STREST

D 5.1

DELIVERABLE

PROJECT INFORMATION

Project Title:	**Harmonized approach to stress tests for critical infrastructures against natural hazards**
Acronym:	**STREST**
Project N°:	603389
Call N°:	FP7-ENV-2013-two-stage
Project start:	01 October 2013
Duration:	36 months

DELIVERABLE INFORMATION

Deliverable Title:	**Report on the proposed engineering risk assessment methodology for stress tests of non-nuclear CIs**
Date of issue:	31 July 2016
Work Package:	WP5 – DESIGNING STRESS TESTS FOR CIs (NON-NUCLEAR)
Editor/Author:	Simona Esposito Bozidar Stojadinovic (ETHZ, IBK)
Reviewer:	Sotirios Argyroudis Stavroula Fotopoulou (AUTH)

REVISION: Final

SEVENTH FRAMEWORK PROGRAMME

Project Coordinator:	Prof. Domenico Giardini
Institution:	ETH Zürich
e-mail:	giardini@sed.ethz.ch
fax:	+ 41 446331065
telephone:	+ 41 446332610

Abstract

An engineering risk-based multi-level stress test, named ST@STREST, is proposed herein aimed at enhancing procedures for evaluation of the risk exposure of critical non-nuclear infrastructures against natural hazards. In order to account for diversity of types of critical infrastructures (CIs), the potential consequence of failure of the CIs, the types of hazards and the available human/financial resources for conducting the stress test, each Stress Test (ST) level is characterized by a different scope (component or system), and by a different complexity of the risk analysis. The ST@STREST workflow is composed of four main phases and nine steps to be conducted sequentially. First the goals, the method, the time frame, and the total costs of the stress test are defined. Then, the stress test is performed at component and system levels; then, the outcomes are checked and compared to the acceptance criteria. A stress test grade is assigned and the global outcome is determined by employing a grading system proposed herein. According to the outcome the parameters of the following evaluation of stress test are adjusted. Finally, the results are reported and communicated to stakeholders and authorities.

Keywords: multi-level, stress test, grading system, acceptance criteria

Acknowledgments

The work presented in this report 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 [FP7/2007-2013] under grant agreement no. 603389. The authors gratefully acknowledge this funding. The methods, results, opinions, findings and conclusions presented in this report are those of the authors and do not necessarily reflect the views of the European Commission.

Deliverable Contributors

ETH, Zurich	Simona Esposito	All
	Bozidar Stojadinovic	All
	Arnaud Mignan	3.4.2.2
University of Ljubljana	Matjaž Dolšek	3.3,3.4.1,3.5, 3.6
	Anze Babič	3.3,3.4.1,3.5, 3.6
INGV	Jacopo Selva	3.2, 3.3, 3.4
	Sarfraz Iqbal	3.2, 3.3, A
GFZ	Fabrice Cotton	3.2
AMRA	Iunio Iervolino	B

Table of Contents

List of Figures

List of Tables

List of Abbreviation

ALARP	As Low As Reasonably Practicable
cEE	classical Expert Elicitation
CI	Critical Infrastructure
ET	Evaluation Team
EL	Effective Level
EU	Epistemic Uncertainty
EU@STREST	Epistemic Uncertainty at STREST
GenMR	Generic Multi-Risk
IR	Internal Review
PBEE	Performance-Based Earthquake Engineering
PEER	Pacific Earthquake Engineering Research
PF	Penalty Factor
PM	Project Manager
PoE	Pool of Experts
PRA	Probabilistic Risk Assessment
QRA	Quantitative Risk Analysis
SBRA	Scenario Based Risk Assessment
SHARE	Seismic Hazard Harmonization in Europe
ST	Stress Test
ST-L	Stress Test Level
STREST	harmonized approach to stress tests for critical infrastructures against natural hazards
ST@STREST	Stress Test at STREST
TI	Technical Integrator
TL	Target Level

1 Introduction

"Critical Infrastructure means an asset, system or part thereof located in Member States which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a result of the failure to maintain those functions" (Art. II a, 2008/ 114/ EC).

As described by the definition above, critical infrastructures (CIs) are of essential importance for the society, but extreme natural events can interrupt services, cause damage, or even destroy such systems, which consequently trigger the disruption of vital socio-economic activities, extensive property damage, or human injuries or loss of lives (Grimaz and Slejko, 2014). Recent catastrophic events showed that these systems cannot recover their functionality back to the pre-disaster state, significantly increasing the concerns of the public.

This can be explained by the increasing complexity and dynamics of infrastructure systems, and by a growing dependency of the society on the infrastructure services. Consequently, the interest of the international academic community, industry and stakeholders in the challenges of understanding and modelling the risk and the resilience of CIs is increasing. The European Programme for Critical Infrastructure Protection (EPCIP) was established in 2006 and it was recently revised to ensure a high degree of protection of EU infrastructures. The result of this process was to establish a working group aimed at increasing the safety and the resilience of these systems, and decreasing the loss of service and the impact to the society. In particular, the EPCIP clearly declared the need to develop stress tests in the context of critical non-nuclear infrastructures, as improvement measure to be applied in the near future.

Actually, the two only domains where the stress test tool has been applied are the nuclear and the banking sectors. Stress test in finance was strongly promoted and emphasized after the 2007 financial crisis, as high analytical tool to assess the stability of some components of an economy (i.e. financial instruments and institutions) that may have an impact on the global economy. In the nuclear sector, after the 2011 Fukushima accident, the European Council mandated the European Commission to review the safety of all nuclear plants on the basis of transparent and comprehensive risk assessment (i.e. stress tests).

In this report an engineering risk-based methodology (ST@STREST) for stress tests of non-nuclear CIs is proposed. The aims of ST@STREST are to verify the safety and the risk of individual components as well as of whole CI system with respect to extreme events and to compare the response of the CI to acceptable values. In particular, a Multi-Level framework has been proposed. Each Level is characterized by different scope (component or system) and by different levels of risk analysis complexity (starting from design codes and ending with state-of-the-art risk analyses, such as cascade modelling). This allows flexibility and application to a broad range of infrastructures. The framework is composed of four main phases and nine steps. First the goals, the method, the time frame, and the total costs of the stress test are defined. Then, the stress test is performed at component and system level; additionally, the outcomes are checked and analyzed. Finally, the results are reported and communicated to stakeholders and authorities.

ST@STREST has been applied and tested in six CIs in Europe (see D6.1 Pitilakis et al. 2016), namely: the ENI/Kuwait oil refinery and petrochemical plant in Milazzo, Italy; the large dams in the Valais region of Switzerland, the major hydrocarbon pipelines in Turkey, the Gasunie national gas storage and distribution network in Holland; the port infrastructures of Thessaloniki in Greece; and the industrial district affected by the 2012 Emilia earthquake in Italy. These case studies are representative of the CIs categories identified in STREST (Fig. 1.1.): 1) individual, single-site infrastructures with high risk and potential for high local impact and regional or global consequences; 2) distributed and/or geographically-extended infrastructures with potentially high economic and environmental impact, 3) distributed, multiple-site infrastructures with low individual impact but large collective impact or dependencies.

- In the following sections, the main aspects of the proposed engineering risk-based methodology (ST@STREST) for stress tests of non-nuclear CIs are presented. First, a review of the most common risk metrics and acceptance criteria adopted in different fields is provided. Then the main characteristics of the proposed methodology are presented. Finally, possible grading and penalty systems are introduced. The grading system allows grading the CI and prescribing how much the safety of the CI should be improved in the periodical verification of the CI. The penalty system, instead, allows defining how reliable are the results of the stress test, and in case it is needed, to penalize the output of the risk assessment.

Fig. 1.1 STREST selected CIs in Europe representative for each of the three identified CI classes: A1) ENI/Kuwait oil refinery and petrochemical plant, Milazzo, Italy: A2) Large dams in the Valais region of Switzerland; B1) Major hydrocarbon pipelines, Turkey; B2) Gasunie national gas storage and distribution network, Holland; B3) Port infrastructures of Thessaloniki, Greece; C1) Industrial district affected by the 2012 Emilia earthquake.

2 Risk assessment of critical infrastructures and acceptance criteria

2.1 RISK ASSESSMENT AND DECISION MAKING FOR CRITICAL INFRASTRUCTURES

Risk is a common notion used in different fields to indicate the uncertainty related to the state of an activity under discussion (Faber and Stewart, 2003). It is typically defined as the expected consequences associated to an activity and it can be expressed with different risk metrics according to the consequences under consideration (e.g. terms of money, number of fatalities etc).

Risk metrics represent the principal tool for quantitative safety risk assessments, decision making, risk communication, and regulatory frameworks. In particular, risk metrics play a key role in setting acceptable risk levels and standards. The definition, the analysis, the requirements and the regulatory have evolved for different engineering disciplines although the theoretical frameworks are similar. The concept of tolerability of risks is fundamental in risk assessment and management; however without a definition of such acceptable risk limits and criteria, risk analysis and assessments may be hampered in term of decision making and formulation of mitigation strategies (Berg, 2010).

In the following, an overview of various risk measures and acceptance criteria used in quantitative risk analysis is provided. The review is mostly based on the paper of Jonkman et al. (2003).

2.1.1 Risk metrics

A risk metric may be defined as a mathematical function of the probability of an event and the consequences of that event. Different risk metrics have been defined in quantitative risk analysis, and they can be generally categorized according to the consequences they consider (Jonkman et al., 2003):

- Fatalities
- Economic damage
- Environmental damage
- Integrated risk measures

The fatality risk is distinguished between individual risk and societal risk.

Individual risk (IR) is defined as the probability that an average unprotected person, permanently present at a certain location, is killed or injured due to an accident resulting from a hazardous activity (Bottelberghs, 2000). The consequence is then represented by a categorical variable that may express different injured severity levels. For a specific injury level *il*, the individual risk can be written as (Broccardo et al., 2017):

$$IR = \sum_{d} P\left(IL = il \mid d\right) P\left(D = d\right) = P\left(IL = il\right) \quad (1.1)$$

where $P(IL = il \mid d)$ is the conditional probability observing the injury level il given a specific damage state of a system, assuming an implicit dependence of damage states from the specific hazard. This expression does not account for whether the individual is actually physically in the proximity of the hazardous area. Bohnenblust (1998) proposed a slightly different version of Eq. (1.1), which accounts for the portion of the time that a person is present in the area.

Societal risk is defined as the relationship between frequency and the number of people suffering from specific level of harm in a given population from the realization of specific hazards (IchemE, 1985). Therefore, individual risk gives the probability of dying on a certain location while societal risk give the number for a whole area, independently where the harm occurs. Different societal risk measures have been formulated in the past years. Some of them are based on individual risk and others are defined based on the cumulative density function (CDF) of the number of fatalities per year ($F_{N_f}(x)$). In the following and in Table 2.1, some of them are summarized.

Simple measures of societal risk based on $F_{N_f}(x)$ are:

i) The expected value of the number of fatalities per year E(N_f) (referred in literature as the potential loss of life, PLL), and expressed as:

$$E(N_f) = \int_0^\infty x dF_{N_f}(x) \mathrm{d}x \tag{1.2}$$

ii) The risk integral (RI), expressed as a function of the two characteristics of the probability distribution function (pdf) of the number of fatalities, E(N_f) and the standard deviation σ (N_f), (Vrijling and Van Gelder, 1997), i.e.:

$$RI = \frac{1}{2}\left(E^2(N_f) + \sigma^2(N_f)\right) = \int_0^\infty x\left(1 - F_{N_f}(x)\right)\mathrm{dx} \tag{1.3}$$

Another version for the risk integral is the so named COMAH risk integral (Carter, 2002) which is defined as:

$$RI_{COMAH} = \int_0^\infty x^\alpha dF_{N_f}(x) \mathrm{d}x \tag{1.4}$$

where α is the aversion coefficient, which is ≥1 and represents the aversion to accidents with many fatalities.

iii) The total risk (TR) , proposed by Vrijling et al. (1995) is the composition of E(N) and σ (N) which is multiplied by a risk aversion factor *k*, i.e.:

$$TR = E(N) + k\sigma(N) \tag{1.5}$$

It is important to note that most of societal risk metrics proposed in literature, which are derived from different disciplines, have different properties and units. This is an obstacle for decision makers because these metrics cannot be compared. To overcome this issue, Broccardo et al. (2017) proposed a generalized metric of societal risk based on simple function space concepts. These spaces are defined with respect to the local cumulative frequency of number of fatalities per year and are equipped with a p-norm used to define flexible, yet unit-consistent risk metrics. The metric is defined for positive variables ($x \in R^{+}$) and is expressed as:

$$R_{SR_p} = \left[\int_x \left| \phi(x) \right|^p \left| \left(1 - F_{N_f}(x) \right) \right| \right]^{\frac{1}{p}}, p \geq 1 \quad (1.6)$$

where $\phi(x)$is a generic function that for example can include risk aversion factors.

Economic risk measures are typically expressed in terms of F-D curve, where F is the probability of exceedance as function of the economic damage, or in terms of expected valued of economic damage E(D).

Some environmental risk measures, have been proposed in literature considering the expected amount of damage to nature, in particular the NORSOK (the competitive standing of the Norwegian offshore sector, 1998) has proposed the probability of exceedance of the time needed by the ecosystem to recover from the damage. Barlettani et al. (1997) proposed the energetic impact index, as a measure of the amount of energy lost per year caused (in Joules) by injured and fatalities expressed as:

$$GPP_{lost} = EPP + GPP^{'}T \quad (1.7)$$

where *EPP* represent the energy loss of the system, *GPP'* the amount of energy needed during the period *T* for recovery of harmed organisms.

Integrated risk measures are usually applied when different kind of consequences are considered in the risk analysis (e.g. fatalities, economic damages, etc.). All the risks can be expressed in monetary terms by determining the willingness to pay for every scenario of different kind of consequences (i.e. fatalities, economic damage, and environmental damage).

2.1.2 Acceptance criteria and existing standards

Acceptance criteria or goals represent an acceptance risk level. These goals may be defined as scalars or continuous measures according to the risk measure adopted. Examples of scalar measures are the annual probability of a risk measure N (e.g. loss of life of a person exposed to hazard), the expected value of N (E(N)) or the composition of the moments of the probability distribution of N (e.g. the total risk TR). As continuous measure, engineers and risk analysts often represent risks and limit on so-called F-N charts, where F represents the cumulative frequency of the risk measure N per given period of time (usually 1 year).

As acceptance scalar criteria, the life safety risk regulations set individual risk criteria, which vary given the nature of the hazard and the exposed individuals. The individual risk criteria can set an absolute limit or specify separate thresholds for the public and the most exposed personnel working at the activity. The Dutch Ministry of Housing, Spatial Planning and Environment has set IR< 10^{-6} (per year) as standard (absolute criteria) for populated areas with a *de minimis*[1] value of risk of 10^{-8}. An example for the other criteria is given by the

---

[1] The "de minimis" level is the upper threshold of the domain where the risks are "below legal concern" (Pate-Cornell, 1994). Any activity whose risk falls below that threshold value can be ignored – no action needs to be taken to manage this de minimis risk.

safety requirements for liquefied natural gas which set IR < 10^{-4} for the employees and IR < 10^{-5} for the population. A slight different standard was proposed by the Dutch Technical Advisory Committee on Water Defenses (TAW, 1985) which consider a policy factor β that varies according to the degree to which participation in the hazardous activity is voluntary or not, IR< $\beta \cdot 10^{-4}$ (per year).

The expected number of fatalities per year, $E(N_f)$ is currently used in the regulation of risk of dams. The British Columbia Hydro (Bowles et al., 1999) proposed as acceptable risk limit, $E(N_f)<10^{-3}$, while the United States Bureau of Reclamation (USBR, 1997) proposed $E(N_f)<10^{-2}$ (fatalities/year). For industrial (and new) facilities, the Dutch regulation proposed that the annual (societal) risk should be lower than 10^{-5} fatalities with a *de minimis* value of 10^{-7}. These standards are relaxed by one order of magnitude for existing facilities (Pate-Cornell, 1994). Vrijling et al (1995) presented a standard for the total risk, TR, considering a policy factor $\beta, TR < \beta \cdot 100$. Acceptance criteria for integrated (scalar) risk measures have been proposed by the TAW (1985) for individual, societal and economic risk.

As continuous acceptance criterion, the F-N curve has been applied in many fields to express and limit the risks, predominantly for hazard installations. In several countries F-N criterion lines limit the risks of various hazardous activities (Fig. 2.1) and they can be described with the following general formula:

$$1 - F_N(x) < \frac{C}{x^n} \qquad (1.8)$$

where *n* is the steepness of the limit line and *C* the constant that determines the position of the limit line. A standard with n equal to 1 is called risk neutral and with n equal to 2 is called risk adverse. This criterion can be translated into a risk criterion for a single installation or location where an activity takes place, considering, under appropriate hypotheses, the factor *C* written as a function of the number of installations N_A:

$$C = \left[\frac{\beta \cdot 100}{k\sqrt{N_A}}\right]^2 \qquad (1.9)$$

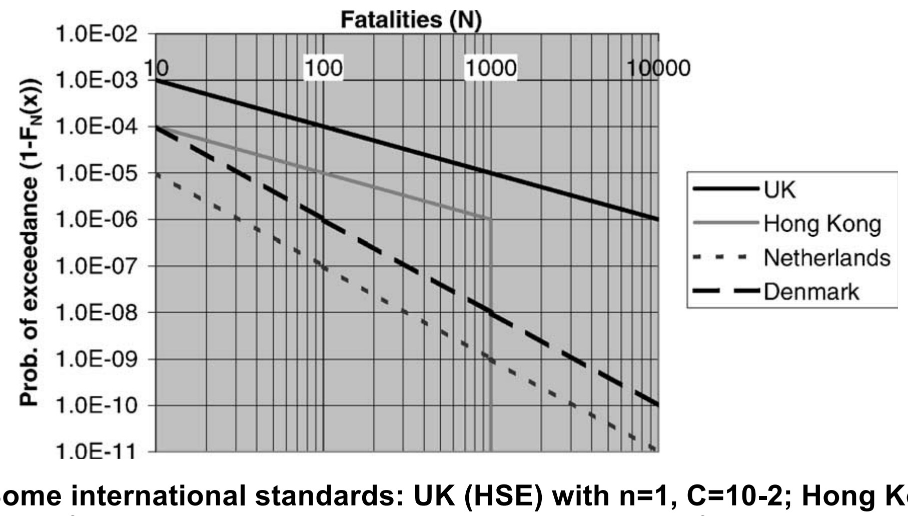

Fig. 2.1 Some international standards: UK (HSE) with n=1, C=10-2; Hong Kong with n =1 and C =10^{-3}; The Netherland (VROM) with n=2 and C=10^{-3} and Denmark with n =2 and C =10^{-2} (Jonkman et al., 2003).

Fig. 2.2 shows some of existing acceptance criteria in the nuclear industry compared by Hakata (2003) and Fig. 2.3 shows instead the F-N chart for risk associated to civil facilities and other large structures (Baecher, 1983). For integrated risk measures, Merz et al. (1995) proposed a framework that limits the risk for man, economy and environment, defining acceptable probability as a function of a specific index value for each type of consequences.

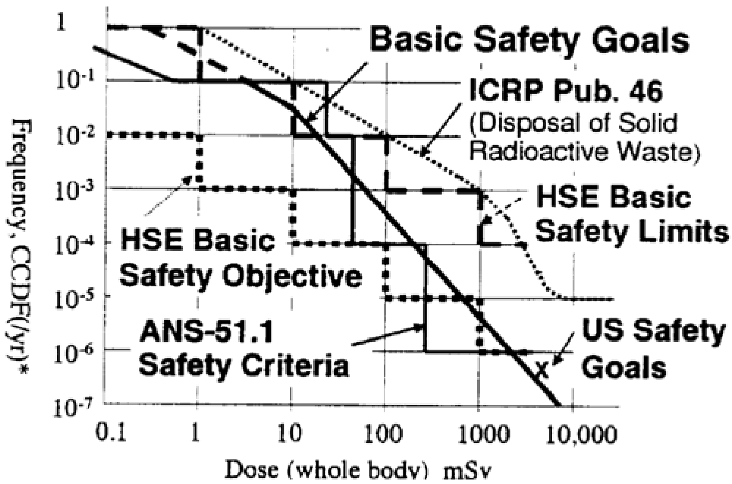

Fig. 2.2 Comparison of different acceptance criteria in the nuclear industry respect to different standards and regulations (source Hakata, 2003).

It is important to note that most risk measures for which acceptable levels have been defined in the past literature, consider only one type of consequences and the majority is limited to consider fatalities in form of individual or societal risk. This is because the number of fatalities is considered as the most important consequence of a disaster (Jonkman et al., 2003).

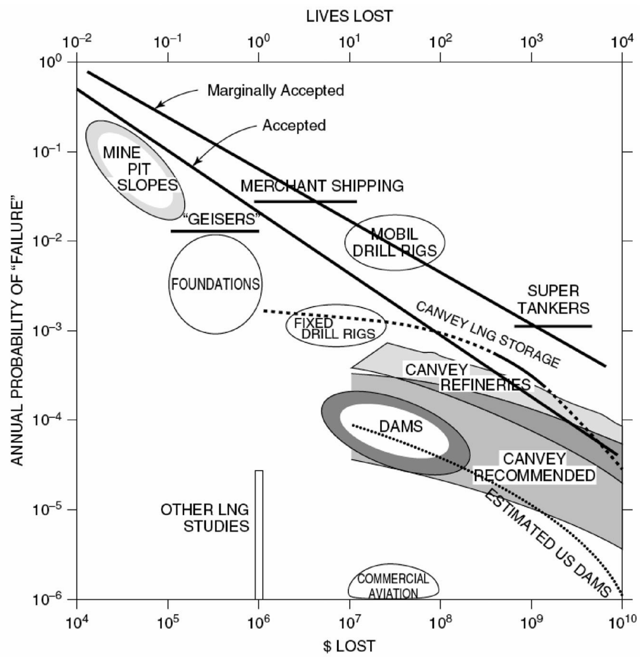

Fig. 2.3 F-N limits for risks associated to civil facilities and other large structures (source Baecher, 1983).

Helm (1996) examined a variety of industrial and other technological perils and assessed the tolerability of these perils as a function of frequency (F) and risk measure (N) in terms of number of deaths. He found that there are four general regions of F-N space (Fig. 2.4) that characterize the tolerability (acceptability) of risk, i.e.

- Intolerable: high frequencies and severe consequences. In this region, “risk cannot be justified except in extraordinary circumstances”.
- Possibly unjustifiable (Upper ALARP): “risk is tolerable only if risk reduction is impractical or if its cost is grossly disproportionate to the improvement gained”. This represents the upper portion of the region Helm denotes ALARP (as low as reasonably practicable). This means that the risk is tolerable as long as all reasonably practical steps are taken to reduce the risk further.
- Lower ALARP: risk is not-negligible, but is “tolerable if cost reduction would exceed the improvement gained”.
- Negligible: below the negligibility line, F and N are low enough for the risk to be considered broadly acceptable.

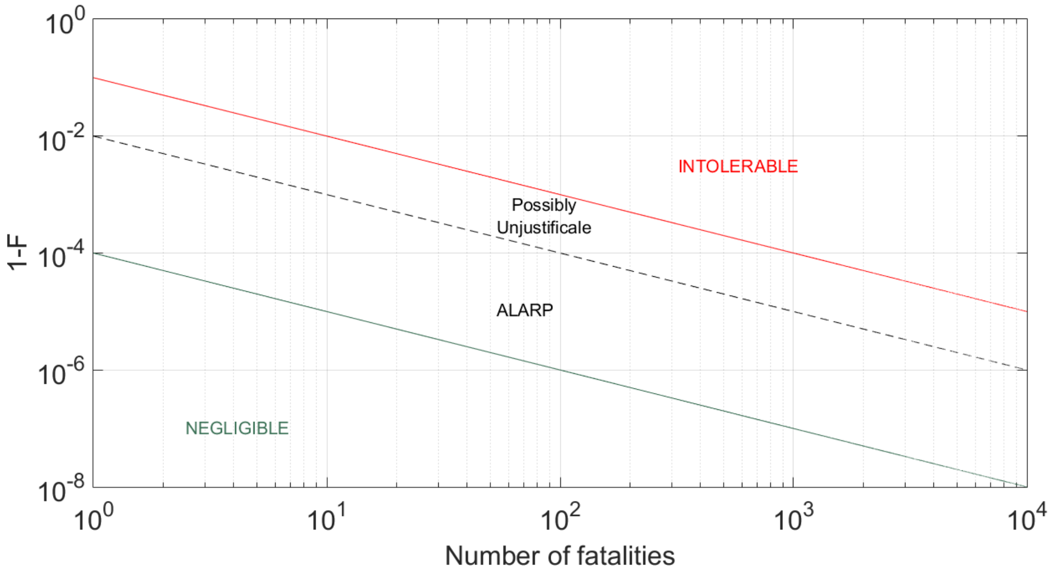

Fig. 2.4 F-N regions and limits derived from risk guidelines developed in the United Kingdom. It depicts risk thresholds in terms of acceptability of deaths from industrial and other accidents (adapted from Helm, 1996).

The relationships showed in Fig. 2.1-2.4 are particularly helpful for helping to assess the necessity of risk mitigation strategies, especially the 4 zone-limits framework proposed by Helm (1996). In fact if a peril exceeds acceptable risk levels, mitigation strategies should be determined in order to reduce the risk in an adequate way. For moderate to-high risk perils, a clear distinction is made between reasonable and unreasonable cost for risk. If the improvements gained do not justify the cost reduction (Upper ALARP region), the risk can be considered tolerable and mitigation strategies are not needed in terms of cost-benefit analysis.

However, it is important to note that distinctions on acceptable risk levels are strongly dependent on legal and political interpretation which may considerably vary from country to country.

Table 2.1 Overview of risk measures and acceptance criteria (adapted from Jonkman et al., 2003)

Risk Measure	Category	Basis of calculation	Consequences	Field of application	Limit
IR	Individual Risk	Probability of death for permanently present person	Death of individual (1 year)	Hazardous Installation in Netherland (VROM)	$<10^{-6}$
IR-TAW	Individual Risk	Probability of death for actually present person	Death of individual (1 year)	Flood studies	$< \beta\ 10^{-4}$
F-N curve	Societal Risk	Probability density function of the number of fatalities	Fatalities (1 year)	International: hazard activities	$1-F_N(x) < \frac{C}{x^n}$ (Figure 3)
E(N)	Societal Risk	Probability density function of the number of fatalities	Fatalities (1 year)	US, Canada: dams	USBR: $<10^{-2}$
RI	Societal Risk	Probability density function of the number of fatalities	Fatalities (1 year)	HSE (UK): land use planning	n.a.
RI_{COMAH}	Societal Risk	Probability density function of the number of fatalities	Fatalities (1 year)	HSE (UK): land use planning near hazardous installation	n.a.
TR	Societal Risk	Probability density function of the number of fatalities	Fatalities (1 year)	NL: studied External safety	$< \beta\ 100$
F-D curve	Economic Risk	Probability density function of economic damage	Economic damage (1 year)	Displays various economic risks	Proposed by Jansen (1988)
E(D)	Economic Risk	Probability density function of economic damage	Economic damage (1 year)	UK and NL: cost benefit for floods, US dams	USBR: $10.000

Recovery time	Environmental Risk	Probability density function of recovery time (T) of the ecosystem	Ecological damage (recovery time T)	NORSOK: oil platform	$1 - F_T(x) < \frac{0.05}{T}$
GPP_{lost}	Environmental Risk	Analysis of the amount of energy lost in the ecosystem	Effect on ecosystem (Joule)	n.a.	n.a.

3 ST@STREST: Stress Test at STREST

3.1 INTRODUCTION

An engineering risk-based methodology for stress testing critical non-nuclear infrastructures, named ST@STREST, has been developed in the scope of the STREST project (Esposito et al., 2017). The aims of the proposed methodology are to assess the performance of individual components as well as of whole CI systems with respect to extreme events, and to compare this response to acceptable values (performance objectives) that are specified at the beginning of the stress test. ST@STREST is based on probabilistic and quantitative methods for best-possible characterization of extreme scenarios and consequences (Cornell and Krawinkler, 2000; Mignan et al., 2014; 2016a).

Further, it is important to note that CIs cannot be tested using only one approach: they differ in the potential consequence of failure, the types of hazards, and the available resources for conducting the stress tests. Therefore, a Multi-Level framework has been proposed (*Section 3.4*). In this framework each Stress Test Level (ST-L) is characterized by different focus (component or system) and by different levels of risk analysis complexity (starting from design codes and ending with state-of-the-art risk analyses, such as cascade modelling, Mignan et al., 2016a). The selection of the appropriate Stress Test Level depends on regulatory requirements, based on the different importance of the CI and the available human/financial resources to perform the stress test.

In order to allow transparency of the process, a description of the assumptions made in connection with the system identification as well as the modeling of consequences and frequencies is foreseen. In fact, all the data, models, methods adopted for the risk assessment and the associated uncertainty are clearly documented and managed by different experts involved in the stress test process, following a pre-defined process for managing the multiple-expert integration (Selva et al., 2015, Selva et al., in prep.). This allows defining how reliable the results of the stress test are (i.e. level of “detail and sophistication”) of the stress test (Section 2.6).

Different experts are involved in the implementation of stress test process and different roles and responsibilities are assigned to different actors, as described in *Section 3.2* and *Section 3.3*. In particular, the several participants may be involved, with different background knowledge, but in specific cases may be reduced to individuals. The size of such groups depends on selected ST-Level (see *Section 3.4*).

The workflow of ST@STREST comprises four phases (Fig. 3.1): Pre-Assessment phase; Assessment phase; Decision phase; and Report phase. In the Pre-Assessment phase all the data available on the CI and on the phenomena of interest (hazard context) is collected. Then, the goal, the time frame, the total costs of the stress test and the most appropriate Stress Test Level to apply to test the CI are defined. In the Assessment phase, the stress test is performed at Component and System Level. In the Decision phase, the stress test outcomes are checked i.e. the results of risk assessment are compared with the objectives defined in Pre-Assessment phase. Then critical events, i.e. events that most likely cause a given level of loss value are identified and risk mitigation strategies and guidelines are formulated based on the identified critical events and presented in the Report phase.

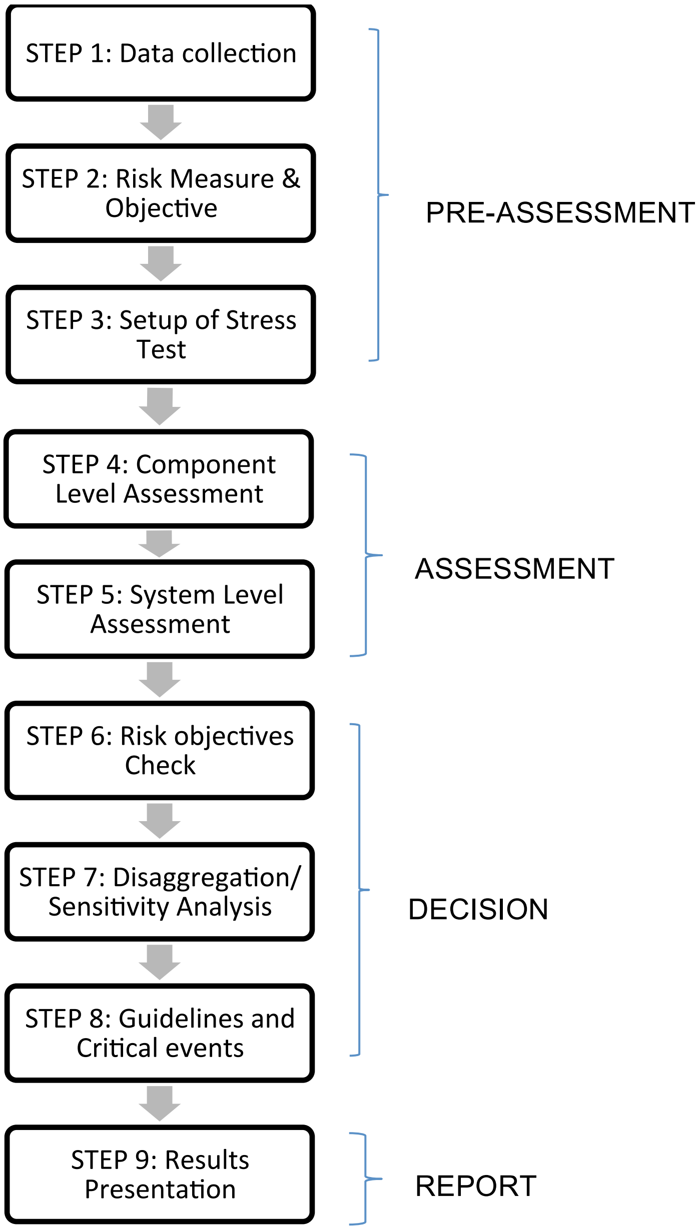

Fig. 3.1 Workflow of ST@STREST methodology

All the aspects characterizing the ST@STREST methodology are described in the following sections, in particular:

- The use of multiple experts (*Section 3.2*): to guarantee the robustness of stress test results, to manage subjective decisions and quantify epistemic uncertainty.
- The workflow of the process (*Section 3.3*): description of the sequence of phases and steps which have to be carried out in a stress test.
- The multi-level framework (*Section 3.4*): the different levels of the analysis to test the CI.

- The grading system (*Section 3.5*): to compare the results of the risk assessment with acceptance criteria and define the outcome of the test.
- The penalty system (*Section 3.6*): to acknowledge the limitation of the methods and models used to assess the performance of the CI and eventually penalize the output of the risk assessment.

3.2 THE USE OF MULTIPLE EXPERTS IN ST@STREST: EU@STREST

The involvement of multiple experts is critical in an assessment when potential controversies exist and the regulatory concerns are relatively high. In order to produce robust and stable results, the integration of experts plays indeed a fundamental role in managing subjective decisions and in quantifying the epistemic uncertainty capturing *'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). To this end, the experts' diverse range of views and opinions, their active involvement, and their formal feedbacks need to be organized into a structured process ensuring transparency, accountability and independency.

Within STREST, a formalized multiple expert integration process has been developed, named EU@STREST (Selva et al. 2015, Selva et al., in prep.), for managing subjective decisions and quantifying the epistemic uncertainty, and have been integrated into the stress test Workflow (*Section 3.3*). The goal of this process is to guarantee the robustness of stress test results, considering the potential limitation in the available budget for non-nuclear critical infrastructures. With respect to the different levels in the SSHAC process developed for nuclear critical infrastructures (SSHAC, 1997), it is located between SSHAC level 2 and 3 in terms of experts interaction, but it also makes an extensive use of classical Expert Elicitations, and it is extended to risk and multi-risk analyses.

The core actors in the multiple expert process are the Project Manager (PM), the Technical Integrator (TI), the Evaluation Team (ET), the Pool of Experts (PoE), and the Internal Reviewers (IR). The interactions among these actors are well-defined in the process. The descriptions and the roles of these actors are given below.

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

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

- *Evaluation Team (ET):* The Evaluation Team is a group of analysts that actually perform the hazard, vulnerability and risk assessments required by the ST, under the guidelines provided by the TI. The team is selected by consensus of the TI and PM, and it may be formed by internal resources and/or external experts. In this sense, the ET represents also the interface between the project and the CI authorities, guaranteeing the successful and reciprocal acknowledgement of choices and results.
- *Pool of Experts (PoE):* This pool is formed only if required by the ST-Level. For all the other ST-Ls, its role is covered by the TI. It has the goal of representing the larger technical community within the process. Two sub-pools are foreseen, which can partially overlap: PoE-H (a pool of hazard analysts) and PoE-V (a pool of vulnerability and risk analysts). The PoE-H should have either site-specific knowledge (e.g., hazards in the area) and/or expertise on a particular methodology and/or procedure useful to the TI and the ET team in developing the community distribution regarding hazard assessments. The PoE-V should have expertise on the specific CI and/or on the typology of CI and/or on a particular methodology and/or procedure useful to the TI and the ET team regarding fragility and vulnerability assessments. Individual experts of the pool may also act as proponent and advocate a particular hypothesis or technical position, in individual communications with the TI (referring to SSHAC (1997) documents, the PoE includes both resource and proponent experts). They participate to the interviewing processes (either in remote or through specific meetings) lead by the TI as pool of experts, providing the TI for their opinions on critical choices/issues. If requested by the CI authorities or if irreconcilable disagreements among the experts of the pool emerge during the interviewing processes (in both PHASE 1 and PHASE 2 of the Workflow, see *Section 3*), the TI and PM may decide to organize meetings with the PoE (or parts of it), in order to openly discuss about controversial issues. In this case, the pool acts as a panel, and TI is responsible for moderating the discussion.
- *Internal Reviewers (IR):* One expert or a group of experts on subject matter under review that independently peer reviews and evaluates the work done by the TI and the ET. This group provides constructive comments and recommendations during the implementation of the project. In particular, IR reviews the coherence between TI choices and PM requests, the TI selection of the PoE in terms of expertise coverage and scientific independence, the fairness of TI integration of PoE feedbacks, and the coherence between TI requests and ET implementations. In particular, IR reviews the project both in terms of technical and procedural aspects of the project (actor's independency, transparency, consistency with the project plan). The IR makes sure that the TI has captured the center, body and range of technically defensible interpretations when epistemic uncertainty is accounted for in the ST level. Note that the IR actively plays an important role during the project and thus is part of the project. If regulators or external authorities foresee an external review of the project results, this further review is performed independently and after the end of the project. Here, the internal review by the IR is considered essential also in this case, in order to increase the likelihood of a successful external ex post review.

The CI authorities select the PM. The PM selects the TI and IR and, jointly with the TI, the components of the ET and of the PoE. 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. In all cases, the size of groups depends on the purpose and given resources of the project.

The PM interacts only with the TI and specifically defines all the questions that the project should answer to, taking care of technical and societal aspects (e.g., selection of the ST level, definition of acceptable risks, etc.). The TI coordinates the scientific process leading to answer to these questions, coordinating the ET in the implementation of the analysis, organizing the interaction with the PoE (through elicitations and individual interactions), and integrating PoE and IR feedbacks into the analysis. The ET implements the analysis, following the TI choices. The IR reviews the whole process, in order to maximize the reliability of the results and to increase their robustness. The basic interactions among the core actors are shown in the Fig. 3.2.

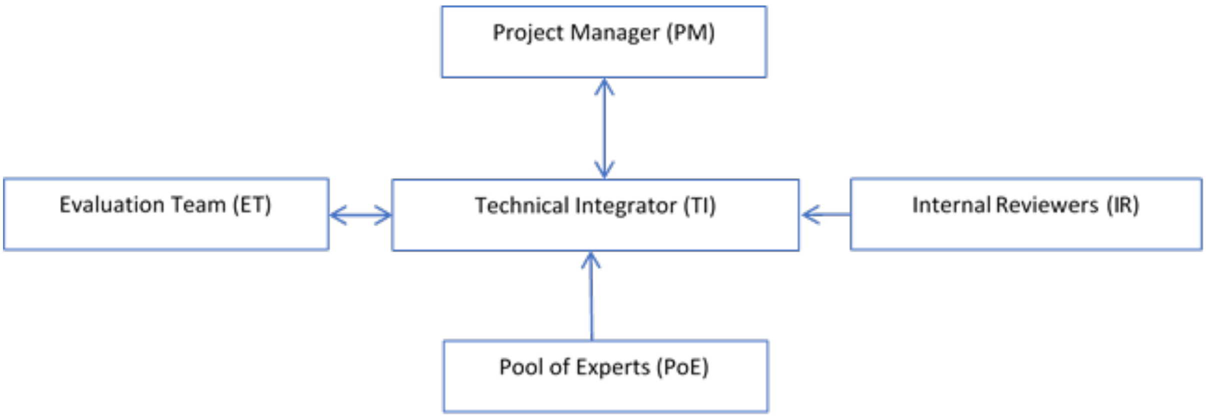

Fig. 3.2 The basic interactions among the core actors in the process of EU@STREST.

3.2.1 Key features of the process

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

Transparency: all choices, data, models and methods are documented; regarding choices, documentation must explicitly report elicitation preparation, formal feedbacks from the PoE, comments from the IR, and consequent decisions of PM and TI;

Independence: The Project Manager (PM), the Technical Integrator (TI), and Internal Reviewers (IR) should be independent;

Responsibility: PM holds the responsibility of the project and about all "political choices" adopted in it (e.g., selection of ST level, of target perils, etc.), taking care of the conformance between the project development and the requests of the funding authorities. TI holds the intellectual ownership of the process and is responsible for all "scientific choices" in the project (e.g., selection of scientific acceptable models, treatment of uncertainty) and for the results. The PoE (if needed in the selected ST-Ls) provides formal inputs to the TI, regarding critical scientific choices and quantifications of the community distribution describing epistemic uncertainty. If the PoE is not needed in the selected ST-Ls, its role is covered by the TI. The ET is responsible for performing the ST following the TI requests. The IR is responsible for the conformance between the scientific development of the project, the ST-L rules, and the EU@STREST guidelines. The PM and TI share the responsibility of the results of the assessment.

Consensus: The PM, TI and IR formally agree on the final products, holding the responsibility of them in their specific roles. Consensus should be reached on both procedural and technical aspects. Regarding the procedural aspects, the consensus should be reached on the completeness of the documentation, the conformance to guidelines and rules, and the fairness of the process. Regarding the scientific aspects, the consensus should be reached on the technical implementation of the analysis as a result of PoE inputs, IR reviews, and PM/TI consequent choices, which take into account also temporal and budget constraints, as well as, potential specific requests of stakeholders.

3.3 WORKFLOW OF ST@STREST

The workflow represents a systematic sequence of steps (processes) which have to be carried out in a stress test. As mentioned before, the ST@STREST workflow comprises four phases: Pre-Assessment phase; Assessment phase; Decision phase; and Report phase. Each phase is subdivided into a number of specific steps, with a total of 9 steps.

In the Pre-Assessment phase all the data available on the CI (risk context) and on the phenomena of interest (hazard context) is collected. Then, the goal (i.e. the risk measures and objectives), the time frame, the total costs of the stress test and the most appropriate Stress Test Level to apply are defined. In the Assessment phase, the stress test is performed at Component and System Level. The performance of each component of the CI and of the whole system is checked according to the Stress Test Level selected in Phase 1. In the Decision Phase, the stress test outcomes are checked i.e. the results of risk assessment are compared with the risk objectives defined in Phase 1. Then critical events, i.e. events that most likely cause the exceedance of a given level of loss value are identified through a disaggregation analysis. Finally, risk mitigation strategies and guidelines are formulated based on the identified critical events. In the Reporting Phase the results are presented to CI authorities and regulators.

The participation of the different actors significantly changes along the different phases of the Stress Test (Fig. 3.3). The PM and TI are the most active participants in the ST workflow. The PM participates in all the steps of the Stress Test until the end (reporting of the results), while the role of TI ends at the end of the Decision phase. The TI is constantly assisted by the ET and supported by the PM, while the level of assistance depends on the ST level. The PoE (if present, see *Section 3.4*) participates in the Assessment and Decision phases. The IR performs a participatory review at the end of Phase 1 and 3. The final agreement, at the end of the Decision phase, is made among the PM, TI and IR.

The workflow and the involvement of main actors and their phase-wise interactions are shown in Fig. 3.3. In the following, a detailed description of the four phases is provided together with a specification of the involvement of the different experts in process.

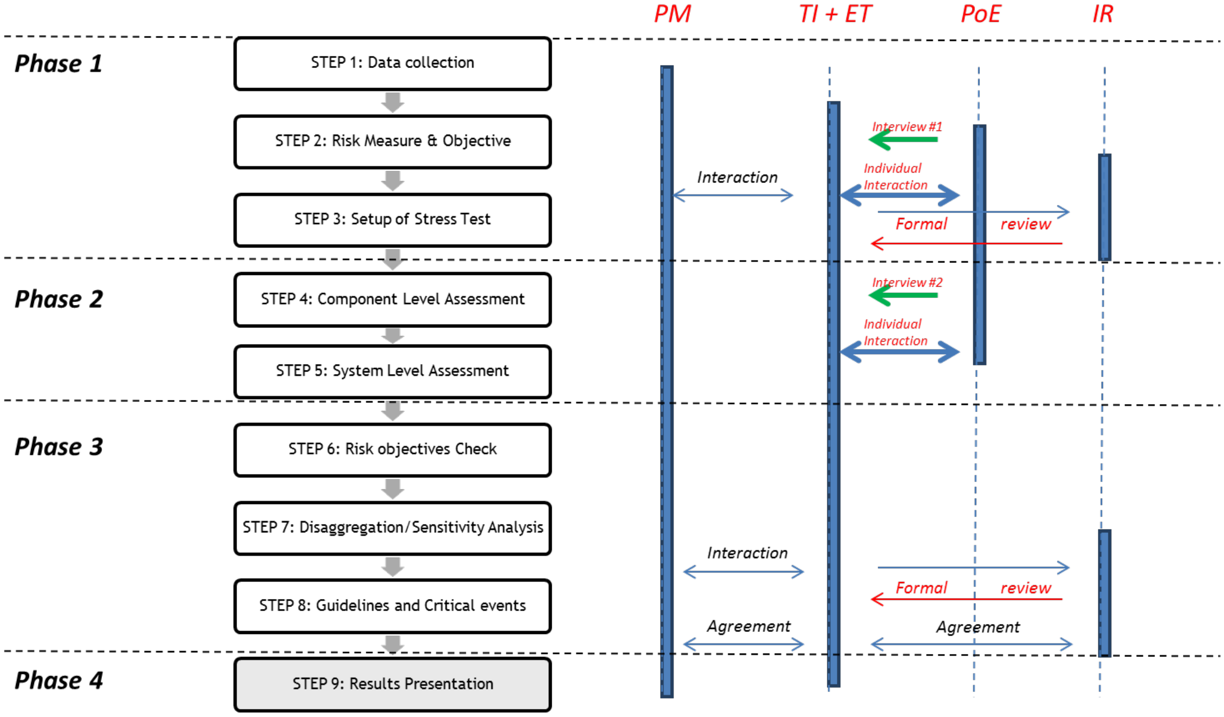

Fig. 3.3 Interaction among the main actors during the multiple-expert process EU@STREST. The PoE is present only in ST sub-levels c and d. For sub-levels a and b, the role of the PoE is assumed directly by the TI.

3.3.1 PHASE 1: Pre-Assessment phase

The Pre-Assessment phase comprises the following three steps:

- *STEP 1) Data collection*: collection of all the data available on the CI (risk context) and on the phenomena of interest (hazard context). Also data coming from Stress Tests performed on other similar CI and/or in the same area are collected. In this step, the participants are selected: the PM selects the TI and the IR; the TI and the PM jointly select the ET. Then, the TI, with the technical assistance of the ET, collects data and relevant information about hazards and CI, and about previous Stress Tests. The TI pre-selects the potential target hazards and the relevant CI components.
- *STEP 2) Risk Measures and Objectives*: definition of one or more risk measures (e.g. fatalities, economic loss, etc.) and objectives (e.g. expected loss, annual probability etc.). This definition is performed by the PM, based on regulatory requirements, the technical and societal aspects and previous Stress Tests.
- *STEP 3) Set-up of the Stress Test*: selection of the Stress Test Level, and Timing and Costs of the project and definition of the "Level of detail" used for the computation of the assessment phase, as presented in Section 2.6. The selection of the ST-Level is made by the PM with the assistance of the TI, based on regulatory requirements.

The finalization of STEP 3 may be a long process and may differ in case the PoE is in place or not (according to the ST- L selected). The presence of the PoE allows for a robust set-up of the ST, based on the quantitative feedbacks of multiple experts. In this case, the PM and

TI set an initial costs and timeframe for the assessments to be performed in STEP 3. The TI selects the PoE and organizes a one-day kick-off meeting with PoE, ET, and PM. With the assistance of PoE, through classical Expert Elicitation, the TI selects the target single and multiple hazards and the relevant CI components and their interactions. If significant disagreements emerge from the elicitation result, the TI may promote specific topical discussions among the members of the PoE, enabling a final decision. Based on this selection, the TI and PM integrate the ET and the PoE to have a complete coverage of the required expertise. The TI collects applicable scientific models and data needed for hazard, vulnerability and risk assessment, with the technical assistance of the ET (and through potential individual interaction with the PoE, if required). At this stage, also potential lacks in modelling procedures are identified by the TI. If technically possible, such lacks should be filled by the TI based on quantification through classical Expert Elicitation of the PoE, which is at this point planned for PHASE 2. Otherwise, a complementary scenario-based assessment should be planned (see *Section 3.4*). The specification about this scenario-based assessment (e.g., the definition of scenarios to be considered) is defined through a specific classical Expert Elicitation planned for PHASE 2.

To complete the planning of actions in PHASE 2, the TI also plans the classical Expert Elicitation of the PoE for ranking alternative models to be used in the stress test, in order to enable the quantification of epistemic uncertainty.

If the selected ST-Level does not foresee the presence of the PoE, this process becomes simplified since all critical decisions are taken directly by one single expert, the TI. The TI selects the target hazards and the relevant CI components. Based on this selection, the TI and PM integrate the ET, to have a complete coverage of the required expertise.

In either case, at the end of these basic choices, the TI collects applicable scientific models and data needed for hazard, vulnerability and risk assessment, with the technical assistance of the ET. Based on this collection, the TI and PM jointly identify the “level of detail and sophistication” used for the computation of the assessment phase (see Section 2.6) based on target costs and model availability. As mentioned above, one of the main goal of this assessment phase is to capture the center and range of technical interpretations that the larger technical community would have if they were to conduct the study. A preliminary sensitivity analysis may help to identify the key parameters which controls the results in order to focus the uncertainty analysis and experts discussions on these key inputs.

All decisions/definitions are specifically documented by the TI. The IR reviews such documents and provides his/her feedbacks regarding the decisions/definitions made thus far. The PM and TI finalize all documents, based on this review. At this point, the final costs and the exact timing for PHASE 2 and PHASE 3 are established. Further, based on the IR review, the PM and TI may evaluate potential changes to the analysis implementation along the assessment phase, in order to avoid potential penalties suggested by the reviewers. In fact, in the case the “level of detail and sophistication” reached in the final implementation is lower than the *level* required, a Penalty System is applied to the output of the risk assessment (STEP 6- *Risk objectives Check).*

3.3.2 PHASE 2: Assessment phase

The Assessment phase is characterized by two steps in which the stress test is performed at Component and System levels according to the Stress Test Level selected in Phase 1. In particular:

- *STEP 4) Component Level Assessment*: the performance of each component of the CI is checked by the hazard-based assessment, design-based assessment or risk-based assessment approach (see *Section 3.4*). This check is performed by the TI or by one expert of the ET selected by the TI.
- *STEP 5) System Level Assessment*: the stress test at the system level is performed. At first, the TI finalizes all the required models. In particular, if the PoE is in place (sub-levels c), the TI organizes the classical Expert Elicitations in order: i) to fill potential methodological gaps, ii) to quantify the potential scenario for the SBHA, and iii) to rank the alternative models to enable the quantification of the epistemic uncertainty. The PoE performs the elicitation in remote. Open discussions among the PoE members (moderated by the TI) are foreseen only if significant disagreements emerge in the elicitation results. If the PoE is not in place but EU assessment is required (sub-level b), the TI directly assigns scores on the selected models for ranking. Then, the ET (coordinated by the TI) actually implements all the required models and performs the assessment. If specific technical problems emerge during the implementation and application, TI may solve them through individual interactions with members of the PoE (if foreseen at the ST-Level).

3.3.3 PHASE 3: Decision phase

The Decision Phase is characterized by three steps:

- *STEP 6) Risk objectives Check*: comparison of results of the Assessment phase with risk objectives. This task is performed by the TI, with the technical assistance of the ET. Depending on the type of risk measures and objectives defined by the PM (F-N curve, expected value, etc.) and on the level of "detail and sophistication" adopted to capture the center and range of technical interpretations, the comparison between results from probabilistic risk assessment with these goals may differ (see *Section 3.6*). One possibility to assess the compliance with risk measures and objectives is presented in *Section 3.5* where the outcome of the stress test is presented by grades (e.g. AA – negligible risk, A – as low as reasonably practicable (ALARP) risk, B – possibly unjustifiable risk, C – intolerable risk).
- *STEP 7) Disaggregation/Sensitivity Analysis*: identification of critical events. This task is performed by the ET coordinated by the TI. Critical events that most likely the exceedance of the considered loss value are identified through a disaggregation analysis (see *Appendix B*) and based on them, risk mitigation strategies and guidelines are then formulated. If specific technical problems emerge during the application, the TI may solve them through individual interactions with the PoE (if present). This step is not mandatory. It depends on the results of STEP 6 (*Risk objectives Check*). For example, if the outcome of STEP 6 is that the critical infrastructure passes the stress tests, performing STEP 7 may be informative, but is not required.
- *STEP 8) Guidelines and Critical events*: risk mitigation strategies and guidelines are formulated based on the identified critical events. This task is performed by the TI, with the technical assistance of the ET.

All the results in all the steps of PHASE 2 and PHASE 3 are specifically documented by the TI. The IR reviews the activities performed in assessments from STEP 4 to STEP 8. The TI, with the technical assistance of the ET, update to the final assessments for such steps accounting for the review. Final assessments and decisions are documented by the TI. Based on such documents, the PM, TI and IR make the final agreement.

3.3.4 PHASE 4: Report phase

The Report phase comprises one step:

- *STEP 9) Results Presentation*: presentation of the outcome of stress test to CI authorities,regulators and community representatives. This presentation is organized and performed by PM and TI. The presentation includes the outcome of stress test in terms of the grade, the critical events, the guidelines for risk mitigation, and level of “detail and sophistication” adopted in the stress test.

Note that the time for this presentation is set in PHASE 1, and it cannot be changed during PHASE 2 and 3.

3.4 MULTI-LEVEL APPROACH: STRESS TEST LEVELS

Due to the diversity of types of critical infrastructures and the potential consequence of failure of the CIs, the types of hazards and the available resources for conducting the stress tests, it is not optimal to require the most general form of the stress test for all possible situations. Therefore, three stress test variants, termed Stress Test Levels (ST-Ls) are proposed:

- Level 1 (ST-L1): single-hazard component check
- Level 2 (ST-L2): single-hazard system-wide risk assessment
- Level 3 (ST-L3): multi-hazard system-wide risk assessment

Each ST-L is characterized by a different objective (component or system) and by a different complexity of the risk analysis (the consideration of multi hazard and multi risk events) as shown in Fig. 3.4.

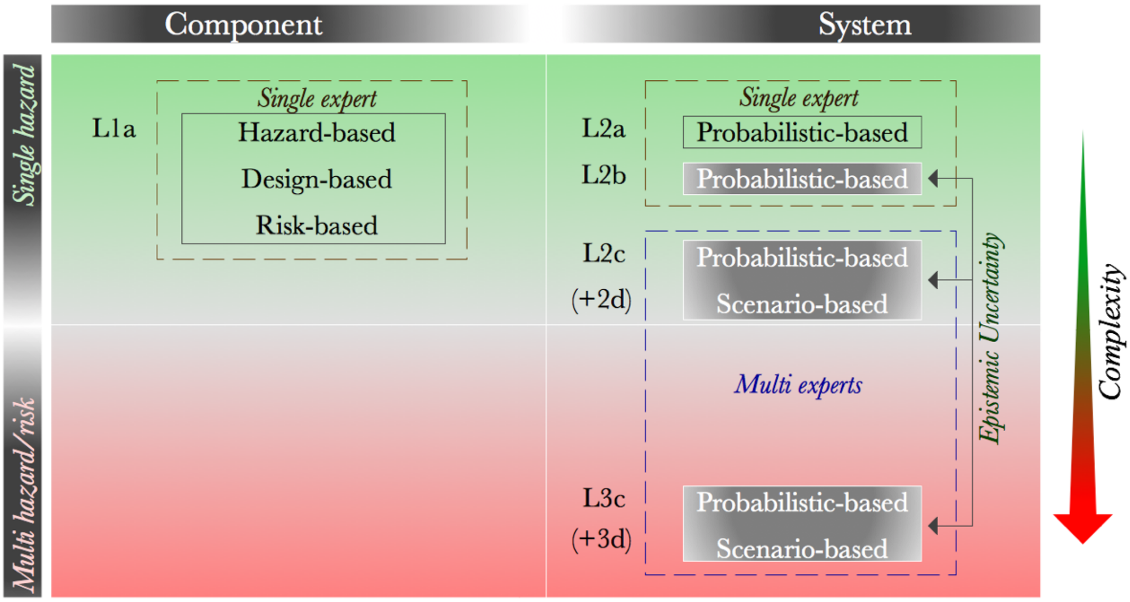

Fig. 3.4 ST-Levels in the ST@STREST methodology

The aim of the ST-L1 (Component Level Assessment) is to check each component of a CI independently in order to show whether the component passes or fails the minimum requirements for its performance, which are defined in current design codes. The performance of each component of the CI is checked for the hazards selected as the most important (e.g. earthquake or flood, etc.).

The stress test at the system level assessment of the critical infrastructure is then foreseen at ST-L2 or ST-L3 where the probabilistic risk analysis of the entire CI (system) is performed. The system level assessment is highly recommended, since it is the only way of make emerging the majority of the mechanisms leading to potential unwanted consequences. However, note that it requires larger knowledge and resources (financial, staff) for conducting the stress test, thus it is not obligatory (if not required by regulations). At these levels, potentially different implementations are possible.

The quantification of EU may not be performed (sub-level a). If performed, it may be either based on the evaluations of a single expert (sub-level b) or of multiple-experts (sub-level c). Indeed, a more accurate quantification of the technical-community distribution describing the EU can be reached if more experts (representing the technical informed community, see SSHAC 1997) are involved in the analysis and, in particular, in dealing with all critical choices. Further, in case specific needs have been identified in the pre-assessment phase (e.g. important methodological/modelling gaps) and such requirements cannot be included into the risk assessment for whatever reason, scenario-based analysis should be also performed as complementary to the ST-L selected (sub-level d). In this case, multiple experts define and evaluate possible scenarios. These additional scenarios are meant to further investigate EU, by including potential processes otherwise neglected only for technical reasons. Thus, it is foreseen on as complementary to a full quantification of EU in a multiple-expert framework.

The system level analysis is thus performed according to

1. The degree of complexity of the analysis (single vs. multi hazards), and

2. The degree of involvement of the technical community in taking critical decisions and in the quantification of the Epistemic Uncertainty (EU) for the computation of risk.

According to these two aspects a subdivision for ST levels has been introduced (Table 3.1, Fig. 3.4).

Table 3.1 ST-Levels subdivision

		Number of Experts		
		Single-expert		Multiple-expert
	Epistemic Uncertainty	No	Yes	Yes
ST-L	1	1a	-	-
	2	2a	2b	2c (+2d)
	3	-	-	3c (+3d)

The selection of the actual procedure to be implemented (row and column in Table 2) is performed in the Pre-Assessment (PHASE 1), STEP 3. These two choices (made by the PM with the assistance of the TI) essentially depend on regulatory requirements, potentially based on the different importance of the CI and the available human/financial resources to perform the stress test. A practical tool to support the choice of the appropriate ST level may be represented by a criticality assessment aimed at identifying and ranking CIs (for example at a national scale). In Appendix A, some existing models available in literature to prioritize CIs at national level are briefly discussed. Further, some key factors that may be considered to define the criticality of the CIs and a possible methodology to rank CIs are presented and discussed.

In the following, a specific description of all ST-Ls and sub-levels is reported.

3.4.1 Component level assessment

At Component level assessment only one implementation is foreseen, i.e. the ST-L1a. This level does not require large knowledge and resources (financial, staff, experts) for conducting the stress test, but it is obligatory because design of (most) CI components is regulated by design codes, and usually, both the data and the experts are available. Further, for some CIs, the computation of system-level analysis (single and multi-risk) could be overly demanding in terms of available knowledge and resources.

Only the TI is required as expert contributing to critical scientific decision, while the whole process may require up to five experts to assist the TI in technical decisions. The TI selects the most important hazard to consider in the component-level analysis but, if more than one hazard is considered critical for the CI under study, more than one Level 1 check should be performed.

Three methods to perform the single-hazard component check are proposed in ST@STREST, and they differ for the complexity and the data needed for the computation. The possible approaches are: the hazard-based assessment, design-based assessment and the risk-based assessment approach. A detailed description is provided in the following.

Later on (*Section 3.4.1.1*), a component level assessment is demonstrated by means of an example of a precast building located in Ljubljana. The results of the assessment are used in *Section 3.5.3* for demonstrating the grading system at the component level.

The performance of the component is checked by comparing the design value of intensity of the hazard which was actually used in the design of the component (building, pipeline, storage tank, etc.), $I_{Design\ phase}$, to the design value of intensity of the hazard prescribed in current regulatory documents or to the value of intensity according to the best possible knowledge, $I_{Assessment\ phase}$. The complexity of such an assessment phase is not high. As a consequence, the level of “detail and sophistication” of this type of assessment is considered moderate, since all other design factors (e.g. minimum requirements for detailing, material safety factors, design procedures, type of analysis, safety margin) and their impact on the performance of the components, which can also change from different versions of regulatory documents, are neglected in the assessment. The outcome of this type of assessment phase is qualitative:

- In compliance with the design level of hazard ($I_{Assessment\ phase} \leq I_{Design\ phase}$),
- Not in compliance with the design level of hazard ($I_{Assessment\ phase} \geq I_{Design\ phase}$),
- The design level of hazard is unknown. This outcome is assigned when there is no regulatory document which would require design of the component for considered type of hazard at the time of performing the stress test.

Hazard-based assessment may be used in cases when the component has not been designed using modern design codes and when the component is not significant for the system response. In such cases, the target level of detail is expected to be set to Moderate (see Section 3.6), which would allow the method to be used. However, if the target level is set to High or Advanced, a more accurate method should be used (i.e. design- or risk-based assessment, respectively) to evaluate the components and avoid imposition of penalty factors. Moreover, due to the trend of increasing design levels of hazard over time, the outcome of the hazard-based assessment is expected, in a vast majority of cases, to be “Not in compliance with the design level of hazard”, which would, again, require a more accurate method to be utilized.

Design-based assessment: The level of “detail and sophistication” of this type of assessment is higher than the previous method since it is based on the design state-of-practice. The expert compares the demand, D, with the capacity, C, (expressed in terms of forces, stresses, deformations or displacements). The assessment can be based on factoring the results from the existing design documentation or by performing design (assessment) of the component according to current state-of-practice. The decision-making regarding the sufficiency of the investigated component is sometimes difficult, since the demand in the design is most often based on linear-elastic analysis while the performance objectives of the component are often associated with their nonlinear behaviour. Alternatively the performance assessment can be based on nonlinear method of analysis. The complexity of this type of assessment may differ, depending on the type of analysis (linear, nonlinear) used. The outcome of the design-based assessment is qualitative:

- In compliance with the code ($D \leq C$),
- Not in compliance with the code ($D \geq C$),

- The design objectives for this type of hazard are not defined. This outcome is assigned when there is no regulatory document which would require design of the component for considered type of hazard at the time of performing the stress test.

Risk-based assessment: The hazard function at the location of the component and the fragility function of the component are required for this type of performance assessment. The level of "detail and sophistication" of this type of assessment varies from Moderate to Advanced, which depends on the level used for evaluation of the hazard function and the fragility function.These two functions can then be convolved in risk integral in order to obtain probability of exceedance of a designated limit state in a period of time (P_{LS}). In general, the risk integral can be solved numerically. Under some conditions, the simple closed-form solutions of risk integral also exist. The target probability of exceedance of a designated limit state for a period of time ($P_{LS,t}$) also has to be defined for each component and different limit states (e.g. loss of function, low/medium/high damage, collapse) if they are considered in this phase of assessment (e.g. probability of exceedance implied by the code). The complexity of risk-based assessment is in general high, but it can be reduced to low when the hazard and fragility functions are already available. Such situation occurs if the ST-L2 or STL3 assessments are also foreseen in the stress test. In this case the ST-L1 assessment and system level assessment should be partly performed in parallel. The outcome of the design-based assessment is quantitative, since the performance of the component is measured by the estimated P_{LS}, which is then used as a basis for the grading (see *Section 3.5*).

The main aspects characterizing the STL-1a are summarized in Table 3.2.

Table 3.2 Main aspects characterizing the Component Level Assessment (STL-1a)

Level	STL-1a
Events considered	Single hazard check. Hazard selected as the most important (e.g. earthquake or flood, etc.). If more than one hazard is important, more than one Level 1 check should be performed.
Number of experts contributing to critical scientific decisions	1 (the TI)
Total number of experts involved in the process	< 5 (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI, and an IR with 1 expert)
Method:	The performance of each component of the CI is checked by the hazard-based assessment, design-based assessment or risk-based assessment approach. Design-based assessment is recommended when only ST-L1 is performed. In the case, when ST-L1 is followed by ST-L2, in which component-specific fragility functions are used, it makes sense to perform risk-based assessment of the components since fragility function are anyway required in ST-L2.
Core actors	PM, TI + ET, IR

3.4.1.1 An example of the component level assessment

In the following the three methods described above for a single hazard check (seismic hazard) are demonstrated by means of an example of a precast building located in Ljubljana (Slovenia).

Description of the component

The component (Fig. 3.5) is a single-storey precast reinforced concrete building with masonry infills on the perimeter. It is located in Ljubljana, Slovenia, and was designed before the introduction of the Eurocode standards. The structure consists of cantilever columns, which are connected by an assembly of roof elements. It has two bays in the X direction and three bays in the Y direction. The distance between the columns in the X and Y directions are 17.4 m and 8.7 m, respectively, whereas the height of the columns amounts to 10.3 m. The critical components of the building are columns and beam-to-column connections. The ratios of longitudinal and transverse reinforcement in all columns amount to 1.29 % and 0.10 %, respectively. No connections between beams and columns are provided. The total mass of the structure amounts to 237 t.

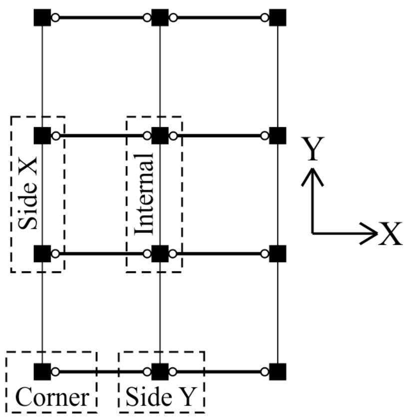

Fig. 3.5 Plan view of the case study building from ST-L1 assessment

Basic information on the seismic hazard

The investigated building (component) is located in Ljubljana, Slovenia. The design peak ground acceleration for the 475 and 2475 year earthquakes amount to 0.25 g and 0.35 g, respectively. The ground is classified as B (CEN, 2005a).

Hazard-based assessment

No information on the hazard used in the design of the component is available. It is therefore concluded that the component was not designed to withstand seismic loading. The outcome of the hazard-based assessment is: The design level of hazard is unknown.

Design-based assessment for Limit State of Near Collapse

In order to conduct the design-based assessment of the component, Eurocode 8 Part 3 (CEN, 2005b) is used. The knowledge level, as required by the code, is identified as »limited«. Consequently, confidence factor amounts to 1.35. The limit state of Near Collapse, which corresponds to the return period of 2475 years, is checked. Lateral force analysis is selected. The detailed explanation of the calculations, which are presented in the following table and equations, is omitted in this report. The reader is thus referred to Eurocode 8 Part 3.

Table 3.3 Detailed explanation of the calculations for the design-based assessment of the component.

Base shear force	$T_1 = C_t \cdot H^{3/4} = 0.05 \cdot 10.3^{3/4} = 0.29s$ $S_e(T_1) = a_g \cdot S \cdot \eta \cdot 2.5 = 1.05g$ $F_b = S_e(T_1) \cdot m = 2441kN$
Distribution of seismic forces on individual columns	$F_i = \frac{F_b}{12} = 203kN$
Torsional effects	$\delta_{X,\text{int}} = 1 + 1.2 \cdot \frac{0}{L_e} = 1 \rightarrow F_{X,\text{int}} = F_i \cdot \delta_{X,\text{int}} = 203kN$ $\delta_{X,ext} = 1 + 1.2 \cdot \frac{L_e/2}{L_e} = 1.6 \rightarrow F_{X,ext} = F_i \cdot \delta_{X,ext} = 325kN$ $\delta_{Y,\text{int}} = 1 + 1.2 \cdot \frac{L_e/6}{L_e} = 1.2 \rightarrow F_{Y,\text{int}} = F_i \cdot \delta_{Y,\text{int}} = 243kN$ $\delta_{Y,ext} = 1 + 1.2 \cdot \frac{L_e/2}{L_e} = 1.6 \rightarrow F_{Y,ext} = F_i \cdot \delta_{Y,ext} = 325kN$

Combination of actions

According to CEN (2004) and CEN (2005a), the following combinations of actions should be taken into account:

Combination 1: $1.0 \cdot G + \Psi_2 \cdot Q + 1.0 \cdot \gamma_I \cdot E_X + 0.3 \cdot \gamma_I \cdot E_Y$ (3.1)

Combination 2: $1.0 \cdot G + \Psi_2 \cdot Q + 0.3 \cdot \gamma_I \cdot E_X + 1.0 \cdot \gamma_I \cdot E_Y$ (3.2)

where G is the characteristic value of permanent action, Ψ_2 is the factor for quasi-permanent value of a variable action, Q is the characteristic value of variable action, γ_I is the importance factor, and Ex and Ey are the characteristic values of seismic action in the X and Y direction, respectively. Importance factor γ_I for the analysed component is equal to 1 and Ψ_2 for category H (roof) is equal to 0. In Tables 3.4 and 3.5 design values of internal forces in columns are presented.

Table 3.4 Design values of internal forces in columns (combination 1)

Column	*N* [kN]	$V_{Ed,X}$ [kN]	$V_{Ed,Y}$ [kN]	$M_{Ed,Y}$ [kNm]	$M_{Ed,X}$ [kNm]
Internal	387	203	73	2092	750
Side X	194	203	98	2092	1010
Side Y	194	325	73	3346	750
Corner	97	325	98	3346	1010

Table 3.5 Design values of internal forces in columns (combination 2)

Column	*N* [kN]	$V_{Ed,X}$ [kN]	$V_{Ed,Y}$ [kN]	$M_{Ed,Y}$ [kNm]	$M_{Ed,X}$ [kNm]
Internal	387	62	243	634	2485
Side X	194	62	325	634	3346
Side Y	194	98	243	1010	2485
Corner	97	98	325	1010	3346

Check of acceptability of type of analysis used for the estimation of design actions

Ductile mechanism is the flexural one. The mean values of concrete compressive strength and reinforcement yield strength are 53 MPa and 456 MPa, respectively. As shown in Table 3.6, the ratios between ρ_{max} and ρ_{min} (see Eurocode 8 Part 3) in both directions are less than 2.5, which means that the use of linear analysis is acceptable.

Table 3.6 Design forces in columns (combination 2)

Column	Capacity of columns in terms of M_{Rd} [kNm]	$\rho_X = \frac{M_{Ed,Y}}{M_{Rd}}$	$\rho_Y = \frac{M_{Ed,X}}{M_{Rd}}$
Internal	411	5.09	6.10
Side X	371	5.64	8.97
Side Y	371	8.97	6.77
Corner	350	9.60	9.60
	ρ_{max}/ρ_{min}	1.89 < 2.5	1.57 < 2.5

Verification of beam-to-column connections

The beam-to-column connections are considered brittle. Their capacity is determined by the following expression:

$$C_{conn} = \frac{N \cdot k_{tr}}{n_{conn}} \tag{3.3}$$

Where N is the axial force in the column, k_{tr} is the coefficient of friction and n_{conn} is the number of connections above the column. Note that this expression is based on the Coulomb friction law and is not taken from CEN (2005b), since friction connections are not included in the code. The coefficient of friction between the column and the beam is taken equal to 0.5 (fib, 2008). The demand on a single connection is determined as follows:

$$\mathrm{D}_{conn} = \frac{F_{Ed}}{n_{conn}} \tag{3.4}$$

where F_{Ed} is the demand on a single column in terms of shear force, which is calculated from the bending capacity of the column. As shown in Table 3.7, connections above the corner columns do not meet requirements, which means that the building does not comply with the code. Further assessment of the columns is not performed since it can be concluded that the outcome of the design-based assessment is: Not in compliance with the code (D ≥ C).

Risk-based assessment

Fragility function for the collapse limit state and hazard function are required in order to estimate the risk. In this case fragility function is determined by conducting non-linear dynamic analyses using a set of hazard consisting ground motions. The numerical model of the building is defined using the principles described in Babič and Dolšek (2016) and Crowley et al. (2015). Based on the results of the numerical simulations a regression analysis is carried out by assuming a lognormal distribution and by using the maximum likelihood method as proposed in previous studies (e.g. Baker, 2015). The geometric mean of the spectral accelerations in both horizontal components at 1.9 s is chosen as an intensity measure. The parameters of the resulting fragility function (Fig. 3.6a), i.e. the median $\overline{IM_C}$ and the standard deviation in the log domain β, are 0.22 g and 0.40, respectively.

Table 3.7 Verification of beam-to-column connections

Column	Capacity of columns in terms of M_{Rc} [kNm] (material characteristics are multiplied by the confidence factor)	D_{conn} [kN]	C_{conn} [kN] (material characteristics are divided by the confidence factor)
Internal	527	26	143
Side X	487	47	72
Side Y	487	24	36
Corner	466	45	36

The seismic hazard curve (Fig. 3.6b) is determined based on the probabilistic seismic hazard analysis (PSHA) used for the development of seismic hazard maps in Slovenia (Lapajne et al. 2003). It is idealized by a straight line on a log-log plot, which can be expressed as follows:

$$\mathrm{H}(IM) = k_0 \cdot IM^{-k} \tag{3.5}$$

Interval from 0.25·$\overline{IM_C}$ and 1.25·$\overline{IM_C}$ was chosen for the idealization of the hazard curve, as proposed by Dolšek and Fajfar (2008). The parameters of the idealized hazard curve $k_{0,C}$ and k_C amount to $4.8 \cdot 10^{-5}$ and 1.75, respectively.

The resulting probability of exceedance of the collapse limit state is determined as follows:

$$P_C = \mathrm{H}(\overline{IM_C})\,\exp(0.5\,\mathrm{k}_C^{\,2}\beta_C^{\,2}) = 8.5 \cdot 10^{-4} \tag{3.6}$$

3.4.2 System level assessment

The system level assessment requires more knowledge and resources for conducting the stress test in respect to the Component Level Assessment, then it is not made obligatory. However, it represents the only way of revealing the majority of the mechanisms leading to potential unwanted consequences. Therefore it is highly recommended.

Different implementations are possible, according to:

- The consideration of a single hazard (STL-2) or multiple-hazard/risks (STL-3).
- The quantification of epistemic uncertainty may not be performed (sub-level a).
- The use of a single expert (sub-level b) or of multiple-experts (sub-level c) to quantify the epistemic uncertainty.

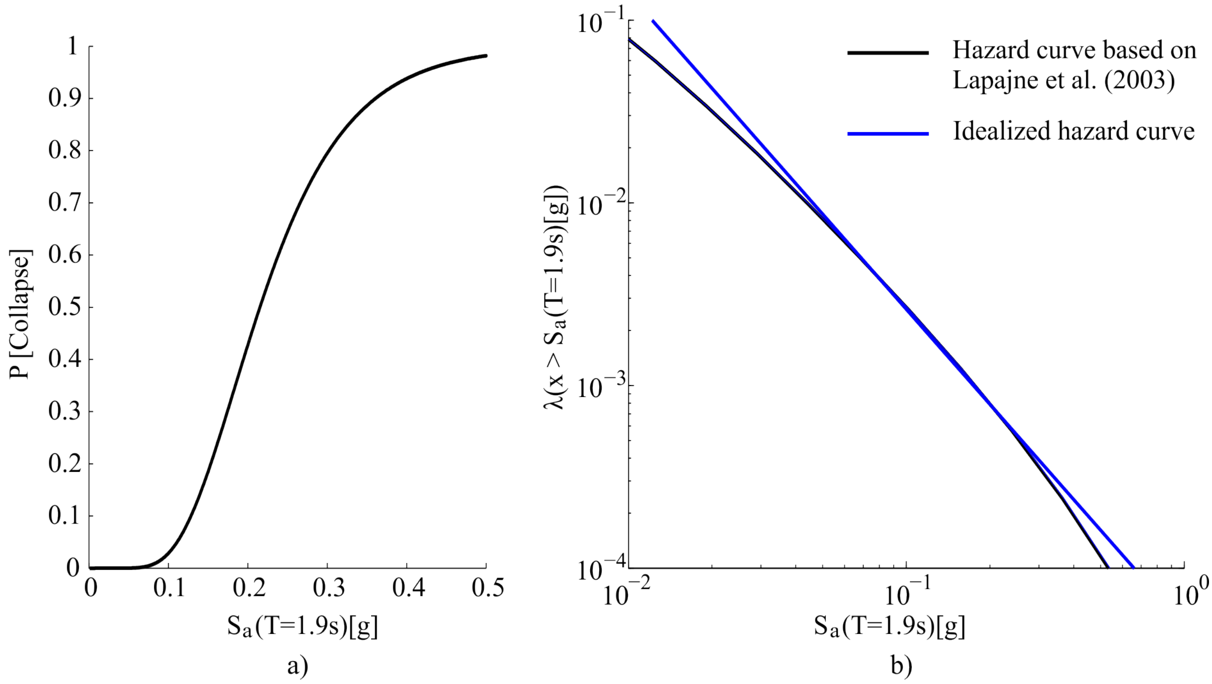

Fig. 3.6 a) Fragility function of the building and b) seismic hazard on the location of the building

3.4.2.1 Single hazard (STL-2)

For the single hazard check, three sublevels are foreseen according to the degree of involvement of the technical community in taking critical decisions and in the quantification of the Epistemic Uncertainty (EU) for the computation of risk. The quantification of EU may not be performed (STL-2a). If performed, it may be either based on the evaluations of a single expert (STL-2b) or of multiple-experts (STL-2c).

As for STL1a for the STL-2a, only the TI is required as expert contributing to critical scientific decision, while the whole process may require up to five experts to assist the TI in technical decisions. ST-L2b, instead, requires the use of up to nine experts (the ET formed by few individuals internal to the CI and a few external experts, and an IR with > 1 expert) to assist the TI. The ST-L2c, requires even more knowledge and resources. In this case more than six experts are required to contribute to scientific decisions (the TI and a PoE formed by at least six experts), while the whole process may require more than ten experts.

Regarding the methods to apply for the risk analysis, for all the sublevels the aim is to evaluate the performance of the whole CI. In generic format, the process may be represented and be independent from the application filed. In particular it can be divided mainly in the following steps (AS/NZS 4360): definition of context, definition of system, hazard identification, analysis of consequences and analysis of probability (or frequency), risk assessment and risk treatment. For each of the main steps several methods/techniques exist and they can be classified as qualitative or quantitative methods (Faber and Stewart, 2003). A list of some of the methods usually applied in risk assessment of engineering facilities is provided in Table 3.8.

In ST@STREST, for all sub-levels of the System level Assessment, probabilistic (i.e. probabilistic risk analysis, PRA) methods are foreseen. PRA is a systematic and comprehensive methodology to evaluate risks associated with every life-cycle aspect of a complex engineered entity, where severity of the consequence(s) and their likelihood of occurrence are both expressed qualitatively (Bedford and Cooke, 2001). It can be also found in the literature under the names of quantitative risk assessment (QRA) or probabilistic safety assessment (PSA).

Table 3.8 List of some of the methods usually applied in risk assessment of engineering facilities.

Method	Type	Description
What if analysis	Qualitative	Determines the system values that have the greatest impact on the results of normal system operation. Input values are varied and ranked in terms of the magnitude and sign of their effects.
Checklist	Qualitative	Hazard and possible consequences identified and arranged in checklist using knowledge gained from the analysis of other similar systems. The list can always be added to and updated by new knowledge and it can be used as input to more rigorous hazard analysis techniques.
HAZOP (Hazard and Operability Studies)	Qualitative	This procedure starts from a fully description of the process and then questioning every part of it in order to determine how many deviations can arise from normal system operation, and where these deviations can arise Once identified, it is determined whether a particular deviation will have a negative effect on the system.
PHA (Preliminary hazard analysis)	Qualitative/Semi-quantitative	PHA identifies all the potential hazards that may lead to an accident. The

		events are then ranked according to their severity.
FMEA (Failure Mode and Effect Analysis)	Quantitative	Identifies potential failures in a system. The system is divided in elements and the functionality of the system is defined in levels. For each element (in sequence) and level, failure modes are identified. The probability of failure for each element may also be estimated and risk can be assessed.
FMECA (Failure Mode and Effect and Criticality Analysis)	Quantitative	Extension of FMEA analysis to include a criticality analysis: the probability of failure modes are charted against the severity of their consequences.
FTA (Fault Tree Analysis)	Qualitative/Quantitative	Fault Tree Analysis (FTA) examines, displays and evaluates failure paths in a system. It follows a logical scheme that links the top event (i.e. the failure) to the causes.
ETA (Event Tree Analysis)	Qualitative/Quantitative	Based on event tree, representation of all the events which can occur in a system after a failure has occurred.
RBD (Reliability block diagram)	Quantitative	Tool that performs system reliability analysis for large and complex systems. It shows the logical connections of components needed to fulfil a specific system function.
Bow-tie	Quantitative/Qualitative	To identify the major accidents and the barriers. Graphical tool used to illustrate an accident scenario, starting from accident causes and ending with its consequences.
Dynamic Bow-tie	Quantitative	Failure probability of primary events of bow-tie, leading to the top event, and the failure probability of safety barriers are periodically updated as new information becomes available over time. The resulting updated bow-tie is used to estimate the posterior probability of the consequences which in turn results in an updated risk profile.

The final result of a PRA is a risk curve and the associated uncertainties (aleatory and epistemic). The risk curve generally represents the frequency of exceeding a consequence value as a function of the consequence values. PRA can be performed for internal initiating events (e.g. system or operator errors) as well as for external initiating events (e.g. natural hazards).

Main applications of PRA have been performed in different fields such as civil, aeronautic, nuclear, and chemical engineering. The specific quantitative method to use ultimately depends upon the context in which the risk is placed (the hazard context), and upon the system under consideration. In civil engineering, PRA methods were developed for the analysis of structural reliability, using analytical or numerical integration, simulation, moment-based methods, or first- and second-order methods (FORM/SORM). In earthquake engineering, the state of the art of probabilistic and quantitative approaches for the estimation of seismic risk relies on performance-based earthquake engineering (PBEE). PBEE is the framework that enables engineers to assess if a new or an existing structure is adequate in the sense that it performs as desired at various levels of seismic excitation. Different analytical approaches to PBEE have been developed in the last years: the approach pursued by the Pacific Earthquake Engineering Research (PEER) Center is the most representative and applied one (Cornell and Krawinkler, 2000). This approach was originally developed for buildings (i.e., point-like structures). However in the years, there was a significant body of research focusing on risk assessment of infrastructure systems and PBEE has extended to spatially distributed systems such as, gas or electric networks (Esposito et al., 2015; Cavalieri et al., 2014) transportation networks (Argyroudis et al., 2015) and telecommunication networks (Esposito et al., in prep.).

For the three CI classes identified in STREST and for the specific hazard considered, the detail list and explanation of all possible methods that may be applied to assess the performance and the risk of the CI, are provided in Deliverable 4.1 (Salzano et al., 2016), 4.2 (Kakderi et al., 2016) and 4.3 (Crowley et al., 2016).

Epistemic Uncertainties are treated only at ST-L2b and ST-L2c. The goal is the assessment of the "community distribution", that is, a distribution describing "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). Here, for "community distribution" is meant the probability distribution representing the epistemic uncertainty within the community. This goal is performed by 1) selecting a number of appropriate alternative scientifically acceptable models, and 2) weighting them according to their subjective credibility. The selection of models may be based on the development of Alternative Trees (more details can be found in Deliverable 3.1, Selva et al. 2015), where the analysis is divided into a number of consecutive steps, and alternative models are defined at each step. The procedure to be followed in these tasks is different for ST-L2b and ST-L2c. In ST-L2b, the TI (supported by the ET) selects the models based on a literature review, and assigns the weights to each one of them. At ST-L2c, a more robust procedure is foreseen (D3.1). In PHASE 1 (pre-assessment), a preliminary list of models is prepared by the TI (supported by the ET), which is formally screened by the PoE and reviewed by the IR. Then, at the beginning of PHASE 2 (assessment), an expert elicitation experiments of the PoE is organized by the TI to assign the weights of the models (for example, following an AHP procedure, see D3.1). Then, the ET implements models and weights in order to produce the "community distribution", implementing methodologies like the Logic Tree (e.g., Bommer and Scherbaum, 2008) or the Ensemble Modelling (Marzocchi et al. 2015). Note that the selection of the models depends on the adopted strategy for their integration (see Selva et al., 2015). For example, Logic Trees require that models form a MECE (Mutually Exclusive and Collectively Exhaustive) set, while Ensemble Modelling simply requires that models form an unbiased set of alternatives representing the epistemic uncertainty into the community.

The main aspects characterizing each sub level of the STL-2 are summarized in Tables 3.9-3.11.

Table 3.9 Main aspects characterizing the System Level Assessment, STL-2a

Level	ST-L2a
Events considered	Single hazard, selected as the most important (e.g earthquake, flood)
Number of experts contributing to critical scientific decisions	1 (the TI)
Total number of experts involved in the process	Up to 5 (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI, and an IR with 1 expert)
Method:	Probabilistic Risk Analysis (PRA, e.g. PBEE framework for seismic hazard)
Core actors	PM, TI + ET, IR

Table 3.10 Main aspects characterizing the System Level Assessment, STL-2b.

Level	ST-L2b
Events considered	Single hazard, selected as the most important (e.g. earthquake, flood)
Number of experts contributing to critical scientific decisions	1 (the TI)
Total number of experts involved in the process	Up to 10 (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI and a few external experts, and an IR with > 1 experts)
Method:	Probabilistic Risk Analysis (PRA, e.g. PBEE framework for seismic hazard) + epistemic uncertainty
Core actors	PM, TI + ET, IR

Table 3.11 Main aspects characterizing the System Level Assessment, STL-2c.

Level	ST-L2c
Events considered	Single hazard, selected as the most important (e.g earthquake, flood)
Number of experts contributing to critical scientific decisions	> 6 (the TI and a PoE formed by > 5 experts)
Total number of experts involved in the process	> 10 (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI and a few external experts, the PoE formed by > 5 experts, and an IR with > 1 experts)

Method:	Probabilistic Risk Analysis (PRA, e.g. PBEE framework for seismic hazard) + epistemic uncertainty
Core actors	PM, TI + ET, PoE , IR

3.4.2.2 Multiple hazards/risks

As for the STL-2c, in this case the process requires more than six experts to contribute to scientific decisions (the TI and a PoE formed by at least six experts), and a total of more than ten experts (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI and a few external experts, the PoE formed by > five experts, and an IR with > one expert) to complete the whole process.

There is no standard approach for multi-risk assessment. Different methods could be used, taken from the scientific literature: e.g., Liu et al. (2015) to identify the multi-risk assessment level required (semi-quantitative vs. quantitative), Marzocchi et al (2012) combined to Selva (2013) when the number of interactions at the hazard and/or risk levels remains limited, and Mignan et al. (2014; 2016a) when the number of interactions becomes consequent (roughly more than 3-4 domino effects). The Bayesian approach of Marzocchi et al. (2012), Marzocchi/Selva has the advantage to build upon the PEER method (Cornell and Krawinkle 2000; Der Kiureghian 2005), which is already well known to seismic engineers and simply relates to the previous stress test levels (L1 and L2). Selva (2013) proposed a method to test potential individual interactions at the risk level (in vulnerability and/or exposure), and to eventually include them into the assessment based on the PEER formula. Moreover, other PEER-based methods, such as damage-dependent vulnerability methods (Iervolino et al., 2016) and loss disaggregation, can easily be added to such a general multi-risk framework. The Generic Multi-Risk (GenMR) framework developed by Mignan et al. (2014), on the other hand, is purely stochastic (variant of a Markov Chain Monte Carlo method) and not derived from existing single-risk assessment approaches. It is therefore more flexible when including a multitude of perils (i.e., not earthquake-focused) but at the same time, requires some adaptation from the modeler to develop a multi-risk model on GenMR (i.e., all events defined in stochastic event set, all interactions defined in a hazard correlation matrix, process memory defined from time-dependent or event-dependent variables). While GenMR could be used for a seismic multi-risk analysis (see Mignan et al., 2015), advantages become more obvious in more complex cases, such as – for example – interactions between different hazards (e.g., earthquake, flooding, erosion) and different infrastructure elements (e.g., hydropower, spillway and bottom outlet failures) at a hydropower dam (Matos et al., 2015; Mignan et al., 2015). Whatever the method used, the final output should be a probabilistic risk result under the form of probabilities of exceeding different loss levels. The multi-risk loss curves shall then be compared to the ones generated in levels L1-L2 and differences at stress test levels identified. The main cause of risk should be investigated, by disaggregation (e.g., Iervolino et al., 2016) or by GenMR time series ranking and metadata analysis (Mignan et al., 2014; Matos et al., 2015; Mignan et al., 2015; 2016a).

The treatment of EUs in ST-L3c is similar to the one described for ST-L2c. In addition, in PHASE 1, it is foreseen that the selection of the hazards and hazard interactions to be included is based on the results of an expert elicitation procedure of the PoE (for example, based on a qualitative risk analysis made through verbal scale, see the case study of the Harbor facilities of Thessaloniki in D6.1, Pitilakis et al. 2016).

The main aspects characterizing the STL-3 are summarized in Table 3.12.

Table 3.12 Main aspects characterizing the System Level Assessment, STL-3c.

Level	ST-L3c
Events considered	Multi-hazards (multi-hazard, i.e. coinciding events and multi-risk)
Number of experts contributing to critical scientific decisions	> 6 (the TI and a PoE formed by > 5 experts)
Total number of experts involved in the process	> 10 (the TI, along with the technical assistance of the ET formed by few individuals internal to the CI and a few external experts, the PoE formed by > 5 experts, and an IR with > 1 experts)
Method:	Multi-risk analysis (extension of PRA methodology for multi-risk) + epistemic uncertainty
Core actors	PM, TI + ET, PoE , IR

3.4.2.3 Scenario-based assessment

Scenario-based analysis may be performed as complementary to STL 2c and 3c, due to methodological gaps identified for specific events/hazards that cannot be formally included into the PRA. This means that it should be considered only if, for technical reasons, one importance phenomenon cannot be included into a formal probabilistic framework (e.g., PRA for ST-L2c). In this case, the choice of performing a scenario-based assessment should be justified and documented by the TI, and reviewed by the IR. If scenario-based assessment is finally selected, the choice of the scenarios should be based on ad-hoc expert elicitation experiments of the PoE (see Selva et al., 2015).

Different strategies can be adopted in organizing the elicitation experiment and in preparing the documentation for the PoE. For example, the hazard correlation matrix (HCM), one of the main inputs to the GenMR framework (see above), can also be used qualitatively to build more or less complex scenarios of cascading hazardous events. The HCM is a square matrix with trigger events defined in rows and target events (the same list of events) in columns. In L3c, each cell of the HCM is defined as a conditional probability of occurrence. In a deterministic view, cells can be filled by plus "+" signs for positive interactions (triggering), minus "-" signs for negative interactions (inhibiting) and empty "Ø" signs for no known interactions (supposedly independent events). The HCM has recently been shown to be a cognitive tool that promotes transformative learning on extreme event cascading. In other words, it allows defining more or less complex scenarios from the association of simple one-to-one interacting couples. Once the modus operandi is understood, more knowledge on multi-risk can be generated (Mignan et al., 2016b). The core actors could use the HCM tool to define the list of relevant events as well as to discuss the space of possible interactions in an intuitive interactive way. L3d scenarios would then emerge from the HCM tool.

The main aspects characterizing the Scenario-based assessment are summarized in Table 3.13.

Table 3.13 Main aspects characterizing the complementary Scenario-based assessment

Level	ST-L2d- ST-L3d
Events considered	Loose "black swan", i.e. events not previously considered (e.g. multi-hazard, correlated events) and for which a PRA is not feasible due to lack of procedures and basic knowledge. This possibility should be confirmed by the IR (Internal Reviewers). The PoE (Pool of Experts) is asked to define such scenarios.
Number of experts contributing to critical scientific decisions	Same of ST-L2c or ST-L3c.
Total number of experts involved in the process	
Method:	Scenario-based risk assessment, SBDA
Core actors	PM, TI + ET, PoE , IR

3.5 GRADING SYSTEM

The first outcome of the stress test, obtained in the STEP 6 - Risk Objectives Check, is determined within a grading system, proposed herein, and is based on the comparison of results of risk assessment with the risk objectives (i.e. acceptance criteria) defined at the beginning of the test (i.e. STEP 2 - Risk Measures and Objectives).

The proposed grading system (Fig. 3.7) is composed of three different outcomes: Pass, Partly Pass and Fail. The CI passes the stress test if it is classified into grade AA or A. The former grade corresponds to negligible risk and is expected to be the goal for new CIs, whereas the latter grade corresponds to risk being as low as reasonably practicable (ALARP, Helm, 1996; Jonkman et al., 2003) and is expected to be the goal for existing CIs. Further, it is proposed that the CI partly passes the stress test if it receives grade B, which corresponds to possibly unjustifiable risk. Finally, the CI fails the stress test if it is given grade C, which corresponds to intolerable risk.

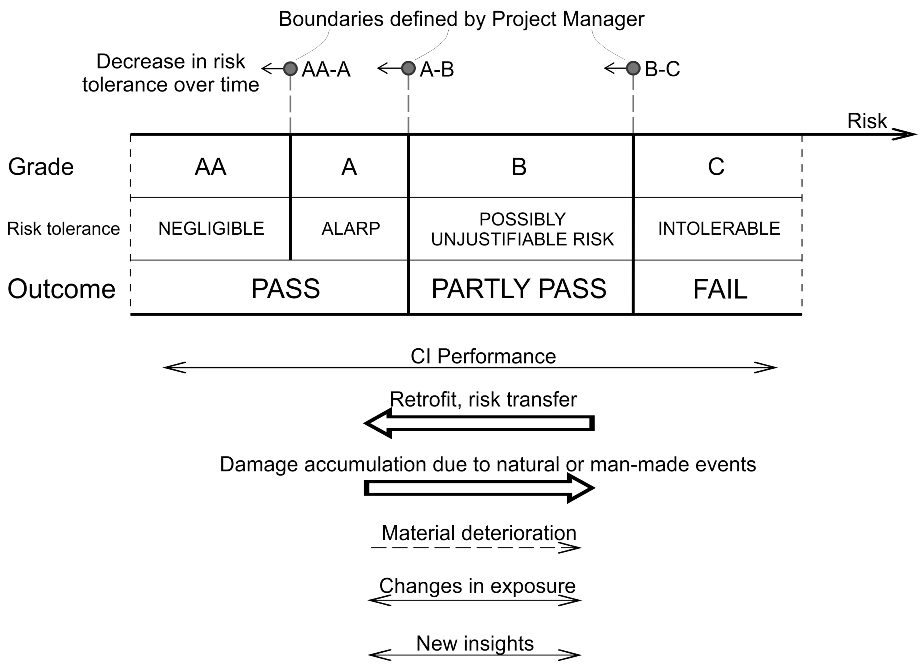

Fig. 3.7 An example of grading system for the global outcome of stress test. The CI may pass, partly pass or fail the stress test.

In the following sections, risk limits and boundaries between grades are first discussed. This is followed by the description of how time domain is applied to the grading system. The guidelines for the grading of individual components are then given and a generalization of the grading system is made in order to apply it to even to those ST levels which take into account epistemic uncertainties and system analysis. Finally a brief discussion is given.

3.5.1 Risk limits and boundaries between grades

The project manager (PM) of the stress test defines the boundaries between grades (i.e. the risk objectives) by following requirements of regulators. The boundaries (i.e. the acceptance risk levels, see *Section 2.1.2)* can be expressed as scalar (Fig. 3.7 and 3.8, top) or continuous (Fig. 3.8, bottom) measures. Examples of the former include the annual probability of the risk measure (e.g. loss of life) and the expected value of the risk measure (e.g. expected number of fatalities per year), whereas the latter is often represented by an F-N curve, where F represents the cumulative frequency of the risk measure N per given period of time. In several countries, an F-N curve is defined as a straight line on a log-log plot. However, the parameters of these curves, as well as parameters of scalar risk objectives (i.e. regulatory boundaries in general) may differ between countries and industries (see *Section 2.1.2*). Harmonizing the risk objectives of risk measures across a range of interests on the European level remains to be done. This is a task for regulatory bodies and for industry association: they should reconcile the societal and industry interest and develop mutually acceptable risk limits. When acceptance criteria are defined as continuous measures, the grade is assigned based on the position of the farthest point from the F-N limits.

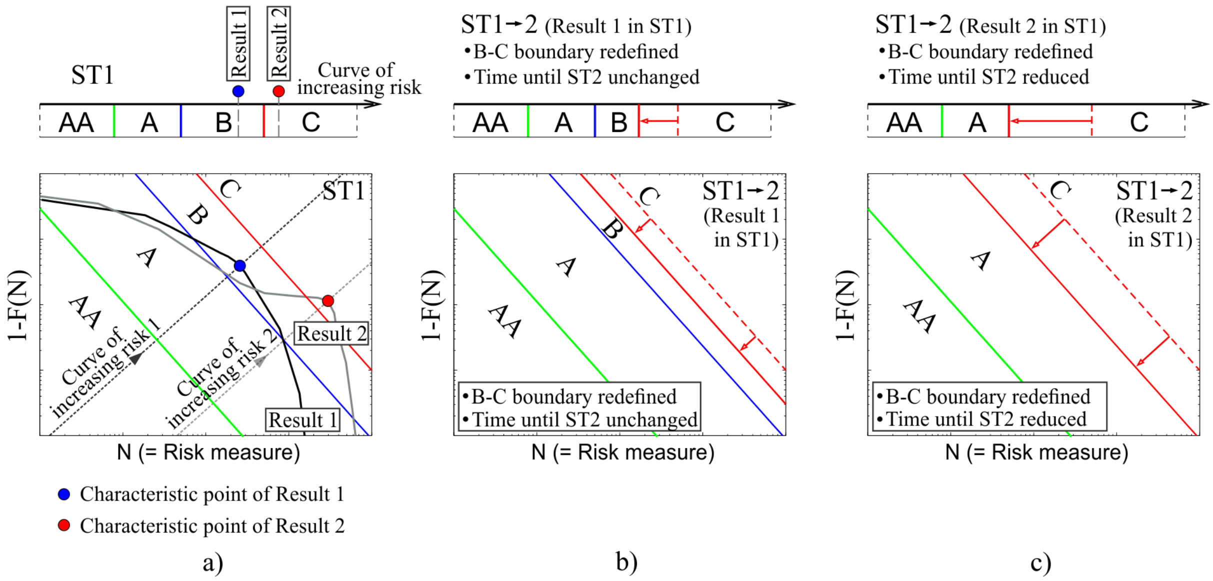

Fig. 3.8 Grading system in time domain using scalar risk objectives (top) and limit F-N curves (bottom): a) two different results of the first evaluation of stress test (ST1), b) redefinition of the parameters of the grading system due to Result 1 in ST1, and c) redefinition of the parameters of the grading system due to Result 2 in ST1.

3.5.2 Grading system in time domain

In general, the CI performance can be understood as time-variant. It may change due to several reasons, such as ageing long-term degradation process (e.g. corrosion, previous hazard events), man-made events (e.g. terroristic attacks), change in exposure (e.g. population) that may increase the probability of failure or loss of functionality during its lifetime (Fig. 3.7). In the proposed grading system, it is foreseen that the performance of the CI or performance objectives can change over time. Consequently, the outcome of the stress test is also time-variant. For this reason, stress test is periodic, which is also accounted for by the grading system. If the CI passes (grade AA or A) a stress test, the risk objectives for the next stress test do not change until the next stress test. The longest time between successive stress tests should be defined by the regulator considering the cumulative risk. However, most of existing CIs will probably obtain grade B or even C, which means that the risk is possibly unjustifiable or intolerable, respectively. In these cases, the grading system has to stimulate stakeholders to upgrade the existing CI or to start planning a new CI in the following stress test cycle. It is proposed that stricter risk objectives are used or that the time between the successive stress tests is reduced in order to make it possible that stakeholders adequately upgrade their CIs in few repetitions of stress test, which means that the CI will eventually obtain grade A or the regulator will require that the operation of CI be terminated.

The basis for redefinition of risk objectives in the next evaluation of stress test is the so-called *characteristic point of risk*. In the case, when scalar risk measures are used, the characteristic point of risk is represented directly by the results of risk assessment (Fig. 3.8, top). In the case when result of risk assessment is expressed by an F-N curve, the characteristic point is defined by one point of the F-N curve. In general, each curve of increasing risk (see Fig. 3.8) results in one point of the F-N curve. The curve of increasing risk, associated with the characteristic point is denoted as the *characteristic curve of increasing risk*. It is recommended that the point associated with the greatest risk above the

ALARP region be selected as the characteristic point (see Fig. 3.8a). In this case the characteristic point is defined as the point of the F-N curve which is the farthest from the limit F-N curve that represents the boundary between grades A and B (A-B boundary) (see blue line in Fig. 3.8a).

Once the characteristic point is determined, the grading system parameters for the next evaluation of stress test can be defined. If the CI obtains grade B in the first evaluation of stress test (ST1, blue dot in Fig. 3.8a), the grading system foresees the reduction of the boundary between grades B and C (B-C boundary) in the next stress test (ST2, Fig. 3.8b). This reduction should be equal to the amount of risk beyond the ALARP region assessed in ST1. This ensures risk equity over two cycles, which may be expressed by the following expression:

$$R_{ST1} - R_{(A\text{-}B)} = R_{(B\text{-}C),\,ST1} - R_{(B\text{-}C),\,ST2} \tag{3.7}$$

where $R_{(A\text{-}B)}$ is the A-B boundary, $R_{(B\text{-}C),ST1}$ and $R_{(B\text{-}C),ST2}$ are the B-C boundary in ST1 and ST2, respectively, and R_{ST1} is the value of risk measure assessed in the ST1. Note that the left side of the Eq. 3.7 is equal to the amount of risk beyond the ALARP region assessed in ST1. Furthermore, if grade C (red dot in Fig. 3.8a) is given in ST1, both the B-C boundary and the period until ST2 are reduced (Fig. 3.8c). In this case, the B-C boundary is set equal to the A-B boundary, since this is the maximum possible reduction of the region of possibly unjustifiable risk. Moreover, the reduced period until ST2 ($t_{cycle,redefined}$) is determined on the basis of equity of risk above the ALARP region over two cycles and can be calculated using the following expression:

$$t_{cycle,redefined} = t_{cycle,initial} \cdot \frac{R_{(B\text{-}C),\,ST1} - R_{(A\text{-}B)}}{R_{ST1} - R_{(A\text{-}B)}} \tag{3.8}$$

where $t_{cycle,initial}$ is the initial amount of time between two stress tests.

3.5.3 Grading of the components

Each component is assessed by at least one method (hazard-based, design-based or risk-based assessment). Objectives of a hazard-based assessment and a design-based assessment are obtained directly from codes, whereas risk objectives need to be defined in Step 2 (*Risk Measures and Objectives*). Similar as in the case of system level assessment, three thresholds need to be defined (between grades AA and A, between grades A and B and between grades B and C).

If a less detailed and sophisticated method assessment (see *Section 3.6*) results in the component not being in compliance with the requirements or the requirements are unknown, a more sophisticated method may be used. For the Component-Level Assessment (STEP 4), three levels of details are defined as Moderate, High and Moderate-Advanced for hazard-based assessment, design-based assessment and risk-based assessment, respectively. Different levels in the case of risk-based assessment exist due to various levels of complexity of hazard and fragility analysis. If the result of a hazard-based assessment or a design-based assessment is that the component is in compliance with the requirements, a grade A is assigned to the component. If these types of assessment result in the component not being in compliance with the requirements or the requirements are unknown, a grade C is assigned to the component, or a higher Level assessment is required. However, the grading system at the component level is equal to that proposed in the case of system-level assessment if the Risk-based assessment is used. The proposed procedure for the progressive approach in the case of the assessment at the level of component and the corresponding grading system is illustrated in Fig. 3.9.

If a component is assigned grade C, mitigation actions need to 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 is used, the mitigation has to be made immediately, as the component is not in compliance with the requirements. If a risk-based assessment is used, the time in which the grade has to be improved is determined on the basis of the amount of risk corresponding to the component reaching the designated limit state (see *Section 3.5.2*).

3.5.3.1 An example of grading of the component

In this section the grading of the components is applied to the example provided in *Section 3.4.1.1*. Risk objectives for the component in terms of probability of collapse, which needed to be defined in Step 2, are as follows: 10^{-6} between grades AA and A, 10^{-4} between grades A and B, and 10^{-3} between grades B and C. The most stringent risk boundary is approximately equal to the target probability of collapse, which is foreseen in building codes for frequent or permanent loads, e.g. in Eurocode 0 (CEN, 2004). Those values of acceptable probability of collapse are within a magnitude of 10^{-6}. Such a low probability of collapse cannot be achieved by employing building codes for earthquake-resistant design since the nature of seismic action is completely different than the nature of frequent or permanent load. The probability of collapse for buildings designed according to Eurocode 8 is around magnitude of 10^{-5}. A significantly larger value of target collapse risk (1% in 50 years ($2\cdot10^{-4}$)) was assumed for new buildings in USA (Luco et al., 2007). As a consequence, the risk boundary between grades A and B was set to 10^{-4}, while the risk boundary between grades B and C was increased 5 fold. The probability of collapse 5% in 50 years approximately corresponds to buildings, which were designed and constructed in the third quarter of 20th century.

The procedure is initiated by performing the hazard-based assessment. Since the design level of hazard is unknown, there are two options: settle with grade C or move on to the design-based assessment. We choose the latter. The design-based assessment results in the component not being in compliance with the code, then two options are possible: settle with grade C or move on to the risk-based assessment. We choose the latter. This results in the probability of collapse equal to $8.5\cdot10^{-4}$. Thus, the component receives grade B, which means that no risk mitigation actions are required, but the threshold between grades B and C will be reduced to $2.5\cdot10^{-4}$ in the next stress test.

3.5.4 Grading of the system taking into account epistemic uncertainties

Grading system as presented in *Sections 3.5.1* and *Section 3.5.2* assumes that no epistemic uncertainties are related to the assessed risk. Since ST-L2c and ST-L3c considers the effect of epistemic uncertainties, the grading system needs to be generalized in a way that it accounts for all possible values of the risk measure. The grading criteria in this case are still a matter of discussion. Currently, it is recommended that the mean value of the risk measure be used. Other options, which should be examined in future studies, are discussed in *Section 3.5.5*.

Furthermore, in the application of the time domain to stress test levels that account for epistemic uncertainties it is assumed that the grading system depends on how probable is that the risk limits are exceeded. For this reason, we determine the left side of Eq. 3.7, i.e. the total value of risk above the ALARP region, as the sum of all possible risk values above the ALARP region (dashed area in Fig. 3.10), which are weighted by their probability:

$$R_{ST1} - R_{(A\text{-}B)} = \int_{R_{(A-B)}}^{\infty} p(R)(R - R_{(A\text{-}B)})\, dR \tag{3.9}$$

In the case of an F-N curve, each curve of increasing risk corresponds to a distribution of points from different F-N curves (Fig. 3.10b). The characteristic curve of increasing risk is the curve, which corresponds to the greatest amount of risk above the ALARP region, i.e. where the integral in Eq. (3.9) produces the highest value.

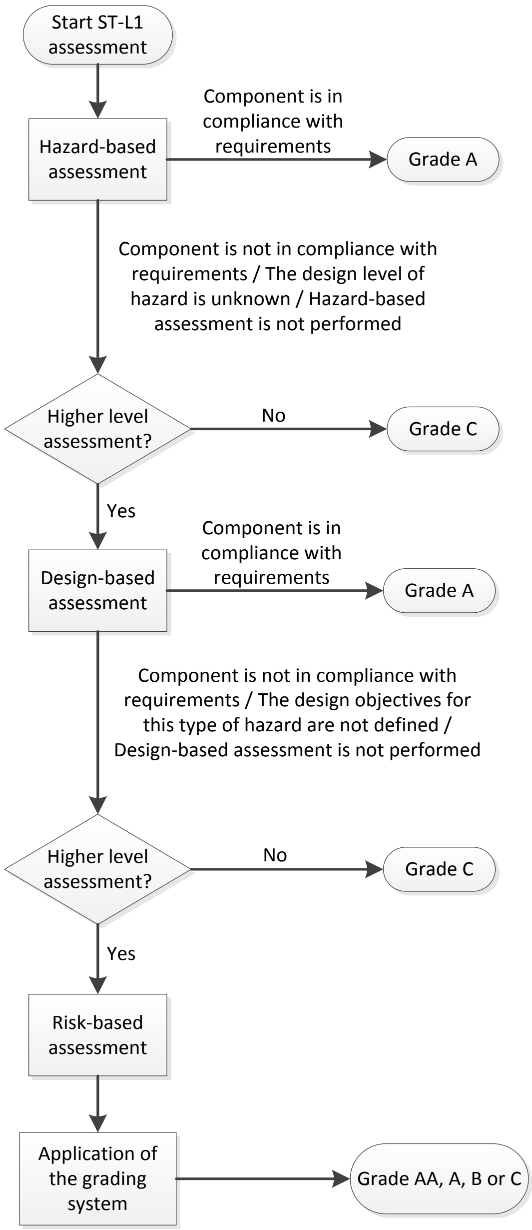

Fig. 3.9 Grading of components of the system (ST-L1)

3.5.5 Discussion and future developments

There are some points of the grading system that need to be discussed and further developed as a part of the future studies.

Firstly, it is yet to be determined how grades of single components should affect the global outcome of stress test. For example, if the CI is assigned grade B in the ST-L2 assessment, the outcome is partly pass. However, one or several components may receive grade C in the component level assessment. It is unclear how this should affect the global outcome. One option would be to change the global outcome of stress test to “fail”, since the stakeholders would be required to reduce risk of those components. However, such an approach may be too conservative. 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 would be defined separately for individual components as well.

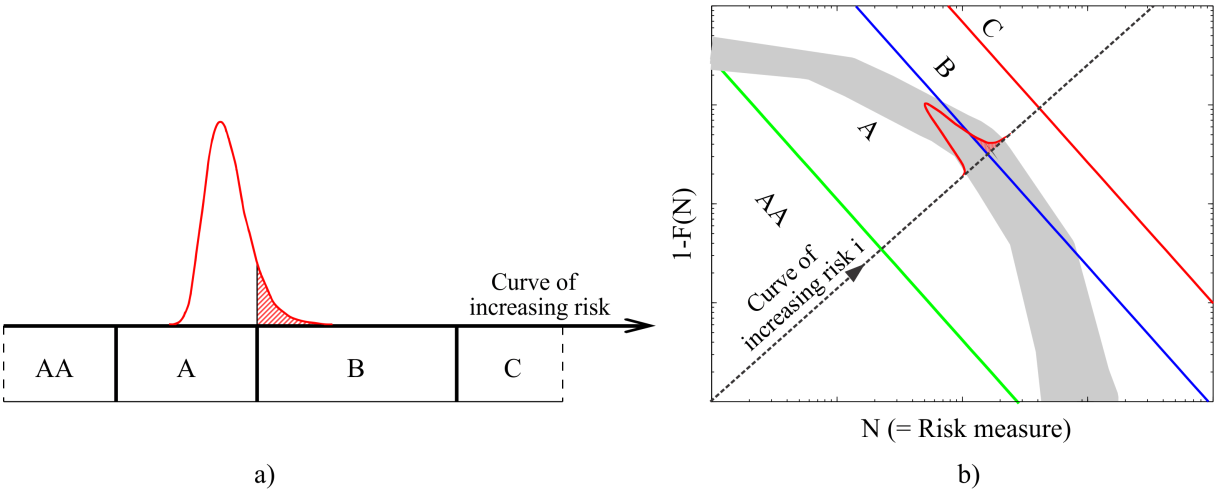

Fig. 3.10 Distribution of a risk measure with boundaries of grades in the case of a) a scalar risk measure and b) an F-N curve

Secondly, in case epistemic uncertainty analysis is of concern, it is currently recommended that the mean value of the designated risk measure is used. However, other options should be discussed. The grade could depend on some other percentiles, which should be determined by the PM. Guidelines for the determination of this percentile should be developed on a comprehensive parametric study as a part of future developments. Grades could also be assigned based on a value of the risk measure corresponding to a specific number of standard deviations above the mean. Again, comprehensive parametric studies would be required to select the appropriate number of standard deviations. Moreover, grades could depend on the type of adjustments of the grading system parameters. For example, if the redefinition of the boundary between grades B and C would be required (based on the amount of risk above the ALARP region, see Fig. 3.8 and 3.10), grade B would be assigned. If the reduction of the time before the next stress test would also be required (again based on the amount of risk above the ALARP region), grade C would be assigned.

Thirdly, the proposed grading system requires boundaries (acceptance criteria) to be defined between regions of negligible, ALARP, possibly unjustifiable and intolerable risk. The PM will often need to rely on their own judgement, when defining these boundaries, especially in societies where regulatory requirements do not yet exist. It is the matter of future developments to create recommendations for the boundaries of different types of performance measures, which can be used by the PM of the stress test as guidelines.

3.6 ST@STREST PENALTY SYSTEM

There is a wide range of methods and models for assessing performance of critical infrastructures against natural hazards. These methods cover different levels of detail and complexity for each hazard, vulnerability, and risk computation. All models are necessarily a simplification of the reality. However, the level of simplification may vary significantly. In fact, different models and methods have to be assumed or introduced to describe how the hazard and vulnerability interact in time and in space. Furthermore, each combination corresponds a different level of detail of the analysis.

For example, regional seismic hazard assessments and site-specific hazard assessments may both represent the input for the risk assessments, however they do differ in the level of details related to the hazard analysis (e.g., the description of the natural variability of sources, the details in modelling the propagation from source to target, etc.). In a similar way, generic fragility functions and element specific fragility function may be used, but again they largely differ in the level of details considered in their quantification. Such differences are expected to significantly influence the reliability of the risk results.

In the STREST methodology, the “level of detail and sophistication” used for the risk computation reflects the level of complexity of the methods adopted for the component and system-level risk assessment. In a general sense, it may be defined as *the trueness and precision, and the repeatability and reproducibility of the results of the risk assessment*.

The selection of the “level of detail and sophistication” to be used in a particular stress test, namely, to perform the hazard and risk analysis, is important because it allows defining how reliable are the results of the Assessment phase of the stress test. At the same time, this is a challenge, since it requires experts that need to have a clear idea about all of the models and methods available in the scientific literature to perform each step of the analysis, i.e. the “*center, body and range”* of the methods and models. The state-of-the-practice methods and models 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 need to characterize the trueness and precision of the state-of-practice methods using multiplicative factors (to shift the mean and adjust trueness) and dispersions (to characterize the precision). More advanced methods should be promoted and less advanced methods should be discouraged by adjusting the factors used to characterize them. Thus, a penalty system is proposed as a part of the ST@STREST methodology.

During the Pre- Assessment Phase (STEP 3: *Set-up of the Stress Test*) the TI and PM select the most appropriate ST-Level for the given CI. As each ST-Level corresponds to a different level of complexity of the hazard and risk analysis, a different level of “detail and sophistication” should be required as a minimum to perform the required analysis.

In particular, in the proposed ST@STREST, a *Target Level (TL)* of “detail and sophistication”, has been associated with each ST-Level, according to the judgement about the complexity of the required hazard and risk analysis. This target value represents the state of knowledge of the community and characterizes the state-of-practice of assessing the CI at the component and the system level.

Then, data, models and methods needed to perform each step of the risk analysis are identified by the TI. These models and methods are characterized by a level of detail that reflects the grade of complexity among the wide range of available methods in the scientific literature. The level of “detail and sophistication” of the Stress Test depends on the specific models selected for the particular test. This selection is mainly based on a scientific ground, but also has practical consequences, such as the requirement of the necessary duration and resources for the stress test. Therefore, the choice of the models should be taken (and documented) jointly by the TI and the PM. Based on the choices made, the TI evaluates the *Effective Level* of detail of the analys*is (EL)*. This assessment is reviewed by the Internal Review (IR) team, and compared with the *TL*. The *EL* should be at least as high as the *TL*.

Based on the IR review, PM and TI may evaluate if changes to the hazard and risk analysis complexity are needed, principally to avoid potential penalties suggested by the reviewers. In fact, if the *EL* attained in the conducted stress test is lower than the *TL* required, the ST@STREST Penalty System is applied.

In the following, the ST@STREST Penalty System, based on the difference between the *EL* and the *TL*, is proposed.

3.6.1 Proposed Penalty System

The proposed ST@STREST penalty system aims to penalize the results of the hazard and risk assessment of the conducted stress test by evaluating a *Penalty Factor* (PF). This factor penalizes simplistic approaches (with respect to the state-of-practice) that cannot guarantee a sufficiently accurate analysis.

The PF is defined by the TI in STEP 6 (*Risk Objectives Check)* of the methodology based on the difference between *EL* and *TL*. Namely, if the *EL* is greater or equal to *TL* the penalty system is not applied.

Levels of “detail and sophistication” and a *Penalty Factor* scheme are proposed in the following. This is just one of the possible schemes that the PM and TI need to determine, the IR to review and confirm, with a possibility to involve the PoE to arrive at the broadest possible consensus. However, the proposed *Levels* and *Penalty Factor* system is general and can be applied in stress test.

3.6.1.1 Proposed Levels

Three categories are defined to describe the trueness, precision, repeatability and reproducibility of the hazard and risk analysis in a stress test:

1. **Advanced**: making use of detailed information and advanced state-of-the-art methods and models in most of the steps of the assessment;
2. **High:** making use of commonly detailed information and state-of-the-practice methods and models in most of the steps of the assessment;
3. **Moderate**: making use of coarse information and simplified methods in most of the steps of the assessment.

Starting from this classification, a *Target Level* has been associated to each ST-Level (Table 3.14) according to the grade of complexity of the risk analysis required. In case a quantitative scale is adopted, a *Factor* interval ($F \in [0,1]$) is set up by the experts and associated to each *Level*. An example is provided in the following:

1. **Advanced** : $F \in [0.7,1]$
2. **High** : $F \in [0.4,0.7)$
3. **Moderate:** $F \in [0.2,0.4)$

These values associated to each level are indicative: in a particular stress test, they need to be determined by consensus between the PM, TI and IR. In general, these values can be studied in more detail, for example, in a study to account for different parameters that affect the results of stress tests. In this case, the resulting *Effective Level* identified for the hazard and risk assessment, i.e. the *EL,* should be at least equal to the lowest *F* (lower bound of the interval) corresponding to the *Target Level*. In this case the *TL* is characterized by lower $\left(TL_{lb}\right)$ and upper $\left(TL_{ub}\right)$bounds.

Table 3.14 Target Levels for each ST-Level

ST-Level	Target Level (TL)
1a	Moderate
2a	Moderate
2b	High
2c	Advanced
2d	Advanced
3c	Advanced
3d	Advanced

3.6.1.2 Effective Level

At component-level (ST-L1) there are three methods to perform the single-hazard component check. These methods differ in the complexity and the data needed for the computation. Therefore, the associated "level of detail and sophistication" is set as follows:

- Hazard-based assessment: **Moderate**
- Design-based assessment: **High**
- Risk-based assessment: **Moderate to Advanced**

The *Effective Level* for the component hazard-based and design-based assessments is moderate (lowest of all possible) and high, respectively. This means that, according to Table 3.14, the hazard-based assessment represents the minimum level of analysis required.

If a risk-based component assessment approach is required, the *Effective Level* may vary according to the level of trueness, precision, repeatability and reproducibility used for the evaluation of hazard and vulnerability. Therefore, the resulting *EL* is a function of the trueness, precision, repeatability and reproducibility of the method adopted for hazard and vulnerability analysis and it may vary from Moderate to Advanced.

For system–level (ST-L2 or ST-L3) stress tests, the evaluation of *EL* is a function of the level of detail selected for each hazard, the method adopted for the epistemic uncertainty quantification, and the method adopted for the multi-hazard/risk evaluation. Furthermore, evaluation of *EL* for each hazard is a function of the level of each step and sub-step needed for the computation of the performance and risk of the CI. In other words, if the computation of risk comprises three principal steps *i* (hazard, vulnerability and risk), and each one of the steps is characterized by *j* different layers*,* the resulting *EL* is a function of the level of "detail and sophistication" of each step *i* and layer *j.* Thus, if a qualitative scale is adopted, the *EL* corresponds to the most frequent (mode) value of the level of detail adopted in each step and layer. If a quantitative scale is adopted (i.e. a quantitative factor is associated with the analysis), the *EL* may be computed (for a single hazard analysis, ST-L2) following Eq. (3.10):

$$EL = W_1 \frac{\sum_{j=1}^{n} w_{1,j} EL_{1,j}}{n} + W_2 \frac{\sum_{j=1}^{m} w_{2,j} EL_{2,j}}{m} + W_3 \frac{\sum_{j=1}^{p} w_{3,j} EL_{3,j}}{p} \qquad (3.10)$$

where *n, m* and *p* are the number of layers in each step (hazard, vulnerability, risk); W_i represent the weight of each step *i* of the risk analysis *and* $w_{i,j}$ the weight of each layer *j* (for each step *i*) set up by experts. If all layers (for each step) are considered equally important, then $w_{1,1} = w_{1,2} = ... = w_{1,\mathrm{n}} = 1$ $w_{2,1} = w_{2,2} = ... = w_{2,\mathrm{m}} = 1$, $w_{3,1} = w_{3,2} = ... = w_{3,\mathrm{p}} = 1$. If all steps are considered equally important, then $W_1 = W_2 = W_3 = 1/3$.

In case of a multi-hazard analysis (ST-L3), AL_E may be obtained as in Eq. (3.11):

$$EL = \frac{H_1 EL^{H_1} + H_2 EL^{H_2} + ... + H_s EL^{H_{1s}}}{s} \qquad (3.11)$$

where H_q represents a weight of each hazard *q* set up by experts. Thus, a multi-hazard *EL* corresponds to the weighted mean of the level of detail evaluated for each hazard EL^{Hq}. If all hazards are considered equally important, then the weights $H_1 = H_2 = ... = H_s = 1$. If the epistemic uncertainty analysis is also of concern, the method of accounting for epistemic uncertainties could be considered as an additional layer.

3.6.1.3 Penalty factor (PF)

The penalty factor (PF) is defined as the difference between the *EL* and the *TL* of the ST level selected.

If a qualitative scale (i.e. Moderate, High, Advanced) is considered, three cases are possible:

a) *TL*=High, *EL* =Moderate
b) *TL* =Advanced, *EL* = High
c) *TL*= Advanced, *EL* =Moderate

The penalty factor may be computed using reference values that may be associated to the three cases, e.g., a) $PF_{H\text{-}M}$=0.2, b) $PF_{A\text{-}H}$=0.2, c) $PF_{A\text{-}M}$=0.4. Again, these values are indicative and they need to be properly set by common experts' consensus.

If the "level of detail and sophistication" is expressed using a quantitative scale, *PF* is defined as the difference between the *EL* and the lower bound of the *TL* of the ST level selected $\left(TL_{lb}^{ST}\right)$,

$$\begin{cases} PF = \left(TL_{lb}^{ST} - EL\right) & if \quad EL < TL_{lb}^{ST} \\ 0 & otherwise \end{cases} \qquad (3.12)$$

Note that the penalty system could be also applied to penalize the CIs that reach the minimum target but just barely, i.e. when $TL_{lb}^{ST} \leq EL < TL_{ub}^{ST}$. In this case, the *PF* may be evaluated considering the upper bound of the *TL* , i.e.,

$$PF = \left(TL_{ub}^{ST} - EL\right) \tag{3.13}$$

3.6.1.4 Penalized Loss, L_P

Consider that the output of the risk assessment at system Level is expressed by the annual exceedance rate of losses (*L*), $\lambda(l)$. For example, in case seismic hazard is of concern, according to the PEER framework (Cornel and Krawinkler, 2000), $\lambda(l)$ is formulated as:

$$\lambda(l) = \int_d \int_{edp} \int_{im} G(l \mid d) \left| dG(d \mid edp) \right| \left| dG(edp \mid im) \right| \left| d\lambda(im) \right| \tag{3.14}$$

where *im* is an intensity measure (e.g., peak ground acceleration, peak ground velocity, spectral acceleration, etc.), *edp* is an engineering demand parameter (e.g., interstorey drift), *d* is a damage measure (e.g., minor, medium extensive, etc.), *l* is the loss variable (e.g., monetary losses, down- town time, etc.), and $G(\mathrm{y} \mid \mathrm{x})$ is a conditional complementary cumulative distribution function (CCDF).

As mentioned before, the risk analysis can be performed adopting different levels of “detail and sophistication”. Thus, it is desirable to include an extra uncertainty, here named penalty uncertainty, which penalize simplistic analysis. Therefore, a new metric is introduced, named penalized loss L_P expressed (in the logarithmic scale) as:

$$\log(L_P) = \log(L) + \varepsilon_P \tag{3.15}$$

where ε_P is the penalty uncertainty. Observe that ε_P acts exactly as a model error. In fact, the objective is to amplify the uncertainties of simplistic approaches that cannot guarantee an analysis with desirable level of “detail and sophistication”.. A convenient choice for the probability distribution of ε_P is the Normal distribution, i.e. $\varepsilon_P \sim N(0, \sigma(l))$, where $\sigma(l)$ is defined as:

$$\sigma(l) = \left| PF \cdot \log(l) \right|, l > 0 \tag{3.16}$$

where *PF* is the penalty factor defined previously. Observe that *PF* acts as a coefficient of variation (c.o.v). Further, in order to focus on the tails of the risk curve, no error is added to the penalty factor for $l = 0$.

Considering that the support of L is usually $[0, +\infty)$ or bounded as $[0, l_{\max}]$, the distribution of ε_P must be truncated according to the support of *L*.

It is of interest to observe that $\sigma(l)$ is proportional to the loss; consequently, the tails are penalized both by the presence of an extra-uncertainty and from a higher $\sigma(l)$.

Then the penalized loss L_P is a new random variable, defined conditionally to the loss value l obtained with the risk assessment. Given this, the conditional cumulative complementary distribution of L_P can be written as:

$$G(l_P \mid l) = 1 - F(l_P \mid l) = P(L_P > l_P \mid L = l) \tag{3.17}$$

and the annual exceedance rate of L_P can be written as:

$$\lambda(l_P) = \int_l G(l_P \mid l) \left| d\lambda(l) \right| \tag{3.18}$$

An example is provided in Fig. 3.11, where the annual exceedance curve of a hypothetical CI has been penalized with different *PF* values. The blue curve corresponds to *PF*=0, i.e. the annual exceedance rate of *L* (Eq. 3.14) where the other three curves represent the annual exceedance rate of the penalized loss L_P expressed in Eq. 3.18.

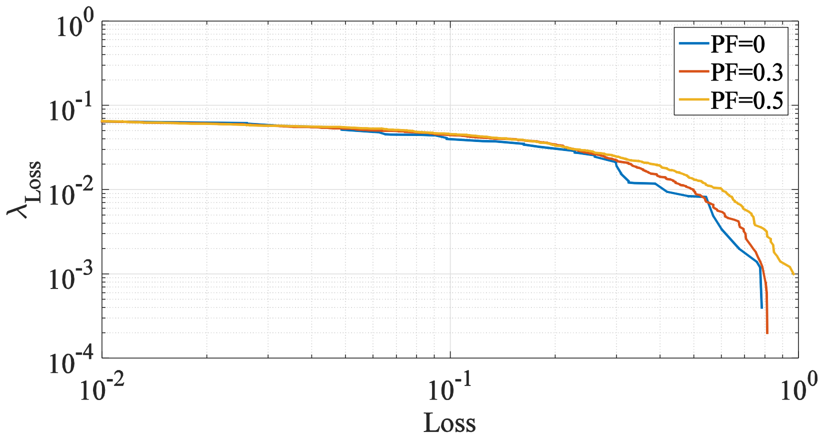

Fig. 3.11 Annual exceedance curves of penalized loss considering different penalty factor values.

3.6.2 Discussion

There are some points of the penalty system that need further discussion and investigations as a part of future studies.

Firstly, the proposed penalty system requires levels of "detail and sophistication" (qualitative and/or quantitative) to be properly set by experts' consensus. Experts must have a clear idea about models and methods available in the scientific literature and their applicability to perform each step of the risk analysis. This may not be feasible for all perils that have to be considered for the stress test. Further, this evaluation should change in each stress test, reflecting the progress of the scientific research. The level of knowledge between two stress tests may change and the levels of "detail and sophistication" scheme should reflect this change.

Secondly, the computation of the *Effective Level* (Eq. 3.10 - 3.11) does not take into account the level of "detail and sophistication "associated to the approach adopted for the multi-risk analysis. This is because the current level of knowledge does not allow ranking these approaches, even though different multi-risk methods have been proposed recently.

Finally, the distribution of the penalized loss has been selected as a Lognormal distribution in this project. Other probability distributions, for example, a Gaussian distribution on the normal scale can be justified as well. Further studies on the determination of the appropriate distribution of the penalized loss should be done.

4 Discussions and future developments

4.1 INCLUDING THE TIME DIMENSION IN ST@STREST

In the interconnected environment that communities present nowadays, when critical infrastructures are affected to extreme events, it seems that these systems are unable to fast recover their functionality either back to the pre-disaster original state, and the relative impact on the whole society can be devastating. The instantaneous loss by itself does not reveal how a community or society responds to a disaster. The *time dimension* represents a key aspect: the time-evolution of community needs and the ability of the critical infrastructures to fulfil these needs (e.g. water, gas, and electricity) is best represented and modelled using the concept of resilience rather than the risk.

The ST@STREST framework proposed herein represent the basis for the development of a new stress test concept that will support decision makers in the evaluation of strategies to improve not only the risk but also the resilience of CIs against natural hazards. In fact, the proposed risk-based methodology (ST@STREST), can be expanded with the aim to test the resilience of CIs to extreme events, i.e. verify the capacity of CIs to anticipate, absorb and adapt to disruptive events to its function, and recover either back to its original state and comparing it to acceptance levels.

The extension of the proposed framework requires mainly the pursuit of the following goals:

- Identification of resilience metrics and standardized methodologies to model the resilience of CIs
- Understanding how community's needs depend on critical infrastructures, defining resilience-based acceptance criteria

The term resilience has increasingly been seen in the research literature and its definition and modelling is the topic of an increasing amount of recent research work. Effort has been devoted to measure the resilience of engineering systems (Hosseini et al, 2016, Francis and Bekera, 2014, Henry and Ramirez-Marquez, 2012, Broccardo et al., 2015, Ouyang and Duenas-Osario 2012, Bruneau et al., 2003), but there is a substantial diversity among the definitions and the modelling of resilience and there is no standardized approaches that suggests how to quantify resilience of CIs in the context of natural hazard.

Regarding the identification of acceptance criteria, the understanding how community's needs depend on the functionality of the CIs represent the key. Businesses activities need suitable facilities and their supply chains and delivery networks; everyone needs a transportation network, electricity, water, gas, and communication networks but, in the aftermath of a disastrous event, some of these services (e.g. water) are more needed.

In the Deliverable 5.4 (Esposito and Stojadinovic, 2016b) these aspects will be argued in more details. An overview of possible approaches to assess the resilience of CI and integrate the evolution in time of the performance of the CI in ST@STREST is presented and discussed.

Further, through the life cycle of the CI, systems operators have the objective to maintain the infrastructure systems and mitigate degradation of system components over time. In the field of civil infrastructures, the life-cycle cost (LCC) concepts and methods have impacted remarkably in the last decades. In particular, in the field of earthquake engineering, seismic risk analysis was established on the basis of LCC. However, it is noted that the seismic risk analysis has not devoted enough attention to the structural maintenance optimization problem (Furuta et al., 2011).

Stress Tests for critical infrastructure systems have herein been proposed as a way to evaluate the performance of these systems against extreme and disastrous events. The outcomes and findings of this tool (such as results of risk analysis, identified mitigation strategies) should be included in the long term maintenance plan in order to increase and optimize the long-term performance of CIs.

In the Deliverable 5.3 (Esposito and Stojadinovic, 2016b) a methodology to integrate stress test outcomes and findings into a unified life-cycle management strategy is proposed and discussed.

5 Conclusions

An engineering risk-based methodology for stress testing critical non-nuclear infrastructures, named ST@STREST, has been developed. In particular, a Multi-Level framework has been proposed where each level is characterized by a different scope (component or system) and by different levels of risk analysis complexity. This allows flexibility and application to a broad range of infrastructures, while producing comparable results. The selection of the appropriate stress test level depends on regulatory requirements, different importance of the CI, and the available human/financial resources to perform the stress test. Further, in order to allow transparency of the stress test process, the data, models, methods adopted for the risk assessment and the associated uncertainty are clearly documented and managed by different experts involved in the stress test process. This allows defining how reliable the results of the stress test are.

The framework is composed of four main phases and nine steps to be conducted sequentially. First the goals, the method, the time frame, and the total costs of the stress test are defined. Then, the stress test is performed at component and system levels; then, the outcomes are checked and compared to the acceptance criteria. A stress test grade is assigned and the global outcome is determined by employing a grading system proposed herein. According to the outcome the parameters of the following evaluation of stress test are adjusted. Further, a penalty system is also proposed to define how reliable are the results of the stress test, and in case it is needed, to penalize simplistic approaches that cannot guarantee an analysis with desirable level of “detail and sophistication”... Finally, the results are reported and communicated to stakeholders and authorities.

ST@STREST has been applied and tested in six critical infrastructures in Europe and it is intended to be improved based on operators feedback.

References

Argyroudis, S., J. Selva, P. Gehl, and K. Pitilakis. 2015. Systemic Seismic Risk Assessment of Road Networks Considering Interactions with the Built Environment. *Computer-Aided Civil and Infrastructure Engineering* 30 (7): 524–540.

AS/NZS 4360. 1999. *Risk management*. Standards Australia, Sydney; 1999.

Babič, A., and M. Dolšek. 2016. Seismic fragility functions of industrial precast building classes. *Engineering Structures*, 118: 357-70.

Baecher, G. B. 1983. Simplified geotechnical data analysis. In *Reliability Theory and Its Application in Structural and Soil Mechanics* (pp. 257-277). Springer Netherlands.

Baker J.W. 2015. Efficient analytical fragility function fitting using dynamic structural analysis. *Earthquake Spectra*, 31: 579-99.

Barlettani M., G. Fruttuoso, and P. Lombardi. 1997. *An ecological approach for risk and impact ranking*, in: B.M. Drottz Sjoberg (Ed.), Proceedings New Risk Frontiers, Society of Risk Analysis, Stockholm, 1997.

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

Berg, H. P. 2010. Quantitative safety goals and criteria as a basis for decision making. *Reliability & Risk Analysis: Theory and Applications*, 1(2): 62-78.

Bohnenblust, H. 1998. *Risk-based decision making in the transportation sector*. In Quantified societal risk and policy making (pp. 132-153). Springer US.

Bommer, J.J. and F. Scherbaum. 2008 The use and misuse of logic-trees in probabilistic seismic hazard analysis. *Earthquake Spectra* 96: 1967–1977

Bommer, J.J., F. Scherbaum, H. Bungum, F. Cotton, F. Sabetta, and N.A. Abrahamson. 2005. On the use of logic trees for ground-motion prediction equations in seismic-hazard analysis. *Bulletin of the Seismological Society of America*. 95: 377-389.

Bottelberghs P.H. 2000. Risk analysis and safety policy developments in The Netherlands, *Journal of Hazardous Materials*, 71: 59–84.

Broccardo, M., P. Galanis, S. Esposito, and B. Stojadinovic. 2015. Resilience-based risk assessment of civil systems using the PEER framework for seismic hazard. *European Safety and Reliability Conference ESREL 2015*, Zurich, Switzerland. 7-10 September.

Broccardo M., L. Danciu, B. Stojadinovic, and S. Wiemer. 2017. Life safety risk metrics as part of a risk governance framework for induced seismicity. *16th World Conference on Earthquake Engineering*, January 9th to 13th 2017, Chile.

Bruneau, M., S. E. Chang, R. T. Eguchi., G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. von Winterfeldt. 2003. A frame-work to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra* 19: 733–752.

Carter D. A. 2002. A worst case methodology for obtaining a rough but rapid indication of the societal risk from a major accident hazard installation, *Journal of Hazardous Materials.* A92: 223–237.

Cavalieri F., P. Franchin, J. A. M. Buritica Cortes, and S.Tesfamariam. 2014. Models for seismic vulnerability analysis of power networks: comparative assessment, *Computer-*

Aided Civil and Infrastructure Engineering. 29(8): 590–607.
CEN 2004. European Standard EN 1990: 2004 Eurocode: Basis of structural design, Brussels, Belgium.
CEN 2005a. European Standard EN 1998-1: 2005 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, Brussels, Belgium.
CEN 2005b. European Standard EN 1998-3: 2005 Eurocode 8: Design of structures for earthquake resistance. Part 3: Assessment and retrofitting of buildings, Brussels, Belgium.
Commission of the European Communities. 2006. A European Programme for Critical Infrastructure Protection, Brussels, December 2006, COM (2006)786 final.
Cornell, C. A., and H. Krawinkler. 2000. Progress and challenges in seismic performance assessment. *PEER Center News*. 3 (2): 1-3.
Crowley H., C. Casotto, M. Dolšek, and A. Babič. 2015. *Guidelines for performance and consequences assessment of multiple-site, low-risk, high-impact, non-nuclear critical infrastructures (exposed to multiple natural hazards, etc.)* Deliverable 4.3 STREST Project.
Dolšek M., and P. Fajfar. 2008. The effect of masonry infills on the seismic response of a four storey reinforced concrete frame-a probabilistic assessment. *Engineering Structures.* 30 (11): 3186–3192.
Esposito, S., I. Iervolino, A. d'Onofrio, A. Santo, P. Franchin, and F. Cavalieri. 2015. Simulation-based seismic risk assessment of a gas distribution network. *Computer-Aided Civil and Infrastructure Engineering DOI*: 10.1111/mice.12105
Esposito, S., and B. Stojadinović. 2016a. *Report on strategies for stress test implementation at community level and strategies to enhance societal resilience using stress tests.* Deliverable 5.4 STREST Project.
Esposito, S., and B. Stojadinović. 2016b. *Tools and strategies to incorporate stress tests into the long-term planning and life cycle management of non-nuclear Cis.* Deliverable 5.3 STREST Project.
Esposito, S., B. Stojadinović, A. Babič, M. Dolšek, I. Safraz, J. Selva, and D. Giardini. 2017 Engineering risk-based methodology for stress testing of critical non-nuclear infrastructures (STREST Project) *16th World Conference on Earthquake Engineering*, January 9th to 13th 2017, Chile.
Esposito, S., A. Botta, M. De Falco, I. Iervolino, A. Pescape, and A. Santo. Seismic Risk Analysis of a Regional Telecommunication Network, *in prep*.
Faber, M. H., M.G. Stewart. 2003. Risk assessment for civil engineering facilities: critical overview and discussion. *Reliability Engineering & System Safety*. 80:173–184.
Fekete, A. 2011. Common criteria for the assessment of critical infrastructures, International *Journal of Disaster Risk Science*, 2: 15–24.
Fib 2008. Structural connections for precast concrete buildings - bulletin N°43. Guide to good practice, Federal Institute of Technology, Lausanne, Switzerland.
Francis R., and B. Bekera 2014. A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety.* 121: 90-103.
Furuta, H., D. M. Frangopol, and K. Nakatsu. 2011. Life-cycle cost of civil infrastructure with emphasis on balancing structural performance and seismic risk of road network. *Structure and Infrastructure Engineering*. 7(1-2): 65-74.
Grimaz, S., D. Slejko. 2014. Seismic hazard for critical facilities. *Bollettino di Geofisica Teorica ed Applicata*, 55 (1): 3-16.

Hakata, T. 2003. Basic considerations on defining safety goals. *Nuclear Technology.* 142: 243 – 249.

Helm, P. 1996. Integrated Risk Management for Natural and Technological Disasters. *Tephra*. 15(1): 4-13.

Henry, D., and J. E. Ramirez-Marquez. 2012. Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering and System Safety*. 99: 114-122.

Hosseini S., K. Barker, and J. E. Ramirez-Marquez. 2016. A review of definitions and measures of system resilience. *Reliability Engineering and System Safety*, 145: 47-61.

Jansen, H.M.A. 1988. Financieel economische risicolimiet (financial economic risk limit, in Dutch), Min. van Volkshuisvesting, Ruimtelijke Ordening en Milieu, directie stralenbescherming, 1988

Jonkman, S. N., P. H. A. J. M. Van Gelder, and J. K. Vrijling. 2003. An overview of quantitative risk measures for loss of life and economic damage. *Journal of Hazardous Materials*. 99 (1): 1-30.

Kabir, G., R. Sadiq, and S. Tesfamariam. 2014. A review of multi-criteria decision-making methods for infrastructure management. *Structure and Infrastructure Engineering*. 10(9): 1176-1210.

Kakderi, K. et al. 2016. *Guidelines for performance and consequences assessment of geographically distributed, non-nuclear critical infrastructures exposed to multiple natural hazards*. Deliverable 4.2 STREST Project.

Institute of Chemical Engineering. 1985. Nomenclature for hazard and risk assessment in the process industries.

Iervolino I. 2016. Soil-invariant seismic hazard and disaggregation. *Bulletin of the Seismological Society of America*. DOI: 10.1785/0120160072 (in press)

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

Lapajne J., B. Šket Motnikar, and P. Zupančič. 2003. Probabilistic seismic hazard assessment methodology for distributed seismicity. *Bulletin Of The Seismological Society of America*. 93 (6): 2502-2515.

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

Luco N, BR Ellingwood, RO Hamburger, JD Hooper, JK Kimball and CA Kircher. 2007. Risk-targeted versus current seismic design maps for the conterminous United States. *Structural Engineers Association of California 76th Annual Convention,* Lake Tahoe, California.Marzocchi, W., A. Garcia-Aristizabal, P. Gasparini, M. L. Mastellone and A. Di Ruocco. 2012. Basic principles of multi-risk assessment: a case study in Italy, Nat. Haz., 62, 551-573, doi: 10.1007/s11069-012-0092-x

Marzocchi W., M. Taroni, and J. Selva. 2015. Accounting for Epistemic Uncertainty in PSHA: Logic Tree and Ensemble Modeling. *Bulletin of the Seismological Society of America*, 105 (4), doi: 10.1785/0120140131.

Matos, J. P., A. Mignan and A. J. Schleiss. 2015. Vulnerability of large dams considering hazard interactions: conceptual application of the Generic Multi-Risk framework, Proceedings of the 13th ICOLD Benchmark Workshop on the Numerical Analysis of Dams, Lausanne, Switzerland, 285-292

Merz, H.A., T. Schneider, and H. Bohnenblust. 1995. Bewertung von technischen Risiken (in German), VDF Hochschulverlag, ETH Zurich, 1995.

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

Mignan, A. et al. 2015. *Report on cascading events and multi-hazard probabilistic scenarios* Deliverable 3.5 STREST Project.

Mignan, A., N. Komendantova, A. Scolobig and K. Fleming (2016a), Multi-Risk Assessment and Governance, Handbook of Disaster Risk Reduction and Management, World Sci. Press & Imperial College Press, London, chapter 16, in press

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

Mota de Sá, F. 2005. Ranking Critical Infrastructures – The Portuguese Methodology. Presidência do Conselho de Ministros, CNPCE.

Moteff, J. 2004. Risk Management and Critical Infrastructure Protection: Assessing, Integrating, and Managing Threats, Vulnerabilities and Consequences: Congressional Research Service, Library of Congress.

OCIPEP 2004. Selection Criteria to Identify and Rank Critical Infrastructure Assets. Public Safety and Emergency Preparedness Canadá, Ottawa, Ontario. Available in http://www.ocipep.gc.ca/critical/nciap/nci_criteria_e.asp

Ouyang M., and L. Duenas-Osario. 2012. A three stage resilience analysis framework for urban infrastructure systems. *Structural Safety*. 36: 23-31.

Pitilakis, et al. 2016. *Integrated report detailing analyses, results and proposed hierarchical set of stress tests for the six CIs covered in STREST*. Deliverable 6.1 STREST Project.

Paté-Cornell, M.E. 1994. Quantitative safety goals for risk management of industrial facilities. *Structural Safety*, 13: 145–157.

Rostek P., J. Markuci., and V. Adamec. 2014. The issue of dependencies when assessing the criticality of infrastructure elements.

Robert, B., M.-H. Senay, M.-È. P. Plamondon, and J.-P. Sabourin. 2003. Characterization and Ranking of Links Connecting Life Support Networks. Ottawa: Public Safety and Emergency Preparedness Canada.

Saaty, T.L. 1980. *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saaty, T. L. 2006. *The analytic network process* (pp. 1-26). Springer US.

Salzano, E. et al. 2016. *Guidelines for performance and consequences assessment of single-site, high-risk, non-nuclear critical infrastructures exposed to multiple natural hazards*. Deliverable 4.2 STREST Project.

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

Selva, J. et al. 2015 *Report on the effects of epistemic uncertainties on the definition of LP-HC events*. Deliverable 3.1 STREST Project.

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

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.
TAW 1985. Technical Advisory Committee on Water Defences, Some considerations of an acceptable level of risk in The Netherlands, TAW, 1985.
Theoharidou, M., P. Kotzanikolaou, and D. Gritzalis. 2009. Towards a Criticality Analysis Methodology: Redefining Risk Analysis for Critical Infrastructure Protection. *Proc. of the 3rd IFIP International Conference on Critical Infrastructure Protection* (CIP-2009), C. Palmer, S. Shenoi (Eds.), Springer, USA, March 2009.
UKCO (United Kingdom. Cabinet Office) 2010. Strategic Framework and Policy Statement on Improving the Resilience of Critical Infrastructure to Disruption from Natural Hazards. London.
US Dept. of Homeland Security, National Infrastructure Protection Plan 2009. Partnering to enhance protection and resiliency, USA, 2009.
Vrijling J.K., W. van Hengel, and R. J. Houben. 1995. A framework for risk evaluation. *Journal of Hazardous Materials* 43: 245–261.
Vrijling, J. K., and P. H. A. J. M. Van Gelder. 1997. Societal risk and the concept of risk aversion. *Advances in Safety and Reliability*. 1: 45-52.
Vrijling, J. K., P. H. A. J. M.van Gelder, L. H. J. Goossens, H. G. Voortman, and M. D. Pandey. 2004. A Framework for Risk Criteria for Critical Infrastructures: Fundamentals and Case Studies in the Netherlands. *Journal of Risk Research.* 7: 569–79.
Zavadskas, E. K., Z. Turskis, and S. Kildienė. 2014. State of art surveys of overviews on MCDM/MADM methods. *Technological and economic development of economy*, 20(1): 165-179.
2008/ 114/ EC 2008. Council Directive of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection.

Appendix A

A Ranking of CIs

Above, the stress test has been classified in three (macro) conceptual frameworks for the safety of non-nuclear Critical Infrastructures (CIs), which are termed Stress Test Levels (ST-Ls). The ST-Levels are defined depending on: i) different objective (component or system), ii) the degree of complexity of the risk analysis (single vs. multi hazards) and iii) the degree of involvement of the technical community in the quantification of the EU for the computation of risk. The selection of the appropriate ST-L and sub-levels is made by the project manager, with the assistance of the Technical Integrator during the Pre-Assessment phase of the stress test. The choice substantially depends on regulatory requirements that should account for the importance/criticality of the type CI. A criticality assessment of the CIs, aimed at identifying and ranking CIs (for example at a national scale) may be a practical tool to support the PM in the choice of the appropriate ST level.

The non-nuclear Critical Infrastructures (CIs) are complex and diverse in nature. It is important to rank them, if the number of CIs being considered is greater than one for performing the stress test. This will support project manager in the pre-assessment phase where to allocate available sources first. The ranking of CIs is a challenging task due to their diverse nature, the potential consequence of failure, the types of hazards posing threat to them, vulnerability state etc.

Different approaches exist in literature to rank them. In the following, we briefly describe the existing models and factors for criticality of CIs available in literature to identify and prioritize CIs at national level. As we are dealing with non-nuclear CIs at European level, we proposed a few criteria to check the criticality of the CIs. Additionally, we also propose a framework - a multi-criterion decision model -that could be employed taking into account the proposed criteria.

A.1 LITERATURE REVIEW FOR IDENTIFICATION AND RANKING OF CIS

A.1.1 Critical factors

Impact factors, also named as critical factors, are criteria used to identify and prioritize CIs to protect them against different kind of risks. Different set of criteria for protection of infrastructure systems have been developed in Canada (Robert et al. 2003), Netherlands (Vrijling et al. 2004), UK (UKCO 2010), US (Moteff 2004, US, 2009), and within EU for all European countries (EC 2006).

In Netherland, the following criteria are used: territorial security (infringement of the Netherlands' territory and the international position), physical safety (fatalities, seriously injured and chronically ill or physical suffering), economic security, ecological security, social and political stability and social psychological impact. All these criteria are evaluated in terms of range and duration.

In US, the National Infrastructure Protection Plan (US, 2009) presents the following criteria to evaluate criticality: public health and safety (effect on human life and physical well-being, e.g. fatalities, injuries/illness) economic (direct and indirect economic losses, e.g. cost to rebuild asset, cost to respond to and recover from attack, downstream costs resulting from disruption of product or service, long-term costs due to environmental damage), psychological (effect on public morale and confidence in national economic and political institutions, and governance/mission (effect on government's or industry's ability to maintain order, deliver minimum essential public services, ensure public health and safety, and carry out national security-related missions.

In UK (UKCO, 2009) the categorization of CIs is done using the Government "Criticality Scale", which assigns categories for different degrees of severity of impact. The Criticality Scale includes three impact dimensions: impact on delivery of the nation's essential services; economic impact (arising from loss of essential service) and impact on life (arising from loss of essential service). The following three factors provide the means to distinguish between different degrees of severity of impact on essential services: the degree of disruption to an essential service, the extent of the disruption, in terms of population impacted or geographical spread and the length of time the disruption persists.

The European Commission (EU, 2006) defines a minimum set of criteria that the member states should take into account on their CI assessments: public effect (number of population affected, loss of life, medical illness, serious injury, evacuation), economic effect (GDP effect, significance of economic loss and/or degradation of products or services), environmental effect (effect on the public and surrounding location), interdependency (between other critical infrastructure elements), (e) political effects (confidence in the ability of government), and psychological effects. These criteria need to be evaluated in relation to the spatial (e.g. local, regional, national and international effect) and time scale (during and after the incident).

The Canadian approach (Robert et al. 2003, OCEPIP, 2004) differentiates from the other because the criteria used are accompanied by impact scale (see section A.1.2). It is characterized by six fundamental criteria: concentration of people and assets (number of potential exposed people); economy impact (direct cost of restoration including critical information and information technology, i.e. service relies on or asset contains critical information and I.T., critical infrastructure sector (service or asset relates to a critical infrastructure sector, interdependency (loss of the asset on other sectors), service delivery (time to repair or replace the asset), public confidence (effect on public confidence in the ability of the relevant government to preserve public health and safety).

Theoharidou et al. (2009) present a summary of critical factors based on a review of current critical infrastructure protection (CIP) approaches enriched with other criteria used in more generic risk methodologies (Table A.1).

The linkage between CIs in terms of interdependencies, dependencies and influences has been emphasized as an important critical factor. In assessing the interdependencies among CIs, the following factors have been found to be very significant (ROSTEK, Petr et al. 2014):

- Time factors;

- o Geographical scale;
- o Cascade effect;
- o Socio-psychological effects;
- o Influence of operational procedures;
- o Trade policy;
- o Back-up and recovery procedures;
- o Government regulation, law, regional policy;
- o Interests of the owners;
- o Services/functions supply;
- o Services/functions demand;
- o Commodities flow, their supply, transfer, storage and consumption;
- o Critical time, critical quality.

Great attention is paid to time factors (so called critical time) when assessing the criticality of the infrastructure element e.g. see Fekete (2011). Time factors represent e.g.: the duration of the failure, the speed of the failure onset, median time to recovery, median time to restoration of function etc.

Table A.1 Target Summary of critical factors used in the recent literature (Theoharidou et al. 2009)

Critical factor	Description
People affected	Number of people that get affected by an incident. It does not evaluate the type of the impact.
Concentration of people	The higher the concentration of people, the greater the potential for catastrophic effects.
Range	It evaluates the geographical scope of an event. It can be quantified in terms of distance (for example {max. 100 km^2, 100 – 1000 km^2, 1000 – 10.000 km^2, > 10.000 km^2} [16]) or in a qualitative way {International, National, Regional, Local}
Economic impact	This criterion measures potential direct economic impact from an incident. It includes the losses to the infrastructure itself from service degradation or loss of assets and information, recovery costs, as well as the estimated loss from cascading effects, so it can be analyzed in these three types of costs.
Interdependency	Likelihood of a high cascading effect resulting from an incident within the sector and across sectors. Types of interdependencies can be (a) physical, (b) cyber, (c) geographic, and (d) logical
Public Confidence	Possible impacts on the public's confidence in the ability of the government to preserve public health and safety, economic security, or to assure the provision of essential services and goods

International relations	Potential effect of an incident in the diplomatic relationships of a state with other countries
Public order	It attempts to estimate the possible implications a loss of a CI may have on the public order of a country. This impacts may be caused both by disclosure of confidential information and of unavailability of critical services to the public (i.e. water supply).
Policy and Operations of Public Service	Ability of the government to maintain its policies and normal operation. It varies from public confidence, as it not evaluates the general belief of the public (psychological effect) but the actual ability of the government to maintain its operations
Safety	It relates to the welfare of individuals when an incident affects the health of the populations; it includes injuries, chronically illnesses and fatalities. It can also refer to pain, grief and suffering of victims
Defense	Possible implications in the ability of a government to protect its population from hostile attacks, either due to unavailability of CIs or by modification or disclosure of critical information
Recovery Time	Time required in order to recover from an incident and it is affected by the availability of substitutes and the cost incurred before the asset or service is restored.
Impact Duration	Duration of the impact effect including long-term effects

A.1.2 Existing models

Different models have been proposed in literature in order to identify and rank CIs at national level. Some of them use criticality assessment as pre-assessment to identify priorities and later merge into a more detailed assessment of risk.

Portuguese Model (Mota de Sá, 2005)

Portuguese model is a single criterion model which is based on the study developed by Mota de Sá (2005). The basis of the model is the robustness and functional methodology, reflected on the mathematical modulation. This modulation includes the construction of an algorithm that allows the evaluation of the cascading effects and interdependencies infrastructure sectors and the infrastructures which belong to the specific sector. The outcome of the Portuguese model is obtained based on the designed algorithm in terms of the subjective probabilities. These probabilities will assist in ranking the more critical infrastructures. The reason to use probabilities consists in the fact that the critical concept includes an open wide of variables (risk, vulnerabilities, threats and consequences) impossible to accurate with precision (Mota de Sá, 2005).

Canadian Model (OCIPEP, 2004)

Canadian model is a multi-criteria model which adopts a group of selected criteria to evaluate the critical infrastructures, in order to capture their level of impact in the well-being of a nation. For each criterion an impact scale is defined, as shown in Table A.2. The infrastructure is classified as “CI” based on the obtained value C_i, that is, the sum of impacts $V_{i,j}$ of each assessment criteria factors (j=1,…,N).

Likewise the Canadian model, many multi-criterion decision models (MCDMs) are used for the ranking or prioritization of elements in different applications (Zavadskas et al., 2014). Some MCDMs are contingent on wholly quantitative and statistical inputs, whereas others, such as the focus explicitly on experts' judgment based on some pre-defined criteria to assign values, to some variables and criteria weights. Most of the decision models have the ability to deal with complex situations through defining important factors such as different criteria and scales, incorporation of different type of data, addressing uncertainties, and involvement several of experts, but the method to employ depends upon the context of the study.
The number of applications of MCDMs is seen to be rise in literature and which also seen in the area of infrastructure management (Kabir et al., 2014). However, in the application of MCDMs, two issues are important to address. The first one is related to selection the critical factors by which CIs will be ranked/identified. As described in the previous section, there are a variety of critical factors that could be considered. However, it again depends on the context of the study and the input requirements of the decision model. Decision maker also plays a prominent role in the selection of critical factors. The second issue concerns the weighting of the defined critical factors. A simple way is to assign weights on equality basis, where all criteria are considered equally. An alternative way to assign weights on different factors is the negotiation among experts involved in the decision making process.

Table A.2 CI Priority Assessment Screening Model

CI Priority Assessment Screening Model				
Impact factor	**Severe**	**High**	**Medium**	**Low**
Score	**15**	**5**	**3**	**1**
Concentration of people and Assets Impact	>10.000 people	Between 1.000 and 10.000 people	Between 100 and 1.000 people	< 100 people
Economic Impact	> $ 1 billion	From $ 100 million to $ 1 billion	From $ 10 million to $ 100 million	< $ 100 million
Critical Infrastructure Sector Impact	Sector may shut down or International	National	Provincial or Regional	Local
Interdependency Impact	Debilitating impact on other sectors	Significant impact or disruption of other sectors	Moderate impact on important missions of other sectors	Minor Impact on important missions of other sectors
Service Impact	High cross-sectorial cost, recovery time longer than 1 year	High cost, long recovery time (months-years)	Medium cost, significant recovery time (days-weeks)	Low cost, brief recovery time (hours-days)
Public Confidence Impact	High national risk & ability control in doubt	Public perceives high national risk & low ability to control risk	Public perceives moderate national risk & moderate ability to control risk	Public perceives low national risk & high ability to control risk

A.2 PROPOSED METHODOLOGY TO RANK CRITICAL INFRASTRCTURES IN ST@STREST

A set of critical factors (criteria) and a multiple-criterion are proposed in the following section. The proposed criteria and decision model are used for the critical assessment of CIs aimed at identifying and ranking CIs (for example at a national and/or European scale), which may provide a practical tool to support the PM in the selection of the appropriate ST level when applying the stress test.

A.2.1 Potential criteria

The proposed twelve potential evaluation criteria are described below. These criteria are divided into four categories which cover different aspects of CIs to evaluate the critical state of the infrastructures. The criteria are compiled on the basis of extensive literature searching to support the PM and these are not the fixed one. The PM can freely add or remove in the set of proposed criteria.

- *Hazard Criteria*

C1 - Natural hazard threats: The internal and external natural hazard threats, which can result in the disruption or destruction of a particular CI.

C2 - Technological threats: The internal and external technological threats, which can result in the disruption or destruction of a particular CI.

- *Vulnerability Criteria*

C3 - Infrastructure sector: The significance or importance of a particular CI sector in which the CI belongs to.

C4 - Structural robustness: The ability to withstand the physical structure of a particular CI against the known and unknown internal and external natural hazards and technological threats.

C5 – Resilience robustness: The ability to anticipate, absorb and adapt to disruptive events to its function, and recover either back to its original state.

- *Impact Criteria*

C6 - Economic effects: The significance of direct economic losses and/or degradation of products or services (e.g. cost to rebuild asset, cost to respond to and recover from natural hazards) as consequences of the disruption or destruction of a particular CI.

C7 – Interdependencies/cascading effects: The significance of indirect economic losses (i.e. downstream costs) of other industrial sectors or companies of the same asset (related to cascading effect resulting from an incident within the sector and across sectors due to interdependencies of the CI with other industries) as a consequence of the disruption of a particular CI.

C8 - Public effects: The significance of the public effects (e.g., physical suffering and disruption of daily life; including the loss of essential services) as consequences of the disruption or distribution of a particular CI.

C9 - Human casualties: The potential number of human casualties in terms of fatalities or injuries as consequences of the disruption or destruction of a particular CI

C10 - Environmental effects: The significance of the potential environmental losses as consequences of the disruption or destruction of a particular CI.

C11 - Scope or spatial effects: The extent of the geographic area which could be affected by as consequences of the disruption or destruction of a particular CI.

- *Resource criteria*

C12 - Available resource and funds to conduct a stress test

A.2.2 Potential framework

An Analytic Hierarchy Process (AHP) (Saaty, 1980) is herein proposed for the ranking/prioritization of non-nuclear CIs. This method has already been suggested in project deliverable 3.1 (Selva et al., 2015) for another application. As the method has numerous applications, here we described it in the context of prioritization of CIs.

The AHP is an effective tool for dealing with complex decision making problem. By reducing the problem to a series of pairwise comparisons, and then synthesizing the results, the AHP helps to capture both subjective and objective aspects of a decision. In addition, the AHP incorporates a useful technique for checking the consistency of the decision maker's evaluations, thus reducing the bias in the decision making process.

The AHP represents the decision problem into a hierarchy and considers: i) a set of evaluation criteria; and ii) a set of alternatives among which the prioritization is to be made. The selection of appropriate criteria is the most important issue in this process.

The prioritization of the alternatives is made based on scores (numerical values), which are estimated using the AHP. The scores are estimated in two steps. First, AHP generates a score for each criterion according to the expert's pairwise comparisons of the all the considered criteria. The estimating un-equal scores on each criterion make it distinguishable from the other existing decision models. Next, for a fixed criterion, the AHP assigns a score to each alternative according to the expert's pairwise comparisons of all the considered alternatives based on that criterion. Finally, the AHP combines the criteria and the alternative score, thus determining a global score for each alternative, and a consequent ranking. The higher the global score of an alternative means the alternative takes the higher place in the ranking. The global score for a given alternative is a weighted sum of the scores it obtained with respect to all the criteria.

In the context of prioritization of CIs, the proposed criteria in the previous section are potential candidates. However, it is up to the analyst(s) to choose appropriate criteria. As mentioned above, the process also requires a set of alternatives from which a CI needs to be identified to be the most critical one. As an example, the set of six CIs in STREST if they are to be ranked in order to implement the stress test.

The Analytic Hierarchy Process assumes that the criteria and alternatives are independent (Saaty, 2006). Simply stated that all aforementioned criteria and CIs are independent from each other, but in many real-world cases, there are interdependencies among the criteria and alternatives. This also means that one of the important proposed criteria that is ''Interdependencies/cascading effects'' cannot be taken into account in AHP. In order to deal with interdependencies among criteria and alternatives, the AHP has explicitly been generalized into another framework called Analytic Network Process (ANP) (Saaty, 2006), which relaxes the assumption of independence. The ANP, like AHP, relies on expert's pairwise comparisons and generates global scores for ranking or prioritization. However, the ANP is represented by a network, rather than a hierarchy.

Appendix B

B Disaggregation Analysis

B.1.1 Loss disaggregation for seismic hazard

In a performance-based earthquake engineering framework, the loss (L) assessment is conveniently split via the terms in the integral of Eq. (B.1). In the equation, $\lambda_{L>l^*}$ is the rate of exceedance of a loss value *l**, ν_i is the rate of occurrence of hazardous event on each of the (usually assumed independent) sources; $f_{IM_i}(x)$ is the probabilistic distribution of event intensity measure or IM (it may be a scalar or a vector depending on the CI) given event occurrence on one of the sources. $f_{ER|IM_i}(y|x)$ is the distribution of the engineering response (ER), that is the performance of the CI, for any possible given value of the event intensity. $P\left[L>l^*|ER=y\right]$ is the exceedance probability of any loss value (consequence) as a function of the performance of the system.[2]

$$\lambda_{L>l^*}=\sum_i \nu_i \cdot \int_{IM}\int_{ER} P\left[L>l^*|ER=y\right]\cdot f_{ER|IM_i}(y|x)\cdot f_{IM_i}(x)\cdot dy\cdot dx$$

(B.1)

It is to finally note that Eq. (B.1) has the minimum number of terms, other representations, with a more refined split of the loss, are also possible. Moreover, adaptions of the same approach to spatially-distributed logically-interdependent CIs are possible keeping the same framework and formal terms.

B.1.1.1 Loss disaggregation in terms of hazard intensity

Once the curve is obtained from Eq. (B.1), it is possible to perform *loss disaggregation*. In the equation $P\left[L>l^*|x\right]$ is the probability of loss exceedance given intensity. Disaggregation is aimed at obtaining the probability that a specific value of a variable involved in the risk assessment is causative for the exceedance of a loss value of interest. This provides a distribution, which enables to identify, for example, the modal values of IM, which are the values of earthquake intensity most likely causing the exceedance of the considered loss value (Fig. B.1).

$$f_{IM|L>l^*}(x)=\frac{\sum_i \nu_i \cdot P\left[L>l^*|x\right]\cdot f_{IM_i}(x)}{\lambda_{L>l^*}} \qquad \text{(B.2)}$$

[2] The equation assumes that the loss is independent of the IM given the system response.

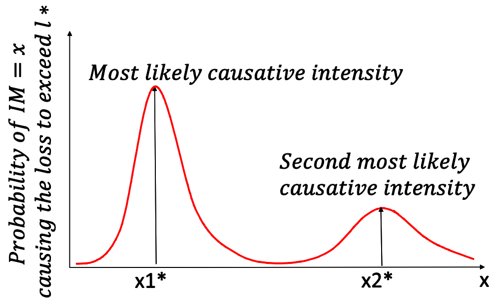

Fig. B.1 Loss disaggregation in terms of event intensity.

B.1.1.2 Loss disaggregation in terms of hazard source and event features

Similar, disaggregation may be computed in terms of magnitude and distance, which may help identifying the sources most likely causing exceedance of a certain loss level, Eq. (B.3); Fig. B.2. In the equation, $P\left[L > l^{*} | w,z\right]$ is the probability of exceedance of the loss given magnitude and distance.

$$f_{M,R|L>l^{*}}(w,z) = \frac{\sum_{i} \nu_i \cdot P\left[L > l^{*} | w,z\right] \cdot f_{M_i,R_i}(w,z)}{\lambda_{L>l}} \quad \text{(B.3)}$$

B.1.1.3 Loss disaggregation in terms of component performance

Even more interestingly, the loss may be disaggregated with respect to system's response, which may help identifying the component the damage of which most likely causes the exceedance of the loss value of interest; Eq. (B.4).

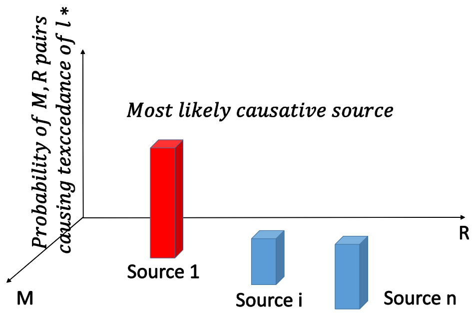

Fig. B.2 Loss disaggregation in terms of hazard source.

In the equation $P\left[L > l^{*} | y\right]$ is the probability of exceedance of the loss given a specific performance of the CI and $f_{ER_i}(y)$ is the distribution of such a performance conditional to an event from the *i-th* source.

$$f_{ER|L>l^{*}}(y) = \frac{\sum_{i} \nu_i \cdot P\left[L > l^{*} | y\right] \cdot f_{ER_i}(y)}{\lambda_{L>l^{*}}}$$

(B.4)

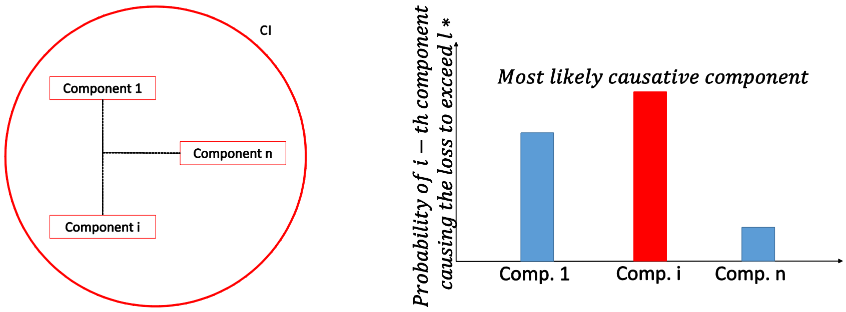

Fig. B.3 Loss disaggregation in terms of component performance.

Disaggregation of probabilistic loss of CIs, in fact, allows to get the most likely causes of any loss level (e.g., one deemed unacceptable) in terms of triggering event and/or CIs component performance.

B.1.2 Disaggregation in the case of epistemic uncertainty

In the case of loss assessment based on logic tree (e.g., Bommer et al., 2005), *k* models of the type in Eq. (B.1) are employed, and the rate of exceedance $\lambda_{L>l^*}$ of a loss threshold is provided by (B.2), where $\lambda_{L>l^*,j}$ is the rate of exceedance according to the *j*-th branch and P_i is its non-negative weight, such that $\sum_{j=1}^{k} P_j = 1$.

$$\lambda_{L>l^*} = \sum_j \lambda_{L>l,j} \cdot P_j \tag{B.5}$$

Consequently, it may be recognized that disaggregation (considering for simplicity one seismic source), for example in terms of magnitude and distance, can be written as in equation (B.6); i.e., Iervolino (2016). In the equation, the *j* subscript indicates the model of the *j*-th branch of the logic tree.

$$f_{M,R|L>l^*}(w,z) = \frac{\sum_j \nu_j \cdot P\left[L>l^* \middle| w,z\right]_j \cdot f_{M_j,R_j}(w,z) \cdot P_j}{\lambda_{L>l^*}} \tag{B.6}$$

Hazard or performance disaggregation, as well as disaggregation of the loss in terms of other random variables involved, can be obtained, when epistemic uncertainty is involved, in analogy with equation (B.6) using the appropriate terms defined in the previous sections.

B.1.3 Discussion

Hazard, performance, and loss disaggregation are recommended in STEP 7 of the methodology presented in the rest of the document, to identify critical events and components, especially when the CI does not pass the stress test. It is, however, informative in any case to identify the causes of undesired event effects.