MULTI-LEVEL RISK-BASED STRESS TEST METHODOLOGY FOR CRITICAL NON-NUCLEAR INFRASTRUCTURE SYSTEMS

Simona ESPOSITO¹, Božidar STOJADINOVIC², Anže BABIĆ³, Matjaž DOLŠEK⁴, Sarfraz IQBAL⁵, Jacopo SELVA⁶, Arnaud MIGNAN⁷, Domenico GIARDINI⁸

ABSTRACT

A more resilient society requires improved and standardized tools for integrated risk management applicable to different type of civil infrastructure systems. Among the most important tools are the stress tests. A need to develop appropriate stress tests for all classes of critical civil infrastructures arose following the post-Fukushima nuclear power plant stress tests. In this paper, a harmonized approach to stress test non-nuclear Critical Infrastructure (CI) systems against natural hazard events is presented and a risk-based multi-level stress test methodology are proposed. A stress test comprises four main phases. First the goals, the methods for the risk analysis, the time frame, and the total costs of the stress test are defined. The stress test is then performed at the component and the system levels, which is followed by verification and analysis of the stress test findings. A penalty system is defined to acknowledge the limitations of the methods used to stress test the CI and adjust the output of the risk assessment. The adjusted results are the input to a grading system, which is utilized to determine the grade and the outcome of the stress test. Finally, the results are reported and communicated to stakeholders and authorities.

Keywords: critical infrastructures, extreme events, stress test, multi-level, grading system

1. INTRODUCTION

Critical infrastructure (CI) systems enable a modern society: these systems provide the essential functions of public safety and support, through their services, the higher-level functions of a community, such as housing, education, healthcare and the economy. Natural hazard events can interrupt services, cause damage, or even destroy CI systems, triggering disruption of vital socio-economic activities, extensive property damage, and/or human injuries or loss of lives. Consequently, the interest in understanding and modelling the risk and the resilience of CIs is increasing. The European Program for Critical Infrastructure Protection (EPCIP) was established in

¹Research Assistant, Department of Civil, Environmental and Geomatic Engineering, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, simona.esposito@ibk.baug.ethz.ch
²Professor, Department of Civil, Environmental and Geomatic Engineering, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, stojadinovic@ibk.baug.ethz.ch
³Research Assistant, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia, Anze.Babic@fgg.uni-lj.si
⁴Professor, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia, Matjaz.Dolsek@fgg.uni-lj.si
⁵Researcher, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Bologna, Bologna, Italy sarfraz.iqbal@ingv.it
⁶Researcher, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Bologna, Bologna, Italy jacopo.selva@ingv.it
⁷Research Assistant, Institute of Geophysics, Department of Earth Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland arnaud.mignan@sed.ethz.ch
⁸Professor, Institute of Geophysics, Department of Earth Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland domenico.giardini@erdw.ethz.ch
2006 and was recently revised to ensure a high degree of protection of European infrastructures. In particular, the EPCIP clearly declared the need to develop a tool, a stress test for critical non-nuclear CIs, as a way to verify their safety and increase their resilience in the near future.

A harmonized and multi-level approach for stress testing critical non-nuclear infrastructure systems is proposed herein. The aims of this methodology are to quantify the safety and the risk of individual components as well as of whole CI systems with respect to natural hazard events, and to compare the behavior of the CI system to acceptable values. The proposed methodology is based on the best-possible characterization of the extreme (and other damaging) scenarios and consequences (Cornell and Krawinkler 2000), including potential multi-hazards (Selva 2013; Liu et al. 2015; Mignan et al. 2014, Mignan et al. 2017) and systemic amplification effects (Esposito et al. 2015; Argyroudis et al. 2015).

2. STRESS TEST METHODOLOGY

The proposed stress test methodology has multiple levels to accommodate great variability of CI systems with respect to types of natural hazards they are vulnerable to, in the potential consequences of their failure, and the available resources for conducting a stress test. Each Stress Test Level (ST-L) is characterized by a different scope (component or system) and by a different level of risk analysis complexity (starting from design code checks and ending with state-of-the-art single and/or multi-hazard probabilistic risk analyses). In this multi-level methodology, three ST-Ls are foreseen: ST-L1, a single-hazard component check; ST-L2, a single-hazard system-wide risk assessment; and ST-L3, a multi-hazard system-wide risk assessment. The selection of the appropriate ST-Ls may depend on the capabilities of the stakeholders, i.e. the available human and financial resources to perform the stress test, and on the regulatory requirements that are based on the different importance of the CI system (Esposito et al. 2016). This multi-level structure provides the necessary flexibility to apply the proposed stress test to a broad range of non-nuclear CIs.

2.1 Multiple-Expert Management Protocol

Engaging multiple experts is critical in a risk assessment when potential controversies exist and the regulatory concerns are relatively high. In order to produce robust and stable results, the integration of expert inputs plays a fundamental role in managing subjective decisions and in quantifying the epistemic uncertainty as “the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study” (SSHAC 1997). To this end, the diverse range of views and opinions of the experts, their active involvement, and their formal feedbacks need to be organized into a structured process ensuring transparency, accountability and independency.

A formalized multiple expert input integration protocol has been developed by Selva et al. (2015) and is integrated into the proposed stress test methodology (Figure 1). This process guarantees the robustness and the transparency of the stress test by guiding actors of the stress test in the assessment of CIs criticality, complexity and ability to conduct hazard and risk analyses, the assessment of subjectivity in decision making, and in the quantification of the epistemic uncertainty in the form of a “community distribution” that, in this context, means “the probability distribution representing the epistemic uncertainty within the community” (Bommer 2012).

The stress test actors are: Project Manager (PM), Technical Integrator (TI), Evaluation Team (ET), Pool of Experts (PoE), and Internal Reviewer(s) (IR). The interactions among these actors during a stress test are structured by the multiple expert input integration protocol to ensure transparency and accountability (Figure 1). Most importantly, the PM, TI and IR must remain independent to guarantee fairness of the stress test outcomes. Different roles and responsibilities are assigned to the different actors, as shortly described below.
The PM is a stakeholder responsible and accountable for the successful implementation of the stress test. The decisions of the PM must be rational and fair to the authorities and the public.

The TI is a lead analyst responsible and accountable for the scientific management of the stress test. The TI is responsible for capturing the views of the informed technical community in the form of a “community distribution”.

The ET is a group of analysts that perform the risk assessment of the CI, following the guidelines provided by the TI. The TI and PM consensually select the ET members from internal (to the CI system) and/or external experts. Thus, the ET represents an interface between the stress test actors and the CI stakeholders, guaranteeing a reciprocal acknowledgement of choices and outcomes.

The PoE has the task of providing blind quantitative input to the TI and PM for managing key critical choices/Issues. The PoE represents the scientific technical community in a stress test. Individual experts of the pool may also act as advocates of a particular hypotheses or technical position in individual communications with the TI. The PoE participates to the interviewing processes led by the TI, based on structured expert elicitation procedures (SSHAC 1997).

The IR is an 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 stress test (i.e. a participatory review). In particular, the IR reviews the following: the coherence between the TI choices and the PM requests and the conformity to the multiple-expert management protocol; 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 deliverables.

The CI authorities select the PM. The PM selects the TI and IR and, jointly with the TI, the members 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 complexity of the stress test and the resources available for it.

The PM interacts mainly with the TI and defines all questions that the stress test should answer, taking care of the technical and societal aspects (e.g., selection of the appropriate ST Level, definition of appropriate hazard and acceptable risk levels, etc.). The TI manages the scientific process and...
coordinates the ET in the implementation of the analyses. The TI also organizes the interaction with the PoE, based on structured group elicitations and individual interactions, and integrates PoE input and IR feedbacks into the analysis. The ET implements the analysis, following the TI choices. The IR reviews the entire stress test process starting from its earliest stages in order to guarantee its transparency, ensure independence of stress test actors, and maximize the reliability of the stress test results. More details on the interaction among the experts and the rationale behind the adopted process are presented in Selva et al. (2015).

2.2 Stress Test Workflow

In a stress test, the workflow is structured in four phases and a systematic sequence of steps (Figure 1). Each phase is subdivided into a number of steps, with a total of nine steps in a stress test.

2.2.1 Pre-Assessment Phase

STEP 1- Data collection: the data available on the CI and on the phenomena of interest are collected. In addition, data coming from stress tests performed on other similar CIs in different locations and other CIs at the same location are collected. In this step, the PM TI, ET and IR are selected, as described above. The TI and the ET collect data and other relevant information about hazards and the CI, and about previous and relevant hazard and risk studies.

STEP 2- Risk Measures and Objectives: the PM defines one or more risk measures (e.g. fatalities, economic losses) and objectives (e.g. expected loss, annual probability etc.) based on the regulatory requirements, the technical and societal considerations, and the outcomes of previous stress tests.

STEP 3- Setup of the Stress Test: the time frame, the total costs of the stress test, and the most appropriate ST-Level and the “level of detail” used for the modeling during the Assessment Phase are defined. Based on the regulatory requirements, the PM selects the ST-L to define the technical implementation of following stress test phases. In this context, the presentation of the outcome of the stress test (Report phase) is set; this date is intended to remain fixed during the stress test. STEP 3 may be a long process and may differ substantially depending if the PoE is in place or not, according to the selected ST-L. The presence of the PoE allows for a robust set-up of the stress test, based on the quantitative feedbacks of multiple experts, but the investment of resources into PoE may not be justified by the ST-L. In either case, the TI collects applicable scientific models and data needed for the stress test, with the technical assistance of the ET. Based on this collection, the TI and PM jointly identify the “level of detail” used in the modeling and analyses in the Assessment phase. Sensitivity analyses may be performed to better support decision-making.

2.2.2 Assessment Phase

STEP 4- Component Level Assessment: seismic performance of each component of the CI is checked by the ET using the hazard-based assessment, design-based assessment or risk-based assessment approach, as described below in Section 2.3.

STEP 5- System Level Assessment: the TI should implement the required models using the assistance of the ET. If the PoE is in place (according to the ST-L selected), the TI organizes the second round of PoE structured expert elicitations to fill the potential methodological gaps, to quantify the potential additional scenarios, and to rank/score the alternative models to enable the quantification of the “community distribution”. Open discussions among the PoE members (moderated by the TI) are foreseen only if significant disagreements emerge from the elicitation results. If the PoE is not in place but the treatment of epistemic uncertainty is required (see Section 2.1), the TI directly assigns scores/ranks to the selected models. Finally, the ET implements all the required models and performs the assessment. If specific technical problems emerge during the implementation and application, the TI may solve them through individual interactions with members of the PoE (if present) and the ET.
2.2.3 Decision Phase

STEP 6 - Risk Objectives Check: the results of risk assessment are compared with the risk objectives defined in the Pre-Assessment phase. Depending on the type of risk measures and objectives defined by the PM, as well as on the methods adopted, the way the results of the risk assessment are compared to the risk objectives may differ. One option, a grading system (Babič and Dolšek, 2016), is presented in Section 4.

STEP 7 - Disaggregation/Sensitivity Analysis: the critical events, i.e. events that most likely cause the exceedance of the considered risk measures and objectives are identified. Based on this, risk mitigation strategies and guidelines are formulated. These events may be identified through a disaggregation analysis (Iervolino, 2016, Esposito et al., 2015). For example, the loss may be disaggregated with respect to system response to identify the component whose damage most likely causes the exceedance of the loss value of interest. As in STEP 5, if specific, technical problems emerge during the risk assessment the TI may solve them through individual interactions with the PoE (if present) and ET. However, note that STEPS 7 is not mandatory in a stress test. Hazard, performance, and loss disaggregation is, nevertheless, recommended to identify the critical events, components and system features, especially if the CI does not pass the stress test.

STEP 8 - Guidelines and Critical events: risk mitigation strategies and guidelines are formulated based on the identified critical events, components and systems. The implementation and the results of the steps of the Assessment and Decision phases are documented by the TI. A preliminary stress test report is prepared. The IR reviews the documentations and provide suggestions. The TI updates the final stress test report and completes the supporting documents. Based on these documents, the PM, TI and IR make the final agreements on the stress test outcome.

2.2.4 Report Phase

The Report phase comprises one step (STEP 9 - Results Presentation) where the results are presented to CI authorities, regulators and the community. This presentation is organized and performed by PM and TI. The presentation includes the outcome of the stress test in terms of the stress test grade, the critical events components and systems, the guidelines for risk mitigation, and “level of detail” adopted in the stress test. Note that the date for this presentation is set in the Pre-Assessment phase, and it cannot be changed during the Assessment and Decision phases.

2.3 Stress Test Levels (ST-Leves)

Non-nuclear CIs, such as transportation, manufacturing, metal production and process, chemical, petro-chemical, power-generation and energy transfer and storage, water or food production CIs, are very diverse: the potential range of consequences of failure of these CIs and the types of hazards they are vulnerable to vary greatly, as do the capability and the available resources for conducting the stress tests. Therefore, it is not optimal to require the most general form of the stress test in all situations. To facilitate conducting stress tests across a broad range of non-nuclear CIS, three ST-Levels are proposed:

- Level 1 (ST-L1): single-hazard component check;
- Level 2 (ST-L2): single-hazard system-wide risk assessment; and

Each ST-L is characterized by a different scope (component or system) and by a different complexity of the risk analysis (e.g. the consideration of single- or multi-hazard and/or single-or multi-risk events) as shown in Figure 2. The aim of the Component Level Assessment (ST-L1) is to check each component of a CI system independently to show if the component passes or fails the minimum requirements for its performance, which are usually defined by the CI design codes, current at the time of the stress test. This level of assessment is obligatory because design of most CI components is
regulated by design codes, and the data and the expertise are readily available. On the other hand, the System Level Assessment (ST-L2 or ST-L3) is generally not compulsory since it requires extensive knowledge of both components and systems and usually necessitates significant resources (e.g., financial, expert staff). Thus, in cases where the System Level Assessment is not required by the regulator, ST-L1 may be the only type of assessment actually performed. However, even if the System Level Assessment is not specifically required, it is highly recommended, since it is the only way to reveal the systemic mechanisms that lead to potentially catastrophic consequences.

System Level Assessment may be performed considering a single hazard (ST-L2) or multiple hazards (ST-L3). Regardless of the level selected, the aim of the assessment is to evaluate performance of the CI using probabilistic methods (i.e., probabilistic risk analysis, PRA). The final result of a PRA is a risk curve or a family of risk curves, and the associated uncertainties (aleatory and epistemic). A risk curve generally represents the frequency of exceeding a given consequence (or loss) value. The specific quantitative PRA method to use depends upon the context in which the risk is placed (the hazard context), and upon the system under consideration. A list of possible methods that may be applied to assess the performance and the risk of the CI may be found in Salzano et al. (2016), Kakderi et al. (2015) and Crowley et al. (2015). It should be noted that there are no standardized approaches for multi-risk assessment.

Different PRAs are possible at levels ST-L2 and ST-L3, resulting in the definition of sub-levels ST-L2a, ST-L2b, ST-L2c, ST-L2d, ST-L3c and ST-L3d (Figure 2). The choice of a sub-level depends on the required degree of engagement in taking critical decisions and in the quantification of the epistemic uncertainties. In other words, different sub-levels correspond to a different number of experts involved in the assessment, whose goal is to assess the “community distribution”, that here means “the probability distribution representing the epistemic uncertainty within the community” (Bommer 2012). This assessment goal is achieved by selecting a number of appropriate alternative, yet scientifically acceptable, models and weighting them according to their subjective credibility. The selection of models may, among many different options, be based on the development of Alternative Trees (Selva et al. 2015), where the analysis is divided into a number of consecutive steps, and alternative models are defined at each step.

3. STRESS TEST PENALTY SYSTEM

There is a broad range of methods and models to assess the natural hazard risk CIs pose. These methods cover different levels of detail and complexity for each hazard, vulnerability, and risk evaluation. All models reflect a strategy of bounded rationality and therefore are necessarily a simplification of reality (Fischhoff, 2015). However, the level of simplification may vary greatly. In fact, different models and methods have to be assumed or introduced to describe how the hazard and the vulnerability interact in time and in space. Furthermore, each combination corresponds to a different level of detail of the analysis.
The “level of detail” used for the risk computation reflects the level of complexity of the methods adopted for the component and system-level risk assessment. It may be defined as a measure of “the trueness and precision, and the repeatability and reproducibility of the results of the risk assessment”. In the implementation of a stress test, the selection of the “level of detail,” (to perform the hazard and the risk analyses) is crucial because it defines a degree of reliability of the Assessment-phase results. In addition of being central to the Assessment phase, determining the appropriate “level of detail” is also a challenge, since it requires experts with deep knowledge of the spectrum of models and methods available in the scientific literature to address a particular hazard and/or risk. 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 involved in a stress test need to characterize the trueness and precision of the state-of-practice methods, while simultaneously promoting more advanced methods and discouraging less advanced methods. Thus, a penalty system is proposed as a part of the stress test methodology.

To facilitate the evaluation of the “level of detail”, a Target Level (TL) is associated with each ST-Level, according to the judgement about the complexity of the required hazard and risk analysis. Such TL 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. The models and methods to perform each step of the stress test are identified by the TI. This selection is based on scientific grounds, but also has practical consequences, such as the duration of and the resources for the stress test. Thus, the “level of detail” of the selected models and methods may be different than the TL. Therefore, the choice of the models should be taken (and documented) jointly by the TI and the PM. Based on this choice, the TI evaluates the Effective Level (EL) of detail of the risk analysis. This assessment is reviewed by the IR, and compared with the TL. The EL should be at least as high as the TL; otherwise, a penalty system must be applied to modulate the outcome of the stress test. Based on the IR review, PM and TI may change the hazard and risk analysis complexity to avoid potential penalties, while considering the resources and time needed to conduct a stress test.

The proposed penalty system aims to penalize simplistic hazard and risk assessment approaches that cannot guarantee a sufficiently accurate risk analysis. It is implemented by assigning a Penalty Factor (PF) in STEP 6 (Risk Objectives Check) based on the difference between EL and TL. The “level of detail” and a Penalty Factor system are proposed in the following. Then, a notion of Penalized Loss is defined and a probabilistic approach is formulated to compute it.

### 3.1 Levels of Detail

Level of detail of the analysis may be described using either qualitative or quantitative scale. Three categories are defined on the qualitative scale:

- **Advanced**: making use of detailed information and advanced methods and models in most of the steps of the analysis; Associated with TLs 2c, 2d, 3c and 3d.
- **High**: making use of less detailed information and less advanced methods and models in most of the steps of the analysis; Associated with TL 2b.
- **Moderate**: making use of non-detailed information and simplified methods in most of the steps of the analysis; Associated with TLs 1a and 2a.

In case a quantitative scale is adopted for TL of the hazard and risk analysis, a factor interval (e.g. \([TL_{lb}, TL_{ub}]\)) should set up by the experts and associated to each ST-Level. In this case, the resulting EL identified for the adopted stress test hazard and risk analysis methods and models should be at least equal to the lower bound of the TL interval, i.e. \(TL_{lb}\). The TL of component level assessment (ST-L1) is evaluated using a qualitative scale. The EL for the component hazard-based is assumed to be moderate and design-based assessments is assumed to be High. If a risk-based component assessment approach is required, the EL may vary from Moderate to Advanced.

For system level (ST-L2 or ST-L3) assessment, the evaluation of EL is a function of the “level of detail” selected for modeling each hazard and risk, the method(s) adopted for the epistemic uncertainty
distribution, where the standard deviation is a function of the uncertainty associated with overcomplexity. To penalize an overly simplistic analysis, a random variable that measures an additional measure (e.g., minor, medium extensive, etc.), and acceleration, etc.), where the output of the risk assessment at system level is expressed by the Normal distribution, where the standard deviation is a function of the “level of detail” adopted in each step and layer. If a quantitative scale is adopted (i.e. a EL quantitative factor is associated with the analysis in each step and layer), the EL may be computed as:

\[ EL = W_1 \sum_{j=1}^{n} w_{1,j}EL_{s,j} + W_2 \sum_{j=1}^{m} w_{2,j}EL_{z,j} + W_3 \sum_{j=1}^{p} w_{3,j}EL_{s,j} \]

where \( n, m, p \) are the number of layers in each step \( i \) (hazard, vulnerability, risk), \( W_i \) represent the weight of each step \( i \) of the risk analysis and \( w_{i,j} \) the weight of each layer \( j \) (for each step \( i \)), and \( EL_{i,j} \) is the Effective Level of the layer \( j \) in the specific step \( i \). The weights and Effective Level values are assigned by the TI and reviewed by the IR. If all layers (for each step) are considered equally important, then \( w_{1,i}=1/n \), \( w_{2,j}=1/m \), \( w_{3,j}=1/p \). 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), EL may be obtained as:

\[ EL = \sum_{s=1}^{S} W^{(H_s)}EL^{(H_s)} \]

where \( W^{(H_s)} \) represents the weight of hazard \( s \) assigned by experts. Thus, a multi-hazard EL corresponds to the weighted mean of the “level of detail” evaluated for each hazard separately. If all hazards are considered equally important, then \( W^{(H_s)}=1/S \). When epistemic uncertainty analysis is of concern, a hierarchical additional layer is added in the same fashion.

### 3.2 Penalty Factor System

A penalty factor \( PF \) is defined as the difference between the EL and the TL of the selected ST level. If a qualitative scale is considered, three cases are possible: i) TL=High, EL=Moderate, ii) TL=Advanced, EL=High, iii) TL=Advanced, EL=Moderate. The PF may be computed using reference values associated with these three cases: i) \( PF_{H,M}=0.2 \), ii) \( PF_{A,H}=0.2 \), iii) \( PF_{A,M}=0.4 \). The values in this example are indicative. In a stress test, the penalty factors need to be set by expert consensus. Consider that the output of the risk assessment at system level is expressed by the annual exceedance rate of loss metric (L), \( \lambda (l) \), after Cornell and Krawinkler (2000):

\[ \lambda (l) = \int_{d} \int_{epd} \sum_{im} G(l \mid d) \mid dG(d \mid epd) \mid dG(epd \mid im) \mid d\lambda (im) \]

where \( im \) is an intensity measure (e.g., peak ground acceleration, peak ground velocity, spectral acceleration, etc.), \( epd \) is an engineering demand parameter (e.g., interstory drift), \( d \) is a damage measure (e.g., minor, medium extensive, etc.), \( l \) is the loss variable (e.g., monetary losses, downtime, etc.), and \( G(y \mid x) \) are the conditional complementary cumulative distribution functions (CCDFs). As stated above, a probabilistic risk analysis can be performed adopting different levels of detail and complexity. To penalize an overly simplistic analysis, a random variable that measures an additional uncertainty associated with overly simplistic analysis methods or models, named Penalty Uncertainty \( \varepsilon_p \), is introduced. Then, a new loss metric is introduced, named Penalized Loss \( L_p \), and expressed on a logarithmic scale as:

\[ \log(L_p) = \log(L) + \varepsilon_p \]

Note that \( \varepsilon_p \) acts as a model error. In fact, it amplifies the uncertainties associated with simplistic analysis models and methods. A possible choice for the probability distribution of \( \varepsilon_p \) is the Normal distribution, where the standard deviation is a function of the Penalty Factor (PF) defined as:

\[ \sigma(l) = |PF \cdot \log(l)|, l > 0 \]
Observe that $PF$ acts as a coefficient of variation. The support of the loss metric $L$ is unbounded or bounded. The distribution of the Penalty Uncertainty $e_p$ must be truncated according to the support of $L$. Further, in order to focus on the tails of the loss curve, no model error is added to $PF$ when $l=0$. It is of interest to observe that $\sigma(l)$ is proportional to the loss; consequently, the tails of the computed loss curves are penalized both by the presence of an extra uncertainty due to overly simplistic analysis methods and models and by a higher $\sigma(l)$. Then, Penalized Loss $L_P$ is a new random variable, defined conditionally with respect to the loss value $l$ of the loss metric $L$ obtained from the probabilistic risk analysis.

4. STRESS TEST GRADING SYSTEM

The principal outcome of a CI stress test is a grade, which is obtained in the STEP 6 (Risk Objectives Check) of a stress test. The grade is based on the comparison of the results of the risk assessment to the risk objectives (i.e. acceptance criteria) defined at the beginning of the stress test in STEP 2 (Risk Measures and Objectives). The grade is obtained using a grading system developed by Babič and Dolšek (2016) and integrated into the proposed stress test methodology (Esposito et al., 2016). The grading system (Figure 3) has three outcomes: Pass, Partly Pass, and Fail. A stress-tested CI passes the stress test if it attains grades AA or A. Grade AA corresponds to negligible risk and is expected to be the risk objective for new CIs. Grade A corresponds to risk being as low as reasonably practicable (ALARP) (Helm, 1996), and is expected to be the risk objective for existing CIs. Grade B corresponds to the existence of possibly unjustifiable risk; in this case, the CI partly passes the stress test. Grade C corresponds to the existence of intolerable risk; in this case, the CI fails the stress test.

![Figure 3. Grading system for the outcome of a stress test.](image)

### 4.1 Stress Test Grade Boundaries

The stress test PM defines the boundaries between grades (i.e. the risk objectives) by following the requirements of the regulators and societally acceptable risk norms. The grade boundaries depend on the type the risk measure used to characterize risk. They can be expressed using point estimates (Figure 5, top row) or continuous functions (Figure 4, bottom row). Examples of point estimates include the annual probability of risk measure exceedance (e.g. loss of life) and the expected value of the risk measure (e.g. expected number of fatalities per year), whereas continuous function examples include a $\lambda$-$L$ curves, where $\lambda$ represents the cumulative frequency of a risk measure ($L$) over a given period of time. Regulatory boundaries may differ between countries and industries. Harmonizing the risk objectives (and risk measures) across a range of critical non-nuclear CIs at the European level remains a challenge. This is a task for both the regulatory bodies and the industry association, who should reconcile the societal and industry interest and develop mutually acceptable risk objectives and, possibly, stress test grade boundaries.
4.2 Grading System in Time Domain

The performance of a CI changes over time, and so does its risk profile. This is due to evolution of the understanding of hazards and risks through new findings, as well as changes of the CI due to use, ageing and long-term degradation processes, effects of previous hazard events, man-made events (e.g., terrorist attacks), and change in the CI-induced risk exposure (e.g., community population increase or decrease). Time variation of CI performances may lead to an increase of the probability of failure, loss of functionality, or exacerbate the consequences of failure during the CI lifetime (Figure 4). This means that a CI that passed a stress test at some point in time may not pass the stress test later on. More important, a CI that partially passed a stress test should be incentivized to reduce the risk it poses within a set time period. Crucially, a CI that failed a stress test must be compelled to design and implement retrofits to urgently reduce the risk it poses or manage these risks using other means, such as reduction or relocation of operations, or transfer of risk using financial measures (such as insurance).

For these reasons, a stress test is designed to be periodic. Instead of a single stress test, the regulator should prescribe a sequence of stress tests designed to ensure continuous reduction of risk posed by the CI. The period between two consecutive stress tests is determined based on equity of the cumulative risk exposure posed by different CIs (Babič and Dolšek, 2016). The proposed grading system is extended to facilitate such risk-driven period stress test organization.

![Figure 4](image-url)

Figure 4. Grading system using point estimates (top row) and continuous functions (bottom row) for risk objectives a) two different results (Result 1 and Result 2) from the first stress test, b) redefinition of the grading system parameters due to Result 1, and c) redefinition of the grading system parameters due to Result 2.

If the CI passes a stress test (obtains a grade AA or A), the risk objectives for the next stress test do not change. The time to the next stress test is set to the longest possible period between two stress tests established by the regulator. Such period may be as long as the expected lifetime of the CI system, or as short as dictated by the changes in the estimates of hazard or changed in the design codes. However, some CIs may obtain grade B (the risk they pose is possibly unjustifiable) or C (the risk they pose is intolerable). In such cases, the grading system stimulates the stakeholders to act to reduce the risk in the following manners: i) by strengthening the grade boundaries for the next stress test; and/or ii) by reducing the time between the successive stress tests; both on the basis of equity of risk above the ALARP region over two cycles (Babič and Dolšek 2016). In particular, if a CI obtains grade B, the boundary between grades B and C is shifted to the left, toward the boundary between grade A and B (Figure 4b). Furthermore, if a CI obtains grade C, the boundary between grades B and C is moved to the boundary between grade A and B, and the period until the next stress test is reduced (Figure 4c). This incentivizes the CI stakeholders to adequately mitigate the risks posed by the CI in as few stress test cycles as possible. It follows that the CI will be upgraded to pass the stress test, or that the regulator will require that the CI ceases operation.
6. CONCLUSIONS

A harmonized and multi-level methodology for stress testing critical non-nuclear infrastructure systems was proposed. The aims of this methodology are to quantify the safety and the risk of individual components as well as of whole CI system with respect to natural hazard events, and to compare the behavior of the CI system to acceptable values. The outcomes of the stress test are intended to support decision makers in the evaluation and management of the risks CI systems pose to communities they serve. However, the instantaneous loss of the service provided by CIs by itself does not reveal how a community served by the CIs responds to a disaster. The time dimension of the recovery process is key: the evolution of community needs and the ability of the CIs to fulfill these needs (e.g. water, gas, and electricity) is best represented and modelled using the concept of resilience rather than of risk.

Conducting a stress test to assess only the risk exposure of a civil infrastructure system, i.e. to relate the losses with the probability of their occurrence due to one or more hazards, does not provide enough information on the ability of the CI system to recovery and function after a disaster. However, the proposed stress test methodology was designed to also serve as a basis for development of a new stress test concept that may support decision makers in the evaluation of strategies to not only decrease the risk, but also to enhance the resilience of CIs against natural hazards. It is clear that a new resilience-oriented stress test methodology for CI systems must include the recovery process and, furthermore, include models of how the systems function and deliver their service to the community, and how the community recovers its needs for such services.

7. ACKNOWLEDGMENTS

The work presented in this paper was conducted within the project “STREST: Harmonized approach to stress tests for civil infrastructures against natural hazards” funded by the European Community’s Seventh Framework Programme under grant agreement no. 603389. The authors gratefully acknowledge this funding. The authors acknowledge the contributions of the Work Package leaders, Mr. Peter Zwicky, Prof. Fabrice Cotton, Prof. Iunio Iervolino, Prof. Kyriazis Pitilakis, Dr. Fabio Taucer and Dr. Sotirios Argyroudis. The methods, results, opinions, findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

8. REFERENCES


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


