

# Intelligent Spatial-Clustering of Seismicity in the Vicinity of the Hellenic Seismic Arc

A. J. Konstantaras<sup>1</sup>, E. Katsifarakis<sup>1</sup>, E. Maravelakis<sup>2</sup>, E. Skounakis<sup>2</sup>, E. Kokkinos<sup>3</sup> & E. Karapidakis<sup>4</sup>

<sup>1</sup> Laboratory of Computational Intelligence and Software Engineering (CISE lab), Technological Educational Institute of Crete, Chania, Greece

<sup>2</sup> Design and Manufacturing Laboratory (DML), Technological Educational Institute of Crete, Chania, Greece

<sup>3</sup> Department of Electronics, T.E.I. Crete, Chania, Greece

Correspondence: A. J. Konstantaras, Laboratory of Computational Intelligence and Software Engineering (CISE lab), Technological Educational Institute of Crete, Romanou 3, Chania 73133, Greece. Tel: 30-282-102-3033. E-mail: akonstantaras@staff.teicrete.gr

Received: January 12, 2012 Accepted: January 29, 2012 Online Published: May 16, 2012

doi:10.5539/esr.v1n2p1

URL: <http://dx.doi.org/10.5539/esr.v1n2p1>

## Abstract

This research paper discusses possible seismic cluster formation and evolution in the vicinity of the Hellenic seismic arc and proposes a graphical user-interface monitoring and analysis tool based on various commercial and self-developed clustering algorithms for cluster discrimination, evolution and visualization. Self-developed algorithms enable the processing of both a) all recorder earthquakes and b) main seismic events alone, excluding foreshocks and aftershocks, by incorporating dynamic filters in space and time. The user can also import external formulae for the computation of the total earthquake preparation time, aftershocks duration and radius of the sphere of earthquake preparation region, and can also select specific regions of interest as well as the entire seismic map. The seismic imaging tool also addresses the concept of topical seismic cluster formation. Seismological maps indicate the presence of several seismic swarms forming within the region of the Hellenic arc, which appear to be either distinct or interacting together in groups of two or more. The identification of the number of possibly individual seismic clusters in a seismological area is a very challenging task by itself, which becomes even more complicated when investigating their outer boundaries especially in the case of multiple interacting clusters. The proposed imaging tool incorporates clustering algorithms that allow the user to apply various techniques for cluster identification, such as density based functions, gradient descent, centre of gravity, evolutionary allocation, and even import expert knowledge regarding the number of individual seismic clusters present.

**Keywords:** seismic epicentre clustering, spatial distribution, distinct earthquake regions, intelligent clustering algorithms

## 1. Introduction

Seismic cluster discrimination is of utmost importance in seismology as it can provide valuable information regarding the topology of the seismic phenomenon in relation with underlying faults. In most cases, little detailed information is readily available regarding the underground structure of a seismogenic region of interest, which in terms of epicenter depth extends from a few meters to several tens of kilometers below sea level. What is made apparent is a distorted reflection of the underlying faults' network on the surface of the planet painted by numerous compact seismic swarms that extend all the way across all active tectonic regions. The fact that underground faults are rarely distinct and in most cases they tend to form large topical or extent interacting networks complicates the process of surface seismic cluster discrimination as it is very difficult to identify discrete cluster boundaries. To make matters even more complex interacting clusters penetrate well into the stronghold vicinity of their companion and vice versa which can result in faulty allocations of seismic events to a particular cluster.

To evaluate this problem this paper applies a number of state-of-the-art clustering algorithms on seismic maps and proposes an intelligent platform allowing clusters to compete for newly existing seismic data points. Seismic data are made available either as distinct points or grouped together in seismic sequences, i.e. foreshocks, main

earthquake and aftershocks, in which case the user can select to work with main earthquakes only. Work can be carried out upon the entire seismic map or within any user-specified orthogonal region of interest. State-of-the-art clustering algorithms, such as fuzzy C-means clustering (Dunn, 1973; Bezdek, 1981), density-based spatial clustering (Ester et al., 1996), quantum clustering (Horn et al., 2002; Horn et al., 2003) and a self-developed dynamic spatial clustering algorithm are then applied comparatively upon the seismic data deriving distinct clusters of various shapes and dimensions. Furthermore, an intelligent platform has been developed to simplify the processes of dynamic evolutionary clustering at the appearance of new earthquakes upon the seismic clustering map by using a semi-automated interface. The platform applies contours and labels initial clusters predefined by the aforementioned clustering algorithms and enables them to compete for cluster allocation of new seismic events presented upon the seismic map using the centre-of-gravity algorithm. This approach allows irregularly shaped clusters to blend together and potential interact with each other depending upon the topology of initial cluster formation and the location of newly emerging earthquakes.

## 2. Epicentre Clustering

In order to identify regions of increased seismic activity we have implemented in our software a series of advanced clustering algorithms to cluster in the spatial domain, earthquake events identified as main events with the dynamic spatial clustering method. Since all spatial clustering algorithms rely on distance calculations between points, our software uses an ellipsoidal earth projected to a plane formula to estimate distance between earthquake events based on their geographical coordinates. The data throughout the paper have been obtained by the National Observatory of Athens Institute of Geodynamics seismicity catalogue (available online: <http://www.gein.noa.gr/services/cat.html>) for the entire Greek vicinity during the year 2009.

### 2.1 Fuzzy C-Means Clustering

The Fuzzy C-Means algorithm (Dunn, 1973; Bezdek, 1981) requires a specific number of cluster centers to be provided on initialization and is implemented so that for each cluster  $k$  the grade of cluster membership  $u_k$  for each earthquake event  $x$  is related to the inverse of its distance from the cluster center (1):

$$u_k(x) = \frac{1}{d(\text{center}_k, x)^2} \quad (1)$$

In our current implementation (Figure 1) each event is assigned to the cluster with the highest grade of membership but since cluster membership ambiguities occur quite often, future implementations could take advantage of partial membership grades.

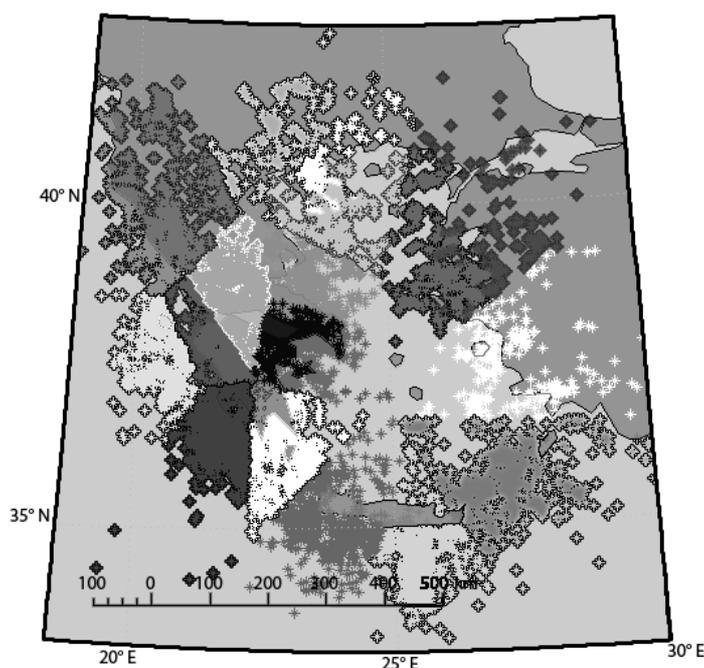


Figure 1. Spatial clustering indicative results using the Fuzzy C-Means algorithm with 15 cluster centres applied upon the entire dataset of earthquake events. Each colour indicates a distinct seismic cluster

### 2.2 Density-Based Spatial Clustering

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm (Ester et al., 1996) (Figure 2) does not require the number of clusters to be specified a priori as opposed to Fuzzy C-Means. Instead it clusters events spatially based on the notions of density reachability, density connectivity with respect to parameters Eps-neighborhood radius and the minimal number of objects considered as a cluster (MinPts). The *Eps neighborhood* of a point  $p$ , denoted by  $NEps(p)$ , is defined by (2) :

$$NEps(p) = \{q \in D \mid \text{dist}(p,q) \leq Eps\} \quad (2)$$

A cluster  $C$  is defined as a non-empty subset of a database of points  $D$  with respect to Eps and MinPts satisfying the following conditions:

- a)  $\forall p, q$ : if  $p \in C$  and  $q$  is density-reachable from  $p$  wrt. Eps and MinPts, then  $q \in C$ .
- b)  $\forall p, q \in C$ :  $p$  is density-connected to  $q$  wrt. Eps and MinPts.

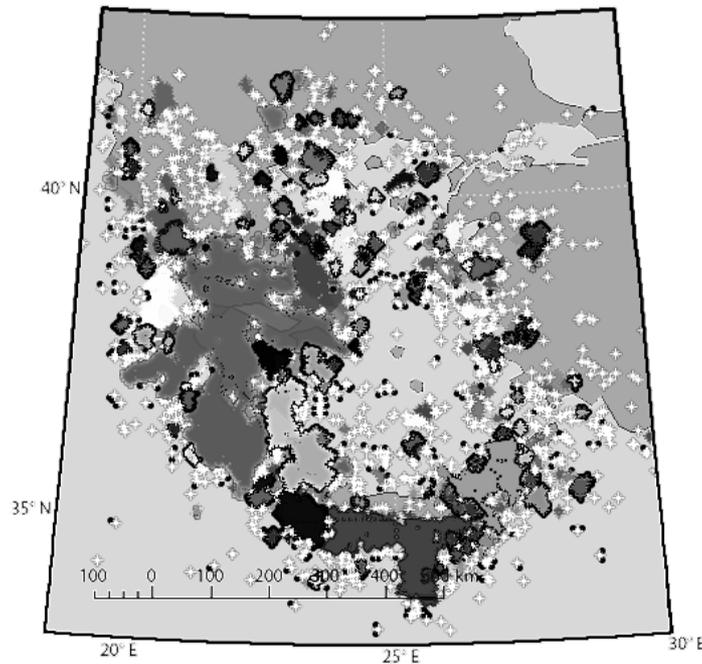


Figure 2. Spatial clustering indicative results using the DBSCAN algorithm applied upon the entire dataset of earthquake events. Yellow indicates noise - each other colour indicates a distinct seismic cluster

### 2.3 Quantum Clustering

The Quantum Clustering (QC) algorithm (Horn & Gottlieb, 2002; Horn & Axel, 2003) (Figure 3) starts out with a Parzen window approach assigning to each data-point a Gaussian of width  $\sigma$  which can be represented up to an overall normalization by the following Parzen-window estimator (3),  $\psi(x)$  :

$$\psi(x) = \sum_i e^{-\frac{(x-x_i)^2}{2\sigma^2}} \quad (3)$$

where  $x_i$  are the data points.

This can serve as a probability density function generating the data. One then proceeds to construct a potential function (4),  $V(x)$  :

$$V(x) = E + \frac{\frac{\sigma^2}{2} \nabla^2 \psi}{\psi} \quad (4)$$

where

$$E = -\min + \frac{\frac{\sigma^2}{2} \nabla^2 \psi}{\psi} \quad (5)$$

thus rendering  $V$  positive definite (5). In fact  $V$  has a global minimum at zero, and grows as a polynomial of second order outside the domain over which the data points are defined. Within this domain,  $V$  develops minima that are identified with cluster centres.

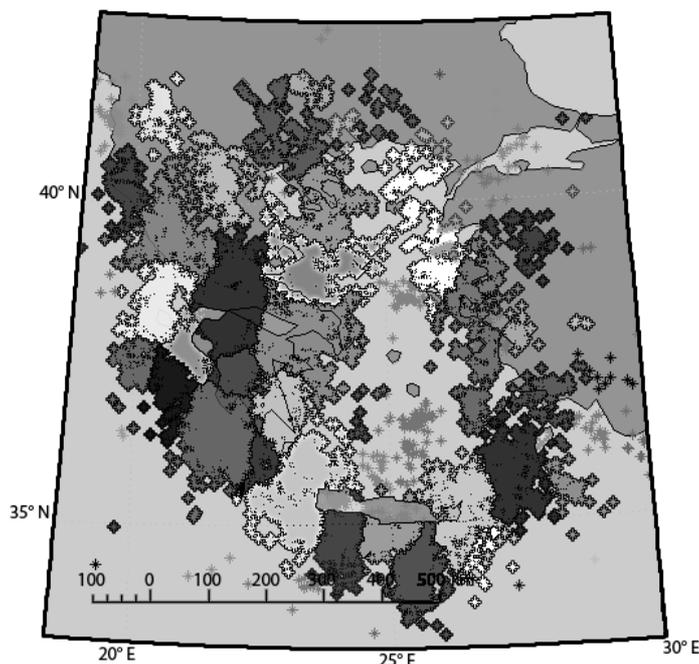


Figure 3. Spatial clustering indicative results using the Quantum Clustering algorithm applied upon the entire dataset of earthquake events. Each colour indicates a distinct seismic cluster

#### 2.4 Dynamic Spatial Clustering

The self-developed dynamic spatial clustering algorithm is based on the concept of earthquake strain radii. The strain radius of an earthquake is defined as the radius of a hypothetical circle, centred at the epicentre of said earthquake, which encloses the zone of effective manifestation of the precursor deformations (Dobrovolsky et al., 1979).

Two different strain radius calculation methods (6, 7) have been implemented into the clustering solution using the following formulas based on event magnitude:

$$a) \rho = 10^{0.43 M} \text{ km (Dobrovolsky et al., 1979)} \quad (6)$$

$$b) \rho = 10^{0.414 M - 1.696} \text{ km (Dobrovolsky et al., 1989)} \quad (7)$$

where  $M$  is the event magnitude.

The time window can be a) either specified as a static period of days common to all events before and after the time of occurrence or b) dynamically calculated (Zubkov, 1987; Stein & Liu, 2009; Alden, 2009) for the time intervals before (8) and after (9) the main earthquake, respectively, where:

$$t_{\text{before}} = 10^{(0.5 M - 2.1)} * 365 \text{ days} \quad (8)$$

$$t_{\text{after}} = 10 \text{ years} \quad (9)$$

where  $t_{\text{after}}$  applies specifically to the current seismological region under investigation.

The dynamic spatial clustering algorithm works as follows:

- 1) Earthquake events are sorted in chronologically ascending order.
- 2) Events that do not belong to a cluster are processed individually, starting with the event that occurred first. That event becomes the currently processed event.
- 3) The strain radius and time window of the currently processed event are calculated. A new cluster is created and all events that occurred within the strain radius area and time window of the currently processed event are assigned to the new cluster.
- 4) The event with the greatest magnitude in the newly created cluster is determined. That event becomes the cluster's main event. If the main event is the currently processed event, the process advances, otherwise the cluster main event becomes the currently processed event. The process returns to step 3).
- 5) The next event in the dataset that does not belong to a cluster becomes the currently processed event and the process returns to step 5.

The above process is repeated until all events are assigned to a cluster.

Since expandability of the software solution was one of the key factors considered at the time of development, additional strain radius and time window calculation methods (Konstantaras et al., 2008) can be easily appended in the future. Furthermore a spatial filter capability is available to reduce large-territory datasets by processing only events in a smaller rectangular geographical region of interest, defined by its upper-left and lower-right corners' latitude and longitude (Figure 4).

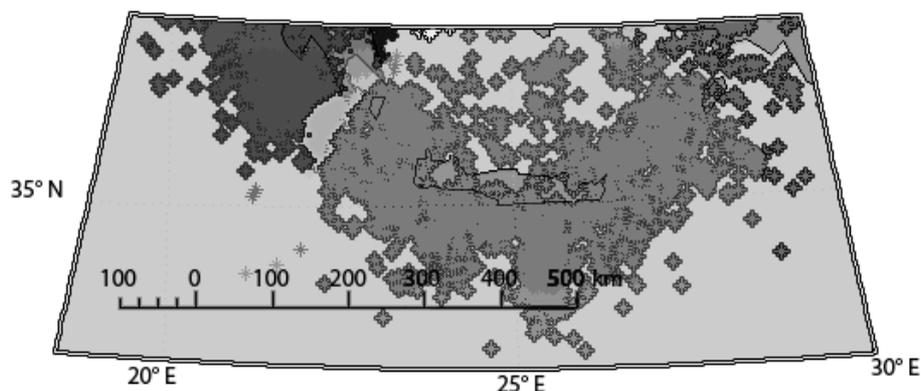


Figure 4. Dynamic spatial clustering applied upon a reduced version of the dataset using the spatial filter feature of our clustering software to focus on a smaller geographical region of interest ranging from 36°80'N, 20°50'E to 34°N, 29°20'E. Each colour indicates a distinct seismic cluster

The dynamic spatial clustering algorithm is an iterative agglomerative clustering algorithm. Initially, strain radius and time-window values are calculated for every event on the dataset and events are ordered by their time of occurrence and processed in the following manner: an unclustered event becomes the centre of a new cluster and all events within its strain radius and time-window become cluster members. At this point a competitive process begins where we look for the event with the highest magnitude within the newly-formed cluster. If that event is not the current cluster centre, a new cluster is formed and the aforementioned competitive process is repeated recursively. With the formation of a new cluster seismic events spatially and temporally located towards the far outer region of the initial cluster with respect to the new spatio-temporal cluster centre might not fall within the strain radius and/or the time window of the new cluster; therefore they remain as members of the initial cluster. Consequently, the dynamic spatial clustering algorithm forms irregularly-shaped seismic clusters allowing cluster interaction by enabling multiple clusters to occupy the same geographical area by exploiting time as an additional physical layer. Indicative results of our dynamic spatial clustering algorithm applied on the National Observatory of Athens Institute of Geodynamics seismicity catalogue (available online: <http://www.gein.noa.gr/services/cat.html>) throughout the Greek vicinity during 2009 using the first strain radius equation can be examined in Figure 5.

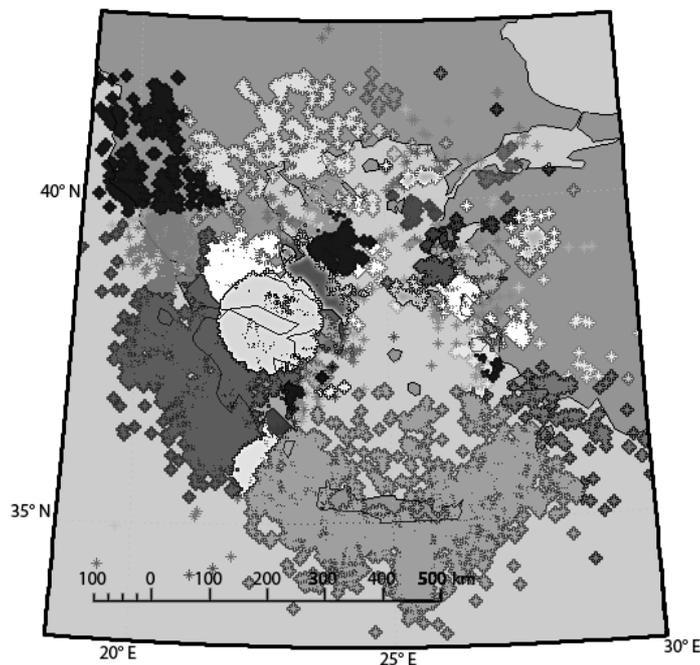


Figure 5. Dynamic spatial clustering applied upon entire dataset of earthquake events. Each colour indicates a distinct seismic cluster, cluster overlapping within the same region but at a different time frame is made apparent by the invasion of different colour data points well into the vicinity of compact clusters

### 3. Evolutionary Clustering Multifunctional Process

In order to identify and extract information from complex seismic images, we have developed the intelligent platform proposed below. This platform simplifies the processes of grouping and labelling several areas into a seismic image by using a semi-automated interface. The platform is based on the .NET framework architecture (available online: <http://msdn.microsoft.com/en-us/kb/kb00829019.aspx>) and can be used in any Windows-based computer with the .NET framework ver. 2.0 or later. However, .NET framework permits, in concept, the porting of the application in many other Linux based environments, using the Mono project (available online: <http://msdn.microsoft.com/en-us/kb/kb00829019.aspx>) functionality and libraries, so the user can run an application using the No-Touch Deployment (available online: <http://msdn.microsoft.com/en-us/kb/kb00829019.aspx>).

Three different phases explain the sequence of the processes shown in Figure 6 and Figure 7.

- **Phase 1:** The seismic data of an area which are depicted on a seismic data map are processed with several segmentation algorithms in order to produce seismic clustering maps such as the one shown in Figure 7 using quantum clustering. This procedure implemented by Phase 1 of the Seismic Monitor Organizer divides a given image into separate regions forming an initial set of clusters.
- **Phase 2:** The resulting seismic clustering map from the above procedure is loaded onto the platform for further analysis. By applying a fast multifunctional version of the classical region-growing segmentation algorithm, we can define precisely and apply contours and labels to various areas of interest.
- **Phase 3:** The region-growing algorithm is assigned to allocate new earthquakes to a particular cluster based upon the magnitude of the centre of gravity of the existing clusters; or create a new cluster if all centers of gravity are above a predefined by the user upper threshold point.

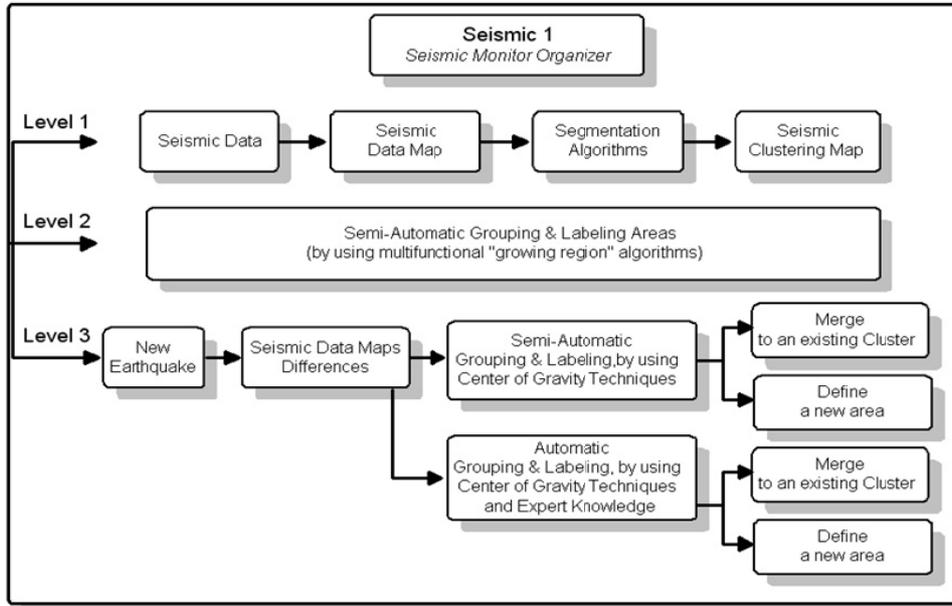


Figure 6. A schematic representation of the proposed platform

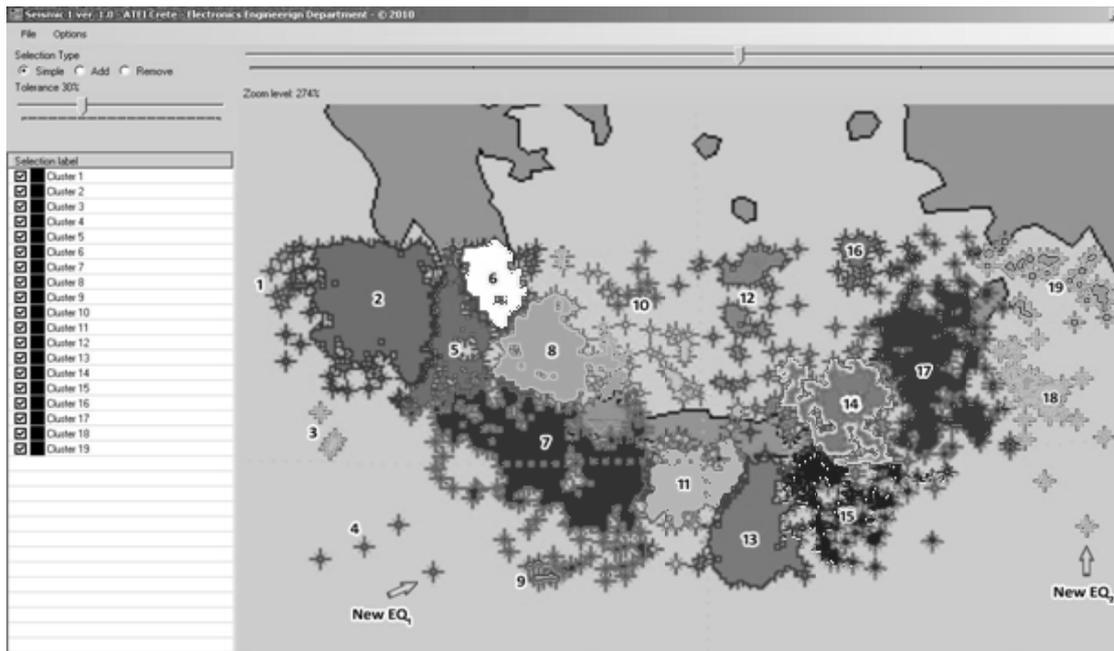


Figure 7. Example illustration of using the Centre of Gravity technique for the classification of newly introduced Earthquakes where New EQ<sub>1</sub> and New EQ<sub>2</sub> are being allocated to clusters “4” and “18”, respectively

### 3.1 The Region-growing Algorithm

The function of the “Region-Growing Algorithm” is based on finding and selecting all the pixels around a pre-specified user-selected initial point that are similar in gray intensity. A tolerance value can also be set by the user to determine how closely to match colours (higher tolerance ends up in a larger selection) (Fujita & Fejri, 2006; Skounakis et al., 2009). All the points selected by the algorithm are automatically stored in an image mask of the same size as the original image. Each pixel in the image mask represents the position of the selected pixels in the original image and is used to label the delineated areas. To enhance its effectiveness, we have implemented a faster version of this algorithm which excludes all the pixels that have already been examined ensuring that the algorithm does not check them again.

When new points are being introduced into the seismic clustering map, e.g. after a new earthquake (see arrows in Figure 7), the platform allows for either their definition as new seismic clusters, or their mergence with one of the predefined seismic cluster. This can be achieved either semi-automatically, allowing for user interaction, or automatically by selecting the cluster with the minimum centre of gravity value for mergence or by defining a new cluster if the value of the centre of gravity exceeds a certain upper threshold induced by the user. The centre of gravity (CoG) of a region defines the centre of a region, and it is important in localizing that region. The centre coordinates  $r_G$  (10) and  $c_G$  (11) are given by:

$$r_G = \frac{1}{N} \sum_{r=0}^{Rows-1} \sum_{c=0}^{Columns-1} r \cdot f(r,c) \quad (10)$$

$$c_G = \frac{1}{N} \sum_{r=0}^{Rows-1} \sum_{c=0}^{Columns-1} c \cdot f(r,c) \quad (11)$$

$r$  and  $c$  are the coordinates of the image,  $N$  represents the number of pixels of the region and  $f(r,c)$  is the image function. These formulae yield an '1' if the current pixel  $(r,c)$  falls within the region of a particular cluster, otherwise a '0' is obtained. The platform calculates the CoG of all the presented areas as well as the distances between the CoG of the new area and the existing ones automatically and presents a list of the proposed areas to be merged in, as shown in Figure 7.

### 3.2 Experimental Procedure

The experimental procedure comprises of six steps:

**Step 1:** The original "Seismic Data Map" is used to produce "Seismic Group Maps" by using several segmentation algorithms.

**Step 2:** A "Seismic Group Map" from the above procedure, is loaded on the platform.

**Step 3:** The right click on the list area of the platform enables the labelling (name & colour) of a newly introduced area on the "Seismic Group Map".

**Step 4:** The zoom and pan controls which are integrated in the platform allow a more detailed view of the region of interest on the map, and render its processing more efficient.

**Step 5:** The application of the region – growing algorithm to a region of interest on the "Seismic Group Map" leads to the definition of that area. The option to adjust the tolerance level of the gray-level differences is available during this process. The "add" and "remove" options can also be used to define more complex areas.

**Step 6:** A switch between "contour" and "area" of the selected areas is also available.

Steps 3 to 6 can be repeated to define new areas on the Seismic Group Map.

## 4. Results

The application of various clustering algorithms on seismic data produces a variety of results with several commons and some distinct differences. Larger seismic clusters appear to be depicted by most if not all of the clustering algorithms. Such an observation strengthens the possibility that these clusters in particular highlight a number of distinct seismic regions possibly baring a largely independent network of underground faults. In several cases there are more than one clustering algorithms depicting similar cluster boundaries at close proximity with one another, which, when applicable, gives a strong indication of the boarder region between neighbouring clusters. The opposite result provides a strong indication that the clusters involved are possible strongly interacting together at some part of their vicinity, which might be the result of inter-crossing underground faults at that particular region. Furthermore, if the clusters involved do not present a dense seismic swarm at some other part of their vicinity then it is possible that the clustering algorithms are at fault and what appears as individual seismic clusters might actually be parts of single broader seismic cluster.

Every clustering algorithm exhibits some particular distinct characteristics in terms of its operation and outcomes worth noting. The fuzzy c-means algorithm does not class any cluster point to a particular cluster. Instead it assigns to it various degrees of membership for a number of clusters. This principle provides a great means of identifying the extent of areas that are in dispute and claimed by various clusters as well as the undisputed strongholds that form the core of each cluster. The fuzzy c-means algorithm also allows for expert knowledge to

be imported in terms of the number of clusters present. The density based clustering algorithm does not necessarily class every cluster point to a particular cluster nor it creates a new cluster for every few scarce points throughout the seismic map. Instead it creates an open cluster to which, in effect, all un-classed seismic events are allocated. This principle reduces significantly the overall number of seismic clusters with respect to other clustering algorithms. The sophistication of the quantum clustering algorithm provides data clusters with more irregular shapes expanding towards areas that could well have been anticipated to belong to a neighbouring cluster. The self developed dynamic spatial clustering algorithm enables multiple clusters to occupy different seismic events located within the same geographical area by exploiting time as an additional physical layer.

The evolutionary clustering multifunctional process is an image processing interactive approach to the problem of cluster formation working in a genesis and growth manner. It relies on an existing clustering algorithm, e.g. quantum clustering, to create early seismic clusters at their infancy, i.e. containing few seismic events, and encompasses a region-growing algorithm to allocate new seismic data points to existing clusters. The region growing algorithm computes the centre of gravity of all existing clusters and measures their distance from the location of the next seismic event. That way the infant clusters compete for every new earthquake thereby growing in size and at various irregular directions. In the case where the distance from the nearest centre of gravity exceeds a maximum user-defined threshold point a new seismic cluster is formed.

## 5. Conclusions

Understanding the seismic phenomenon remains an open front in the scientific community. An insight on the environment that hosts the phenomenon could provide valuable information regarding its nature, genesis and propagation to the surface. Satellite images and Earth-based investigations of the Earth's interior have revealed the existence of an underground faults network almost throughout the most active seismic regions on the planet but they are all conducted at a distance from the source of interest. The work carried out on this paper acts complementary to these methods and aims to decode information carried through directly from the source to the surface of the planet via seismic waves. The numerous seismic events recorded on the surface of the planet form a distorted reflection of the underlying faults network. Perhaps working together these techniques could strengthen the validity of observations regarding the identification of the number of possibly individual or interacting together seismic clusters of various seismological regions as well as their outer boundaries and areas of interaction.

## References

- Alden, A. (2009). How long do aftershocks last. Retrieved from <http://geology.about.com/b/2009/11/04/how-long-do-aftershocks-last.htm>
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.
- Dobrovolsky, I. P., Gershenzon, N. I., & Gokhberg, M. B. (1989). Theory of electrokinetic effects occurring at the final stage in the preparation of a tectonic earthquake. *Physics of the Earth and Planetary Interiors*, 57, 144-156. [http://dx.doi.org/10.1016/0031-9201\(89\)90224-0](http://dx.doi.org/10.1016/0031-9201(89)90224-0)
- Dobrovolsky, I. P., Zubkov, S. I., & Miachkin, V. I. (1979). Estimation of the size of Earthquake Preparation Zones. *Pure and Applied Geophysics*, 117(5), 1025-1044. <http://dx.doi.org/10.1007/BF00876083>
- Dunn, J. (1973). A fuzzy relative of the Isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*, 3(3), 32-57. <http://dx.doi.org/10.1080/01969727308546046>
- 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)*, 226-231.
- Fujita, H., & Fejri, M. (2006). New trends in software methodologies, tools and techniques: How to create the magic wand?. *Proceedings of the Fifth SoMeT-06*. pp. 127.
- Horn, D., & Axel, I. (2003). Novel clustering algorithm for microarray expression data in a truncated SVD space. *Bioinformatics*, 19, 1110-1115. <http://dx.doi.org/10.1093/bioinformatics/btg053>
- Horn, D., & Gottlieb, A. (2002). Algorithm for Data Clustering in Pattern Recognition Problems Based on Quantum Mechanics. *Physical Review Letters*, 88(1), 18702. <http://dx.doi.org/10.1103/PhysRevLett.88.018702>
- 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. <http://dx.doi.org/10.1109/LGRS.2008.916069>

- Skounakis, E., Sakkalis, V., Marias, K., Banitsas, K., & Graf, N. (2009). DoctorEye: A multifunctional open platform for fast annotation and visualization of tumours in medical images. *IEEE Proceedings*, 3759-3762.
- Stein, S., & Liu, M. (2009). Long aftershock sequences within continents and implications for earthquake hazard assessment. *Nature*, 462, 87-89. <http://dx.doi.org/10.1038/nature08502>
- Zubkov, S. I. (1987). The appearance times of earthquake precursors. *Izv. Akad. Nauk SSSR Fiz. Zemli (Solid Earth)*, 5, 87-91.