Seismic vulnerability assessment at urban scale using data mining and GIScience technology: application to Urumqi (China)

Yaohui Liu, Zhiqiang Li, Benyong Wei, Xiaoli Li and Bo Fu
Institute of Geology, China Earthquake Administration, Beijing, China

ABSTRACT
Seismic vulnerability assessments play a significant role in comprehensive risk mitigation efforts and seismic emergency planning, especially for urban areas with a high population density and a complex construction environment. Traditional approaches such as in situ fieldwork are accurate for conducting seismic vulnerability assessments of buildings; however, they are too much time and cost-consuming, especially in moderate to low seismic hazard regions. To address this issue, an integrated approach for a macroseismic vulnerability assessment composed of data mining methods and GIScience technology was presented and applied to Urumqi, China. First, vulnerability proxies were established via in situ data of buildings in the Tianshan District with an EMS-98 vulnerability classification scheme and two data mining methods, namely, support vector machine and association rule learning methods. Then, vulnerability proxies were applied to the Urumqi database, and the accuracy was validated. Finally, seismic risk maps were constructed through data consisting of direct damage to buildings and human casualties. The results indicated that the two data mining methods could achieve desirable accuracies and stabilities when estimating the seismic vulnerability. The seismic risk of Urumqi was estimated as Slight with a predicted number of 61,380 homeless people for a seismic intensity scenario of VIII.

ARTICLE HISTORY
Received 23 January 2018
Accepted 7 September 2018

KEYWORDS
seismic vulnerability; data mining; GIScience; EMS-98; Urumqi

1. Introduction
Natural hazards represent an ever-present threat to human life as well as to the physical infrastructure and global economy; among them, earthquakes constitute extremely serious and deadly natural disasters worldwide (Wei et al. 2013; Wei et al. 2017). China, which is highly prone to earthquakes, frequently suffers from destructive and catastrophic earthquake activity that results in the serious loss of both life and property (An et al. 2015). Approximately 1,600,000 deaths have been caused by...
earthquakes worldwide, 600,000 of which have been reported in China since the beginning of the twentieth century (Wu et al. 2014). Four major earthquakes have occurred within the last 10 years in China, namely, the 2008 $M_s 8.0$ Wenchuan earthquake, which caused more than 87,000 casualties and 845.1 billion Yuan (120 billion dollars) in direct economic losses (Tessfamariam and Liu 2010), the 2010 $M_s 7.1$ Yushu earthquake, which caused the deaths of 2,698 people (Qu et al. 2011), the 2013 $M_s 7.0$ Lushan earthquake, which killed 196 people and injured an additional 11,826 (Jia et al. 2014), and the 2014 $M_s 6.5$ Ludian earthquake, which caused 617 deaths and injured 3,143 people (Lin et al. 2015). China has recently experienced rapid and massive urban expansion coincident with tremendous economic, social and cultural development (Huang et al. 2017). Unfortunately, this accelerated and extensive growth of urban sprawl and the associated rapid changes to urban structures has induced substantially greater seismic risks at a uniquely high level (Zhou et al. 2015; Qi et al. 2016). Moreover, the accurate prediction of earthquakes is still a difficult and potentially impossible task (Jia et al. 2014; Ji et al. 2016). Consequently, reducing the seismic risk has become a priority for local authorities to maintain the well-being and safety of the local populations and to ensure economic and social security. Accordingly, to provide stakeholders and decision-makers with the means to develop mitigation strategies and safety measures for future earthquakes, urban-scale damage scenarios and seismic vulnerability assessments are greatly important, and they can help to raise the public awareness of seismic hazards prior to the onset of expected disasters (Panagiota et al. 2013).

A detailed seismic vulnerability assessment is a technically complex and expensive procedure that involves an enormous number of buildings in addition to their geometrical features and heterogeneities with regard to their structural design and seismic response characteristics within urban areas (Guettiche et al. 2017). Over the past few decades, several methodologies (both qualitative and quantitative), such as those proposed by the Gruppo Nazionale per Difesa dai Terremoti in Italy (Benedetti and Petrini 1984) and the Federal Emergency Management Agency (FEMA) in the U.S.A. (Kircher et al. 2006) in addition to the European RISK-UE method (Lagomarsino and Giovinazzi 2006; Mouroux and Le Brun 2008), have been developed according to the resources and scale of the assessment. These methodologies, which can be fine-tuned to represent the characteristics of regional building vulnerabilities so long as a detailed building inventory is available, have been widely applied to many areas of the world. The most reliable and detailed building inventories are provided via in situ field surveys, though they are usually quite expensive. Furthermore, these difficulties are additionally serious in regions of moderate seismic hazards where the resources that can be mobilized for seismic evaluations are rather limited (Guéguen et al. 2007; Riedel et al. 2015).

Recently, simplified macro-scale methods have been adopted to reduce the costs of the in situ field surveys and improve the efficiency seismic vulnerability and risk assessments at the urban scale. Such initiatives seek to simplify the visual screening stage by considering only the key parameters whose contributions to the seismic vulnerability are significant (Guéguen et al. 2007) or by employing remote sensing and geographic information system (GIS) methods (Mück et al. 2012; Wieland et al. 2012;
Geiß and Taubenböck 2013; Kaushik and Dasgupta 2013; Geiß et al. 2014; Moradi et al. 2014; Geiß et al. 2015a; Geiß et al. 2015b; Su et al. 2015; Costanzo et al. 2016; Dhar et al. 2016; Klotz et al. 2016; Ghorbanzadeh et al. 2017). Data mining methods have also been developed to ascertain the best proxy that links the features of building, which are easily assessed using remote sensing and civil engineering methods, with their seismic vulnerabilities (Şen 2010, 2011; Chen et al. 2012; Siraj et al. 2014; Wu H et al. 2014; Riedel et al. 2015; Campostrini et al. 2017; Ghorbanzadeh et al. 2017; Guettiche et al. 2017). The synergistic use of in situ fieldwork, data mining methodologies and GIScience technology constitutes a promising yet challenging approach for reducing the costs of in situ building surveys, which could become more efficient and accurate through an association with vulnerability index methods (VIM) in addition to macroseismic vulnerability and risk assessments in moderate-risk, seismic-prone regions.

Here, an integrated approach combining data mining, remote sensing interpretation, GIS-based mapping, and VIMs is tested and validated through a macroseismic vulnerability assessment of the city of Urumqi. First, the city of Urumqi is presented along with the Urumqi database. Second, the support vector machine (SVM) and the association rule learning (ARL) methods are presented and applied to in situ data to derive two vulnerability proxies for Urumqi. Based on these two vulnerability proxies and the Urumqi database, vulnerability maps of the city are generated within a GIS environment. Finally, the seismic damage within the city is further presented in association with the distribution of the expected damage grade throughout Urumqi in addition to the estimated number of human casualties.

2. Study area and data

2.1. Study area

The city of Urumqi, which is situated in Northwest China, has experienced considerable changes to its urban development over the last 20 years, and it currently serves as a regional, cultural, political, and commercial centre. By the end of 2014, the city hosted approximately 3,550,000 inhabitants and an administrative area of 10,989 square kilometres (2015 Census). As presented in Figure 1, the study area of this article, which encompasses the centre of Urumqi exhibits a probabilistic seismic intensity of VIII (Xu et al. 2016) and constitutes an urban area where comprehensive disaster prevention methods are crucial.

Since the beginning of the twentieth century, Urumqi has witnessed seven earthquakes with magnitudes of $M_s \geq 5.0$. The strongest recorded event was the $M_s$ 6.0 earthquake that struck in 1934 (Shen and Song 2008). Recent GPS measurements indicated that the compressional deformation of the Tianshan area is accommodated primarily by a series of W-E-trending fold-thrust belts with a slip rate of 1–2 mm/a and that Urumqi is moving toward the NE at a rate of 4–5 mm/a (Yang et al. 2008; Wu et al. 2017). The city is currently characterized by moderate to low seismic activity, and no major earthquake has occurred recently. However, no detailed quantitative assessments have been conducted within the city to date.
2.2. Data

Building inventories are considered the most important factors in seismic vulnerability and risk assessments. The database used in this study is constituted of two parts: the Urumqi database and the in situ data of the Tianshan District.

2.2.1. Urumqi database

The Urumqi database refers to the Urumqi earthquake emergency foundation database, which comprises social, economic, and population information in addition to a building inventory and city map as well as natural geographic landforms, key object locations, rescue team information, relief communication data, and earthquake
preplanning data (Zhong et al. 2002; Xu et al. 2016). The building inventory and population information were employed in this study to conduct a seismic vulnerability assessment; meanwhile, the other elements, which would only slightly affect the results of the assessment, were not included.

In China, urban expansion processes are closely related to national macroscopic policies. Moreover, urban residential communities typically refer to various types of relatively independent residential areas throughout the city (Liu et al. 2016; Lixin et al. 2017). Considering the above characteristics and the ambitions of the macroseismic vulnerability assessment in this study, the basic Urumqi database is subdivided into multiple sections according to several elements: urbanization, urban planning, land use and land cover, district limits, and other factors regarding building construction. Each section is represented by high building and population densities, unified structures, and a small mobility. Only residential dwellings are considered in this study; the information regarding the period of construction and number of floors for each building is available within the Urumqi database. Ultimately, 1,289 residential sections were identified and characterized as homogeneous for the following seismic vulnerability assessment.

2.2.2. In situ data
As demonstrated in any seismic vulnerability assessment approach, the classification of the seismic vulnerability fundamentally depends on the material and type of a structure while taking a variety of additional factors (e.g., constructional and architectural features, quality of material, period of construction, and state of preservation), which may impact the seismic performance of the building, into consideration (Santos et al. 2013; Geiß et al. 2014; Geiß and Taubenböck 2017). Therefore, the contents of the in situ data in this study include various parameters, including the type of material, number of floors, period of construction, and roof type (Matsui 2007; Riedel et al. 2014). Based on these parameters, each residential section was categorized into typical vulnerability classes according to the European Macroseismic Scale 98 (EMS-98) (Grüenthal 1998). The EMS-98 standard was primarily developed to analyze typical structural susceptibilities that could be expected throughout Europe; however, since the building vulnerability was considered when defining the intensity, the EMS-98 vulnerability classification could be used to represent the seismic damage in the target region for a given intensity (Riedel et al. 2015). This standard has been widely utilized as a useful vulnerability classification in other areas around the world (Tertulliani et al. 2011). According to the EMS-98, different types of buildings are classified into one of six vulnerability classes ranging from A (highest vulnerability) to F (lowest vulnerability). These different building types experience different levels of damage to a particular earthquake intensity depending on their inherent characteristics. According to the EMS-98, each building is labeled with a damage grade ranging from DG1 to DG5, where DG1 is the least vulnerable and DG5 is the most vulnerable.

The in situ data consisting of seismic vulnerabilities were collected in the Tianshan District in August 2017. A building-to-building survey was performed in a small area (approximately 1530 × 1200 m), including all buildings within the surveyed area (Figure 2). During the in situ fieldwork, we investigated the following attributes: the type of structure, period of construction, number of floors, and land use and land
cover. In general, the basic census data and the field survey conducted in Urumqi revealed the following dominant building types as presented in Figure 3 (Lagomarsino and Giovinazzi 2006; Su et al. 2015; Xu et al. 2016): adobe (M2), wooden slabs (M3.1), unreinforced masonry (M3.3), reinforced masonry (M4), reinforced concrete (RC1), and concrete shear walls (RC2). Based on the historic urban planning and urbanization, the period of construction was divided into four categories, namely, pre-1960, 1960–1981, 1982–2002, 2002+. The number of floors was defined according to the interval given in the RISK-UE method (Lagomarsino and Giovinazzi 2006; Mouroux and Le Brun 2008): low-rise (1–2 stories), mid-rise (3–5 stories), and high-rise (>6 stories). Through the use of high-resolution remote sensing images (from Google Earth) and manual image interpretation, the roof types were identified and categorized into two groups: flat and sloped. The building classes were divided into residential and non-residential types. Soil amplification factors were particularly removed from consideration in this study.

In total, 703 buildings were surveyed in the study area and their detailed construction and conservation properties were recorded. The buildings were then grouped according to the following attributes: the type of structure, period of construction, number of floors, land use and land cover, and roof type. Nearly all of the 703 buildings were constructed subsequent to 1982; only 1 building (0.1% in total) was built prior to 1960, whereas 12 (1.7%) were constructed in 1960–1981, 360 (51.2%) were constructed during 1982–2001, and 330 (46.9%) were built after 2002. A total of 179 (1.7%) buildings were low-rise structures (less than 2 floors), while 424 were mid-rise structures (3–6 floors), and only 100 buildings were high-rise (7 or more floors). Through the manual interpretation of remote sensing imagery, 46 of the buildings were identified as having sloped roofs, whereas 657 of the buildings had flat-type roofs. During the field survey in the study area, 616 buildings were identified as residential while the rest of the 87 buildings were non-residential. The distribution and
proportion of these different attributes were mapped within a GIS environment, the results of which are shown in Figures 4 and 5.

3. Methods

Data mining, which is an interdisciplinary subfield of computer science, is a computational process used to discover patterns within large datasets through a combination of machine learning, statistics, and database systems (Linoff and Berry 1997). The overall goal of the data mining process is to extract information from a dataset and transform it into an understandable structure for further use. This technology is important for various research fields, including genetics, mathematics, cybernetics, and marketing (Riedel et al. 2014). Various types of learning algorithms such as association rules, Bayesian classifiers, decision rules, decision trees, neural networks, and SVMs can be used for data mining (Fayyad et al. 1996).

Two data mining algorithms and VIMs are briefly introduced in this section. The schematic workflow of this study is presented in Figure 6.
Figure 4. Distributions of different attributes within the in situ database of the buildings.
3.1. Support vector machine (SVM)

The SVM model, which was proposed by Cortes and Vapnik (1995), was originally defined for the classification of linearly separable classes of objects. Subsequently, the SVM model has been widely employed in decision-making and prediction endeavors due to its considerable efficiency in dealing with linearly inseparable and high-dimension datasets (Chen et al. 2012; Kavzoglu et al. 2014).

Figure 5. Proportions of different attributes within the in situ database of the buildings.

3.1. Support vector machine (SVM)

The SVM model, which was proposed by Cortes and Vapnik (1995), was originally defined for the classification of linearly separable classes of objects. Subsequently, the SVM model has been widely employed in decision-making and prediction endeavors due to its considerable efficiency in dealing with linearly inseparable and high-dimension datasets (Chen et al. 2012; Kavzoglu et al. 2014).
Consider a set of training dataset vectors \( x_i \) \((i = 1, 2, \ldots, n)\) consisting of two classes denoted as \( y_i = \pm 1 \). The SVM model seeks an \( n \)-dimensional hyperplane that differentiates different classes of labeled samples according to their maximum gap. Two class data points are set to the farthest distance from the classifier hyperplane, which can be a flat or curved surface and can be mathematically expressed as follows:
The selection of the kernel function in an SVM model is crucial for the classification results (Xu and Xu 2012). In this work, radial basis function (RBF) kernels were implemented as follows:

\[ K(x_i, x_j) = e^{-\gamma (x_i - x_j)^2} \]  

(2)

where \( \gamma \) is the kernel function parameter that is entered manually during the calculation.

Lastly, we established the SVM model through RBF kernel function, which involves training and testing sets. Each instance in the training set has one class label (damage grade) and four attributes (type of structure, number of floors, period of construction, and roof type). Within a supervised classification framework, an SVM statistical learning algorithm is used on the dataset to label the buildings according to the desired EMS-98 standard for seismic vulnerability classes. Having found the best hyperplane (using only the training set), accuracy is then measured by creating a confusion matrix and calculating the ratio between the sum of the diagonal values (correct classification) over the sum of all the elements in the matrix.

### 3.2. Association rule learning (ARL)

ARL is a rule-based data mining method used to reveal interesting relationships among variables within a large database (Piatetsky-Shapiro 1991) using some measures of interestingness (Agrawal et al. 1993). Recently, some researchers (Riedel et al. 2014; Riedel et al. 2015; Guettiche et al. 2017) applied this method to conduct seismic vulnerability assessments since it establishes correlations using mathematical algorithms (i.e. if/then statements) between the basic attributes of buildings and the EMS-98 vulnerability classes.

In this study, we applied a simplified ARL method to the in situ Urumqi database. The conditional probabilities between the EMS-98 vulnerability classes (target \( Y_i = A, B, C, D, \) and \( E \)) and the basic information regarding the buildings (attribute \( X \)) are derived to obtain the in situ vulnerability proxy. ARL also takes the form \( X \rightarrow Y_i \), where \( X \) (consequent) and \( Y_i \) (antecedent) represent two sets of independent items. Each relationship between \( X \) and \( Y_i \) can be presented in a binary format \([0, 1]\): knowing the building attribute \( X \), the probability of belonging to the class \( Y_i \) is defined as:

\[ P(Y_i|X) = \frac{P(Y_i|X)}{P(X)} \]  

(3)

In practice, \( P(Y_i|X) \) can be calculated as follows:

\[ P(Y_i|X) = \frac{N_{XY}}{N_X} \]  

(4)
where $N_X$ is the total number of buildings with the attribute $X$, and $N_{XY}$ is the number of buildings with the attribute $X$ belonging to the class $Y_i$. One limitation of the ARL method is the significant risk of false associations that might arise when searching through massive numbers of possible associations, as the results would consequently include inconsistencies (Guettiche et al. 2017).

### 3.3. Vulnerability index method

The vulnerability index method (VIM) adopted in this work to assess the seismic vulnerabilities of buildings in Urumqi is based on the RISK-UE project (Lagomarsino and Giovinazzi 2006), which was launched in 1999 and involved seven cities throughout Europe and around the Mediterranean Sea, namely, Barcelona (Spain), Bitola (Macedonia), Bucharest (Romania), Catania (Italy), Nice (France), Sofia (Bulgaria), and Thessaloniki (Greece). Due to its compatibility with the EMS-98 standard, the VIM has been widely applied to other areas worldwide, including Iran (Omidvar et al. 2012), Grenoble in France (Riedel et al. 2014; Riedel et al. 2015), Faro in Portugal (Maio et al. 2016), Horta in Portugal (Ferreira et al. 2016), Sion and Martigny in Switzerland (Lestuzzi et al. 2016), and Al Hoceima in Morocco (Cherif et al. 2016). The VIM requires that seismicity must be interpreted as a vulnerability index with values ranging from 0 (least vulnerable) to 1 (most vulnerable) based on the macroseismic intensity and the resistance of a structure. The procedure for determining the damage can be categorized into three steps as discussed hereafter.

Step 1: estimation of the vulnerability index $V_I$

A building typology matrix (BTM) is a typological classification that reflects the differences among various types of structures that are expected to display similar seismic performances (Table 1). Basic vulnerability indices are attributed to each typology class (Mouroux and Le Brun 2008) corresponding to the median value $V_{I,BTM}^*$ and to the lower ($V_{I,BTM}^-$) and upper ($V_{I,BTM}^+$) bounds of the possible values of the vulnerability index. Modifier factors are then applied to $V_I^*$ in consideration of various physical factors, including the height, irregularities, and position. The final vulnerability index is the sum of all possible factors as follows:

$$V_I = V_I^* + \Delta V_R + \Delta V_m$$  \hspace{1cm} (5)

where $V_I^*$ is the vulnerability index corresponding to the building classification, $\Delta V_R$ is a regional vulnerability factor considering the characteristics of the region or building period, and $\Delta V_m$ is the seismic behavior modifier that includes all other

| Typology | Description | $V_{I,BTM}^-$ | $V_{I,BTM}^*$ | $V_{I,BTM}^+$ | $V_{I,BTM}^{max}$ |
|---|---|---|---|---|---|
| M2 | Adobe (earth bricks) | -0.02 | 0.047 | 0.386 | 0.67 | 0.86 |
| M3.1 | Wooden slabs (hardwood) | 0.14 | 0.33 | 0.451 | 0.633 | 0.7 |
| M3.3 | Composite steel and masonry slabs | 0.46 | 0.527 | 0.704 | 0.83 | 1.02 |
| M4 | Reinforced/confined masonry walls | 0.46 | 0.65 | 0.74 | 0.83 | 1.02 |
| RC1 | Concrete moment frames | -0.02 | 0.047 | 0.442 | 0.8 | 1.02 |
| RC2 | Concrete shear walls | -0.02 | 0.047 | 0.386 | 0.67 | 0.86 |
aspects of the seismic performance. The regional vulnerability modifier and the seismic behavior modifiers, which could characterize the regional difference of seismic vulnerability, would help to improve the accuracy of the seismic vulnerability assessment. For macroseismic vulnerability assessment at urban scale in this study, the regional vulnerability modifier and the seismic behavior modifiers were taken to be equal to zero.

Step 2: estimation of the mean damage grade $\mu_D$

A mean damage grade $\mu_D$ is defined to characterize the expected damage using the following equation:

$$\mu_D = 2.5 \left[ 1 + \tanh \left( \frac{I + 6.25V_I - 13.1}{\phi} \right) \right]$$

(6)

where $I$ is the macroseismic intensity, which usually ranges from V (5) to XII (12), and $\phi$ is the ductility index, which is evaluated by considering the building typology and constructive features (Lagomarsino 2006; Lagomarsino and Giovinazzi 2006). For residential buildings, $\phi$ takes a value of 2.3 (Lantada et al. 2009).

The weighted mean damage index $\bar{\mu}_D$ can be calculated using the following equation:

$$\bar{\mu}_D = \sum_{k=0}^{5} p_k k$$

(7)

where the integer $k$ ranges from 0 to 5 for the damage stakes, and $p_k$ represents the corresponding probability of occurrence for each damage stake.

The EMS-98 damage grades are considered to describe the observed physical damage to the structural and non-structural components of buildings. Five damage grades are identified according to Lagomarsino and Giovinazzi (2006) such that $D_k$ ($k = 0, 1, 2, 3, 4, 5$): $D_0$ (None), $D_1$ (Slight), $D_2$ (Moderate), $D_3$ (Heavy), $D_4$ (Very Heavy), and $D_5$ (Destruction).

Step 3: estimation of the damage distribution

The expected probability of occurrence of a damage grade for any degree of seismic intensity is assumed to follow a binomial distribution. The probability of each damage grade can be calculated as follows:

$$P(D_k) = \frac{5!}{k!(5-k)!} \left( \frac{\mu_D}{5} \right)^k \left( 1 - \frac{\mu_D}{5} \right)^{5-k}$$

(8)

where $!$ indicates the factorial operator.

4. Results and discussion

4.1. SVM application

4.1.1. Training samples

After the buildings in the in situ database were assigned to the EMS-98 classes based on their basic attributes, we applied the SVM classification. The entire database was
separately divided into two groups: training sets and testing sets (training sets are not included within the testing sets). To confirm the uncertainties in the prediction results, we randomly selected a training set and repeated each experiment 1000 times (with 1,000 independent training and testing divisions). Subsequently, we counted the median and overall accuracy of each experiment and created an accuracy histogram.

First, the four attributes of the buildings and the five EMS-98 classes were utilized to identify the optimal size of the training sets. The numbers of training buildings are 62, 123, 185, 247, 308, 370, and 431, which respectively correspond to 10, 20, 30, 40, 50, 60, and 70% of the entire training set. Our results clearly demonstrate that the accuracy initially increased and the dispersion initially decreased as the training set size percentage increased. The maximum attainable accuracy was reached when the training set size was increased to 40%, and the influence of the increasing set size was lessened. After increasing beyond 40%, the precision gradually diminished and the deviation increased (Figure 7). Hence, a training size of 40% was determined to represent the best training set size, and thus, this size is adopted for the calculations in the following experiments.

The classification accuracy depends on the different combinations of datasets employed. Therefore, different combinations of building attributes were tested with a 40% training size. The mean accuracy was then calculated after 1000 iterations with each combination, the results of which are presented in Figure 8. Figure 8a demonstrates that the mean accuracy is approximately 65.9% when considering only two attributes (the period of construction and number of floors) and then sharply falls to 41.1% for the combination of two different attributes (the roof type and the number of floors) (Figure 8b). By adding the roof type, which was obtained through the manual interpretation of remote sensing imagery, the mean accuracy is slightly increased to 66.4% (Figure 8c). Through a comparison of Figure 8a, b, and c, it is clear that the roof type is not completely independent since the associated accuracy is not substantially improved. Thus, after replacing the roof type with the type of structure, the

![Figure 7. Evolution of the accuracy and dispersion among the different training set sizes.](image)
mean accuracy is enhanced drastically to 78.9% (Figure 8d). This indicates that the type of structure is a fundamental factor for a seismic vulnerability assessment, which is in accordance with the findings of previous studies (Lang and Bachmann 2004; Wu et al. 2014; Riedel et al. 2015; Guettiche et al. 2017), further revealing that high-resolution remote sensing imagery can allow for the extraction of building details and aid

Figure 8. Overall accuracy of different attribute combinations. (a) Period of construction and number of floors. (b) Roof type and number of floors. (c) Roof type, number of floors and period of construction. (d) Number of floors, period of construction and type of structure. (e) All four attributes (number of floors, period of construction, type of structure, and roof type) with all five classes. (f) All four attributes with the vulnerability classes merged into only three classes: (A-B), (C-D), and (E-F).
the accurate assessment of building seismic vulnerabilities. Figure 8e shows that the mean accuracy of the classification when considering all four attributes (i.e. the number of floors, period of construction, type of structure, and roof type) and all five EMS-98 classes climbs to 79.7%. Since the SVM classifier could not classify the hyperplane with an extreme accuracy and always failed to differentiate nearby classes clearly, the effect of merging the classes was studied by reducing the six classes to only three. After Classes A and B were combined into Class 1, C and D were combined into Class 2, and E and F were combined into Class 3, the accuracy rose from 79.7 to 83.8% (Figure 8f). In summary, Figure 8 shows that a greater amount of building information can help improve the overall accuracy of the learning phase.

In addition, we also calculated an example of the confusion matrix using four attributes and five classes from the in situ database (Table 2). The values on the diagonal are the buildings for which the classes were assigned correctly. The overall accuracy of the classification was 79.7%. Meanwhile, the accuracy was lower for Classes A and C; 2 Class A buildings were correctly classified, but two buildings were wrongly classified into Class B. In addition, 55 Class C buildings were correctly classified while 36 and 11 buildings were classified as Class B and C, respectively. Previous research (Riedel et al. 2015) noted that it is difficult to distinguish between buildings of Class A and B because of their equivalent vulnerabilities according to the EMS-98 standard. However, considering the fact that Urumqi is a relatively newly constructed city where Class A buildings are rare, we can neglect the influence of an incorrect classification of Class A structures. The other three classes display better classification results. For Class B, 87 buildings were correctly classified while 12 buildings were classified as Class C. 44 Class D buildings were correctly classified while 14 buildings were classified as Class E. In addition, 105 Class E buildings were correctly classified while 1 building was classified as Class D. The classification accuracies for Class A, B, C, D, and E were respectively 50, 87.9, 53.9, 75.9, and 99.1%.

### 4.1.2. Validation and application to Urumqi

A vulnerability proxy was defined after the SVM training phase by considering only two building attributes (the period of construction and number of floors). Table 3 displays the conditional classification value in the EMS-98 vulnerability classes, including all 12 combinations of the two attributes (i.e. for four types of construction

---

**Table 2. Example of a confusion matrix obtained using the SVM method with the in situ database.**

| EMS-98 Vulnerability (classified) | A | B | C | D | E | Sum | Accuracy (%) |
|----------------------------------|---|---|---|---|---|-----|-------------|
| **Vulnerability (field surveyed)** |   |   |   |   |   |     |             |
| A                                | 2 | 0 | 0 | 0 | 0 | 2   | 100         |
| B                                | 2 | 87| 36| 0 | 0 | 125 | 69.6        |
| C                                | 0 | 12| 55| 0 | 0 | 67  | 82.1        |
| D                                | 0 | 0 | 11| 44| 1 | 56  | 78.6        |
| E                                | 0 | 0 | 0 | 14| 105| 119 | 88.2        |
| **Sum**                          | 4 | 99| 102| 58| 106| 369 |             |
| **Accuracy (%)**                 | 50| 87.9|53.9|75.9|99.1|     |             |

EMS-98 classes (A–E) and four attributes (the number of floors, period of construction, type of structure, and roof type).
periods and three ranges for the number of floors). For instance, a randomly selected building within the in situ area of the Tianshan District that is known to have been built after 2002 and has more than seven floors has a 25% probability of being in Class D and a 75% probability of being in Class E.

Then, we estimated the seismic vulnerability of the entire city by applying the SVM vulnerability proxy to the Urumqi database. The need for only two attributes eliminates one of the main difficulties related to any vulnerability study while remaining relevant at the urban scale. All of the available data were collected and integrated into a GIS environment, which can perform building analyses and present the associated distributions. Figure 9 shows the different vulnerability distributions throughout Urumqi calculated using the vulnerability proxy in Table 3. The results clearly indicate that the whole city is generally less vulnerable since the building vulnerability is mostly higher than Class C. This suggests that the city has a generally high resistance to earthquakes. The lowest vulnerability is discovered both within the city centre, since the building structures therein change and update rapidly, and around the periphery of the city, especially in the Toutunhe District and Xinshi District, where an earthquake resistance project with carried with the building renewal. In contrast, the highest vulnerabilities are mainly detected in the urban and rural areas, where the building structures are unconfined and there are fewer urban processes than in the city centre.

### 4.2. ARL application

Using two attributes (the period of construction and number of floors) and ARL with 616 buildings from the in situ database, we derived the ARL vulnerability proxy (Table 4).

From Table 4, we can conclude that all of the buildings built prior to 1960 with less than two floors all belong to Class A. No buildings were constructed prior to 1960 with three or more floors. This phenomenon corresponds to the urban planning and construction history of Urumqi.

After the establishment of the vulnerability proxy, we then utilized the following formula (Riedel et al. 2015) to obtain the distribution of the EMS-98 classes throughout the whole city:

| Attributes               | P(A) | P(B) | P(C) | P(D) | P(E) |
|--------------------------|------|------|------|------|------|
| < 1960 and ≤ 2 fl.       | 1.000| 0.000| 0.000| 0.000| 0.000|
| (1960–1981) and ≤ 2 fl.  | 0.000| 1.000| 0.000| 0.000| 0.000|
| (1982–2001) and ≤ 2 fl.  | 0.000| 0.179| 0.821| 0.000| 0.000|
| > 2002 and ≤ 2 fl.       | 0.000| 0.000| 0.136| 0.636| 0.228|
| < 1960 and (3–6) fl.     | /    | /    | /    | /    | /    |
| (1960–1981) and (3–6) fl.| 0.000| 0.211| 0.474| 0.315| 0.000|
| (1982–2001) and (3–6) fl.| 0.000| 0.366| 0.549| 0.085| 0.000|
| > 2002 and (3–6) fl.     | 0.000| 0.000| 0.015| 0.353| 0.632|
| < 1960 and >7 fl.        | /    | /    | /    | /    | /    |
| (1960–1981) and >7 fl.   | 0.000| 0.412| 0.588| 0.000| 0.000|
| (1982–2001) and >7 fl.   | 0.000| 0.200| 0.300| 0.500| 0.000|
| > 2002 and >7 fl.        | 0.000| 0.000| 0.000| 0.250| 0.750|
Figure 9. Spatial distribution of the seismic vulnerability in Urumqi using the SVM method.

Table 4. Conditional probabilities obtained by applying ARL to the Urumqi database.

| Attributes | $P(A)$ | $P(B)$ | $P(C)$ | $P(D)$ | $P(E)$ |
|------------|--------|--------|--------|--------|--------|
| <1960 and ≤2 fl. | 1.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| (1960–1981) and ≤2 fl. | 0.154  | 0.846  | 0.000  | 0.000  | 0.000  |
| (1982–2001) and ≤2 fl. | 0.000  | 0.200  | 0.800  | 0.000  | 0.000  |
| >2002 and ≤2 fl. | 0.000  | 0.000  | 0.149  | 0.489  | 0.362  |
| <1960 and (3–6) fl. | /      | /      | /      | /      | /      |
| (1960–1981) and (3–6) fl. | 0.000  | 0.333  | 0.667  | 0.000  | 0.000  |
| (1982–2001) and (3–6) fl. | 0.000  | 0.555  | 0.361  | 0.084  | 0.000  |
| >2002 and (3–6) fl. | 0.000  | 0.000  | 0.018  | 0.309  | 0.673  |
| <1960 and >7 fl. | /      | /      | /      | /      | /      |
| (1960–1981) and >7 fl. | 0.000  | 0.000  | 0.600  | 0.400  | 0.000  |
| (1982–2001) and >7 fl. | 0.000  | 0.786  | 0.143  | 0.071  | 0.000  |
| >2002 and >7 fl. | 0.000  | 0.000  | 0.022  | 0.455  | 0.523  |
where \( P_j(X) \) is the probability of containing a vulnerability class \( X_i = (A, B, C, D, E) \) in each homogeneous area, \( N_{ji} \) is the number of buildings with the attribute \( Y_i \) in section \( j \), \( N \) is the total number of buildings in \( j \), and \( P(X|Y_i) \) is the value of the probability given by the vulnerability proxy for the \( X \rightarrow Y_i \) association in Table 4.

Figure 10 shows the vulnerability distributions in Urumqi calculated using the ARL method. Similar seismic vulnerability distributions and trends can be observed by comparing Figures 9 and 10, indicating that both data mining methods can achieve a desirable accuracy and stability for an estimation of the seismic vulnerability.

To further investigate the differences between the SVM and ARL methods, we calculated the ratio between the two methods using the estimated results of the

\[
P_j(X) = \frac{\sum_{i}^{12} N_{ji}P(X|Y_i)}{N}
\]
vulnerability distributions. As illustrated in Figure 11, the ARL/SVM ratios for Class A, B, and C are all above 1, indicating that the ARL method returns a greater number of more vulnerable buildings. In contrast, ARL calculated fewer buildings for the less vulnerable classes. This suggests that a greater degree of seismic damage would be estimated using the ARL method than with the SVM method for a particular seismic intensity over a given large-scale region.

4.3. Seismic risk scenarios for Urumqi

Seismic risk assessments incorporate both vulnerability assessments and hazard analyses. To reduce the seismic vulnerabilities of buildings and lower the loss of human life, we attempted to assess the percentages of the degrees of damage and loss of life in Urumqi.

4.3.1. Overall vulnerability map of Urumqi

The distribution of the mean vulnerability index is exhibited in the overall vulnerability map (Figure 12). The vulnerability index ranges from 0.3 to 0.8 for the Urumqi database with a mean value of 0.51. In general, the central urban areas have the lowest overall vulnerability with values of less than 0.55. In contrast, higher vulnerabilities are distributed along the peripheral zones of the Midong, Tianshan, and Saybagh Districts, where the buildings were erected with a lower quality of construction and fewer seismic reinforcements. Some particularly vulnerable sections with vulnerability values exceeding 0.7 can be found scattered throughout the city. This is primarily due to the old age of the buildings and the fact that they have not been renovated.

4.3.2. Direct physical damage

Once the vulnerability has been defined, the mean damage grade can be calculated for the different macroseismic intensities using formula 8. Table 5 shows the mean damage index values and their corresponding damage states. Only the intensity is considered in this study, while other additional factors such as site effects, soil
nonlinearity or triggered effects (e.g. landslides) that might affect the seismic hazard behavior are not considered.

GIS spatial analysis is employed to spatially represent the damage distributions of human settlements, thereby enabling the identification of more vulnerable areas and buildings, which can be crucial for urban management and planning in addition to protection strategies (Ferreira et al. 2016). Figure 13 displays the mean damage grade.
distribution of Urumqi for an earthquake scenario with a given intensity of VIII. The values of physical damage range from 0.44 to 4.24 with a mean value of 0.83. This corresponds to a Slight damage state according Table 5. Figure 14 shows the proportion of damage states among the residential districts of Urumqi; the differences among the damage states in each district are not obvious, as 80% of the sections in each district exhibit values between None and Slight. According to the distribution of residential areas, fewer sections in Urumqi County exhibit a better seismic performance.

4.3.3. Human casualties and homelessness

The main purpose of an earthquake protection programme is to ensure human safety and reduce injuries. Human losses are estimated by considering building damage as the root cause for fatalities and injuries (Coburn and Spence 2003); however, this is not consistent with the results presented in Section 4. Therefore, no casualties are expected for a seismic intensity scenario of VIII in this study.

The HAZUS methodology presented by FEMA was adopted in this study to estimate the probabilities of casualties and homelessness. This method considers 100% of the residential units located in buildings within the Very Heavy and Destruction damage grades and 90% of those units within the Substantial to Heavy damage grade to be uninhabitable.

The total number of uninhabitable residential units due to structural damage is computed using the following equation:

\[
\%MF = 100\%D_{MF} + 100\%VH_{MF} + 90\%H_{MF}
\]

\[
UNU_{SD} = U_{MF} \times \%MF
\]  

(10)

where \(U_{MF}\) represents the total number of multi-family residential units, and \(D_{MF}\), \(VH_{MF}\), and \(H_{MF}\) represent the damage grade probabilities for the Destruction, Very Heavy and Substantial to Heavy structural damage states, respectively.

The total number of persons relocated from each building with a typology \(P_{UND}\) is obtained using the following relationship:

\[
P_{UND} = P_h \times UNU_{SD}
\]  

(11)

where \(P_h\) is the number of persons assumed to live in each household of the building.
Figure 15 presents the amount and distribution of homeless people expected following a seismic intensity scenario of VIII. The total number of homeless people estimated in this scenario is 61,380. A large number of homeless people are located in the Xinshi, Saybagh, and Tianshan Districts, which are densely populated and rarely contain reinforced buildings. In contrast, the number of homeless people is relatively small in both the Toutunhe District and Urumqi County.

5. Conclusions

Rapid socioeconomic growth in earthquake-prone areas can cause rapid exposure and changes to the seismic vulnerability and risk of loss. Accordingly, seismic vulnerability and risk assessments at the urban scale have proven to be significant and effective for not only urban planners and decision-makers in the development of corresponding strategies for earthquake disaster reduction and immigration but also public awareness with regard to earthquake prevention. Detailed and real-time building inventories, however, are not easily obtained, especially for developing countries characterized by dynamic urban growth with unplanned and highly vulnerable settlements. To address this problem, an integrated approach for a macroseismic vulnerability assessment combining data mining methods with GIScience technology was tested and applied to the city of Urumqi, China. This combination aspires to take full advantage of existing data and extract ‘hidden’ information from datasets. In this study, the possibility of using relatively few building attributes to represent and analyze the seismic risk at the urban scale was proven efficiently.

Using the information available in the in situ area of the Tianshan District with two data mining methods (i.e. the SVM and ARL methods), two vulnerability proxies were derived to create relationships between combinations of building attributes and their most likely EMS-98 vulnerability classes. These proxies were derived during the learning phase and then applied to the whole basic Urumqi database. The accuracies of the SVM and ARL methods were evaluated, the results of which showed that both
approaches provide similar results. This suggests that both data mining methods were successfully applied to the database of Urumqi and achieved desirable accuracies and stabilities for the estimation of the seismic vulnerability. For a particular seismic intensity within a given large-scale region, the ARL method would predict a greater seismic damage than the SVM method. Meanwhile, few differences were considered for the building structures between those suitable for RISK-UE and the existing structures of Urumqi. A seismic risk assessment was performed for Urumqi using the seismic hazard scenarios, the results of which revealed that the seismic risk of Urumqi is Slight with a predicted number of 61,380 homeless people for a seismic intensity scenario of VIII, since Urumqi is not exposed to soil effects and most of the buildings (especially those built before 2002) have low vulnerability indices. Future research will conduct in situ field surveys in other districts of Urumqi to expand the training sets and the testing sets, in order to improve the reliability and minimize the uncertainty of these methods. Additionally, updates to the Urumqi database are also critical to identifying more precise results of the seismic vulnerability assessments.

In conclusion, we confirmed that a seismic vulnerability assessment using data mining methods and GIScience technology can provide a relevant estimation of the seismic damage at a far lower cost than that of a conventional method (e.g. in situ fieldwork). The results of this study represent a potential guide for earthquake protection and risk management endeavors for the city of Urumqi as well as a powerful tool for urban development. Moreover, the original results described in this study in conjunction with the complete and detailed technical information concerning the expected damage and damage to the populace will be delivered to the civil protection services of the municipality of Urumqi, who will utilize these findings to update the emergency plans for the city.

**Geolocation information**

Urumqi, China, Asia.

**Acknowledgements**

We sincerely thank the anonymous reviewers and associated editors for their valuable and constructive comments and suggestions that helped to improve this article.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by the National Natural Science Foundation of China under Grant [number 41661134013]; under Grant [number 41601567]; under Grant [number 41601390]; the Key Special Fund for the Study on Rapid Assessment of Multi-source Earthquake Loss under Grant [number 201308018-5].
References

Agrawal R, Imieliński T, Swami AN. 1993. Mining Association Rules Between Sets of Items in Large Databases. SIGMOD Conference.

An J, Bai X, Xu J, Nie G, Wang X. 2015. Prediction of highway blockage caused by earthquake-induced landslides for improving earthquake emergency response. Nat Hazards. 79(1):511–536.

Benedetti D, Petrini V. 1984. Sulla vulnerabilita sismica di edifici in muratura: un metodo di valutazione. A method for evaluating the seismic vulnerability of masonry buildings. *Lindustria Delle Costruzioni*. 18:66–78.

Camposotrini GP, Taffarel S, Bettiol G, Valluzzi MR, Da Porto F, Modena C. 2017. A Bayesian approach to rapid seismic vulnerability assessment at urban scale. *Int J Archit Heritage*. 12:1–11.

Chen CS, Cheng MY, Wu YW. 2012. Seismic assessment of school buildings in Taiwan using the evolutionary support vector machine inference system. *Exp Syst Appl*. 39(4):4102–4110.

Cherif S-E, Chourak M, Abed M, Pujades L. 2016. Seismic risk in the city of Al Hoceima (north of Morocco) using the vulnerability index method, applied in Risk-UE project. *Nat Hazards*. 85(1):329–347.

Coburn A, Spence R. 2003. *Earthquake protection*. New York: Wiley.

Cortes C, Vapnik V. 1995. Support-vector networks. *Mach Learn*. 20(3):273–297.

Costanzo A, Montuori A, Silva J, Silvestri M, Musacchio M, Doumaz F, Stramondo S, Buongiorno M. 2016. The combined use of airborne remote sensing techniques within a GIS environment for the seismic vulnerability assessment of urban areas: an operational application. *Remote Sens*. 8(2):146.

Dhar S, Rai AK, Nayak P. 2016. Estimation of seismic hazard in Odisha by remote sensing and GIS techniques. *Nat Hazards*. 86(2):695–709.

Fayyad U, Piatetsky-Shapiro G, Smyth P. 1996. *From data mining to knowledge discovery: an overview. Advances in knowledge discovery and data mining*. Menlo Park: American Association for Artificial Intelligence.

Ferreira TM, Maio R, Vicente R. 2016. Seismic vulnerability assessment of the old city centre of Horta, Azores: calibration and application of a seismic vulnerability index method. *Bull Earthq Eng*. 15(7):2879–2899.

Geiß C, Pelizari PA, Marconcini M, Sengara W, Edwards M, Lakes T, Taubenböck H. 2015a. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS J Photogramm Remote Sens*. 104:175–188.

Geiß C, Taubenböck H. 2013. Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Nat Hazards*. 68(1):7–48.

Geiß C, Taubenböck H. 2017. One step back for a leap forward: toward operational measurements of elements at risk. *Nat Hazards*. 86(1):1–6.

Geiß C, Taubenböck H, Tyagunov S, Tisch A, Post J, Lakes T. 2014. Assessment of seismic building vulnerability from space. *Earthq Spectra*. 30(4):1553–1583.

Geiß C, Wurm M, Breunig M, Felbier A, Taubenbock H. 2015b. Normalization of TanDEM-X DSM data in urban environments with morphological filters. *IEEE Trans Geosci Remote Sens*. 53(8):4348–4362.

Ghorbanzadeh O, Feizizadeh B, Blaschke T. 2017. Multi-criteria risk evaluation by integrating an analytical network process approach into GIS-based sensitivity and uncertainty analyses. *Geomatics Natural Hazards and Risk*. 9:127–151.
Grünthal G. 1998. European macroseismic scale 1998. European Seismological Commission (ESC).

Guéguen P, Michel C, LeCorre L. 2007. A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: application to Grenoble (France). Bull Earthq Eng. 5(3):467–490.

Guettiche A, Guéguen P, Mimoune M. 2017. Seismic vulnerability assessment using association rule learning: application to the city of Constantine, Algeria. Nat Hazards. 86(3):1223–1245.

Huang X, Xia J, Xiao R, He T. 2017. Urban expansion patterns of 291 Chinese cities, 1990–2015. Int J Digit Earth. 57:2802–2812.

Ji J, Xu Z, Gao X. 2016. Analysis of seismic vulnerability of rural houses in China. In: D’Amico S, editor. Earthquakes and their impact on society. Springer Natural Hazards. Cham: Springer; p. 363–375.

Jia H, Pan D, Wang J-A, Zhang W-C. 2016. Risk mapping of integrated natural disasters in China. Nat Hazards. 80(3):2023–2035.

Jia K, Zhou S, Zhuang J, Jiang C. 2014. Possibility of the Independence between the 2013 Lushan Earthquake and the 2008 Wenchuan Earthquake on Longmen Shan Fault, Sichuan, China. Seismol Res Lett. 85(1):60–67.

Kaushik HB, Dasgupta K. 2013. Assessment of Seismic Vulnerability of Structures in Sikkim, India, Based on Damage Observation during Two Recent Earthquakes. J Perform Constr Facil. 27(6):697–720.

Kavzoglu T, Sahin EK, Colkesen I. 2014. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides. 11(3):425–439.

Kircher CA, Whitman RV, Holmes WT. 2006. HAZUS earthquake loss estimation methods. Nat Hazards Rev. 7(2):45–59.

Klotz M, Kemper T, Geiß C, Esch T, Taubenböck H. 2016. How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. Remote Sens Environ. 178:191–212.

Lagomarsino S. 2006. On the vulnerability assessment of monumental buildings. Bull Earthq Eng. 4(4):445–463.

Lagomarsino S, Giovinazzi S. 2006. Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. Bull Earthq Eng. 4(4):415–443.

Lang K, Bachmann H. 2004. On the seismic vulnerability of existing buildings: a case study of the city of Basel. Earthq Spectra. 20(1):43–66.

Lantada N, Irizarry J, Barbat AH, Goula X, Roca A, Susagna T, Pujades LG. 2009. Seismic hazard and risk scenarios for Barcelona, Spain, using the Risk-UE vulnerability index method. Bull Earthq Eng. 8(2):201–229.

Lestuzzi P, Podestà S, Luchini C, Garofano A, Kazantzidou-Firtinidou D, Bozzano C, Bischof P, Haffer A, Rouiller J-D. 2016. Seismic vulnerability assessment at urban scale for two typical Swiss cities using Risk-UE methodology. Nat Hazards. 84(1):249–269.

Lin X, Zhang H, Chen H, Chen H, Lin J. 2015. Field investigation on severely damaged aseismic buildings in 2014 Ludian earthquake. Earthq Eng Eng Vib. 14(1):169–176.

Linnof GS, Berry MJA. 1997. Data mining technique: for marketing, sales, and customer support. Indianapolis: Wiley.

Liu F, Zhang Z, Wang X. 2016. Forms of urban expansion of Chinese municipalities and provincial capitals, 1970s–2013. Remote Sens. 8(11):930.

Lixin Y, Ke C, Xiaoying C, Yueling S, Xiaoqing C, Ye H. 2017. Analysis of social vulnerability of residential community to hazards in Tianjin, China. Nat Hazards. 87(2):1223–1243.

Mück M, Taubenböck H, Post J, Wegscheider S, Strunz G, Sumaryono S, Ismail FA. 2012. Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data. Nat Hazards. 68(1):97–114.

Maio R, Ferreira TM, Vicente R, Estêvão J. 2016. Seismic vulnerability assessment of historical urban centres: case study of the old city centre of Faro, Portugal. J Risk Res. 19(5):551–580.
Matsui T. 2007. Sloshing in a cylindrical liquid storage tank with a single-deck type floating roof under seismic excitation. *J Press Vessel Technol*. 131: 021303.

Moradi M, Delavar MR, Moshiri B. 2014. A GIS-based multi-criteria decision-making approach for seismic vulnerability assessment using quantifier-guided OWA operator: a case study of Tehran, Iran. *Ann GIS*. 21(3):209–222.

Mouroux P, Le Brun B. 2008. Risk-UE project: an advanced approach to earthquake risk scenarios with application to different European towns. In: Oliveira CS, Roca A, Goula X, editors. *Assessing and managing earthquake risk*. Dordrecht: Springer; p. 479–508.

Omidvar B, Garmiri B, Derakhshan S. 2012. Experimental vulnerability curves for the residential buildings of Iran. *Nat Hazards*. 60(2):345–365.

Panagiota M, Jocelyn C, Erwan P, Philippe G. 2013. A support vector regression approach for building seismic vulnerability assessment and evaluation from remote sensing and in-situ data. *Geoscience and Remote Sensing Symposium*.

Piatetsky-Shapiro G. 1991. Discovery, analysis, and presentation of strong rules. In: *Knowledge discovery in databases*. Cambridge, MA: AAAI/MIT Press.

Qi W, Su G, Sun L, Yang F, Wu Y. 2016. Internet+” approach to mapping exposure and seismic vulnerability of buildings in a context of rapid socioeconomic growth: a case study in Tangshan, China. *Nat Hazards*. 86(S1):107–139.

Qu Y, Huang C, Zhang P, Zhang J. 2011. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In: CSCW ’11 Proceedings of the ACM 2011 conference on Computer supported cooperative work; p. 25–34.

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. *Nat Hazards*. 76(2):1111–1141.

S¸en Z. 2010. Rapid visual earthquake hazard evaluation of existing buildings by fuzzy logic modeling. *Exp Syst Appl*. 37(8):5653–5660.

S¸en Z. 2011. Supervised fuzzy logic modeling for building earthquake hazard assessment. *Exp Syst Appl*. 38(12):14564–14573.

Shen J, Song H. 2008. Brief introduction on the predominant results of the active fault detecting and seismic risk assessment in Urumqi City. *Seismol Geol*. 30(1):273–288.

Siraj T, Tesfamariam S, Dueñas-Osorio L. 2014. Seismic risk assessment of high-voltage transformers using Bayesian belief networks. *Struct Infrastruct Eng*. 11(7):929–943.

Su G, Qi W, Zhang S, Sim T, Liu X, Sun R, Sun L, Jin Y. 2015. An integrated method combining remote sensing data and local knowledge for the large-scale estimation of seismic loss risks to buildings in the context of rapid socioeconomic growth: a case study in Tangshan, China. *Remote Sens*. 7(3):2543–2601.

Tertulliani A, Arcoraci L, Berardi M, Bernardini F, Camassi R, Castellano C, Del Mese S, Ercolani E, Graziani L, Leschiutta I, et al. 2011. An application of EMS98 in a medium-sized city: The case of L’Aquila (Central Italy) after the April 6, 2009 Mw 6.3 earthquake. *Bull Earthq Eng*. 9(1):67–80.

Tesfamariam S, Liu Z. 2010. Earthquake induced damage classification for reinforced concrete buildings. *Struct Saf*. 32(2):154–164.

Wei B, Nie G, Su G, Sun L, Bai X, Qi W. 2017. Risk assessment of people trapped in earthquake based on km grid: a case study of the 2014 Ludian earthquake, China. *Geomat Nat Hazards Risk*. 8(2):1289–1305.

Wei B, Su G, Liu F. 2013. Public response to earthquake disaster: a case study in Yushu Tibetan Autonomous Prefecture. *Nat Hazards*. 69(1):441–458.
Wieland M, Pittore M, Parolai S, Zschau J, Moldobekov B, Begaliev U. 2012. Estimating building inventory for rapid seismic vulnerability assessment: towards an integrated approach based on multi-source imaging. *Soil Dyn Earthq Eng*. 36:70–83.

Wu H, Cheng Z, Shi W, Miao Z, Xu C. 2014. An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. *Nat Hazards*. 71(1):151–174.

Wu Y, Zhao Q, Zhang B, Wu W. 2017. Characterizing the seasonal crustal motion in tianshan area using GPS, GRACE and surface loading models. *Remote Sens*. 9(12):1303.

Xu C, Xu XW. 2012. Spatial prediction models for seismic landslides based on support vector machine and varied kernel functions: a case study of the 14 April 2010 Yushu earthquake in China. *Chin J Geophys*. 55(6):666–679.

Xu J, An J, Nie G. 2016. A quick earthquake disaster loss assessment method supported by dasymetric data for emergency response in China. *Nat Hazards Earth Syst Sci*. 16(3):885–899.

Yang S, Li J, Wang Q. 2008. The deformation pattern and fault rate in the Tianshan Mountains inferred from GPS observations. *Sci China Ser D-Earth Sci*. 51(8):1064–1080.

Zhong NG, Ying CJ, Qiang LZ, Wu SG, Guo GJ, Min LH. 2002. The construction of basic database for earthquake emergency response. *Earthquake*. 22(3):105–112.

Zhou Y, Liu Y, Wu W, Li N. 2015. Integrated risk assessment of multi-hazards in China. *Nat Hazards*. 78(1):257–280.