Nat. Hazards Earth Syst. Sci., 13, 2337–2351, 2013
www.nat-hazards-earth-syst-sci.net/13/2337/2013/
doi:10.5194/nhess-13-2337-2013

Natural Hazards and Earth System Sciences
Open Access

Seismic zones for Azores based on statistical criteria

M. C. M. Rodrigues[1] **and C. S. Oliveira**[2]

[1]Faculdade de Ciências e Tecnologia, Caparica, Portugal
[2]Instituto Superior Técnico, Lisbon, Portugal

Correspondence to: M. C. M. Rodrigues (mcmr@fct.unl.pt)

Received: 6 December 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –
Revised: 18 July 2013 – Accepted: 23 July 2013 – Published: 24 September 2013

Abstract. The objective of this paper is to define seismic zones in the Azores based on statistical criteria. These seismic zones will likely be used in seismic simulations of occurrences in the Azores Archipelago.

The data used in this work cover the time period from 1915 to 2011. The Azores region was divided into $1^{\circ} \times 1^{\circ}$ area units, for which the seismicity and the maximum magnitudes of events were calculated.

The seismicity, the largest earthquakes recorded and the geological characteristics of the region were used to group these area units because similar seismic zones must delineate areas with homogeneous seismic characteristics. We have identified seven seismic zones.

To verify that the defined areas differ statistically, we considered the following dissimilarity measures (variables): time, size and seismic conditions – the number of seismic events with specific characteristics.

Statistical tests, particularly goodness-of-fit tests, allowed us to conclude that, considering these three variables, the seven earthquake zones defined here are statistically distinct.

1 Introduction

The Azores Archipelago is located at the triple junction of the Mid-Atlantic Rift, where the Eurasian, Nubian, and American Plates meet.

The intense seismic activity in the region has been studied by many authors (e.g., Bezzeghoud et al., 2008; Borges et al., 2008).

As shown in Fig. 1a, the Azores consists of nine islands distributed among three different groups: the islands of Flores and Corvo, constituting the Western Group; the islands of Terceira, Graciosa, São Jorge, Faial and Pico, which are part of the Central Group; and the islands of São Miguel and Santa Maria in the Eastern Group.

Figure 1b shows epicenters in the Azores between 1915 and 2011, and Fig. 1c shows a zoomed-in map of epicenters of the islands.

The aim of this study is to define the seismic zones of the Azores, which will later be used for seismic simulations of the region.

We define several zones that express differences in seismicity, while allowing for a model that is not overly complex. Seismic zones are defined by polygons that delineate areas of homogeneous seismicity characteristics. They are also based on differences in geology and tectonics, but seismicity is the main characteristic in defining them (e.g., Reiter, 1991; Kagan et al., 2010).

In this work, the main criterion to define the zones is the recorded seismicity, as different zones should exhibit different statistical characteristics. The number of events is the most important variable used in this study; magnitude is also used, although large events may be infrequent.

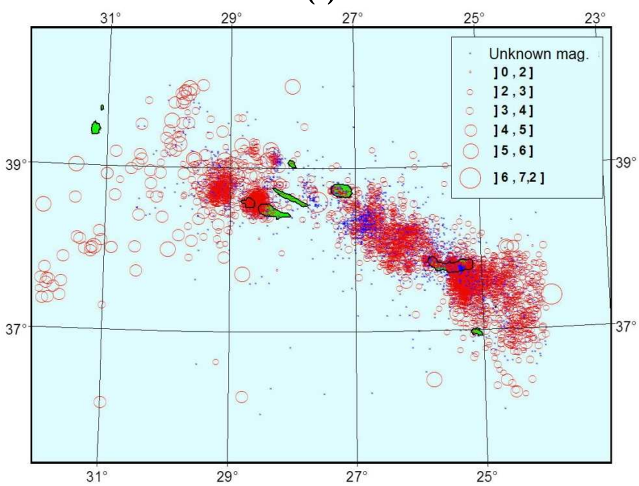
Nunes et al. (2000) delineated a 28-seismic-zone model based on the distribution of epicenters and on the tectonics of Azores region. Due to the lack of seismic data, the model was simplified for use in hazard assessment to include nine main zones in order to allow a reliable statistical characterization of the model (Carvalho et al., 2001).

With the upgrade of the seismological network in the Azores in recent decades, seismic data have become more reliable and complete for magnitudes greater than $M_{\mathrm{L}} = 3$. This allows for more robust statistical analyses than were possible in the past.

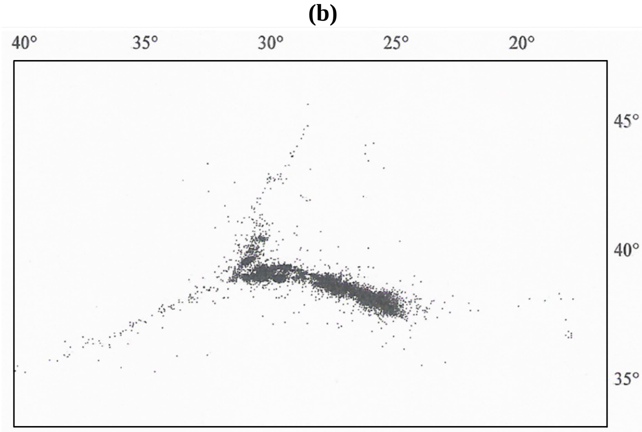
The seismic zones cover geophysical units where data are available. For each unit, we computed the following:

– The number of events.

(a)



(b)



(c)

Fig. 1. **(a)** The Azores Archipelago, **(b)** epicentral map for 1915–2011 and **(c)** zoom of epicenters for the islands.

– The maximum magnitude recorded.

We grouped the 28 areas of Nunes et al. (2000) into seven zones that exhibit different characteristics. We used several statistic tests (parametric and nonparametric) to confirm whether these seven zones were significantly different.

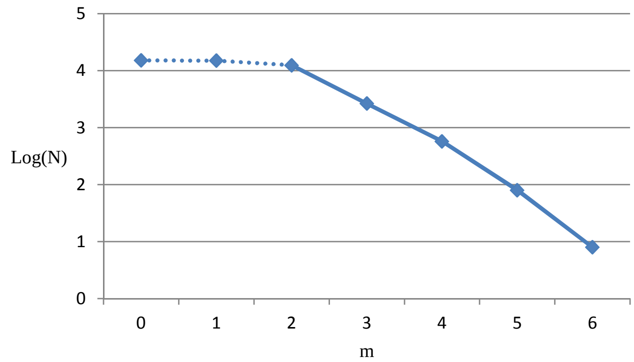


Fig. 2. Gutenberg–Richter plot for the Azores region showing all catalog seismicity.

2 Data

The data used in this work were gathered from two sources. For the period 1915–1998, we used the catalog of Nunes et al. (2004), and for the period 1999–2011, data were directly obtained from Instituto de Meteorologia (2011).

The earlier period covers the region encompassed by 11.50° W–42.86° W and 10.80° N–47.54° N. A total of 9214 earthquake records are available, of which 5456 have information on Richter-scale magnitudes.

The catalog for the later period covers the area within 21.31° W–35.42° W and 34.30° N–45.57° N, and contains 9608 earthquakes, all of which include magnitude information.

Table 1 summarizes the main characteristics of the data used.

The data were analyzed as a whole, including foreshocks and aftershocks. Fig. 2 shows a Gutenberg–Richter plot, which indicates that the dataset is not complete. Many small-magnitude events occur in the sea, far from the seismic network, and thus are not recorded. According to the Gutenberg–Richter law, a linear trend should exist between Log N and m:

$$\text{Log}N(m) = a - b \times m, \qquad (1)$$

where N is the number of events of magnitude greater than m, and a and b are constants fitted to the data.

Removing earthquakes smaller than magnitude 2, a least-squares approximation leads to

$$\text{Log}N(m) = 5.77611 - 0.79\ \text{m}, \qquad (2)$$

with a correlation coefficient $R = -0.996$, which indicates a significant linear correlation and that the catalog is complete for earthquakes with magnitude larger than 2.

If we consider only events with magnitudes greater than 2, much of the dataset would be lost (the value 2 corresponds approximately to the 0.40 quantile of magnitude; see Table 4), and the aim of this paper is not to estimate the

Table 1. Data characteristics.

Data characteristics	
Period of time covered (in years)	1915 to 2011
Total number of records	18822
Records containing information of magnitude	15065
Records without information of magnitude	3757
Records containing information of magnitude and intensity	499
Records with intensity information and without information of magnitude	247

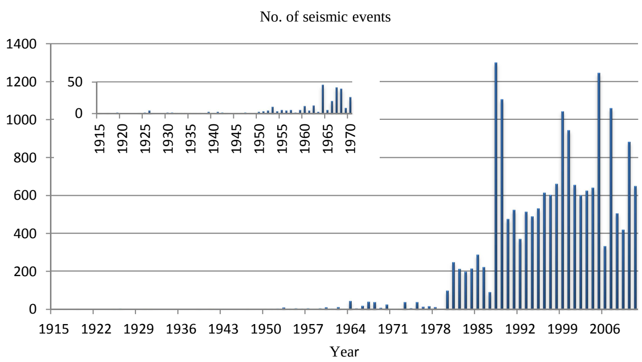


Fig. 3. Annual seismicity.

Table 2. Statistics of the AS.

Statistic	Value
Mean	200.39
Standard deviation	330.76
Skewness	1.63
Kurtosis	4.76
Minimum	0
Quantile	
0.1	1
0.2	1
0.3	2
0.4	4
0.5	7
0.6	28.6
0.7	212.8
0.8	499.8
0.9	648.2
1 (max.)	1300
Number of years	95

constants a and b of the Gutenberg–Richter law. Therefore, we consider all earthquakes with a catalog magnitude greater than 0, which corresponds to events for which the magnitude has not been determined.

3 Exploratory data analysis

We used the R® software (e.g., Dalgaard, 2008; Venables et al., 2011) to perform the statistical analysis in this study. For some calculations, we also used the Turbo Pascal® software.

3.1 Annual seismicity

The seismic records contain information about the year, month, day, hour, minute and second of each event. For straightforward computation, time was converted into decimal years.

Consider the variable annual seismicity (AS), which represents the number of earthquakes that occurred in one year.

Figure 3 displays the AS for the period from 1915 to 2011.

The AS is very heterogeneous throughout the study period, and it appears to increase in 1960. This increase reflects the expansion of the seismic network in the Azores Archipelago.

Table 2 shows the main statistical properties of the AS for the period of 1915–2011.

Figure 4 shows a histogram of the AS, which is highly variable, varying between fewer than 200 earthquakes in one year to more than 1000.

The heterogeneity of the data suggests that we should use records from 1960 onwards, as this produces a dataset that best reflects the actual seismicity.

3.2 Statistical study of some characteristics of seismic events

Each earthquake can be characterized by three variables: time, size and space.

The variable time (Dt) is characterized by the time intervals between consecutive earthquakes, the variable size (S) is the Richter magnitude associated with an earthquake and the space variable (Sp) gives the number of the zone corresponding to the epicenter of the earthquake. However at this stage, Sp is characterized by the latitude and longitude of each earthquake.

Figure 5 describes a schematic representation of the seismic process of occurrences, where S_i represents the size of earthquake i, Dt_i the time interval between this event and the preceding one $(i-1)$ and Sp_i the location of event i.

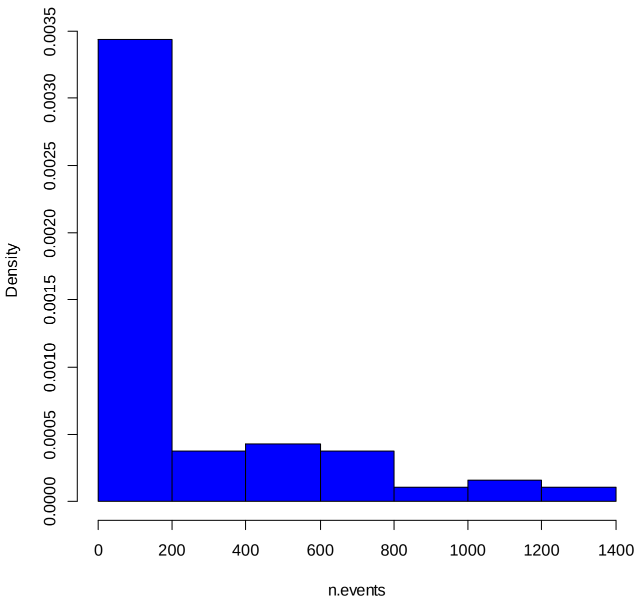


Fig. 4. Histogram of the number of seismic events per year.

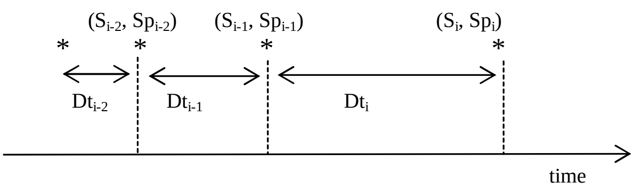


Fig. 5. Schematic representation of the seismic process of occurrences.

3.2.1 Study of the time variable

As previously stated, Dt represents the time intervals between consecutive earthquakes, expressed in years, during the time period of 1915–2011. Let Dt_{60} be the variable that typifies the time intervals between consecutive earthquakes between 1960 and 2011.

Table 3 shows the statistics of the variables Dt and Dt_{60}, with the largest difference observed in their maximum values. While the maximum value of Dt is approximately 5 yr, the maximum value of Dt_{60} is only 0.78 yr, which is less than nine months. The mean Dt is approximately half of the mean Dt_{60}. Comparing the quantiles of these two variables, there is no significant difference below the 0.90 quantile, indicating that the major difference is in the maximum values of the variables.

Figure 6a and b present the histograms of Dt and Dt_{60}, respectively, which show clear difference between the two random variables.

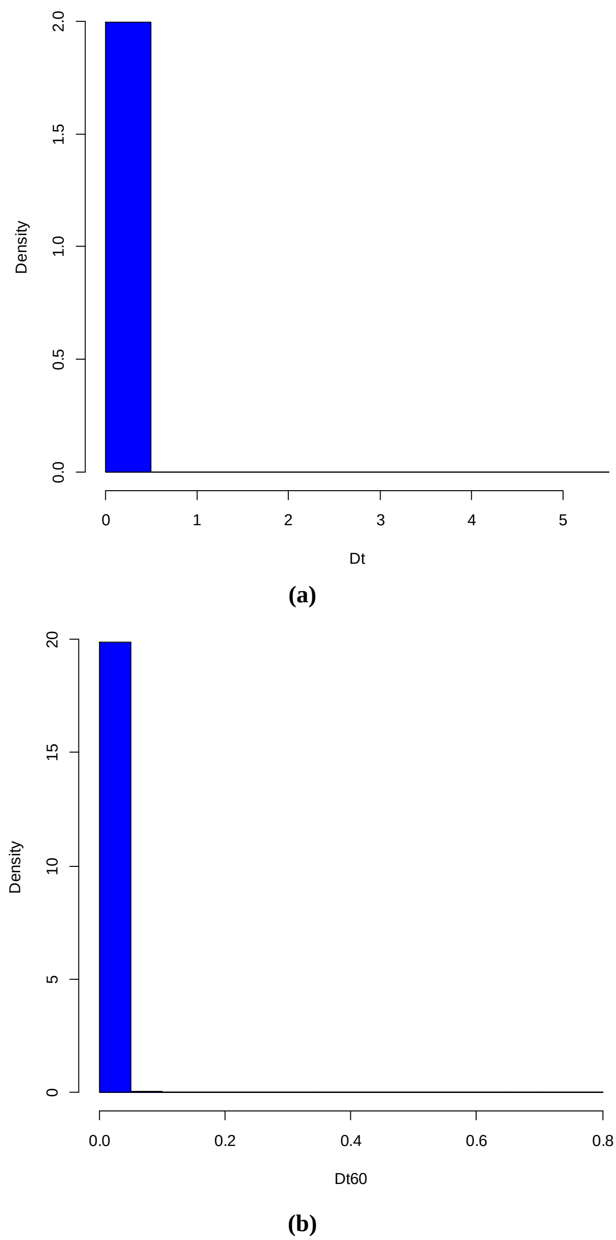


Fig. 6. (a) Histogram of Dt, and **(b)** histogram of Dt_{60}.

3.2.2 Study of the size variable

As described in Sect. 3.2.1, *S* represents the size of each earthquake between 1915 and 2011.

As shown in Table 1, 3757 seismic records do not include magnitude; the magnitudes are null values in the catalog. If these records are not removed, they will influence the statistics of *S*.

In addition, if the earthquakes with null magnitudes were ignored, the time intervals between consecutive events would increase.

Table 3. Statistics of Dt and Dt_{60}.

Statistics of Dt		Statistics of Dt_{60}	
Mean	0.0050207	Mean	0.0027532
Standard deviation	0.0632967	Standard deviation	0.0180847
Skewness	44.61	Skewness	24.8
Kurtosis	2823.92	Kurtosis	798.7
Minimum	0.00000000	Minimum	0.00000000
Quantile		Quantile	
0.1	0.00000770	0.1	0.00000760
0.2	0.00003982	0.2	0.00003884
0.3	0.00011980	0.3	0.00011790
0.4	0.00027390	0.4	0.00027020
0.5	0.00053885	0.5	0.00053270
0.6	0.00093420	0.6	0.00091922
0.7	0.00157952	0.7	0.00155820
0.8	0.00263610	0.8	0.00260176
0.9	0.00492949	0.9	0.00479008
1 (max.)	5.0392669	1 (max.)	0.77849250
Total number of records	18 821	Total number of records	18 733

Let S_{w0} represent the earthquake magnitudes, excluding the zero values of each earthquake between 1915 and 2011.

Table 4 summarizes the statistics calculated for S and S_{w0}.

As expected, S_{w0} has a larger mean than S, and the standard deviation of S_{w0} is less than for S. The quantiles of S_{w0} are greater than similar quantiles of S, except for the 1.0 quantile (maximum).

Figure 7a presents a histogram of the absolute frequencies of S. The large number of zero magnitudes is due to earthquakes with unknown magnitudes.

The histogram displayed in Fig. 7b shows the asymmetry of the probability density function of S_{w0}, with a significant tail for large values of S_{w0} and a positive skewness coefficient.

4 Definition of seismic zones

As previously described, the main goal of this study is to identify regions with significant differences in seismicity. We use the number of events and the *maximum magnitude recorded* to identify these differences. The region included in the dataset was divided into $1° \times 1°$ area units, and the number of earthquakes recorded in the period from 1915 to 2011 was computed for each area unit. Let Sq represent the number of events between 1915 and 2011 in each area unit.

Figure 8a shows the values of Sq for the region bounded by 40° W–15° W and 30° N–47° N.

There is a band of increased seismicity (values above the 0.8 quantile of AS) with an approximately WNW–ESE orientation, which covers the Eastern and Central groups of the Azores Archipelago, as well as the NW Faial region, the trench west of Graciosa, the D. João de Castro Bank and the Hirondelle Trench.

However, for roughly half of this band, there is a slight decrease in the AS of the region bounded by 36° N–39° N and 27° W–28° W.

A region with high values of AS, although lower than for the WNW–ESE band, is oriented approximately SSW–NNE, and includes the islands of the Western Group and the northern Mid-Atlantic Ridge.

East of the WNW–ESE band, there is a region with nearly E–W orientation, in which the AS is also elevated.

The maximum magnitude recorded was also computed for each area unit during the study period.

Figure 8b shows that the WNW–ESE and SSW–NNE bands of seismicity also have higher maximum recorded magnitudes, with the largest magnitude, 8.2, recorded in the E–W band.

In the WNW–ESE band, two centers of high magnitudes are highlighted, one of which covers the Central Group of the Archipelago, particularly the western region of Faial Island, and the other covers the Eastern Group of the Archipelago, with an emphasis on São Miguel Island.

The seismic zones were defined by aggregating area units according to the aforementioned patterns, with an emphasis on the seismicity and the maximum magnitude recorded. Differences in geomorphology were also taken into account.

In the region within 11.50° W–42.54° W, 10.80° N–47.54° N, the following seven seismic zones were defined (Table 5, Fig. 9a and b):

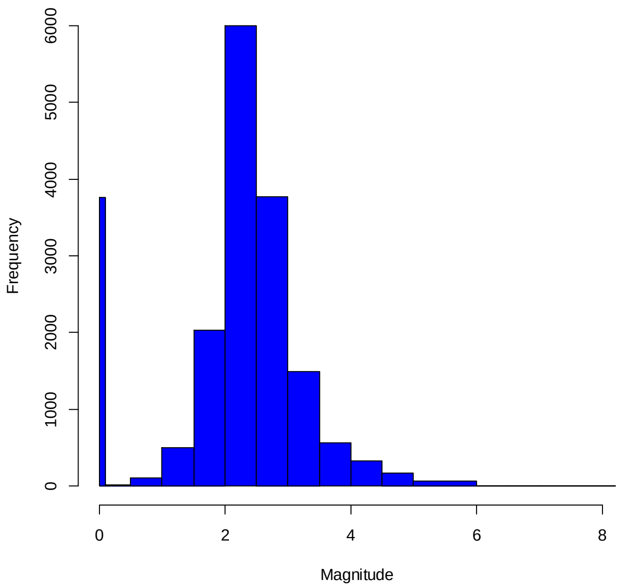
Zone 1 comprises the Western Group of the Azores Archipelago and is situated NW of the Mid-Atlantic Ridge. It presents low values of seismicity, and the maximum

Table 4. Statistics of S and S_{w0}.

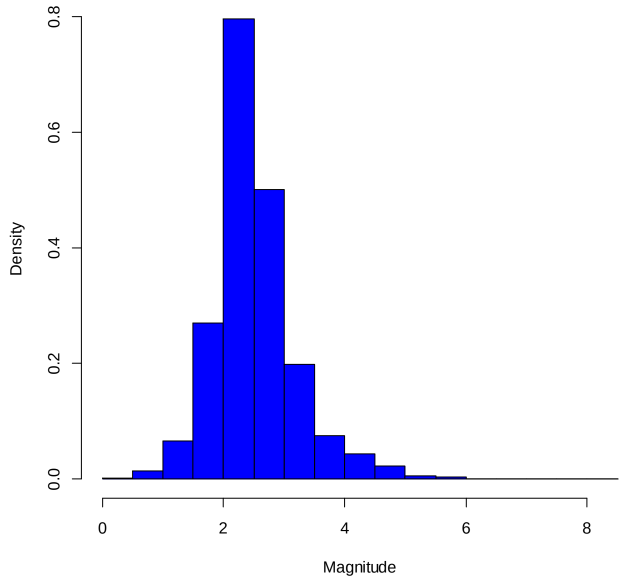
Statistics of S		Statistics of S_{w0}	
Mean	2.05	Mean	2.56
Standard deviation	1.19	Standard deviation	0.68
Skewness	−0.45	Skewness	1.11
Kurtosis	2.88	Kurtosis	6.21
Minimum	0.0	Minimum	0.2
Quantile		Quantile	
0.1	0.0	0.1	2.0
0.2	0.5	0.2	2.1
0.3	2.0	0.3	2.2
0.4	2.1	0.4	2.3
0.5	2.3	0.5	2.4
0.6	2.4	0.6	2.6
0.7	2.6	0.7	2.8
0.8	2.9	0.8	3.0
0.9	3.2	0.9	3.4
1 (max.)	8.2	1 (max.)	8.2
Total number of records	18 822	Total number of records	15 065

Table 5. Seismic zones.

Zone	Definition	Designation
Zone 1	[lat ≤ 39 and lat ≥ (long + 70)] OR [lat >39 and long ≤ (−31)]	West of Mid-Atlantic Ridge
Zone 2	[lat ≤ 39 and lat <(long + 70) and lat ≥ (long + 68)] OR [lat >39 and long >(−31) and lat ≥ (long + 68)]	Mid-Atlantic Ridge
Zone 3	lat ≥ (−0.4 × long + 28.2) and lat <(long + 68) and lat ≥ 38	Northeast of Mid-Atlantic Ridge
Zone 4	lat <(−0.4 × long + 28.2) and lat ≥ (−0.2 × long + 32) and lat <(long + 68) and lat ≥ (1.5 × long + 78.75)	Azores Island Central Group
Zone 5	lat <(1.5 × long + 78.75) and long ≤ (−24.5) and lat <(−0.4 × long + 28.2) and lat ≥ (−0.833 × long + 14.583)	Azores Island Eastern Group
Zone 6	long >(−24.5) and lat <38 and lat ≥ 35	Gloria Fault
Zone 7	lat <(−0.2 × long + 32) and lat <(long + 68) and lat <(−0.833 × long + 14.583)	South of Azores Islands



(a)



(b)

Fig. 7. (a) Histogram of absolute frequencies of S, and **(b)** histogram of S_{w0}.

magnitude recorded is 6.2. The islands of Flores and Corvo are in this zone.

Zone 2 is a maritime zone corresponding to the Mid-Atlantic Ridge and its transform faults to the north. This zone also comprises the North Azores Fracture Zone. It has high levels of seismicity and a maximum magnitude of 6.0.

Zone 3 is a maritime zone with very low seismicity, located NE of the Central and Eastern groups of the Archipelago and east of the Mid-Atlantic Ridge. The maximum magnitude recorded is 4.7, the lowest maximum magnitude for all zones.

Zone 4 encompasses the Central Group of the Archipelago, west of Capelinhos and the Terceira Rift central sector. It features very high seismicity and a maximum magnitude of 6.0. Compared to the maximum magnitudes recorded in other zones, this magnitude is not very high, indicating that the main characteristic of this zone is the high seismicity and not its maximum magnitude. This zone contains five islands: Faial, Pico, São Jorge, Terceira and Graciosa.

Zone 5 comprises the Eastern Group of the Archipelago, the Hirondelle Trench and the D. João de Castro Bank. It has the highest seismicity of all seven zones, and the maximum magnitude recorded is 7.0. This zone is characterized not only by its high seismicity but also by its high maximum magnitude recorded. This zone contains two islands: São Miguel and Santa Maria.

Zone 6 is a maritime zone and includes the Gloria Fault. The seismicity is moderate, but this zone has the highest magnitude of all zones: 8.2. It is characterized by a moderate number of earthquakes, which can be of relatively high magnitude.

Zone 7 is a maritime zone and is the furthest south of all seismic zones. It has the lowest seismicity, and the maximum magnitude recorded is 6.1.

Zones 1, 3 and 7 include small numbers of events compared to the other seismic zones. Therefore, they are considered to be *background zones*.

The statistical study focuses primarily on zones 2, 4, 5 and 6, although all zones were examined initially.

We calculated the number of earthquakes recorded between 1915 and 2011 for each seismic zone. Table 6 and Fig. 10 summarize the results.

4.1 Statistical study of the time and size variables for each seismic zone

In the statistical study of the time variable, characterized by the time intervals between consecutive events, only the period 1960–2011 was considered.

For the size variable, data from all time periods were considered, but the null values were not taken into consideration.

4.1.1 Time

Consider $\mathrm{Dt}_{60,i}, i \in \{1,2,3,4,5,6,7\}$, the variable that represents the time interval between an event and its previous event, both in zone i, in 1960 and later.

Table 7 summarizes the statistics calculated for $\mathrm{Dt}_{60,i}, i \in \{1,2,3,4,5,6,7\}$.

4.1.2 Size

Let $S_{w0,i}$ represent the nonzero magnitudes in the zone i, $i \in \{1,2,3,4,5,6,7\}$.

Table 8 condenses the statistics computed for $S_{w0,i}, i \in \{1,2,3,4,5,6,7\}$.

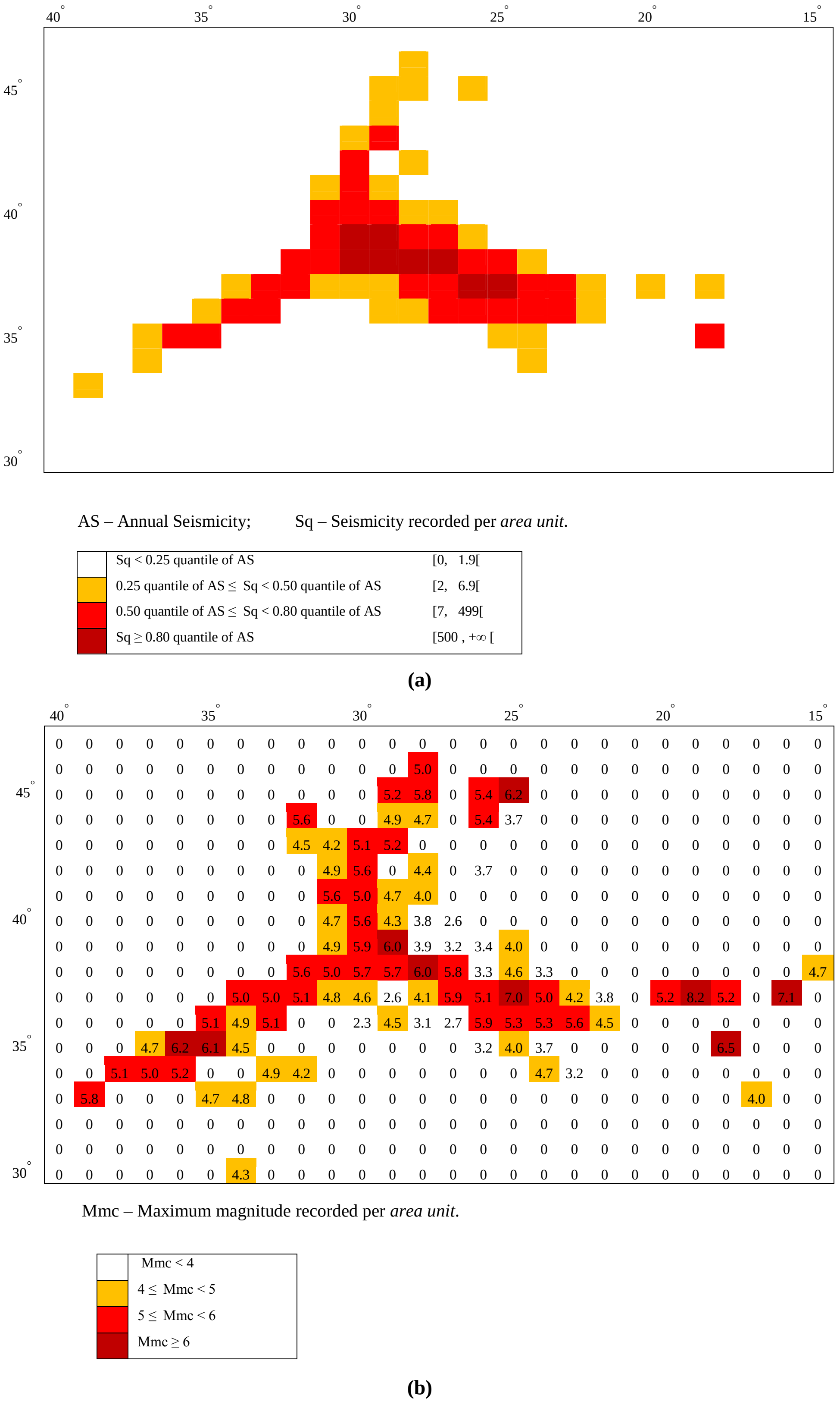
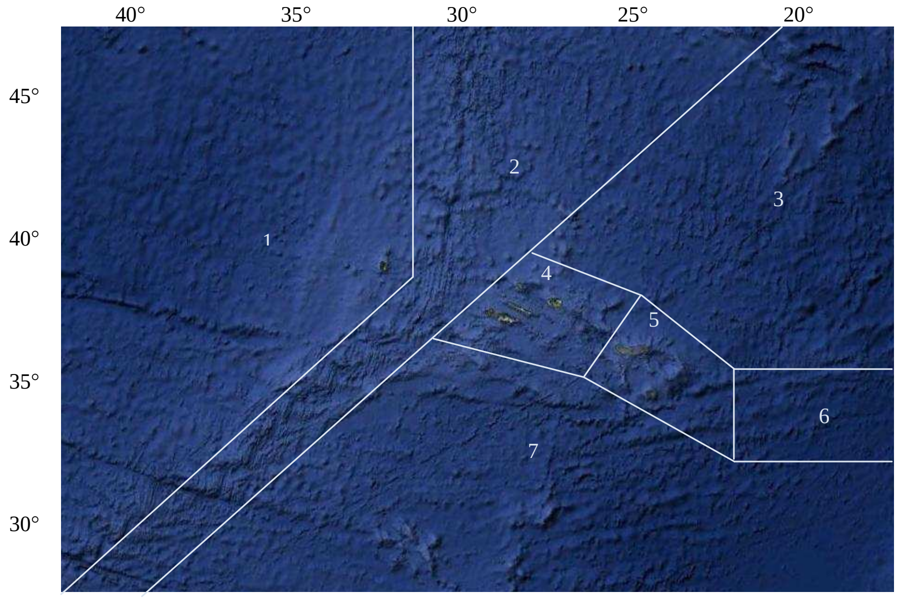
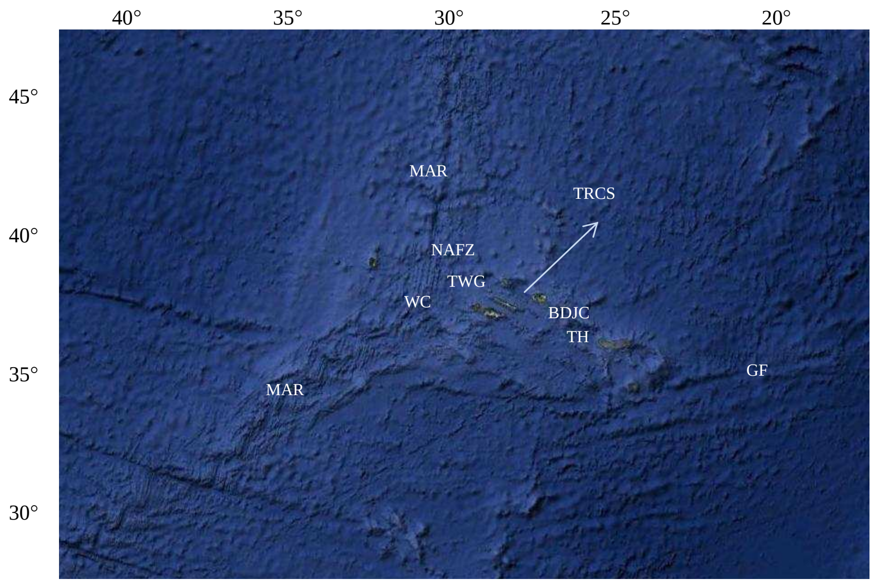


Fig. 8. (a) Seismicity recorded per area unit, and **(b)** maximum magnitude recorded per area unit.



(a)



Mid-Atlantic Ridge (MAR); West of Capelinhos (WC); North Azores Fracture Zone (NAFZ); Bank D. João de Castro (BDJC); Trench Hirondelle (TH); Trench West of Graciosa (TWG); Terceira Rift Central Sector (TRCS); Gloria Fault (GF).

(b)

Fig. 9. (a) Schematic representation of the defined zones, and **(b)** morphological features of the study area.

Table 6. The number of seismic events in each seismic zone.

Zone	1	2	3	4	5	6	7	Total
Obs.	201	1847	65	6009	9948	727	25	18 822

Table 7. Statistics of $Dt_{60,i}$.

	Mean	Standard deviation	Number of records	Minimum	Maximum	95 % confidence interval for the mean
$Dt_{60,1}$	0.2683907	0.5484976	192	0.0000000	3.6578259	[0.1907628, 0.3460186]
$Dt_{60,2}$	0.0281624	0.1090001	1831	0.0000000	1.7924995	[0.0231697, 0.0331551]
$Dt_{60,3}$	0.8320763	1.6549227	60	0.0004793	10.6083954	[0.4133231, 1.2508295]
$Dt_{60,4}$	0.0085003	0.0671134	5981	0.0000000	2.4777918	[0.0067994, 0.0102012]
$Dt_{60,5}$	0.0051966	0.0462983	9921	0.0000000	2.3838838	[0.0042855, 0.0061077]
$Dt_{60,6}$	0.0631802	0.2491379	720	0.0000019	4.9323605	[0.0449820, 0.0813784]
$Dt_{60,7}$	2.2769963	2.1647580	22	0.0566912	8.3945568	[1.3170182, 3.2369744]

Table 8. Statistics of $S_{w0,i}$.

	Mean	Standard deviation	Number of records	Minimum	Maximum	95 % confidence interval for the mean
$S_{w0,1}$	4.5	0.7	174	2.8	6.2	[4.4, 4.6]
$S_{w0,2}$	3.0	0.7	1703	1.4	6.0	[3.0, 3.0]
$S_{w0,3}$	2.9	0.6	38	2.0	4.7	[2.7, 3.1]
$S_{w0,4}$	2.5	0.5	5407	0.2	6.0	[2.5, 2.5]
$S_{w0,5}$	2.4	0.7	7087	0.2	7.0	[2.4, 2.4]
$S_{w0,6}$	2.9	0.7	642	2.0	8.2	[2.8, 3.0]
$S_{w0,7}$	4.1	1.2	14	1.9	6.1	[3.4, 4.8]

5 Methodology for the dissimilation of seismic zones

For the region covered by the data, area units were aggregated by their identical characteristics, resulting in the seven distinct zones.

In the following tests, the aim was to quantitatively show whether the variables corresponding to these areas were significantly different.

If the variables time, size and seismic conditions, which will be explained latter, differ significantly for each defined area, then statistical tests must indicate that these samples come from different populations.

As the seismic zones 1, 3 and 7 are markedly different from other areas based on their reduced seismicity, they are considered to be background zones. It was unnecessary to carry out statistical tests for these zones, and our statistical study focuses on zones 2, 4, 5 and 6.

5.1 Statistical tests

Zones 2, 4, 5 and 6 were first studied together. We used a chi-square test for r independent samples to investigate whether the r populations from which r samples were extracted were the same; that is, we tested the null hypothesis of the variables corresponding to the different zones being taken from the same population.

If the test conclusion was a clear rejection of the null hypothesis, it would not be necessary to use additional tests for r samples, otherwise we must use, for example, the Kruskal–Wallis test (see Siegel and Castellan, 1988).

If a nonparametric test for r samples leads to the rejection of the null hypothesis, the variables cannot come from the same population, but it remains unclear as to whether all come from distinctly different populations. To investigate whether there are samples with the same distribution, we can compare any pair of the r samples using a nonparametric test for pairs of samples.

In this case, we can use the chi-square test for two independent samples or the Kolmogorov–Smirnov two-sample test (e.g., Conover, 1999), with the latter preferable because it is more powerful; see Appendix A2 for a detailed description of these methods.

6 Experiments carried out

6.1 Testing differences in time

The chi-square test for r independent samples was used to verify whether the samples formed by $Dt_{60,j}$, $j \in \{2, 4, 5, 6\}$ can be extracted from the same population.

> Null hypothesis, H0: $Dt_{60,2}$, $Dt_{60,4}$, $Dt_{60,5}$ and $Dt_{60,6}$ have the same distribution.
>
> Alternative hypothesis, H1: $Dt_{60,2}$, $Dt_{60,4}$, $Dt_{60,5}$ and $Dt_{60,6}$ do not have the same distribution.

The data may be grouped into classes. Ten classes bounded by the deciles of Dt_{60} have been adopted (see Table 3).

The meanings of O_{ij}, E_{ij}, C_k and n_r are explained in Appendix A1.

Table 9. Summary of results obtained in the chi-square test of time variable.

Zone		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	n_r
2	O_{ij}	87	224	185	139	97	76	93	86	102	742	1831
	E_{ij}	166.0	173.0	151.0	144.1	133.2	131.3	136.1	140.5	184.9	470.9	
4	O_{ij}	269	436	504	552	526	575	550	548	640	1381	5981
	E_{ij}	542.3	565.3	493.3	470.6	435.0	428.8	444.7	459.0	603.8	1538.3	
5	O_{ij}	1314	1078	821	747	706	652	706	757	1075	2065	9921
	E_{ij}	899.5	937.6	818.3	780.6	721.5	711.3	737.6	761.3	1001.6	2551.6	
6	O_{ij}	3	6	12	14	13	20	23	25	46	558	720
	E_{ij}	65.3	68.0	59.4	56.7	52.4	51.6	53.5	55.2	72.7	185.2	
C_k		1673	1744	1522	1452	1342	1323	1372	1416	1863	4746	18 453

The results obtained in the chi-square test (Table 9) reveal significant differences between the observed and expected values, leading to the rejection of the null hypothesis.

The computation of the test statistic by Eq. (A1) – $T = 441\,301$ with a 0.95 quantile of χ^2_{27} of 40.11 and a 0.99 quantile of 46.96 – indicates that we should reject the null hypothesis. As expected, we can conclude that the samples do not have the same distribution.

Given the large difference between the critical values and the test statistic, it was not necessary to carry out more tests using multiple samples.

The rejection of the null hypothesis only means that the samples do not have the same distribution, but they do not determine whether the samples have distinctly different distributions.

In cases such as this, Siegel and Castellan (1988) recommend investigating whether there are any samples with the same distribution. For this purpose, it is adequate to use the Kolmogorov–Smirnov two-sample test, in which we compare $C_2^4 = 6$ pairs of samples.

Hypothesis

H0: $\mathrm{Dt}_{60,i}$, $\mathrm{Dt}_{60,j}$, $i \in \{2,4,5\}$, $j \in \{4,5,6\}$, $i \neq j$ have the same distribution.

H1: $\mathrm{Dt}_{60,i}$, $\mathrm{Dt}_{60,j}$, $i \in \{2,4,5\}$, $j \in \{4,5,6\}$, $i \neq j$ do not have the same distribution.

Test statistics were computed using Eq. (A3). Table 10 summarizes the obtained results.

In all comparisons, the null hypothesis was rejected; that is, the statistical distributions of the variables $\mathrm{Dt}_{60,i}, i \in \{2,4,5,6\}$ are different.

However, for the comparison of zones 4 and 5, the test statistic is equal to the critical value for a significance level of 1 %. This means that although the empirical distributions of these two populations differ significantly, the difference is smaller than that obtained for other pairs of samples.

To dispel any doubt concerning the possible (but unlikely) similarity between the distributions of $\mathrm{Dt}_{60,4}$ and $\mathrm{Dt}_{60,5}$, a parametric test using the average of these two variables was conducted. The t test for two populations (variances unknown and unequal) (see Kanji, 1993) allows testing if the mean of the two variables may be considered equal. For details on t tests, see Appendix A3.

The test can be applied because the size of the samples is large.

Let μ_4 and μ_5 be the means of the variables $\mathrm{Dt}_{60,4}$ and $\mathrm{Dt}_{60,5}$.

Null hypothesis

H0: $\mu_4 = \mu_{5.}$

The test statistic t has a Student's t distribution with v degrees of freedom. Applying Eqs. (A6) and (A7), one obtains, respectively, $t = 3.356$ and $v = 9436$.

The Student's t variable with n degrees of freedom approaches the standard normal distribution as n approaches infinity. Let $Z_{1-\alpha/2}$ be the 1-$\alpha/2$ quantile of the normal standard distributions: $Z_{0.975} = 1.96$ and $Z_{0.995} = 2.58$.

As t is much greater than the critical value, the null hypothesis can be rejected for the significance levels of 5 % and 1 %.

We conclude that the statistical distributions of $\mathrm{Dt}_{60,i}$, $i \in \{2,4,5,6\}$ differ significantly.

6.2 Testing differences in size

As was performed for $\mathrm{Dt}_{60,j}$, $j \in \{2,4,5,6\}$, the variables $S_{w0,i}$, $i \in \{2,4,5,6\}$ were compared as a whole using the chi-square test for independent samples, and pairs were later compared.

Hypothesis

H0: $S_{w0,2}$, $S_{w0,4}$, $S_{w0,5}$ and $S_{w0,6}$ have the same distribution;

H1: $S_{w0,2}$, $S_{w0,4}$, $S_{w0,5}$ and $S_{w0,6}$ do not have the same distribution.

Data can be grouped into classes. Ten classes bounded by the deciles of S_{w0} were adopted (see Table 4), but classes 1 and 2 were joined because they have few expected values.

Table 10. Summary of results obtained in the Kolmogorov–Smirnov test for time.

Zones compared	2 and 4	2 and 5	2 and 6	4 and 5	4 and 6	5 and 6
n_1	1831	1831	1831	5981	5981	9921
n_2	5981	9921	720	9921	720	720
D	0.155	0.174	0.265	0.027	0.406	0.433
Critical value ($\alpha = 0.05$)	0.036	0.035	0.060	0.022	0.054	0.052
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0
Critical value ($\alpha = 0.01$)	0.044	0.042	0.073	0.027	0.065	0.064
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0

Table 11 summarizes the results obtained for the chi-square test for independent samples.

Computing the test statistic using Eq. (A1), we obtain $T = 1781.1$, with a 0.95 quantile of χ^2_{24} of 36.42 and a 0.99 quantile of 42.98; we reject the null hypothesis.

Therefore, $S_{w0,i}, i \in \{2,4,5,6\}$ do not come from the same population.

To investigate whether the samples arise from the same population, they were compared in pairs using the Kolmogorov–Smirnov two-sample test.

Null hypothesis

H0: $S_{w0,i}$, $S_{w0,j}, i \in \{2,4,5\}, j \in \{4,5,6\}, i \neq j$ have the same distribution;

H1: $S_{w0,i}$, $S_{w0,j}, i \in \{2,4,5\}, j \in \{4,5,6\}, i \neq j$ do not have the same distribution.

Table 12 summarizes the results obtained for the Kolmogorov–Smirnov test.

In all comparisons, the null hypothesis was rejected; that is, the statistical distributions of the variables $S_{w0,i}, i \in \{2,4,5,6\}$ are different, demonstrating that for the size variable, seismic zones differ significantly. In this case, performing additional tests is unnecessary.

6.3 Testing seismic conditions dissimilarity

For each seismic zone, all earthquakes belong to one of four seismic conditions:

1. A recent event (i.e., $\mathrm{Dt}_{60,i} \leq 0.50$ quantile of Dt_{60}) with a magnitude that is not large (i.e., $S_{w0,i} \leq 0.80$ quantile of S_{w0});

2. Not a recent event (i.e., $\mathrm{Dt}_{60,i} > 0.50$ quantile of Dt_{60}) and with a large magnitude (i.e., $S_{w0,i} > 0.80$ quantile of S_{w0});

3. A recent event (i.e., $\mathrm{Dt}_{60,i} \leq 0.50$) with a large magnitude (i.e., $S_{w0,i} > 0.80$ quantile of S_{w0});

4. Not a recent event (i.e., $\mathrm{Dt}_{60,i} > 0.50$ quantile of Dt_{60}) and with a magnitude that is not large (i.e., $S_{w0,i} \leq 0.80$ quantile of S_{w0}) have the correct boundaries.

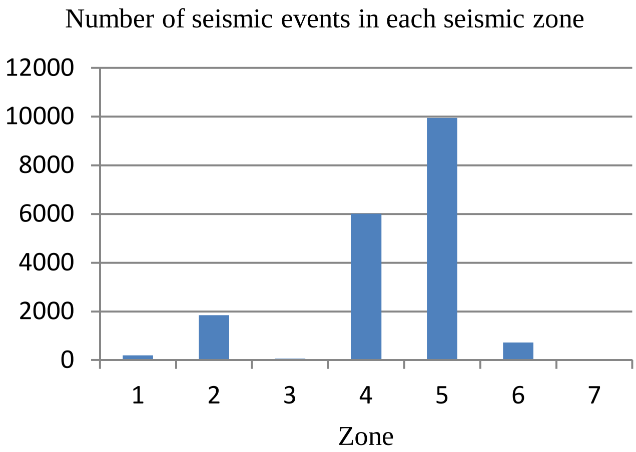


Fig. 10. A plot showing the number of seismic events in each seismic zone.

Let $\mathrm{cd}_i, i \in \{2,4,5,6\}$ represent the seismic condition of each earthquake that occurred in zone i. This variable can assume only values of 1, 2, 3 and 4, corresponding to the four seismic conditions.

Figure 11 summarizes the results obtained for zones 2, 4, 5 and 6.

To verify that the samples formed by $\mathrm{cd}_i, i \in \{2,4,5,6\}$ can be extracted from the same population, a chi-square test for independent samples was used.

Figure 11 strongly implies that the test leads to the rejection of the null hypothesis. Indeed, there is only some similarity in the distribution of cd_i in zones 4 and 5.

Hypothesis

H0: $\mathrm{cd}_i, i \in \{2,4,5,6\}$ have the same distribution;

H1: $\mathrm{cd}_i, i \in \{2,4,5,6\}$ do not have the same distribution.

Table 13 summarizes the results obtained for the chi-square test.

Calculating the test statistic using Eq. (A1), we obtain $T = 1810.4$, with a 0.95 quantile of χ^2_9 of 16.92 and a 0.99 quantile of 21.67. Therefore, we reject the null hypothesis

Table 11. Summary of results applying the chi-square test for size.

Zone		Class 1-2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	n_r
2	O_{ij}	44	44	66	84	180	268	241	348	428	1703
	E_{ij}	304.2	150.7	150.0	143.6	243.5	208.0	160.9	180.1	162.0	
4	O_{ij}	820	633	582	524	785	698	511	507	347	5407
	E_{ij}	966.0	478.4	476.2	455.8	773.2	660.3	510.9	571.7	514.5	
5	O_{ij}	1765	606	626	608	1067	765	565	587	498	7087
	E_{ij}	1266	627	624	597	1013	865	670	749	674	
6	O_{ij}	22	30	33	35	90	81	85	127	139	642
	E_{ij}	114.7	56.8	56.5	54.1	91.8	78.4	60.7	67.9	61.1	
C_k		2651	1313	1307	1251	2122	1822	1402	1569	1412	14 839

Table 12. Summary of results obtained in Kolmogorov–Smirnov test for size.

Zones compared	2 and 4	2 and 5	2 and 6	4 and 5	4 and 6	5 and 6
n_1	1703	1703	1703	5407	5407	7087
n_2	5407	7087	642	7087	642	642
D	0.373	0.414	0.082	0.136	0.297	0.335
Critical value ($\alpha = 0.05$)	0.038	0.037	0.063	0.025	0.057	0.056
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0
Critical value ($\alpha = 0.01$)	0.046	0.045	0.076	0.030	0.069	0.068
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0

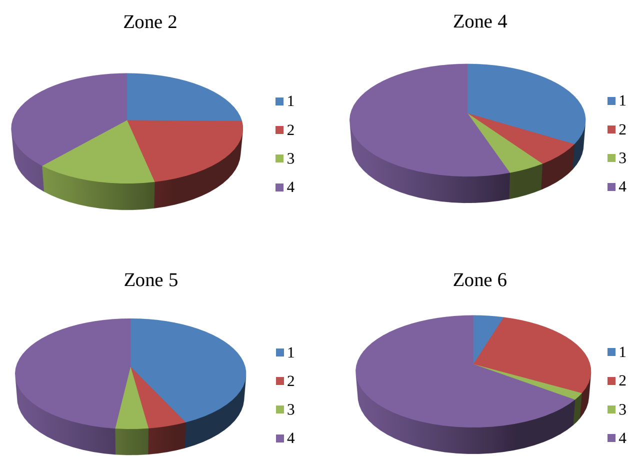


Fig. 11. Graphical representation of $cd_i, i \in \{2, 4, 5, 6\}$.

and conclude that the samples do not have the same distribution.

This means that the distributions of seismic conditions are not the same in zones 2, 4, 5 and 6.

To investigate whether samples of the seismic conditions are from the same population, they were compared in pairs using the Kolmogorov–Smirnov two-sample test.

Hypothesis

H0: cd_i, cd_j, $i \in \{2, 4, 5\}$, $j \in \{4, 5, 6\}$, $i \neq j$ have the same distribution;

H1: cd_i, cd_j, $i \in \{2, 4, 5\}$, $j \in \{4, 5, 6\}$, $i \neq j$ do not have the same distribution.

Table 14 summarizes the results obtained in the Kolmogorov–Smirnov test.

In all comparisons, the null hypothesis was rejected; that is, the statistical distributions of $cd_i, i \in \{2, 4, 5, 6\}$ are different, demonstrating that the seismic conditions of the seismic zones differ significantly.

We also tested the dissimilarity of seismic conditions using a similar procedure that differs only in using the Dt_{60} quantile of 0.80 instead of 0.50. This provided similar results.

7 Conclusions

In this study, we defined seismic zones for the Azores region. We first divided the area into $1^\circ \times 1^\circ$ area units. For each area unit, the seismicity and maximum magnitude recorded were computed.

These two variables were used with the geological characteristics of the region to group area units with similar characteristics; we identified seven seismic zones.

Statistical tests, particularly goodness-of-fit tests, were used, allowing for us to conclude that the variables time, size

Table 13. Summary of results applying the chi-square test for the seismic condition.

Zone		cd_1	cd_2	cd_3	cd_4	n_r
2	O_{ij}	462	393	270	706	1831
	E_{ij}	672.5	145.6	95.3	917.6	
4	O_{ij}	2016	391	274	3300	5981
	E_{ij}	2196.9	475.5	311.2	2997.5	
5	O_{ij}	4266	478	402	4775	9921
	E_{ij}	3644.1	788.7	516.1	4972.1	
6	O_{ij}	34	205	14	467	720
	E_{ij}	264.5	57.2	37.5	360.8	
C_k		6778	1467	970	9248	18 453

Table 14. Summary of results obtained in the Kolmogorov–Smirnov test for the seismic condition.

Zones compared	2 and 4	2 and 5	2 and 6	4 and 5	4 and 6	5 and 6
n_1	1831	1831	1831	5981	5981	9921
n_2	5981	9921	720	9921	720	720
D	0.166	0.178	0.263	0.093	0.290	0.383
Critical value ($\alpha = 0.05$)	0.036	0.035	0.060	0.022	0.054	0.052
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0
Critical value ($\alpha = 0.01$)	0.044	0.041	0.072	0.027	0.064	0.063
Conclusion	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0	Rej. H0

and seismic conditions describing the seven seismic zones differ significantly.

The results of this study will likely be used in future seismic modeling of occurrences in the region.

Appendix A

Statistical tests

A1 Chi-square test for *r* independent samples

The data consist of r independent random samples of sizes $n_1, n_2, \ldots n_r$.

Let $F_1(x)$, $F_2(x)$, ..., $F_r(x)$ represent their respective distribution functions. Each observation can be classified as exactly one of the k categories or classes.

Null hypothesis (H0): $F_1(x) = F_2(x) = \ldots = F_r(x)$.

Let O_{ij} represent the observed number of cells (i, j). The total number of observations is denoted by N. Therefore, $N = n_1 + n_2 + \ldots + n_r$.

Let C_j be the total number of observations in the jth class ($j = 1, 2, \ldots, k$), such that $C_j = O_{1_j} + O_{2_j} + \ldots + O_{rj}$, $j = 1, 2, \ldots, k$.

Table A1. Notation used in the chi-square test for r independent samples.

Sample	Class 1	Class 2	...	Class k	*Totals*
1	O_{11}	O_{12}	...	O_{1k}	n_1
2	O_{21}	O_{22}	...	O_{2k}	n_2
...	...	...	...	...	
r	O_{r1}	O_{r2}	...	O_{rk}	n_r
Totals	C_1	C_2	...	C_k	N

Test statistic

$$T = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}, \tag{A1}$$

$$\text{where } E_{ij} = \frac{n_i C_j}{N}. \tag{A2}$$

The term E_{ij} represents the expected number of observations in cell (i, j) if H0 is true. That is, if H0 is true, the number of observations in cell (i, j) should be close to the ith sample size n_i multiplied by the proportion C_j/N.

It can be shown that the sampling distribution of T is approximately chi-square distributed with

$(k-1).(r-1)$ degrees of freedom, $\chi 2_{(k-1).(r-1)}$.

Let α be the level of significance, i.e., the maximum probability of rejecting a true null hypothesis.

Decision rule

Reject H0 if T exceeds the 1-α quantile of the variable $\chi^2_{(k-1)(r-1)}$; otherwise do not reject H0.

A2 Kolmogorov–Smirnov two-sample test

The Kolmogorov–Smirnov test checks whether two samples were extracted from the same population. The bilateral test is sensitive to any difference in location, dispersion or asymmetry.

The Kolmogorov–Smirnov test aims to assess agreement between two cumulative distribution functions.

The data consist of two independent random samples of sizes n_1 and n_2. Let $F_1(x)$ be the empirical distribution function based on the one random sample $X_1, X_2, \ldots, X_{n1}$, and let $F_2(x)$ be the empirical distribution function based on the other random sample $Y_1, Y_2, \ldots, Y_{n2}$. In order for this test to be precise, the variables must also be continuous.

Hypothesis: (two-sided test)

H0: $F_1(x) = F_2(x)$ for all x from $-\infty$ to $+\infty$;

H1: $F_1(x) \neq F_2(x)$ for at least one value x.

Test statistic: for the two-sided test, the test statistic, D, is

$$D = \sup_x |F_1(x) - F_2(x)|. \quad \text{(A3)}$$

Decision rule: reject H0 at the level of significance α if the test statistic, D, exceeds its 1-α quantile.

For great samples and for $\alpha = 0.05$, the 1-α quantile of D is

$$1.36\sqrt{\frac{n1+n1}{n1.n2}}, \quad \text{(A4)}$$

and for $\alpha = 0.01$, the (1-α) quantile of D is

$$1.63\sqrt{\frac{n1+n1}{n1.n2}}. \quad \text{(A5)}$$

A3 *t* test for two population means (variances unknown and unequal)

Consider two populations with means of $\mu 1$ and $\mu 2$. Independent random samples of size n_1 and n_2 are taken from sets with means $\bar{x}1$ and $\bar{x}2$ and variances $s1^2$ and $s2^2$. The populations may be normally distributed, or the sample sizes may be sufficiently large (see Kanji, 1993).

Null hypothesis: $\mu 1 = \mu 2$.

Test statistic: the variable

$$t = \frac{(\bar{x}1 - \bar{x}2) - (\mu 1 - \mu 2)}{\left[\frac{s1^2}{n1} + \frac{s2^2}{n2}\right]^{\frac{1}{2}}} \quad \text{(A6)}$$

has a Student's t distribution with v degrees of freedom, given by

$$v = \left\{ \frac{\left[\frac{s1^2}{n1} + \frac{s2^2}{n2}\right]^2}{\frac{s1^4}{n1^2(n1+1)} + \frac{s2^4}{n2^2(n2+1)}} \right\} - 2. \quad \text{(A7)}$$

Decision rule: reject H0 at the level of significance α if the absolute value of t exceeds its 1-$\alpha/2$ quantile.

Acknowledgements. The authors would like to acknowledge the comments and recommendations made by two anonymous referees and to Oded Katz, Editor of NHESS, which allowed improvements to the original manuscript.

Edited by: O Katz
Reviewed by: two anonymous referees

References

Bezzeghoud, M., Borges, J. F., Caldeira, B., Buforn, E., and Udias, A.: Seismic Activity in the Azores region in the context of the Western part of the Eurasia-Nubia Plate boundary. Proc. International Seminar on Seismic Risk, Azores, Portugal, paper 7 (CDROM), 1–11, 2008.

Borges, J. F., Bezzeghoud, M., Caldeira, B., and Buforn, E.: Recent seismic activity in the Azores Region. Proc. International Seminar on Seismic Risk, Azores, Portugal, paper 8 (CDROM), 1–5, 2008.

Carvalho, A., Sousa, M. L., Oliveira, C. S., Campos-Costa, A., Nunes, J. C., and Forjaz, V. H.: Seismic hazard for Central Group of the Azores Islands, B. Geofis. Teor. Appl., 42, 89–105, 2001.

Conover, W. J.: Practical nonparametric statistics. John Wiley & Sons, New York, 1999.

Dalgaard, P.: Introductory Statistics with R, Springer, New York, 2008.

Kagan, Y. Y., Bird, P., and Jackson, D. D.: Earthquake Patterns in Diverse Tectonic Zones of the Globe, Pure Appl. Geophys., 167, 721–741, 2010.

Kanji, G. K.: 100 Statistical Tests, London, SAGE Publications, 1993.

IM – Instituto de Meteorologia: available at: http://www.meteo.pt/pt/sismologia/actividade/ (last access: January 2011), 2011.

Nunes J. C., Forjaz, V. H., and Oliveira, C. S.: Zonas de geração sísmica para o estudo da casualidade do Grupo Central do Arquipélago dos Açores, PPERCAS Project, Report no. 3/2000, Azores University, Ponta Delgada, Portugal, 2000.

Nunes, J. C., Forjaz, V. H., and Oliveira, C. S.: Catálogo Sísmico da Região dos Açores, Lisbon, Portugal, (CDROM), 2004.

Reiter, L.: Earthquake Hazard Analysis: Issues and Insights, Columbia University Press, USA, 1991.

Siegel, S. and Castellan, N. J.: Nonparametric statistics for the behavioral sciences, McGraw-Hill, New York, 1988.

Venables, W. N., Smith, D. M., and R Development Core Team: An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics, available at: http://cran.r-project.org/doc/manuals/R-intro.pdf (last access: November 2012), 2011.