Computers and Geosciences 110 (2018) 90–95

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

A density-based clustering algorithm for earthquake zoning

Sanja Scitovski

University of Osijek, Trg Sv. Trojstva 3, HR-31 000 Osijek, Croatia

ARTICLE INFO

Keywords:
Earthquake zoning
Density-based clustering
Big data
Rough-DBSCAN

ABSTRACT

A possibility of applying the density-based clustering algorithm Rough-DBSCAN for earthquake zoning is considered in the paper. By using density-based clustering for earthquake zoning it is possible to recognize nonconvex shapes, what gives much more realistic results. Special attention is thereby paid to the problem of determining the corresponding value of the parameter ε in the algorithm. The size of the parameter ε significantly influences the recognizing number and configuration of earthquake zones. A method for selecting the parameter ε in the case of big data is also proposed. The method is applied to the problem of earthquake data zoning in a wider area of the Republic of Croatia.

1. Introduction

In this paper, we consider a problem of seismogenic zoning in some bounded area (see e.g. Markušić and Herak (1998); Martínez-Álvarez et al. (2015); Morales-Esteban et al. (2010, 2014); Scitovski and Scitovski (2013)). It is well known that seismic moments can be considered as stationary Poisson processes with a fixed occurrence rate over time (Cho et al., 2010), and that devastating earthquakes usually occur without warning and in seconds they can destroy whole cities and severely injure or even kill thousands of inhabitants. Hence it is important to regularly monitor the occurrence of earthquakes and to study their characteristics. The well-known Gutenberg-Richter Law is often used in various studies of seismic activity, e.g. Asencio-Cortés et al. (2017) have studied different seismogenic zones in a wider area of the Republic of Croatia (hereinafter referred to as: Croatia) in terms of earthquake predictability.

Seismic activity in a wider area of Croatia is considered in this paper. Namely, due to its nonconvex geographical shape, in order to analyze seismogenic zones of Croatia, the whole area of Bosnia and Herzegovina and parts of Montenegro, Serbia, Italy and Slovenia should be taken into consideration. Data on seismic activity in a wider area of Croatia can be downloaded free of charge from U.S. Geological Survey http://earthquake.usgs.gov/. These data are of the form: *Year/Month/Day/hh/mm/ss/Latitude (φ)/Longitude (λ)/Depth/Magnitude (M_i)/*. Similarly to Scitovski and Scitovski (2013), based on such data the set

$$\mathcal{A} = \left\{a^i = (\lambda_i, \varphi_i) \in \mathbb{R}^2 : L_\lambda \le \lambda_i \le U_\lambda, \quad L_\varphi \le \varphi_i \le U_\varphi\right\} \tag{1}$$

is defined, which contains earthquake locations determined by longitude λ_i and latitude φ_i. Furthermore, to each point a^i the weight $w_i > 0$ is associates, which is defined as the magnitude M_i of the earthquake in the point a^i. In this case, $\mathcal{A} \subset [12.5, 21] \times [41.5, 47.5]$ (see Fig. 1a), and the data number is $|\mathcal{A}| = 8744$. In the numerical experiments given below, only 5324 data with the magnitude ≥ 3 will be used, among which there are 4051 data of magnitude <4, 1124 data of magnitude between 4 and 5, and 149 data of magnitude ≥ 5 (see Fig. 1b).

If the rectangle $R = [L_\lambda, U_\lambda] \times [L_\varphi, U_\varphi]$ is relatively small (such that relative distances in this rectangle do not significantly differ from relative distances in the corresponding rectangle in the Gauss-Krüger coordinate system), then all evaluations can be carried out directly with the data from the set $\mathcal{A}$. Otherwise it would be necessary to transform the data set in the Gauss-Krüger coordinate system. In the aforementioned case of data from a wider area of Croatia, the rectangle $R = [12.5, 21] \times [41.5, 47.5]$ can be considered as relatively small and no transition to the Gauss-Krüger coordinate system is necessary.

The problem of determining seismogenic zones can be considered as a classical data clustering problem. A partition of the set $\mathcal{A}$ into k disjoint subsets $\pi_1, \dots, \pi_k$, $1 \le k \le |\mathcal{A}|$, such that

$$\bigcup_{i=1}^k \pi_i = \mathcal{A}, \quad \pi_r \cap \pi_s = \emptyset, \quad r \ne s, \quad |\pi_j| \ge 1, \quad j = 1, \dots, k, \tag{2}$$

will be denoted by $\Pi = \{\pi_1, \dots, \pi_k\}$ and the set of all such partitions will be denoted by $\mathcal{P}(\mathcal{A}, k)$. The elements $\pi_1, \dots, \pi_k$ of the partition Π are called *clusters in* $\mathbb{R}^2$.

If $d: \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$, $\mathbb{R}_+ = [0, +\infty\rangle$ is some distance-like function (see e.g. Kogan (2007)), then by introducing the objective function $F: \mathbb{R}^2 \to \mathbb{R}_+$,

E-mail address: sscitov@unios.hr.

http://dx.doi.org/10.1016/j.cageo.2017.08.014
Received 25 May 2017; Received in revised form 24 July 2017; Accepted 20 August 2017
Available online 30 August 2017

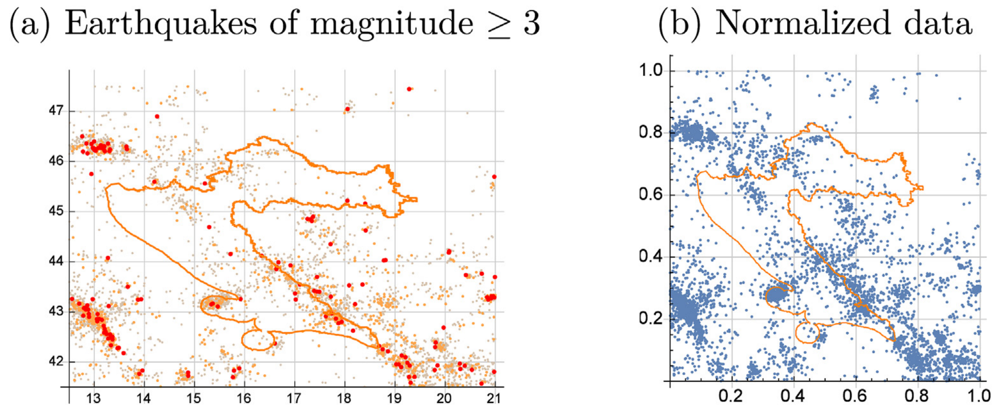

Fig. 1. Earthquake locations in a wider area of Croatia since 1950: magnitude ≥ 5 (red points), magnitude between 4 and 5 (orange points), and magnitude < 4 (brown points), and a normalized data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$F(c_1, \ldots, c_k) = \sum_{i=1}^{m} \min_{1 \le j \le k} d(c_j, a^i), \tag{3}$$

the following global optimization problem can be formulated:

$$\underset{c_j \in conv(\mathscr{A})}{argmin}\ F(c_1, \ldots, c_k), \tag{4}$$

where $conv(\mathscr{A})$ is a convex hull of the set $\mathscr{A}$ (see e.g. Grbić et al. (2013); Morales-Esteban et al. (2014); Scitovski and Scitovski (2013); Scitovski (2017)).

The paper is organized as follows. Some previous results concerning seismogenic zoning which can be found in the literature are mentioned in the next section. In Section 3, first the procedure of normalizing the data is given, and then the density-based algorithm for earthquake zoning is described, with special attention paid to the selection of parameters in the algorithm. Section 4 presents the application of the proposed algorithm to earthquake zoning in a wider area of Croatia and also discusses the case of a big data set.

2. Related works concerning seismogenic zoning

Problem (4) is a complex global optimization problem for the solution of which numerous methods can be found in the literature (see e.g. Bezdek et al. (2005); Kogan (2007); Theodoridis and Koutroumbas (2009); Zaki and W.M. (2014)). The problem of seismic zoning is most commonly solved by using the Least Squares (LS) distance-like function d_{LS} or the ordinary Euclidean distance function d_2 (see Kogan (2007); Scitovski and Scitovski (2013)). In this case, the obtained zones of seismic activity are of spherical form. In Morales-Esteban et al. (2014), a new efficient algorithm using the adoptive Mahalanobis distance-like function was applied to seismic catalogues of Croatia and the Iberian Peninsula. In that case, the obtained zones of seismic activity were of elliptical form. For real earthquake data sets, these forms are too idealized and do not approximate well the current situation in space. In Reyes and Cárdenas (2010), a study of seismic zoning for continental Chile based on a neural network is presented and Parvez (2013) provides significant contributions in the field of seismic zonation and microzonation studies in the Indian subcontinent.

Another important shortcoming of seismic zoning when solving problem (4) by using various distance-like functions is how to determine the most appropriate number of clusters in an optimal partition. Namely, it is well known that if the number of clusters is not given in advance, to recognize the right number of clusters is a complex problem in cluster analysis (see e.g. Kogan (2007); Vendramin et al. (2009)). This problem is solved by using various indexes (see e.g. Morales-Esteban et al. (2014); Scitovski and Scitovski (2013); Vendramin et al. (2009)), but the results are not always reliable enough.

For the foregoing reasons, it makes sense to consider the application of density-based clustering techniques to the problem of seismic zoning. These techniques are based on the idea that objects which form a dense region should be grouped together into one cluster. One of the most widely used density-based clustering algorithms, i.e., the `DBSCAN` (Density-Based Spatial Clustering of Applications with Noise), was originally proposed in Ester et al. (1996).

The `DBSCAN` algorithm has rapidly gained popularity and has been applied in various areas of applications, such as medical images, geology, spam detection, etc. (see Birant and Kut (2007); Jiang et al. (2011); Karami and Johansson (2014)). This algorithm is well described in the papers: Birant and Kut (2007), Mimaroglu and Aksehirli (2011), Zaki and Meira (2014), where corresponding pseudocodes can also be found. In the case of large data sets, there are efficient modifications, such as the `Rough-DBSCAN` algorithm (see Viswanath and Babu (2009)).

It makes sense to apply the `DBSCAN` algorithm if the set $\mathscr{A}$ has a large number of points (in the literature, test examples mostly contain between 3000 and 700000 points in two-dimensional space). The most important advantages of this algorithm are as follows:

- the possibility of recognizing non-convex clusters,
- the partition with the most appropriate number of clusters is obtained automatically, and
- it is not necessary to use indexes for defining an appropriate number of clusters in a partition.

3. Earthquake zoning

Let $\mathscr{A}$ be the set defined in (1) and for each $a^i \in \mathscr{A}$ let the weight $w_i > 0$ be defined as the magnitude of the earthquake in the point a^i. We will try to identify zones of those earthquakes that took place in the reference area.

3.1. Data preparation

As already mentioned, our set $\mathscr{A}$ is contained in the rectangle $R = [L_\lambda, U_\lambda] \times [L_\varphi, U_\varphi]$ with sides of generally unequal length. Therefore, first the set $\mathscr{A}$ will be normalized. This can be done by transforming the set $\mathscr{A}$ into the set $\mathscr{B} = \{T(a^i) : a^i \in \mathscr{A}\} \subset [0,1]^2$ by using the mapping $T : R \to [0,1]^2$, where (see e.g. Scitovski and Šarlija (2014))

$$T(x) = D(x-u), \quad D = \text{diag}\left(\frac{1}{U_\lambda - L_\lambda}, \frac{1}{U_\varphi - L_\varphi}\right), \quad u = (L_\lambda, L_\varphi)^T. \tag{5}$$

After all necessary evaluations on the set $\mathscr{B}$ are done, the obtained results will be transformed again into the rectangle R by using the inverse

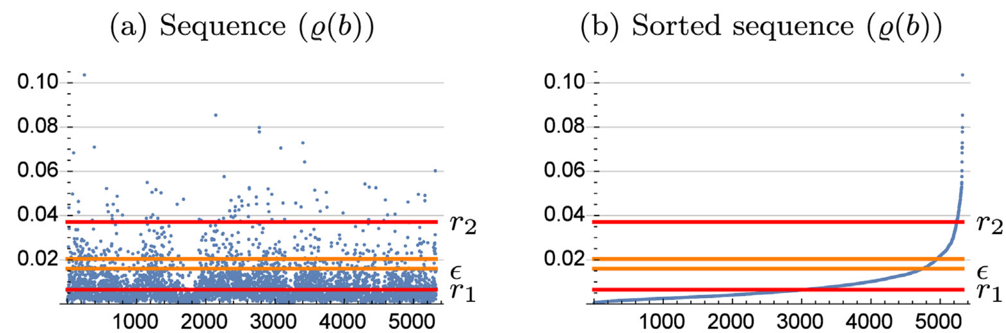

Fig. 2. The choice of the parameter ε. Red lines denote centroids r_1, r_2 of the clusters R_1, R_2, and orange lines denote the length of the right edge of the interval around the centroid containing 95% and 99% of elements of that cluster, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mapping $T^{-1} : [0,1]^2 \to R$, $T^{-1}(x) = D^{-1}x + u$.

For example, the corresponding normalized set $\mathscr{B}$ for data related to seismic activity in a wider area of Croatia is shown in Fig. 1b.

3.2. A density-based algorithm for earthquake zoning

In the set $\mathscr{B}$ constructed in this way we will find disjoint clusters $\pi_1, \ldots, \pi_k$ and the set of noise $\mathscr{N}$ by using the DBSCAN algorithm. In doing so, the Euclidean distance function d_2 will be used.

For the chosen $\varepsilon > 0$, the algorithm is run by selecting the point $b^0 \in \mathscr{B}$ in whose ε-neighborhood

$$\mathscr{N}_\varepsilon\left(B, b^0\right) = \left\{b \in \mathscr{B} : d_2\left(b^0, b\right) < \varepsilon\right\} \tag{6}$$

there are as many points from the set $\mathscr{B}$ as possible. The minimum number of points that could be found in that ε-neighborhood will be denoted by *MinPts*. Many different proposals for the selection of the parameter *MinPts* can be found in the literature. The heuristic suggests that *MinPts* should be approximately $\ln(m)$ (see Birant and Kut (2007)). In order to reduce computational complexity, for the data with two features ($\mathscr{A} \subset \mathbb{R}^2$) *MinPts* is usually fixed to 4 (Jiang et al., 2011). The point $b^0 \in \mathscr{B}$ for which the requirement $|\mathscr{N}_\varepsilon(B, b^0)| \geq$ *MinPts* is met is called a `Core Point`.

The chosen point b^0 will be the first point of the cluster π_1, that will be further supplemented such that it includes (Ester et al., 1996):

- all points $p \in \mathscr{B}$ located within the ε-neighborhood of b^0. These points are said to be directly density-reachable from b^0
- all points $p \in \mathscr{B}$ for which there is a chain of objects $p_\varepsilon, \ldots, p_n \in \mathscr{B}$, p1 = b0 and pn = p such that pi+1 is directly density-reachable from pi with respect to ε and *MinPts*, for $1 \leq i \leq n$. These points are said to be density-reachable.

The points of the cluster π_1 defined in such a way are said to be density-connected, and the cluster itself has the following features (Ester et al., 1996):

(i) (`Maximality`) If $q \in \pi$ and p is density-reachable from q with respect to ε and *MinPts*, then $p \in \pi$;
(ii) (`Connectivity`) Points p and q are density-connected with respect to ε and *MinPts.*

Furthermore, a point $p \in \mathscr{B}$ is a border point if it is not a core point, but density-reachable from another core point, i.e. $1 < |N_\varepsilon(\mathscr{A}, p)| <$ *MinPts*. Note that border points form a *cluster edge*. Noise is a set of points which are neither core points nor border points.

In the next step, we look at the set $\mathscr{B} \backslash \pi_1$ and repeat the previous procedure starting with a new core point. The procedure is repeated as long as there is a possibility of choosing a core point in the rest of the set $\mathscr{B}$ in whose ε-neighborhood there are at least *MinPts* points from $\mathscr{B}$.

3.2.1. Selection of the parameter ε

Several different proposals for the selection of the parameter ε can be found in the literature (see e.g. Andrade et al. (2013); Ester et al. (1996); Jiang et al. (2011); Karami and Johansson (2014)).

In our paper, an exact method for determining the size of the parameter ε will be proposed provided that there is prior knowledge of the parameter *MinPts*. In Section 4.2, this method is modified for big data sets. First, for each $b \in \mathscr{B}$ we determine radius $\varrho(b) > 0$ of the circle

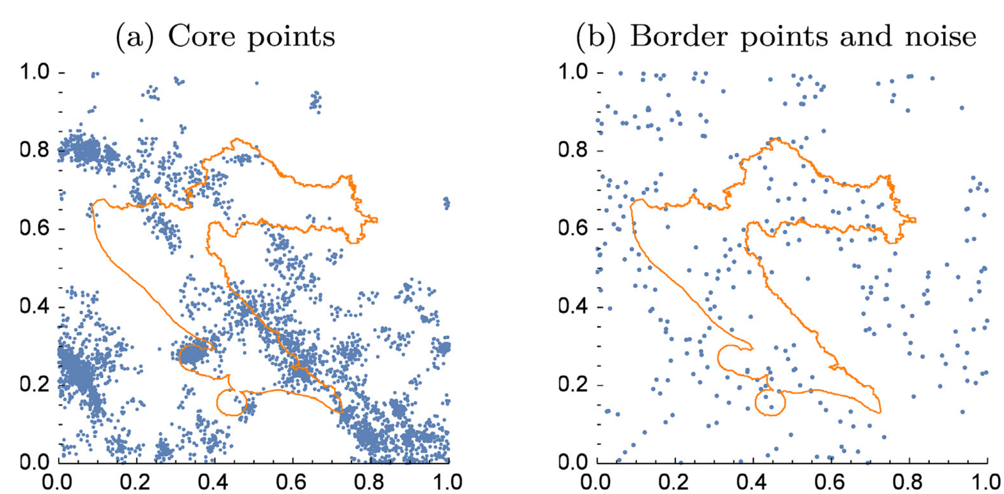

Fig. 3. Separation of the set of core points from the set of border points and noise.

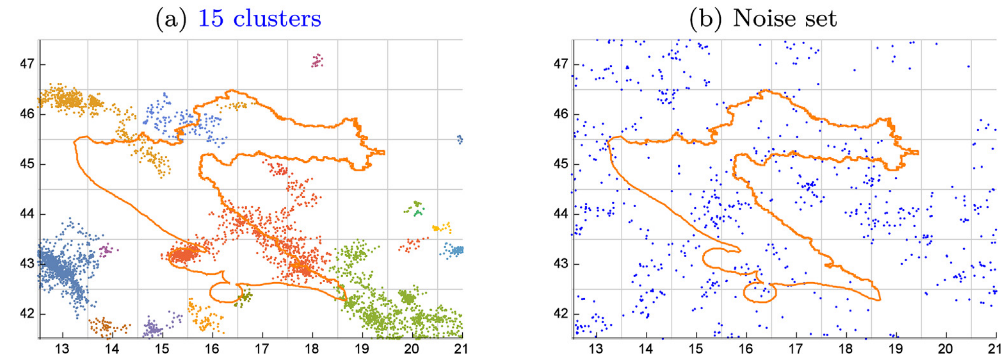

Fig. 4. The results of the `Rough-DBSCAN` algorithm with *MinPts* = 4 and $\varepsilon = 0.016$.

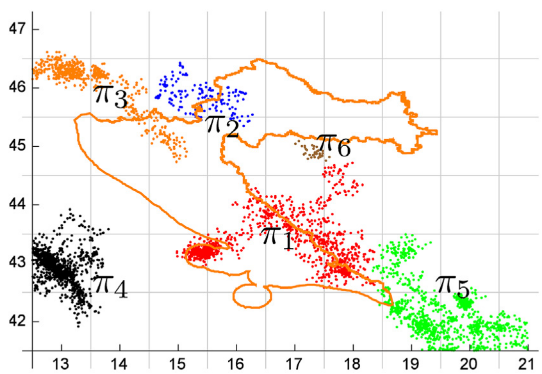

Fig. 5. Significant earthquake zones for the Republic of Croatia.

containing *MinPts* elements of the set $\mathscr{B}$. The set $\Re$ of all radii obtained in such a way will be grouped into two clusters $R_1(r_1)$ and R_2 (r_2), where $r_1 < r_2$ are corresponding centroids by using the LS-distance-like function $d_{LS}\colon \mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}_+$, $d_{LS}(a,b) = \| a-b\|_2^2$. This can be achieved by applying the one-dimensional center-based LS-clustering method using the efficient `SymDIRECT` algorithm proposed in Grbić et al. (2013) or the `SepDIRECT` algorithm proposed in Scitovski (2017). It is important to note that the LS-distance-like function must be used, which successfully separates small values of the radii from large ones (see Sabo et al. (2013); Sabo and Scitovski (2015)). Namely, in this way, most of the values of the radii of circles assigned to core points will be grouped around the smaller centroid r_1, while the radii of circles assigned to border points and noise will be grouped around the larger centroid r_2. If in the larger cluster R_1 the smallest interval around the centroid r_1 containing 95% elements of that cluster is found, then it would make sense to take the value of our parameter ε as the length of the right edge of that interval.

Example 1. *Let us consider the set* $\mathscr{A}$ *of* 5324 *earthquake locations of magnitude* ≥ 3 *in a wider area of Croatia since 1950* (see Fig. 1a) *and the corresponding normalized set* $\mathscr{B}$ (Fig. 1b).

For each $b^i \in \mathscr{B}$, Fig. 2a shows a corresponding radius $\rho(b^i)$ of the circle containing $MinPts = 4$ elements of the set $\mathscr{B}$, whereas Fig. 2b gives a sorted sequence of these radii $(\varrho(b^i))$. Grouping this sequence of radii $(\varrho(b^i))$ into two clusters by using the LS-distance-like function yields the centroid $r_1 = 0.00646$ of the cluster R_1 consisting of 5002 relatively small radii and the centroid $r_2 = 0.03289$ of the cluster R_2 consisting of 322 relatively large radii. The smallest interval around the centroid r_1 containing 95% of elements of the cluster R1 is $\langle -0.003089, 0.0160158\rangle$. Thus in this case we choose $\varepsilon = 0.016$ (see Fig. 2a–b), although a slightly higher value would be acceptable, too.

Fig. 3a shows the points from the set $\mathscr{B}$ associated to the cluster R_1 (core points), and Fig. 3b shows the points from the set $\mathscr{B}$ associated to the cluster R_2 (border points and noise).

3.2.2. *Implementation of the* `DBSCAN` *algorithm*

The `DBSCAN` algorithm has already been described at the beginning of this section. For the purpose of calculation, a corresponding *Mathematica*-program was constructed which uses the `Rough-DBSCAN` algorithm (Viswanath and Babu, 2009).

An obvious advantage of the `DBSCAN` algorithm for earthquake zoning is the possibility of recognizing non-convex clusters. In addition to that, the algorithm is constructed such that it automatically determines the number of disjoint clusters and the set of noise $\mathscr{N}$, too.

Possible disadvantages of the `DBSCAN` algorithm are that it is not easy to determine proper values for ε and *MinPts*. Computational complexity without a special structure is $\mathscr{O}(m^2)$. Therefore, the algorithm can take a large amount of time and because of that, it is not suitable to work with very large data sets. But if a spatial index is used (Viswanath and Babu, 2009), the complexity can be reduced to $\mathscr{O}(m \log m)$. It should also be taken into account when the border points of two clusters are relatively close as it can happen that these clusters are connected.

4. Application of the `Rough-DBSCAN` algorithm to earthquake zoning

Previously in Section 2 and Subsection 3.2.2, we have mentioned the reasons why it would make sense to apply the `DBSCAN` algorithm to earthquake zoning. In the following subsection, this algorithm will be applied to the earthquake data from a wider area of Croatia, and Subsection 4.2 will consider problems that can occur in the case of a significantly larger data set.

4.1. *An example: earthquake zoning in a wider area of Croatia*

The `DBSCAN` algorithm will be tested on the set $\mathscr{A}$ which consists of 5324 earthquake locations of magnitude ≥ 3 in a wider area of Croatia since 1950 (see Fig. 1a). The set $\mathscr{A}$ will first be normalized according to Subsection 3.1, and then the set $\mathscr{B} = T(\mathscr{A})\subset[0,1]^2$ will be considered (see Fig. 1b). Parameters *MinPts* = 4 and $\varepsilon = 0.016$ are chosen as in Example 1.

By implementing the `DBSCAN` algorithm on the set $\mathscr{B}$ with $MinPts = 4$ and $\varepsilon = 0.016$, 15 clusters with 1081, 999, 1081, 541, 156, 69, 41, 57, 87, 40, 19, 80, 31, 24 and 12 elements will be obtained, respectively, and they are shown by different colors in Fig. 4a. The remaining 1006 points represent the noise set (see Fig. 4b).

Significant earthquake zones for the Republic of Croatia are the clusters shown in Fig. 5:

- π_1: the red cluster with 1081 points (South Croatia);
- π_2: the blue cluster with 156 points (the vicinity of Zagreb);
- π_3: the orange cluster with 541 points (North Italy and Istria);
- π_4: the black cluster with 999 points (Central Italy);
- π_5: the green cluster with 1081 points (Montenegro and North Albania);

Table 1
Selection of the parameter ε from the sample.

	$\lvert R_1\rvert$	r_1	$\lvert R_2\rvert$	r_2	ε(95%)	ε(99%)	CPU(1) (sec)	CPU(2) (sec)
Set $\mathscr{A}$	5002	0.0065	322	0.0329	0.016	0.020	151.94	9.42
$T_1 = (0.7, 0.2)$	888	0.0057	71	0.0200	0.010	0.012	3.70	3.73
$T_2 = (0.6, 0.3)$	659	0.0064	135	0.0172	0.011	0.040	2.42	3.42
$T_3 = (0.35, 0.65)$	305	0.0094	43	0.0268	0.0166	0.0175	0.42	1.81
$T_4 = (0.25, 0.75)$	376	0.0085	34	0.0292	0.016	0.018	0.58	1.95

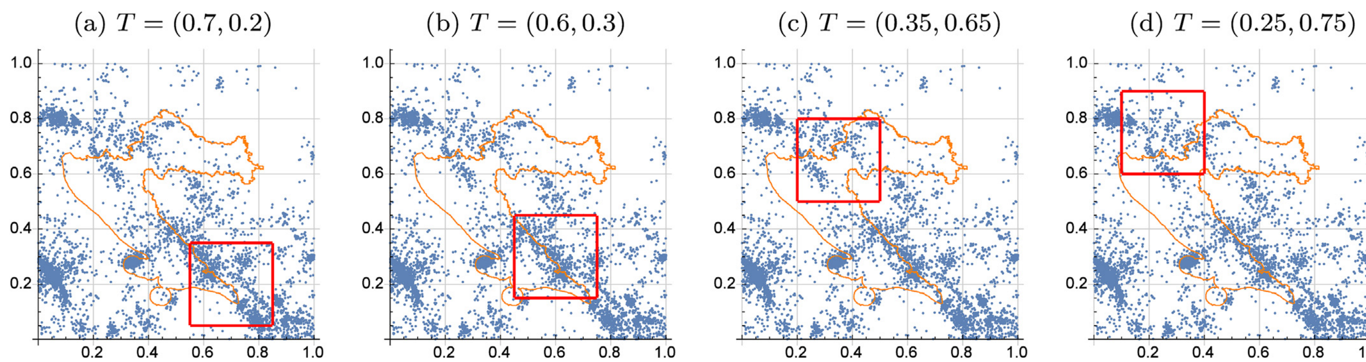

Fig. 6. Selection of different windows in the square $[0, 1]^2$.

π_6: the brown cluster with 40 points (North Bosnia).

In Markušić and Herak (1998), this area is divided into 17 seismic zones by using data from the Croatian earthquake catalogue (see Herak et al. (1996)). Significant similarity with our results can be mentioned.

4.2. *Application of the* DBSCAN *algorithm to a very large earthquake data set*

If we would like to zone a normalized set $\mathscr{B}$ which has significantly more data, the first problem might occur when determining an appropriate parameter ε. Namely, it is estimated that computational complexity of the procedure described in Subsection 3.2.1 is $\mathscr{O}(m^2)$ and hence large CPU-time can be required for the described procedure of searching for the parameter ε.

In order to shorten that time, the following can be done. Within the square $[0, 1]^2$ containing the set $\mathscr{B}$, let us choose a window with approximately 10% of the area of the square $[0, 1]^2$. This can be done by selecting the point $T \in [0,1]^2$ and the number $\rho = 0.15$ and defining the window

$$W_T = \left\{b^i \in \mathscr{B} : \| T - b^i\|_\infty \le \rho\right\}. \tag{7}$$

If the procedure described in Subsection 3.2.1 is applied to the window W_T, the approximation ε_T of the parameter ε will be obtained in a significantly shorter period of time (see CPU1-time in Table 1). The procedure can be repeated several times.

The described procedure will be illustrated on the set $\mathscr{B}$ from Example 1. This will enable us to compare the obtained results with the ε-value 0.016 obtained in Subsection 3.2.1.

An experiment will be conducted for estimating the parameter ε with four points: $T_1 = (0.7, 0.2)$ (Fig. 6a), $T_2 = (0.6, 0.3)$ (Fig. 6b), $T_3 = (0.35, 0.65)$ (Fig. 6c), $T_4 = (0.25, 0.75)$ (Fig. 6d) and the result will be compared with the parameter ε obtained based upon all points of the set $\mathscr{B}$.

The results are shown in Table 1. For each point T_i, the number of elements $|R_1|$ is given as well as the centroid r_1 of a larger cluster, the number of elements $|R_2|$ and the centroid r_2 of the smaller cluster, the length of the right edge of the smallest interval around the centroid r_1 containing 95% and 99% elements of the cluster R_1, respectively, and CPU(1)-time necessary for calculating the set of all radii $\mathfrak{R}$ and CPU(2)-time necessary for the implementation of the SymDIRECT algorithm on the set $\mathfrak{R}$. It can be seen that the obtained approximations of the parameter ε are acceptable if window (7) is selected representatively. Also, the necessary CPU-time is significantly shorter in relation to the CPU-time for the whole set $\mathscr{A}$.

5. Conclusions

The advantages of the density-based clustering algorithm DBSCAN for earthquake zoning can be seen in the possibility of recognizing non-convex clusters and in the fact that the partition with the most appropriate number of clusters is obtained automatically without using indexes. High computational complexity and rather long implementation in the case of big data sets can be listed as shortcomings of this method. This shortcoming is significantly reduced by using the Rough-DBSCAN modification. The parameter ε estimation method proved to be good and to considerably accelerate the calculation process.

Acknowledgements

The author would like to thank the referees and the journal editors for their careful reading of the paper and insightful comments that helped us to improve the paper.

References

Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., Rocha, L., 2013. G-DBSCAN: a GPU accelerated algorithm for density-based clustering. Procedia Comput. Sci. 18, 369–378.
Asencio-Cortés, G., Scitovski, S., Scitovski, R., Martinez-Álvarez, F., 2017. Temporal analysis of Croatian seismogenic zones to improve earthquake magnitude prediction. Earth Sci. Inf. http://dx.doi.org/10.1007/s12145-017-0295-5.
Bezdek, J.C., Keller, J., Krisnapuram, R., Pal, N.R., 2005. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Springer, NewYork.
Birant, D., Kut, A., 2007. ST-DBSCAN: an algorithm for clustering spatial-temporal data. Data & Knowl. Eng. 60, 208–221.
Cho, N.F., Tiampo, K.F., Mckinnon, S.D., Vallejos, J.A., Klein, W., Dominguez, R., 2010. A simple metric to quantify seismicity clustering. Nonlin. Process. Geophys. 17, 293–302.
Ester, M., Krieogel, H., Sander, J., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, pp. 226–231.
Grbić, R., Nyarko, E.K., Scitovski, R., 2013. A modification of the DIRECT method for Lipschitz global optimization for a symmetric function. J. Glob. Optim. 57, 1193–1212.

Herak, M., Herak, D., Markušić, 1996. Revision of the earthquake catalogue and seismicity of Croatia, 1908–1992. Terra Nova 8, 86–94.

Jiang, H., Li, J., Yi, S., Wang, X., Hu, X., 2011. A new hybrid method based on partitioning-based DBSCAN and ant clustering. Expert Syst. Appl. 38, 9373–9381.

Karami, A., Johansson, R., 2014. Choosing DBSCAN parameters automatically using differential evolution. Int. J. Comput. Appl. 91, 1–11.

Kogan, J., 2007. Introduction to Clustering Large and High-dimensional Data. Cambridge University Press, New York.

Markušić, S., Herak, M., 1998. Seismic zoning of Croatia. Nat. Hazards 18, 269–285.

Martínez-Álvarez, F., Morales-Esteban, A., Reyes, J., Gutiérrez-Avilés, D., Escudero, C.R., 2015. A novel method for seismogenic zoning based on triclustering: application to the Iberian Peninsula. Entropy 17, 5000–5021.

Mimaroglu, S., Aksehirli, E., 2011. Improving DBSCAN's execution time by using a pruning technique on bit vectors. Pattern Recognit. Lett. 32, 1572–1580.

Morales-Esteban, A., Martínez-Álvarez, F., Scitovski, S., Scitovski, R., 2014. A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. Comput. Geosci. 73, 132–141.

Morales-Esteban, A., Martínez-Álvarez, F., Troncoso, A., Justo, J., Rubio-Escudero, C., 2010. Pattern recognition to forecast seismic time series. Expert Syst. Appl. 37, 8333–8342.

Parvez, I.A., 2013. New approaches for seismic hazard studies in the indian subcontinent. Nat. Hazards Risk 4, 299–319.

Reyes, J., Cárdenas, V.H., 2010. A Chilean seismic regionalization through a Kohonen neural network. Neural Comput. Appl. 19, 1081–1087.

Sabo, K., Scitovski, R., 2015. An approach to cluster separability in a partition. Inf. Sci. 305, 208–218.

Sabo, K., Scitovski, R., Vazler, I., 2013. One-dimensional center-based l_1-clustering method. Optim. Lett. 7, 5–22.

Scitovski, R., 2017. A new global optimization method for a symmetric Lipschitz continuous function and application to searching for a globally optimal partition of a one-dimensional set. J. Glob. Optim. 68, 713–727. http://dx.doi.org/10.1007/s10898-017-0510-4.

Scitovski, R., Scitovski, S., 2013. A fast partitioning algorithm and its application to earthquake investigation. Comput. Geosci. 59, 124–131.

Scitovski, S., Šarlija, N., 2014. Cluster analysis in retail segmentation for credit scoring. Croat. Oper. Res. Rev. 5, 235–245.

Theodoridis, S., Koutroumbas, K., 2009. Pattern Recognition, fourth ed. Academic Press, Burlington.

Vendramin, L., Campello, R.J.G.B., Hruschka, E.R., 2009. On the comparison of relative clustering validity criteria. In: Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 – May 2, 2009. SIAM, Sparks, Nevada, USA, pp. 733–744.

Viswanath, P., Babu, V.S., 2009. Rough-DBSCAN: a fast hybrid density based clustering method for large data sets. Pattern Recognit. Lett. 30, 1477–1488.

Zaki, M.J., Meira Jr., W., 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, New York.