

# Generating a seismogenic source zone model for the Pyrenees: A GIS-assisted triclustering approach

José L. Amaro-Mellado <sup>a,b</sup>, Laura Melgar-García <sup>c</sup>, Cristina Rubio-Escudero <sup>d</sup>, David Gutiérrez-Avilés <sup>c,\*</sup>

<sup>a</sup> Department of Graphic Engineering, University of Seville, ES-41092, Seville, Spain

<sup>b</sup> National Geographic Institute of Spain, Andalusia Division, ES-41013 Seville, Spain

<sup>c</sup> Data Science & Big Data Lab, Pablo de Olavide University, ES-41013 Seville, Spain

<sup>d</sup> Department of Computer Languages and Systems, University of Seville, ES-41012, Seville, Spain

## ABSTRACT

**Keywords:**  
Seismic sources  
GIS  
Data science  
Triclustering

Seismogenic source zone models, including the delineation and the characterization, still have a role to play in seismic hazard calculations, particularly in regions with moderate or low to moderate seismicity. Seismic source zones establish areas with common tectonic and seismic characteristics, described by a unique magnitude–frequency distribution. Their definition can be addressed from different views. Traditionally, the source zones have been geographically outlined from seismotectonic, geological structures, and earthquake catalogs. Geographic information systems (GIS) can be of great help in their definition, as they deal rigorously and less ambiguously with the available geographical data. Moreover, novel computer science approaches are now being employed in their definition. The Pyrenees mountain range – in southwest Europe – is located in a region characterized by low to moderate seismicity. In this study, a method based purely on seismic catalogs, managed with a GIS and a triclustering algorithm, were used to delineate seismogenic zones in the Pyrenees. Based on an updated, reviewed, declustered, extensive, and homogeneous earthquake catalog (including detailed information about each event such as date and time, hypocentral location, and size), a triclustering algorithm has been applied to generate the seismogenic zones. The method seeks seismicity patterns in a quasi-objective manner following an initial assessment as to the best suited seismic parameters. The eight zones identified as part of this study are represented on maps to be analyzed, being the zone covered by the Arudy–Arette region to Bagnères de Bigorre as the one with the highest seismic hazard potential.

## 1. Introduction

Seismogenic source zones are necessary for applying the most widely used seismic hazard calculation method proposed by [Cornell \(1968\)](#) and [McGuire \(1976\)](#). Originally, this method contemplated the zones as an alternative to model the probabilistic space of distance-to-source relationships with the area under analysis, particularly when associations between seismicity and faults capable of triggering an earthquake are not apparent ([García-Mayordomo, 2015](#)). In addition, the method conceived a fault as a linear element, but it has evolved to the use of fault as zones (even as 3D planes). This second option is much more widely used today. Currently, there is a tendency to make greater use of faults, but through a characterization collected in a database. These records include geological and seismic data, such as the slip rate, the fault length, the depth, recurrence periods, or maximum expected magnitude ([García-Mayordomo et al., 2012a](#)).

The main assumptions of the Cornell-McGuire method regarding the occurrence of seismicity are the following ([García-Mayordomo, 2015](#)):

1. Earthquakes are independent random events (Poissonian).
2. The probability of earthquakes occurrence is the same throughout the entire area.
3. The size of earthquakes is related to their frequency through the Gutenberg–Richter Law ([Gutenberg and Richter, 1944](#)) of each source zone, usually limited to a maximum magnitude value.
4. The seismic activity rate of each zone is constant over time.

These cannot be strictly fulfilled in practice since the mechanism of accumulation and release of energy in earthquakes is a memory process with long-range dependence ([Barani et al., 2018](#)), and the rate of seismic activity is variable over time. However, in seismic engineering,

\* Corresponding author.

E-mail addresses: [jamaro@us.es](mailto:jamaro@us.es) (J.L. Amaro-Mellado), [lmelgar@upo.es](mailto:lmelgar@upo.es) (L. Melgar-García), [crubioescudero@us.es](mailto:crubioescudero@us.es) (C. Rubio-Escudero), [dgutavi@upo.es](mailto:dgutavi@upo.es) (D. Gutiérrez-Avilés).

seismogenic source zones are sufficiently adequate to estimate the probability according to different levels of ground shaking (García-Mayordomo, 2015). In this case, a memory process refers to the fact that shocks do not happen independently at any time. By contrast, they remember the occurrence (time and magnitude) of the last earthquake (Corral, 2006). The main strength of using zones is that they are efficient because their associated parameters are somehow easily obtained, and their calculations are fast. Thus, they have been used for over 50 years so far.

Seismic source delimitation is a crucial task when a Probabilistic Seismic Hazard Analysis (hereinafter, PSHA) is conducted (Morales-Esteban et al., 2014). Once they have been geometrically outlined, their activity is characterized based on the recorded seismicity. Both stages define the seismic potential (geometric shape and size, maximum expected magnitude, seismic activity, and frequency distribution in size). Subsequently, the equations for predicting strong movements should be considered, and, finally, the annual probability of exceeding a certain magnitude earthquake should be obtained.

Despite the importance of the seismogenic zonation to conduct a seismic hazard analysis, the criteria used for its definition are often not uniform. It is a very complex process and depends strongly on the analyst group's criteria (Amaro-Mellado et al., 2017). There are two main groups within these criteria: one based exclusively on the distribution of seismicity and the other on geological domains or structures. The criterion, or its weight, depends on the background of the analyst who defines the zones. On the one hand, much weight is placed on the characterization of sources, it will mainly be based on seismicity. This can lead to inconsistent limits from a geological point of view but valid for the calculation of hazards. On the other hand, if the analyst relies on the location of the geological structure, two opposing situations arise. In the first, areas are based exclusively on the limits of large geological units, even if they have no relation to current tectonic activity. In the second, they only consider areas with active faults, even though the recorded seismicity does not allow for statistically significant characterization.

Computational robust methods are becoming increasingly popular, as they can be applied in different geographical and geological contexts (Morales-Esteban et al., 2014; Martínez-Álvarez et al., 2015; Scitovski, 2018). Besides, when handling and representing georeferenced spatial information such as the distribution of epicenters, the use of a geographic information system (hereinafter, GIS) is powerful. Its ability to manage, represent, and generate geographical information is a useful tool for integrating different geographic data kinds.

This work aims at proposing a seismic source model for the Pyrenees range in southwest Europe. To this end, the use of a GIS and a tricluster algorithm, called TriGen (Gutiérrez-Avilés et al., 2014), has been applied to derive seismicity patterns from a seismic catalog. Later, the source models have been drawn. Finally, seismic parameters have been obtained from these zones, and they have been analyzed. The range of seismicity in the Pyrenees is estimated as low to moderate and diffuse, where earthquakes with a magnitude larger than or equal to 5.0 ( $M5+$ ) are infrequent. Besides, in these areas, obtaining a correlation between epicenters and faults is difficult (Drouet et al., 2020). Therefore, seismic zonings usage to drive a PSHA is adequate (Martínez-Álvarez et al., 2015; Amaro-Mellado and Tien Bui, 2020).

## 2. Related work

In this section, research made in both domains presented in this study is described: firstly, the studies in seismogenic zonings and, secondly, the triclustering approach.

### 2.1. Previous seismic zonations for the Pyrenees

The Pyrenees are located in a border-area between France, Spain, and Andorra, so each of these countries have studied their seismicity. They are also included in some European seismicity projects as SHARE (Woessner et al., 2011; Stucchi et al., 2013), or SIGMA (Pecker et al., 2017).

In the Spanish seismic sources, the Pyrenees are an important region due to their activity, and they are included in different seismic zonations. For example, that conducted by Martín (1984), who established 27 zones for the whole Iberian Peninsula and adjacent area; Mezcua et al. (2011) addressed a PSHA as a combination from the one defined by Molina (1998) and the CODE one (from NCSE-2002 Ministerio de Fomento (Gobierno de España), 2002, which is very similar to Martín, 1984). More recently, in the frame of the Updated Map of PSHA for Spain, both García-Mayordomo et al. (2012b) and Bernal (2011) defined a set of seismic sources. Later, Morales-Esteban et al. (2014) Morales-Esteban et al. undertook a zonation for the whole Peninsula from Mahalanobis distances, based only on an earthquake catalog (including information on the shock such as date and time, hypocentral location, or size) from 1978 to 2012. In Martínez-Álvarez et al. (2015), Martínez-Álvarez et al. used the triclustering methodology to obtain seismic sources for the whole Iberian Peninsula.

French institutions have also defined different seismogenic zonations. In Metropolitan France, the first PSHA map was deployed in 2002 (Martin et al., 2002) and gave the foundations for the French zoning (Drouet et al., 2020). Moreover, in 2004, Marin depicted PSHA maps for France (Marin et al., 2004). Later, Baize et al. (2013) proposed a new seismotectonic zoning for Metropolitan France based on geological, seismological, and tectonic data. Recently, Drouet et al. (2020) constructed a seismotectonic model (GEOTER) for Metropolitan France, considering geological, structural, neotectonic, geophysical, and seismological data. Besides, some seismic zonations have been conducted specifically for the Pyrenees. In Njike-Kassala et al. (1992), the authors established nine zones to calculate the  $b$ -value of the Gutenberg and Richter relation (Gutenberg and Richter, 1944). Later, Secanell et al. (2008) defined a seismotectonic zonation as a result of the ISARD project, held by France and Spain. Finally, Rigo et al. (2015) proposed eight zones to estimate stress tensors locally.

Finally, the Institut d'Estudis Andorrans holds an internet site (<https://www.iea.ad/sismoweb>, last accessed on November 2020) with information on Andorran seismicity.

### 2.2. Triclustering

Regarding the second area of study of this research, the state of the art of triclustering techniques is presented. The traditional clustering approach is a data mining technique that works by grouping objects according to a predefined similarity in a one-dimensional space. When dealing with subspace clustering, the problem can be addressed with biclustering (Pontes et al., 2015) or triclustering (Zhao and Zaki, 2005) if the context data is structured over two or three dimensions, respectively.

In the last decade, the tricluster algorithm has evolved, and many different algorithms have been proposed. For example, some triclustering algorithms are based on iterative searches, distribution parameter identification, biclustering algorithms, pattern mining procedures, or evolutionary multiobjective optimization (Henriques and Madeira, 2018).

In particular, in Liu et al. (2008) and Bhar et al. (2015) some triclustering algorithms based on multiobjective techniques can be found. Furthermore, classic bio-inspired meta-heuristics as genetic algorithms (Holland, 1992) or newer, as the COVID-19 propagation model (Martínez-Álvarez et al., 2020) emerge as a proper method to optimize multiobjective functions. In this sense, the TriGen algorithm presented in Gutiérrez-Avilés et al. (2014) is based on a population

of individuals and some genetic operators. Different fitness functions are also proposed in order to assess each research problem most accurately. TriGen is the triclustering algorithm used in this study and is further explained in 4.3. Besides developing triclustering methodologies, another open line of investigation is to measure the quality of the triclustering algorithms' solutions. In this sense, we can find correlation-based measures inspired by Pearson and Filon (1898), and Spearman (Spearman, 1910) correlation indexes. In Gutiérrez-Avilés and Rubio-Escudero (2014) the authors propose a three-dimensional version of the mean squared residue, a classical biclustering evaluation measure (Cheng and Church, 2000). An evaluation measure based on the last squared lines depicted by the discovered patterns in the triclusters was presented in Gutiérrez-Aviles and Rubio-Escudero (2014). Finally, in Gutiérrez-Avilés and Rubio-Escudero (2015), the authors present a new approach to measure the tricluster quality based on the graphical representation of it.

Triclustering has many applications: biological as Li and Tuck (2009) that combines gene regulator information with expression data; medical as Melgar-García et al. (2020) that applies triclustering in the streaming environment to high-content screening images or the analysis of social networks as Gnatyshak et al. (2012). The domain of seismic data has not yet been deeply analyzed with triclustering techniques. To the best of our knowledge, only the work in Martínez-Álvarez et al. (2015) applies triclustering techniques, particularly, the TriGen algorithm, to data of the Iberian Peninsula, in order to discover possible seismogenic zones.

### 3. Seismicity and geological settings

The Pyrenees mountain range, constituted from the collision between Iberian and Eurasian plates, was formed in Alpine times (Vissers and Meijer, 2012); however, current deformation rates are relatively low. Spanning over 450 km in E-W direction and 150 km in N-S (Rigo et al., 2018), the Pyrenees have nearly cylindrical symmetry (Njike-Kassala et al., 1992). It is bordered to the north by the North Pyrenean Fault, which matches a 10–15 km vertical Moho offset, where the crust is thicker on the Iberian plate (Gallart et al., 1981). The North Pyrenean Fault extends along the whole chain in E–W direction, and it is supposed to be the boundary between the two plates (Njike-Kassala et al., 1992), and currently, between the well-known North Pyrenean Zone and the Axial Zone. Upon Vissers and Meijer (2012), the Pyrenees are characterized by, from north to south: first, a north-directed thrust-belt including Mesozoic and Tertiary sediments of the Aquitanian Basin; second, the North Pyrenean Zone, formed mainly by Paleozoic basement and Mesozoic sediments; third, the Axial Zone consisting of Paleozoic rocks; and, finally, the southern Pyrenees where the southward thrust Mesozoic and Tertiary sediments prevail, and they part of the Ebro foreland basin system.

Regarding seismicity, it is very complex (Souriau et al., 2014), and it can be considered as low to moderate (Amaro-Mellado and Tien Bui, 2020). In the western part, the seismicity is generally concentrated along the North Pyrenean Zone. In contrast, it is more diffuse in the eastern part and not particularly associated with the known faults (Njike-Kassala et al., 1992). Besides, the identification and parametrization of seismogenic structures have to face some challenges. The most important ones are the lack of surface faulting related to shocks in the last centuries, the low deformation rates, or the fact of being a political-border region (partially solved by the projects above mentioned) (Lacan and Ortuño, 2012).

Over time, as can be deduced from earthquake catalogs, some shocks have produced severe damages, mainly in the historical period, even with estimated magnitudes between 6 and 7 or with MSK intensities between VIII and IX (Lacan and Ortuño, 2012). Among others, the 1373 earthquake ( $I_o = VIII - IX$ ) by Maladetta Massif, in the Central Pyrenees; those happened during 1427 Amer and 1428 Querulbs, in Catalonia (up to  $I_o = IX$ ), in the eastern Pyrenees;

the 1660 Bigorre earthquake ( $I_o = IX$ ), near Lourdes, and the most recent (1750) in Juncalas ( $I_o = VIII$ ), close to the Bigorre one. In the instrumental period, some relevant shocks (magnitude equal to or larger than 5) have occurred, such as 1967 Arette (M5.3), in the western Pyrenees, and 1996 St-Paul Fenouillet (M5.0), in the Agly Massif, in the east (Amaro-Mellado and Tien Bui, 2020).

Some specific studies have been driven to study the seismicity of the whole Pyrenees Njike-Kassala et al. (1992), Secanell et al. (2007), Rigo et al. (2018), Amaro-Mellado and Tien Bui (2020). Besides, the Arudy–Arette region has been analyzed in-depth by Gallart et al. (1980) and Sylvander et al. (2008).

Fig. 1 shows a general view of the major events that have shocked the Pyrenees from the raw data obtained from the earthquake catalog published by the Instituto Geográfico Nacional (IGN)–National Geographic Institute of Spain (NGIS) (Insitituto Geográfico Nacional, 2020). The map only represents the events with a magnitude larger than or equal to 3.5, or macroseismic intensity,  $I_o$ , larger than or equal to “V” in macroseismic scale. If both data appear, the magnitude one prevails.

## 4. Methods

In this section, the methodology driven to generate the seismogenic sources for the Pyrenees are described. Thus, the seismic dataset used in this study is presented; later, the data processing undertaken in this work is explained; and finally, a summary of the TriGen algorithm is set out.

### 4.1. Generation of the earthquake catalog

The dataset employed in this study comes from the NGIS earthquake catalog (Insitituto Geográfico Nacional, 2020). In-depth research on the NGIS catalog is presented in González (2017). In particular, that compiled in Amaro-Mellado and Tien Bui (2020) has been chosen in this work. In this catalog, starting from the NIGS earthquake database, different steps were taken to generate an updated, reviewed, homogeneous, declustered, and extensive catalog, including events with a magnitude larger than or equal to 2.0; from 1373 to 2019; and the geographical extension was limited by 2.5° W and 3.5° E meridians, and 41° N and 44° N parallels.

The steps driven to compile the catalog were the following:

1. Data from other sources, such as specific studies or other catalogs, were considered to re-evaluate the magnitude of some shocks, particularly the major ones.
2. To conduct a seismic analysis, the size of the events must be homogeneous. In this sense, the vast majority of the researches use  $M_w$  (Hanks and Kanamori, 1979), due to its direct relationship with the released energy in the rupture by scalar seismic moments. Besides, it does not get saturated for the largest events (Das et al., 2019). Over time (from 1373 to 2019), different kinds of sizes have been assigned to the earthquakes (macroseismic intensities and several magnitude types). Herein, to homogenize the size,  $M_w$  was assigned to all the events, according to Cabañas et al. (2015). Besides, after estimating a lower magnitude of completeness (1.8), the cut-off magnitude was set as 2.0.
3. As stated in Section 1, the earthquakes must be independent events to fulfill the Poisson distribution. To this end, foreshocks, aftershocks, and seismic swarms must be eliminated. In seismicity, this process is called declustering. In that work (Amaro-Mellado and Tien Bui, 2020), it was based on the definition of temporal and spatial windows following the method proposed by Gardner and Knopoff (1974).

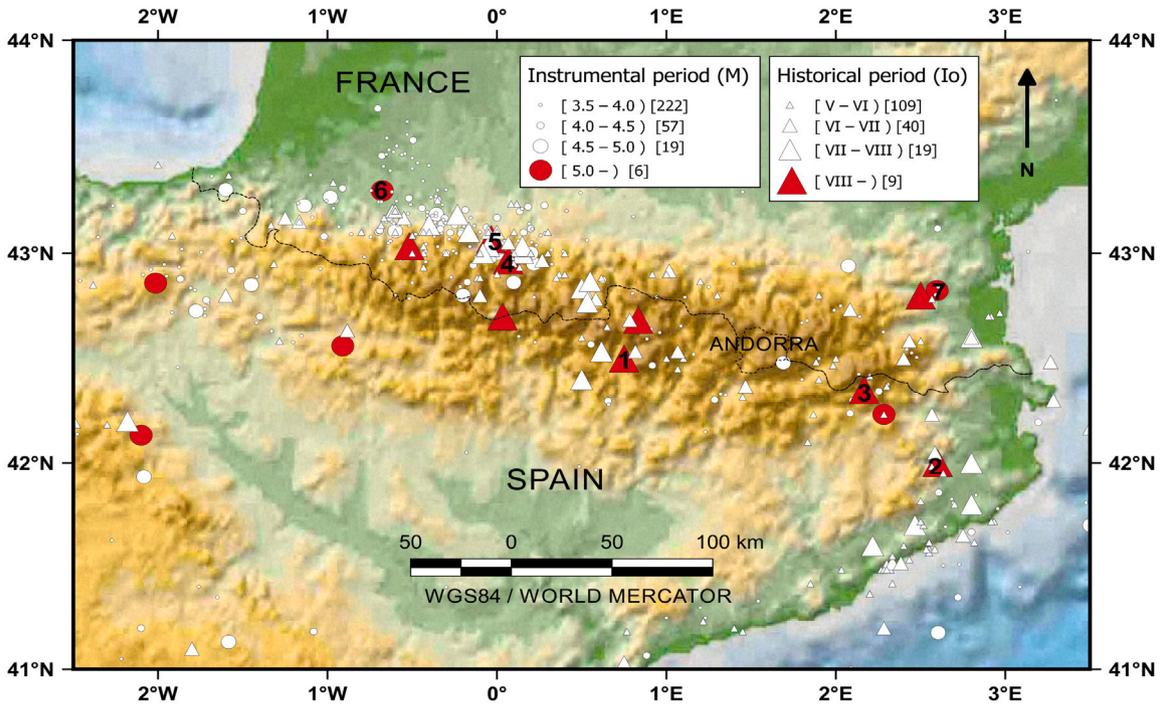


Fig. 1. Seismicity in the Pyrenees from the raw NGIS earthquake catalog. Events instrumentally recorded (Circles); events with only macroseismic intensity (triangles). Above referred earthquakes have been numbered: 1—Maladetta(1373); 2—Amer(1427); 3—Queralbs(1428); 4—Bigorre(1660); 5—Juncalas(1750); 6—Arette(1967); 7—St-Paul Fenouillet(1996).

4. Although it is not a specific task in a catalog generation, the year of completeness related to different levels of magnitude ( $M_c$ —year of completeness) were defined: M2—2013; M3—1978; M4—1943; and, M5—1810.
5. The final catalog, after removing the events deeper than 65 km, since they are not important for the seismic hazard of the region (IGN-UPM-WorkingGroup, 2013), consists of 7706 earthquakes.

The present work aims at defining a source zone model, so such a low cut-off value as 2.0 should not be established. Therefore, the cut-off magnitude has been set as 2.5, as pointed out in other works (Njike-Kassala et al., 1992; Talbi et al., 2013). Besides, it has been considered in a recent work for the region (Drouet et al., 2020). This value is judged as a good trade-off between the seismic hazard and the low to moderate Pyrenees seismicity.

The final catalog of this work consists of 3500 earthquakes, with  $M_w$  larger than or equal to 2.5, and it is shown in Fig. 2.

#### 4.2. Dataset construction

The seismic data, analyzed in the previous section, must be converted into a 3-D dataset (or cube of data) in order to be interpreted by the TriGen algorithm. Herein, several new attributes must be generated to transform a 2D dataset into a 3D one. Firstly, using a GIS, all data included in the catalog have been sorted into  $30 \times 20$  cells, each representing an area of approximately  $16.7 \text{ km} \times 16.4 \text{ km}$ . Afterward, each cell is characterized as a set of features. These features, originally defined by literature (Scitovski and Scitovski, 2013; Reyes and Cárdenas, 2010), have been re-adapted to the Pyrenees seismicity and are:

1.  $M_{max}$ : maximum earthquake magnitude recorded in the cell.
2.  $D$ : mean depth of all earthquakes' epicenters recorded in the cell.
3.  $M_{2,9}$ : number of earthquakes occurring with a magnitude larger than or equal to 2.9 in the cell considered.
4.  $M_{3,3}$ : number of earthquakes occurring with a magnitude larger than or equal to 3.3 in the cell considered.

5.  $M_{3,7}$ : number of earthquakes occurring with a magnitude larger than or equal to 3.7 in the cell considered.
6.  $M_{4,1}$ : number of earthquakes occurring with a magnitude larger than or equal to 4.1 in the cell considered.
7.  $N$ : total number of earthquakes occurring.

As can be seen in Fig. 3, the result of applying this data transformation is a data cube (or 3D dataset) where every cell is defined by three coordinates  $x$ ,  $y$  and  $f$ . The  $x$  coordinate is in  $[1, 30]$  representing the relative latitude, the  $y$  coordinate is in  $[1, 20]$  representing the relative longitude, and, the  $f$  coordinate is in  $\{M_{max}, D, M_{2,9}, M_{3,3}, M_{3,7}, M_{4,1}\}$  and represents the particular feature. Therefore, as an example, the cell  $C_{1,2,M_{3,7}}$  represents the number of earthquakes occurring with a magnitude larger or equal to 3.7 in the cell with relative latitude 1 and relative longitude 2.

#### 4.3. The TriGen algorithm

The triclustering algorithm used in this research is TriGen (Gutiérrez-Avilés et al., 2014), which stands for Triclustering Genetic based algorithm. As it can be appreciated in Fig. 4(a), the TriGen algorithm receives the 3D input dataset as well as the control parameters to yield a set of tricluster solutions. To do this, TriGen performs an evolutionary heuristic search with a population of individuals that evolves using genetic operators during a specific number of generations with the main goal of optimizing an evaluation function. Considering that an individual is a potential solution tricluster, in this case, each individual represents a zone of the Pyrenees.

Focusing on the algorithm, TriGen repeats for each tricluster solution ( $N$  control parameter) the following process. Firstly, it creates an initial population that is evaluated based on a fitness function that has to be optimized. Afterward, four genetics operators are applied to the initial population during a number of generations ( $G$  control parameter) in order to select the best individual that is going to be the solution tricluster. The four genetics operators that are repeated are selection, crossover, mutation, and evaluation.

Each operator has specific characteristics as were deeply analyzed in Gutiérrez-Avilés et al. (2014), Gutiérrez-Avilés and Rubio-Escudero (2015):

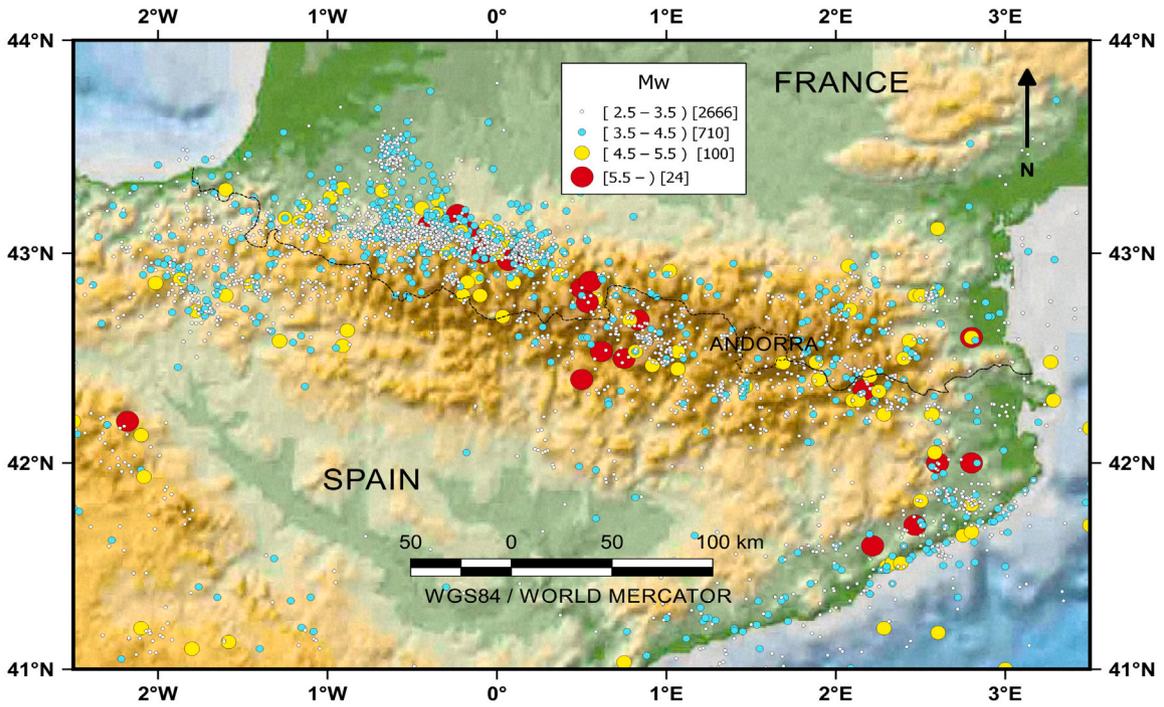


Fig. 2. Catalog of the work. The size of all the events is given as moment magnitude ( $M_w$ ).

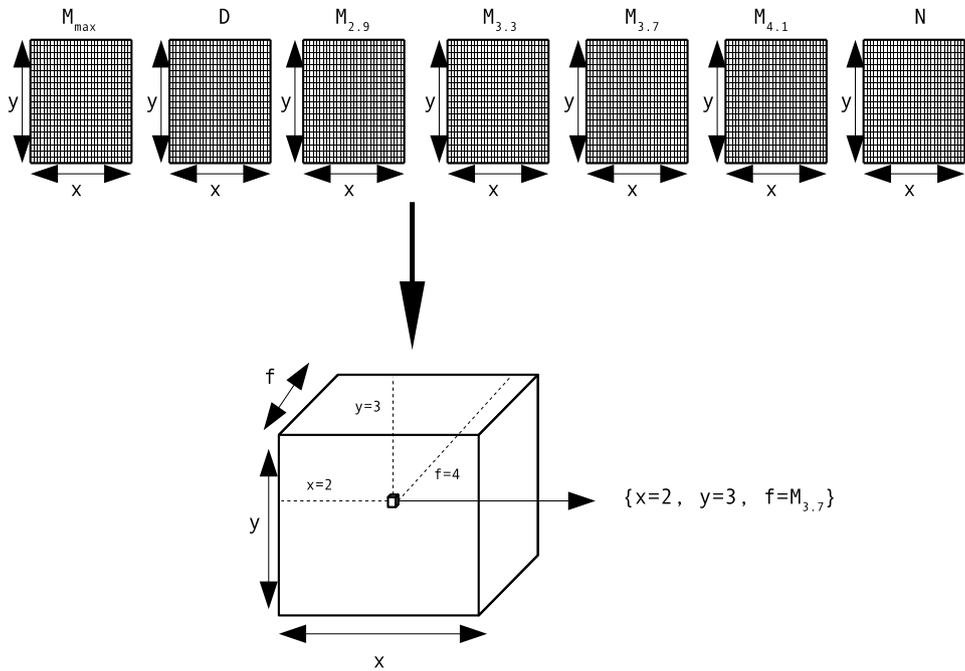


Fig. 3. TriGen input dataset.

- The initial population is created differently to search for the first solution tricluster and all the other  $N-1$  solution triclusters searches. The first one is produced with a random subset of all three data coordinates. For the next solution triclusters, some individuals are also randomly generated, and the other ones are created considering not to overlap the previously founded solutions.
- The evaluation operator is very useful in promoting the best individuals to continue in the next generation and selecting the final solution tricluster. TriGen offers three different fitness function: the mean square residue measure ( $MSR$ ) (Gutiérrez-Avilés and

Rubio-Escudero, 2014), the least square lines ( $LSL$ ) (Gutiérrez-Avilés and Rubio-Escudero, 2014), and the multi slope measure ( $MSL$ ) (Gutiérrez-Avilés and Rubio-Escudero, 2015). For this particular case, the  $MSL$  fitness function is selected. In general, the  $MSL$  function is based on the resemblances of the slope angles formed by the expression values of the tricluster coordinates.

- The selection operator is the first operator computed in the loop of  $G$  generations. The tournament selection is implemented, and so the selected individual is promoted to the next generation.

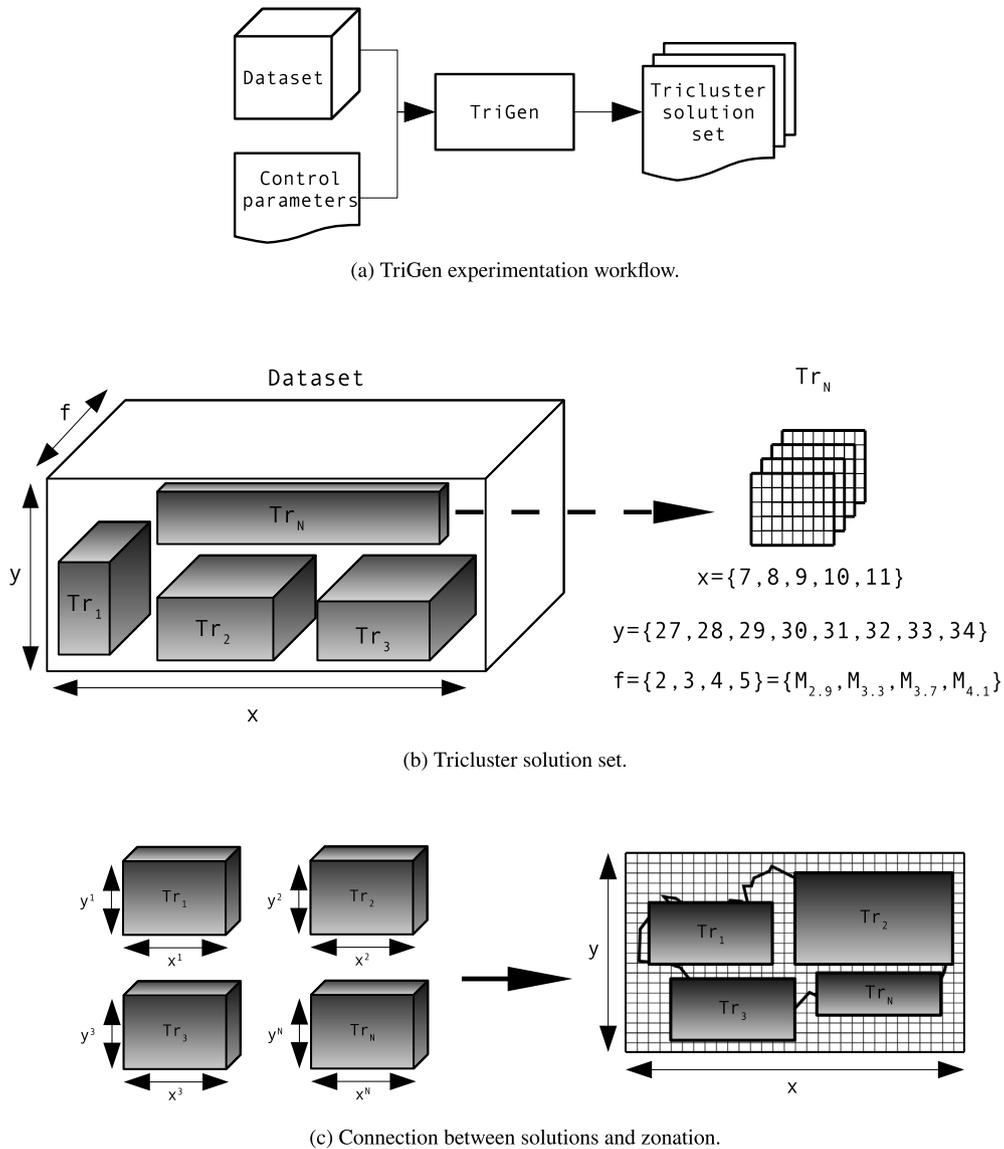


Fig. 4. TriGen algorithm overview.

- The crossover operator applied is the one point crossing technique. In particular, two selected individuals (by a probability of crossover parameter  $p_c$ ) mix their coordinates randomly and generate two new individuals (their children).
- The mutation operator also considers a probability of mutation parameter  $p_m$  that if it is satisfied, applies one of the six possible actions: add or remove one of the three coordinates of data. The goal of the mutation operator is to guarantee variability for the next generations.

The algorithm ends when all the best solution triclusters are found. The produced tricluster solution set represents a partition of the 3D input dataset, whereas each tricluster is a subset of  $x$  coordinates,  $y$  coordinates, and  $f$  coordinates. Therefore, as it is represented in the example in Fig. 4(b), a tricluster  $Tr_N$  of the solution set is a dataset subset composed by the  $x$  coordinates from 7 to 11, the  $y$  coordinates from 27 to 34 and, the  $\{2, 3, 4, 5\}$  features, that is, the  $M_{2.9}$ ,  $M_{3.3}$ ,  $M_{3.7}$  and  $M_{4.1}$  features. To yield the zonation discovered by the algorithm, for each tricluster, its  $x$  and  $y$  coordinates will be an area of the complete analyzed grid.

## 5. Results and discussion

In this section, the experimental workflow carried out in this research is explained, as well as an analysis of the obtained results. For this purpose, the experimental setup of the TriGen algorithm, followed to produce the zonation, is described in Section 5.1. Finally, the discovered zone and a critical discussion are addressed in Section 5.2.

### 5.1. TriGen experimental setup

The TriGen algorithm has been executed four times. As there is a component of randomness in each execution, different executions can lead to different solutions. From each tricluster solution set of each TriGen run, the non-overlapping zones are selected to yield the final zonation presented in the next Section 5.2. As can be seen in Table 1, for each run (from  $R_1$  to  $R_4$ ), a particular parameter configuration has been used as the input dataset described in Section 4.2, in order to obtain tricluster solutions in the widest range possible.

For all runs, TriGen has been set up to find ten solutions ( $N$ ) to guide the algorithm to cover the  $X - Y$  space. The number of generations ( $G$ ) and the number of individuals in the first population ( $I$ ) has been  $\{20, 30\}$ , and  $\{15, 20\}$  respectively. This is considered a proper

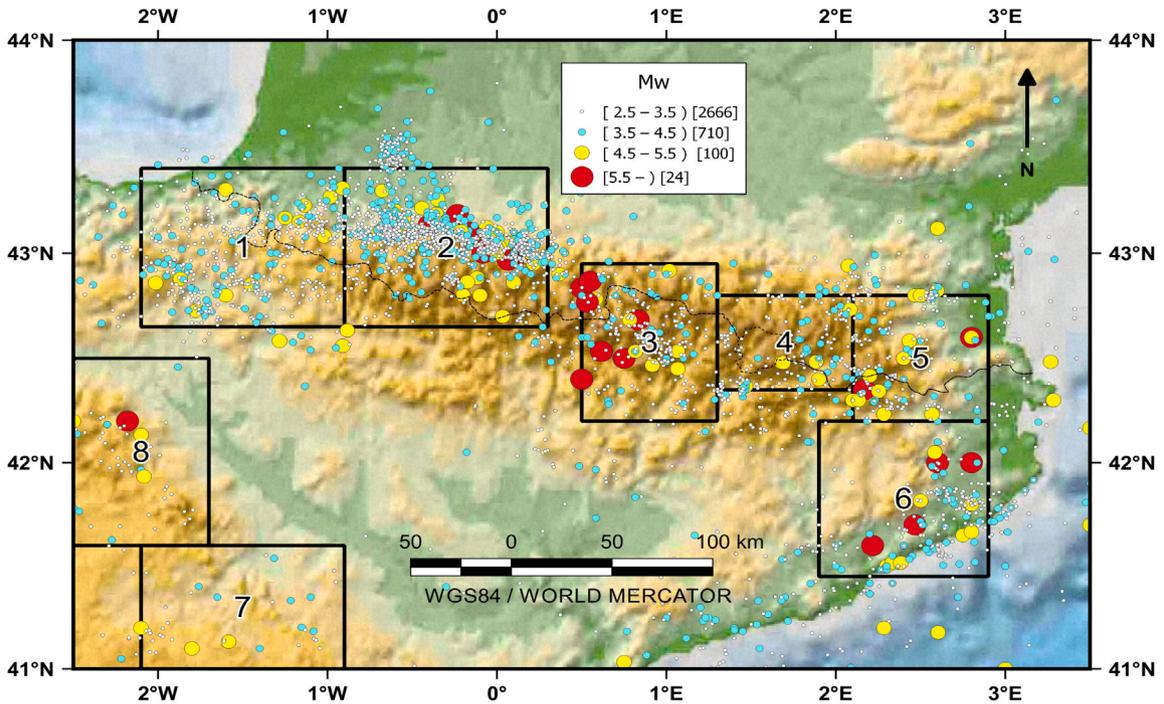


Fig. 5. Zonation proposal (eight zones). The size of all the events is given as moment magnitude ( $M_w$ ). Source zone #1: Atlantic Pyrenees; #2 Arudy–Arette–Bagnères de Bigorre; #3: Central Pyrenees; #4: Andorra; #5: Easternmost Pyrenees; #6: Eastern Catalonia; #7: Central Iberian System; #8: Northeast Iberian System.

Table 1  
TriGen parameters setup.

Parameter	$R_1$	$R_2$	$R_3$	$R_4$
$N$	10	10	10	10
$G$	20	30	20	30
$I$	15	20	20	15
$p_c$	0.2	0.2	0.8	0.8
$p_m$	0.5	0.5	0.1	0.1

balance between intensification and diversification of the evolutionary process. This fact has been taken into account when the crossover and mutation parameters were set. Firstly,  $p_c$  varies in  $\{0.2, 0.8\}$  to obtain high intensified and low diversified solutions in the  $p_c = 0.2$  runs and low intensified and high diversified solutions in the  $p_c = 0.8$  runs. In a similar way, with  $p_m$  varying in  $\{0.1, 0.5\}$ , to get solutions with high and low variability.

## 5.2. Zonation obtained and its representation

With the solutions obtained by the application TriGen, following the procedure described in Section 5.1, the zones with less than 25 events have been ruled out, according to Bender (1983) and Skordas and Kulhánek (1992). The use of a GIS (QGIS, <https://www.qgis.org>, last accessed November 2020) allows a rigorous representation, and the data can be handled for further analysis. The resulting zonation, as well as the epicenters distribution, is shown in Fig. 5.

### 5.2.1. Seismic parameters of the obtained zones

Once the zones have been delineated according to solutions provided by TriGen, the next step is to seismically characterize each zone, in which some seismic parameters must be uniform, namely, the  $b$ -value, or the annual rate, as well as the maximum magnitude. Therefore, they must be estimated for the proposed zones.

Thus, the Gutenberg–Richter law (Gutenberg and Richter, 1944, 1954) holds:

$$\log_{10} N(M) = a - bM \quad (1)$$

where the  $b$ -value estimates the relationship between small and large earthquakes, and it is related to the physics of the source. The lower its value is, the more energy can be accumulated.  $N$  is the number of events with a magnitude larger than or equal to a cut-off magnitude  $M$ , and  $a$  refers to the seismic productivity.

The Maximum-Likelihood-Estimate proposed by Bender (1983) and Aki (1965) for binned data ( $\Delta$ ) has been considered, in terms of  $b$ -value.

$$b = \frac{1}{\ln(10)(\bar{M} - M_{min} - \frac{\Delta}{2})} \quad (2)$$

where  $\bar{M}$  is the average magnitude;  $M_{min}$  is the magnitude of completeness, which means all the events of magnitude larger than or equal to this value have been recorded from the reference date;

In this work,  $M_{min}$  has been set as 3.0, and the corresponding year of completeness is 1978, as stated in Section 4.1, and the  $\Delta$  value for this work is 0.1. Besides, it is interesting to give uncertainty parameters in geology (Bárdossy and Fodor, 2001). In this case, the  $b$ -value uncertainty (Sigma  $b$ ) has been calculated by the approach suggested by Kijko and Smit (2012).

Another relevant parameter is the annual rate referred to a unit area ( $km^2$ ). It means the number of events exceeding a threshold per year and related to  $km^2$ .

$$\lambda(M_{min}) = \frac{N}{t} \quad (3)$$

where  $t$  is the time-lapse. In this research, from the beginning of 1978 until the end of 2019, i.e., 42 years.

When this value is normalized with the surface, by  $km^2$ , and multiplied by 10,000, it is called  $AR3$ , in this paper.

$$AR3 = 10,000 \frac{\lambda(M_{min})}{km^2} \quad (4)$$

Finally, the maximum magnitude  $M_{max}$  is calculated for each zone. The resulting values can be found in Table 2.

**Table 2**  
Obtained zones. Seismic parameters.

Zone	Name	Area (km <sup>2</sup> )	<i>b</i> -value	<i>AR</i> <sub>3</sub>	<i>M</i> <sub>max</sub>	<i>Depth</i> <sub>mean</sub>	Sigma <i>b</i>
1	Atlantic Pyrenees	8157	1.01	5.43	5.2	8.4	0.07
2	Arudy–Arette–Bagnères de Bigorre	8157	0.99	14.54	6.7	6.0	0.04
3	Central Pyrenees	5478	1.20	3.30	6.2	5.7	0.14
4	Andorra	3287	1.32	3.19	5.0	5.7	0.20
5	Easternmost Pyrenees	4388	1.17	3.91	6.5	5.8	0.14
6	Eastern Catalonia	6931	1.34	1.89	5.8	6.3	0.18
7	Central Iberian System	6709	1.13	0.50	5.2	8.6	0.30
8	Northeast Iberian System	6630	1.40	0.61	5.7	8.0	0.34

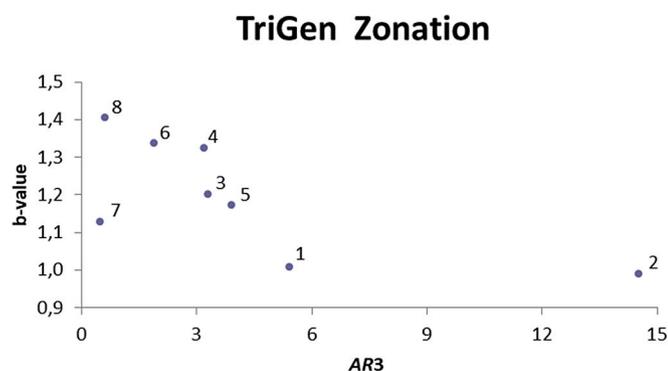


Fig. 6. *AR*<sub>3</sub> vs *b*-value.

### 5.3. Discussion

After calculating some of the most relevant seismic parameters, the analysis of the results is driven.

First, in order to check that the zones are seismically different from one another, a graphical representation of the *AR*<sub>3</sub> and *b*-value is presented in Fig. 6. In addition, to support this discrimination, different color maps have been deployed. Herein, Fig. 7 shows the *b*-value variation; *M*<sub>max</sub>; in Fig. 8 the *AR*<sub>3</sub> is depicted; Fig. 9 represents the *M*<sub>max</sub>; and finally, the mean depth is illustrated in Fig. 10.

Second, the analysis of each zone is undertaken.

Zone 1 is located in the Atlantic Pyrenees and presents one of the lowest *b*-values and the highest annual rate. However, the maximum magnitude is one of the smallest of the proposed zones (5.2), so the seismic hazard can be considered low to moderate in the Pyrenees region.

The most seismically hazardous zone is undoubtedly the number 2, which runs from the Arudy–Arette region to the Bagnères de Bigorre environment. This is due to the combination of the minimum *b*-value, and both maximum annual rate (the highest in the Iberian Peninsula) and *M*<sub>max</sub>. Therefore, this zone deserves special attention when a PSHA is driven in the Pyrenees.

Zone 3 is situated in the Central Pyrenees, in the Axial Zone. This area, which includes the Maladetta Massif, has been shocked by some of the strongest events in the Pyrenees (M6.2). Besides, it presents medium values in both annual rate and *b*-value. Herein, the seismic hazard can be defined as moderate; it might be even said moderate to high.

The smallest zone, numbered 4, is located in the Andorra environment, and not significant earthquakes have been taken place in it. Its *b*-value is high, and the annual rate is medium.

Like Zone 3, the 5 one, in the easternmost Pyrenees, is characterized by a medium *b*-value and annual rate but has suffered the second most energetic earthquake in the whole belt. Therefore, the seismic hazard can be estimated as moderate to high.

In zone 6, fully contained in eastern Catalonia (Spain), although the *M*<sub>max</sub> is not relatively high, four events with a magnitude exceeding 5.5 have occurred. Besides, it shows a high *b*-value, and the annual rate is low to moderate so that the seismicity can be rated as moderate. The epicenter distribution may indicate that this zone could be extended along the Catalanian Coastal Range to the southwest or consider a new one.

The following zones, 7 and 8, could not be closely related to the Pyrenees, but in its environment, in the Iberian System. Besides, both zones are in the limit to be seen as a seismic zone, individually. Both zones have the lowest annual rate by far, and the maximum magnitude is not notably high (5.7 and 5.2). Regarding the *b*-value, zone 7 (Central) shows an average value, and zone 8 (Northeast) the highest in the region. Given the few events available for *b*-value calculations, these are not entirely meaningful. The epicenter distribution points out that both zones could correspond to the same seismic zone if a proper rotation were to be carried out.

Although it is a debated topic, the highest seismic potential of source zone #2 (Arudy–Arette–Bagnères de Bigorre) can be related to the presence of several major faults. For example, the 25–30 km-long Herrère right lateral fault (Arudy) or the Lourdes reverse faults (60 km in three segments). Besides, the relevant seismicity of source zone #3 is related to the North Maladetta normal fault (30 km long) or Coronas normal fault (10 km) (Lacan and Ortuño, 2012). It is also remarkable that in zone 5 (Easternmost Pyrenees), the more destructive earthquake could be caused by the Vallfogona or Ribes–Camprodon thrusts or the Amer normal fault. The latter fault generated the major 1428 Queralbs earthquake (in zone 6) (Perea, 2009).

Regarding the method limitations, one of them would be that it only produces “rectangular zones” along parallels and meridians arches, as the original TriGen algorithm works that way. Thus, to overcome this issue, the results could be improved using rotated polygons, which could have been useful to extend zone 6, or join zones 7 and 8. Another method limitation is related to the fact that conceptually, an active fault should not be cut by a source zone boundary (Hamdache and Peláez, 2019), so it could be considered in further research.

## 6. Conclusions

In this research, a seismic zonation has been proposed for the Pyrenees mountain range. The zoning is purely based on the application of the TriGen algorithm to an earthquake catalog and a grid division. Herein, a GIS has been used to manage and represent both the input and the output data. The method aims to be as objective as possible, seeking patterns for the best considered seismic indicators and minimizing human interaction.

The triclustering methodology has been the primary procedure to yield the zonation and applied to a pre-processed 3D dataset using the TriGen algorithm. It is based on the genetic algorithm model. Thus, from an initial population of potential solutions, by applying genetic operators (crossovers, mutations), it evolves to optimal solutions based on the *M**S**L* fitness function.

The starting point is an updated, reliable, homogeneous, extensive, and with independent events earthquake catalog, coupled with a geographical grid defined by the GIS. This grid cell size has been set out as a nearly square shape in 16.7 km × 16.4 km, a trade-off between the coverage and the low to moderate Pyrenees seismicity. Then, as a previous step to run the evolutionary algorithm, the suggested best indicators have been adapted to the Pyrenean seismicity. Finally, the algorithm

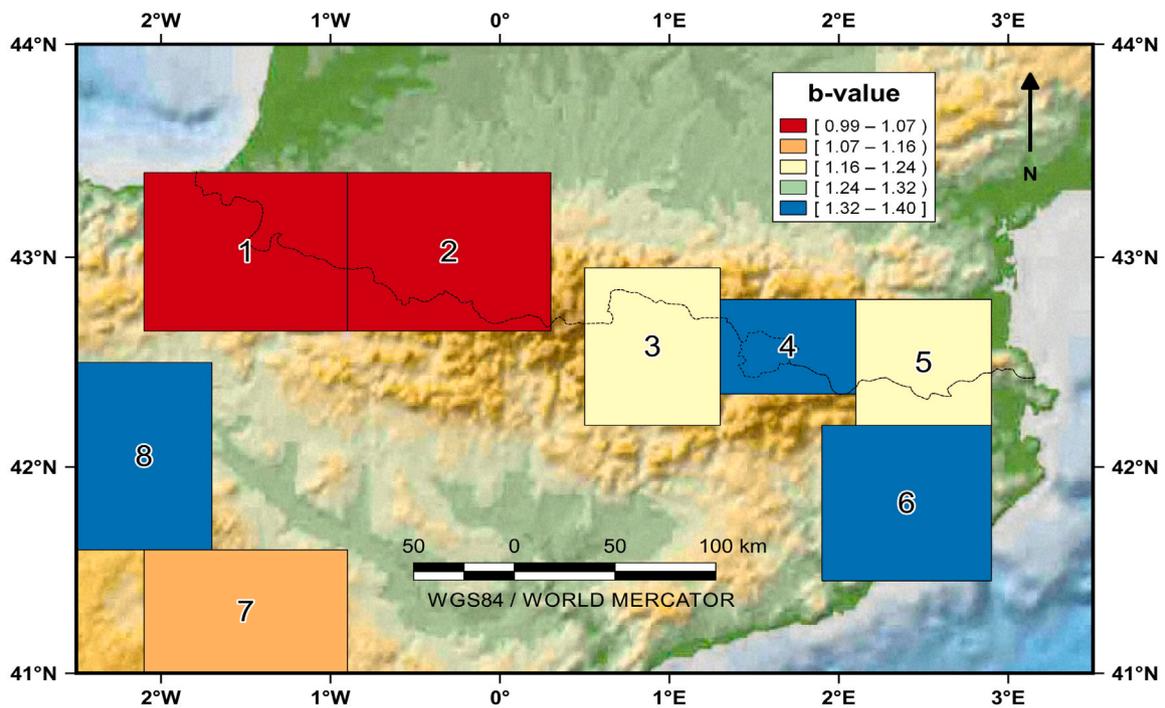


Fig. 7. Color map showing the  $b$ -value for the obtained source zones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

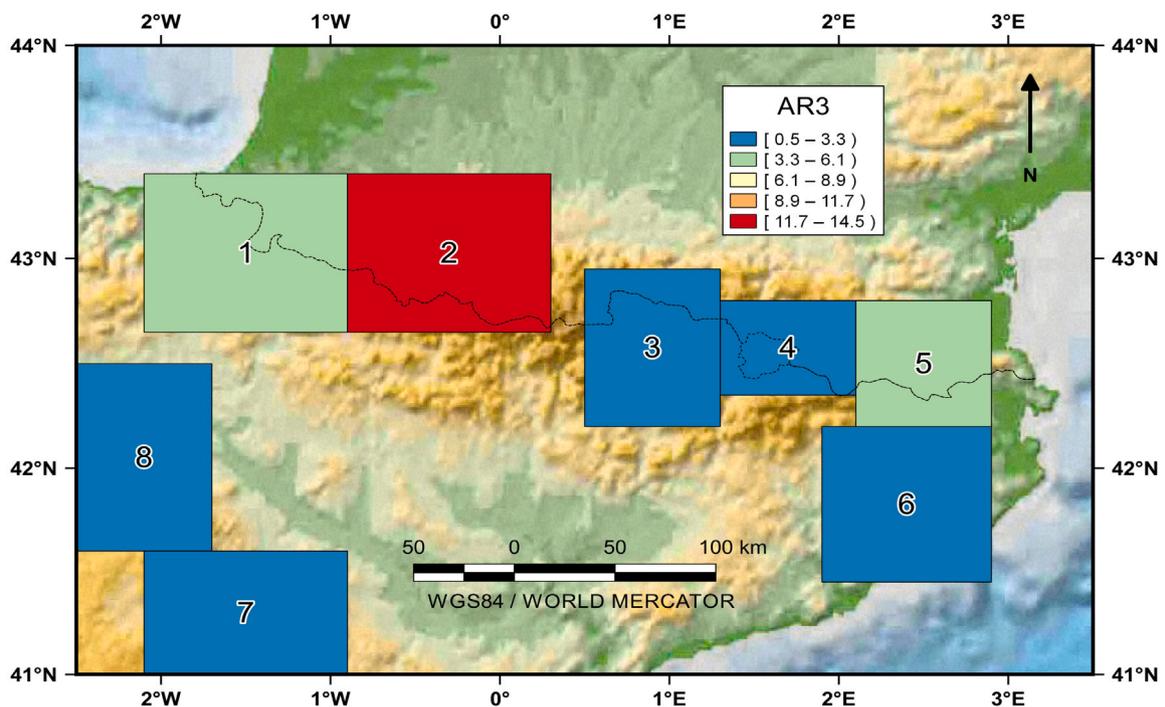


Fig. 8. Color map showing the  $AR3$  for the obtained source zones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

has generated the eight seismogenic zones, and their geometries have been integrated into a GIS.

The following step is characterizing the zones seismically. To this end, the most relevant seismic parameters have been calculated for each source zone by a GIS to further analysis. This analysis reveals that the source zone covered by the Arudy–Arette–Bagnères de Bigorre region is the most relevant seismically by far, due to its minimum  $b$ -value, the maximum annual rate, and the highest  $M_{max}$ . Therefore,

this zone deserves special attention when a PSHA is driven in the Pyrenees. The seismic hazard is also remarkable in both the Axial Zone, that includes the Maladetta Massif, and in the easternmost Pyrenees, along the Spain–France border, including  $M6+$  events over time. In the Catalanian Coastal Range, the zone rated as medium seismicity could be extended through this range to the southwest. In the Atlantic Pyrenees and the Andorra environment, the seismic activity is less pronounced. Finally, in the southwest of the studied region (even

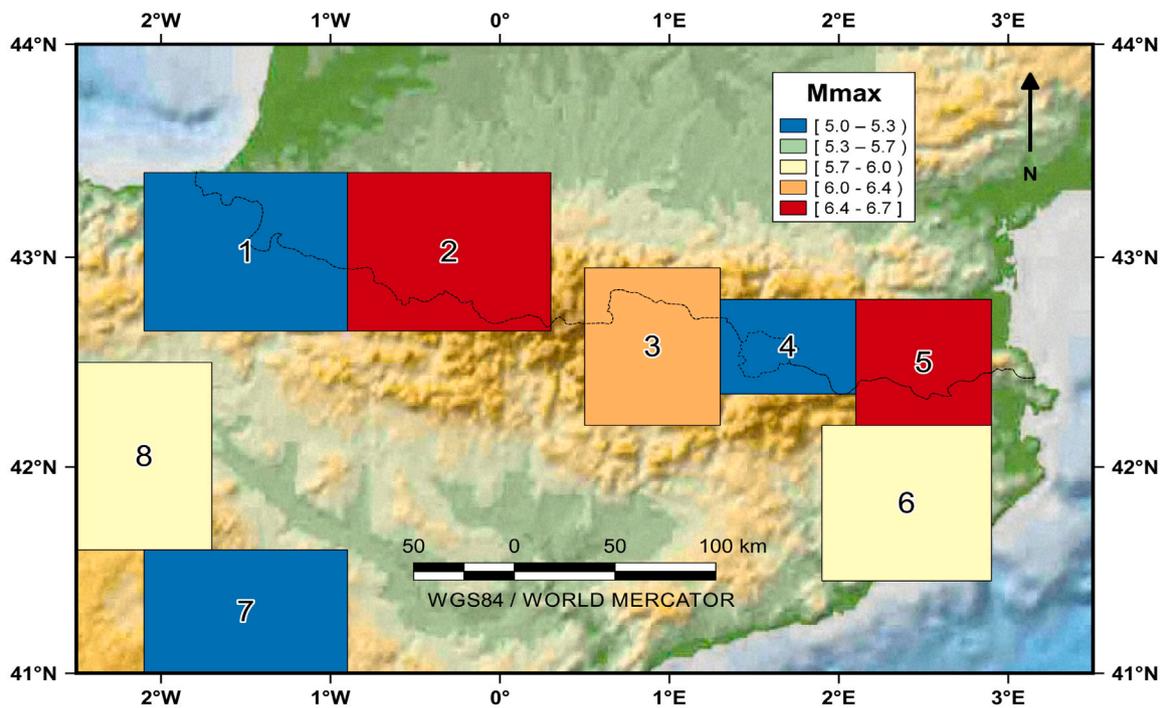


Fig. 9. Color map showing the  $M_{max}$  for the obtained source zones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

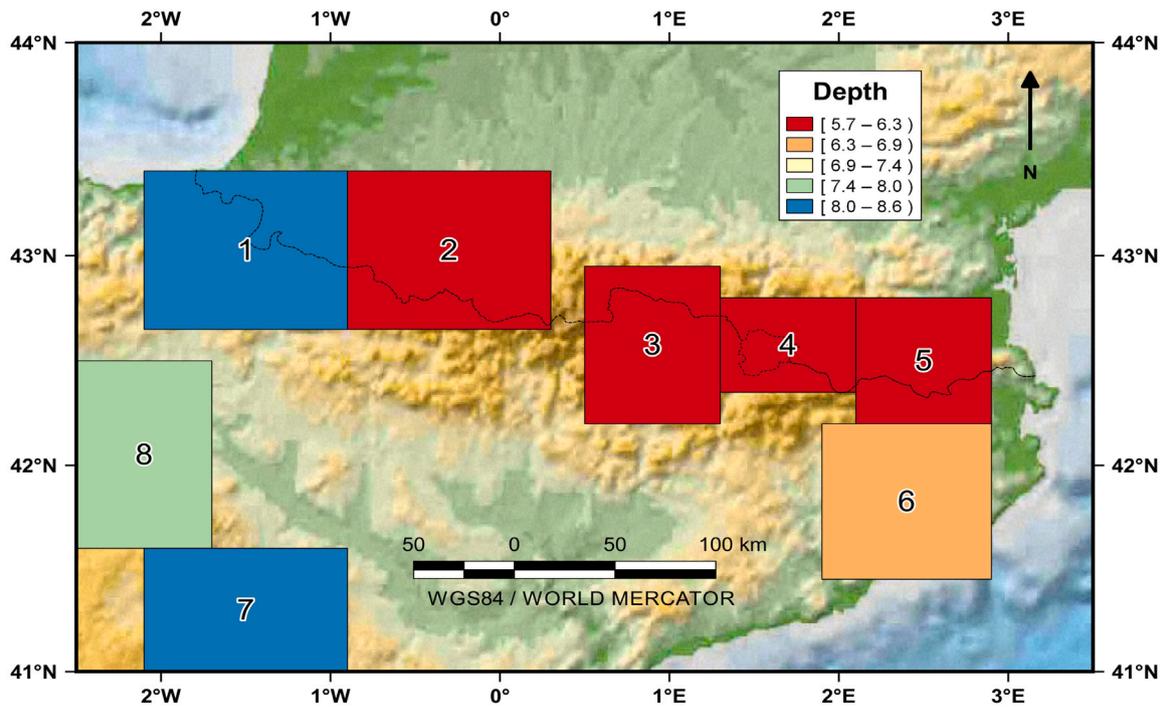


Fig. 10. Color map showing the mean depth for the obtained source zones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outside the Pyrenees environment), the seismic activity is significantly lower, and the two obtained zones should be merged into one.

The method has shown to be efficient as the delineated zones are different from one another and cover the vast majority of the region's epicenter. Besides, the most relevant point is that the procedure is almost human-free-action.

#### CRediT authorship contribution statement

**José L. Amaro-Mellado:** Conceive and design the experiments, Retrieve and analyze the data, Writing - original draft, Contribution to the writing of the manuscript, Agreement with manuscript results and conclusions. **Laura Melgar-García:** Contribution to the writing

of the manuscript, Agreement with manuscript results and conclusions. **Cristina Rubio-Escudero:** Contribution to the writing of the manuscript, Agreement with manuscript results and conclusions, Development of the structure and arguments of the paper, Revision and approbation of the final manuscript. **David Gutiérrez-Avilés:** Conceive and design the experiments, Retrieve and analyze the data, Writing - original draft, Contribution to the writing of the manuscript, Agreement with manuscript results and conclusions.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors want to thank the financial support given by the European Commission 0313-PERSISTAH project, the Spanish Ministry of Economy and Competitiveness project TIN2017-88209-C2, and the Junta de Andalucía US-1263341 project.

## Computer code availability

TriGen 3.5, GPL v3 license, <https://github.com/davgutavi/trlab-trigen>.

## References

Aki, K., 1965. Maximum likelihood estimate of  $b$  in the formula  $\log N = a - bM$  and its confidence limits. *Bull. Earthq. Res. Inst.* 43, 237–239.

Amaro-Mellado, J.L., Morales-Esteban, A., Asencio-Cortés, G., Martínez-Álvarez, F., 2017. Comparing seismic parameters for different source zone models in the Iberian Peninsula. *Tectonophysics* 717 (July), 449–472.

Amaro-Mellado, J.L., Tien Bui, D., 2020. GIS-based mapping of seismic parameters for the pyrenees. *ISPRS Int. J. Geo-Inf.* 9 (7), 452.

Baize, S., Cushing, E.M., Lemeille, F., Jomard, H., 2013. Updated seismotectonic zoning scheme of Metropolitan France, with reference to geologic and seismotectonic data. *Bull. Soc. Geol. France* 184 (3), 225–259.

Barani, S., Mascandola, C., Riccomagno, E., Spallarossa, D., Albarello, D., Ferretti, G., Scafidi, D., Augliera, P., Massa, M., 2018. Long-range dependence in earthquake-moment release and implications for earthquake occurrence probability. *Sci. Rep.* 8 (1), 1–11.

Bárdossy, G., Fodor, J., 2001. Traditional and new ways to handle uncertainty in geology. *Nat. Resour. Res.* 10 (3), 179–187.

Bender, B., 1983. Maximum likelihood estimation of  $b$  values for magnitude grouped data. *Bull. Seismol. Soc. Am.* 73 (3), 831–851.

Bernal, A., 2011. Anexo I del informe técnico IGN-PSE. ZF. P03. IGN-PSE. ZF. P03.

Bhar, A., Haubrock, M., Mukhopadhyay, A., Wingender, E., 2015. Multiobjective triclustering of time-series transcriptome data reveals key genes of biological processes. *BMC Bioinformatics* 16, 200.

Cabañas, L., Rivas-Medina, A., Martínez-Solares, J.M., Gaspar-Escribano, J.M., Benito, B., Antón, R., Ruiz-Barajas, S., 2015. Relationships between  $M_w$  and other earthquake size parameters in the Spanish IGN seismic catalog. *Pure Appl. Geophys.* 172 (9), 2397–2410.

Cheng, Y., Church, G.M., 2000. Biclustering of expression data. In: *International Conference on Intelligent Systems for Molecular Biology*. pp. 93–103.

Cornell, C.A., 1968. Engineering seismic risk analysis. *Bull. Seismol. Soc. Am.* 58, 1583–1606.

Corral, Á., 2006. Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophysics* 424 (3–4), 177–193.

Das, R., Sharma, M.L., Wason, H.R., Choudhury, D., Gonzalez, G., 2019. A seismic moment magnitude scale. *Bull. Seismol. Soc. Am.* 109 (4), 1542–1555.

Drouet, S., Ameri, G., Le Dortz, K., Secanell, R., Senfaute, G., 2020. A probabilistic seismic hazard map for the metropolitan France. *Bull. Earthq. Eng.* 18 (5), 1865–1898.

Gallart, J., Banda, E., Daignières, M., 1981. Crustal structure of the Paleozoic Axial Zone of the Pyrenees and transition to the North Pyrenean Zone. *Ann. Géophys.* 37 (3), 457–480.

Gallart, J., Daignières, M., Banda, E., Suriñach, E., Hirn, A., 1980. The eastern Pyrenean domain: lateral variations at crust-mantle level. *Ann. Geophys.* 36 (2), 141–158.

García-Mayordomo, J., 2015. Creación de un modelo de zonas sísmicas para el cálculo del mapa de peligrosidad sísmica de España. In: *Riesgos Geológicos/Geotecnia n° 5*, Instituto Geológico y Minero de España, p. 125.

García-Mayordomo, J., Insua-Arévalo, J.M., Martínez-Díaz, J.J., Jiménez-Díaz, A., Martín-Banda, R., Martín-Alfageme, S., Álvarez-Gómez, J.A., Rodríguez-Peces, M., Pérez-López, R., Rodríguez-Pascua, M.A., Masana, E., Perea, H., Martín-González, F., Giner-Robles, J., Nemser, E.S., Cabral, J., 2012a. The quaternary active faults database of Iberia (QAFI v.2.0). *J. Iberian Geol.* 38 (1), 285–302.

García-Mayordomo, J., Martínez-Díaz, J.J., Capote, R., Martín-Banda, R., Insua-Arévalo, J.M., Álvarez-Gómez, J.A., Perea, H., González, A., Lafuente, P., Martínez-González, F., Pérez-López, R., Rodríguez-Pascua, M.A., Giner-Robles, J., Azañón, J., Masana, E., Moreno, X., Benito, B., Rivas, A., Gaspar-Escribano, J.G., Cabañas, L., Vilanova, S., Fonseca, J., Nemser, E., Baize, S., 2012b. Modelo de zonas sísmicas para el cálculo de la peligrosidad sísmica en España. In: *Actas de la 7 Asamblea Geodesia y Geofísica*. pp. 23–28.

Gardner, J.K., Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian. *Bull. Seismol. Soc. Am.* 64 (5), 1363–1367.

Gnatyshak, D., Ignatov, D., Semenov, A., Poelmans, J., 2012. Gaining insight in social networks with biclustering and triclustering. In: *Perspectives in Business Informatics Research*. In: *Lecture Notes in Business Information Processing*, vol. 128, pp. 162–171.

González, Á., 2017. The Spanish National Earthquake Catalogue: Evolution, precision and completeness. *J. Seismol.* 21 (3), 435–471.

Gutenberg, B., Richter, C.F., 1944. Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* 34, 185–188.

Gutenberg, B., Richter, C.F., 1954. *Seismicity of the Earth*. Princeton University.

Gutiérrez-Avilés, D., Rubio-Escudero, C., 2014. LSL: A new measure to evaluate triclusters. In: *2014 IEEE International Conference on Bioinformatics and Biomedicine*. BIBM, IEEE, pp. 30–37.

Gutiérrez-Avilés, D., Rubio-Escudero, C., 2014. Mining 3D patterns from gene expression temporal data: A new tricluster evaluation measure. *Sci. World J.* 2014, 1–16.

Gutiérrez-Avilés, D., Rubio-Escudero, C., 2015. MSL: A measure to evaluate three-dimensional patterns in gene expression data. *Evol. Bioinform.* 11, 121–135.

Gutiérrez-Avilés, D., Rubio-Escudero, C., Martínez-Álvarez, F., Riquelme, J.C., 2014. TriGen: A genetic algorithm to mine triclusters in temporal gene expression data. *Neurocomputing* 132, 42–53.

Hamdache, M., Peláez, J.A., 2019. Comment on the paper “seismic hazard analysis of surface level, using topographic condition in the northeast of Algeria” by Mouloud Hamidatou, Mohammedi Yahia, Abdelkrim Yelles-Chaouche, Itharam Thallak, Dietrich Stromeyer, Saad Lebdioui, Fabrice Cotton. *Pure Appl. Geophys.*

Hanks, T., Kanamori, H., 1979. Moment magnitude scale. *J. Geophys. Res.* 84 (B5), 2348–2350.

Henriques, R.U.I., Madeira, S.C., 2018. Triclustering algorithms for three-dimensional data analysis: A comprehensive survey. *ACM Comput. Surv.* 51 (5).

Holland, J.H., 1992. Genetic algorithms. *Sci. Am.* 267 (1), 66–73.

IGN-UPM-WorkingGroup, 2013. Actualización de mapas de peligrosidad sísmica 2012. Instituto Geográfico Nacional, p. 267.

Instituto Geográfico Nacional, 2020. Catálogo de terremotos. Instituto Geográfico Nacional, URL: <http://www.ign.es/web/ign/portal/sis-catalogo-terremotos>.

Kijko, A., Smit, A., 2012. Extension of the Aki-Utsu  $b$ -value estimator for incomplete catalogs. *Bull. Seismol. Soc. Am.* 102 (3), 1283–1287.

Lacan, P., Ortuño, M., 2012. Active Tectonics of the Pyrenees: A review. *J. Iberian Geol.* 38 (1), 9–30.

Li, A., Tuck, D., 2009. An effective tri-clustering algorithm combining expression data with gene regulation information. *Gene Regul. Syst. Biol.* 3, 49–64.

Liu, J., Li, Z., Hu, X., Chen, Y., 2008. Multi-objective evolutionary algorithm for mining 3D clusters in gene-sample-time microarray data. In: *2008 IEEE International Conference on Granular Computing*. pp. 442–447.

Marin, S., Avouac, J.P., Nicolas, M., Schlupp, A., 2004. A probabilistic approach to seismic hazard in metropolitan France. *Bull. Seismol. Soc. Am.* 94 (6), 2137–2163.

Martín, A.J., 1984. Riesgo sísmico en la península Ibérica (Ph.D.). Instituto Geográfico Nacional.

Martin, C., Secanell, R., Combes, P., Lignon, G., 2002. Preliminary probabilistic seismic hazards assessment of France. In: *12th European Conference in Earthquake Engineering*. London. p. 870.

Martínez-Álvarez, F., Asencio-Cortés, G., Torres, J.F., Gutiérrez-Avilés, D., Melgar-García, L., Pérez-Chacón, R., Rubio-Escudero, C., Riquelme, J.C., Troncoso, A., 2020. Coronavirus optimization algorithm: A bioinspired metaheuristic based on the COVID-19 propagation model. *Big Data* 8 (4), 308–322.

Martínez-Álvarez, F., Gutiérrez-Avilés, D., Morales-Esteban, A., Reyes, J., Amaro-Mellado, J., Rubio-Escudero, C., 2015. A novel method for seismogenic zoning based on triclustering: Application to the Iberian peninsula. *Entropy* 17 (12), 5000–5021.

McGuire, R.K., 1976. FORTRAN Computer Program for Seismic Risk Analysis. Technical Report 76-67, US Geological Survey Open-File Report, p. 90.

Melgar-García, L., Gutiérrez-Avilés, D., Rubio-Escudero, C., Troncoso, A., 2020. High-content screening images streaming analysis using the STriGen methodology. In: *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. SAC '20, Association for Computing Machinery, pp. 537–539.

Mezcuá, J., Rueda, J., García Blanco, R.M., 2011. A new probabilistic seismic hazard study of Spain. *Nat. Hazards* 59 (2), 1087–1108.

Ministerio de Fomento (Gobierno de España), 2002. Norma de la Construcción Sismorresistente Española (NCSE-02). Boletín Oficial del Estado.

- Molina, S., 1998. Sismotectónica y peligrosidad sísmica del área de contacto entre Iberia y África. Universidad de Granada, Spain.
- Morales-Esteban, A., Martínez-Álvarez, F., Scitovskis, S., Scitovski, R., 2014. A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Comput. Geosci.* 73, 132–141.
- Njike-Kassala, J.D., Souriau, A., Gagnepain-Beyneix, J., Martel, L., Vadell, M., 1992. Frequency-magnitude relationship and Poisson's ratio in the Pyrenees, in relation to earthquake distribution. *Tectonophysics* 215 (3–4), 363–369.
- Pearson, K., Filon, L.N.G., 1898. Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Phil. Trans. R. Soc. Lond. Ser. A* 229–311, Containing Papers of a Mathematical or Physical Character.
- Pecker, A., Faccioli, E., Gurpinar, A., Martin, C., Renault, P.L.A., 2017. An Overview of the SIGMA Research Project: A European Approach to Seismic Hazard Analysis. In: *Geotechnical, Geological and Earthquake Engineering*, vol. 42, Springer, p. 172.
- Perea, H., 2009. The Catalan seismic crisis (1427 and 1428; NE Iberian Peninsula): Geological sources and earthquake triggering. *J. Geodyn.* 47 (5), 259–270.
- Pontes, B., Giráldez, R., Aguilar-Ruiz, J.S., 2015. Biclustering on expression data: A review. *J. Biomed. Inform.* 57, 163–180.
- Reyes, J., Cárdenas, V.H., 2010. A Chilean seismic regionalization through a Kohonen neural network. *Neural Comput. Appl.* 19 (7), 1081–1087.
- Rigo, A., Souriau, A., Sylvander, M., 2018. Spatial variations of b-value and crustal stress in the Pyrenees. *J. Seismol.* 22 (1), 337–352.
- Rigo, A., Vernant, P., Feigl, K.L., Goula, X., Khazaradze, G., Talaya, J., Morel, L., Nicolas, J., Baize, S., Chery, J., Sylvander, M., 2015. Present-day deformation of the Pyrenees revealed by GPS surveying and earthquake focal mechanisms until 2011. *Geophys. J. Int.* 201 (2), 947–964.
- Scitovskis, S., 2018. A density-based clustering algorithm for earthquake zoning. *Comput. Geosci.* 110, 90–95.
- Scitovski, R., Scitovskis, S., 2013. A fast partitioning algorithm and its application to earthquake investigation. *Comput. Geosci.* 59, 124–131.
- Secanell, R., Bertil, D., Martin, C., Goula, X., Susagna, T., Tapia, M., Dominique, P., Carbon, D., Fleta, J., 2008. Probabilistic seismic hazard assessment of the Pyrenean region. *J. Seismol.* 12 (3), 323–341.
- Secanell, R., Martin, C., Goula, X., Susagna, T., Tapia, M., Bertil, D., 2007. Evaluación probabilista de la peligrosidad sísmica de la región pirenaica. In: 3º Congreso Nacional de Ingeniería Sísmica, no. 1. Asociación Española de Ingeniería Sísmica, pp. 1–17.
- Skordas, E., Kulhánek, O., 1992. Spatial and temporal variations of Fennoscandian seismicity. *Geophys. J. Int.* 111 (3), 577–588.
- Souriau, A., Rigo, A., Sylvander, M., Benahmed, S., Grimaud, F., 2014. Seismicity in central-western Pyrenees (France): A consequence of the subsidence of dense exhumed bodies. *Tectonophysics* 621, 123–131.
- Spearman, C., 1910. Correlation calculated from faulty data. *Br. J. Psychol.*, 1904–1920 3 (3), 271–295.
- Stucchi, M., Rovida, A., Gomez Capera, A.A., Alexandre, P., Camelbeeck, T., Demircioglu, M.B., Gasperini, P., Kouskouna, V., Musson, R.M.W., Radulian, M., Sesetyan, K., Vilanova, S., Baumont, D., Bungum, H., Fäh, D., Lenhardt, W., Makropoulos, K., Martinez Solares, J.M., Scotti, O., Živčić, M., Albin, P., Batllo, J., Papaioannou, C., Tatevossian, R., Locati, M., Meletti, C., Viganò, D., Giardini, D., 2013. The SHARE European earthquake catalogue (SHEEC) 1000-1899. *J. Seismol.* 17 (2), 523–544.
- Sylvander, M., Souriau, A., Rigo, A., Tocheport, A., Toutain, J.P., Ponsolles, C., Benahmed, S., 2008. The 2006 November, M L = 5.0 earthquake near Lourdes (France): new evidence for NS extension across the Pyrenees. *Geophys. J. Int.* 175 (2), 649–664.
- Talbi, A., Nanjo, K., Satake, K., Zhuang, J., Hamdache, M., 2013. Comparison of seismicity declustering methods using a probabilistic measure of clustering. *J. Seismol.* 17 (3), 1041–1061.
- Vissers, R.L.M., Meijer, P.T., 2012. Iberian plate kinematics and Alpine collision in the Pyrenees. *Earth-Sci. Rev.* 114 (1–2), 61–83.
- Woessner, J., Danciu, L., Kästli, P., Monelli, D., 2011. Grant agreement no. 226967 seismic hazard harmonization in Europe project acronym: SHARE. pp. 1–23, SHARE, 226967.
- Zhao, L., Zaki, M., 2005. TRICLUSTER: an effective algorithm for mining coherent clusters in 3D microarray data. In: *Proc. of the 2005 ACM SIGMOD International Conference on Management of Data*. pp. 694–705.