# A comprehensive synthetic database of global seismic losses covering the period 1967-2018.

Cyrielle Dollet[1], Philippe Guéguen[1], Andres Hernandez[1,2]

[1]ISTerre, Université Grenoble Alpes, CNRS, IRD, Université Savoie Mont-Blanc, Université Gustave Eiffel
[2]Fugro, France

Bulletin of Earthquake Engineering volume 21, pages 4265–4288 (2023)
https://doi.org/10.1007/s10518-023-01695-x

**Corresponding author:**
Philippe Guéguen
ISTerre
Philippe.gueguen@univ-grenoble-alpes.fr

**Abstract**

This work aims to construct a synthetic database of human and economic seismic losses. For weak-to-moderate magnitude and older earthquakes, the catalogs of losses are incomplete, which limits the creation of probabilistic based loss models. Furthermore, the number of earthquakes involving losses has increased in recent years, following a non-stationary Poisson distribution with a rate proportional to the exposed population and GDP. First, this study involved defining a series of empirical models (from definition of magnitude to losses) tested by the likelihood method applied to data from 377 earthquakes with variables related to exposure (exposed population and exposed GDP) and consequences (economic losses, number of fatalities and injuries). For these 377 earthquakes, the spatial variation of the hazard was deduced from USGS ShakeMaps and the social and economic losses evaluated were made stationary by taking into account exposure evolving over time. We then built a synthetic database of seismic losses from the ISC-GEM catalog of epicenters, which is assumed to be complete and homogeneous since 1967 for magnitudes > 5. The combination of the 377 events and the synthetic data indicates that earthquakes of magnitudes [5.5; 6.9] represent 36% of all economic losses, 56% of all fatalities, and 71% of injuries. An occurrence model was then designed to predict the evolution of losses over the next years.

**Keywords**: risk, losses, moderate earthquakes, database

## 1. Introduction

Major seismic catastrophes generate huge economic losses and are usually associated with the most severe earthquakes (usually M>7, Holzer and Savage, 2013). However, with the growing concentration of populations and economies in increasingly dense urban areas, even moderate earthquakes (magnitude around 6) can cause significant direct and indirect economic losses and a number of victims that, although lower, remains unacceptable (e.g., Aquila/Italy, 2009, M 6.3, 309 victims, 0.1% GDP - Gross Domestic Product; Christchurch/New Zealand, 2011, M 6.3, 168 victims, 10% GDP; Emilia-Romana/Italy, 2012, M6.1, 27 victims, 0.1% GDP). To evaluate the risks represented by these seismic events globally or within a specific region, we adapted a procedure used for probabilistic seismic hazard analysis (PSHA) to determine the annual probability (or rate) of suffering losses.

Past events naturally provide key information for modeling what might happen in the future. To perform a probabilistic risk analysis for a given region, a catalog of losses associated with earthquakes must be as homogeneous as possible, covering the longest possible time period (to have enough observations to extrapolate to the lower rates of occurrence in which we are interested), and from the smallest magnitude (to be as complete as possible). To avoid certain conventional biases that are also found in PSHA, such as the fusion of catalogs of different origins or the conversion of basic parameters (e.g., magnitude conversion), global catalogs are used to ensure the homogeneity of the earthquake information and associated losses, and to reduce the bias of the frequency-loss distributions. The complete ISC-GEM catalog (2019) for global earthquakes of magnitudes 5 to 8 for the period 1967-2015 (Di Giacomo et al., 2018) shows that moderate magnitudes (around 6) occur frequently. At the same time, Nievas et al. (2020a; 2020b) showed that the information concerning moderate earthquake losses is incomplete. Dollet and Guéguen (2022) made the same observation by assessing the global occurrence models for human and economic losses due to earthquakes from 1967 to 2018, considering the exposed GDP ($GDP_{Exp}$) and population ($POP_{Exp}$) and by exploring the information in international loss databases (e.g., EM-Dat, NOAA, etc.). These international databases list the consequences of the most recent seismic events, but the magnitude of completeness associated with the losses remains high (Nichols and Beaver, 2008; Holzer and Savage, 2013; Dollet and Guéguen, 2022). Finally, although earthquake occurrence follows a stationary Poisson distribution, seismic losses obey a non-stationary distribution with a rate of occurrence proportional to the increase in population (Holzer and Savage, 2013) and therefore the economic assets exposed.

Most loss estimation models are derived by regression from hazard-, exposure- and consequences-related parameters (e.g. among others, Nichols and Beavers, 2003; Jaiswal and Wald, 2010; Heatwole and Rose, 2013; Guettiche and al., 2017). For the hazard, the ISC-GEM catalog (2019) provides earthquake magnitude and location. However, unlike magnitude, macroseismic intensity (not recorded in ISC-GEM, 2019) provides spatially variable parameter of hazard, to be crossed with a spatially variable exposure (Jaiswal and Wald, 2010; 2013). Consequences depend on exposure, such as exposed population and regional GDP or GDP per capita. These two parameters are essential for predictions (e.g. among others, Christoskov and Samardjieva, 1984; Cha, 1998; Badal and Samardjieva, 2002; Wyss and Trendafilosky, 2011; Spence and al., 2011; Jaiswal and Wald, 2013; Heatwole and Rose, 2013 ; Guettiche et al., 2017) and vary over time, making the process non-stationary (Holzer and Savage, 2013). Dollet and Guéguen (2022) therefore compiled a database (LEQ377) associating ShakeMaps and economic and human losses in relation to the exposed population and exposed GDP at the time of the earthquake. They also concluded on the incompleteness of the losses reported in the international catalogs, in particular for low to moderate magnitude earthquakes, essential to the definition of occurrence models.

The objective of this study is to build a complete and homogeneous catalog of seismic losses for M>5 earthquakes to assess annual occurrence rates. Using the data available in the LEQ377 database (Dollet and Guéguen, 2022), we followed a step-by-step procedure based on traditional practices used for performance-based evaluation or seismic hazard probability i.e., by estimating macroseismic intensities from magnitudes, then losses for a given exposure. Conversion or prediction models were then derived from the observation data compiled by Dollet and Guéguen (2022) for earthquakes of magnitudes 5-8 over the period 1967-2018 with epicentral intensity larger than V, testing models efficiency at each step using the likelihood method (Scherbaum et al., 2004). The models were then applied to the complete ISC-GEM database (2019) for magnitudes >5 for the period 1967-2018 (Di Giacomo et al., 2018) to get a synthetic catalogue of losses. Finally, annual probabilities (or occurrence rate models) for human and economic losses are discussed.

## 2. Data

The data used here to develop the seismic loss models were taken from the LEQ445 database (Dollet and Guéguen, 2022). This database includes seismic events having on the one hand, a spatial representation of the macroseismic intensity in the form of Shakemap and, on the other hand, causing social or economic losses recorded over the period 1967-2018.

Hazard-related data were taken from the Atlas of ShakeMaps (Wald et al., 1999; Allen et al., 2009). ShakeMaps are produced for each earthquake along with information on date of occurrence, location (latitude, longitude, depth) and magnitude. In total, 377 earthquakes were included (catalog LEQ377), with magnitudes ranging from 5 to 8 and an epicentral intensity (I0) above or equal to V, i.e., liable to cause damage (Musson et al., 2010). The observed post-seismic macroseismic intensities would be preferable in this study, due to the inherent uncertainties of the ShakeMaps. However, the ShakeMaps have the advantage of being available for all 377 earthquakes in the catalog (actually the selection criterion for these earthquakes according to Dollet and Guéguen, 2022) and give the spatial variability of ground motion (and thus of the exposure model related to the exposed area for I0> V) in a uniform way whatever the earthquake.

Consequences-related data (human and economic) were merged from authoritative international databases (e.g., NOAA, 2018; EM-DAT, 2018; Desinventar, 2018…). For the LEQ377 catalog, the parameters considered are the fatalities F (corresponding to 272 events), injured J (288 events) and direct economic losses L$ (288 events). The economic losses were homogenized and adjusted to a US$ reference year. Usually, economic losses should have been adjusted according to a country-specific price index and for the time of the earthquake. This index is not avalaible for each time/country dependent earthquake of the LEQ377 catalogue. In Dollet and Guéguen (2022), an average index was then calculated on consumer prices (CPI) and the construction index (CI) for France, as provided by INSEE (French national institute of statistics and economic studies), and for the USA by the United States Census Bureau (Dollet and Guéguen, 2022). These indexes are available for 2016, which is the reference year used in this study for the adjustment of economic losses. Dollet and Guéguen (2022) showed that the LEQ377 events producing the most cumulative losses (76%) correspond to events of magnitudes between 5.6 and 7.3.

To get the losses to obey a stationary process, the losses for the year of the earthquake were scaling according to exposure at the time. The exposure data – exposed population and GDP– were calculated and adjusted to the date of the earthquake (Dollet and Guéguen, 2022) using European Commission spatially-distributed demographic data (Eurostat, 2018), with a resolution of 1km, for the year 2015. The conversion factor between 2015 and the year of the earthquake was calculated based on the United Nations' population table per country for the period 1950-2015 (UN, 2019). Dollet and Guéguen (2022) preferred to calculate exposed GDP per capita in US$2016 to measure the development of the area concerned; according to Schumacher and Strobl (2011), this provides better statistical significance for loss results. GDP per capita is the country's GDP divided by its number of inhabitants. This index is expressed in US$ 2010 for each country for the years 1960 to 2018 by the World Bank (World Bank, 2019). Considering GDP to be evenly distributed per country, the GDP per capita is calculated according to the population in the year of the earthquake, then converted into $2016 using the economic conversion index.

Figure 1 shows the distribution of social and economic losses by epicentral intensity (I0) for the LEQ377 database. The earthquakes with I0 between VII and IX are the largest contributors (in number) to the earthquakes. The social losses (fatalities F and injuries J) and economic losses (L$) of the earthquakes I0>=VII represent 99% and 83% of total losses in the LEQ377 database, respectively. Events of intensities I0 between V and VI represent just 1% of total social losses and 17% of economic losses. The dispersion is high, regardless of the intensities and losses considered, with many atypical values (outside the fourth quartile).

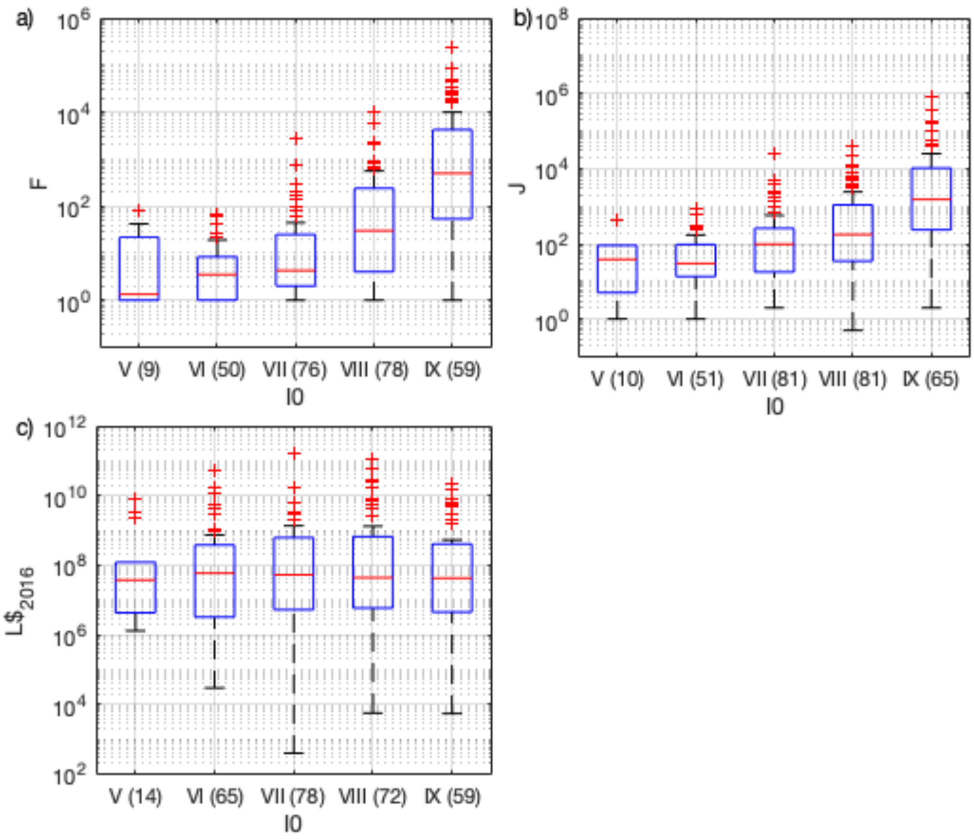


**Figure 1**. Social and economic losses by epicentral intensity I0 of earthquakes in the LEQ377 database. a) Number of fatalities (F); b) Number of people injured (J); c) Economic losses (L$_{2016}$). The number of events per intensity I0 is indicated in brackets.

In LEQ377, earthquake losses are associated with the total losses, i.e. considering direct and indirect losses, distinction rarely provided in the international losses databases. By consequent, loss values in LEQ377 may include losses from secondary effects that can introduce a bias in the model (see Daniell et al., 2017 for the contribution of secondary effets). .

## 3. Method

Figure 2 shows the four steps applied to the LEQ377 database to develop the conversion or loss prediction models used to compile the synthetic database: (1) Magnitude to intensity conversion; (2) Exposed area assessment; (3) Exposure values assessment; (4) Economic and social losses assessment.

At each step, the model efficiency was evaluated using indicators from the likelihood method proposed by Scherbaum et al. (2004) for ground motion prediction models testing. The quality of the model ("goodness-of-fit") is obtained by qualifying the adjustment of the model and estimating the extent to which the statistical model hypotheses are met. Scherbaum et al., (2004) thus combined the properties of the residual distributions with a likelihood measurement (LH). The residuals are normalized to obtain a zero mean and unit variance distribution. The quality of a model in relation to the data is ultimately the probability that the absolute value of a random sample of the normalized distribution falls between the absolute value of a particular observation $|z_0|$ and ∞. Considering the two distribution tails of the error function Erf(z), the likelihood value LH($|z_0|$) is obtained thus (Scherbaum et al., 2004):

$$\mathrm{LH}(|z0|) = \mathrm{Erf}\left(\frac{|z0|}{\sqrt{2}}, \infty\right) = \frac{2}{\sqrt{\pi}} \int_{|z0|\sqrt{2}}^{\infty} e^{-t^2} dt \qquad (1)$$

Here, we assume that each model developed in steps 1 to 4 can be described by a log-normal distribution. Although the hypotheses of the model correspond exactly for samples taken from a unit variance normal distribution, the samples of the random variable LH are also distributed between 0 and 1. This makes it easy to quantify the adjustment quality using the characteristics of the residual distribution and the properties of the LH values (Scherbaum et al., 2004).

The absolute values of the mean (Mean-NRES), median (Med-NRES), standard deviation (Std-NRES) of the residuals and of the median of the LH values (Med-LH) thus indicate the central tendency and the diffusion of the distribution. Using these values, Scherbaum et al., (2004) classified the models into three categories of goodness-of-fit (Table 1). The additional category D (unacceptable) applies if the indicators do not meet any criteria of the categories A, B and C. Note that these criteria are determined based on data, i.e., they measure the quality of the model only within the limits of the data available (Scherbaum et al., 2004).

**Table 1**. Model classification according to the quality criteria defined by Scherbaum et al. (2004). Med-LH: median value of LH; σN: normalized standard deviation of the sample; Mean-NRES, Med-NRES, Std-NRES: mean, median and standard deviation of the residuals.

| Category | Acceptance | Med-LH | Mean-NRES | Med-NRES | Std-NRES | σN |
|---|---|---|---|---|---|---|
| C | Low | ≥ 0.2 | ≤ 0.75 | ≤ 0.75 | ≤ 0.75 | ≤ 1.5 |
| B | Middle | ≥ 0.3 | ≤ 0.5 | ≤ 0.5 | ≤ 0.5 | ≤ 1.25 |
| A | High | ≥ 0.4 | ≤ 0.25 | ≤ 0.25 | ≤ 0.25 | ≤ 1.125 |

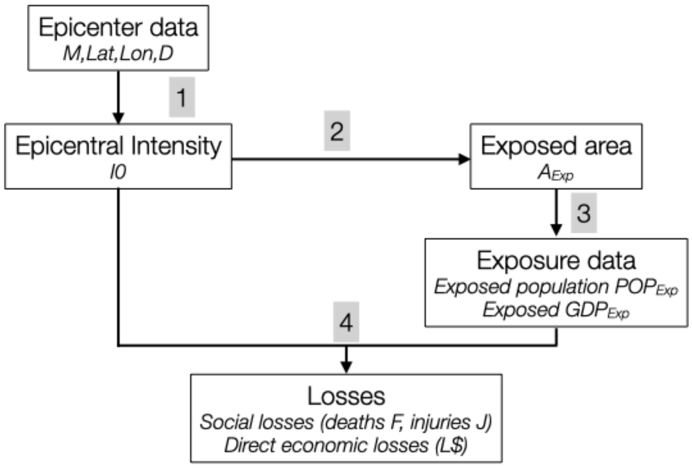


**Figure 2**. Schematic view of the 4-step process followed to establish the predictive equation for earthquakes. M, Lat,Lon,D: magnitude, latitude, longitude and depth of each earthquake. Numbers 1 to 4 correspond to the steps of the process described in section 3.

## 3.1 Step 1 – Epicentral intensity prediction equation

The epicentral intensity I0 of the events in the LEQ377 database is given by the ShakeMaps. The magnitude (M)/depth (D) pair is converted to I0 by adjusting an intensity prediction equation using the functional form as follow (e.g., Atkinson and Wald, 2007; Bindi et al., 2011):

$$I0 = a + bM + c * \log(D) + \sigma \quad (2)$$

where a, b and c are the regression coefficients and σ is the standard deviation. Four relationships are derived, considering different classes of magnitude: [5; 6[, [6; 7[, [7; 8] and [5; 8]. Figures 3 and 4 show the LH and residual distributions for [5; 8] and for each magnitude class, respectively. Table 2 summarizes the coefficients of the model equation (Eq. 2), the rank of the equations, and the adjustment quality indicators according to the LH method. For [5; 8] (Fig. 3), the median of the LH distribution is 0.29, and the absolute values of the mean and median of the normalized residuals are 0 and 0.08, respectively (Table 2), with a model ultimately ranked intermediate (B). The distribution of the normalized residuals (Figure 3a) shows that the variance of the sample tested is greater than the model variance.

**Table 2**. Rankings of different epicentral intensity prediction equation to model the LEQ377 dataset of Dollet and Gueguen (2022).

| M Ranking | Reg. coef Eq. 2 | | Rank | Med-LH | Med-NRES | Mean-NRES | Std-NRES | No. of events |
|---|---|---|---|---|---|---|---|---|
| [5-8] | a | 1.35 | B | 0.29 | 0.08 | 0 | 1.41 | 377 |
| | b | 1.22 | | | | | | |
| | c | -0.69 | | | | | | |
| | σ | 0.8 | | | | | | |
| [5-6[ | a | 2.19 | C | 0.19 | 0.03 | 0 | 2.56 | 85 |
| | b | 0.89 | | | | | | |
| | c | -0.30 | | | | | | |
| | σ | 0.72 | | | | | | |
| [6-7[ | a | 0.77 | B | 0.24 | 0.18 | 0 | 1.48 | 187 |
| | b | 1.43 | | | | | | |
| | c | -0.95 | | | | | | |
| | σ | 0.82 | | | | | | |
| [7-8] | a | 5.40 | C | 0.20 | 0.03 | 0 | 1.67 | 105 |
| | b | 0.64 | | | | | | |
| | c | -0.64 | | | | | | |
| | σ | 0.72 | | | | | | |

For the other magnitude ranges (Fig. 4, Tab. 2), the median of the LH distribution is 0.19, 0.24 and 0.20 for [5-6[, [6-7[ and [7-8], respectively. Considering the absolute value of the mean (0) and the median of the normalized residuals (0.03) for [5-6[ and [7-8], the models are ranked C. In conclusion, for the sake of simplicity, we will use the model [5-8] (rank B) equation without distinguishing the magnitude ranges.

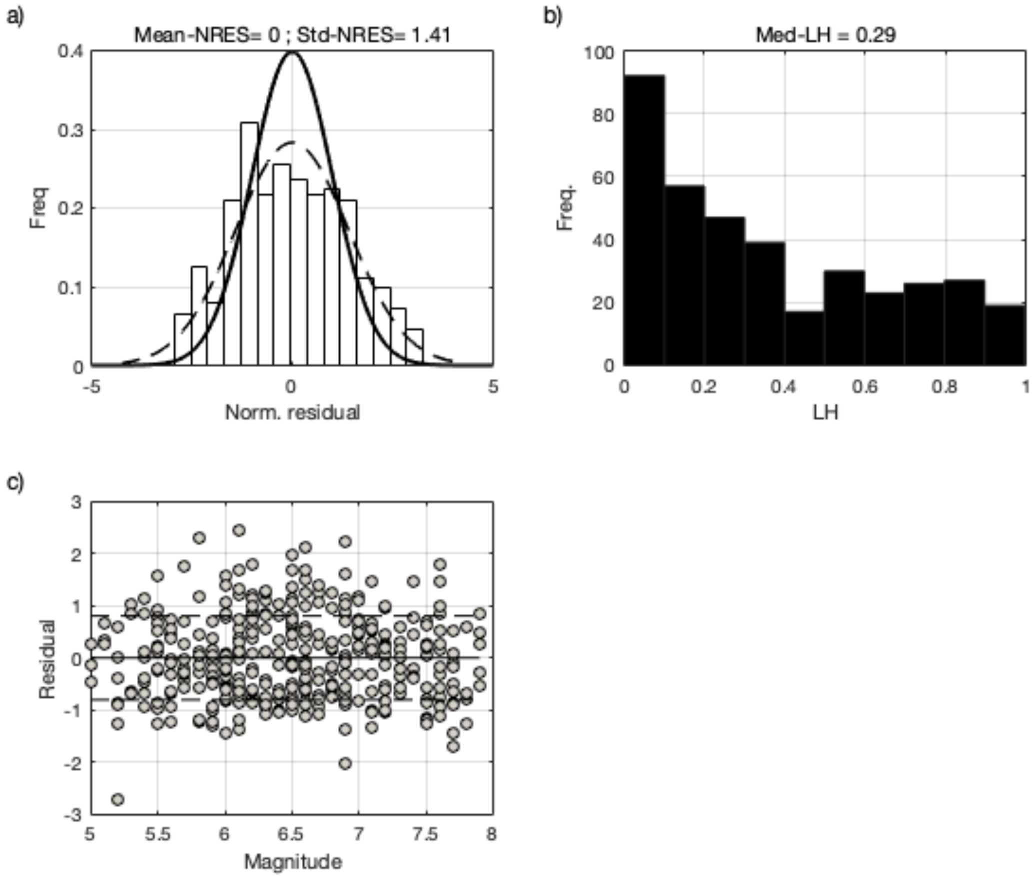


**Figure 3**. Distribution of residuals (a), corresponding LH values (b) and residuals as a function of magnitude (c) for the earthquakes with magnitude ranges [5; 8] from the LEQ377 database. The mean and standard deviation values of the residual distribution and the median value of the resulting LH-value distribution are given at the top of the panel (a) and (b) respectively. The two distribution functions in (a) indicate the unit variance normal distribution (continuous) and the actual residual distribution (dashed). The continuous and dashed horizontal lines in (c) indicate mean +/-std of the residual distribution as a function of magnitude.

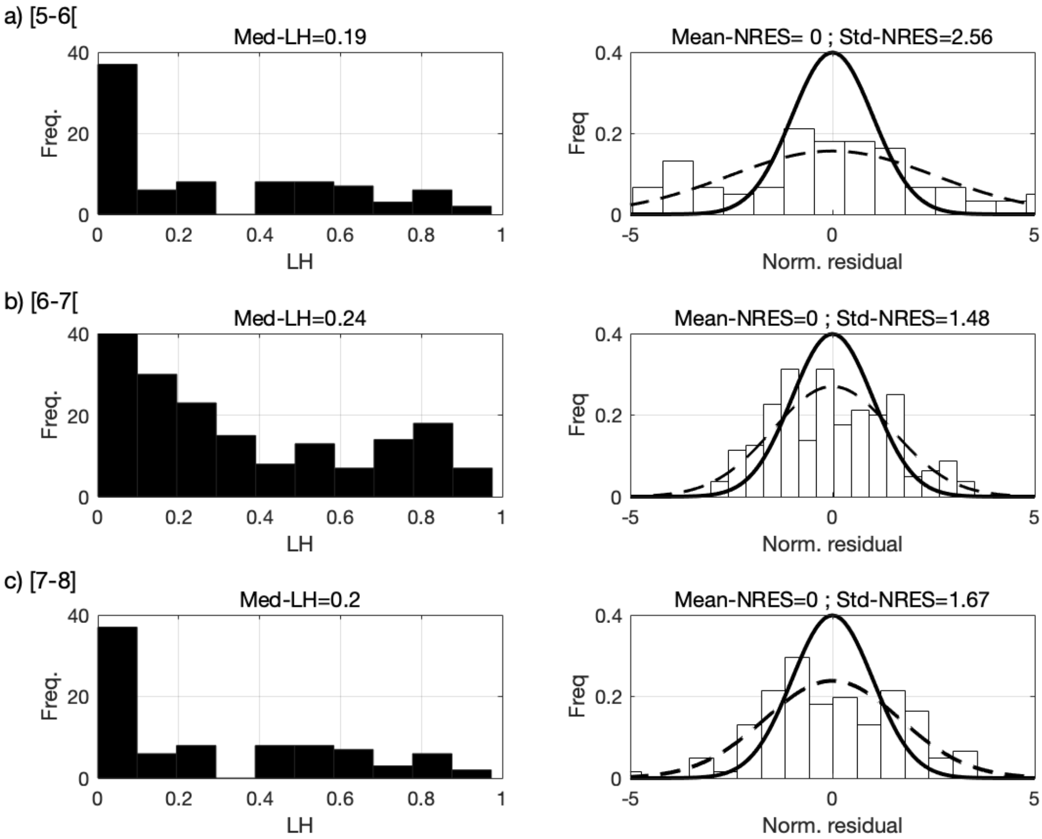


**Figure 4**. Same as Figs. 3a and 3b for magnitude ranges (a) [5-6[; (b) [6-7[; (c) [7-8].

### 3.2 Step 2 – Prediction of exposed areas as a function of I0

To estimate the exposure values, the exposed areas for each macro-seismic intensity are given by the ShakeMaps of the LEQ377 events. As already confirmed by Dollet and Guéguen (2022), the higher the magnitude (M)/depth(D) ratio, the higher the epicentral intensity I0 and the larger the exposed area for each level of macroseismic intensity (e.g., among others Levret and al., 1994; Bakun and Scotti, 2006). Using exposed areas, the epicentral distance and depht is then considered. Without prior functional forms known, we started by testing two conventional forms taken from seismic motion (Douglas, 2003) or intensity (Bakun and Scotti, 2006) predictions:

$$\log\left(\mathrm{A}_{\mathrm{exp/I}}\right) = \mathrm{a(I)} * \mathrm{I0} + \mathrm{b(I)} * \mathrm{D} + \mathrm{c(I)} * \log(\mathrm{D}) \quad (3a)$$

$$\log\left(\mathrm{A}_{\mathrm{exp/I}}\right) = \mathrm{a(I)} * \mathrm{D} + \mathrm{b(I)} * \log(\mathrm{D}) + \sigma \quad (3b)$$

where $\mathrm{A}_{\mathrm{exp/I}}$ is the cumulative exposed area in $km^2$ (i.e., $\mathrm{A}_{\mathrm{exp/I}} = \sum_{I}^{I0} A_{exp/i}$ for macroseismic intensity I>V), a, b and c are the coefficients of the model, σ is the standard deviation, and D is distance in km. We chose to consider cumulative area, because the losses reported in LEQ377 are not given by intensity. For every I0, I range from V to I0.

Without going into further detail, the model (Eq. 3a) does not meet the LH criteria and the quality classes are low (C) to unacceptable (D). Only equation 3b will therefore be discussed hereinafter.

The coefficients a, b, and c of equation 3b depend upon macroseismic intensity: we therefore developed 15 relationships calculating the cumulative exposed area for each macroseismic intensity and for I0 between [V; IX] (Tab. 3). With one exception, the models were all ranked A or B. For example, Figure 5 shows for I0=VII the normalized distribution of residuals (Fig. 5a), the corresponding LH values (Fig. 5b) and the value of the residuals as a function of depth (Fig. 5c) for each macroseismic intensity I≤I0. For $\mathrm{I} = \mathrm{V}\left(\mathrm{i.e.}, \sum_{\mathrm{V}}^{\mathrm{VII}} \mathrm{A}_{\mathrm{exp}/\mathrm{i}}\right)$ to $\mathrm{I} = \mathrm{VII}\left(\mathrm{i.e.}, \mathrm{A}_{\mathrm{exp/VII}}\right)$, the mean residual and the standard deviation of the distribution are 0.1, 0, 0, and 0.75, 0.87 and 1.65, respectively (Table 3). Figure 5c shows an under-estimation of the exposed areas for D>30km (Fig. 5c). The distribution medians of the LH values (Fig. 5b) are 0.55, 0.51 and 0.26 for I=V, I=VI and I=VII, respectively. Bearing in mind that the absolute value of the median of the normalized residuals is 0.06 for I=V, 0.02 for I=VI and 0.05 for I=VII, the exposed area prediction model for I0=VII ranked A (Tab. 3).

Similarly, for I0=VI (Fig. 6), the median value of LH is 0.69 and 0.46 for I=V and VI, respectively. The absolute values of the mean and the mean of the normalized residuals are 0.2 and 0.04 for I=V and 0.2 and 0.18 for I=VI, respectively. Finally, the exposed area model for I0=VI ranked A (Tab. 3).

Note that the normalized residual distributions (Fig. 5a and 6a) are narrower than a unit normal distribution. The sample variance is therefore smaller than that of the model, characterized with bias in terms of mean (Scherbaum and al., 2004). The median of LH is above 0.5 and the associated LH distributions are asymmetrical, i.e., the models under-estimate the cumulative exposed area. The distribution of LH values appears asymmetrical for all models when intensity I is strictly less than I0, with an LH median value above 0.5. When macroseismic intensity I is I0, the frequency of the low values of LH increases and the median of the LH distribution falls below 0.5. According to Table 3, the exposed area models for each I0 rank medium (B) to high (A), except for $\mathrm{A_{exp/VIII}}\ \mathrm{for\ I0 = VIII}$, although we have no explanation for this.

**Table 3.** Rankings of the relationships for cumulative exposed areas as a function of I0 to model the LEQ377 dataset of Dollet and Gueguen (2022).

| I0 | $A_{exp}$ | Reg. coef Eq. 3b a | b | Rank | Med-LH | Med-NRES | Mean-NRES | Std-NRES | Nbr. of events |
|---|---|---|---|---|---|---|---|---|---|
| V | $\mathrm{A_{exp/V}}$ | -0.012 | 1.96 | B | 0.27 | 0.24 | 0.09 | 1.50 | 41 |
| VI | $\sum_{\mathrm{V}}^{\mathrm{VI}} \mathrm{A_{exp/i}}$ | -0.037 | 3.18 | A | 0.69 | 0.04 | 0.20 | 0.63 | 112 |
| | $\mathrm{A_{exp/VI}}$ | -0.031 | 2.25 | A | 0.46 | 0.18 | 0.23 | 1.18 | |
| VII | $\sum_{\mathrm{V}}^{\mathrm{VII}} \mathrm{A_{exp/i}}$ | -0.098 | 4.20 | A | 0.55 | 0.06 | 0.10 | 0.75 | 108 |
| | $\sum_{\mathrm{VI}}^{\mathrm{VII}} \mathrm{A_{exp/i}}$ | -0.069 | 3.45 | A | 0.51 | 0.02 | 0.09 | 0.87 | |
| | $\mathrm{A_{exp/VII}}$ | -0.030 | 2.19 | B | 0.26 | 0.05 | 0.06 | 1.65 | |
| VIII | $\sum_{\mathrm{V}}^{\mathrm{VIII}} \mathrm{A_{exp/i}}$ | -0.090 | 4.43 | A | 0.68 | 0.24 | 0.18 | 0.60 | 83 |
| | $\sum_{\mathrm{VI}}^{\mathrm{VIII}} \mathrm{A_{exp/i}}$ | -0.069 | 3.84 | A | 0.63 | 0.23 | 0.16 | 0.65 | |
| | $\sum_{\mathrm{VII}}^{\mathrm{VIII}} \mathrm{A_{exp/i}}$ | -0.050 | 3.18 | A | 0.59 | 0.15 | 0.15 | 0.75 | |
| | $\mathrm{A_{exp/VIII}}$ | -0.041 | 2.12 | C | 0.22 | 0.26 | 0.11 | 1.62 | |
| IX | $\sum_{\mathrm{V}}^{\mathrm{IX}} \mathrm{A_{exp/i}}$ | -0.34 | 6.49 | A | 0.72 | 0.20 | 0.19 | 0.42 | 33 |
| | $\sum_{\mathrm{VI}}^{\mathrm{IX}} \mathrm{A_{exp/i}}$ | -0.27 | 5.64 | A | 0.76 | 0.14 | 0.18 | 0.49 | |
| | $\sum_{\mathrm{VII}}^{\mathrm{IX}} \mathrm{A_{exp/i}}$ | -0.21 | 4.87 | A | 0.78 | 0.08 | 0.18 | 0.56 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\sum_{VIII}^{IX} A_{exp/i}$ | -0.14 | 3.93 | A | 0.78 | 0.02 | 0.17 | 0.69 |
| $A_{exp/VIII}$ | 0.021 | 1.85 | A | 0.55 | 0.15 | 0.02 | 1.40 |

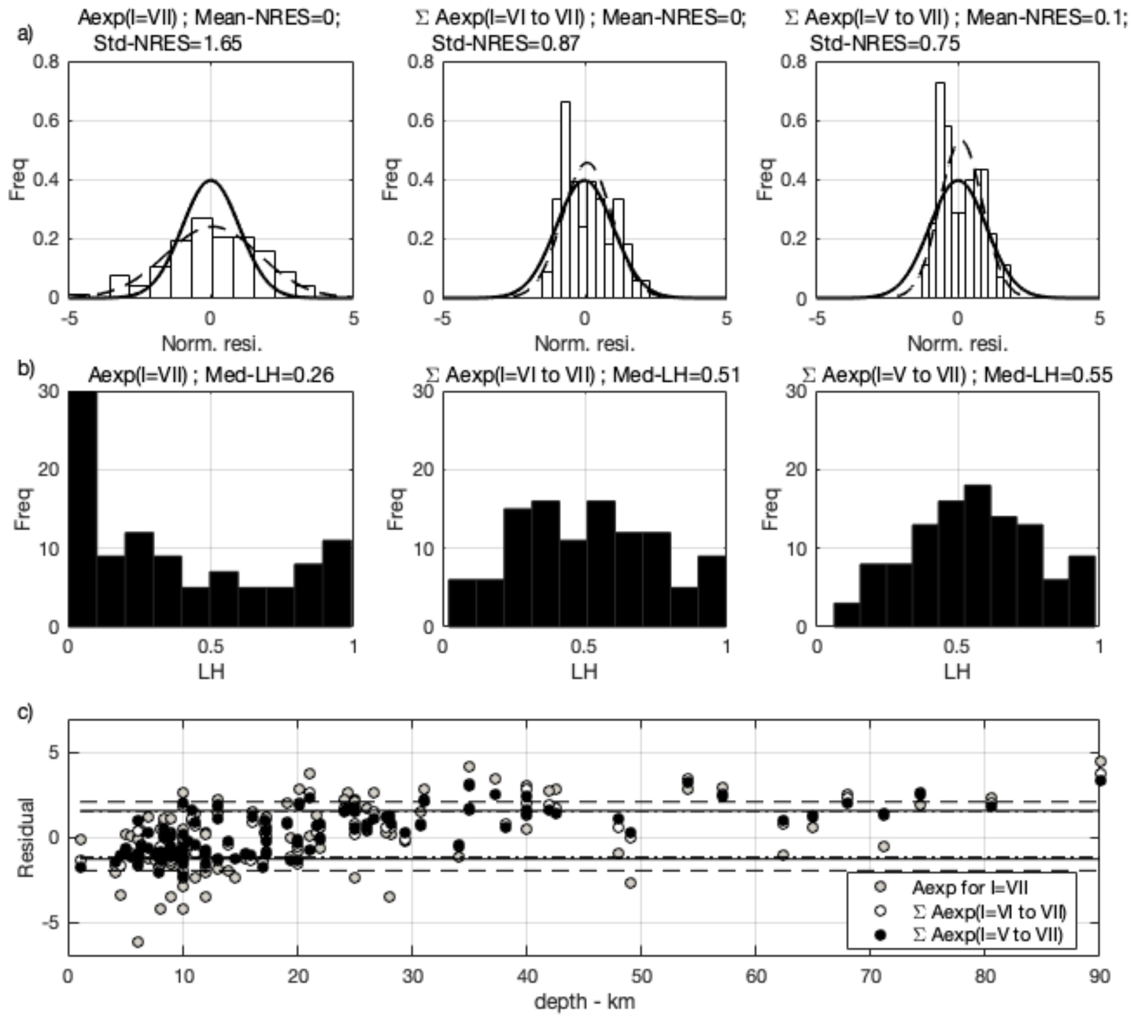


**Figure 5**. Distribution of residuals (a), corresponding LH values (b) and residuals as a function of depth (c) for I0=VII earthquakes in the LEQ377 database. The mean and standard deviation values of the residual distribution (a) and the median value of the resulting LH-value distribution (b) are given at the top of the panels. The two distribution functions in (a) indicate the unit variance normal distribution (continuous) and the actual residual distribution (dashed). The horizontal dashed lines in (c) indicate +/-std of the residual distribution as a function of depth.

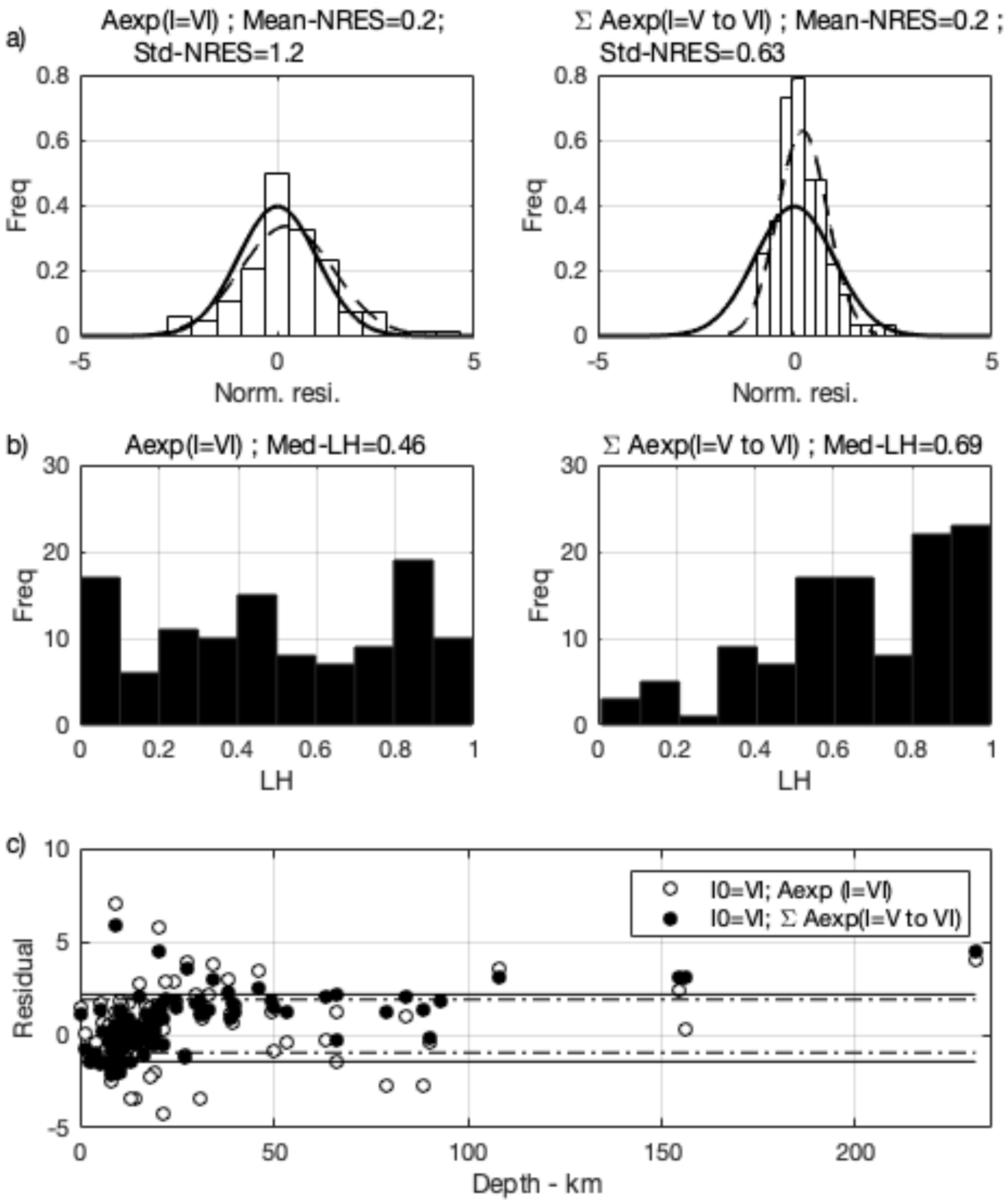


**Figure 6.** Same as Fig. 5 for I0=VI

### 3.3. Step 3 – Loss models

For seismic losses, it is common practice to consider a model with the following functional form (e.g., Heatwole and Rose, 2013; Guettiche and al., 2017):

$$\log10(\mathrm{L}) = \mathrm{c}_1 * \log10(\mathrm{I0}) + \mathrm{c}_2 * \log10(\mathrm{X}_2) + \cdots + \mathrm{c}_\mathrm{i} * \log10(\mathrm{X}_\mathrm{i}) + \sigma \quad (4)$$

where log10(L) represents the logarithmic function of the economic or human consequences, $c_i$ are the regression coefficients and i is the number of variables $X$. This step requires consideration of the exposure variables (exposed population and exposed GDP per capita) for each macroseismic intensity, already calculated for LEQ337 (Dollet and Guéguen, 2022) based on (1) the georeferenced population grid 2015 (Eurostat, 2018) and the demographic growth conversion factor (UN, 2019); (2) GDP per capita of the countries affected by the earthquake (Word Bank, 2019) in US$ in the year of the earthquake, converted to US$(2016) using the average economic index proposed by Dollet and Guéguen (2022). To test the models, only the overall losses associated with each earthquake are given, with no distinction of losses per macroseismic intensity (not available in the international loss databases) but considering exposure for each macroseismic intensity.

**3.3.1** Prediction models for economic losses (L$2016) as a function of I0 and GDPexp

In the LEQ377 database, economic losses are indicated for 288 earthquakes. We derived the model (called DOL22) from Eq. 4, considering cumulative GDP per capita calculated as the sum of exposed GDP per capita for each macroseismic intensity I between [Xi=I; i=V to I0]. Figure 7a shows the distribution of residuals and LH values for the derived model, compared with two existing models produced by Guettiche et al. (2017) (GUE17, Fig. 7b) and Heatwole and Rose (2013) (HEA13, Fig. 7c):

GUE17 $$\log10(\mathrm{L}) = \mathrm{c}_1 * \log10(\mathrm{I0}) + \mathrm{c}_2 * \log10(\mathrm{GDP}_{\mathrm{expTotal}}) + \mathrm{c}_3 \quad (5a)$$

HEA13 $$\log(\mathrm{L}) = \mathrm{c}_1 * \log(\mathrm{Mag}) + \mathrm{c}_2 * \log(\mathrm{Pop}_{\mathrm{expTotal}}) + \mathrm{c}_3 \quad (5b)$$

The LH values of the DOL22 model (Table 4) give a median of 0.30, and the absolute values of the mean and the median of the normalized residuals are 0.03 and 0.02, respectively (model rank B). Note that the distribution of normalized residuals (Fig. 7a, left panel) has a higher variance than the model variance, the low values of LH have an increasing frequency and the LH distribution has a median value that falls below 0.5. The associated mean residual is 0.0, with a standard deviation of the distribution of 1.51.

The coefficient c1 of the DOL22 model (Eq. 4, Table 4) is high, indicating that each increase in I0 causes a large increase in economic losses. This suggests that the economic losses due to earthquakes are much more dependent on the parameter defining the hazard than on the parameters defining exposure.

**Table 4**. Classification of economic loss prediction models (losses in million US$2016) DOL22 (this study), Guettiche et al., (2017, GUE17) and Heatwole and Rose (2013, HEA13). The coefficients are those of equations 4 and 5.

| I0 | Reg. coef Eq. 4 and 5 | | Rank | Med-LH | Med-NRES | Mean-NRES | Std-NRES |
|---|---|---|---|---|---|---|---|
| DOL22 | c1 | 5.57 | B | 0.30 | 0.02 | 0.03 | 1.51 |
| | c2 | 0.21 | | | | | |
| | c3 | 0.05 | | | | | |
| | c4 | 0.02 | | | | | |
| | c5 | 0.11 | | | | | |
| | c6 | 0.06 | | | | | |
| | σ | 1.21 | | | | | |
| GUE17 | $c_1$ | 5.09 | A | 0.52 | 0.52 | 0.51 | 0.77 |
| | $c_2$ | 0.61 | | | | | |
| | $c_3$ | -1.16 | | | | | |
| | σ | 0.56 | | | | | |
| HEA13 | $c_1$ | 8.4 | A | 0.51 | 0.13 | 0.12 | 1.00 |
| | $c_2$ | 0.9 | | | | | |
| | $c_3$ | -8.9 | | | | | |
| | σ | 1.5 | | | | | |

Models GUE17 and HEA13 are ranked A (Tab. 4). However, Figure 7 shows a highly dispersed normal law for HEA13 ($\mu$=-0.4, $\sigma$=2.82) (Fig. 7c) and GUE17 ($\mu$=-0.93, $\sigma$=1.59) (Fig. 7b), although $\mu$=0 and $\sigma$=1.21 for model DOL22 (Figure 9a). Thus, although our model is ranked below GUE17 and HEA13, the consideration of local exposure improves the distribution of economic loss residuals.

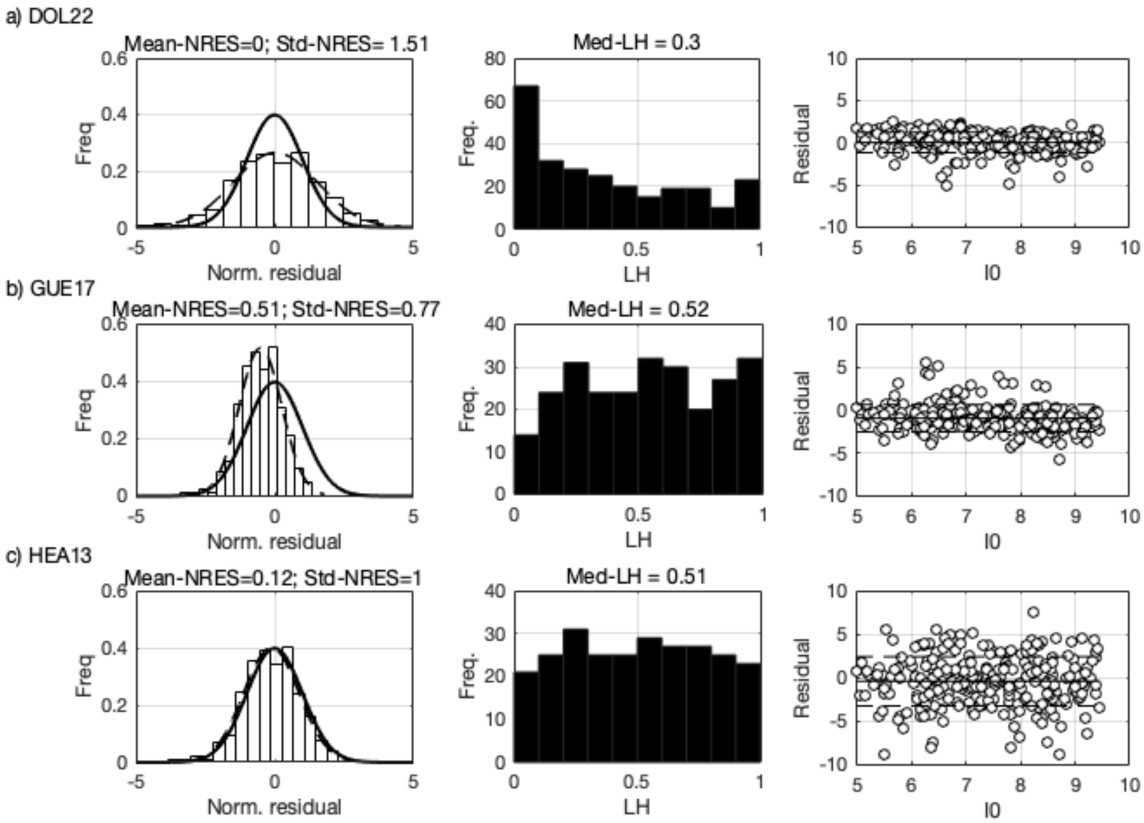


**Figure 7.** Distribution of residuals (left panel), corresponding LH values (middle panel) and residuals as a function of I0 (right panel) for the economic losses of the LEQ377 database, considering DOL22 (a), GUE17 (b), and HEA13 (c). Same symbols and legends as in Fig. 5.

### 3.3.2. Prediction models for fatalities F as a function of I0 and POPexp

Figure 8 and Table 5 summarize the results ranking of model DOL22 (Fig. 8a) and models GUE17 (Figure 8b) and BAD05 (Badal et al. 2005; Figure 8c) for the prediction of fatalities. The equations of models GUE17 and BAD05 are:

GUE17 $$\log10(\mathrm{F}) = \mathrm{c}_1 * \log10(\mathrm{I0}) + \mathrm{c}_2 \quad (6a)$$

BAD05 $$\log(\mathrm{F}) = \mathrm{c}_1(\text{Pop density}) * \mathrm{M} + \mathrm{c}_2(\text{Pop density}) \quad (6b)$$

The median of the LH value distribution (0.26), and the absolute values of the mean and the mean of the normalized residuals (0 and 0.39) rank DOL22 as intermediate (B). Figure 9a shows that the variance of the tested sample is higher than the variance of the model. The positive mean residual indicates that the model prediction model under-estimates the number of fatalities. This under-estimation is most obvious for intensities I0 greater than or equal to VIII (Fig. 8a, right panel). The model adjustment could be improved with loss data for each macroseismic intensity, but these are not currently available in the international databases. Unlike the economic model, the coefficient values $c_i$ (in absolute value) rank the same for I0 and for the variables related to exposure, suggesting that human losses depend as much on the parameter related to the hazard event as on the parameter related to exposure.

**Table 5**. Classification of prediction models for number of fatalities DOL22 (this study), GUE17 (Guettiche et al., 2017) and BAD05 (Badal et al., 2005). For DOL22, the coefficients $c_i$ (i>1) correspond to the exposure values Xi (with i ∈ [V; I0]). For BAD05, coefficients c1 and c2 depend on population density and are not indicated in the table (see Tab. II in Badal et al. 2005).

| I0 | Reg. coef Eq. 4 and 6 | | Rank | Med-LH | Med-NRES | Mean-NRES | Std-NRES |
|---|---|---|---|---|---|---|---|
| DOL22 | $c_1$ | -0.37 | B | 0.26 | 0.39 | 0 | 1.44 |
| | $c_2$ | 0.18 | | | | | |
| | $c_3$ | 0.02 | | | | | |
| | $c_4$ | 0.08 | | | | | |
| | $c_5$ | 0.24 | | | | | |
| | $c_6$ | 0.15 | | | | | |
| | $\sigma$ | 0.89 | | | | | |
| GUE17 | $c_1$ | 12.8 | C | 0.24 | 0.31 | 0.09 | 1.5 |
| | $c_2$ | -9.85 | | | | | |
| | $\sigma$ | 0.89 | | | | | |
| BAD05 | $\sigma$ | 1.72 | D | 0.12 | 1.42 | 1.09 | 1.20 |

Models BAD05 and GUE17, which consider overall exposure, rank much lower (D and C, respectively). In particular, $\mu$=-1.13 and $\sigma$=1.72 for BAD05, $\mu$=0.08 and $\sigma$=1.06 for GUE17 (in comparison with $\mu$=0 and $\sigma$=0.89 for DOL22).

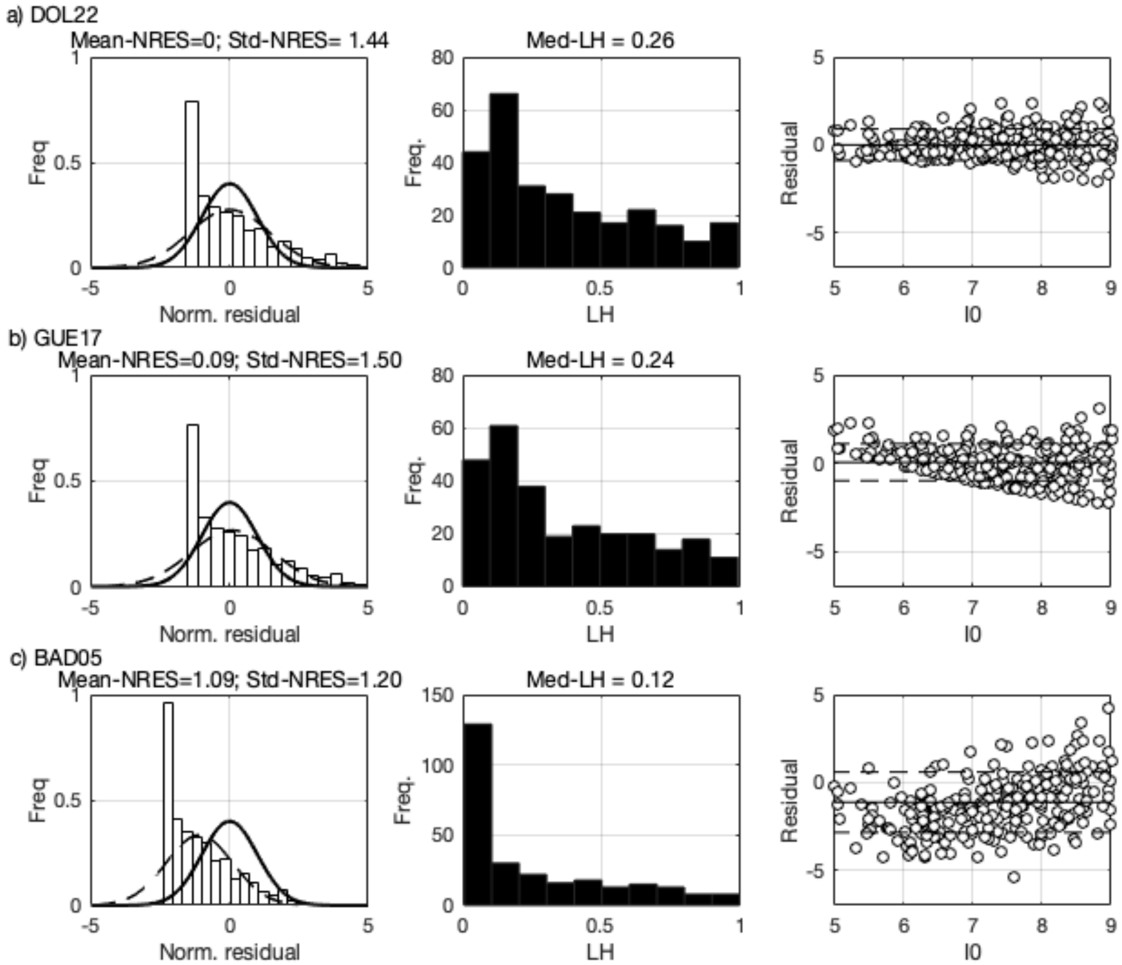


**Figure 8**. Same as Fig. 7 for fatalities and considering DOL22 (a), GUE17 (b) and BAD05 (c).

### 3.3.3. Prediction model for number of people injured as a function of I0 and POPexp

As for the number of fatalities, our hypothesis is that the variables of exposure per intensity improve the prediction of the number of people injured. Figure 9 and Table 6 summarize the results that rank the model derived in this study (Fig. 9a, DOL22) and compared with the Guettiche et al. (2017) model (Fig. 9b, GUE17) for injuries, according to the equation:

GUE17 $\log10(\mathrm{J}) = c_1 * \log10(\mathrm{I0}) + c_2 + \sigma$ (7)

where J is the number of people injured. The median of the LH value distribution (0.31), and the absolute values of the mean and median of the normalized residuals (0 and 0.13) rank DOL22 as intermediate (B). The positive mean residual indicates that the model under-estimates the number of people injured, which does not appear to depend upon intensity I0.

**Table 6**. Classification of the prediction models for the number of people injured: DOL22 (this study) and GUE17 (Guettiche et al., 2017). For DOL22, the coefficients $c_i$ (i>1) correspond to the exposure values Xi (with i ∈ [V; I0]).

| I0 | Reg. coef Eq. 4 and 6 | | Rank | Med-LH | Med-NRES | Mean-NRES | Std-NRES |
|---|---|---|---|---|---|---|---|
| DOL22 | $c_1$ | 0.36 | B | 0.31 | 0.13 | 0 | 1.51 |
| | $c_2$ | 0.26 | | | | | |
| | $c_3$ | -9e-4 | | | | | |
| | $c_4$ | 0.07 | | | | | |
| | $c_5$ | 0.15 | | | | | |
| | $c_6$ | 0.09 | | | | | |
| | σ | 0.86 | | | | | |
| GUE17 | $c_1$ | 9.88 | C | 0.27 | 0.03 | 0.18 | 1.72 |
| | $c_2$ | -6.52 | | | | | |
| | σ | 0.78 | | | | | |

Figure 9 confirms the hypothesis that consideration of the exposed population per intensity improves the prediction of the number of people injured, as model GUE17 is ranked C. The residuals of models DOL22 and GUE17 (Fig. 9, right panel) are normally distributed with a zero mean and unit variance (μ=0 and σ=0.86 for model DOL22 and μ=0.12 and σ=0.99 for model GUE17). The improvement obtained by considering exposure per intensity is visible on the dispersion.

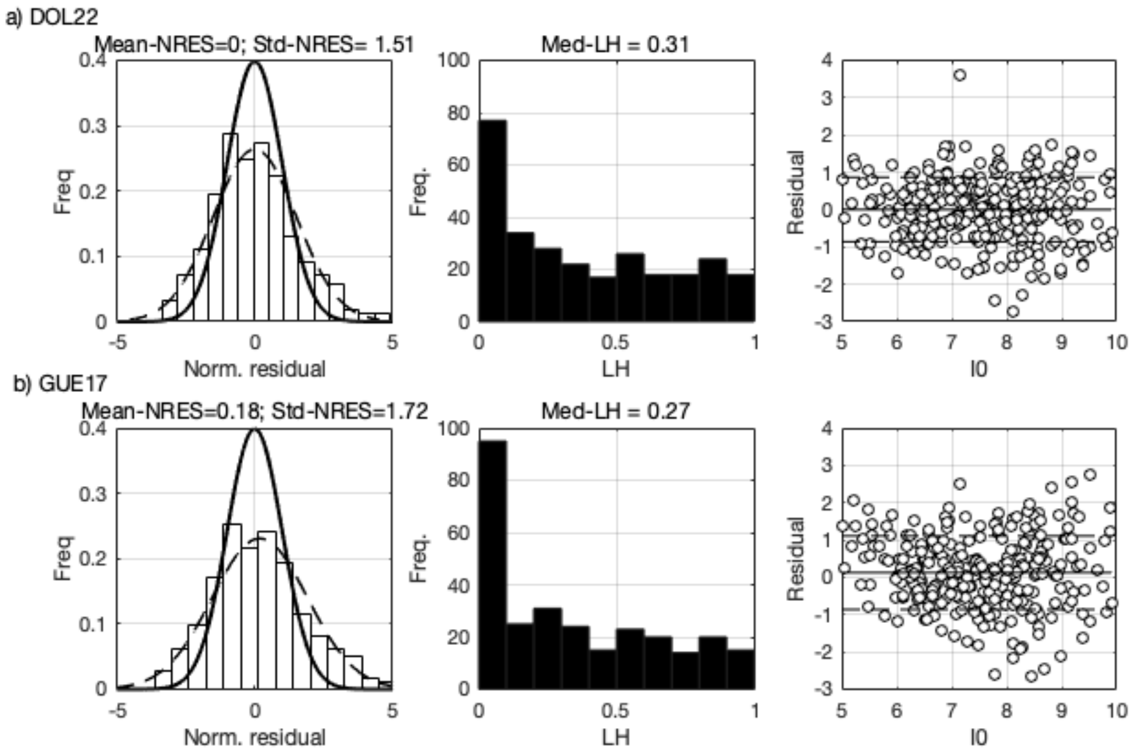


**Figure 9**. Same as Fig. 7 for injuries and considering DOL22 (a) and GUE17 (b).

## 4. Construction of the synthetic database

To compensate for the lack of loss data for low magnitudes, we built a database according to the procedure defined in Fig. 2. First, we calculated I0 values according to Eq. 2 from M and the locations in the ISC-GEM catalog (2019), which is considered complete for M>5 for the period 1900-2015 (Di Giacomo et al., 2018). A total of 17,721 seismic events with I0 ≥ V (i.e., with possible losses), M ∈ [5.5; 8[ over the period 1967-2015 were considered.

Next, the Aexp values derived from Eq. 3b were calculated for all the earthquakes, supposing an equivalent concentric circular surface area for each intensity. Only the earthquakes whose Aexp I>V affected a populated area were retained, population being estimated by crossing Aexp I>V with the georeferenced population table 2015 (European Commission, 2019) for each event. 7,515 events (42%) were thus retained, forming the dataset herein referred to as ISC7515, with the following distribution of number of events per intensity: [V-VI[: 1,797 events; [VI-VII[: 4,220 events; [VII-VIII[: 1,289 events; [VIII-IX[:196 events; [IX-X[: 13 events.

Finally, the exposed GDP per intensity was calculated according to the method proposed by Dollet and Guéguen (2022) using GDP per capita for the country affected (World Bank, 2019) in US$, and the exposed population in 2010, converted to the year 2016 using the average economic index. For earthquakes affecting several countries, exposure values in each country are considered and merged accordingly.

Figure 10 shows the distribution of the number of earthquakes per year, by depth and by magnitude, in the ISC7515 database. As for the LEQ377 database of earthquakes and losses observed (Dollet and Guéguen, 2021), a few noteworthy trends are visible for ISC7515 (Fig. 10): these are the superficial earthquakes that can make the largest contribution to losses; most of the earthquakes (68%) able to cause losses according to the selection criteria applied by this study are events of moderate magnitude, between [5.5 and 6.4].

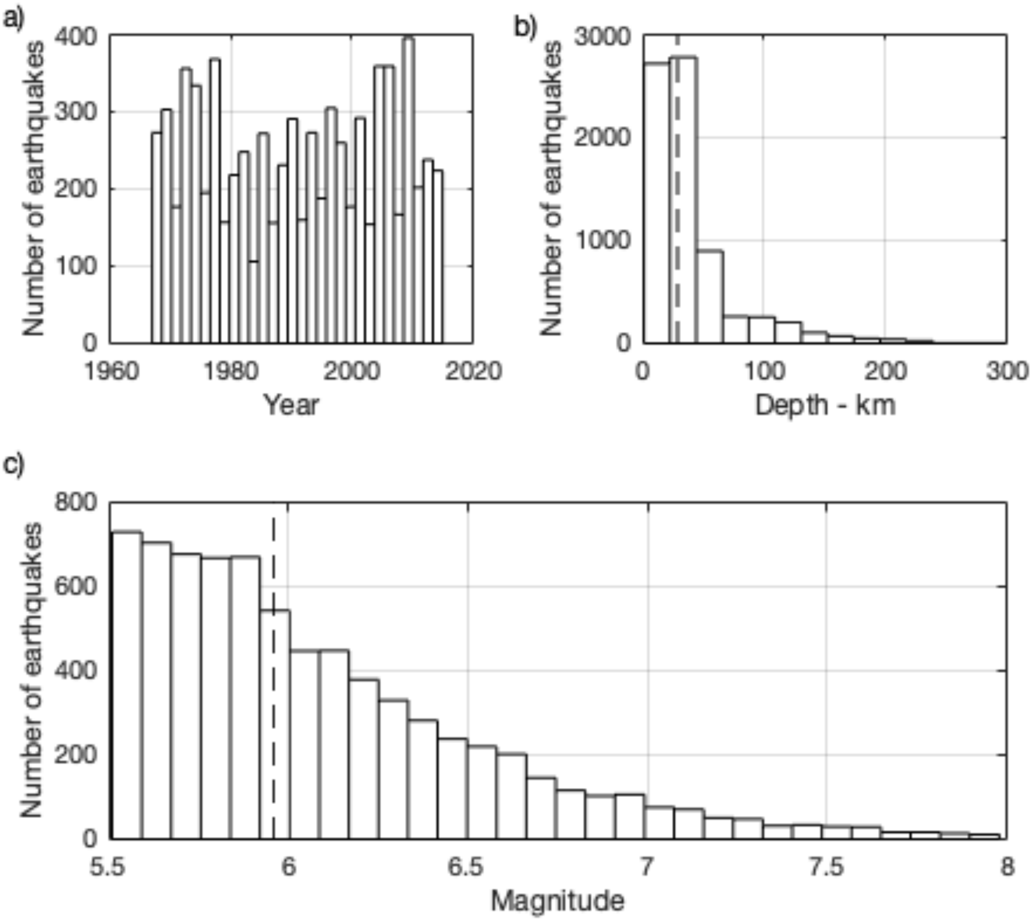


**Figure 10**. Distribution of the number of earthquakes in database ISC7515 a) by year, b) by depth, c) by magnitude. The black dashed lines on b) and c) represent the median values of depth (28.9 km) and magnitude (6.0).

Losses were finally calculated using the Eq. 4 models with the regression coefficients by intensity I0 of Tables 4 to 6. Concerning GDP per capita, only 6,809 seismic events were considered, as the World Bank tables (2019) do not provide GDP values for some exposed countries.

### 4.1 Test on LEQ377

The distribution of the synthetic loss residuals for the LEQ377 events is given in Fig. 11. The synthetic data under-estimate the probable losses. The distributions of the residuals for social and economic losses follow a normal value distribution $\mu = 0.4$; $\sigma = 1.14$ for F, $\mu = 0.31$; $\sigma = 1.02$ for J and $\mu = 0.45$; $\sigma = 1.17$ for L\$. Two hypotheses might explain this under-estimation: (1) distribution of losses related to damage to buildings not considered in this study (because of missing information in the international database) although the losses are highly correlated with structural collapse (e.g., Coburn and Spence, 2003; Riedel and al., 2014; 2015; Guettiche and al., 2017); (2) high uncertainty related to the evaluation of exposed areas per intensity, with consequences on the estimations of exposed populations and GDP. These uncertainties might also include the values given in the ShakeMaps.

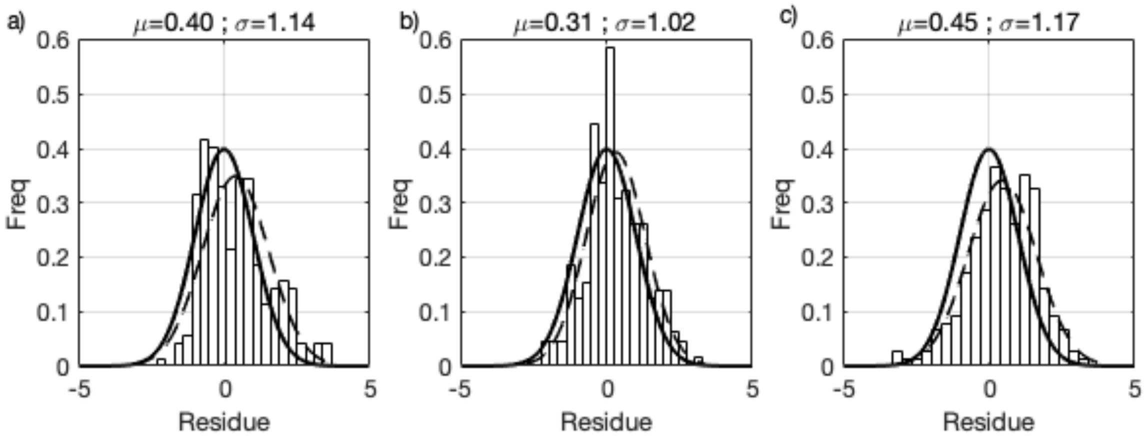


**Figure 11.** Distribution of the residuals of (a) fatalities, (b) number of people injured, and (c) economic losses of seismic events in the LEQ377 database. Residue = log10(obs)-log10(pred).

### 4.2 Synthetic losses

Table 7 summarizes the statistical indicators associated with the social and economic losses calculated from the synthetic database ISC7515. Total economic losses represent US$(2016) 87.736 billion, with a mean loss of around US$(2016) 12.88 million. The Kurtosis are very high, i.e., the distribution tails include more observations than in a Gaussian distribution, and there are a lot of values a long way from the mean. This means that the strongest (M≥7) and low-probability earthquakes contribute significantly to economic losses (64%), and social losses (44% for fatalities and 29% for people injured). However, the cumulative losses for earthquakes between 5.5 and 7.0 (i.e., 95% of ISC7515 events) that correspond to weak-to-moderate earthquakes, contribute very significantly to the global seismic losses: they represent 36% of all economic losses, 56% of all fatalities, and 71% of people injured. Note also that weak-to-moderate earthquakes may cause higher cumulative social losses but lower cumulative economic losses than stronger magnitude earthquakes.

**Table 7**. Economic losses L$ (US$2016), number of fatalities and people injured calculated for the events in the synthetic database ISC7515

| | Fatalities | Injuries | Economic L$2016 |
|---|---|---|---|
| Total value | 54,713 | 366,559 | 87,736,086,272 |
| Mean | 7 | 49 | 12,885,312 |
| Standard deviation | 79.7 | 198.1 | 145,854,342 |
| Kurtosis | 3,379.9 | 2,145.8 | 1,775.2 |
| Median | 3.5 | 28.5 | 1,950,616 |
| Number of events | 7,515 | 7,515 | 6,809 |

## 5. Annual rate of exceedance of losses

Assuming a complete synthetic loss catalog, we then derived an exceedance model for the 7,515 seismic events between 1967 and 2018, according to the following functional form inspired by the seminal Guttenberg-Richter model:

$$\log_{10} N(\mathrm{Y} > \mathrm{X}) = \mathrm{a} - \mathrm{b} * \log_{10}\mathrm{X} \quad (9)$$

where N(Y>X) represents the number N of earthquakes with a ratio Y of social losses to exposed population ($F/POP_{exp}$ or $J/POP_{exp}$) or economic losses ($L\$/GDP_{exp}$) greater than or equal to a given loss X. Note that Eq. 9 predicts losses with no upper limit. However, physical constraints related to exposure (population and GDP) make this unrealistic, i.e., in the spirit of the maximal magnitude of the Guttenberg-Richter model defined by the finite size of the source faults of a given region. For this reason, limited loss models, called bounded loss exceedance models, are computed with the number of deaths and injuries limited by the value of the exposed population and the economic losses limited by twice the exposed GDP.

Figure 12 represents the cumulative distribution of the unbounded and bounded synthetic losses, computed following the maximum likelihood estimation-based method proposed by Ogata and Katsura (1993). Completeness value Lc and regression coefficients (Eq. 9) are given in Table 8.

**Table 8**. Completeness and regression coefficients (Eq. 9) of the synthetic database (from Ogata and Katsura 1993 method).

| | Fatalities *Boun./Unboun.* | Injuries *Boun./Unboun.* | Economic *Boun./Unboun.* | Magnitude |
|---|---|---|---|---|
| Completeness Lc or Mc | $3.3\ 10^{-5}/3.1\ 10^{-5}$ | $2.3\ 10^{-4}/1.9\ 10^{-4}$ | $2.3\ 10^{-4}/1.7\ 10^{-4}$ | 5.08 |
| b | 0.40/0.39 | 0.36/0.33 | 0.36/0.31 | 0.77 |

The b values lesser than 1 indicate the relative ratio of small and large losses, which confirms the preponderance of small-to-weak cumulative losses in the overall losses. The completeness values for bounded and unbounded synthetic losses (Lc) do not show a significant difference, while the b values changes do. The long return period model values are constrained by a small number of earthquakes due to the relatively short period of the considered catalog. Therefore, their impact on intermediate loss values (i.e., short return period) remains limited and confirms the relevant information provided by this model for weak-to-moderate earthquakes.

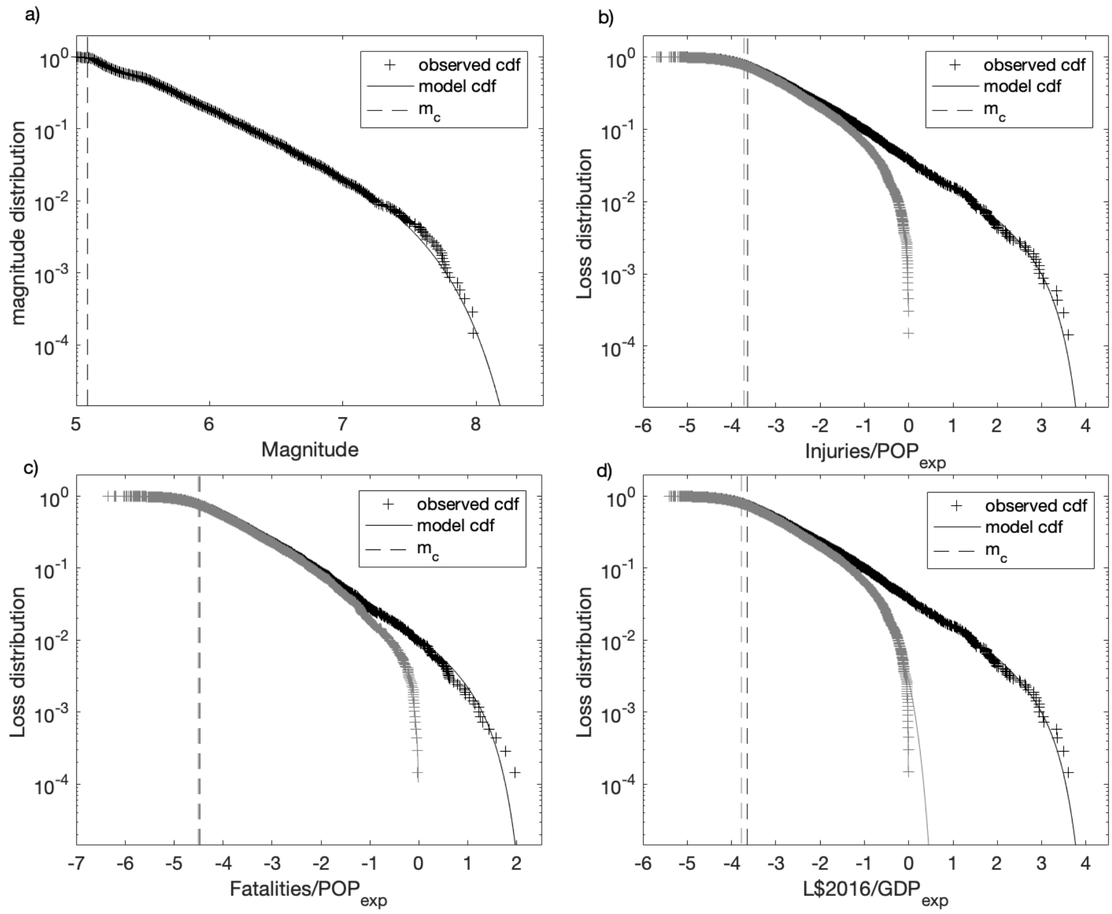


**Figure 12**. Frequency distribution of events (cumulative distribution) estimated following the method proposed by Ogata and Katsura (1993) for (a) the magnitude of the events, (b) the injuries and (c) fatalities normalized by the exposed population, and (d) the economic losses normalized by the exposed GDP along with bounded (gray) and unbounded (black) recurrence laws fit to the synthetics. Vertical dashed lines represent the completeness values.

From Eq. 9, the annual rates of loss exceedance derived from Fig. 12 are given in Fig. 13, considering the weak-to-moderate earthquake (M between 5.5 and 7.0). We assume that event frequency is independent from the date of the most recent earthquake, for which the Poisson model is used. Unlike Holzer and Savage (2013), who use a non-stationary Poisson model for losses to take into account the worldwide population growth, herein, the variables are expressed according to exposure (exposed population and exposed GDP), i.e., assuming a rate independent of global population growth. The probability P of observing at least a given loss value in a period of time t is given by:

$$P = 1 - e^{-\lambda.\Delta t} \qquad (10)$$

The annual rate of losses over the period 1967-2018 is shown in Figure 13 for $F/Pop_{exp}$, $J/Pop_{exp}$ and $L\$/GDP_{exp}$.

For example, from Figure 13 and Eq. 10, the global annual probabilities we obtain are:

P=0.27 for $F/Pop_{exp}$=0.5 i.e., 50% of the exposed population has a probability of 27% to die each year;

P=0.50 for $J/Pop_{exp}$=0.5 i.e., 50% of the exposed population has a probability of 50% to be injured each year;

P=0.99 for $L\$/GDP_{exp}$=0.5 i.e., 50% of the exposed GDP has a probability of 99% to be affected each year.

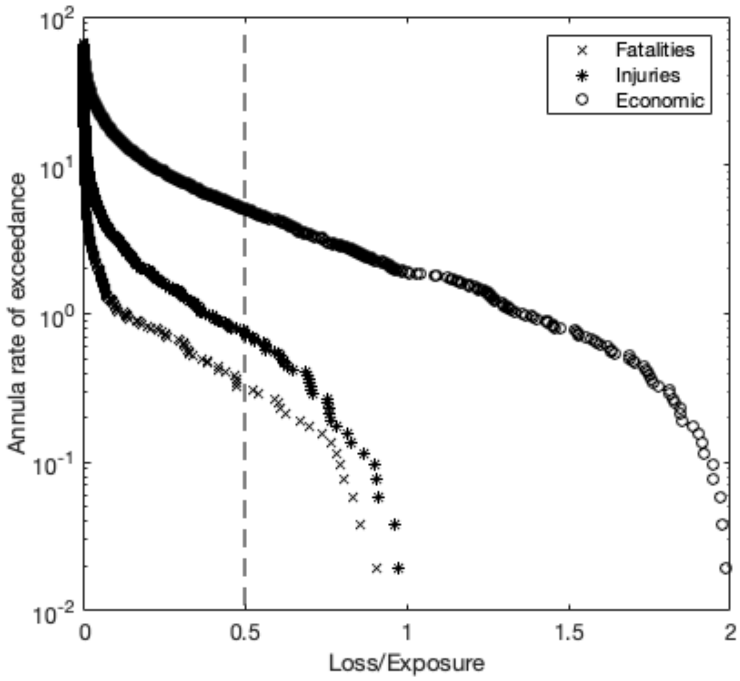


**Figure 13**. Annual rate of excedance for social and economic losses, considering magnitude [5.5;7.0[ earthquakes over the period 1967-2018. Dashed line

corresponds to 50% of exposure.

These results should be taken with caution, as the uncertainties (and their effects) estimated at each stage of the generation of the synthetic loss database have an important impact on the final assessment. Once each earthquake has been defined, intensities are estimated to estimate the level of shaking and its spatial distribution. The spatial cross-correlation of the aleatory (random) variability in the ground motion model should ideally be accounted for, by repeating the scenario event many times, and producing hundreds of possible simulation of the ground motion over the area of interest, and then hundreds of possible losses. Same repeating simulation should be done for each input parameters of the databases and the at each step of the model developed herein. The mean and standard deviation of the loss for the whole exposure model could then be estimated. Nevertheless, this first attempt to build a synthetic seismic losses catalogue as complete as possible at a global scale allows giving a level of risk, in particular for weak-to-moderate earthquakes which remain the least documented events over the period concerned, even though they contribute (in cumulative terms) significantly.

## 6. Conclusions

In this study, data for 7,515 global seismic events were collected to create a synthetic database of losses associated with earthquakes since 1967 with magnitudes between 5.5 and 8, macroseismic intensities above IV, and affecting populations. At each step of the process, prediction models were developed using data for which all the variables were available (concerning the hazard event, exposure and losses) and tested by the LH method (Scherbaum et al., 2004).

Like Dollet and Guéguen (2022), we observed that the strongest (M≥7) (low probabilities-high-consequences) earthquakes make a significant contribution to economic losses (64%), and social losses (44% fatalities and 29% people injured). However, weak-to-moderate earthquakes [5.5; 6.9] also make a sizeable contribution. Compared with high magnitude earthquakes (M≥7), the lesser consequences of these more frequent earthquakes add up to represent a large proportion of social losses (56% of all deaths and 71% of people injured), while the strongest earthquakes make the largest contribution to economic losses (64% of all economic losses). Efforts must be continued to increase the number of post-seismic reports concerning events of moderate magnitude that cause losses and describe parameters related to the hazard event itself, its consequences, and exposure, and detailing more the direct/indirect losses ratio or even cascading effect that may spread outside directly exposed areas.

To calculate the losses, variables related to exposure for each macroseismic intensity produce models with better estimations (less uncertainty). This might be further improved if the post-seismic observations of economic and social losses were recorded by macroseismic intensity and not just overall. The models could also be improved by taking into account variables concerning the seismic vulnerability of buildings, which were not considered in this study. These exposure parameters demand a detailed analysis of the zone considered to assess the vulnerability of both assets and people.
Finally, economic and social loss occurrence models have been produced to enable the estimation of the probability of fatalities and injuries compared with the exposed population and economic losses compared with the exposed GDP. This loss assessment method provides a stationary distribution of the earthquakes causing losses, assuming a homogeneous distribution of exposures per intensity and over time.

With the uncertainty values being evaluated at each step of the process, it will be possible to estimate the errors and test the sensitivity of the results at each step. This missing issue is not considered in this study and must be analyzed separately.

## 7. Funding

This work was supported by the Fondation MAIF (URBASIS-Décision: Analyse multi-critères de la réglementation parasismique applicable aux bâtiments publics. Responsabilité acceptable), the European Union’s H2020 research and innovation programme under the Maria Sklodowska-Curie (URBASIS-EU, grant agreement N°

813137) and funding from Labex OSUG@2020 (Investissements d'avenir, ANR10-LABX56).

## 8. Competing interests

The authors have no relevant financial or non-financial interests to disclose.

## 9. Author contributions

Philippe Guéguen contributed to the study conception and design, analysis of the data and results and commented on previous versions of the manuscript. Material preparation, data collection and analysis were performed by Cyrielle Dollet. The first draft of the manuscript was written by Cyrielle Dollet. Andres Hernandez contributed to the occurrence model section 7. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## 10. References


Allen, T. I., Wald, D. J., Earle, P. S., Marano, K. D., Hotovec, A. J., Lin, K., & Hearne, M. G. (2009). An Atlas of ShakeMaps and population exposure catalog for earthquake loss modeling. *Bulletin of Earthquake Engineering*, *7*(3), 701-718.

Atkinson, G. M., & Wald, D. J. (2007). "Did You Feel It?" intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, *78*(3), 362-368.

Badal, J., & Samardzhieva, E. (2002). Prognostic estimations of casualties caused by strong seismic impacts. *Bulletin of the Seismological Society of America*, *92*(6), 2310-2322.

Badal, J., Vázquez-Prada, M., & González, Á. (2005). Preliminary quantitative assessment of earthquake casualties and damages. *Natural Hazards*, *34*(3), 353-374.

Bakun, W. H., & Scotti, O. (2006). Regional intensity attenuation models for France and the estimation of magnitude and location of historical earthquakes. *Geophysical Journal International*, *164*(3), 596-610.

Bindi, D., Parolai, S., Oth, A., Abdrakhmatov, K., Muraliev, A., & Zschau, J. (2011). Intensity prediction equations for Central Asia. *Geophysical Journal International*, *187*(1), 327-337.

Cha, L. S. (1998). Assessment of global seismic loss based on macroeconomic indicators. *Natural Hazards*, *17*(3), 269-283.

Christoskov, L., & Samardjieva, E. (1984). An approach for estimation of the possible number of casualties during strong earthquakes. *Bulg Geophys J*, *4*, 94-106.

Coburn, A., & Spence, R. (2003). *Earthquake protection*. John Wiley & Sons.

Daniell, J.E., Schaefer, A.M., & Wenzel, F. (2017). Losses Associated with Secondary Effects in Earthquakes. *Front. Built Environ*. https://doi.org/10.3389/fbuil.2017.00030

Desinventar (2018) Sendai Framework for disaster risk reduction database. http://www.desinventar.net/DesInventar/results.jsp. Last access: June 2018

Di Giacomo, D., Engdahl, E. R., & Storchak, D. A. (2018). The ISC-GEM earthquake catalogue (1904–2014): status after the extension project. *Earth System Science Data*, *10*(4), 1877-1899.

Dollet, C., & Guéguen, P. (2022). Global occurrence models for human and economic losses due to earthquakes (1967–2018) considering exposed GDP and population. *Natural Hazards*, *110*(1), 349-372.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews*, *61*(1-2), 43-104.

EM-DAT (2022) EM-DAT: International disaster database. Université Catholique de Louvain, Belgium. http://www.emdat.be. Last Access : June 2018

Eurostat (2018). https://ec.europa.eu/eurostat/fr/data/browse-statistics-by-theme. Last Access: June 2018

Guettiche, A., Guéguen, P., & Mimoune, M. (2017). Economic and human loss empirical models for earthquakes in the mediterranean region, with particular focus on Algeria. *International Journal of Disaster Risk Science*, *8*(4), 415-434.

Heatwole, N., & Rose, A. (2013). A reduced-form rapid economic consequence estimating model: Application to property damage from US earthquakes. *International Journal of Disaster Risk Science*, *4*(1), 20-32.

Holzer, T. L., & Savage, J. C. (2013). Global earthquake fatalities and population. *Earthquake Spectra*, *29*(1), 155-175.

ISC-GEM Global Instrumental Earthquake Catalogue (2019) - Version 6.0, 7 March 2019 - http://doi.org/10.31905/D808B825

Jaiswal, K., & Wald, D. (2010). An empirical model for global earthquake fatality estimation. *Earthquake Spectra*, *26*(4), 1017-1037.

Jaiswal, K., & Wald, D. J. (2013). Estimating economic losses from earthquakes using an empirical approach. *Earthquake Spectra*, *29*(1), 309-324.

Levret, A., Backe, J. C., & Cushing, M. (1994). Atlas of macroseismic maps for French earthquakes with their principal characteristics. *Natural hazards*, *10*(1), 19-46.

Musson, R. M., Grünthal, G., & Stucchi, M. (2010). The comparison of macroseismic intensity scales. *Journal of Seismology*, *14*(2), 413-428.

Nichols, J. M., & Beavers, J. E. (2008). World earthquake fatalities from the past: implications for the present and future. *Natural Hazards Review*, *9*(4), 179.

Nievas, C. I., Bommer, J. J., Crowley, H., & van Elk, J. (2020a). Global occurrence and impact of small-to-medium magnitude earthquakes: a statistical analysis. *Bulletin of Earthquake Engineering*, *18*(1), 1-35.

Nievas, C. I., Bommer, J. J., Crowley, H., van Elk, J., Ntinalexis, M., & Sangirardi, M. (2020b). A database of damaging small-to-medium magnitude earthquakes. *Journal of Seismology*, *24*(2), 263-292.

National Geophysical Data Center / World Data Service (NGDC/WDS): NCEI/WDS Global Significant Earthquake Database. NOAA National Centers for Environmental Information. doi:10.7289/V5TD9V7K [Last access : June 2022]

Ogata, Y., & Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, *113*(3), 727-738.

Riedel, I., Gueguen, P., Dunand, F., & Cottaz, S. (2014). Macroscale vulnerability assessment of cities using association rule learning. *Seismological Research Letters*, *85*(2), 295-305.

Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., & Chanussot, J. (2015). Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine methods. *Natural hazards*, *76*(2), 1111-1141.

Scherbaum, F., Cotton, F., & Smit, P. (2004). On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion. *Bulletin of the seismological society of America*, *94*(6), 2164-2185.

Spence, R., So, E., Jenkins, S., Coburn, A., & Ruffle, S. (2011a). A global earthquake building damage and casualty database. In *Human Casualties in Earthquakes* (pp. 65-79). Springer, Dordrecht.

Schumacher, I., & Strobl, E. (2011). Economic development and losses due to natural disasters: The role of hazard exposure. *Ecological Economics*, *72*, 97-105.

UN (2019) United Nations DESA / Population Division. https://population.un.org/wpp/. Last access: Jul 2019

Wald, D. J., Quitoriano, V., Heaton, T. H., Kanamori, H., Scrivner, C. W., & Worden, C. B. (1999). TriNet "ShakeMaps": Rapid generation of peak ground motion and intensity maps for earthquakes in southern California. *Earthquake Spectra*, *15*(3), 537-555.

World Bank (2019) The World Bank IBRD/IDA. https://data.worldbank.org/indicator/sp.pop.totl. Last access: Jul 2019

Wyss, M., & Trendafiloski, G. (2011). Trends in the casualty ratio of injured to fatalities in earthquakes. In *Human casualties in earthquakes* (pp. 267-274). Springer, Dordrecht.