fabrice Cotton, Prof. Iunio Iervolino, Prof. Kyriazis Pitilakis, Dr. Fabio Taucer and Dr. Sotirios Argyroudis. The methods, results, opinions, findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

Data Availability Statement

All data, models, and code generated or used during the study appear in the submitted article.

References

Aspinall, W.P., and Cooke, R.M. (2013) *Quantifying scientific uncertainty from expert judgement elicitation, in Risk and Uncertainty Assessment for Natural Hazards* [Eds. J. Rougier, S. Sparks, L.J. Hill] Cambridge University Press. https://doi.org/10.1017/CBO9781139047562.005

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

Babič A., and Dolšek, M. (2016). *Grading system in time domain in ST@STREST*. Faculty of Civil and Geodetic Engineering, University of Ljubljana.

Babič, A., and Dolšek, M. (2019). “A five-grade grading system for the evaluation and communication of short-term and long-term risk posed by natural hazards." *Structural Safety*, https://doi.org/10.1016/j.strusafe.2018.12.006, 78: 48–62.

Bedford T., Cooke R. (2001). *Probabilistic Risk Analysis*. 1st ed. Cambridge: Cambridge University Press, 2001. Cambridge Books.

Bommer J.J., and Scherbaum, F. (2008). “The use and misuse of logic-trees in probabilistic seismic hazard analysis.” *Earthquake Spectra*, https://doi.org/10.1193/1.2977755 , 96: 1967–1977.

Bommer J.J. (2012). “Challenges of Building Logic Trees for Probabilistic Seismic Hazard Analysis.” *Earthquake Spectra*, https://doi.org/10.1193/1.4000079, 28:1723-1735, ISSN: 8755-2930.

Broccardo M., Galanis, P., Esposito, S., and Stojadinović, B. (2015). *Probabilistic resilience assessment of civil systems: Analysis and validity of the PEER framework*. European Safety and Reliability Conference ESREL 2015, Zurich, Switzerland, September 7-10, 2015

Cornell, C., and Krawinkler, H. (2000). *Progress and challenges in seismic performance assessment*. PEER Center News, 3 (2).

Cotton F., Guéguen P., Abspoel-Bukman L., Akkar S., Aristizabal C., Bard P.-Y., Beauval C., Cheng Y., Courage W., Crowley H., Erdik M., Fotopoulou S., Giardini D., Hollender F., Karafagka S., Kotha S.R., Iqbal S.M., Manakou M., Marzocchi W., Mignan A., Miraglia S., Pitilakis K., Riedel I., Selva J., Smerzini C., Taroni M., Tsionis G., Uçkan E., and Weatherill G. (2016). *Reference Report 2: Guidelines for harmonized hazard assessment for LP-HC events*, EUR 28325 EN, http://dx.doi.org/10.2788/666080

Crowley, H., Casotto, C., Dolšek, M., and Babič, A. (2015). *Deliverable D4.3: Guidelines for performance and consequences assessment of multiple-site, low-risk, high-impact, nonnuclear critical infrastructures (exposed to multiple natural hazards, etc.)*. STREST project: Harmonized approach to stress tests for critical infrastructures against natural hazards. http://www.strest-eu.org/opencms/opencms/results/

ENSREG. (2011) "EU Stress test Specification", https://ec.europa.eu/energy/sites/ener/files/documents/20110525_eu_stress_tests_specifications.pdf

EPCIP (2013) Commission staff working document on a new approach to the European Programme for Critical Infrastructure Protection Making European Critical Infrastructures more secure, Brussels, 28.8.2013 SWD(2013) available at: https://ec.europa.eu/energy/sites/ener/files/documents/20130828_epcip_commission_staff_working_document.pdf.

Esposito, S., Stojadinović, B., Mignan, A., Dolšek, M., Babič, A., Selva, J., Iqbal, S., Cotton, F., and Iervolino, I. (2016). *Deliverable D5.1: Report on the proposed engineering risk assessment methodology for stress tests of non-nuclear CIs*. STREST project: Harmonized approach to stress tests for critical infrastructures against natural hazards. http://www.strest-eu.org/opencms/opencms/results/

Esposito, S., Stojadinović, B., Babič, A., Dolšek, M., Iqbal, S., Selva, J., and Giardini, D. (2017). "Engineering risk-based methodology for stress testing of critical non-nuclear infrastructures (STREST Project)*” 16th World Conference on Earthquake Engineering (16WCEE)* Santiago, Chile, 9-13, January 2017.

Esposito, S., Iervolino, I., d’Onofrio, A., Santo, A., Franchin, P., and Cavalieri, F. (2015). “Simulation-based seismic risk assessment of a gas distribution network.” *Computer Aided Civil Infrastructure Engineering*, https://doi.org/10.1111/mice.12105, 30: 508-523.

EUCO 10/1/11. (2011) “Conclusions of the European council of 24/25 march 2011” paragraph 31, https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/120296.pdf

Field, E. H., Jordan, T. H., Page, M.T., Milner, K. R., Shaw B. E., Dawson, T. E., Biasi, G. P., Parsons, T, Hardebeck, J. L., Michael, A. J., Weldon, R. J. II, Powers. P. M., Johnson, K. M., Zeng, Y., Felzer, K. R. Erst van der, N., Madden, C., Arrowsmith R., Werner, M. J., Thatcher, W. R. (2017). “A Synoptic View of the Third Uniform California Earthquake Rupture Forecast (UCERF3)”, Seismological Research Letters, https://doi.org/10.1785/0220170045, 88 (5): 1259-1267.

Fischhoff, B. (2015) “The realities of risk-cost-benefit analysis.” *Science*, https://doi.org/10.1126/science.aaa6516, 350 (6260): aaa6516.

Helm, P. (1996) “Integrated Risk Management for Natural and Technological Disasters.” *Tephra*, https://www.civildefence.govt.nz/assets/Uploads/publications/tephra-june-1996.pdf 15: 4-13.

Iervolino, I. (2016). “Soil-invariant seismic hazard and disaggregation.” *Bulletin of the Seismological Society of America*, https://doi.org/10.1785/0120160072, 4:1900–1907.

Jonkman, S.N., Van Gelder, P.H.A.J.M., and Vrijling, J.K. (2003). “An overview of quantitative risk measures for loss of life and economic damage.” *Journal of Hazardous Materials*, https://doi.org/10.1016/S0304-3894(02)00283-2 , 99:1-30.

Kakderi, K., Fotopoulou, S., Argyroudis, S., Karafagka, S., Pitilakis, K., Anastasiadis, A., Smerzini, C., Selva, J., Giannopoulos, G., Galbusera, L., Courage, W., Reinders, J., Cheng, Y., Akkar, S., Erdik, M., and Uckan, E. (2015). *Deliverable D4.2: Guidelines for performance and consequences assessment of geographically distributed, non-nuclear critical infrastructures exposed to multiple natural hazards*. STREST project: Harmonized approach to stress tests for critical infrastructures against natural hazards. http://www.strest-eu.org/opencms/opencms/results/

Liu, Z., Nadim, F., Garcia-Aristizabal, A., Mignan, A., Fleming, K., and Luna, B.Q. (2015). “A three-level framework for multi-risk assessment.” *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, https://doi:10.1080/17499518.2015.1041989, 9: 59-74.

Marzocchi, W., Taroni, M., and Selva, J. (2015). “Accounting for Epistemic Uncertainty in PSHA: Logic Tree and Ensemble Modeling.” *Bulletin of the Seismological Society of America*, https://doi.org/10.1785/0120140131, 10A5: 2151-2159.

Marzocchi, W., Garcia-Aristizabal, A., Gasparini, P., Mastellone, M.L., and Di Ruocco, A. (2012). “Basic principles of multi-risk assessment: a case study in Italy.” *Natural Hazards*, https://doi.org/10.1007/s11069-012-0092-x, 62: 551-573.

Mignan A, Wiemer S, and Giardini D (2014) “The quantification of low-probability–high-consequences events: part I. A generic multi-risk approach.” *Natural Hazards*, https://doi.org/10.1007/s11069-014-1178-4, 73: 1999-2022.

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

Mignan, A., Komendantova, N,. Scolobig, A., and Fleming, K. (2017). *Multi-Risk Assessment and Governance, Handbook of Disaster Risk Reduction and Management*, World Sci. Press & Imperial College Press, London, chapter 14, https://doi:10.1142/9789813207950, 357-381.

Mignan, A., Danciu, L., and Giardini, D. (2018). “Considering large earthquake clustering in seismic risk analysis.” *Natural Hazards*, https://doi:10.1007/s11069-016-2549-9, 91(Suppl 1): 149.

Morgan, M.G. (2014) “Use (and abuse) of expert elicitation in support of decision making for public policy”, Proceedings of the US National Academy Sciences. 2014 May 20; https://doi:10.1073/pnas.1319946111 , 111(20):7176-84.

Pitilakis, K., Argyroudis, S. Fotopooulou, S., Karafagka, S., Esposito, S., Stojadinović, B., Giardini, D., Dolšek, M., Babič ,A., Selva, J., Iqbal, S., Volpe, M., Tonini, R., Romano, F., Brizuela, B., Piatanesi, A., Basili, R., Lorito, S., Salzano, E., Basco, A., Crowley, H., Rodrigues, D., Matos, J.P., Schleiss, A.J., Courage, W., Reinders, J., Akkar, S., Uckan, E., and Erdik, M. (2016). *Reference Report 5: Guidelines for stress-test design for non-nuclear critical infrastructures and systems: Applications*. EUR 28342 EN, https://doi10.2788/85094 .

Pitilakis, K., Argyroudis, S., Fotopoulou, S., Karafagka, S., Selva, J., Salzano , E., Basco, A., Crowley, H., Rodrigues, D., Matos, J.P., Schleiss, A.J., Courage, W., Reinder,s J., Akkar, S,. Cheng, Y., Uckan, E., Erdik, M.(2018). A Multi-Level Stress Test Methodology: Application to Six Critical Infrastructures in Europe, *16th European Conference on Earthquake Engineering*, Thessaloniki, Greece, 18-21, June 2018.

Quagliariello, M. (2009). *Stress-testing the banking system: methodologies and applications*. Cambridge. University Press, 2009.

Rodrigues, D., Crowley, H. and Silva, V. (2018) "Earthquake loss assessment of precast RC industrial structures in Tuscany (Italy)", *Bull Earthquake Eng*, https://doi.org/10.1007/s10518-017-0195-6 , 16: 203.

Salzano, E., Basco, A., Karafagka, S., Fotopoulou, S., Pitilakis, K., Anastasiadis, A., Matos, J.P., and Schleiss, A.J. (2016). *Deliverable D4.1: Guidelines for performance and consequences assessment of single-site, high-risk, non-nuclear critical infrastructures exposed to multiple natural hazards*. STREST project: Harmonized approach to stress tests for critical infrastructures against natural hazards. http://www.strest-eu.org/opencms/opencms/results/

Selva, J. (2013). “Long-term multi-risk assessment: statistical treatment of interaction among risks.” *Natural Hazards*, https://doi:10.1007/s11069-013-0599-9, 67:701.

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

Selva, J., Iqbal, S., Cotton, F., Giardini, D., Esposito, S., Stojadinović, B., Argyroudis, S., Pitilakis, K., and Mignan, A. (2019) The Epistemic Uncertainty Process (EU-P): management of critical choices and uncertainty in single and multi-hazard/risk assessments based on multiple experts, *working paper*.

SSHAC (1997). “Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts (No. U.S. Nuclear Regulatory Commission Report, NUREG/CR-6372).” U.S. Nuclear Regulatory Commission Report, NUREG/CR-6372. Washington, D.C.

Stojadinović B., Esposito S., Babič A., Cotton F., Dolšek M., Giardini D., Iqbal S., Mignan A., Selva J., and Tsionis G. (2016). *Reference Report 4: Guidelines for stress-test design for* *non-nuclear critical infrastructures and systems: Methodology*, EUR 28342 EN, https://doi:10.2788/659118.

Tsionis, G., Pinto, A., Giardini, D., Mignan, A., Cotton, F., Iervolino, I. Pitalakis, K., Stojadinović, B., Zwicky, P., Danciu, L. (2016). *Reference Report 6: Harmonized approach to stress tests for critical infrastructures against natural hazards.* EUR 28343 EN, http://dx.doi.org/10.2788/17066

Woessner, J., Danciu, L., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M.B., Hiemer, S., Meletti, C., Musson, R., Rovida, A., Sesetyan, K., Stucchi, M. & the SHARE Consortium, 2015. "The 2013 European seismic hazard model: key components and results. *Bulletin of Earthquake Engineering*, https://doi.org/10.1007/s10518-015-9795-1 13, 3553-3596.