Expert Systems with Applications 40 (2013) 4183–4189

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

“Seismic-mass” density-based algorithm for spatio-temporal clustering

G. Georgoulas [a], A. Konstantaras [b,*], E. Katsifarakis [b], C.D. Stylios [a], E. Maravelakis [b], G.J. Vachtsevanos [c]

[a] Technological Educational Institute of Epirus, Arta 47100, Greece
[b] Technological Educational Institute of Crete, Romanou 3, Chania 73133, Greece
[c] Georgia Institute of Technology, North Ave., Atlanta 30332, USA

ARTICLE INFO

Keywords:
Regional spatio-temporal clustering
Seismic-mass
Density based algorithms
Agglomerative algorithms
Distinct seismic regions classification

ABSTRACT

In this research work a new hybrid approach to spatio-temporal seismic clustering is proposed. The method builds upon a novel density based clustering scheme that explicitly takes into account earthquake's magnitude during the density estimation. The new density based clustering algorithm considers both time and spatial information during cluster formation. Therefore clusters lie in a spatio-temporal space. A hierarchical agglomerative clustering algorithm acts upon the identified clusters after dropping the time information in order to come up only with the spatial description of seismic events. The approach is demonstrated using data from the vicinity of the Hellenic seismic arc in order to enable its comparison with some of the state-of-the-art distinct seismic region identification methodologies. The presented results indicate that the combination of the two clustering stages could be potentially used for an automatic definition of major seismic sources.

1. Introduction

Modeling the behavior of the earthquake phenomenon remains an open front and top challenge in geosciences. More importantly, understanding the underlying physics of the earthquake phenomenon should be the first priority before trying to build any mathematical model. What makes this task even more difficult is that the only data readily presented to mankind is the output of the phenomenon, in other words only the seismic events.

A quick examination of geographical maps with the epicenters of earthquakes marked on them reveals a strong tendency of these points to form compact clusters of irregular shapes and various sizes often traversing with other clusters. Seismic cluster formation is believed to be due to underlying geological natural hazards, which: (a) act as the energy storage elements of the phenomenon and (b) tend to form a complex network of numerous interacting faults (Vallianatos & Tzanis, 1998). “Earthquakes are correlated in space and time over large distances” (Saleur, Sammis, & Sornette, 1996). This implies that seismic sequences are not formatted randomly but they follow a spatial pattern with consequent triggering of events. In other words geological natural hazards rarely appear on their own, instead they tend to form a complex network of numerous interacting faults. Even though the physical/geological mechanisms that account for the formation of this spatio-temporal phenomenon is literally unobservable, its structure can be indirectly monitored based on the seismic activity of “neighboring” areas or using geological and geophysical characteristics of the region (Zamani & Hashemi, 2004).

One of the first steps of seismic hazard analysis is to identify seismogenic zones which can be further divided into smaller subzones (seismic sources) based on various seismological criteria (Morrato et al., 2007). Seismic zoning is usually performed based on expert knowledge (Papaioannou & Papazachos, 2000). However expert knowledge can sometimes be subjective (Zamani & Hashemi, 2004) and as that lately quantitative methods and more specifically, clustering applications (Jain, 2010) have emerged as an alternative for seismotectonic zone delineation with the hope to provide a more objective approach.

Even though clustering in seismology usually refers to temporal association of seismic events creating a temporal sequence (Console, Murru, & Catalli, 2006; Pondard, Armijo, King, Meyer, & Flerit 2007; Zobin 1996) attempting to explore/explain the triggering mechanism and the causality effect of seismic events within the context of seismic zoning we are also concerned with the spatial clustering of events and their spatial origin. Various clustering methods have been employed with different inputs but with the same more or less goal: to identify “uniform” seismic areas that can be used for further hazard analysis.

Algorithms have been developed for delineation of seismic source zones using a modified *K*-means algorithm that takes into account the spatial orientation of earthquake hypocentres (Burton, Weatherill, Karnawati, & Pramumijoyo, 2008; Weatherill & Burton, 2009, 2010) with applications to the Aegean region and in Java

* Corresponding author. Address: Laboratory of Computational Intelligence and Software Engineering, Department of Electronics, Technological Educational Institute of Crete, Romanou 3, Chania 73133, Greece. Tel.: +30 28210 23033.
E-mail address: akonstantaras@chania.teicrete.gr (A. Konstantaras).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.01.028

island. A series of papers report the use of a geographical grid superimposed on Iran and based on geophysical, geological and other information that merge sites of the grid which share similar characteristics. More specifically in Zamani and Hashemi (2004) a hierarchical clustering algorithm, the Ward's method (Theodoridis & Koutroumbas, 2008), was employed for the identification of similar sites of the grid using 12 geological and 13 geophysical parameters. In Zamani, Nedaei, and Boostani (2009) a self-organized map was trained using an augmented set of 49 attributes (14 geological, 6 seismicity related and 29 attributes of geophysical nature) with the same grid set-up as in Zamani and Hashemi (2004) while the same approach was tested in Zamani, Khalili, and Gerami (2011) using statistical evaluation methods for the selection of the optimum number of clusters. In Ansari, Noorzad, and Zafarani (2009) a fuzzy clustering algorithm capable of identifying elliptical clusters along with the epicenter locations of seismic events was used to cluster a seismic catalog of Iran. (Konstantaras, Vallianatos, Varley, & Makris, 2008; Konstantaras, Varley, et al., 2007) applied an agglomerative algorithm that exploits both the spatial and temporal dimension of seismic events to cluster seismic data in the Hellenic arc. In Mukhopadhyay, Fnais, Mukhopadhyay, and Dasgupta (2010) a density based clustering method is employed to seismic data coming from the Burmese–Andaman Arc System (BAAS) and the West Sunda Arc (WSA) whereas the same approach is employed in Mukhopadhyay, Mukhopadhyay, and Dasgupta (2011) for the case of Arabian Sea Triple Junction. In Hernandez and Sallis (2011) a mixture modeling cluster approach is applied to seismic data coming from Chile resulting in soft clustering of the seismic catalog. A similar approach was followed by Kayabol (Kayabol, 2012) where a constrained Finite Mixture Model used the epicenters of seismic events from the Kashmir area in Pakistan as inputs to the clustering paradigm.

The identification of valid seismic clusters can aid the development of seismic zone formation that can be subsequently used in probabilistic seismic hazard analysis. Temporal as well as spatial information can be fused looking for consistent patterns emerging from interconnected underlying faults triggering consecutive main seismic events in a chain effect fashion. Along this path algorithms must be developed based on historical data trying to: (a) identify clusters that are forming elongated and sometimes irregularly shaped patterns, and (b) identify seismic events that belong to the same seismic chain, grouping together pre- and post-seismic events to individual main earthquakes.

In this research work we propose a two stage clustering method that is based on a modified density based clustering algorithm and a hierarchical agglomerative scheme. The proposed method is applied to data coming from the region of the Hellenic seismic arc with results indicating that the framework could potentially be used along with probabilistic paradigms for the development of a spatio-temporal model that could cover the specific area.

The rest of the paper is structured as follows: Section 2 describes the proposed algorithm. Section 3 presents the results of our method and Section 5 concludes the paper also providing directions for future research.

2. The hybrid clustering scheme

In a simple marking on an atlas of the earthquake epicenters, it can be seen even by the naked eye that there is a spatial structure with sometimes irregular, elongated shapes. In Fig. 1 this is shown for the case of the Hellenic seismic arc, where only events with $M_S > 4$ are depicted. As it can be seen there are a number of points aligned along well known geological faults (even though there are some that are "off the mark"). This means that clustering algorithms such as the well-known k-means that tend to form

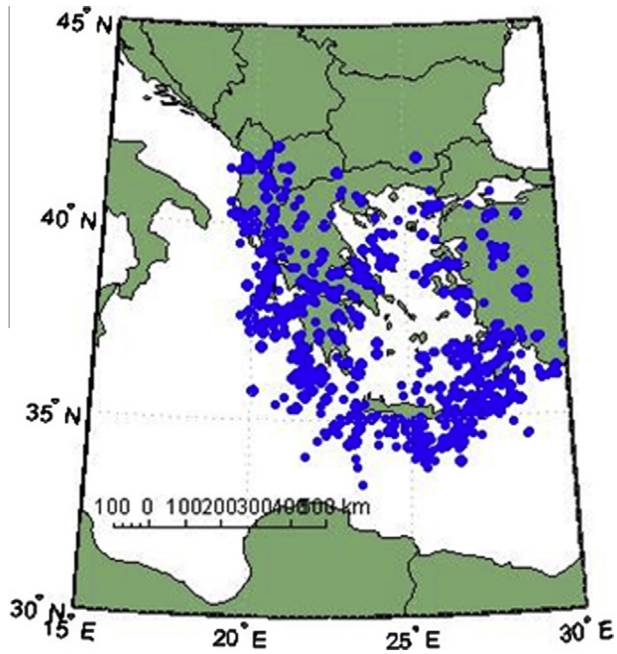

Fig. 1. Epicenters of earthquakes with $M > 4$ for the period 2000–2010.

"well-shaped" clusters may not suffice for the problem at hand and other families of unsupervised pattern recognition methods might be a better choice (Jain, 2010). Therefore algorithms that are not affected by the shape of the clusters and the number of instances belonging to each cluster, such as density based algorithms could be used.

Clustering is a method to uncover inherent grouping in a set of observations. Based on different underlying principals and assumptions, a number of clustering algorithms have been proposed over the years with some of them aiming to improve existing algorithms and some of them rising from the need to meet the requirements of new data sources, like those coming from social networks etc. (Jain, 2010). In the case of seismic data, even the way that we often depict them (using markers that are proportional to the magnitude of the event) makes it obvious that the size of an event matters and should be taken into account during clustering.

In this work we propose a modified version of the well know DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm (Ester, Kriegel, Sander, & Xu, 1996) which we call it Seismic Mass DBSCAN (SM-DBSCAN) that takes into account the peculiarity of seismic data that was mentioned in the previous paragraph: each seismic event (earthquake) is characterized apart from its spatial and time dimensions, also by its magnitude. In other words the "importance" of two events taking place at the same location depends on their respective magnitude. Therefore treating each event "equally" might not be the best way when performing clustering. Along this line of thought, (Weatherill & Burton, 2009, 2010) suggest that "the enhanced influence of strong earthquakes" in the cluster analysis should improve the stability of the algorithm" but in their work the proposed weighing scheme does not depend on the seismic size but on the underlying fault length.

The SM-DBSCAN operates both in space and time thus creating clusters that may overlap. In order to merge these closely related clusters which are usually separated in time we apply an agglomerative hierarchical scheme after dropping the space dimension and focusing solely on the space dimension since in this work we are mainly interested in the development of a model for spatial source identification. Both stages of the method are described in the following subsections.

2.1. The SM-DBSCAN algorithm

The traditional/conventional DBSCAN algorithm clusters datasets' events based on the "density" of data occurrence (Ester et al., 1996). More specifically it estimates the density of data using a Parzen window like approach, i.e. uses a predefined neighborhood radius to estimate the density, in order to discover regions which contain a "significant" number of data. The neighborhood parameter (*Eps*) as well the minimum number of data samples (MinPts) must be provided by the user and are application specific. As it can be easily seen each data point is treated equally within the DBSCAN framework. In the modified version of the DBSCAN we propose to weight each seismic event by its magnitude. We propose in other words a notion of "seismic mass". Therefore a region is "dense" provided that the accumulated "seismic masses" exceed a predefined threshold *M*, which replaces the MinPts parameter in the traditional/conventional DBSCAN. The concept of the inclusion of the size of the seismic event in the calculation of the density is depicted for a hypothetical one dimensional case; Fig. 2a, depicts the estimated density by simply counting the occurrence of data points along the *x*-axis whilst Fig. 2b depicts the estimated density also considering the fact that the central points correspond to events with much higher "mass".

Within this context a *core* point is one that has within its neighbor a number of points that have a cumulative seismic mass exceeding *M*. If a point is not a core point then it is either a *boarder* point or a *noise* point. The discrimination between core points and noise points is based on the notion of density reachability (Ester et al. 1996). The proposed SM-DBSCAN algorithm operates in a similar manner to the DBSCAN algorithm as shown below, using the notation proposed by Theodoridis and Koutroumbas (2008):

Let X be the set of seismic events to be clustered and X_{un} be the set of points belonging to X that have not yet been considered and m denote the number of clusters.

SM-DBSCAN Algorithm

- ●Set $X_{un} = X$
- ●Set $m = 0$
- ●While $X_{un} \neq \varnothing$ do
  - ○Arbitrarily select a $\boldsymbol{x} \in X_{un}$
  - ○If $\boldsymbol{x}$ is a noncore point (i.e. the cumulative masses of the points within *Eps*) then
    - ■Mark $\boldsymbol{x}$ as a noise point
    - ■$X_{un} = X_{un} - \{\boldsymbol{x}\}$
  - ○If $\mathbf{x}$ is a core point then
    - ■$m = m + 1$
    - ■Determine all density-reachable point in X from $\boldsymbol{x}$
    - ■Assign $\boldsymbol{x}$ and the previous points to the cluster C_m. The boarder points that may have been marked as noise as also assigned to C_m.
    - ■$X_{un} = X_{un} - C_m$
  - ○End {if}
- ●End {while}

We must note that the main difference between the SM-DBSCAN and the DBSCAN lies in the calculation of the density. Therefore SM-DBSCAN inherits both the advantages and the disadvantages of the conventional DBSCAN. As a result one of the potential advantages of SM-DBSCAN is that as in the case of DBSCAN the number of clusters has not to be set beforehand and moreover the algorithm does not impose any restriction to the shape of resulting clusters. The latter can be quite useful as it will be shown in Section 2.2. On the other hand there are two parameters that have to be set by the user that heavily influence the result of the clustering procedure. The effect of them as well as the effect of scaling of the original data is presented in Section 3.

2.2. Second stage agglomerative hierarchical clustering

SB-DBSCAN uses as input both the spatial information of the epicenters (latitude and longitude) as well as a third dimension, the chronological occurrence of the event. With the inclusion of time we have the formation of clusters in a three dimensional space and on one hand we can have clusters of events grouped together representing seismic sequences corresponding to the same main earthquake while on the other hand we can have clusters that if projected on the space plane will overlap.

Therefore the inclusion of time as an extra dimension can resolve the issue of closely located faults by enabling clusters to overlap spatially. They are in other words clusters that come from the same fault but at different time instances. This is illustrated in Fig. 3 using artificial data where the three presented clusters overlap spatially.

In order to come up with a seismic zoning procedure these timely-separated clusters (with usually irregular shapes) could be combined to define the common seismic zone. Since this is by itself a hierarchical bottom-up approach, a hierarchical single linkage agglomerative procedure (Jain, 2010, Minskin 2011, Theodoridis & Koutroumbas, 2008) was imported following the application of the SM-DBSCAN hereby dropping the time information related to the identified clusters and utilizing only the spatial related dimensions. Due to the fact that we need to group together overlapping clusters, single linkage is most likely to make clusters to merge. This is due to the way the proximity between clusters is calculated in the single linkage case and more specifically in the case that proximity is a dissimilarity measure (Theodoridis & Koutroumbas, 2008):

If D_i, D_j are two sets of vectors, then the min (single linkage) proximity function is given:

$$P^{ss}_{\min}(D_i, D_j) = \min_{\mathbf{x}\in D_i, \mathbf{y}\in D_j} p(\mathbf{x}, \mathbf{y})$$

where $p(\boldsymbol{x}, \boldsymbol{y})$ is the proximity measure between two data points.

In other words this is defined as the proximity between the closest two points that are in different clusters. The reason that single linkage approach "guarantees" that overlapping clusters will merge before we start merging spatially separated clusters comes also as a by-product of the way that SM-DBSCAN creates clusters at the first place: SM-DBSCAN identifies "dense" clusters and it is therefore highly likely that within these dense regions the calculated proximity measure (distance in our case) will be very small (close to 0).

The whole approach is summarized as follows:

Step 1: Apply the SM-DBSCAN algorithm to the seismic catalog.
Step 2: Drop the time information.
Step 3: Calculate the proximity between the clusters identified in step 1 based on their spatial coordinates.
Step 4: Merge the closest clusters.
Step 5: Update the proximities.
Step 6: Go to step 4 unless a predefined number of cluster remains.

3. Results

The region of the Hellenic seismic arc is the most seismological active part of Europe as it can be seen in Fig. 4 where all seismic events with a magnitude greater than 1 and for the time period 2000–2010 are marked. This figure makes apparent that the inclusion of for- and after-shocks can obscure the underlying structure. Failing to identify the earthquakes that are actually the main ones within small chains of for- and after-shocks and resorting to use the entire volume of archived data can erroneously lead to the

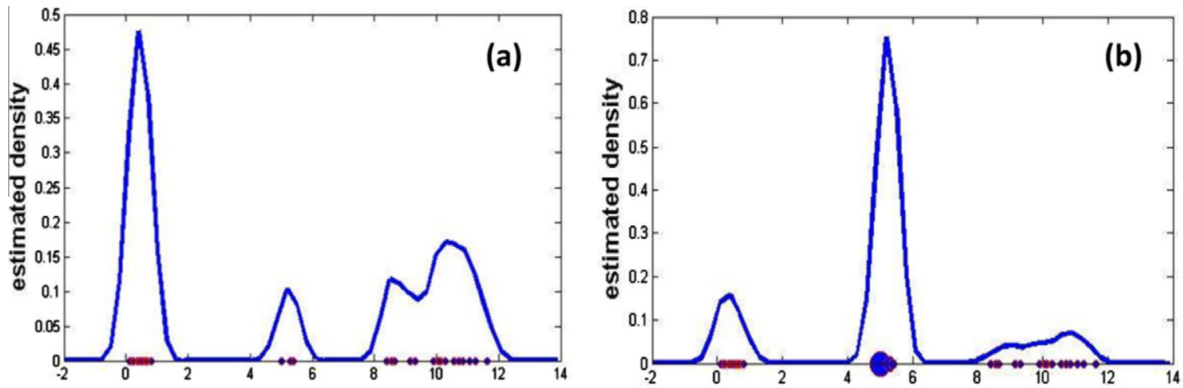

Fig. 2. Density estimation, with and without consideration of the respective "mass".

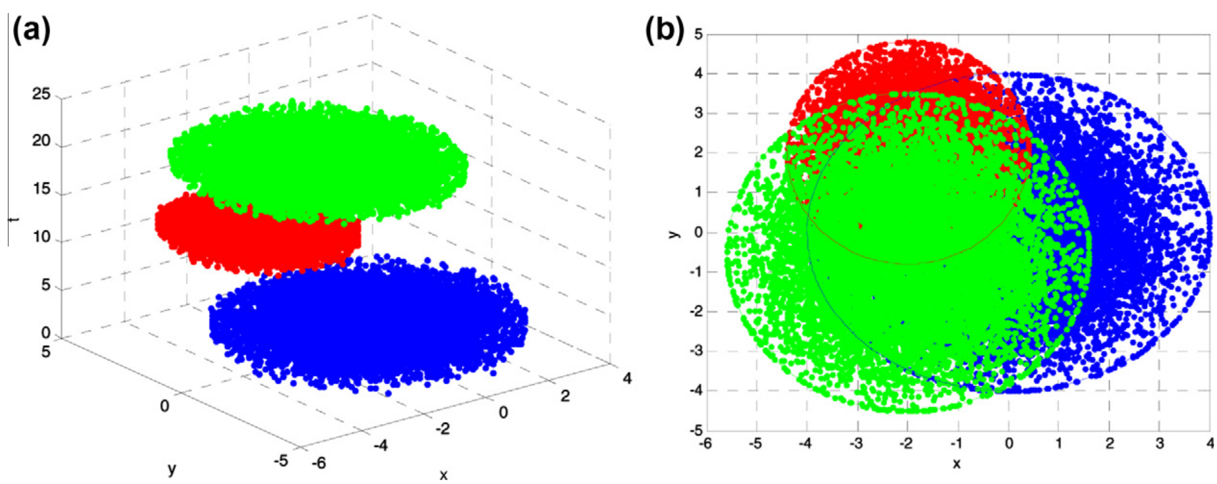

Fig. 3. Illustration of the effect of using the time dimension (a) and the result of dropping the time dimension (b). The three distinctive clusters that are well formed due to the different time occurrences heavily overlap when we move to spatial coordinates.

identifications of patterns of short-spaced seismic events attributed to smaller earthquakes accompanying individual main seismic events, instead of identifying possible patterns of the aforementioned broader relation. Therefore in most clustering algorithm those events are eliminated using for example the approach proposed by Reasenberg (Reasenberg, 1985). In our approach this matter is treated implicitly by the parameter *M* that forbids small magnitude events to blow up the volume of the cluster unless they are really close to the main event.

The cumulative seismic mass parameter *M* can control the volume of clusters. Apart from *M* the user has also to select the radius for the SM-DBSCAN and the number of clusters at the agglomerative hierarchical level as well as the appropriate scaling of the time dimension (no scaling was performed in the spatial coordinates since they are of the same order of magnitude). As it is shown in the following figures the results heavily depend on the selection of the aforementioned parameters. A number of configurations were tested and some of the results are depicted in the following figures. Apart from these three parameters a number of temporal units (scaling) were also employed ranging from 5 min to 30 min temporal units to evaluate its effect on the formation of the clusters.

Despite that the cluster formation is controlled by a complex interplay of the aforementioned parameters, there are some general patterns that can guide the selection of parameter values that yield "reasonable" results. More specifically selecting too small a

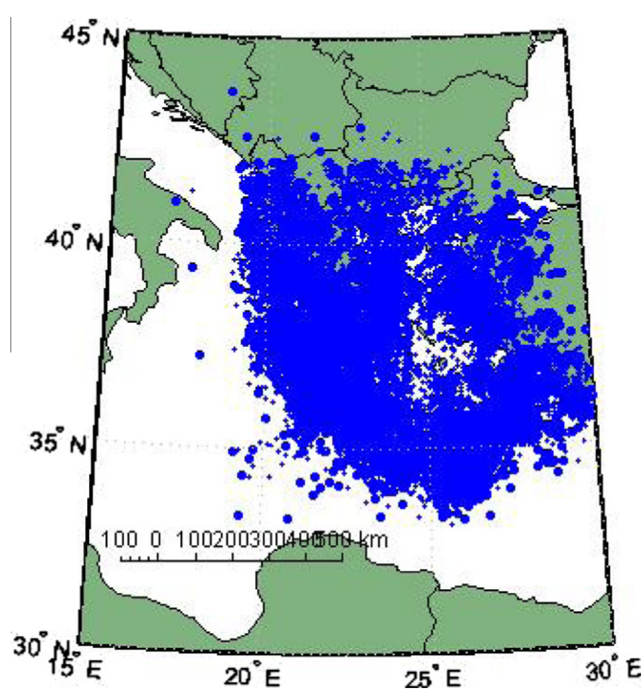

Fig. 4. Epicenters of earthquakes with $M > 1$ for the period 2000–2010.

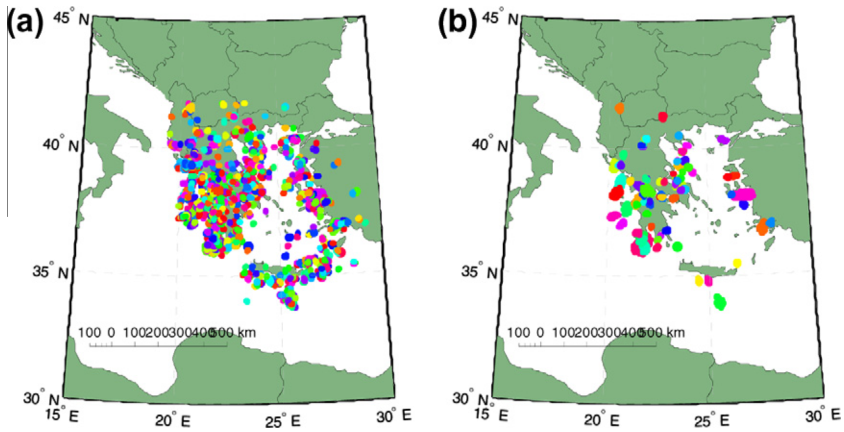

Fig. 5. The results of the application of the SM-DBSCAN for (a) $M = 10$, $Eps = 0.05$ and time unit $T = 5$ min and (b) $M = 50$, $Eps = 0.05$ and time unit $T = 5$. Two small a number for M fails to filter out the for- and the after-shocks, whereas too large a values leaves large areas (with known underlying fault sources) empty.

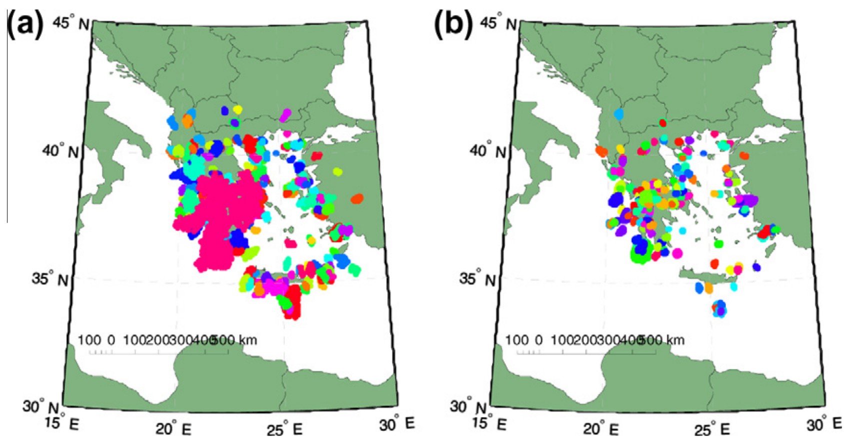

Fig. 6. The results of the application of the SM-DBSCAN for (a) $M = 30$, $Eps = 0.12$ and time unit $T = 5$ min and (b) $M = 50$, $Eps = 0.05$ and time unit $T = 5$. Two large a radius creates clusters that span more than one seismic sources, whereas too small a radius can lead to very narrow clusters.

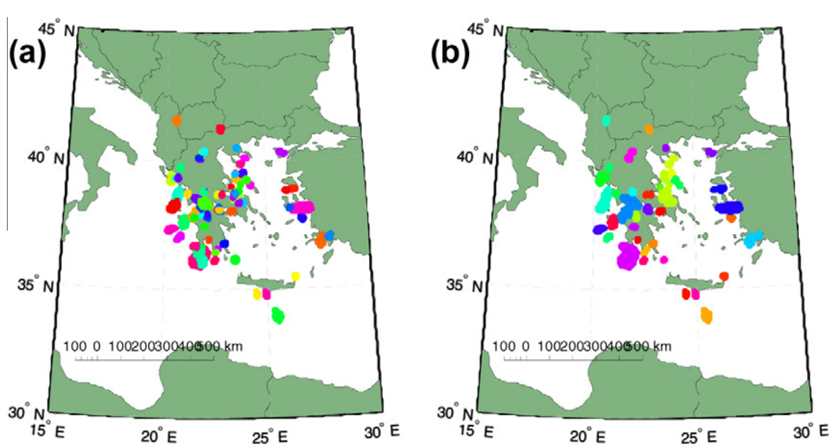

Fig. 7. The results of the application of the first stage (a) SM-DBSCAN with $M = 50$, $Eps = 0.05$ and time unit $T = 5$ min and (b) the second stage agglomerative clustering with 39 clusters. The selection of a lower number of clusters in the second stage produces a merging region of the Ionian and Corinthian bay faults.

value for the aggregated seismic parameter M a very large number of clusters will be formed (Fig. 5a) whereas too big a value results in very few clusters leaving "empty" areas where known faults exist (Fig. 5b). A large radius creates large clusters that artificially span areas that are known to belong to different faults (Fig. 6a) while on the other hand a small radius can create many small clusters that actually come from the same fault (Fig. 6b). Similar behavior we have for the selection of the time unit (scaling of the time dimension). Too large a unit creates very large clusters whereas too small a value leads to the formation of many small clusters.

In the case where there are many clusters the application of the second stage of the proposed approach, the use of agglomerative clustering, can lead to meaningful segregation; whereas in the case where the output of the first stage has produced artificially very large clusters it is not possible to compensate for them. Therefore careful tuning is needed and as a rule of thumb a larger number of clusters at the first stage is preferred over a smaller one.

Our observations suggest that more than one parameter setting can lead to similar results consistently identifying specific seismic formations. A quite high value for *M* and small values for *Eps* and *T* create quite concrete clusters at the first stage which are combined using the agglomerative scheme at the second stage to create a final set of 35–40 clusters Fig. 7. This is a bit higher than the number of clusters proposed by Weatherill and Burton (2009) and a bit less than the number of seismic zones proposed in Papazachos, Makropoulos, Latoussakis, & Theodulidis (1989).

4. Conclusions

In this work a novel approach to seismic event clustering is proposed. The approach is based on a modified implementation of the well-known DBSCAN algorithm that introduces the concept of accumulated seismic mass for the isolation of clusters of seismic events in time and space, called SM-DBSCAN, and then it employs single linkage agglomerative hierarchical clustering for a second level spatial clustering. The presented results show that the method is capable of finding irregularly shaped clusters which is a useful feature since fault seismic zones are rarely well-shaped. Moreover the algorithm was able to recognize among others, the south Cretan and the Ionian region as individual seismic regions in accordance with empirical results on distinctive seismic zones (Drakatos & Latoussakis, 2001; Papaioannou & Papazachos, 2000). In addition, although these are still preliminary results, it is worth noting that the algorithm perceives a small area south-east of Peloponnesus as a further individual seismic region separating the above two. Underground faults cartography (Seismotectonic map of Greece with seismo-geologic elements, 1989) of that region indicates the presence of two sets of parallel and in close proximity with one another underground faults extending throughout the region's vicinity, which do not appear to interact together with any of the neighboring underground faults (Konstantaras, in press; Konstantaras, Makris, Vallianatos, & Varley, 2007). Moreover since the usefulness of clustering applications should be considered within the specific domain context (Guyon, Von Luxburg, & Williamson 2009), the aforementioned results suggest that the novel proposed scheme could be a valuable tool for the automatic annotation of seismic catalogs after the appropriate parameter tuning.

Research to follow shall investigate the spatial distribution of seismic events produced by our proposed framework in order to test the causality effect implied by what is known as the Domino theory (Bürgmann, 2009; Olson & Allen, 2005). We will examine whether this physical bondage results in a pattern on the spatial distribution of the seismic sequence and at what geographical scale. Potential confirmation of the causality between adjacent seismic regions might provide significant information about the geographical scale upon which the theory of domino extents, at particular seismological areas. Moreover we shall use temporal pattern recognition techniques within the spatial framework of the individual cluster to model the sequence of events based on intelligent and soft computing systems. Furthermore, we shall also attempt to develop a method capable of automatically assigning earthquakes to seismic sources and subsequently "optimize" the involved parameters by correlating the resulted clusters with known faults as well as along with the development of predictive models and their capability to model local seismic (major) events and global models under the prism of the domino theory.

References

Ansari, A., Noorzad, A., & Zafarani, H. (2009). Clustering analysis of the seismic catalog of Iran. *Computers & Geosciences, 35*(3), 475–486.

Bürgmann, R. (2009). Earthquakes: Imperfect dominoes. *Nature Geoscience, 2*(2), 87–88.

Burton, P. W., Weatherill, G., Karnawati, D, & Pramumijoyo, S. (2008). Seismic hazard assessment and zoning in java: New and alternative probabilistic assessment models. *International Conference on Earthquake Engineering and Disaster Mitigation, Jakarta*.

Console, R., Murru, M., & Catalli, F. (2006). Physical and stochastic models of earthquake clustering. *Tectonophysics, 417*(1), 141–153.

Drakatos, G., & Latoussakis, J. (2001). A catalog of aftershock sequences in Greece (1971–1997): Their spatial and temporal characteristics. *Journal of Seismology, 5*, 137–145.

Ester, M., Kriegel, H., Sander, J, & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, (pp. 226–231).

Guyon, I., Von Luxburg, U., & Williamson, R. C. (2009). Clustering: Science or art. *NIPS 2009 Workshop on Clustering Theory*.

Hernandez, S., & Sallis, P. (2011). Modelling seismic activity using a Bayesian nonparametric method. *International Journal of Geology, 5*(4), 126–130.

Jain, A. K. (2010). Data clustering: 50 Years beyond K-means. *Pattern Recognition Letters*, 651–666.

Kayabol, K. (2012). A latent variable Bayesian approach to spatial clustering with background noise probability, networks and algorithms [PNA]. PNA-1106 (pp. 1–11).

Konstantaras, A. J. (in press). Classification of distinct seismic regions and regional temporal modelling of seismicity in the vicinity of the Hellenic seismic arc. *IEEE Selected Topics in Applied Earth Observations and Remote Sensing*, http://dx.doi.org/10.1109/JSTARS.2012.2227244.

Konstantaras, A., Makris, J. P., Vallianatos, F., & Varley, M. R. (2007). On the electric field transient anomaly observed at the time of the Kythira M 6.9 earthquake on January 2006. *Natural Hazards and Earth System Science, 7*(6), 677–682.

Konstantaras, A., Vallianatos, F., Varley, M. R., & Makris, J. P. (2008). Soft-computing modelling of seismicity in the southern Hellenic arc. *IEEE Geoscience and Remote Sensing Letters, 5*(3), 323–327.

Konstantaras, A., Varley, M. R., Vallianatos, F., Makris, J. P., Collins, G., & Holifield, P. (2007). Detection of weak seismo-electric signals upon the recordings of the electrotelluric field by means of neuro-fuzzy technology. *IEEE Geoscience and Remote Sensing Letters, 4*(1), 161–165.

Morrato, L., Orlrcka-Sikora, G., Costa, P., Suhadolc, Ch., Papaioannou, Ch., & Papazachos, C. B. (2007). A deterministic seismic hazard analysis for shallow earthquakes in Greece. *Tectonophysics, 442*(1–4), 66–82.

Mukhopadhyay, B., Fnais, M., Mukhopadhyay, M., & Dasgupta, S. (2010). Seismic cluster analysis for the Burmese–Andaman and West Sunda Arc: Insight into subduction kinematics and seismic potentiality. *Geomatics, Natural Hazards and Risk, 1*(4), 283–314.

Mukhopadhyay, B., Mukhopadhyay, M., & Dasgupta, S. (2011). Seismic clusters and their characteristics at the Arabian sea triple junction: Supportive evidences for plate margin deformations. *Journal Geological Society of India, 78*, 131–146.

Olson, E. L., & Allen, R. M. (2005). The deterministic nature of earthquake rupture. *Nature, 438*(7065), 212–215.

Papaioannou, Ch. A., & Papazachos, B. C. (2000). Time-independent and time-dependent seismic hazard in Greece based on seismogenic sources. *Bulletin of the Seismological Society of America, 90*, 22–33.

Papazachos, B. C., Makropoulos, K. C., Latoussakis, J, & Theodulidis, N.P. (1989). Elaboration of a map of seismic hazard in Greece. *2nd Report for the Program of Organization of Antiseismic Design and Protection*, (pp. 1–24).

Pondard, N., Armijo, R., King, G. C., Meyer, B., & Flerit, F. (2007). Fault interactions in the Sea of Marmara pull-apart (North Anatolian Fault): Earthquake clustering and propagating earthquake sequences. *Geophysical Journal International, 171*(3), 1185–1197.

Reasenberg, P. A. (1985). Second-order moment of central Californiaseismicity, 1969–1982. *Journal of Geophysical Research, 90*, 5479–5495.

Saleur, H., Sammis, C. G., & Sornette, D. (1996). Renormalization group theory of earthquakes. *Nonlinear Processes in Geophysics, 3*(2), 102–109.

Seismotectonic map of Greece with seismo-geologic elements (1989). Institute of Geology and Mineral Exploration [Online].

Theodoridis, S., & Koutroumbas, K. (2008). *Pattern recognition* (4th ed.). Academic Press.

Vallianatos, F., & Tzanis, A. (1998). Electric current generation associated with the deformation rate of a solid: Preseismic and coseismic signals. *Physics and Chemistry of the Earth, 23*(9–10), 933–938.

Weatherill, G., & Burton, P. W. (2009). Delineation of shallow seismic source zones using k-means cluster analysis, with application to the Aegean region. *Geophysical Journal International, 176*(2), 565–588.

Weatherhill, G., & Burton, P. W. (2010). An alternative approach to probabilistic seismichazard analysis in the Aegean region using Monte Carlo simulation. *Tectonophysics, 492*(1–4), 253–278.

Zamani, A., & Hashemi, N. (2004). Computer-based self-organized tectonic zoning: A tentative pattern recognition for Iran. *Computers & Geosciences, 30*, 705–718.

Zamani, A., Khalili, M., & Gerami, A. (2011). Computer-based self-organized tectonic zoning revisited: Scientific criterion for determining the optimum number of zones. *Tectonophysics, 510*, 207–216.

Zamani, A., Nedaei, M., & Boostani, R. (2009). Tectonic zoning of Iran based on self-organizing map. *Journal of Applied Sciences, 9*, 4099–4114.

Zobin, V. M. (1996). Earthquake clustering in shallow subduction zones: Kamchatka and Mexico. *Physics of the earth and planetary interiors, 97*(1), 205–218.