16th World Conference on Earthquake, 16WCEE 2017

Santiago Chile, January 9th to 13th 2017

Paper N° 4247

Registration Code: S-F1463151090

ENGINEERING RISK-BASED METHODOLOGY FOR STRESS TESTING OF CRITICAL NON-NUCLEAR INFRASTRUCTURES (STREST PROJECT)

S. Esposito[(1)], B. Stojadinovic[(2)], A. Babič[(3)], M. Dolšek[(4)], S. Iqbal[(5)], J. Selva[(6)], D. Giardini[(7)]

[(1)] *Senior Research Assistant, Institut für Baustatik und Konstruktion, ETH Zürich, simona.esposito@ibk.baug.ethz.ch*
[(2)] *Professor, Institut für Baustatik und Konstruktion, ETH Zürich, stojadinovic@ibk.baug.ethz.ch*
[(3)] *PhD Student, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Anze.Babic@fgg.uni-lj.si*
[(4)] *Professor, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Matjaz.Dolsek@fgg.uni-lj.si*
[(5)] *Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, sarfraz.iqbal@ingv.it*
[(6)] *Researcher, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, jacopo.selva@ingv.it*
[(7)] *Professor, Institute of Geophysics, ETH Zürich, domenico.giardini@erdw.ethz.ch*

Abstract

An engineering risk-based multi-level stress test, named ST@STREST, is proposed aimed at enhancing procedures for evaluation of the risk exposure of critical non-nuclear infrastructures against natural hazards in the context of the European project STREST (2013-2016, within the European Union's Seventh Framework Programme) presented in a companion paper. 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 consists of four phases: Pre-Assessment, the Assessment, the Decision and the Report phase. The phases are performed in sequence. In the Pre-Assessment phase all the data available on the CI (risk context) and on the phenomena of interest (hazard context) are 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 to test the CI are defined. This process is crucial and requires interaction among different experts and a detailed technical evaluation of the available information. In the Assessment phase both component and system ST levels (selected in the Pre-assessment phase) are performed. At the component level, initial design demand levels for each component of the CI are compared with the best available information about their capacity at the time of the stress test. Then, the systemic probabilistic risk analysis of the entire CI (system level) is performed. This process requires interaction with experts, to review and finalize the result of the assessment. In the Decision phase, results of the Assessment phase are compared to the risk objectives defined in the Pre-Assessment phase. The outcome of the Decision phase is the global outcome of the stress test. In this paper, a system to grade the outcome of a stress test is proposed. It is foreseen that CI can pass, partly pass, or fail the stress test. The CI partly passes the stress test when the assessed risk is considered possibly unjustifiable with respect to the risk objective defined in the Pre-Assessment phase. In this case, the proposed grading system prescribes by how much the safety of the CI should be improved until the next periodical verification of the CI. Finally, in the Decision Phase, critical events that most likely cause the exceedance of a given level of loss value are identified through a disaggregation analysis. Risk mitigation strategies and guidelines are formulated based on the identified critical events. In the Report Phase, the experts present the stress test results to CI authorities and regulators. The presentation includes the outcome of stress test in terms of the grade, the critical trigger events, the guidelines for risk mitigation, and the level of "detail and sophistication" of the methods adopted in the stress test.

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

1. Introduction

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 or disruption of vital socio-economic activities, extensive property damage, or human injuries or loss of lives [1]. 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 dynamicity of infrastructure systems, and by a growing dependency of the society on the infrastructure services. Consequently, the interest of the international academic community in the challenges of understanding and modelling the risk and the resilience of critical infrastructures 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. In response to the EPCIP requirements, the research project FP7 STREST (2013-2016, http://www.strest-eu.org), has been funded by the European Commission with the aim to develop a methodology to carry out appropriate stress tests for different categories of non-nuclear CIs. In the context of this project, a harmonized hazard and vulnerability assessment approach for stress testing critical non-nuclear infrastructures has been developed and is presented in a companied paper [2].

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 will be applied and tested in six critical infrastructures in Europe, 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: 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, the main aspects of the proposed engineering risk-based methodology (ST@STREST) for stress tests of non-nuclear CIs are presented. First, the workflow and the interaction among the main actors of the process are discussed. Then the multi-level approach and the different levels of analysis are presented. Finally, a possible grading system for testing the CI is introduced. This system allows to grade the CI and prescribe how much the safety of the CI should be improved in the periodical verification of the CI.

2. Stress tests of critical non-nuclear infrastructures: ST@STREST

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. The aims of the proposed methodology are to assess the performance of individual components as well as of whole CI system 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 [3, 4]. 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). 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 [4]). 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. A criticality assessment of the CIs, aimed at identifying and ranking CIs (for example at a national scale), may represent a practical tool to support the choice of the appropriate ST level [5].

Further, 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 (Section 2.1) , following a pre-defined process for managing the multiple-expert integration [6]. This allows to define how reliable the results of the stress test are in terms of Accuracy[1] of the test [5].

The workflow of ST@STREST comprises four phases: Pre-Assessment phase; Assessment phase; Decision phase; and Report phase. 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, 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.

2.1 Multiple-expert Integration

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'* [7]. 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 granting transparency, accountability and independency.

For stress tests, a formalized multiple expert integration process has been developed [6], and integrated into the stress test Workflow (Section 2.2). 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 [7], it is located between SSHAC level 2 and 3, it makes an extensive use of classical Expert Elicitations, and it is extended to risk analyses.

Different experts are involved in the implementation of stress test process namely the Project Manager (PM), the Technical Integrator (TI), the Evaluation Team (ET), the Pool of Experts (PoE), and the Internal

[1] The Accuracy is related to the "Level of detail" used for the computation of risk assessment of the STREST methodology, i.e.the methods and models selected for assessing the performance of critical infrastructures against natural hazards.

Reviewers (IR). All these actors should interact along the stress test to assure transparency and accountability of the different actors, while PM, TI and IR should be independent to guarantee fairness of the results. Different roles and responsibilities are assigned to different actors, as described in the following.

The PM 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 TI 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 a community knowledge distribution, to be implemented in the risk calculations. The ET is a group of analysts that actually perform the risk assessment required, 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. The ET represents also the interface between the project and the CI authorities, guaranteeing the successful and reciprocal acknowledgement of choices and results. The PoE has the goal of representing the larger technical community within the process. 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 [7] documents, the PoE includes both resource and proponent experts). They participate to the interviewing processes lead by the TI, providing the TI for their opinions on critical choices/issues. The IR is 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 expertise coverage and scientific independence, the fairness of TI integration of PoE feedbacks, and the coherence between TI requests and ET implementations.

The participation of the different actors significantly changes along the different phases of the Stress Test (Fig. 1). 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 ET and supported by the PM, while the level of assistance depends on the ST level. The PoE (if present, see Section 3) 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.

2.2 ST@STREST Workflow

The workflow represents a systematic sequence of steps (processes) which have to be carried out in a stress test. The ST@STREST workflow proposed (reported in Fig. 1), comprises four phases and each phase is subdivided in a number of specific steps, with a total of 9 steps.

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.

PHASE 1: Pre-Assessment phase

The Pre-Assessment phase comprises three steps. First, the data available on the CI (risk context) and on the phenomena of interest (hazard context) is collected (*STEP 1- Data collection*). Also data coming from Stress Tests performed on other similar CI and other CIs in the same location is 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. In *STEP 2- Risk Measures and Objectives*, the goal of the project is defined. In particular, one or more risk measures (e.g. fatalities, economic losses) and objectives (e.g. expected loss, annual probability etc.) are defined by the PM, based on regulatory requirements, technical and societal considerations and previous Stress Tests. Then (*STEP 3- Set-up of the Stress Test*), the time frame, the total costs of the stress test and the most appropriate Stress Test Level and level of "detail and sophistication" to apply to test the CI (to follow) are defined. The selection of the

ST-L is made by the PM with the assistance of the TI, based on regulatory requirements. The conclusion of *STEP 3* may take time and may differ in case the PoE is in place or not (according to the ST- L selected).

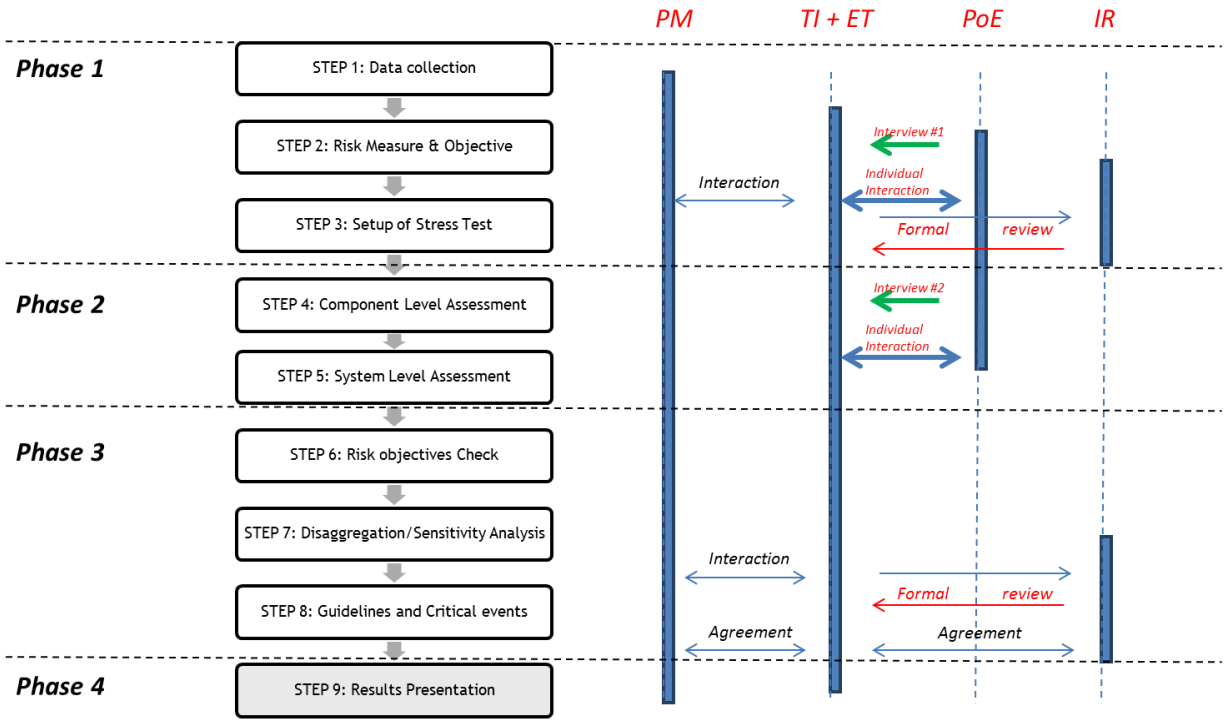

Fig. 1 – Workflow of ST@STREST methodology and interaction among the main actors during the multiple-expert process EU@STREST

PHASE 2: Assessment phase

The Assessment phase is characterized by two steps in which the stress test is performed at Component (*STEP 4 - Component Level Assessment)* and System (*STEP 5 - System Level Assessment*) levels according to the Stress Test Level selected in Phase 1. At Component level, 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). This check is performed by the TI or by one expert of the ET selected by the TI. At System level, first, the TI finalizes the required models. In particular, if the PoE is in place (sub-levels c), the TI organizes the classical Expert Elicitations i) to fill potential methodological gaps, ii) to quantify the potential scenario and iii) to rank/score the alternative models to enable the quantification of the epistemic uncertainty. The members of the PoE perform the elicitation separately. 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/ranks the selected models. Then, the TI actually implements all the required models and performs the assessment, with the technical assistance of the ET. 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, see Section 3).

PHASE 3: Decision phase

In the Decision Phase, the stress test outcomes are determined i.e. the results of risk assessment are compared with the risk objectives defined in Phase 1 (*STEP 6 - Risk Objectives Check*). 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 mean, 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. One possibility to assess the compliance with risk measures and objectives is presented in Section 4 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). 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 (*STEP 7 - Disaggregation/Sensitivity Analysis*). Finally, risk mitigation strategies and guidelines are formulated based on the identified critical events (*STEP 8 - Guidelines and 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 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.

PHASE 4: Report phase

The Report phase comprises one step (*STEP 9 - Results Presentation*) where the results are presented to CI authorities, community representatives and regulators. 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 the level of "detail and sophistication" of the methods adopted in the stress test. Note that the time for this presentation is set in Phase 1, and it cannot be changed during Phases 2 and 3.

3. 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 scope (component or system) and by a different complexity of the risk analysis (e.g. the consideration of multi hazard and multi risk events) as shown in Fig. 2.

The aim of the ST-L1 (Component Level Assessment) is to check each component of a critical infrastructure 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.). At component-level there are three methods to perform the single-hazard component check. These methods 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.

In the hazard-based assessment, 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, $I_{Assessment\ phase}$. In the design-based assessment 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. In the risk-based assessment, the hazard function at the location of the component and the fragility function of the component are required to evaluate the probability of meeting the risk acceptance criteria.

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. More details on the possible approaches for ST-L1 assessment can be found in [5].

Since a CI is a system of interacting components, ST-L1 is inherently not adequate. Nevertheless, ST-L1 is obligatory because design of (most) CI components is regulated by design codes, and the data and the expertize 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. The outcome of the ST-L1 is most often qualitative, e.g. component is compliant with the current regulation, component is not compliant with current regulation or the regulation does not yet exist for this type of component or type of hazard.

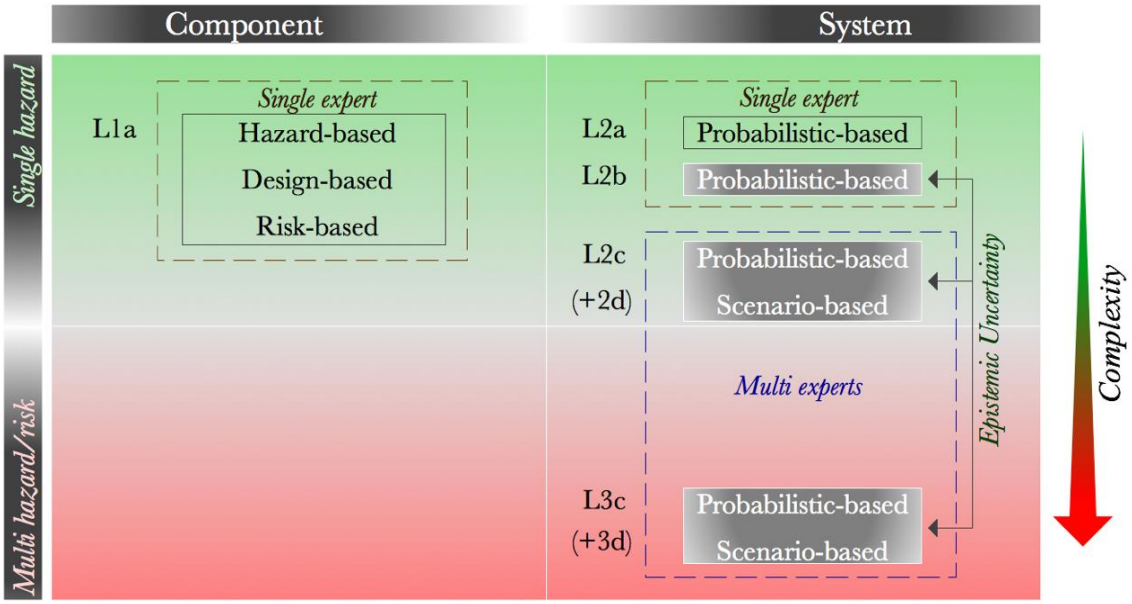

Fig. 2 – ST-Levels in the ST@STREST methodology

The stress test at the system level assessment of the critical infrastructure is 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 revealing the majority of the mechanisms leading to potential unwanted consequences. However, note that it requires more knowledge and resources (financial, staff) for conducting the stress test, thus it is not made obligatory (if not required by regulations). At these levels, potentially different implementations are possible. The quantification of epistemic uncertainty 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 knowledge distribution (describing the epistemic uncertainty) can be reached if more experts are involved in the analysis and, in particular, in dealing with all critical choices. A description of possible methods to threat epistemic uncertainty in each ST-Level and sub-level can be found in [5, 7].

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 system ST-L selected (sub-level d) . Levels 2d and 3d are complementary to L2c and L3c, respectively. In this case, multiple experts define and evaluate possible scenarios that, for whatever reason, cannot be included into probabilistic risk analysis. 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 [7]. These additional scenarios are meant to further investigate the epistemic uncertainty by including events otherwise neglected only for technical reasons. Indeed,

L2d and L3d are performed to evaluate the potential impact of epistemic gaps identified by experts, eventually increasing the capability of exploring the effective epistemic uncertainty. Thus, it is foreseen only as complementary to a full quantification of epistemic uncertainty 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 for the computation of risk. According to these two aspects a subdivision for ST levels has been introduced (Table 1, Fig. 2). The selection of the actual procedure to be implemented (row and column in Table 1) is performed in the Pre-Assessment (Phase 1). These two choices essentially depend on regulatory requirements, on the different importance of the CI, and on the available human/financial resources to perform the stress test. A criticality assessment of the CIs, aimed at identifying and ranking CIs, may represent a practical tool to support the choice of the appropriate ST level [5].

Table 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)

4. Proposed 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) is composed of three different outcomes: Pass, Partly Pass and Fail. The CI passes the stress test if it 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, [8]) 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.

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 can be expressed as scalar (Fig. 3 and 4, top) or continuous (Fig. 4, 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 [9-13]. 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.

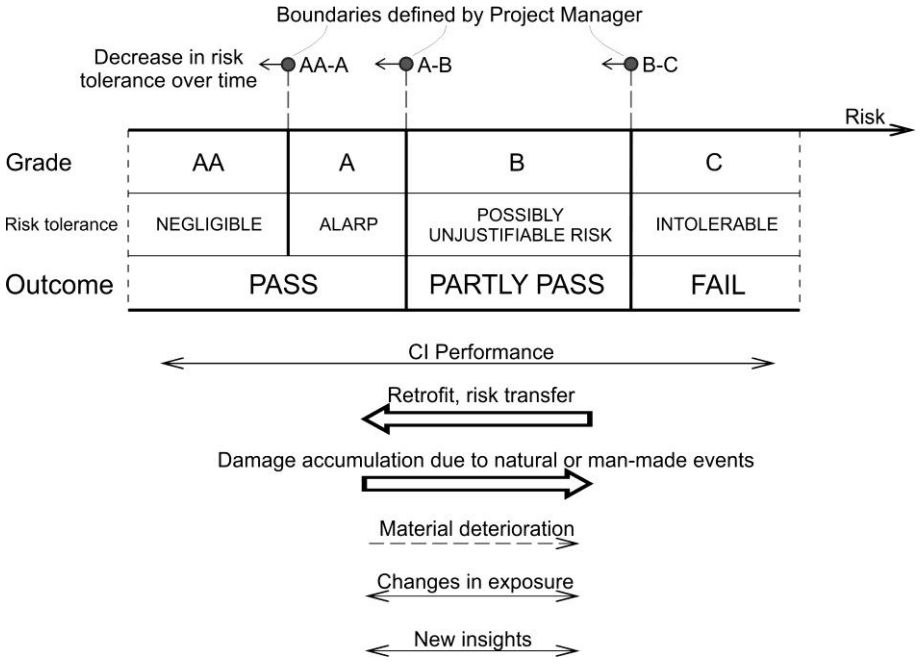

Fig. 3 – Grading system for the global outcome of stress test

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), 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 their lifetime. (Fig. 3). 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. 4, 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. It is recommended that the point associated with the greatest risk above the ALARP region be selected. 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. 4a).

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. 4a), the grading system foresees the reduction of the boundary between grades B and C (B-C boundary) in the next stress test (ST2, Fig. 4b). This reduction should be equal to the amount of risk beyond the ALARP region assessed in ST1. This ensures risk equity over two cycles. Furthermore, if grade C (red dot in Fig. 4a) is given in ST1, both the B-C boundary and the period until ST2 are reduced (Fig. 4c). 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:

$$t_{cycle,redefined} = t_{cycle,initial} \cdot \frac{R_1}{R_2} \quad (1)$$

where $t_{cycle,initial}$ is the initial amount of time between two evaluations of stress test, R_1 is the initial amount of risk between the B-C boundary and the A-B boundary (distance between the red and the blue line in Fig. 4a) and R_2 is the amount of the risk above the ALARP region assessed in ST1 (distance between the red dot and the blue line in Fig. 4a).

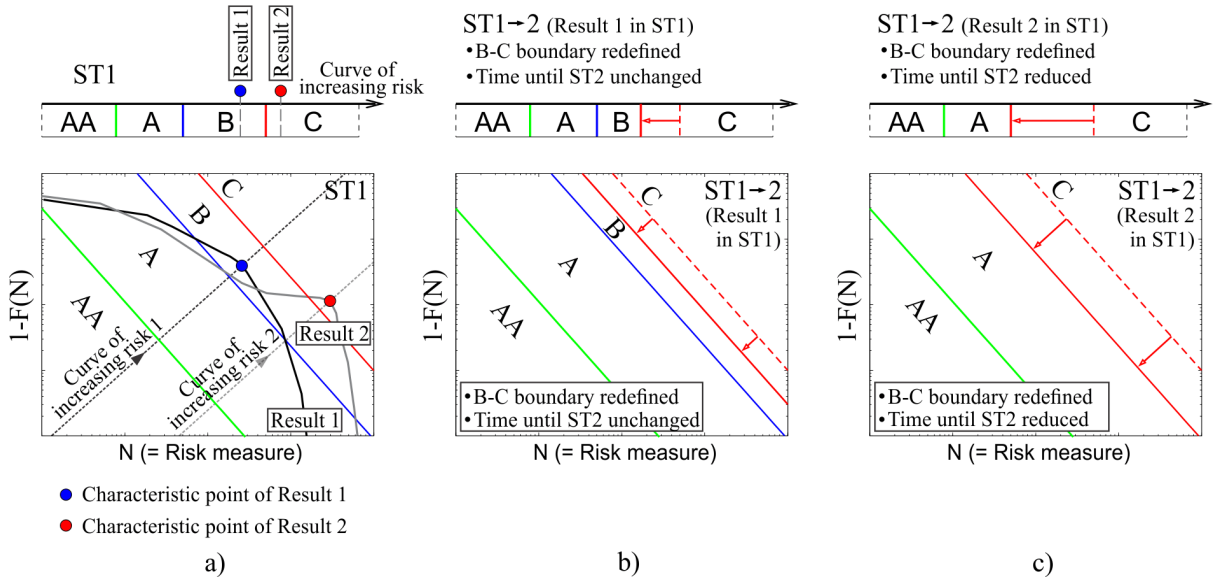

Fig. 4 – Grading system in time domain using scalar risk objectives (top) and limit F-N curves (bottom): a) two different results of the first evaluation of stress test (ST1), b) redefinition of the parameters of the grading system due to Result 1 in ST1, and c) redefinition of the parameters of the grading system due to Result 2 in ST1.

In the proposed model, the mean values of risk assessment results are used. It is a matter of future research to include the effect of epistemic uncertainties on the outcome of stress test. The proposed grading system requires boundaries 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.

5. Conclusions

The research project FP7 STREST (2013-2016, http://www.strest-eu.org), has been funded by the European Commission, with the aim to develop a methodology to carry out appropriate stress tests for different categories of non-nuclear CIs. So far, the only two domains where the stress test tool has been well developed are the nuclear and the banking sectors.

In the context of this project, an engineering risk-based methodology for stress test 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 to define 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. Finally, the results are reported and communicated to stakeholders and authorities.ST@STREST will be applied and tested in six critical infrastructures in Europe and it is intended to be improved based on operators 'feedback.

Acknowledgements

The work presented in this paper was conducted within the project "STREST: Harmonized approach to stress tests for civil infrastructures against natural hazards" funded by the European Community's Seventh Framework Programme under grant agreement no. 603389. The authors gratefully acknowledge this funding. The authors acknowledge the contributions of the project manager, Dr. Arnaud Mignan, and the Work Package leaders, Mr. Peter Zwicky, Prof. Fabrice Cotton, Prof. Iunio Iervolino, Prof. Kyriazis Pitalakis, 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.

6. References

[1] Grimaz S, Slejko D (2014): Seismic hazard for critical facilities. *Bollettino di Geofisica Teorica ed Applicata*, **55** (1), 3-16.

[2] Mignan A, Giardini D, Iervolino I, Stojadinović B, Pitilakis K (2017): Harmonized approach to stress tests for critical infrastructures against natural hazards (STREST) *16th World Conference on Earthquake Engineering*, January 9th to 13th 2017, Chile.

[3] Cornell C A, Krawinkler H (2000): Progress and challenges in seismic performance assessment. *PEER Center News*, **3** (2), 1-3.

[4] Mignan A, Wiemer S, Giardini D (2014): The quantification of low-probability–high-consequences events: part I. A generic multi-risk approach. *Natural Hazards*, **73**, 1999-2022.

[5] Esposito S, Stojadinovic B, Mignan A, Dolšek M, Babič A, Selva J, Iqbal S, Cotton F, Iervolino I (2016): Report on the proposed engineering risk assessment methodology for stress tests of non-nuclear CIs. *Deliverable 5.1 STREST Project*.

[6] Selva J, Iqbal S, Taroni M, Marzocchi W, Cotton F, Courage W, Abspoel-Bukman L, Miraglia S, Mignan A, Pitilakis K, Argyroudis S, Kakderi K, Pitilakis D, Tsinidis G, Smerzini C (2015): Report on the effects of epistemic uncertainties on the definition of LP-HC events, *Deliverable 3.1 STREST Project*.

[7] 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.

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

[9] Bowles D S, Anderson L R, Evelyn J B, Glover T F, van Dorpe D M (1999): Alamo dam demonstration risk assessment, ASDSO meeting, http://www.engineering.usu.edu/uwrl/www/faculty/DSB/alamo.html.

[10] Health & Safety Executive (HSE) (1989): *Risk criteria for land use planning in the vicinity of major industrial Hazards*. HSE Books ISBN 0-11-885491-7.

[11] MHLUPE (1988): Dutch National Environment Plan. Ministry of Housing Land Use Planning and Environment (MHLUPE), The Hague.

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

[13] Whitman RV (1984): Evaluating calculated risk in geotechnical engineering, *Journal of Geotechnical Engineer*, ASCE, **110** (2), 145-188.