16th World Conference on Earthquake, 16WCEE 2017
Santiago Chile, January 9th to 13th 2017
Paper N°3580
Registration Code: S-T1464682820

Individual and societal risk metrics as parts of a risk governance framework for induced seismicity

Marco Broccardo[(1)] , Laurentiu Danciu [(2)], Bozidar Stojadinovic[(4)], Stefan Wiemer[(5)]

[(1,3,4,5)]*Swiss Competence Centers for Energy Research, SCCER, ETH Zürich.*
[(2,5)] *Swiss Seismological Service, SED, ETH Zürich*
[(3)] *Institute of Geophysics, ETH Zürich*
[(1,4)] *Chair of Structural Dynamics & Earthquake Engineering, ETH Zürich*

Abstract
This study serves three purposes: (i) to review appropriate metrics for life safety (ii), to provide the outline of a robust mathematical framework quantifying and comparing risk safety metrics, and (iii) to provide upper bounds for aggregate measures of risk. In recent decades, the significant increase in seismicity caused by anthropogenic activities such as hydraulic fracturing, fluid injections, and mining, has posed the challenge of establishing a framework governing the risks. Risk metrics, which provide the baseline for risk management and decision-making, are a pivotal component of the risk governance framework. There is a broad spectrum of metrics that can be implemented; however, all were developed for rather different disciplines. In this study, we select the ones that are more closely related to the context of induced seismicity. Next, we introduce a general mathematical framework to define generalized risk metrics, which is based on the concept of norms in functional analysis. This framework allows a uniform comparison of different individual and societal risk metrics, and comparison across different activities and technologies. The second part of the paper focuses on the computational framework for the calculation of individual and societal risk. The framework is based on the well-established PSHA analysis and the PEER probabilistic performance-based seismic evaluation framework. The final part of the paper deals with the challenge of aggregate risk measures by providing upper bounds that can be compared with prescribed life safety criteria.

Key words: risk metrics, induced seismicity, risk assessment, aggregate risk

1. Introduction

The significant increase in seismicity, caused by anthropogenic activities such as hydraulic fracturing, fluid injections, and mining, has posed the challenge of establishing a risk governance framework to manage the risk related to such activities. One of the main differences between natural seismicity and induced seismicity is that while the first one is only related to natural processes, the second one is a combination of anthropogenic activities and natural processes. When considering hazardous anthropogenic activities, the first major distinction is between physical and non-physical risk. Examples of non-physical risk are vibrations felt, noise, public campaign against the project, NIMBI, etc. These risks are difficult, and in some cases impossible, to quantify. Within this setting, the view presented here maintains that an effective approach should gravitate towards risk mitigation rather than risk assessment. Nevertheless, the physical risk faced by the exposed communities must be quantitatively assessed. In a classical protocol for anthropogenic hazardous activities, the physical risk is divided into two major categories, i.e., fatalities and injuries, and economic losses. Thus, any risk analysis for induced seismicity should encompass both life safety and economical risk. Injury and fatality risk is generally assessed through the computation of the individual and societal risks, while economic risk is based on loss curves. In the induced seismicity case, injury and fatality risk is essentially related to building safety. However, building safety is regulated throughout building codes that do not explicitly account for individual and societal risks. For a given structure, building codes define acceptable levels of probability of structural failure [1], which is either expressed as annual probability of failure or probability of failure over the life time of the facility.

Therefore, there is a need to establish a framework to build a regulatory policy for induced seismicity risks. Within this task, risk metrics are the fundamental tools to measure risk, and to set safety standards for decision making. Since the risk metrics are necessarily an output of a risk analysis and are a required input for the decision making process, it is important that all the parties involved agree on their selection. Ideally, this processes should be undertaken prior risk analysis, since the outcomes are affected by the chosen metrics. For the sake of simplicity and synthesis, this paper focuses only on risk metrics for injuries and fatalities. However, the concepts for threating injury and fatality risk can be easily extended to economic risks.

The first part of the paper focuses on a critical review of a selected set of existing risk metrics for individual and societal risk. The review is mostly based on the work of Jonkman et al. [2]. Next, we introduce a mathematical framework based on concepts of functional analysis, which allows to define generalized risk metrics. Metrics defined within this framework are unit consistent (i.e. all expressed with the same chosen units), and convex. The convexity property is of particular importance for determining upper bounds for aggregate risk, and for risk optimization problems since they guarantee a unique optimum point.

The second part of this paper outlines a computational framework for the calculation of individual and societal risk. The suggested framework is based on the combination of classical probabilistic seismic hazard analysis (PSHA) analysis [3] and PEER (Pacific Earthquake Engineering Research) Center probabilistic performance-based seismic evaluation framework [4]. However, the focus of this study gravitates more towards the second component. It is shown that individual risk can be computed directly within this framework, although important challenges are still an open research topic. In particular, the hazard component of induced seismicity is inherently a time-non-stationary problem, while the natural seismicity is usually regarded as a time stationary problem. Moreover, the vulnerability part still represents an open challenge since the state of art literature focuses on macro-seismic (e.g. [5]), not micro-seismic, reliability analysis.

While the individual risk is a point-wise risk measure, societal risk is inherently a spatial aggregate risk measure. Then, the last part of the paper deals with the calculation of aggregate risk measures. When aggregate injuries and fatalities (or economical losses) are of interest, we essentially consider the problem of summing spatially correlated random variables. In this last part, we outline two upper bounds for aggregate fatalities or losses derived directly from the local marginal distributions of fatalities or losses. It is shown that any convex risk metric applied on the aggregate upper bound represents also an upper bound to the true risk. These risk upper bounds can be checked versus predefined risk safety thresholds.

2. Metrics for injury and fatality risk

In this section, we review the most common metrics for injury and fatality risks. Risk metrics are the principal tool for quantitative safety risk assessments, decision making, risk communication, and regulatory frameworks. In particular, in the latter, the risk metrics play a key role in setting acceptable risk levels and acceptance criteria in the standards. Generally, a risk metric is a mathematical mapping of the consequence of an event and its probability of occurrence,

$$\mathcal{R} = \varphi(consequence, probability), \tag{1}$$

where $\mathcal{R}$ is the risk measure, and $\varphi(\cdot)$ represents a defined mapping. In this study we focus on quantitative risk measures for injuries and fatalities and distinguish between individual risk and societal risk.

2.1 Individual risk metrics

Individual risk is the annual frequency at which a statistically average individual is expected to experience death or a given level of injury from the realization of a given hazard [7]. In this case, the consequence is a categorical variable, here named IL, expressing different injury severity levels. Normally, the probability distribution of IL is a multinomial with parameters $\pi_{il} = P(IL = il)$, which represent the probability of observing a specific injury level.

For a prescribed injury level il and a given hazard intensity measure, here denoted as im, the individual risk is ideally expressed as

$$\mathcal{R}_{IR} = \int_{im} P(IL = il|im)\,|dG_{IM}| = \pi_{il} = P(IL = il), \tag{2}$$

where $P(IL = il|h)$ is the conditional probability of observing the injury level il given a hazard intensity im, $G_H(h) = 1 - F_H(h)$ is the complementary cumulative distribution function (CCDF) of the hazard intensity, and $F_H(h)$ the cumulative distribution function (CDF) of the hazard intensity.

The probability of a given injury level is often conditioned to the damage state of a structure, D, rather than to the hazard level im. In this case, the individual risk can be written as

$$\mathcal{R}_{IR} = \sum_d P(IL = il|d)P(D = d) = \pi_{il} = P(IL = il). \tag{3}$$

In (3), we assume an implicit dependence of the damage states from the hazard.

Equations (2) and (3) do not account for whether the individual is actually physically in the proximity of the hazardous area. In [6], a slightly different version of (2) and (3) is proposed to account for the portion of the time that a person is present in the hazardous area. This is simply derived by multiplying (2) or (3) by the probability of a statistically average person is in the hazardous area in the course of one year. The individual risk is usually represented by iso-risk contour plots to facilitate land-use planning applications (Fig. 1.I). There are several others individual risk metrics, which differ depending on the application field. Here, we selected the one that is often used in land-use planning, and adopt it as appropriate for decisions concerning projects associated with induced seismicity. For a comprehensive review, the reader should consult [8].

The life-safety based risk regulations set individual risk acceptable criteria, which vary given the nature of the hazard and the exposed individuals. Individual risk limits are enforced to ensure that persons living in proximity or working either at or close to a hazardous activity are not exposed to an unacceptable risk. The individual risk criteria can set an absolute limit (by considering a statistically average person), or specify separate thresholds for the public and the most exposed personnel working at the activity. An example of absolute criteria is the limit $\mathcal{R}_{\mathrm{IR}} \leq 10^{-6}$(per year) set by the Dutch Ministry of Housing, Spatial planning and Environment (VROM) for new installations in populated areas [2]. An example for the other criteria is given by the safety requirements for liquefied natural gas industry which set $\mathcal{R}_{\mathrm{IR}} \leq 10^{-4}$ for the employees and $\mathcal{R}_{\mathrm{IR}} \leq 10^{-5}$ for the population. Within the field of technologies related to induced seismicity, no regulatory standards exist yet. To the best of our knowledge, the only recommended limit is $\mathcal{R}_{\mathrm{IR}} \leq 10^{-5}$, as indicated by the Commissie Meijdam in the Hazard and Risk Assessment for Induced Seismicity in Groningen, 2015 [9].

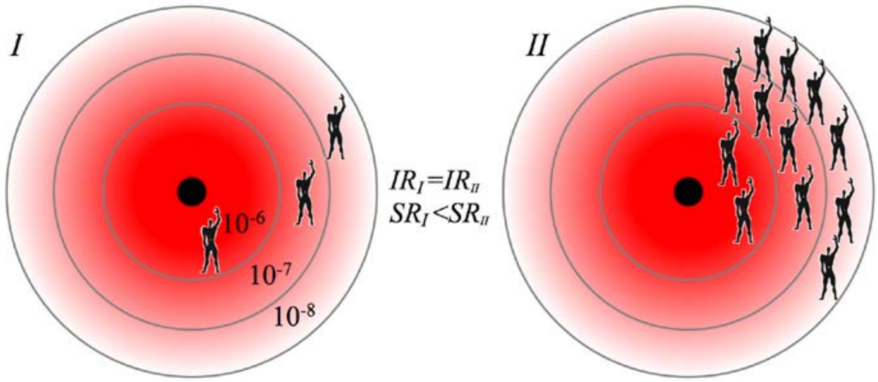

Figure1 I) Individual risk, iso-risk contour plots II) Societal risk, (Adaptation from [2])

2.2 Societal risk metrics

Societal risk is defined as the relationship between the frequency and the number of people suffering a given level of injury (or dying) from the realization of specified hazards [9]. Whereas the individual risk mirrors the severity of the hazards, it does not account for the number of individuals exposed to the hazard (Fig. 1). The societal risk aims to account for this effect by considering all exposed individuals present in the proximity of the hazardous activity. It is clear that while the individual risk is a point-wise measure, the societal risk involves an area/density integration. This allows a measure that better accounts for the risks of calamitous accidents, which may impact a large number of persons at once. Hereafter for simplicity, we refer only to fatality societal risk, as usually done in practice. It is clear that the same metrics can be applied to different level of injuries.

Some commonly used risk societal risk metrics can be derived directly from the individual risk maps. A first example of these metrics is the aggregate weighted risk (AWR), [10], which is defined as

$$\mathcal{R}_{AWR} = \int\int_A \mathcal{R}_{IR}(x,y)\rho_h(x,y)dxdy, \tag{4}$$

where $\mathcal{R}_{IR}(x,y)$ is the individual risk per year at a given location (x,y), $\rho_h(x,y)$ is the number of houses for that given location, and A is the total area of interest. In common practice, A is divided into a finite number of zones, usually defined by the exposure asset, and Eq. (4) is computed as sum $\mathcal{R}_{AWR} = \sum_j \mathcal{R}_{IR}(j)\rho_h(j)$, where $\rho_h(j)$ is the number of houses for the j^{th} zone.

Laheij et al. [11] define the societal risk as the expected number of fatalities per year

$$\mathcal{R}_{SR} = E[N] = \int\int_A \mathcal{R}_{IR}(x,y)\rho_p(x,y)dxdy, \tag{5}$$

where N is a random variable representing the number of fatalities per year, $E[\cdot]$ is the expectation operator, and $\rho_p(x,y)$ is the population density at a given location (x,y).

Equations (4) and (5), are based on individual risk maps, which are computed point-wise. However, most of the societal risk metrics used in practice are defined based on the CCDF of the number of fatalities per year. The relationship between the iso-risk contours and the CCDF of the number of fatalities is rather difficult to obtain because it depends on the probabilistic spatial correlation of the problem. Generally, the CCDF of the number of fatalities, here denoted with G_N, is computed via numerical methods and represented graphically using the so-called FN-curves. FN-curves are exactly the CCDF of the number of fatalities per year, i.e.

$$G_N(x) = 1 - P(N > x) = 1 - F_N(x) = \int_x^\infty dF_N(x), \tag{6}$$

where $F_N(x)$ is the cumulative distribution function (CDF) of the number of fatalities per year. The complementary cumulative form is preferred to ensure monotonically decreasing curves. The simplest and widely used societal risk metric is the expected number of fatalities per year, also known as potential loss of life (PLL),

$$\mathcal{R}_{SR_{PLL}} = E[N] = \int_0^\infty x dF_N(x) = -\int_0^\infty x dG_N(x) = -[xG_N(x)]_0^\infty + \int_0^\infty G_N(x)dn = \int_0^\infty G_N(x)dn, \tag{7}$$

where it is assumed that $\lim_{x\to\infty} xG_N(x) = 0$, a condition that is always satisfied since the total number of persons present in the area, n_{tot}, is clearly an upper bound of $G_N(x)$. By systematically and partially integrating (7), it is easy to show that

$$E[N^m] = \int_0^\infty x^m dF_N(x) = \int_0^\infty x^m |dG_N(x)| = m\int_0^\infty x^{m-1} G_N(x)dx. \tag{8}$$

Noteworthy: the area under the FN curve is the expected number of fatalities, and the m^{th} moment of N is a weighted $(m-1)^{th}$ moment of the FN curve.

In [12], the UK's health and safety executive (HSE) the societal risk is defined as

$$\mathcal{R}_{SR_{HSE}} = \int_0^\infty xG_N(x)dn. \tag{9}$$

Given (8) it is easy to see that $\mathcal{R}_{SR_{HSE}} = 1/2E[N^2]$. Moreover, it is well known that $\mathrm{Var}[N] = E[N^2] - E^2[N]$, where $\mathrm{Var}[\cdot]$ is the variance operator, so that (8) can be written as

$$\mathcal{R}_{SR_{HSE}} = \frac{1}{2}(\mathrm{Var}[N] + E[N^2]). \tag{10}$$

A similar risk metric, named total risk measure, is proposed by Vrijling at al. [13] as

$$\mathcal{R}_{SR_{TR}} = E[N] + k\sigma(N), \tag{11}$$

where $\sigma(\cdot) = \sqrt{\mathrm{Var}[\cdot]}$ is the standard deviation operator, and k is a risk aversion factor.

Another version for the societal risk is the so named COMAH risk metric [14] which is defined as

$$\mathcal{R}_{SR_{COMACH}} = \int_0^\infty x^\alpha dF_N(x), \tag{12}$$

where $\alpha \geq 1$ is a tail-weight factor which lends weight to $G_N(x)$'s right tail. Essentially, α represents the aversion to large accidents, which might involve a large number of persons. The HSE suggests a practical value of $\alpha = 1.4$. Observe that for $\alpha \in \mathbb{N}^+$, (12) is equivalent to (8).

Bohnenblust [6] proposes the following expression as a societal risk metric

$$\mathcal{R}_{SR_{PC}} = \int_0^\infty xU(x)dF_N(x), \tag{13}$$

where $U(x)$ is a monotonic increasing function named the risk aversion function. This risk metric is also known as a perceived collecting risk. A similar metric, which encompasses both a tail-weight factor, α, and a risk aversion function, $\phi(x)$, is proposed by Kroon and Hoej [15]

$$\mathcal{R}_{SR_U} = \int_0^\infty x^\alpha U(x)dF_N(x). \tag{14}$$

The authors named (14) as expected disutility of the system; (14) can be read as generalized metric and different authors propose different values for α (commonly $\alpha \in [1,2]$), and different expressions for the function $U(x)$.

Although all of these risk metrics are attempts to quantify the societal risks, they cannot be compared because they are defined in different units. For example, (7) and (11) are expressed in the same units, which are the number of fatalities; however, all the other metrics have their own units, which are difficult to pinpoint. This is not only an obstacle for a direct comparison of the metrics, but also for a cross-comparison of the risk between different hazardous activities. Moreover, there is a lack of a robust mathematical framework, which could be useful to address critical issues such as the convexity of the metrics. For example, the convexity of (12), (13), and (14) depends both from the choice of α and $U(x)$.

In the next subsection, we will tackle these issues by proposing a consistent risk metric framework, which is based on simple function space concepts. The purpose is not to define a new risk metric, but rather a consistent mathematical framework, which allows the parties involved in the decision-making process to choose a suitable metric.

2.2 Generalized risk metric for societal risk

A framework for induced seismicity risk metrics has the following desirable features:

- Dimension consistency, with units that are easy to identify, e.g. number of fatalities or losses.
- Simplicity: to be easily adopted and used by different communities.
- Flexibility: to include a tail-weight factor and/or a risk aversion function.
- Mathematical robustness, i.e., based on functional spaces concepts.
- Convexity, which offers a great advantage in defining upper bounds and in risk optimization problems.

To fulfill this list, we start by defining the functional space in which the mapping expressed in Eq. (1) is operating. For this purpose, consider the functional space $\mathcal{L}^p_{G_N}(\mathbb{R}^+)$ of p-integrable functions with respect to $f_N(x) = |dG_N(x)/dx|$, i.e.

$$\mathcal{L}^p_{G_N}(\mathbb{R}^+) = \left[\phi(x): x \in \mathbb{R}^+ \to \mathbb{R}^+, \int_x |\phi(x)|^p \, |dG_N(x)| < \infty\right], \text{for } p \geq 1 \tag{15}$$

where $\phi(x)$ is a generic function, and $f_N(x)$ is the probability density function (PDF) of the number of fatalities.. It follows that the only requirement for $\phi(x)$ is to be finite for $x \in [0, n_{tot}]$. Given this, we define the p-norm of a chosen function $\phi(x)$ in the functional space $\mathcal{L}^p_{G_N}(\mathbb{R}^+)$, as a societal risk metric, i.e.

$$\mathcal{R}_{SR_p}[\phi(x)] = \|\phi(x)\|_{p,G_n} = \left[\int_x |\phi(x)|^p \, |dG_N(x)|\right]^{\frac{1}{p}}, \text{ for } x \in \mathbb{R}^+, \text{and } p \geq 1. \tag{16}$$

The mathematical properties of this metric are equivalent to the properties of any p-norm defined in a $\mathcal{L}^p_G$ space, which are:

- Scalability: $\mathcal{R}_{SR_p}[a\phi(x)] = |a|\mathcal{R}_{SR_p}[\phi(x)]$.
- Subadditivity (convexity): $\mathcal{R}_{SR_p}[\phi(x) + \gamma(x)] \leq \mathcal{R}_{SR_p}[\phi(x)] + \mathcal{R}_{SR_p}[\gamma(x)]$.
- Separate points: $\mathcal{R}_{SR_p}[f(x)] = 0 \Leftrightarrow f(x) = 0$.

Moreover, the rank of the norm p is equivalent to the tail factor α, and $\phi(x)$ can arbitrarily include a risk aversion function. An important observation is that (16) is generally unit consistent. It follows that a comparison between different metrics and different hazardous activities is consistent. We can easily rewrite all the previous norms within this framework. For example, (7) is simply

$$\mathcal{R}_{SR_{PLL}} = E[N] = \|x\|_{1,G_n} = \mathcal{R}_{SR_{p=1}}[x], \tag{17}$$

and (9)

$$\mathcal{R}_{SR_{HSE}} = 2\left(\mathcal{R}_{SR_{p=2}}[x]\right)^2 = 2\left(\|x\|_{2,G_n}\right)^2 \quad . \tag{18}$$

As we already mentioned, (17) and (18) are not unit comparable; however, within the $\mathcal{L}^p_{G_N}$ space we can redefine each metric in consistent units (e.g. number of fatalities). For example, a unit-consistent version of (18) is simply

$$\mathcal{R}^c_{SR_{HSE}} = \mathcal{R}_{SR_{p=2}}\left[\sqrt{2}x\right] = \sqrt{2}\|x\|_{2,G_n}, \tag{19}$$

where the suffix $\cdot^c$ is used to indicate a unit consistent metric. Following the same logic, we can define the unit-consistent version of $\mathcal{R}_{SR_{COMACH}}$ as

$$\mathcal{R}^c_{SR_{COMACH}} = \mathcal{R}_{SR_p}[x] = \|x\|_{p,G_n}, \tag{20}$$

where $p \geq 1$, is equivalent to the tail factor α. The unit consistent version of the perceiving risk $\mathcal{R}_{SR_{PC}}$ is

$$\mathcal{R}^c_{SR_{PC}} = \mathcal{R}_{SR_{p=1}}[xU(x)] = \|xU(x)\|_{1,G_n}, \tag{21}$$

and the unit consistent version of $\mathcal{R}_{SR_U}$ is

$$\mathcal{R}^{c}_{SR_U} = \mathcal{R}_{SR_p}\left[xU^{\frac{1}{p}}(x)\right] = \left\|xU^{\frac{1}{p}}(x)\right\|_{p,G_n}. \tag{22}$$

Finally, it is easy to see that $\mathcal{R}_{SR_{TR}}$ is

$$\mathcal{R}^{c}_{SR_{TR}} = \mathcal{R}_{SR_{TR}} = \|x\|_{1,G_n} + k\left\|(x - \|x\|_{1,G_n})\right\|_{2,G_n}. \tag{23}$$

All metrics are legitimate choices for quantifying societal risk; however, we suggest adopting a consistent metric, i.e., one of the (17)-(23) equations. Moreover, the introduced framework allows a wide range of flexibility for choosing both a tail factor (here equivalent to p) and a risk aversion function $U(x)$.

3. Individual and societal risk computation for induced seismicity

In the previous sections we reviewed a list of common risk metrics to address individual and societal risk. In this section, we focus on the general framework for their computation. In the case of induced seismicity, individual and societal risks are direct consequences of damages and/or failures of building structures due to seismic activities. In fact, risk standards for structural safety are mostly regulated through building codes. Most of them do not specifically address the risk for individuals, but rather prescribe an acceptable level of probability of structural failure. This can be either an annual probability of failure, or the probability of failure over the lifetime of the structure. Consequently, the first step for individual and societal risk computation is the quantification of the likelihood of potential structural damages, followed by the transformation of this likelihood into the likelihood of injuries or fatalities. Therefore, it is natural to use the well established procedures of earthquake engineering to estimate both the likelihood of the hazard and the building structural failure probabilities. The classical probabilistic seismic hazard analysis (PSHA) [3] is the primary tool to estimate the likelihood of the hazard component. However, because of the inherent non-stationary nature of induced seismicity compared to the assumed stationary nature of natural seismicity, some adaptations are required. Time-dependent PSHA models have been studied [16,17,18]; however, they need to be tailored for each specific case. This is due to the significant, case-specific differences across earthquake occurrence models [19]. A thorough discussion of the probabilistic hazard computation for induced seismicity is beyond the scope of this paper, for which we suggest the following readings [18, 19].

For a given structure, the well-established procedure to compute mean annual rates of a given decision variable is the PEER probabilistic performance-based seismic evaluation framework [18]

$$\lambda(dv) = \int_{dg}\int_{edp}\int_{im} G(dv|dg)|dG(dv|dg)|\,|dG(dg|edp)||dG(edp|im)||d\lambda(im)|, \tag{23}$$

where im is an hazard intensity measure (e.g., peak ground acceleration, peak ground velocity, spectral acceleration, etc.), edp is an engineering demand parameter (e.g., interstory drift, maximum displacement etc.), dg is a damage grade measure (e.g., minor, medium extensive, collapse etc.), dv is a decision variable (e.g., number fatalities, monetary losses, etc.), $\lambda(x)$ is a mean annual rate of events exceeding a given threshold, and $G(y|x) = P(Y \geq y|X = x)$ is a conditional CCDF. Often in common practice, dg is a discrete variable that is related deterministically to edp. For example, dg_I (damage grade I, equivalent to light damage) is defined by $EDP \geq u_y$, where u_y is a given threshold (e.g. the elastic displacement limit of the structural behavior of the structure), and dg_{IV} (damage grade IV, equivalent to collapse) is defined by $dg_{IV} \geq u_u$, where u_u is the ultimate displacement threshold just before structural collapse occurs. Given this, (24) is usually used to compute mean annual rate of damage grade as

$$\lambda(dg) = \int_{im} G(dg|im)\,|d\lambda(im)|, \tag{24}$$

where $G(dg|im) = P(DG \geq dg|im) = P(EDP \geq u_u|im)$ is the fragility function, which represents the probability of exceeding a damage grade dg, given a level of hazard im. In the last decades, extensive efforts have been conducted to compute fragility functions for natural seismicity and for a wide range of building structures and civil infrastructure system components. However, induced seismicity presents new challenges related to the combination of lower intensity vibrations and higher earthquakes rates. At the present time, fragility functions for induced seismicity are still an open research topic.

The individual risk, $\mathcal{R}_{\mathrm{IR}}$ can be computed from (24) by following standard earthquake-injury classification (e.g. HAZUS 2003). In earthquake-injury classifications, each injury level occurrence probability is defined conditioned to a certain damage grade [20], i.e $P(il|dg)$. Given this, the mean annual rate of injury level il is defined as

$$\lambda(il) = \sum_{dg} \int_{im} P(il|dg)|dG(dg|im)|\,|d\lambda(im)|. \tag{25}$$

This rate of injury level can be scaled by the amount of time in a year an average person is present in the structure. Then, the rate can be transformed into annual probabilities by assuming a time recurrence model that follows a Poisson process with rate $\lambda(il)$, i.e.

$$\mathcal{R}_{IR} = P(IL = il|t = [1\text{ year}]) = 1 - \exp\big(-\lambda(il)\big). \tag{26}$$

When the intensity measure rate is time dependent, i.e, $\lambda(im, t)$, the same framework applies to determine $\lambda(il, t)$. Then, by assuming a time recurrence model that follows an inhomogeneous Poisson process,

$$\mathcal{R}_{IR}(T) = P(IL = il|t = T) = 1 - \exp\left(-\int_0^T \lambda(il, t)dt\right), \tag{27}$$

where T is the observational time period. Observe that both (25) and (26) imply that all the uncertainties are ergodic, i.e. renewable. While this is generally a realistic assumption for natural seismicity, it is an open question for induced seismicity. It is shown [21] that the ergodic approximation inherent in (25) is an upper bound that converges to the true value for lower probabilities. For a thorough discussion of this topic, the reader should consult [21, 22].

As stated in Section 2, the individual risk $\mathcal{R}_{IR}$ is locally defined for a single individual and does not give information regarding the number of persons in proximity of the hazard. A first estimate of the societal risk is calculated by multiplying the individual risk by the average number of building occupants, and by integrating this procedure for the buildings present in the hazardous area (essentially Eq. (4)). However, this procedure neglects both the hazard spatial correlation and the injuries correlation. The hazard spatial correlation refers to the observation that ground motions intensities appear in geographical clusters, i.e. two neighboring locations most likely experience similar intensities measures. The injuries/fatalities correlation refers to the observation that injuries severity appears in clusters, e.g. collapses of high occupancy buildings lead to a highly concentrated number of fatalities. A second method to aggregate fatalities or losses is to divide the area of interest in zones based on the exposure of assets, and compute local FN or losses curves. Then, the aggregate FN/loss-curve is derived by summing random variables with prescribed marginal distributions. In the third method, the aggregate FN-curve is usually determined by simulations that use a spatial correlation model for the hazard component [23, 24] and possibly a correctly correlated injury model.

4. Upper bounds for aggregate societal risk

When aggregate injuries and fatalities (or economic losses) are of interest, we essentially consider the problem of summing random variables. The random variables, here for simplicity denoted as X_{i}, are either number of fatalities or losses for a given location, and X_i is a vector of X_{i}. Then, the random variable sum $S = \sum_i X_i$ represents either the aggregate number of fatalities or the aggregate loss. In this section, we focus on the problem of determining lower and upper probabilistic bounds for the CCDF of the random variable S when we know the marginal CCDF of X_i, but the joint distribution is either unspecified or costly to determine. Observe that the CCDFs are simply the local FN curves or loss curves that can be determined by state-of-art PSRA and risk analysis software (e.g. [25]).

A trivial but yet important observation must be made when aggregating fatalities and losses arising from sum of dependent random variables. The magnitude of importance of the correlation structure between $\boldsymbol{X_i}$s is related to the selection of the risk metric. For example, if the expected number of fatalities (Eq. 7), or expected loss is of interest, the knowledge of the joint distribution of the $\boldsymbol{X_i}$s is unnecessary. In fact $\boldsymbol{E}[\boldsymbol{S}] = \sum_{\boldsymbol{i}} \boldsymbol{E}[\boldsymbol{X_i}]$ regardless of the joint distribution of the $\boldsymbol{X_i}$s. On the other hand, the joint distribution becomes increasingly important when the risk metrics is nonlinear, which is the case for all the other metrics of Section 2. In the following, we assume that the random variables are at most positively correlated. This is corroborated by physical evidence from seismic events, i.e., two closely spaced random variables are expected to have similar values, while sufficiently dispersed random variables are expected to have "more" independent values.

Since we assume that the random variables are positively correlated, a probabilistic lower bound for the aggregate CCDF is simply given by considering all the random variables statistically independent. Given this, we can write

$$G_{\underline{S}}^{lb}(x) = \mathcal{F}^{-1}[\mathcal{F}[G_I(x)](\omega) \cdot \ldots \cdot \mathcal{F}[G_i(x)](\omega) \cdot \ldots \cdot \mathcal{F}[f_i(-x)]](x), \tag{28}$$

where $\boldsymbol{\mathcal{F}}[\cdot]$ indicates the Fourier transform operator, $\boldsymbol{G_i(x)}$ is the CCDF of the random variable $\boldsymbol{X_i}$, $\boldsymbol{G_{\underline{S}}^{lb}(x)}$ is the CCDF of the random variable $\underline{\boldsymbol{S}}$, the suffix $\cdot^{\boldsymbol{lb}}$ and the underscore $\underline{\cdot}$ stand for lower bound, $\boldsymbol{f_i(x)}$ is the PDF of the random variable $\boldsymbol{X_i}$, and $\boldsymbol{I}$ is the total number of locations. Eq. (28) is derived based on the following recursive relationship $\boldsymbol{G_{i+1}(x) = G_i(x) * f_{i-1}(-x)}$ (where $*$ denotes convolution) for the sum of consecutive statistically independent random variables, and by using the Fourier transforms to convert convolutions into multiplications.

A first probabilistic upper bound for the sum of random variables with given marginal distribution and unknown joint distribution is given in the context of financial mathematics in [27, 28]. In order to determine the $G_{\overline{S}}^{ub}(x)$, we first need to provide the definition of *convex order sense* [27].

Definition 1: The random variable S precedes in the *convex order sense* the random variable $\overline{S}$, in notation $S \leq_{CX} \overline{S}$ if and only if[1]

$$E[S] = E[\,\overline{S}], \tag{29}$$

and

$$\int_x^\infty G_S\left(x'\right)dx' \leq \int_x^\infty G_{\overline{S}}^{ub}\left(x'\right)dx' \tag{30}$$

Observe that Eq. (29) is a measure of the weight of the right tail of S and $\overline{S}$. Moreover, since $E[S] = \int_0^\infty G_S\left(x'\right)dx'$, (30) implies that $G_S(x)$ and $G_{\overline{S}}^{ub}(x)$ should cross at least once. Shaked and Shanthikumar [29] have proven the following important relationship

$$S \leq_{CX} \overline{S} \Leftrightarrow E[\phi(S)] \leq E\left[\phi\left(\overline{S}\right)\right], \tag{31}$$

for *all convex* functions $\phi(\cdot)$, provided the expectation exists. Observe that since $E[\phi(S)] = \int_0^\infty \phi(S)|dG_S|$ and the p-norms, $\|\cdot\|_{p,G_S}$, and $\|\cdot\|_{p,G_{\overline{S}}^{ub}}$, are convex functions defined in the spaces $\mathcal{L}_{G_S}^p(\mathbb{R}^+)$ and $\mathcal{L}_{G_{\overline{S}}^{ub}}^p(\mathbb{R}^+)$, we can write the following relationship

$$S \leq_{CX} \overline{S} \Longrightarrow \|\cdot\|_{p,G_S} \leq \|\cdot\|_{p,G_{\overline{S}}^{ub}}. \tag{32}$$

This has an important implication for all the consistent risk metrics defined in Section 2; in fact, if they are applied on $\overline{S}$, they represent an upper bound of the real risk. Observe that Eq. (32) is not a reversible implication, so p-norms cannot be used to imply convex order sense.

Goovaerts et al. (2000) have proven the following convex order relationship

$$S = \sum_{i=1}^{I} X_i \leq_{CX} \sum_{i=1}^{I} F_{X_i}^{-1}(U) \triangleq \overline{S}, \tag{33}$$

where U is a uniform random variable over the interval [0,1], and $F_{X_i}^{-1}(\cdot)$ is the inverse CDF of the X_i random variable. Notice that the component of vector $[F_{X_1}^{-1}(U), \ldots, F_{X_I}^{-1}(U)]$ are highly dependent, since all of the

[1] The original definition involves two generic random variable X, Y. Here, we use S and $\overline{S}$ only to remark the final goal of the approximation.

components are non-decreasing functions of a single random variable. The joint probability distribution of (33) is known as the comonotonous joint distribution. It can be easily proven that the inverse CDF of a sum of comonotonous random variables is simply the sum of the inverse of the marginal. Then we can write

$$F_{\bar{S}}^{-1}(p) = \sum_{i=1}^{I} F_{X_i}^{-1}(p), \tag{34}$$

where, in this case, $p \in [0,1]$ is the percentile of the distribution (to not be confused with with the p-norm). Relationship (34) can be interpreted as the maximum risky combination for the sum of random variables with the prescribed marginal distributions and unknown correlation. Relationship (34) is important, since $G_{\bar{S}}^{ub}(x) = 1 - (F_{\bar{S}}^{-1}(p))^{-1}$ completely defines the $\mathcal{L}_{G_{\bar{S}}^{ub}}^{p}(\mathbb{R}^+)$ space. It follows that any risk metric based on the framework of Section 2 and computed in the $\mathcal{L}_{G_{\bar{S}}^{ub}}^{p}(\mathbb{R}^+)$ space is an upper bound of the real risk computed on the undefined $\mathcal{L}_{G_S}^{p}(\mathbb{R}^+)$ space. As a consequence, if the upper bound is below a predefined safety margin, no further computation is required.

Relationship (34) is a sensible choice for approximating the unknown CCDF of S. However, there is an improved upper bound that can be derived when the following conditions apply: i) all the random variables X_is are conditioned on a parent random variable, here named M, ii) all the conditional CCDF $G_{X_i|M=m}$ are known, and iii) the marginal G_M is given. The parent random variable name, M, has been chosen since in seismic risk analysis the earthquake magnitude has exactly these characteristics. By defining the random variables $\psi_i(U,M)$, where $\psi_i(u,m) = F_{X_i|m}^{-1}(u)$, U is a uniform (0,1) distribution and statistically independent from M, Dhaene et al. [27] have proven the following relationship

$$S \leq_{CX} \sum_{i=1}^{I} \psi_i(U,M) \triangleq \bar{\bar{S}} \leq_{CX} \bar{S}. \tag{35}$$

It can be also shown that $[\psi_1(U,M), \dots, \psi_i(U,M), \dots \psi_I(U,M)]$ is a series of comonotic random variables. It follows that

$$F_{\bar{\bar{S}}|M=m}^{-1}(p) = \textstyle\sum_{i=1}^{I} F_{X_i|M=m}^{-1}(p) \text{ with } p \in [0,1] \tag{36}$$

and after inverting (36), we obtain the CCDF of $\bar{\bar{S}}$ as

$$G_{\bar{\bar{S}}}^{ub}(x) = 1 - \int_0^{\infty} F_{\bar{\bar{S}}|M=m}(x) dF_M(m) \tag{37}$$

Relationship (37) is an improved upper bound to the real societal risk. As we have seen above, $G_{\bar{\bar{S}}}^{ub}(x)$ completely defines a $\mathcal{L}_{G_{\bar{\bar{S}}}^{ub}}^{p}(\mathbb{R}^+)$ space which is “closer” to the true $\mathcal{L}_{G_S}^{p}(\mathbb{R}^+)$ space. It follows that any risk metric based on the framework of Section 2 and computed in the $\mathcal{L}_{G_{\bar{\bar{S}}}^{ub}}^{p}(\mathbb{R}^+)$ space is an improved upper bound of the true risk computed on the undefined $\mathcal{L}_{G_S}^{p}(\mathbb{R}^+)$ space. In practice, a first risk upper bound can be easily computed by computing any convex risk metric in the $\mathcal{L}_{G_{\bar{S}}^{ub}}^{p}(\mathbb{R}^+)$ space. If this upper bound satisfies a predefined risk threshold, then no further computations are required.

5. Conclusion

The first part of this paper presented a short review of the commonly used risk metrics measuring individual and societal risk. It is shown that most of the metrics, which are derived from different disciplines, have different properties and units. This is an obstacle for a selection of a suitable metric for the field of interest, for metrics with different units cannot be compared. Moreover, it limits cross-comparisons among different fields. To overcome this issue, we proposed a mathematical framework based on functional spaces. These spaces are defined with respect to the local FN or loss curve distributions, and they are equipped with a p-norm. We showed that the concept of p-norm can be used to define flexible, yet unit-consistent risk metrics. Risk metrics defined within this framework are convex. Convexity is a desirable property for many applications, such as risk aggregation or risk optimization problems, as allows for bounding the risk metrics.

In the second part of this paper, we proposed a framework for individual risk calculation based on the combination of PSHA analysis and the PEER probabilistic performance-based seismic evaluation framework. This combination is a standard procedure for probabilistic seismic risk analysis, where injury and fatality risk is not regulated in terms of individual or societal risk, but rather based on probabilities of failure of structures. For this reason, we outlined a simple adaptation to compute individual risk. Furthermore, we indicate the main shortcomings of the framework, which are related to the non-stationarity nature of the induced seismicity, and to the currently available structural fragility functions, which are calibrated for macro-seismic events.

Finally, in the last section, we address the problem of determining spatial aggregate risk measure such as societal risk. Often, aggregating risk arising from different sources boils down to the problem of determining the marginal distribution of a sum of correlated random variables. Generally, a spatial region of interest is divided into zones based on the asset exposure. Then, for each zone, local FN and loss distributions are computed. Based on these local marginal distributions we defined a series of comonotonous random variables, which precede the real original random variables in convex order. The marginal of the sum of comonotonous random variables can be easily determined. It is shown that any convex risk metric expressed in the functional space defined with respect to the surrogate marginal is an upper bound of the real risk. An improved upper bound can be determined if the correlation structure of the problem can be defined. In particular, in seismic risk analyses, all the local FN or losses distributions can be conditioned to a parent random variable, which is the magnitude of the event. By following the same procedure of determining a series of conditional comonotonous random variables, we have shown that it is possible to obtain improved upper bounds for any convex risk metric expressed in the space defined by the improved surrogated distribution. Upper bounds to the real risk are important, for they can be checked against predefined safety thresholds.

6. References

[1] Galanis, P. H., & Moehle, J. P. (2015). Development of Collapse Indicators for Risk Assessment of Older-Type Reinforced Concrete Buildings. *Earthquake Spectra*, *31*(4), 1991-2006.

[2] 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.

[3] McGuire, R. K. (2004). *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute.

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

[5] Galanis, P., Shin, Y. B., & Moehle, J. P. (2016). Laboratory and computer simulations of reinforced concrete frames with different ductility. *Earthquake Engineering & Structural Dynamics*.

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

[7] Institute of Chemical Engineering (1992). Nomenclature for hazard and risk assessment in the process industries.

[8] Bedford, T., & Cooke, R. (2001). Probabilistic risk analysis: foundations and methods. Cambridge University Press, New York, 2001.

[9] Hazard and Risk Assessment for Induced Seismicity in Groningen Interim Update November 2015. http://www.namplatform.nl/feiten-en-cijfers/feiten-en-cijfers-onderzoeksrapporten.html#iframe-L2VtYmVkL2NvbXBvbmVudC8/aWQ9b25kZXJ6b2Vrc3JhcHBvcnRlbg==

[10] Piers, M. (1998). Methods and models for the assessment of third party risk due to aircraft accidents in the vicinity of airports and their implications for societal risk. In *Quantified Societal Risk and Policy Making* (pp. 166-204). Springer US.

[11] Laheij, G. M. H., Post, J. G., & Ale, B. J. M. (2000). Standard methods for land-use planning to determine the effects on societal risk. *Journal of Hazardous Materials*, *71*(1), 269-282.

[12] Hirst, I. L., & 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*, *92*(3), 223-237.

[13] Vrijling, J. K., Van Hengel, W., & Houben, R. J. (1995). A framework for risk evaluation. *Journal of hazardous materials*, *43*(3), 245-261.
[14] Hirst, I. L., & 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*, *92*(3), 223-237.
[15] Kroon, I. B., & Hoej, N. P. (2001). Application of risk aversion for engineering decision making. *Safety, risk and reliabilitytrends in engineering, Malta*.
[16] Akinci, A., Galadini, F., Pantosti, D., Petersen, M., Malagnini, L., & Perkins, D. (2009). Effect of time dependence on probabilistic seismic-hazard maps and deaggregation for the Central Apennines, Italy. *Bulletin of the Seismological Society of America*, *99*(2A), 585-610.
[17] Convertito, V., Maercklin, N., Sharma, N., & Zollo, A. (2012). From induced seismicity to direct time-dependent seismic hazard. *Bulletin of the Seismological Society of America*, *102*(6), 2563-2573.
[18] Mignan, A., Landtwing, D., Kästli, P., Mena, B., & Wiemer, S. (2015). Induced seismicity risk analysis of the 2006 Basel, Switzerland, Enhanced Geothermal System project: Influence of uncertainties on risk mitigation.*Geothermics*, *53*, 133-146.
[19] Bommer, J. J., Crowley, H., & Pinho, R. (2015). A risk-mitigation approach to the management of induced seismicity. *Journal of Seismology*, *19*(2), 623-646.
[20] Galanis, P., Sycheva, A., Broccardo M., Mimra, W., Stojadinovic, B. (2015). A Risk Management Approach for the Swiss Seismic Hazard. *Safety and and Reliability of Complex Engineered Systems: ESREL 2015, 331*.
[21] Der Kiureghian, A. (2005). Non-ergodicity and PEER's framework formula. *Earthquake engineering & structural dynamics*, *34*(13), 1643-1652.
[22] Broccardo, M., Galanis, P., Esposito, S., & Stojadinovic, B. (2015). Probabilistic resilience assessment of civil systems: Analysis and validity of the PEER framework. *Safety and Reliability of Complex Engineered Systems: ESREL 2015*, 331.
[23] Park, J., Bazzurro, P., & Baker, J. W. (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Applications of statistics and probability in civil engineering. Taylor & Francis Group, London*, 1-8.
[24] Esposito, S., & Iervolino, I. (2011). PGA and PGV spatial correlation models based on European multievent datasets. *Bulletin of the Seismological Society of America*, *101*(5), 2532-2541.
[25] Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards*, *72*(3), 1409-1427.
[26] Kaas, R., Dhaene, J., & Goovaerts, M. J. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, *27*(2), 151-168.
[27] Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., & Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, *31*(1), 3-33.
[28] Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., & Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: applications.*Insurance: Mathematics and Economics*, *31*(2), 133-161.
[29] Shaked, M., & Shanthikumar, J. G. (Eds.). (1994). *Stochastic orders and their applications*. Academic Pr.