

Macro-scale vulnerability assessment of cities using Association Rule Learning

Ismael Riedel, Philippe Gueguen, François Dunand, Stéphanie Cottaz

► **To cite this version:**

Ismael Riedel, Philippe Gueguen, François Dunand, Stéphanie Cottaz. Macro-scale vulnerability assessment of cities using Association Rule Learning. *Seismological Research Letters*, Seismological Society of America, 2014, 85 (2), pp.295-305. <10.1785/0220130148>. <hal-01015723>

HAL Id: hal-01015723

<https://hal.archives-ouvertes.fr/hal-01015723>

Submitted on 27 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Macro-scale vulnerability assessment of cities using Association Rule Learning

I. Riedel¹, P. Guéguen¹, F. Dunand², S. Cottaz¹

1. ISTERre. CNRS/IFSTTAR. Université Joseph Fourier Grenoble I. France

2. Geoter SA. France

Seismological Research Letters, 85(2): 295-305. [10.1785/0220130148](https://doi.org/10.1785/0220130148)

Corresponding author:

Philippe Gueguen

ISTerre

BP 53

38041 Grenoble cedex 9

France

email: philippe.gueguen@ujf-grenoble.fr

1. Introduction

After most moderate-to-strong earthquakes causing considerable damage (e.g. recent earthquakes such as those of Boumerdes in 2003, Bam in 2003, L'Aquila in 2009, Haïti in 2010, etc.), the losses observed remind local authorities and decision-makers that reducing seismic risk is essential for the well-being and safety of local populations as well as for economic and social stability. The anticipation and simulation of the consequences of an earthquake scenario require knowledge of the probabilistic seismic hazard as well as a representation of the capacity of structures to support the seismic ground motion: this is the objective of seismic vulnerability assessments. Such assessments allow the estimation of probable damage at a large overall scale (country, region, town) and (1) give information on the most vulnerable building categories that must benefit from priority reinforcement; (2) inform local authorities on the level of risk to which the population is exposed compared with other natural or domestic hazards which are more frequent and therefore more easily appreciated by the population, particularly in moderate seismic hazard countries (Dunand and Gueguen, 2012; Lestuzzi et al., 2009); and finally (3) allow anticipation of the actions and reactions for crisis management by local or wider communities (Jaiswal et al., 2010). Coupled with real-time seismic ground motion estimates (e.g., Wald et al., 1999; Worden et al., 2010), macro-scale vulnerability data is crucial information for the early assessment of damage, as proposed for specific facilities by Wald et al. (2008).

Old structures, designed before the application of earthquake design rules, and present everywhere, are certainly a critical element affecting the extent of loss and fatalities. Many empirical methods

for vulnerability assessment have been published, most of them calibrated on post-earthquake observations (e.g., GNDT, 1993; Hazus, 1997; Spence and Lebrun, 2006). They give the probability of reaching a certain level of damage for a given class of buildings and a given seismic demand. Recent initiatives as part of the Global Earthquake Model project (GEM, <http://www.globalquakemodel.org>) also attempt to provide worldwide vulnerability assessment. This macro model aims to propose and develop global procedures for deriving vulnerability functions for a wide variety of building types defined according to their structural characteristics referenced by the World Housing Encyclopaedia (<http://www.world-housing.net>) as the building taxonomy (Porter et al., 2001).

The key issues and drawbacks of these methods and initiatives are the variability of the response of existing buildings to earthquakes, the inadequate information concerning the seismic behaviour of old design buildings as well as the quality of construction materials and the lack of observation data for adjusting empirical methods to the highest damage grade. All these issues introduce a large epistemic uncertainty to seismic vulnerability assessment and seismic risk analysis. Recent initiatives have been proposed to complete the lack of empirical observations for the highest damage level based on hybrid methods (Kappos et al., 2006) or to reduce epistemic uncertainties related to knowledge of building model using operative modal analysis (Michel et al., 2012). Nevertheless, these issues remain critical for moderate seismic prone regions because the mobilization of resources for seismic evaluation is limited, even though the seismic hazard is not negligible. For example, France is rightly regarded as a country with moderate hazard. However, several destructive historical earthquakes have occurred in the past, the most recent (1909 Lambesc earthquake) killing 42 people and causing thousand millions euros of losses in the southeastern region of France. Even moderate earthquakes in the Pyrenees (Arudy Ossau-1980, $M_L = 5.1$) and the Alps (Annecy 1996, $M_L = 4.8$) have caused several million euros of damage as observed in other seismic prone regions (e.g., Pierre and Montagne, 2004). In this context, vulnerability

assessment studies have been conducted in France, focused on large exposed cities and applying traditional empirical methods. However, the application of empirical methods requires so much information that the evaluation struggles to find sufficient political motivation and financial resources for accomplishing the seismic inventory of buildings. Consequently, the structural characteristics required for the seismic vulnerability assessment of existing buildings are not available for all exposed urban areas of the country, even though seismic exposure is higher than in the past and a repetition of historic earthquakes may provide more casualties and economic losses due to the urbanization growth over the past century, and as suggested by Jackson (2006) for the worldwide situation.

To overcome the lack of building information at the macro scale, one solution would be to assess vulnerability not considering the information required but the information already available in a region or country. In this paper, a datamining method based on Association Rule Learning (ARL) is applied to define a vulnerability proxy between the elementary characteristics of buildings and the vulnerability classes of the European Macroseismic Scale EMS98 (Grunthal, 1998). The method was applied to the Grenoble city test-bed described in the first part of this paper. The ARL method is then presented and a vulnerability proxy was derived for a Grenoble city-like environment. The vulnerability proxy is tested in Nice in the third part, a city that has been the subject of a vulnerability study (Spence and Lebrun, 2006). Finally, the damage produced by historic earthquakes was computed, considering the (equivalent) earthquake-era and the present-day urbanization for simulating seismic damage.

2. The Grenoble test-bed area

For the VULNERALP project (Gueguen et al., 2007), a simplified empirical method was proposed and tested in Grenoble, one of France's most exposed cities to earthquakes. By sending experts into the field, basic information was collected, assigning elementary structural characteristics to existing buildings and finally constituting the Grenoble building vulnerability database. The main pieces of information were [date of construction] ranked by period, [number of floors] ranked by categories and [roof shape] (flat or slope). In addition, experts associated a type of building according to the EMS98 typology with the most likely vulnerability class. Building vulnerability was established as belonging to a category of buildings (the typology EMS98) with 6 classes from A (the most vulnerable) to F (the less vulnerable). 3860 buildings were characterized, with elementary attributes considered certain, and classified according to the EMS98 class of vulnerability. Some attributes, e.g. [number of storeys] and [period of construction] are available in the INSEE (National Institute for Statistics and Economic Studies, <http://www.insee.fr>) database, grouped by geolocalized cells called IRIS2000. These units were defined in 1999 for the population census. The name IRIS2000 (IRIS in the rest of this document) alludes both to the year of establishment and the size, corresponding to 2,000 inhabitants and representing the unit for geographical data distribution. Municipalities of at least 10,000 inhabitants and most municipalities of between 5,000 and 10,000 inhabitants are divided into IRIS. By extension, municipalities with less than 2,000 inhabitants are considered as an IRIS unit. France has 50,100 IRIS units plus 700 in the overseas region.

Buildings per IRIS are grouped by attributes: the number of storeys divided into three categories (<4, [4-8] and >8 floors) and the period of construction, according to social and urban evolutions in France (<1915, [1915-1948], [1949-1967], [1967-1974], [1975-1982], [1983-1989] and >1990). In Grenoble, the INSEE database contains 6,214 buildings: 3,664 buildings with fewer than four floors; 2,020 buildings with between 4 and 8 floors; 530 buildings with more than 9 floors. The distribution per period of construction is as follows: 1,543 buildings constructed before 1915; 1,180 buildings between 1915 and 1948; 2,034 buildings between 1949 and 1967; 525 buildings between

1968 and 1974; 263 buildings between 1975 and 1982; 335 buildings between 1983 and 1989; 334 buildings after 1990. Figure 1 shows the geolocalization of the buildings by IRIS as function of the period of construction (Fig. 1A) and the number of floors (Fig. 1B). As already reported by Gueguen et al. (2007), the characteristics of Grenoble's urbanization observed using INSEE data are partly controlled by the population's growth (Parent, 1982), and its history.

3. Association rule learning

Datamining is the process that attempts to discover patterns in large data sets. It utilizes methods of artificial intelligence, machine learning, statistics, and database systems. The overall goal of datamining techniques is to extract information from a data set and transform it into an understandable structure for further use. Agrawal et al. (1993) introduced Association Rules Learning (ARL) as *if/then* statements that help reveal relationships between seemingly unrelated data in a relational database or other information repository. By an association rule, we mean an implication of the form $Y \implies X_i$, where Y is a set of items that belongs to a database of attributes and X_i is a single item in the database and not present in Y . Each relationship between Y and X_i is represented as a binary vector, equal to 0 if Y and X_i are not related and to 1 otherwise. The ARL method defines the relationships (or proxy) between Y and X_i once a learning phase has been completed on a subset of the database.

In order to develop a vulnerability proxy, we tested the simplified ARL method using Grenoble's database of buildings. Structural information (attributes Y) and EMS98 vulnerability classes (item X_i) allow definition of a conditional matrix between X_i and Y (the learning phase). The conditional probability of having a class $X=\{A, B, C, D, E\}$ knowing that an event Y has a non-zero probability (or the probability of X_i , given Y) is the number denoted by $P(X | Y)$ and defined by:

$$P(X|Y) = P(X \cap Y)/P(Y) \quad (1)$$

X and Y are not totally independent because according to almost all empirical methods, vulnerability also depends on elementary structural characteristics. Knowing that we randomly selected one building in the city for the variable "a building belonging to class X", the probability P (X | Y) of experiencing an attribute Y is:

$$P(X|Y) = \frac{N_{XY}}{N_Y} \quad (2)$$

where N_{XY} is the number of buildings belonging to class X with attribute Y, and N_Y the total number of buildings with attribute Y. In Grenoble, the data set used for the learning sample does not include EMS98 class E or F. Moreover, no information is available on the quality of the expert survey. Therefore, the most likely class of vulnerability was considered herein; further analysis will be carried out in the future, based on more enhanced methods, integrating the most probable EMS98 vulnerability interval.

As often in practice, the number of buildings is not enough to define the most efficient ARL possible. To ensure that our training sample was sufficient, we compared the distribution of the structural parameters used for the learning phase and given by the VULNERALP expert survey, with the INSEE database. We focused on the two attributes available in INSEE, with the objective of extending the association to the whole French territory. Nevertheless, the shape of the roof (flat or slope) can also be considered for learning, with the perspective of defining a [roof shape] attribute Y and the associated vulnerability class X_i through remote sensing. At the beginning of the VULNERALP project, [period of construction] attribute ranks were defined according to the historical evolution of the urbanization and development of the construction code, while the

[number of floors] attribute was defined according to the traditional interval given by the GNDT method. Therefore, we considered the ground truth (i.e., the VULNERALP database) as the reference model for which the EMS98 classification was available. We then re-defined the building distribution in each INSEE interval for the [period of construction] and [number of floors] attributes, according to the VULNERALP intervals, considering a constant annual rate of urbanization in each INSEE period, together with an uniform distribution of buildings per number of floors. Table 1 shows that buildings randomly surveyed in the city of Grenoble for the VULNERALP project are representative of the area's urbanization as a whole, their distributions being quite similar (maximum difference observed 10%).

The vulnerability class X_i was then associated to attributes using the ARL methods, knowing Y attributes and using Eq. 2. Considering each attribute, a conditional probability relationship was obtained and used as a vulnerability proxy. For example, the probability $P(B|Y_1)$ that buildings are in EMS98 class B, knowing that Y_1 is [building constructed before 1945] is 0.838. Table 2 summarizes the Grenoble vulnerability matrix (GVM) of each conditional probability of being in EMS98 class X , knowing information related to Y , and called the GVM proxy hereafter. We note that the majority of buildings constructed before 1945 are in class B, while those in the period [1945 - 2000] are primarily Class C. For buildings built after 2000, a similar distribution between Classes C and D is observed. Furthermore, Y_i items can be crossed to introduce additional building characteristics for refining probability $P(X | Y)$. All possible combinations using the 3 basic parameters (i.e., number of storeys, period of construction and roof shape) were taken into account but only the two parameters found in INSEE Database were considered for this paper (Tab. 2).

Finally, having accomplished the learning phase giving the GVM proxy, the second phase can be implemented to obtain the geolocalized distribution of classes X_i in each IRIS, knowing Y for the whole French territory (Eq. 2). For example, since the INSEE database provides the numbers of

buildings with fewer than 3 floors (N_1 , attribute Y_1), [3-5] floors (N_2 , attribute Y_2) and more than 5 floors (N_3 , attribute Y_3), the probability $P(A)$ in each IRIS is given by the formula:

$$P(A) = (N_1 P(A|Y_1) + N_2 P(A|Y_2) + N_3 P(A|Y_3)) / (N_1 + N_2 + N_3) \quad (3)$$

The twelve combinations crossing [number of floors] and [period of construction] (Tab. 2), used for the learning phase according to the VULNERALP database and applied to the INSEE intervals for the application phase, are given by the following formula:

$$P_j(X) = \frac{\sum_i^{12} N_{ji} P(X|Y_i)}{N} \quad (4)$$

with $P_j(X)$ as the probability of having vulnerability class $X_i = (A, B, C, D)$ in each j IRIS cell, N_{ji} the number of buildings with attribute Y_i within j , N the total number of buildings in IRIS j and $P(X_i|Y)$ the value of the probability given by the GVM proxy for the $X \Rightarrow Y_i$ association (Tab. 2).

Table 3 compares the vulnerability class distribution in Grenoble using VULNERALP data and the GVM proxy applied to the INSEE database. Similar results were obtained. For example, the probability of having a building in Grenoble in class B is 0.428 and 0.423 for INSEE and VULNERALP datasets, respectively. Figure 2 shows the vulnerability classes in Grenoble computed using the GVM proxy, considering [number of floors] and [period of construction]. The main trends as those reported by Gueguen et al. (2007) and Michel et al. (2012) are also observed in Fig. 2: the highest vulnerability in the historic down-town area, the lowest at the periphery (particularly towards the south) and heterogeneous intermediate districts covering all periods of urbanization and mixing masonry and reinforced concrete buildings.

The GVM proxy established during the learning phase provides the distribution of the number of buildings by vulnerability class for a given geographic area (IRIS). This relationship can give

vulnerability class distribution in each IRIS for the whole of France, based only on national census (INSEE) or/and remote sensing information (introducing the [shape of the roof] attribute to the analysis), and above all assuming Grenoble city-like urbanization. This assumption will be tested in the following sections.

4. Validation of the GVM proxy method in the city of Nice

The city of Nice, another of France's most exposed cities to seismic risk, has undergone numerous vulnerability evaluations (e.g., Bard et al., 2005; Spence and Lebrun, 2006). In order to validate the GVM proxy, seismic damage in Nice was predicted using the GVM proxy applied to INSEE data and with the vulnerability indexes obtained by the RiskUE method.

The RiskUE analysis (Lagomarsino and Giovinazzi, 2006) assigns a vulnerability index (I_v). I_v is defined by the structural characteristics of the buildings, adjusted according to damage observed during earthquakes in Italy. The damage for a given seismic demand expressed as macroseismic intensity (I_{EMS98}) is given by the formula:

$$\mu D = 2.5 \left[1 + \tanh \frac{I_{EMS98} + 6.25 I_v - 13.1}{2.3} \right] \quad (5)$$

where μD is the average observed damage in buildings. EMS98 characterizes damage according to 6 levels (D_k with $k = 0,1,2,3,4,5$), ranging from D_0 (no damage) to D_5 (complete destruction). To take into account the variability of the damage grade level k in a set of buildings, Lagomarsino and Giovinazzi (2006) assume a binomial distribution of damage level. Therefore, the probability $P(D_k)$ of observing each damage level D_k ($k = 0$ to 5) for a given damage μD is evaluated according to the probability function of the binomial distribution, namely:

$$P(D_k) = \frac{5!}{k!(5-k)!} \left(\frac{\mu D}{5}\right)^k \left(1 - \frac{\mu D}{5}\right)^{5-k} \quad (6)$$

This gives the distribution of buildings for each level of damage considering a set of buildings, or the probability of observing each level of damage for a given building. In Nice, the RiskUE project identified 27 zones (Z_{RiskUE}) considered homogeneous for vulnerability assessment (Fig. 3A). A random sample of buildings was selected to assess the vulnerability of each zone, with I_v between 0.365 and 0.849. Every Z_{RiskUE} was then geolocalized and characterized by a surface area (square-metre) and an average vulnerability I_v .

The EMS98 vulnerability classes spatially distributed and deduced from the GVM proxy are given in Fig. 3B. Most of the buildings are in Classes B and C (96%) and 4% are in classes A and D. The traditional trends observed in European urban centers are also observed in Nice, namely the more vulnerable historic down-town area with 60-80% in Class B, more modern and less vulnerable suburb areas (mostly class C). As INSEE and RiskUE zoning rules are not the same, data was fused by reducing the information to a common reference to enable comparison (i.e., data alignment process). Thus, the city was divided into elementary zones i as small as possible, so that $Z_i = Z_{RiskUE} \cap Z_{IRIS}$ ($i = 1:457$) ultimately corresponding to 457 elementary zones. For each Z_i , RiskUE vulnerabilities (I_v) were assigned assuming homogeneous distribution of vulnerability, i.e.:

$$I_{v_i} = I_{v_{RiskUE}}; Z_i \in Z_{RiskUE} \quad (7)$$

Meanwhile, the distribution of building classes in the area Z_i belonging to Z_{IRIS} was calculated in proportion to the surface ratio between Z_i and Z_{IRIS} , i.e.:

$$P(X|Y)_i = P(X|Y)_{IRIS} \frac{S_i}{S_{IRIS}}; Z_i \in Z_{IRIS} \quad (8)$$

The data alignment process resulted in some vulnerability information being different between RiskUE and the GVM proxy. This is the case of Z_{RiskUE} #9, for example, which is characterized by

low vulnerability since it covers recent urbanization while the Z_{IRIS} crossing this zone covers more or less all of the historic down-town area with high vulnerability. For this reason, the comparison between RiskUE and the GVM proxy must be made at a macro scale rather than zone by zone.

The EMS98 intensity scale contains 12 degrees, i.e. I (not felt) to XII (general catastrophe), based on the perception of the earthquake and the degree of damage observed, taking into account the class of the buildings. For example, damage to buildings appears from intensity V, with D_1 damage grade affecting some buildings of classes A and B (Grunthal, 1998). The number is defined by a literal qualification, associated with a damage percentage as proposed by Lagomarsino and Giovinazzi (2006): "some" (5%), "many" (35%) and "most" (80%). On this basis, damage matrices are established giving the occurrence probability distribution $P(D=D_k)$ for each intensity as a function of building vulnerability (Fig. 4). All the buildings must be described according to D_k , i.e. $\sum P(D=D_k) = 1$ ($k=0.5$), so we assume the damage matrices can be extended to replace the missing information in the description of the intensities, according to the description of damage in EMS98, i.e. a monotonically decreasing function at a higher damage level D_k and corresponding to the lowest intensity, a normal distribution of probabilities around the mean damage grade for an intermediate level of damage, and a monotonically increasing function at a low damage level D_k , typical of higher intensities. For example, for buildings in Class A and intensity VII, EMS98 mentions that "many (35%) buildings in vulnerability class A suffer damage of grade 3, few (5%) grade 4." The remaining 60% are distributed over the lower levels of damage to propose a continuous, smoothed probability function of damage, as shown in Fig. 4.

Figure 5 shows the probability of occurrence of damage D_k for intensities V to X, computed with RiskUE and the GVM proxy, for each Z_i and averaged for the scale of the city. Unlike the RiskUE method, which gives a non-zero probability of damage, the probability for the highest D_k and lowest intensity are associated with zero due to discrete representation of the damage grade in the

EMS98 method. Slight differences are observed at the macro scale, except for the lowest intensities. Nevertheless, for the intensities causing damage, the orders of magnitude of the damage occurrence probability are quite similar. Although the GVM proxy was defined for a Grenoble-like environment, the damage prediction provides reliable information at the first order and allows generalization to the whole of the French territory, ultimately producing a geo-localized assessment of vulnerability.

5. A new look at France's historic earthquakes and predicting their consequences in 2013.

In 1909, the historic Lambesc earthquake, southeastern France, was probably the strongest earthquake in recent history. This earthquake produced macroseismic intensities of VIII in the epicentral area (Fig. 6), 30km from Marseille, France's second city in terms of population and economic activity. Its magnitude was recently re-appraised and estimated at around 6.0 (Baroux et al., 2004). Because of its shallow depth (less than 10km), it was felt more than 300km from the epicenter, affecting a rural area. 46 casualties were reported after the event and referenced in the SisFrance database (Scotti et al., 2004), occurring in different towns located less than 50km from the epicenter, and about 250 injured persons were reported. In term of loss, damage to the buildings was serious, as summarized by Lambert (1997), in different cities within the region affected. Although there is no detailed quantitative information in the historic description of damage, this information can be used as ground truth for comparing predictions and observations. This earthquake is all the more important since it served as a scenario in 1982 to forecast seismic losses and casualties taking into account urbanisation evolutions between 1909 and 1982, (MED82). The results provided information that increased the awareness of the authorities, an element (among others) that led to the establishment of the modern national earthquake rules for construction design, published 15 years later.

In our analysis, we consider a square surface including all sectors with a macroseismic intensity above V and some with an intensity between III and IV (Fig. 6). In total, the studied area represents 4254 IRIS covering a large part of southeastern France. Since the INSEE database does not give information on the inventory of past – and now inexistent – buildings, we assume that the number of buildings per attribute corresponds to the buildings that were present in each period. We thus assume a slow rate of replacement and are able to provide an approximate simulation of the damage produced by the Lambesc earthquake in 1909, considering buildings before 1915 and existing in 2008 as those present in 1909. The temporal evolution of seismic vulnerability was thus assessed for different periods of construction (before 1945, before 1970, before 2000 and before 2008) in order to visualize the effects of the rate of urbanization on seismic vulnerability (Fig. 7). Before 1945, the majority of buildings belong to class B (e.g., $P(B|before\ 1945) = 0.847$). The probabilities for high vulnerability classes are reduced with time, class A and B changing from 0.072 to 0.026 and 0.847 to 0.474, respectively, between 1945 and 2008. On the contrary, for the same two periods, $P(C)$ and $P(D)$ change from 0.053 to 0.460 and 0 to 0.039, respectively, reflecting the construction of new buildings in the studied area, these buildings being less vulnerable. The evolution of the number of buildings for all IRIS is significant, with more than 130,000 new constructions between 1945 and 2008, which is also coherent with the urbanization rate observed in Grenoble and reflecting the post World War II needs for housing. Furthermore, new buildings are less vulnerable, thanks to the use of reinforced concrete rather than masonry and the application of new building codes after the 1970s.

Although the probability of the same earthquake re-occurring is extremely low in a diffuse seismicity region, this technique eliminates the difficulties of simulating ground motion using GMICEs or GMPEs, including site and source effects, by considering the 1909 macroseismic intensity curves (Fig. 6) as representing the seismic demand. Damage probability in each grade is

computed with Eqs. 3 to 6. by crossing the GVM proxy applied to the INSEE attributes and using the 1909 macroseismic intensity curves as seismic demand. The number of buildings in each damage grade is displayed on Fig. 8, grouped into three classes according to the EMS98 scale: no or slight damage (D0+D1), moderate damage (D2+D3) and heavy damage (D4+D5). However, to compare with historical description available in villages or cities, damage is computed separating slight damage (D1+D2) and strong damage (D3+D4+D5). Figure 8A represents the number of buildings in each class of damage for the 1909 earthquake affecting dwellings built before 1915. The highest damage computed is localized close to the epicenter, 150 buildings would suffer heavy damage, while 3,890 are estimated as suffering moderate damage, the rest being distributed over the studied area. The historic information shows a reliable estimate of the consequences of the damage: for example, considering the highest damage level (D3+D4+D5), we obtained 1,110 buildings while historic information indicates several thousand. The synthesis of the comparison is given in the following:

Lambesc ($R_{epic}=5\text{km}$)

Observed: 600 damaged buildings. 50 destroyed;

GVM proxy: 257 (D1+D2) and 195 (D3+D4+D5) buildings;

Rognes ($R_{epic}=3\text{km}$)

Observed: 250 damaged buildings;

GVM proxy: 124 (D1+D2) and 95 (D3+D4+D5) buildings;

Saint-Canat ($R_{epic}=4\text{km}$)

Observed: 310 damaged buildings. 50 being heavily;

GVM proxy: 105 (D1+D2) and 80 (D3+D4+D5) buildings;

La Roque d'Anthéon ($R_{epic}=7\text{km}$)

Observed: 110 heavily;

GVM proxy: 142 (D1+D2) and 16 (D3+D4+D5) buildings;

Aix-en-Provence ($R_{\text{epic}}=20\text{km}$)

Observed: 1500 damaged buildings;

GVM proxy: 1915 (D1+D2) damaged buildings.

Slight differences exist, which may reflect the iso-intensity curves considered as seismic ground motion, without consideration of local site effects observed in the zone, and the differences between the 2008 inventory of buildings built before 1915 and the state of urbanization in 1909. Nevertheless, we can assume that the damage obtained by this method is relevant in terms of damage estimation at the macro scale; the lack of more accurate descriptions of historical damage and information on urbanization at the time prevents better comparison.

In 1982, the results of a simulation of the Lambesc earthquake performed by the ministry in charge of natural hazard indicated 1.8% of buildings collapsed (D5), 85.8% of buildings damaged (D1 to D4) and 12.4% with no damage (D0). Keeping the same assumptions as for the 1909 estimate (i.e. 2008 inventory of building before 1982 and 1909 iso-intensities), the GVM proxy applied to buildings built before 1982 gives the same order of magnitude for the damage distribution, i.e. 0.8, 81.5 and 17.7% of buildings collapsed, damaged and with no damage, respectively.

Finally, the simulation can be continued by forecasting the impact of a future earthquake having the same characteristics as the 1909 Lambesc earthquake (i.e., same location and same macroseismic intensities) on the state of urbanisation in 2008. In 2008, the area had more than 1.77 million buildings and a population of more than seven million, 94% of buildings had between 0 and 2 floors, and 50% were built between 1945 and 2000. Less than 2% of the buildings had more than 5 floors and approximately 4% had between 3 and 5 floors. The probability distribution of the vulnerability classes B and C is high: $P(B) = 0.474$ and $P(C) = 0.460$, while the probability of being

in class A or class D is relatively low ($P(A)=0.026$ and $P(D)=0.039$). If the 1909 earthquake re-occurred in 2008, nearly 96% of the buildings present in the studied area would not be damaged while about 66,000 buildings (4%) would be affected with different levels of severity. The epicentral area (intensity VII and VIII) includes more than 31,600 damaged buildings, i.e. representing 61% of the buildings (Fig. 8): 370 buildings show strong levels of damage (D4 and D5), only a dozen buildings are completely destroyed (D5) and more than 15,000 buildings suffer moderate damage (D2+D3). Largest damage are expected in main cities: 82% of the buildings in Lambesc, 62% suffering minor damage and 20% suffering extensive structural damage, Aix-en-Provence (15km) with nearly 30% of the city damaged, 8% with damage greater than D3, Marseille, Nimes or Avignon with between 2 and 4% of slightly damaged buildings (D1+D2).

6. Conclusions

The aim of this paper is to validate a macro-scale methodology for seismic vulnerability assessment, in a situation where only a poor description of construction characteristics is available for a large number of the buildings studied. In order to eliminate these difficulties, the basic idea of our approach is to extract information on the vulnerability of urbanization from elementary attributes of buildings. In a moderate seismic prone region, where it is often difficult to mobilize resources for the reduction of seismic risk, the idea of using readily-available data to expand the assessment to any given region is obviously of interest. Using the information available in Grenoble, we proposed a vulnerability proxy (GVM proxy) defined using the Association Rule Learning method. This proxy was first defined (learning phase) using the Grenoble database with the elementary attributes describing the buildings as listed in the French national census. It was then validated in Nice and finally tested for the historic Lambesc earthquake which struck south-eastern

France in 1909. Although the attributes describing the buildings are very basic, the analyses for Nice and Lambesc provided results that confirm the pertinence of our solution.

The state of urbanization in Grenoble may be different from that of other cities in France, however, our investigation shows that when applied, the proxy provides reliable estimates of damage for earthquake scenarios. In Nice, a more sophisticated method (RiskUE), based on a relatively detailed description of structural features, produced similar levels of damage across the city, macroseismic intensity being considered as the seismic demand parameter. In Lambesc, simulation of the 1909 historic earthquake and comparison with a previous simulation conducted in 1982, show that the damage estimates using the vulnerability proxy are relevant. Because of the lack of elements of comparison and the lack of details about historic damage, it is difficult to quantify the assessment errors that might be obtained for a given earthquake. However, the datamining method ARL appears to be well suited to the large-scale assessment of seismic vulnerability and thus to the simulation of seismic damage. We can thus highlight certain obvious trends, such as the reduction in the proportion of vulnerable buildings with the development of urbanization. We can also confirm the increasing effects of earthquakes in terms of damage, mainly due to the explosion of urbanization and urban concentrations in certain areas prone to seismic hazard, For example, in the Lambesc region, if the 1909 earthquake had occurred in 2013, there would be serious consequences in terms of casualties and economic losses: 370 buildings would suffer strong levels of damage (D4 and D5), a dozen buildings would be completely destroyed (D5) and more than 15,000 buildings affected by moderate damage (D2+D3). All things considered, we observe damage comparable to those observed during the earthquakes of similar magnitude in L'Aquila in Italy or Christchurch, New-Zealand.

While the application of this proxy to other regions outside France remains to be tested, a machine or proxy may be developed to estimate damage in a broader region. These are essential elements to enable the evaluation of economic and human losses. Once the distribution of vulnerability classes

is known, the consequences in terms of damage can be simulated rapidly after an earthquake, providing an additional element to the simulation of ground motion via Shakemap for a seismic warning system.

Acknowledgement

INSEE data were prepared and provided by the Centre Maurice Halbwachs (CMH). This work was supported by the French Research National Agency (ANR) under the RiskNat program (project URBASIS n°ANR-09- RISK-009) and by the Fondation MAIF.

References

- Agrawal, R., T. Imieliński, and A. Swami (1993). Mining association rules between sets of items in large databases, in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*. pp. 207. doi:10.1145/170035.17007.
- Bard, P.-Y., A.-M. Duval, E. Bertrand, J.-F. Vassiliadès, S. Vidal, C. Thibault, B. Guyet, J.-P. Mèneroud, P. Gueguen, P. Foin, F. Dunand, S. Bonnefoy-Claudet, and G. Vettori (2005). *Le risque Sismique à Nice: apport méthodologique, résultats et perspectives opérationnelles*. Final technical report of the GEMGEP project, CETE-Méditerranée, 52 pages
- Baroux, E., N.A. Pino, G. Valensise, O. Scotti, and M.E. Cushing (2003). Source parameters of the 11 June 1909, Lambesc (Provence, southeastern France) earthquake: a reappraisal based on macroseismic, seismological, and geodetic observations, *Journal of Geophysical Research* **108** (B9) 2,454. doi:10.1029/2002JB002348.
- Coburn, A., and R. Spence (2002). *Earthquake protection*. John Wiley and Sons. Ltd. 2nd edition. 420 pages.
- Dunand, F. and P. Gueguen (2012). Comparison between seismic and domestic risk in moderate seismic hazard prone region: the Grenoble City (France) test site, *Natural Hazards Earth Systems Sciences* **12** 511-526. doi:10.5194/nhess-12-511-2012
- GNDT (1993). *Rischio sismico di edifici pubblici Parte I: aspetti metodologici*. Centro Servizi Quasco, Bologna, 1993.

- Gueguen, P., C. Michel, and L. LeCorre (2007). A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: application to Grenoble (France), *Bulletin of Earthquake Engineering*, **4**(3) 467-490. doi: [10.1007/s10518-007-9036-3](https://doi.org/10.1007/s10518-007-9036-3)
- Global Earthquake Model (Online). (<http://www.globalquakemodel.org>)
- Grunthal, G. (1998). *European Macroseismic Scale*. Centre européen de Géodynamique et de Séismologie, Luxembourg, vol. 15, 1998.
- HAZUS (1997). *Earthquake loss estimation methodology*. Hazus technical manuals, National Institute of Building Science, Federal Emergency Management Agency (FEMA), Washington, 1997.
- Jackson, J. (2006). Fatal attraction : living with earthquakes, the growth of villages into megacities, and earthquake vulnerability in the modern world, *Philosophical Transactions of the Royal Society*, **364**(1845) 1,911- 1,925.
- Jaiswal, K., D.J. Wald, and K.A. Porter (2010). A global building inventory for earthquake loss estimation and risk management, *Earthquake Spectra* **26**(3) 731-748.
- Kappos, A.J., G. Panagopoulos, C. Panagiotopoulos, and G. Penelis (2006). A hybrid method for the vulnerability assessment of R/C and URM buildings, *Bulletin of Earthquake Engineering* **4**(4) 391-413.
- Lagamarsino, S., and S. Giovinazzi (2006). Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings, *Bulletin of Earthquake Engineering* **4**(4) 415–443.
- Lambert, J. (1997). *Les tremblements de terre en France : hier, aujourd'hui, demain....* BRGM Eds. Orléans (France), 196 pages.
- Lestuzzi, P., B. Duvernay, R. Peruzzi, and A. Schmid (2009). Evaluation sismique de bâtiments existants – approche basée sur le risque introduite en Suisse. *Annales du bâtiment et des travaux publics* **5** 30-36.
- Michel, C., P. Gueguen, and M. Causse (2012). Seismic vulnerability assessment to slight damage based on experimental modal parameters, *Earthquake Engineering and Structural Dynamics* **41**(1) 81-98.
- Parent, J.-F. (1982). *Grenoble, deux siècles d'urbanisme*. Presse Universitaire de Grenoble, 187 pages.
- Pierre, J.P., and M. Montagne (2004). The 20 April 2002, Mw 5.0 Au Sable Forks, New York, earthquake: a supplementary source of knowledge on earthquake damage to lifelines and buildings in Eastern North America, *Seismological Research Letters* **75**(5) 626-635.

- Porter, K.A., A.S. Kiremidjian, and J.S. LeGrue (2001). Assembly-based vulnerability of buildings and its use in performance evaluation, *Earthquake Spectra* **17**(2) 291-312.
- Scotti, O., D. Baumont, G. Quenet, and A. Levret (2004). The French macroseismic database SISFRANCE: objectives, results and perspectives, *Annals of geophysics* **47**(2/3) 571-581. doi: 10.4401/ag-3323.
- Spence, R., and B. Lebrun (2006). Earthquake scenarios for European cities: the risk-UE project, *Bulletin of Earthquake Engineering*, **4** special issue.
- Wald, D.J., K.-W. Lin, K.A. Porter, and L. Turner (2008). ShakeCast: automating and improving the use of ShakeMap for post-earthquake decision-making and response, *Earthquake Spectra* **24**(2) 533-553.
- Wald, D.J., V. Quitoriano, T.H. Heaton, H. Kanamori, C.W. Scrivner, and C.B. Worden (1999) TriNet ShakeMaps: rapid generation of peak ground motion and intensity maps for earthquakes in Southern California, *Earthquake Spectra* **15**(3) 537-555.
- Worden, C.B., D.J. Wald, T.I. Allen, K. Lin, D. Garcia, and G. Cua (2010). A revised ground motion and intensity interpolation scheme for shakemap, *Bulletin of the Seismological Society of America* **100**(6) 3,083-3,096.
- World Housing Encyclopedia (Online). <http://www.world-housing.net>

Tables

Table 1 - Number of buildings in Grenoble per attribute [roof number] and [period of construction] included in the VULNERALP and INSEE database.

Attributes	VULNERALP	INSEE
<i>Number of floors</i>		
< 3 floors	31.38 %	35.38 %
[3-5] floors	34.22 %	31.71 %
> 5 floors	34.4 %	32.91 %
<i>Epoch of construction</i>		
< 1945	34.61 %	42.11 %
[1945 - 1970[36.45 %	37.29 %
[1970 - 2000]	28.19 %	17.95%
>2000	0.75 %	2.69 %

Table 2 - Conditional probabilities for each EMS98 vulnerability class according to building attributes obtained by the learning phase applied to the VULNERALP database (GVM proxy).

INSEE Attributes	P(A)	P(B)	P(C)	P(D)
<3 floors - Before 1945	0.077	0.874	0.049	0.000
<3 floors - [1945-1970[0.000	0.647	0.353	0.000
<3 floors - [1970-2000[0.003	0.179	0.819	0.000
<3 floors - After 2000	0.000	0.071	0.500	0.429
[3 - 5] floors - Before 1945	0.055	0.884	0.060	0.000
[3 - 5] floors - [1945-1970[0.000	0.271	0.729	0.000
[3 - 5] floors - [1970-2000[0.000	0.052	0.948	0.000
[3 - 5] floors - After 2000	0.000	0.182	0.364	0.455
>5 floors - Before 1945	0.011	0.563	0.425	0.000
>5 floors - [1945-1970[0.000	0.089	0.911	0.000
>5 floors - [1970-2000[0.000	0.042	0.958	0.000
>5 floors - After 2000	0.000	0.000	0.250	0.750

Table 3 - Conditional probabilities for each EMS98 vulnerability class computed from the VULNERALP database and applying the GVM proxy to the INSEE database in Grenoble.

	INSEE	VULNERALP
P(A)	0.0219	0.0215
P(B)	0.4275	0.4323
P(C)	0.5468	0.5913
P(D)	0.0038	0.0038

Figure captions

Figure 1. INSEE national census information in Grenoble (France), grouped by geolocalized cells and representing (A) date of construction and (B) number of floors.

Figure 2. Distribution of the EMS98 vulnerability class in Grenoble computed with the GVM proxy considering INSEE attributes. i.e. date of construction and number of floors.

Figure 3. Application of the GVM proxy to the city of Nice. (A) Distribution of seismic vulnerability computed by the RiskUE method. Each zone was referenced by a number (B) Distribution of the EMS98 vulnerability classes in Nice computed by the GVM proxy.

Figure 4. Translation of the EMS98 damage classification (D0 to D5) according to vulnerability class and macroseismic intensity.

Figure 5. Prediction of damage in Nice using RISKUE and GVM proxy methods for intensity scenarios ranging from V to X.

Figure 6. Iso-values of macroseismic intensities produced by the historic 1909 Lambesc earthquake (Source SisFrance).

Figure 7. Time evolution of the vulnerability class probability of the area affected by the Lambesc earthquake.

Figure 8. Evaluation of the level of damage for the Lambesc earthquake scenario considering (A) 1909 urbanization and (B) 2008 urbanization. Damage is grouped by D0+D1 (left), D2+D3 (middle) and D4+D5 (right column) EMS98 damage scale. The colour scale gives the number of buildings per damage levels.