

Article

A Novel Method for Seismogenic Zoning Based on Triclustering: Application to the Iberian Peninsula

Francisco Martínez-Álvarez ^{1,†,*}, David Gutiérrez-Avilés ^{2,†}, Antonio Morales-Esteban ³, Jorge Reyes ⁴, José L. Amaro-Mellado ⁵ and Cristina Rubio-Escudero ²

¹ Department of Computer Science, Pablo de Olavide University of Seville, Seville, 41013, Spain;

² Department of Computer Science, University of Seville, Seville, 41012, Spain; E-Mails: davgutavi@alum.us.es (D.G.-A.); crubioescudero@us.es (C.R.-E.)

³ Department of Building Structures and Geotechnical Engineering, University of Seville, Seville 41012, Spain; E-Mail: ame@us.es

⁴ TGT-NT2 Labs, Santiago de Chile, 7820090, Chile; E-Mail: daneel@geofisica.cl

⁵ Department of Graphics Engineering, University of Seville, Seville, 41012, Spain; E-Mail: jamaro@us.es

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: fmaralv@upo.es; Tel.: +34-954977370.

Academic Editor: Raúl Alcaraz Martínez

Received: 23 June 2015 / Accepted: 5 July 2015 / Published: 16 July 2015

Abstract: A previous definition of seismogenic zones is required to do a probabilistic seismic hazard analysis for areas of spread and low seismic activity. Traditional zoning methods are based on the available seismic catalog and the geological structures. It is admitted that thermal and resistant parameters of the crust provide better criteria for zoning. Nonetheless, the working out of the rheological profiles causes a great uncertainty. This has generated inconsistencies, as different zones have been proposed for the same area. A new method for seismogenic zoning by means of triclustering is proposed in this research. The main advantage is that it is solely based on seismic data. Almost no human decision is made, and therefore, the method is nearly non-biased. To assess its performance, the method has been applied to the Iberian Peninsula, which is characterized by the occurrence of small to moderate magnitude earthquakes. The catalog of the National Geographic Institute of Spain has been used. The output map is checked for validity with the geology. Moreover, a

geographic information system has been used for two purposes. First, the obtained zones have been depicted within it. Second, the data have been used to calculate the seismic parameters (b-value, annual rate). Finally, the results have been compared to Kohonen's self-organizing maps.

Keywords: triclustering; seismic zoning; self-organizing maps; Iberian Peninsula

1. Introduction

The goal of a probabilistic seismic hazard analysis (PSHA) is to quantify the rate of exceeding various ground motion levels at a site, given all possible earthquakes. Therefore, PSHA is a methodology that estimates the likelihood that various levels of earthquake-caused ground motions will be exceeded at a given location in a given future time period. The results are expressed as estimated probabilities per unit of time or estimated frequencies [1].

Milne and Davenport [2] developed an empirical statistical method. However, this method has been replaced by the numerical models initiated by Cornell [3]. The most complete information about PSHA can be found in the Senior Seismic Hazard Analysis Committee report [1].

A key point in order to do a PSHA is to define the geometry of the seismic sources. This is depicting the seismic zones in a map and representing the locations within the Earth's crust that have uniform seismicity characteristics.

A seismic source is a region that has relatively uniform seismic characteristics, and it is different from neighboring sources. It is possible to allow for some variation of a- and b-values within a given seismic source. Nevertheless, the maximum magnitude and the earthquake recurrence are assumed to be uniform.

Modern PSHA mainly considers two kinds of seismic sources: seismogenic zones and large faults. In modern PSHA, both models are applied together, forming hybrid models. Up to a threshold magnitude, seismic occurrence is dominated by zones and, for larger earthquakes, by main faults.

The geometry of seismic sources is the aspect of PSHA that depends more on the earthquake environment considered. In high-activity regions, the locations and geometries of seismic sources (faults) are less uncertain than their recurrence rates. The recurrence rates are normally based on geologic data. In low-activity regions (normally of small to moderate magnitude earthquakes), the geometry of the seismic sources is often highly uncertain (area sources). The recurrence rates are obtained from seismic data.

Fault sources must be plotted in a map. The intersection of shallow faults with the ground surface is depicted in a fault map. The shallowest extent of blind surfaces must be also drawn in the map. Relevant faults are main faults with quaternary activity evidence.

Contrariwise, for small to moderate earthquake magnitude zones, it is difficult to identify faults. Due to this, area sources have been conceived of for PSHA. These are a simplified representation of one or more seismogenic structures whose location is unknown [1]. Area sources are defined by their

boundaries depicted on a map. These boundaries classify zones according to their maximum magnitude and recurrence rate. Area source boundaries can be defined by a variety of characteristics, including concentrations of seismicity and changes in tectonic and geologic boundaries. Moreover, it is also advisable to consider the depth distribution of seismicity to identify anomalies relative to other regions. Finally, the focal mechanism must be considered.

Seismogenic zones have traditionally been defined from the spatial distribution of seismic epicenters and shallow geological structures. This procedure is based on the available seismic database. In moderately active zones, the database can be incomplete. Likewise, the geological structures considered are not related to the current tectonic regime many times (paleogeographic criteria, pre-alpine structures, *etc.*). This has caused that the same area can have as many seismogenic zones as authors.

The analysis of the thermal and resistant parameters (thickness variations, pre-existing faults, thermal inhomogeneities, rheological inhomogeneities, constant temperature or constant heat flux, plain strain or plain stress models) of the crust provides a better criterion for the definition of the seismic sources, rather than correlation between epicenters and geologic structures. Nonetheless, there is great uncertainty in the construction of rheological profiles [4].

Morales-Esteban *et al.* [5] studied the seismicity of the Iberian Peninsula (IP). They concluded that its seismicity is moderate, generally characterized by earthquakes less than 5.0. Large earthquakes happen after long periods of time. The seismicity is caused by the convergence between Eurasia and Africa's plates, and it is spread over large areas. These characteristics make it necessary to define seismogenic zones before doing a PSHA.

The goal of this research is to introduce a new seismogenic zoning method. For that purpose, a computational technique based on triclustering, TriGen [6,7], has been applied. Triclustering mines 3D datasets in order to extract patterns of coherent behavior. TriGen is applied to data obtained from the catalog of the National Geographic Institute of Spain (NGIS), due to its ability to find compact areas regardless the spatial coordinates. Later, the obtained zones are correlated with the geology. This way, the bias is greatly eliminated, as the method is solely based on seismic data. Almost no human decision is made. The subsequent correlation with the geology is used to check the validity of the obtained zones. This method could be used in other areas of the world with similar tectonic characteristics (small to moderate earthquakes, individual faults that are hard to identify and spread seismicity).

The rest of the work is structured as follows. In Section 2, some related works regarding both the seismogenic zoning of the IP and the use of triclustering algorithms are described. Section 3 describes the seismicity of the Iberian Peninsula. In Section 4, a description of the methodology used to generate the seismogenic zoning of the Iberian Peninsula is detailed. Section 5 shows the results obtained. A comparison to another zoning method, Kohonen's self-organizing maps, can be found in Section 6. In Section 7, a critical review of the method is provided. Finally, the conclusions can be found in Section 8.

2. Related Works

This section provides a comprehensive discussion of the most relevant works published regarding both previous seismogenic zoning proposed for the Iberian Peninsula (Section 2.1) and the use of triclustering

in other fields (Section 2.2). From its reading, it can be concluded that the latest zoning carried out has not been automatically generated. Furthermore, triclustering is used in this work, for the first time, for a geoscience-related issue.

2.1. Previous Seismogenic Zoning of the Iberian Peninsula

The seismicity of the IP is moderate (from the energetic point of view), but active in some zones. Therefore, in order to do a PSHA, a seismogenic zoning is needed beforehand. A thorough search of previous seismic zoning of the IP in the literature is shown next. Martín [8] defined 27 seismogenic zones for the Iberian Peninsula. The data used for this zoning are based on the catalog of earthquakes and the isoseismal maps of the NGIS. The depicting of the seismogenic zones was done according to geological and homogenous geophysical criteria. Giner *et al.* [9] also used the earthquake catalog with intensities for mainland Spain. Mezcua *et al.* [10] proposed a PSHA for mainland Spain based on zones previously proposed by various authors. Finally, another clustering-based approach was recently applied to the Iberian Peninsula [11].

Partial zoning of the IP is described next. Sanz de Galdeano *et al.* [12] studied the Granada Basin. García-Mayordomo [13] defined five zones for the southeast of Spain. A hybrid model based on faults and area sources was used. Giner *et al.* [9] and Buforn *et al.* [14] zoned southern and eastern Spain. López-Casado *et al.* López-Casado *et al.* [15] provided four zones for the Betic Cordillera, Rif and nearby regions. The *a* and *b* parameters of the GR law were used. The maps were depicted according to the time-division proposed by the authors. This is historic (1390–1910) and instrumental (1910–1992, 1930–1992 and 1960–1992). In the last zoning proposed (1960–1992), only earthquakes of intermediate depth were used. Muñoz and Udías [16] depicted their seismic zoning considering the main geological features, particularly recent fracturing. Finally, a team of experts from Spain and France [17] proposed a unified seismic hazard zoning for the Pyrenean region. Vilanova and Fonseca [18] proposed two zones for the Portuguese mainland and the nearby Atlantic Ocean.

2.2. Use of Triclustering

Triclustering algorithms have been successfully applied to several research fields. A short, but interesting survey on triclustering applied to biological time series can be found in [19]. It might provide a good starting point for researchers for which this research area is novel.

Zhao and Zaki [20] presented the triCluster algorithm in 2005, which was originally devoted to discover patterns in 3D gene expression data. Along with the algorithm, several metrics were also introduced to evaluate its performance. A revised and generalized version, called g-triCluster, was proposed the following year [21]. In this case, the novelty claimed was that more coherent triclusters could be found, as well as being robust to noise. The same algorithm was parallelized and adapted to be used in streaming in [22]. Consequently, a great improvement in terms of computational cost was achieved.

Evolutionary computation has been also applied to the discovery of 3D clusters. In particular, a multi-objective algorithm for optimizing conflicting constraints was introduced in 2008 [23], with application to gene expression data.

Another strategy to discover triclusters in continuous data can be explored in [24]. In particular, the authors were concerned about discovering subspaces with a significant number of items, one of the main drawbacks typically found in approaches based on triclustering. At the same conference, another approach focusing on triclustering to analyze continuous data was presented [25]. The main goal reached by the authors was the discovery of triclusters with low variance.

In 2015, a novel approach for dealing with fuzzy triclusters was proposed [26]. The experimental studies were carried out over MovieLens data. Results were compared to other co-clustering methods showing remarkable improvement in terms of accuracy.

The work in [27] introduced an approach to deal with non-negative matrix tri-factorization, in order to enhance the performance of three factors as a co-clustering method. It was applied to several common datasets, achieving better inter-triclustering distances for all of them.

Ignatov *et al.* [28] presented several definitions of optimal patterns for experimental comparison of five triclustering algorithms. Their extensive experimentation let the reader choose the most suitable solution for a given dataset, by means of the principle of Pareto-optimality.

Finally, following the increasing popularity of the big data field, a novel, efficient algorithm, in terms of time and memory, with reduced computational cost, was presented in [29]. The authors also affirmed that the algorithm can be easily parallelized to become adapted to the new big data programming paradigms.

3. Seismicity and Tectonics of the Iberian Peninsula

The IP (Spain, Portugal, Andorra and Gibraltar) is located in the Eurasian plate next to the contact with African's plate. The seismicity is moderate [5,30], and it is associated with convergence between these plates. Likewise, it is distributed over a wide area of deformation, as expected in a continent-continent collision. Long recurrence periods separate large events [31,32]. Most earthquakes are at a shallow depth ($h < 40$ km). However, there is a significant seismic activity at intermediate ($40 < h < 130$ km) and greater depths ($h \leq 650$ km) [33]. Historical seismicity studies report the occurrence of large earthquakes in the south of Iberia. Instrumental magnitudes larger than 5.5 have mostly occurred in the Gulf of Cadiz and north of Africa. The largest recent earthquake was the Goringe Bank (1969, $M_s = 8.0$), and the largest earthquake in the Gulf of Cadiz (1964) was of a magnitude 6.2 [30].

The plate limit is not homogenous, with constant oceanic and continental areas in contact. There is a consistent N-S to NW-SE orientation of P axes that corresponds to the plate convergence direction. This coexists with E-W- to NE-SW-directed horizontal tension in the Betics, the Alboran Sea and northern Morocco. The area of contact between the IP and the northwest of Africa is the most complicated area. This area is surrounded on both sides by frequent seismic activity with very large earthquakes [34]. Large earthquakes, such as the Lisbon 1531, 1755 and 1909 earthquakes, have happened in this area.

The regional seismicity is diffuse and is not clearly aligned with the current limit between the two plates [35]. The seismic activity extends to far away inter-plate areas, such as the northeast and the center of the IP.

Great seismic activity is located in the east of the Gibraltar Arc. It is spread over a large area approximately 500 km wide, centered at the Alboran Sea. It contains parts of the southeast of Spain, the

north of Morocco and Algeria [36]. At the west of Gibraltar, most earthquakes happen in the southern coast of Portugal, in the surroundings of the limit between the Azores and Gibraltar plates. Other seismic sources include the northwest of Spain, southern Portugal and the Pyrenees. Earthquakes are infrequent in other locations. This seismicity happens over several geo-tectonic regions with different rheology and tectonics.

There are three main tectonic regimes at the IP: first, the Hercynian block, known as the Iberian Massif, inactive Variscan orogeny; second, the Atlantic continental margin off Portugal and Spain; and third, the Alpine young orogenic belts formed by the Pyrenean fold and the Betic foldchain.

The core of the IP consists of a Hercynian block, known as the Iberian Massif. It is formed by various pieces that were assembled to form the block. The external part of the orogeny is the Cantabrian Zone. This is just deformed in the upper crustal layers. Contrariwise, the west Asturian Leonese Zone and Central Iberian Zone layers are more deeply deformed and metamorphosed. These three zones are part of a terrane. The other two zones, Ossa-Morena and south Portugal, are two different terranes that have become fastened. The Iberian Massif was marginally affected by the Alpine deformation. It is formed by folded Paleozoic and Precambrian rocks partially covered by Mesozoic and Tertiary deposits to the east. Nonetheless, it outcrops through the Iberian Chain and the Catalanian Coastal Ranges.

The IP is delimited to the west by the continental boundary formed by the magma-poor opening of the Atlantic Ocean. It is a unique 100 km-wide zone of exhumed continental mantle.

The Alpine orogenic belts are formed with thickening crust. They are located in the old limit between the African and the Iberian plates (the Pyrenees), on the northeast of the Iberian Massif. Currently, it is located in the limit between Africa and the Iberia plates (the orogenic belt Betica-Riff-Tell), on the southeast of the Iberian Massif. The inner depressions of the Cenozoic extensive tectonic regime include the folded depressions in the lower areas of the Alpine belt (Ebro depression at the south of the Pyrenees and the Guadalquivir depression at the north of the Betica), the larger inner depressions of the Tajo and the Duero and, in a lesser number, the intra-orogenic depressions in both the Hercynian and Alpine regions. Farther away, an local-scale extensive tectonic regime is active in the western part of the Mediterranean Sea. It is located in the Liguro-Provencal depression, Valencia depression, the depression at the south of the Balearic Islands located in the east of the IP and the Alboran depression at the south of Iberia [37].

The Alboran Sea is from the lower Pliocene and is located in the limit between Eurasia and Africa plates. It is especially relevant for the proper understanding of the collision between plates. The continental crust in the Alboran depression has become considerably thinner (15 km compared with over 35 km at the center of the Betica [38,39]). The transition between the Alboran depression and the Neogene oceanic crust to the east is gradual [40]. The formation of the Alboran depression in a compressive stress regime can be explained by the elimination of the lower parts of the crust; nevertheless, the mechanism is still not clear.

4. Method

This section describes the methodology used to generate the seismogenic zoning of the Iberian Peninsula. In this sense, the features of the database used, as well as all information related to the

catalog are introduced in Section 4.1. Section 4.2 describes the data processing carried out in this work. Finally, a brief description of the TriGen algorithm can be found in Section 4.3.

4.1. Data Description and Retrieval

The database used for this study is the catalog of the NGIS. From 1961, the NGIS has routinely calculated the magnitude of the earthquakes placed between 35°N to 44°N and 10°W to 5°E. It also produces weekly, monthly and definitive catalogs from 1981 until today.

It uses five different correlations for the estimation of the magnitudes:

- (1) Duration magnitude, $M_D(M - M_S)$.
- (2) Surface-wave magnitude, $m_{b,L_g}(M - M_S)$.
- (3) Body-wave magnitude, $m_b(V-C)$ [41].
- (4) Surface-wave magnitude, $m_{b,L_g}(L)$.
- (5) Moment magnitude, M_w [42].

A thorough description of these correlations is shown in Morales-Esteban *et al.* [5]. The estimation of the magnitudes in the catalog is not homogenous, as the earthquakes before 1962 (M_D) were calculated using a different procedure. In order to work out the b-value from Equations (4) and (5), the database must be complete. To guarantee the completeness of the data, the analysis should only include earthquakes recorded after the year of completeness of the catalog. This is the year from which all of the earthquakes of a magnitude larger to or equal to the minimum magnitude have been recorded. The works in [43,44] point out that the year of completeness of the catalog is 1978 and that the minimum magnitude is 3.0. Only earthquakes from 1978 and onwards are used in this study. The surface-wave magnitude (m_b, L_g , hereinafter M) has been used in this study.

The procedure used for the location of the epicenters is described in Mezcua *et al.* [45]. Older epicenters have been determined using isoseismal maps. Later, the application HYPO 71, based on the time of arrival of the seismic waves to the stations and a model of the crust, has been used.

4.2. Data Transformation

Before TriGen can be applied, the 2D database must be transformed into a 3D one. Moreover, several new attributes must be generated, so that the inputs can provide meaningful information.

First, all data included in the catalog have been sorted into 60×90 cells, representing each an area of approximately 20×20 km². The next step is to characterize each cell in terms of a set of features. According to Reyes and Cárdenas [46], Scitovski and Scitovski [47], who performed a Chilean and Croatian seismic zonification in 2010 and 2013, respectively, the features generated are:

- (1) Total number of earthquakes occurring, N .
- (2) Number of earthquakes occurring with a magnitude larger or equal to 3.5 in the cell considered, $M_{3.5}$.

- (3) Number of earthquakes occurring with a magnitude larger or equal to 4.0 in the cell considered, $M_{4.0}$.
- (4) Number of earthquakes occurring with a magnitude larger or equal to 4.5 in the cell considered, $M_{4.5}$.
- (5) Number of earthquakes occurring with a magnitude larger or equal to 5.0 in the cell considered, $M_{5.0}$.
- (6) Mean depth of all earthquakes' epicenters recorded in the cell, \overline{D} .
- (7) Maximum earthquake magnitude recorded in the cell, M_{max} .

Therefore, every cell is characterized by:

$$C_{i,j} = \{N^*, M_{3.5}^*, M_{4.0}^*, M_{4.5}^*, M_{5.0}^*, \overline{D}^*, M_{max}^*\} \tag{1}$$

where $i \in [1, 90]$ and $j \in [1, 60]$ denote the relative latitude and longitude. Cell (1, 1) corresponds to coordinates (12°W, 33°N), whereas cell (90, 60) stands for coordinates (6°E, 45°N). Figure 1 depicts the data structure before being processed by TriGen.

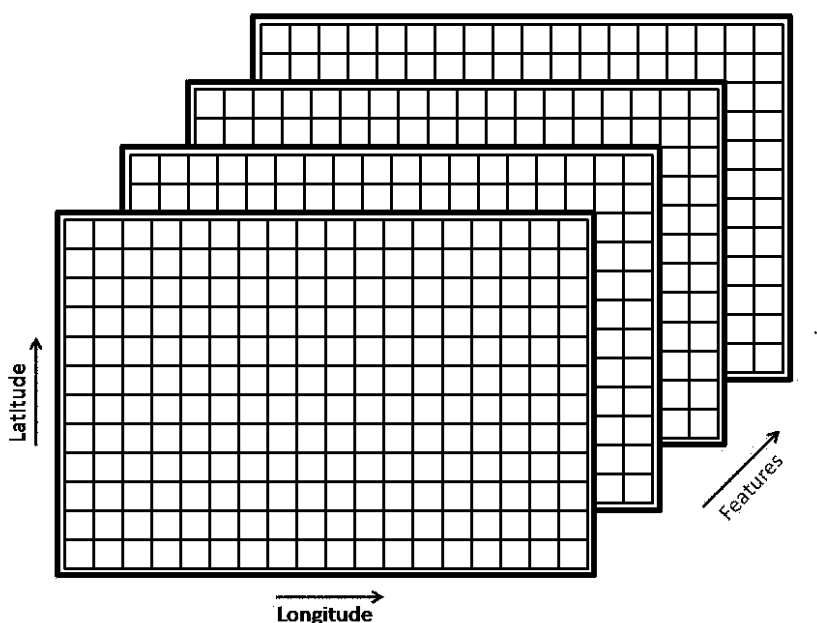


Figure 1. Data transformation to serve as TriGen’s inputs.

It is worth noting that all of the aforementioned features have been calculated as follows. Let $C_{i,j}$ be the target cell. To calculate every feature, the 24 closest surrounding cells and $C_{i,j}$ itself are considered. That is, the set of cells included in the square with boundaries $C_{i-2,j-2}$ (lower left corner) and $C_{i+2,j+2}$ (upper right corner) is used to estimate all features for cell $C_{i,j}$. The cells involved in the calculation of a target cell are illustrated in Figure 2.

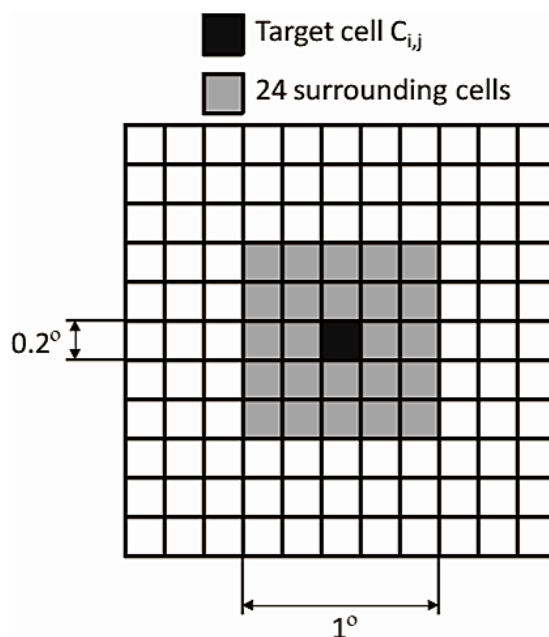


Figure 2. Twenty-five cells used to calculate the features for a target cell.

4.3. Description of the TriGen Algorithm

Triclustering is a computational technique in the data mining field. The entire catalog of techniques and procedures of data mining have a common and general goal: extracting hidden, valid and useful information from large volumes of data. Triclustering is the evolution of a widely-known computational technique, clustering, which involves grouping data based on features' similarity. Clustering works in a 1D space, while triclustering works in a 3D space.

TriGen [6] is an algorithm that applies a triclustering technique based on the paradigm of evolutionary algorithms, in particular genetic algorithms [48]. Genetic algorithms perform a heuristic search that mimics the process of natural selection, since a population of individuals is created and evolves throughout several generations using crossover and mutation operators, and the individuals are evaluated based on a function to optimize. For this particular problem, each individual represents a zoning proposal for the IP; the genetic materials of each individual are the longitude and latitude coordinates; and the fitness function maximizes the similarity among all seven features for each cell contained in the zoning. All zones contained in the population are evolved, and after some generations, the optimal one is selected. The flowchart of the algorithm is illustrated in Figure 3.

TriGen's steps can be described as: the population initialization step in which the initial population will be created taking into account overlap with previously found solutions; an evaluation step in which each individual will be evaluated using the fitness function; a selection step, which serves to decide which individuals will survive to the next generation; crossover, which creates the necessary connections between pairs of individuals to share new genetic material; and finally, mutation, which performs punctual changes to individuals to ensure genetic variability of future generations.

Each of these operators are now briefly described; for further details please refer to [6]. The initial population is randomly generated. Part of the individuals are purely randomly generated, that is a random

subset of longitudes, latitudes and features are assigned. The rest of the individuals are also randomly generated, but paying attention to overlapping with previously found solutions. This is done to promote the visiting of non-explored areas and the widest area possible of the database.

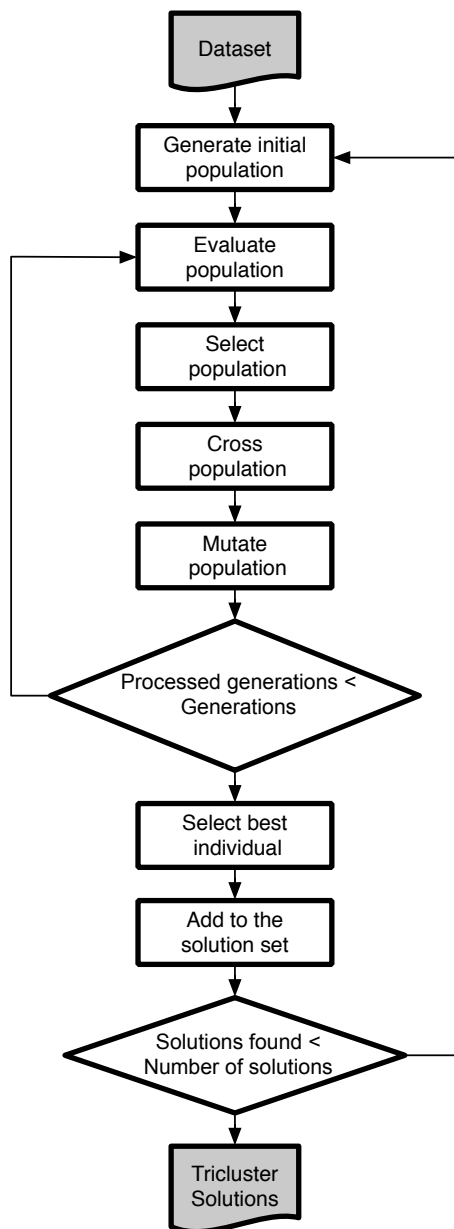


Figure 3. TriGen algorithm flowchart.

As for the crossover operator, two individuals (or parents) are combined to create two new individuals (or offspring). They are chosen based on a probability of crossover parameter, denoted by p_c . Their genetic materials (latitude and longitude) are randomly combined, mixing the coordinates in both children.

An individual can be mutated according to a probability of mutation, p_m . The p_m condition is verified for every individual, and if it is satisfied, one out of six possible actions is taken. These actions are: add/remove a new random latitude coordinate or add/remove a new longitude coordinate.

The selection operator was implemented following the roulette wheel selection method [49].

Another key question is the selection of the fitness function. The fitness of each individual allows the algorithm to determine the best candidates to remain in subsequent generations. For this particular case, the authors in [6] proposed a measure based on a three-dimensional adaptation of the mean square residue measure (MSR), which is a classical clustering measure for gene expression analysis [50], referred to as f_{MSR} . This measure maximizes the coherence in the behavior of the seven features for all of the cells contained in the zoning. Additionally, the same authors proposed an evaluation measure for triclusters called the multi-slope measure, based on the similarity among the angles of the slopes formed by each profile formed by the genes, conditions and times of the tricluster in [51].

5. Results

This section describes the results achieved by the application of TriGen to the Iberian Peninsula earthquake database. First, the TriGen parameter setup for this particular case is shown. Then, the zones generated by TriGen are introduced. Finally, the results are correlated with the geology for validation.

5.1. Applying TriGen to Discover Seismogenic Zones

TriGen has been applied to the transformed database described in Section 4.2. By applying this algorithm, we aim to find meaningful patterns for the whole search space, in this case the Iberian Peninsula. In particular, the discovery of zones with similar geophysical properties among the 5400 cells (60×90) is the output of such an algorithm.

Note that TriGen was originally conceived of to deal with temporal data, that is the third dimension being time. In this context, the third dimension is composed of the seven features calculated for every latitude and longitude. In other words, these seven features could be considered seven time stamps for the tuple $\{longitude, latitude\}$ introduced in [6]. Since TriGen is a general-purpose algorithm, it can be applied to any problem formulated on three dimensions. The interpretation of each dimension is up to the context of application.

Table 1. TriGen's parameter configuration.

Parameter	Value
Number of solutions	10
Generations	200
Number of individuals in the population	100
Selection rate p_c	0.5
Mutation probability p_m	0.3
Presence of longitude w_{lgt}	0.4
Presence of latitude w_{ltt}	0.4
Presence of features w_f	1.0
Overlapping prevention in longitude wd_{lgt}	1.0
Overlapping prevention in latitude wd_{ltt}	1.0
Overlapping prevention in features wd_f	0.0

TriGen can find triclusters, including different numbers of features. However, in this particular research area of the IP with a similar behavior, all seven features under study are intended to be found. Therefore, the algorithm has been forced to generate triclusters including all of the features.

Table 1 shows the selected TriGen setup parameters. The size of the longitude and latitude dimensions are very similar (60 and 90, respectively). Consequently, the same weight for w_{lgt} and w_{lnt} , 0.4, has been set. The weight for w_f has been set to 1.0, since it is intended that all features appear in all solutions. The overlapping prevention values have been set to 1.0 for longitude, wd_{lgt} , and latitude, wd_{lnt} , since all of the cells in the IP dataset have to appear in the solution in order to partition the whole space. Nonetheless, the overlapping prevention value for features, wd_f , has been set to 0.0 to allow all features to appear in all solutions.

5.2. Zones Obtained by the Method and Their Representation in a Geographic Information System

The definition of a geographic information system (GIS) is still an open issue of discussion among experts. Its own name provides some answers. Geographic relates to attributes and spatial relationships of positions in space. Information is the useful data that are attached to these positions. System suggests numerous components, which in the case of a GIS are users, hardware, software and data. Therefore, in one sense, a GIS is simply a software package. In another, it is a computer database containing information about the Earth. In yet another sense, a GIS is an integrated system of software and geo-referenced data for planning [52].

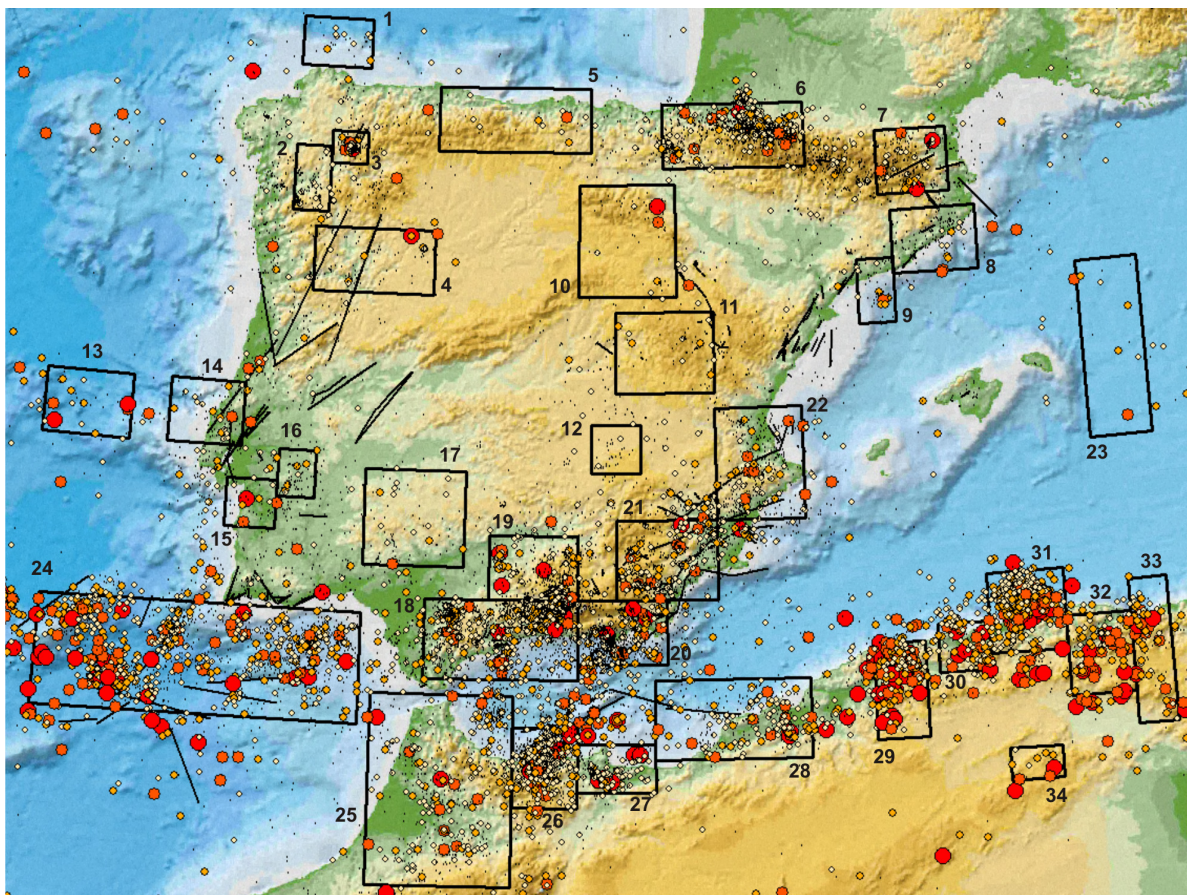


Figure 4. Seismic zones of the Iberian Peninsula (IP) obtained with triclustering.

A GIS has been created with the data from the Spanish catalog over a physical map. The known quaternary faults, which are the ones that generate seismic activity, have also been plotted. The location of these faults has been provided by the Spanish Mining and Geological Institute. Figure 4 depicts the 34 zones obtained and their representation in the GIS. Earthquakes have been characterized as described next: earthquakes of a magnitude in $[2.5, 3.4]$ as points; $[3.5, 3.9]$ in yellow circles; $[4.0, 4.4]$ in light-orange circles; $[4.5, 4.9]$ in orange circles; and ≥ 5 in red circles. Moreover, circles are bigger according to magnitude. The good correlation obtained between the zones and the seismic activity can be observed. Table 2 lists the geological description of every zone. The advantage of creating a GIS is that a database can be obtained from the information of the map. In this case, the b-value and the annual rate by km^2 for every zone have been calculated from it.

Table 2. Seismogenic zones proposed for the Iberian Peninsula.

Zone	Description	b-value	Annual rate/ km^2
1	Zone to the north of Galicia	1.06	2.77×10^{-3}
2	Center and south of Galicia	1.95	1.36×10^{-2}
3	East of Galicia	1.33	1.39×10^{-1}
4	Trás-Os-Montes	1.26	2.63×10^{-3}
5	Cantabrian Mountains	1.15	9.60×10^{-4}
6	West of the Pyrenees	1.30	2.04×10^{-2}
7	East of the Pyrenees	1.45	1.41×10^{-2}
8	North of the Catalan Coastal Range	1.26	2.92×10^{-3}
9	South of the Catalan Coastal Range	0.95	4.03×10^{-3}
10	North of the Iberian mountain mass	3.10	7.24×10^{-4}
11	Center of the Iberian mountain mass	1.32	2.05×10^{-3}
12	South of the Iberian mountain mass	2.42	5.27×10^{-3}
13	Zone to the west of Portugal	0.61	2.40×10^{-3}
14	Mesozoic basins on the west Portugal fringe	1.22	3.89×10^{-3}
15	Cenozoic basins on the lower Tagus Basin	1.12	3.62×10^{-3}
16	Cenozoic basins on the Tagus Basin	1.16	7.05×10^{-3}
17	Ossa-Morena zone	1.05	1.56×10^{-3}
18	Southwest of the Betic	1.30	2.31×10^{-2}
19	North of the Betic	1.31	1.25×10^{-2}
20	Southeast of the Betic	1.26	2.72×10^{-2}
21	Northeast of the Betic	1.24	1.18×10^{-2}
22	Valencia Basin	1.20	1.57×10^{-3}
23	Zone to the east of the Balearic Islands	0.56	6.68×10^{-4}
24	Azores-Gibraltar fault	0.79	1.58×10^{-2}
25	Rif	0.91	6.44×10^{-3}
26	East of the Rif	0.91	1.27×10^{-1}
27	North of the middle Atlas	0.87	3.50×10^{-2}
28	Western part of the Tell	0.73	4.66×10^{-3}
29	West of the Tell	0.44	2.70×10^{-2}
30	Northwest of the Tell	0.54	1.92×10^{-2}
31	North of the Tell	0.47	7.69×10^{-2}
32	Northeast of the Tell	0.46	7.38×10^{-3}
33	East of the Tell	0.45	7.65×10^{-3}
34	South of the Tell	0.34	3.07×10^{-3}

5.3. Correlation with Geology

Once the method has been applied and the zones have been obtained, the correlation of the zones with the geology must be verified. In this sense, Figure 5 illustrates the annual rate of earthquake occurrence, as well as the associated b-value for every zone discovered by TriGen.

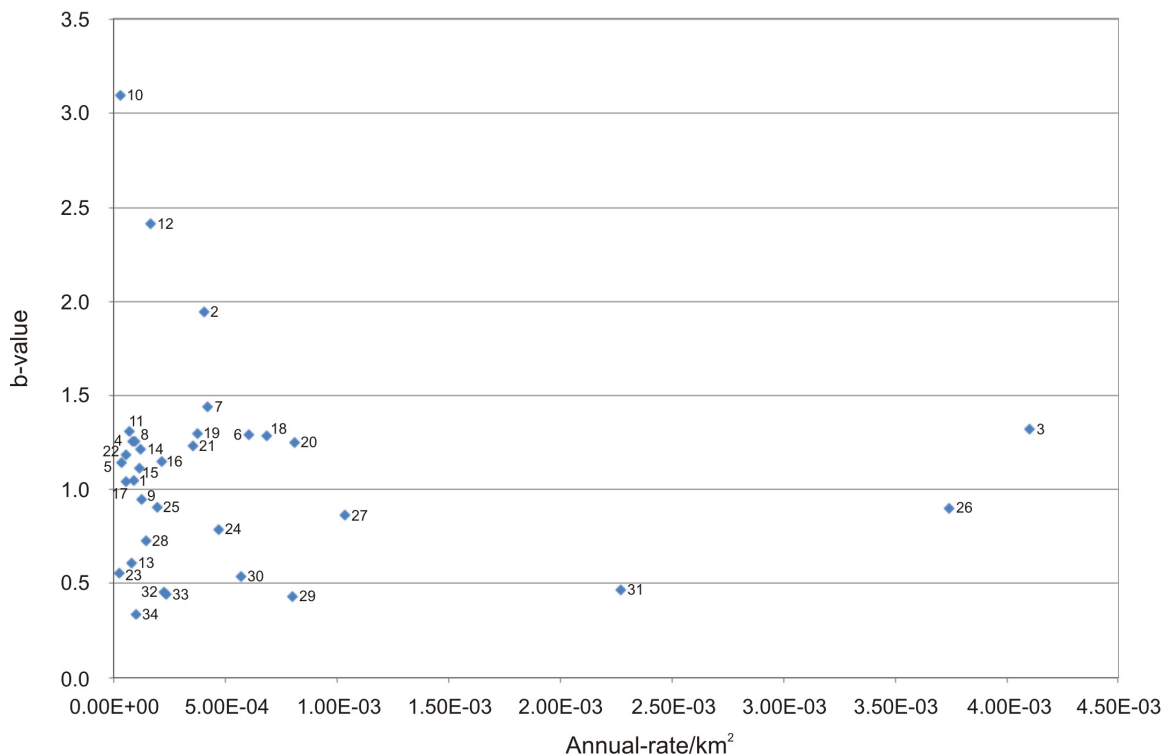


Figure 5. Earthquake annual rate and b-value associated with every zone.

Zone 1 (Z1) is a zone of moderate seismicity and average b-value. It is placed in the north of Galicia. In Z2, the center and south of Galicia, large earthquakes are improbable, and the rate of events is moderate. Contrariwise, in Z3 (east of Galicia), large earthquakes are more probable, and this is the zone with the highest annual rate of earthquakes by km². Z4 corresponds to Trás-Os-Montes and is a zone of moderate seismicity.

In the Cantabrian Mountains (Z5), earthquakes are infrequent and of moderate magnitude. Z6 and Z7 correspond to the west and the east of the Pyrenees, respectively. They present a similar annual rate of earthquakes and b-value. The Catalan Coastal Range is seismically divided into two: north (Z8) and south (Z9). The southern one has a slightly more frequent annual rate of earthquakes and the lowest b-value.

The Iberian mountain mass is separated into three zones: north (Z10), center (Z11) and south (Z12). The annual rate of earthquakes is moderate, and large earthquakes are improbable. Z13 is a zone located in the west of the coast of Portugal. Large earthquakes are frequent. However, its annual rate is low.

Z14 corresponds to the Mesozoic basins on the west of the Portugal fringe, and it is placed in the west of Z13. Contrarily, large earthquakes are not so frequent. Z15 and Z16 are placed on the Cenozoic basins of the lower Tagus Basin and Tagus Basin, respectively. The b-value is around one, and the annual rate of earthquakes is moderate. Nevertheless, the annual rate of Z16 is twice the rate of Z15.

The Ossa-Morena zone (Z17) presents moderate seismicity. The Betics are divided into four zones (Z18, Z19, Z20 and Z21). It can be observed that the b-value ranges between 1.24 and 1.31. The annual rate of earthquakes is quite high. The zones located in the south (Z18 and Z20) and closer to the contact area between the plates have a higher annual rate of earthquakes.

Z22 is the Valencia Basin and has a moderate seismicity. Z23 is placed on the east of the Balearic Islands. It presents a very low b-value, so the relative number of small and large events is similar. Nonetheless, not many earthquakes hit that zone.

Z24, the Azores-Gibraltar fault, is a very relevant zone, as large earthquakes are known to have occurred within it. Moreover, it is very close to the south of Portugal and the southwest of Spain, and large earthquakes are known to have affected the inland. It is the largest zone depicted and presents a high seismicity rate. Z25 and Z26, the Rif and the east of the Rif, respectively, have the same b-value, but the annual rate of Z26 is much higher. Z27 is located to the east of Z26. It also presents a similar b-value to the two previous ones. It has an outstanding annual rate, but not as high as the east of the Rif zone. To the east, the b-value declines. It is 0.73 for Z28. This zone corresponds to the western part of the Tell.

The b-values of Z29, Z30, Z31, Z32 and Z33 range between 0.44 and 0.54, which points out that large earthquakes are frequent. If the annual rate of earthquakes is examined, it can be observed that it is similar for Z29 (west of the Tell) and Z30 (northwest of the Tell). Maybe they could be joined into a single zone. The most active zone of this set is Z31 (the north of the Tell), one of the most seismically-hazardous zones examined in this paper. Similarly to Z29 and Z30, the northeast of the Tell (Z32) and the east of the Tell (Z33) have an almost coincident b-value and annual rate, which suggest that they could form a single zone. Finally, Z34, placed at the south of the Tell, has the lowest b-value worked out: 0.34.

6. Comparison to Kohonen's Self-Organizing Maps

This section is to compare the zoning proposed with the zoning obtained with Kohonen's self-organizing maps (SOM). SOM [53] is a special neural network suitable for spatial distribution discrimination. It is trained using unsupervised learning. It generates maps based on a low-dimensional (typically two dimensions) representation of the input space.

The network must be fed with a large number of training vectors that represent the differences exhibited among data during the mapping. In this way, SOM can find the best distribution of zones, based on these training vectors. Since SOM was successfully applied to generate seismic hazard maps in Chile [46], it has been applied to assess its performance for the Iberian Peninsula. The same input features used for TriGen (see Section 4.3) have been used. Three thousand epochs of 5400 iterations have been used. This number was empirically determined since the use of 1000, 2000, 3000, 5000 and 10,000 epochs did not substantially change the results, and 3000 epochs was advised in [46].

The number of clusters to be generated is still an open question, and there is no general agreement to determine the optimal number for every case. For this reason, the proposal introduced in [54] has been applied. This approach combines three well-known indices (silhouette, Davies–Bouldin and Dunn) in a majority vote system. Such indices have been selected due to the different nature of the objectives

they maximize. The application of this procedure to the transformed database determined that the best number of partitions is 10.

Figure 6 depicts the map obtained when the 7D vectors have been divided into 10 classes. Note that these seven classes did not include the spatial coordinates, longitude and latitude, as input features. Many isolated zones have been labeled similarly. This means that far away zones present similar magnitude histograms. A common procedure in geology is softening the borders of these maps. However, as mentioned before, this procedure is not done in this study.

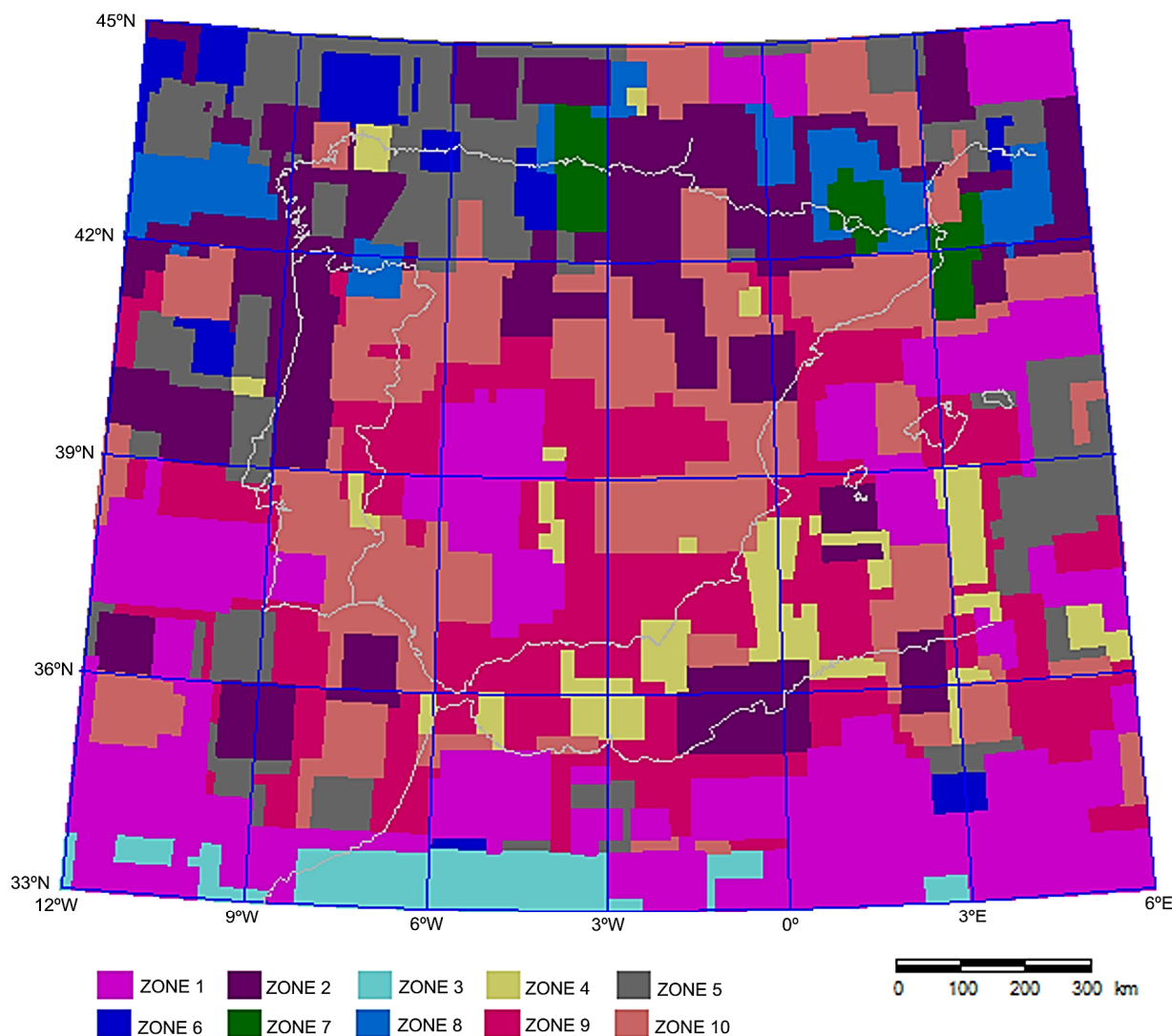


Figure 6. Self-organizing maps (SOM) applied to the Iberian Peninsula without using spatial coordinates as input features, for 10 clusters.

It can be observed that there is not a continuous structure among the zones found by SOM, and isolated and small zones have been generated with no apparent correlation with the geology. Obviously, these results cannot be considered satisfactory, since they should delimitate particular areas with meaningful information, which is not the case.

To improve the quality of the results, a second scenario has been considered. Similar to [46], the coordinates of every cell have also been used as input features (see Figure 7). This new scenario provided a more compact zoning.

If the map shown in Figure 7 is examined in detail, some vague correlations with geology can be made. Zones 1 and 6 in the south of the IP correspond to the Azores-Gibraltar fault and the north of the Africans plate. Zones 1 and 7 in the center of the IP show the low seismic activity zones of the Central Massif and the Mesozoic basins. It can be observed that Zone 5 coincides with the Alboran Sea, the southeast of the IP and the south of the Balearic Islands. The Pyrenees present a very complex structure represented by labels 2, 9 and 10, similarly to the zoning in [43]. The Azores-Gibraltar fault, Z24 in this study, has been characterized by labels 1, 5 and 8. Seismicity in Portugal was mainly labeled 7 and 8.

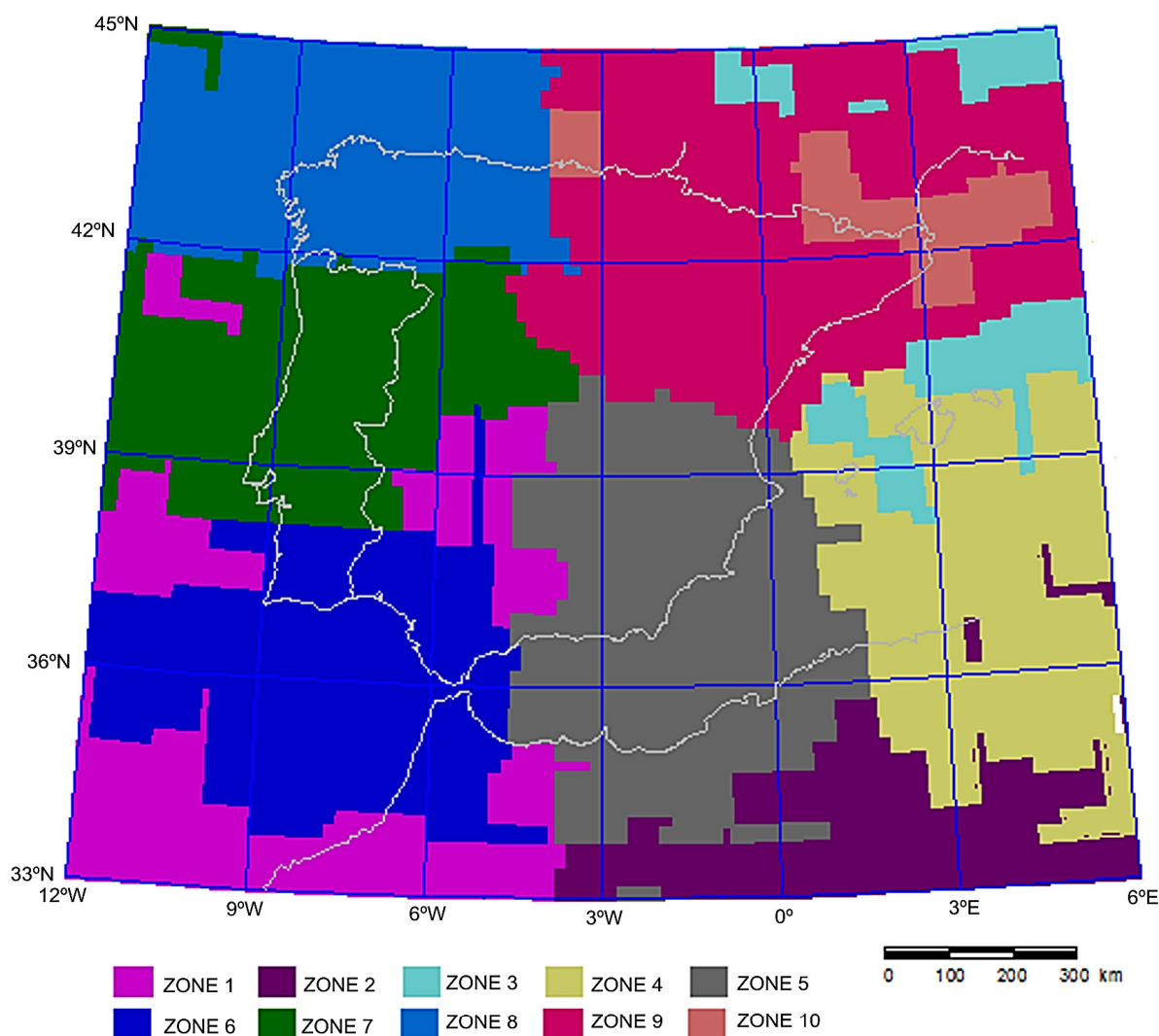


Figure 7. SOM applied to the Iberian Peninsula using spatial coordinates as input features, for 10 clusters.

It can be observed that the results provided by the SOM are not as satisfactory as expected. In [46], the authors properly zoned the whole of continental Chile into just six zones. The coordinates for every

cell have been used as inputs, and the division was done considering cells of $1^\circ \times 1^\circ$. In this study, 34 zones have been considered for the IP, and the size of the cells is $0.2^\circ \times 0.2^\circ$.

This could mean that SOM is only valid for zoning high seismic activity areas, such as Chile, into a small number of zones. However, for zoning the IP, a larger number of zones is required. This is due to the fact that the seismicity of the IP is spread over a wide area of deformation. A larger number of zones is required owing to its complexity and a smaller-sized cell is needed.

7. Critical Discussion

The objective of this section is to provide a critical review of the method. The main features of the method and its application to be IP are discussed.

A key issue on seismic zoning is that every zone must be different from the adjacent ones. In order to obtain accurate results, the b-value must be estimated properly. Therefore, only earthquakes after the year of completeness of the catalog must be used. Furthermore, all of the magnitudes must have been calculated with the same procedure. For that purpose, only earthquakes from 1978 onwards have been used in this study. However, a catalog that gathers the largest possible earthquakes must be of a very long duration, probably of hundreds of years. In this research, the catalog has been limited to the 34-year period between 1978 and 2013. The historical seismicity has not been considered. The previous zoning proposed for the IP, described in Section 2.1, used the historical seismicity. For the authors of this research, it is more important to present a robust method than its application to the IP. This is the first step in this sense.

The method only generates rectangular zones, a restriction associated with the types of solutions that TriGen provides. Further research must be done in this sense. Nevertheless, from the analysis exposed in Section 5.3 and from Figure 4, it can be concluded that the results are consistent and that areas of different seismicity have been delimited. Moreover, they correspond with different tectonic units or parts of a tectonic unit with different seismic activity.

The main advantage of this method is that the depicting of the zones is done automatically by the algorithm and that almost no human decision is necessary.

8. Conclusions

A novel seismogenic zoning method has been proposed in this work. To accomplish such a task, a computationally-efficient method based on triclustering has been used. It has been applied to the IP in order to observe its performance. One of the main achievements of this work lies in having avoided the subjective interpretation of experts, as usually happens in this context. That is, the new proposed zones share similar histograms and physical properties and, intentionally, have not been smoothed. To assess TriGen's performance, the well-known self-organizing maps have also been applied to the Iberian Peninsula data for two different scenarios, since they have already been used satisfactorily in other areas. Although the results obtained showed a certain degree of correlation with the underlying geology, they could not be considered as satisfactory as those obtained with TriGen.

Acknowledgments

The authors would like to thank the Spanish Ministry of Economy and Competitiveness, Junta de Andalucía and Pablo de Olavide University for the support under Projects TIN2014-55894-C2-2-R, P12-TIC-1728 and APPB813097, respectively.

Author Contributions

The experimental setup was designed by Francisco Martínez-Álvarez, Antonio Morales-Esteban and Jorge Reyes. David Gutiérrez-Avilés and Cristina Rubio-Escudero carried out experiments. The geographic information system was designed and created by José L. Amaro-Mellado. All authors contributed to the writing of the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Committee, S.S.H.A. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts. *US Nuclear Regulatory Commission Report* **1997**, pp. CR-6372.
2. Milne, W.G.; Davenport, A.G. Distribution of earthquake risk in Canada. *Bull. Seismol. Soc. Am.* **1969**, *59*, 729–754.
3. Cornell, C.A. Engineering seismic risk analysis. *Bull. Seismol. Soc. Am.* **1968**, *58*, 729–754.
4. Fernández, M.; Ranalli, G. The role of rheology in extensional basin formation modelling. *Tectonophysics* **1997**, *282*, 129–145.
5. Morales-Esteban, A.; Martínez-Álvarez, F.; Reyes, J. Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence. *Tectonophysics* **2013**, *593*, 121–134.
6. Gutiérrez-Avilés, D.; Rubio-Escudero, C.; Martínez-Álvarez, F.; Riquelme, J.C. TriGen: A genetic algorithm to mine triclusters in temporal gene expression data. *Neurocomputing* **2014**, *132*, 42–53.
7. Gutiérrez-Avilés, D.; Rubio-Escudero, C. Mining 3D patterns from gene expression temporal data: A new tricluster evaluation measure. *Sci. World J.* **2014**, *ID624371*, 1–16.
8. Martín, A.J. Riesgo sísmico en la Península Ibérica. Ph.D. Thesis, Polytechnical University of Madrid, Spain, 1984.
9. Giner, J.J.; Molina, S.; Delgado, J.; Jáuregui, P.J. Mixing methodologies in seismic hazard assessment via a logic tree procedure: an application for Eastern Spain. *Nat. Hazards* **2003**, *25*, 59–81.
10. Mezcuá, J.; Rueda, J.; García-Blanco, R.M. A new probabilistic seismic hazard study of Spain. *Nat. Hazards* **2011**, *59*, 1087–1108.
11. Morales-Esteban, A.; Martínez-Álvarez, F.; Scitovski, R.; Scitovski, S. A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Comput. Geosci.* **2014**, *73*, 132–141.

12. Sanz de Galdeano, C.; Peláez, J.A.; López-Casado, C. Seismic potential of the main active faults in the Granada Basin (southern Spain). *Pure Appl. Geophys.* **2003**, *160*, 1537–1556.
13. García-Mayordomo, J. Considering geological data and geologically based criteria in seismic hazard analysis of moderate activity regions: I. Definition and characterization of seismogenic sources. *Geogaceta* **2007**, *41*, 87–90.
14. Buforn, E.; Benito, B.; Sanz de Galdeano, C.; del Fresno, C.; Muñoz, D.; Rodríguez, I. Study of the damaging earthquakes of 1911, 1999 and 2002 in the Murcia, Southeastern Spain, region: seism-tectonic and seismic-risk implications. *Bull. Seismol. Soc. Am.* **2005**, *95*, 549–567.
15. López-Casado, C.; de Galdeano, C.S.; Delgado, J.; Peinado, M.A. The b parameter in the Betic Cordillera, Rif and nearby sectors. Relations with the tectonics of the region. *Tectonophysics* **1995**, *248*, 277–292.
16. Muñoz, D.; Udías, A. Earthquake occurrence and seismic zonation in South Spain. In Proceedings of the World Conference On Earthquake Engineering, Madrid, Spain, 19-24 July 1992; pp. 483–487.
17. Secanell, R.; Bertil, D.; Matin, C.; Goula, X.; Susagna, T.; Tapia, M.; Dominique, P.; Carbon, D.; Fleta, J. Probabilistic seismic hazard assessment of the Pyrenean region. *J. Seismol.* **2008**, *12*, 323–341.
18. Vilanova, S.P.; Fonseca, F.B.D. Probabilistic seismic hazard assessment for Portugal. *Bull. Seismol. Soc. Am.* **2007**, *97*, 1702–1717.
19. Mahanta, P.; Ahmed, H.A.; Bhattacharyya, D.K.; Kalita, J.K. Triclustering in gene expression data analysis: A selected survey. In Proceedings of the IEEE International Conference on Emerging Trends and Applications in Computer Science, Shillong, India, 4–6 March 2011; pp. 1–6.
20. Zhao, L.; Zaki, M.J. TRICLUSTER: an effective algorithm for mining coherent clusters in 3D microarray data. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Baltimore, MD, USA, 13–17 June 2005; pp. 694–705.
21. Jiang, H.; Zhou, S.; Guan, J.; Zheng, Y. gTRICLUSTER: A More General and Effective 3D Clustering Algorithm for Gene-Sample-Time Microarray Data. *Lect. Notes Comput. Sci.* **2006**, *3916*, 48–59.
22. Araújo, R.B.; Ferreira, G.H.T.; Orair, G.H.; Meira, W.; Ferreira, R.A.C.; Neto, D.O.G.; Zaki, M.J. The ParTriCluster algorithm for gene expression analysis. *Int. J. Parallel. Program.* **2008**, *36*, 226–249.
23. Li, Z.; Hu, X.; Chen, Y. Multi-objective evolutionary algorithm for mining 3D clusters in gene-sample-time microarray data. In Proceedings of the IEEE International Conference on Granular Computing, Hangzhou, China, 26–28 December 2008; pp. 442–447.
24. Sim, K.; Aung, Z. Discovering Correlated Subspace Clusters in 3D Continuous-Valued Data. In Proceedings of the IEEE International Conference on Data Mining, Sydney, Australia, 14–17 December 2010; pp. 471–480.
25. Hu, Z.; Bhatnagar, R. Algorithm for Discovering Low-Variance 3-Clusters from Real-Valued Datasets. In Proceedings of the IEEE International Conference on Data Mining, Sydney, Australia, 14–17 December 2010; pp. 236–245.

26. Liu, Y.; Yang, T.; Fu, L. A partitioning based algorithm to fuzzy tricluster. *Math. Probl. Eng.* **2015**, ID 235790.
27. del Buono, N.; Pio, G. Non-negative matrix tri-factorization for co-clustering: An analysis of the block matrix. *Infor. Sci.* **2015**, *301*, 13–26.
28. Ignatov, D.I.; Gnatyshak, D.V.; Kuznetsov, S.O.; Mirkin, B.G. Triadic formal concept analysis and triclustering: searching for optimal patterns. *Mach. Learn.* **2015**, *99*, 1–32.
29. Gnatyshak, D.V. A single-pass triclustering algorithm. *Automat. Doc. Math. Ling.* **2015**, *49*, 27–41.
30. Negredo, A.M.; Bird, P.; de Galdeano, C.S.; Buforn, E. Neotectonic modeling of the Ibero-Maghrebian region. *J. Geophys. Res.* **2002**, *107*, 1–15.
31. Buforn, E.; Udías, A.; Colombás, M.A. Seismicity, source mechanisms and seismotectonics of the Azores-Gibraltar plate boundary. *Tectonophysics* **1988**, *152*, 89–118.
32. Buforn, E.; Udías, A.; Mezcuca, J. Seismicity and focal mechanisms in south Spain. *Bull. Seismol. Soc. Am.* **1988**, *78*, 2008–2024.
33. Morales, J.; Serrano, I.; Vidal, F.; Torcal, F. The depth of the earthquake activity in the Central Betics (southern Spain). *Geophys. Res. Lett.* **1997**, *24*, 3284–3292.
34. Vanucci, G.; Gasperini, P. The new release of the database of Earthquake Mechanisms of the Mediterranean Area (EMMA Version 2). *Ann. Geophys.* **2004**, *47*, 307–334.
35. Buforn, E.; de Galdeano, C.S.; Udías, A. Seismotectonics of the Ibero-Maghrebian region. *Tectonophysics* **1995**, *248*, 247–261.
36. Jiménez-Munt, I.; Fernández, M.; Torné, M.; Bird, P. The transition from linear to diffuse plate boundary in the Azores-Gibraltar region: results from a thin-sheet model. *Earth Planet. Sci. Lett.* **2001**, *192*, 175–189.
37. Vegas, R.; Banda, E. Tectonic framework and Alpine evolution of the Iberian Peninsula. *Earth Evol. Sci.* **1982**, *4*, 320–343.
38. Banda, E.; Gallart, J.; García-Dueñas, V.; Dobeitia, J.J.; Makris, J. Lateral variation of the crust in the Iberian Peninsula. New evidence from the Betic Cordillera. *Tectonophysics* **1993**, *221*, 53–66.
39. Casas, A.; Carbó, A. Deep structure of the Betic Cordillera derived from a complete Bouguer anomaly map. *J. Geodynamics* **1990**, *12*, 137–147.
40. Torné, M.; Fernández, M.; Comas, M.C.; Soto, J.I. Lithospheric structure beneath the Alboran Basin: Results from 3D gravity modeling and tectonic relevance. *J. Geophys. Res.* **2000**, *105*, 3209–3228.
41. Veith, K.F.; Clawson, G. Magnitude from short period p -wave data. *Bull. Seismol. Soc. Am.* **1972**, *62*, 435–452.
42. Hanks, T.C.; Kanamori, H. A moment magnitude scale. *J. Geophys. Res.* **1979**, *84*, 23480–23500.
43. Morales-Esteban, A.; Martínez-Álvarez, F.; Troncoso, A.; de Justo, J.L.; Rubio-Escudero, C. Pattern recognition to forecast seismic time series. *Expert. Syst. Appl.* **2010**, *37*, 8333–8342.
44. Martínez-Álvarez, F.; Reyes, J.; Morales-Esteban, A.; Rubio-Escudero, C. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. *Knowl. Base. Syst.* **2013**, *50*, 198–210.

45. Mezcuca, J.; Rueda, J.; Blanco, R.M.G. Reevaluation of historic earthquakes in Spain. *Seismol. Res. Lett.* **2004**, *75*, 189–204.
46. Reyes, J.; Cárdenas, V. A Chilean seismic regionalization through a Kohonen neural network. *Neur. Comput. Appl.* **2010**, *19*, 1081–1087.
47. Scitovski, R.; Scitovski, S. A fast partitioning algorithm and its application to earthquake investigation. *Comput. Geosci.* **2013**, *59*, 124–131.
48. Goldberg, D.E. *Genetic Algorithms*; Pearson Education: London, UK, 2006.
49. Martínez-Ballesteros, M.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. Selecting the best measures to discover quantitative association rules. *Neurocomputing* **2014**, *126*, 3–14.
50. Cheng, Y.; Church, G.M. Biclustering of expression data. In Proceedings of the International Conference on Intelligent Systems for Molecular Biology, La Jolla/San Diego, CA, USA, 19–23 August 2000; pp. 93–103.
51. Gutiérrez-Avilés, D.; Rubio-Escudero, C. MSL: A measure to evaluate three-dimensional patterns in gene expression data. *Evol. Bioinformatics* **2015**, *11*, 121–135.
52. Burrough, P.A.; McDonnell, R.A.; Lloyd, C.D. *Principles of geographical information systems*; Oxford University Press: Oxford, UK, 2015.
53. Kohonen, T. *Self-Organizing Maps*; Springer: Berlin, Germany, 2010.
54. Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C.; Aguilar-Ruiz, J.S. Energy time series forecasting based on pattern sequence similarity. *Pattern Recogn. Lett.* **2011**, *32*, 1662–1665.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).