Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space

Witold Dzwinel[1], David A.Yuen[2],
Krzysztof Boryczko[1,2], Yehuda Ben-Zion[3], Shoichi Yoshioka[4], Takeo Ito[5]

[1] *AGH Institute of Computer Science, al. Mickiewicza 30, 30-059, Kraków, Poland*
[2] *Minnesota Supercomputing Institute, Univ. of Minnesota, Minneapolis, MN 55455, USA*
[3] *Department of Earth Sciences, University of Southern California, Los Angeles, CA 98809, USA*
[4] *Dept. of Earth and Planteary Sciences, Kyushu University, Fukuoka, 812-8581, Japan*
[5] *Graduate School of Environmental Studies, Nagoya, University, Furo-cho, Nagoya, Aichi, 464-8602, Japan*

Abstract

A novel technique based on cluster analysis of the multi-resolutional structure of earthquake patterns is developed and applied to observed and synthetic seismic catalogs. The observed data represent seismic activities situated around the Japanese islands in the 1997-2003 time interval. The synthetic data were generated by numerical simulations for various cases of a heterogeneous fault governed by 3-D elastic dislocation and power-law creep. At the highest resolution, we analyze the local cluster structure in the data space of seismic events for the two types of catalogs by using an agglomerative clustering algorithm. We demonstrate that small magnitude events produce local spatio-temporal patches corresponding to neighboring large events. Seismic events, quantized in space and time, generate the multi-dimensional feature space of the earthquake parameters. Using a non-hierarchical clustering algorithm and multi-dimensional scaling, we explore the multitudinous earthquakes by real-time 3-D visualization and inspection of multivariate clusters. At the resolutions characteristic of the earthquake parameters, all of the ongoing seismicity before and after largest events accumulate to a global structure consisting of a few separate clusters in the feature space. We show that by combining the clustering results from low and high resolution spaces, we can recognize precursory events more precisely and decode vital information that cannot be discerned at a single level of resolution.

Keywords: Earthquakes, clustering, earthquake prediction, visualization, data mining

(*Submitted to* ***Earth and Planetary Sci. Letters***, August, 2003)

1 Introduction

The understanding of earthquake dynamics and development of forecasting algorithms require a knowledge and skill in both measurement and analysis that cover various types of data, such as seismic, electromagnetic, gravitational [Song and Simons, 2003], geodetic, geochemical, etc. The Gutenberg-Richter power-law distribution of earthquake sizes implies that the largest events are surrounded (in space and time) by a large number of small events (e.g. [Wesnousky, 1994; Ben-Zion and Rice, 1995]). The multi-dimensional and multi-resolutional structure of this global cluster depend strongly on geological and geophysical conditions [Miller et. al., 1999; Ben-Zion and Lyakhovsky, 2002] past seismic activities [Rundle et. al., 2001], closely associated events (e.g., volcano eruptions) and time sequence of the earthquakes forming isolated events, patches, swarms etc.

Investigations on earthquake predictions are based on the assumption that all of the regional factors can be filtered out and general information about the earthquake precursory patterns can be extracted [Geller et. al., 1996]. This extraction process is usually performed by using classical statistical or pattern recognition methodology. Feature extraction involves a pre-selection process of various statistical properties of data and generation of a set of the seismicity parameters [Eneva and Ben-Zion, 1997a,b], which correspond to linearly independent coordinates in the feature space. The seismicity parameters in the form of time series can be analyzed by using various pattern recognition techniques ranging from fuzzy sets theory and expert systems (e.g. [Wang and Gengfeng, 1996], multi-dimensional wavelets [Enescu et. al, 2002; Erlebacher and Yuen, 2003] to neural networks [Joswig 1990; Dowla 1995; Tiira, 1999; Rundle et. al., 2001].

The prediction of the earthquakes is a very difficult and challenging task; we cannot operate on only one level of resolution. The coarse graining of the original data can destroy the local dependences between the events and the isolated earthquakes by neglecting, e.g., their local spatial localization. In this manner, the subtle correlations between the earthquakes and preceding patches of events can be dissolved in the background of uncorrelated and noisy data. In [Dzwinel et. al., 2003] we have proposed a new approach employing clustering for multivariate analysis of seismic data. We can extract local spatio-temporal clusters of low magnitude events and recognize correlations between the clusters and the earthquakes. We showed that these clusters could reflect clearly the short-term trends in seismic activities followed by isolated large events. However, local clustering of seismic events is not able to extract an overall picture concerning the precursory patterns.

We do not use a standard software package. Our goal is to construct an interactive system for data-mining, which allows one to match the most appropriate clustering schemes on the structure of actual seismic data. Our data-mining techniques, include not only various clustering algorithms but also feature extraction and visualization techniques. This present approach is different from the work of [Dzwinel et. al., 2003]. In this paper we propose a novel muti-resolutional approach, which combines local clustering techniques in the data space with a non-hierarchical clustering in the feature space. We use multi-dimensional scaling procedures for visualization of multi-dimensional events in 3-D space. This visual analysis helps greatly in detecting of subtle structures, which escape the classical clustering techniques.

We used our methodology for analyzing the observed [Ito and Yoshioka, 2002; Toda et. al., 2002;] and synthetic [Ben-Zion, 1996] earthquake data. The observed data represent seismic activities of the Japanese islands in 1997-2003 time interval. The synthetic catalogs correspond to various cases of a large heterogeneous fault zone in elastic half-space.

The paper is constructed as follows. First, we present our knowledge of multi-resolutional and visual clustering and data-mining. Then we describe the clustering and multi-dimensional scaling methods we used for recognition of seismic anomalies. We discuss the results of data analysis both for the actual data set and for the synthetic data catalogs. We show that the multi-resolutional approach can indeed improve greatly the accuracy of earthquakes prediction. Finally, we discuss the conclusions and prospects.

2 Methodologies

2.1 Multi-resolutional analysis of multi-dimensional data

In Fig.1, we show that the seismic data can be analyzed in different resolutions associated with two types of spaces:

1. data space F of <u>data vectors</u> $\boldsymbol{f}_i$ (i=1,...,N) – which correspond to the data describing a single seismic event,
2. feature space W <u>of time events</u> $\boldsymbol{F}_j$ (j=1,...,M) and M<<N – an abstract space resulting from a non-linear transformation $\boldsymbol{L}$ [F]®W representing the feature generation procedure.

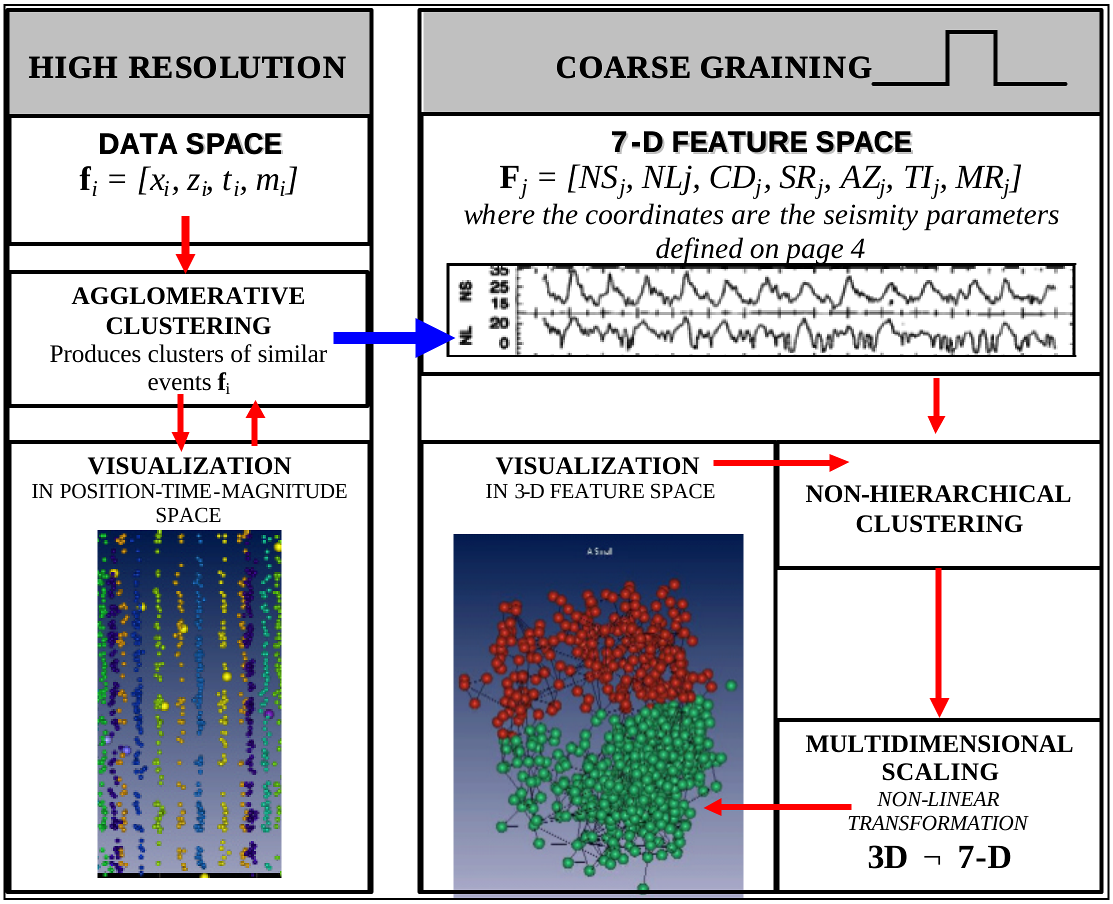

Fig.1 Schematic diagram of multi-resolutional analysis of seismic events

At the highest level of resolution, a single seismic event i can be represented as a multi-dimensional data vector $\boldsymbol{f}_i = [m_i, z_i, \mathbf{x}_i, t_i]$ where: m_i is the magnitude while $\mathbf{x}_i$, z_i, t_i – its epicentral coordinate, depth and the time of occurrence, respectively. As shown in [Dzwinel et. al., 2003], we can analyze these data locally by looking for clusters with similar (or dissimilar) events using the agglomerative clustering procedures [Andenberg 1973; Gowda and Krishna, 1978; Jain and Dubes 1988; Theodoris and Koutroumbas 1998]. The search for similar data is limited to successive time stripes $\mathbf{t}_k$ with the same width $\Delta\tau$. We are seeking for the neighbors of event i, only in $\mathbf{t}_k$ and the previous $\mathbf{t}_{k-1}$ time intervals. The point j belongs to the nearest neighbors of event i if some list of conditions is fulfilled (see [Dzwinel et. al., 2003]). This allows us to identify the correlated patches of events. They reflect clearly the short-term trends in seismic activity initiated by rapid changes in local activity generated by strong events. By combining similarity and dissimilarity measures [Theodoris and Koutroumbas 1998] between the data vectors, we can extract also the patches of small magnitude events corresponding to the isolated large earthquakes. This type of data analysis extracts information on the local properties of seismic patterns [Dzwinel et. al., 2003]. At the same time, this yields a large number of extraneous clusters, which produce unreliable information over a long timescale. We cannot also extract general knowledge about the data, which requires the detection of long-range spatial and temporal correlations.

This knowledge has to be extracted from a global data structure in a low resolution space W. We achieve this by using coarse graining procedure ***L***[F]. This averages out the noise and fine detailed modes of the data vector components.

The coordinates in the low resolution feature space are defined as seismicity parameters. Originally they are computed as time and space averages in a sliding time window with a length ΔT and time step dt, i.e.,:

$$\boldsymbol{a}_i = \int_{t_0-\Delta T/2} \int_X a(t, X) H(t_0 + i \cdot dt - \frac{\Delta T}{2}, t) dt dX \tag{1}$$

$$H(t, t) = \begin{cases} 1 & \text{for } -\frac{\Delta T}{2} < t - t < \frac{\Delta T}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where α = {NS, NL, CD, SR, AZ, TI, MR}. The seismicity parameters in α are defined as follows [Eneva and Ben-Zion, 1997a,b]:

1. Degree of spatial non-randomness at short (**NS**) and at long distances (**NL**) – represents the difference between distributions of event distances and distances between randomly distributed points. NS and NL represent the portions of events involved in anomalies in short distances and long distances, respectively (0-5 km and 60-65 km for the synthetic catalogs [Eneva and Ben-Zion, 1997]).
2. Spatial correlation dimension (**CD**) - calculated on the basis of correlation integrals based on interevent distances.
3. Degree of spatial repetitiveness (**SR**) – contains the spatio-temporal components and represents the tendency of events with similar magnitudes to have nearly the same locations of hypocenters.
4. Average depth (**AZ**)
5. Time interval for the occurrence of a constant number of events (**TI**) - represents time intervals for the occurrence of the 100 events in each group.
6. Ratio of two number of events falling into two different magnitude ranges **MR**=$N(m \geq M_0)/N(m<M_0)$.

We have introduced an additional parameter **M**, which is not used in data processing and which displays the maximum magnitude of events in the moving time window.

We focus our data analysis on the time series of seven seismicity parameters, which create the abstract 7-dimensional feature space of time events $\boldsymbol{F}_t = (NS_t, NL_t, CD_t, SR_t, AZ_t, TI_t, MR_t)$ where t are discretized moments of time. These events produce clusters, which correspond to similar (or dissimilar) fragments of 7-dimensional time series. Thus the clusters will have information about the anomalies reflected by all the 7 parameters in the same moment of time. Because the number of clusters is generally unknown, and most of the clustering methods are not able to extract the clusters of complicated shapes and densities accurately [Ertoz et. al., 2003], we propose to visualize the clustering structure in the feature space. We use multi-dimensional scaling transformation ***MS*** [W]® w, which maps the 7-D feature space W into its image in a 3-D space w [Jain and Dubes 1988; Siedlecki et. al., 1988; Theodoris and Koutroumbas 1998; Dzwinel, 1994; Dzwinel and Blasiak, 1999]. From the high-resolution 3-D visualization, one can discern clearly how strong the clusters are and how they are positioned with respect to each other. This will allow use to tune better clustering parameters or select different clustering algorithm that matches better the clustered structures.

2.2 Clustering schemes

Clustering analysis is a mathematical concept, whose main useful role is to extract the most similar (or dissimilar) separated sets of objects according to a given similarity (or dissimilarity) measure [Andenberg, 1973]. This concept has been used for many years in pattern recognition. Nowadays clustering and other feature extraction algorithms are recognized as important tools for revealing coherent features in the earth sciences [Rundle et., al., 1997, 2000, Freed and Lin, 2001] and in data mining [Xiaowei et. al., 1999, Grossman et. al., 2001, Hand et. al., 2001, Hastie et. al., 2001]. Depending on the data structures and goals of classification, different clustering schemes must be applied [Gowda and Krishna, 1978, Karypis and Kumar, 1999].

In this new approach we use two different classes of clustering algorithms for different resolution levels. In data space we use local agglomerative schemes, such as modified mutual nearest neighbor algorithm **mnn** [Gowda and Krishna, 1978; Karypis et. al., 1999; Boryczko et. al., 2003]. This type of clustering extracts better the localized clusters in the high resolution data space.

In the feature space we are searching for global clusters of time events comprising the similar events from the whole time interval. The non-hierarchical clustering algorithms are used mainly for extracting compact clusters by using global knowledge about the data structure. We use improved **k-means** based schemes [Theodoris and Koutroumbas, 1998], such as a suite of **moving** schemes [Ismail and Kamel, 1989], which uses the **k-menas** procedure plus four strategies of its tuning by moving the data vectors between clusters to obtain a more precise location of the minimum of the goal function:

$$J(w,z) = \sum_{j} \sum_{i \in C_j} \left| x_i - z_j \right|^2 \tag{3}$$

where $\mathbf{z}_j$ is the position of the center of mass of the cluster j, but $\mathbf{x}_i$ are the feature vectors closest to $\mathbf{z}_j$. To find a global minimum of function *J(),* we repeat many times the clustering procedures for different initial conditions. The new initial configuration is constructed in a special way from the previous results by using the methods from [Ismail and Kamel, 1989; Zhang and Boyle, 1991]. The cluster structure with the lowest minimum is selected.

2.3 Multi-dimensional scaling (MDS)

Multi-dimensional scaling (MDS) [Jain and Dubes 1988; Siedlecki et. al., 1988; Theodoris and Koutroumbas 1998; Dzwinel, 1994; Dzwinel and Blasiak, 1999] deals with the transformation **MS** of N-dimensional feature space to a space with a reduced dimensionality, i.e., $\mathbf{MS}(\Re^N) \rightarrow \Re^n$ and $n<<N$. In visualization we can assume that n=3. A non-linear mapping method is constructed on the basis of minimizing of the "stress" function criterion.

$$E = \sum_i \sum_j D_{ij}^{2 \cdot w \cdot m} \left(D_{ij}^2 - r_{ij}^2 \right)^m \tag{4}$$

The criterion means that the distances r_{ij} between the two transformed $\mathbf{MS}(\mathbf{F}_j)$ and $\mathbf{MS}(\mathbf{F}_i)$ time events in a lower dimension, e.g., in 3-D space, approximate the respective distances D_{ij} between $\boldsymbol{F}_i$ and $\boldsymbol{F}_j$ in the original 7-dimensional space in the sense of minimum squared error. The parameters w,m (m>1;$w \in \{-1,0,1\}$) represent the error function parameters, which allows for different views of the data in the reduced space.

3 Description of the data

We data-mine and analyze observed and synthetic earthquake catalogs in different time intervals. The observed data represents (Fig.2a) seismic activities of the Japanese islands collected by the Japan Meteorological Agency (JMA). The seismic events shown in Fig.2a, were detected during 5 years time interval from 1997 to 2003. The data set consists of 4×10^4 seismic events with magnitudes m, position in space (latitude X, longitude Y, depth z) and occurrence time t. The lowest magnitudes were determined by using a detection level, estimated from the Gutenberg-Richter frequency-size disribution. Thus the statistical completeness of the earthquakes above the detection level assures that there is no significant lack of events in space and time.

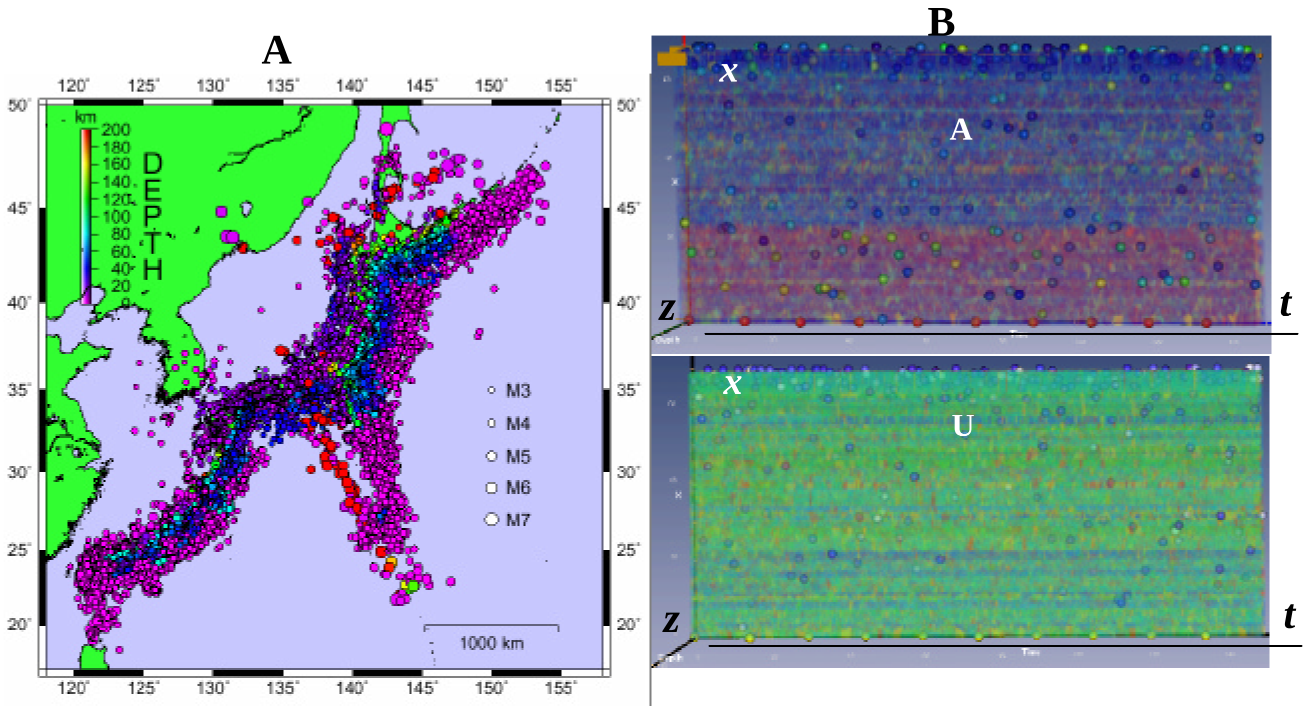

Fig.2 A) Seismic activities around the Japanese Archipelago. We use the hypocentral data provided by the Japan Meteorological Agency (JMA). The magnitude of the earthquakes (JMA magnitude) and their depth are represented by differences of the radius of the circle and colors, respectively. B) The synthetic 2-D raw data visualized by using the Amira visualization package [www.amiravis.com] for A and U data sets. Large events are shown as spheres on the background of smaller events.

The synthetic catalogs (Fig.2b) were obtained by numerical simulations of seismicity on a heterogeneous fault governed by 3-D elastic dislocation theory, power-law creep and boundary conditions corresponding to the central San Andreas Fault [Ben-Zion, 1996]. The model yields the earthquake distribution in space (horizontal distance X, depth z), time t and magnitude m size. A large 1857-type event is imposed at the south in the beginning of the simulation and a large 1906-type event is imposed at the north 50 years later. We study catalogs from four different model realizations representing various levels of fault zone disorder. These are models with statistically uniform brittle properties (U), with a Parkfield type Asperity (A), with fractal brittle properties (F), and with multi-size-heterogeneities (M). These models and various statistical properties of the catalogues have been discussed in greater detail elsewhere [Ben-Zion, 1996, Eneva and Ben-Zion, 1997]. The time interval covers every event, which occurred during the last 150 years of simulated fault activity and this period contains from 1-3×10^4 events with magnitudes from [3.3-6.8] interval.

The seismicity parameters were obtained by averaging the data using the sliding time window of the constant width ΔT and shift dt (see Eqs.1-2): ΔT= 10 days and dt= 2 days for Japanese data and ΔT= 10 months and dt= 2 months for synthetic data. Each parameter in the clustering was normalized to the standard deviation.

4 Results of Clustering

Clustering results of the observed Japanese seismic catalogs both in raw data and in feature spaces are shown in Fig.3. At the highest resolution level a single seismic event i can be represented as a multi-dimensional data vector $\boldsymbol{f}_i = [m_i, z_i, X_i, Y_i, t_i]$ where: m_i is the magnitude, X_i - the latitude, Y_i - the longitude,

z_i and t_i –the depth and the time of occurrence, respectively. The seismic events are visualized with the Amira package in Figs.3a,b as an irregular cloud consisted of colored dots with (z,x,t) coordinates.

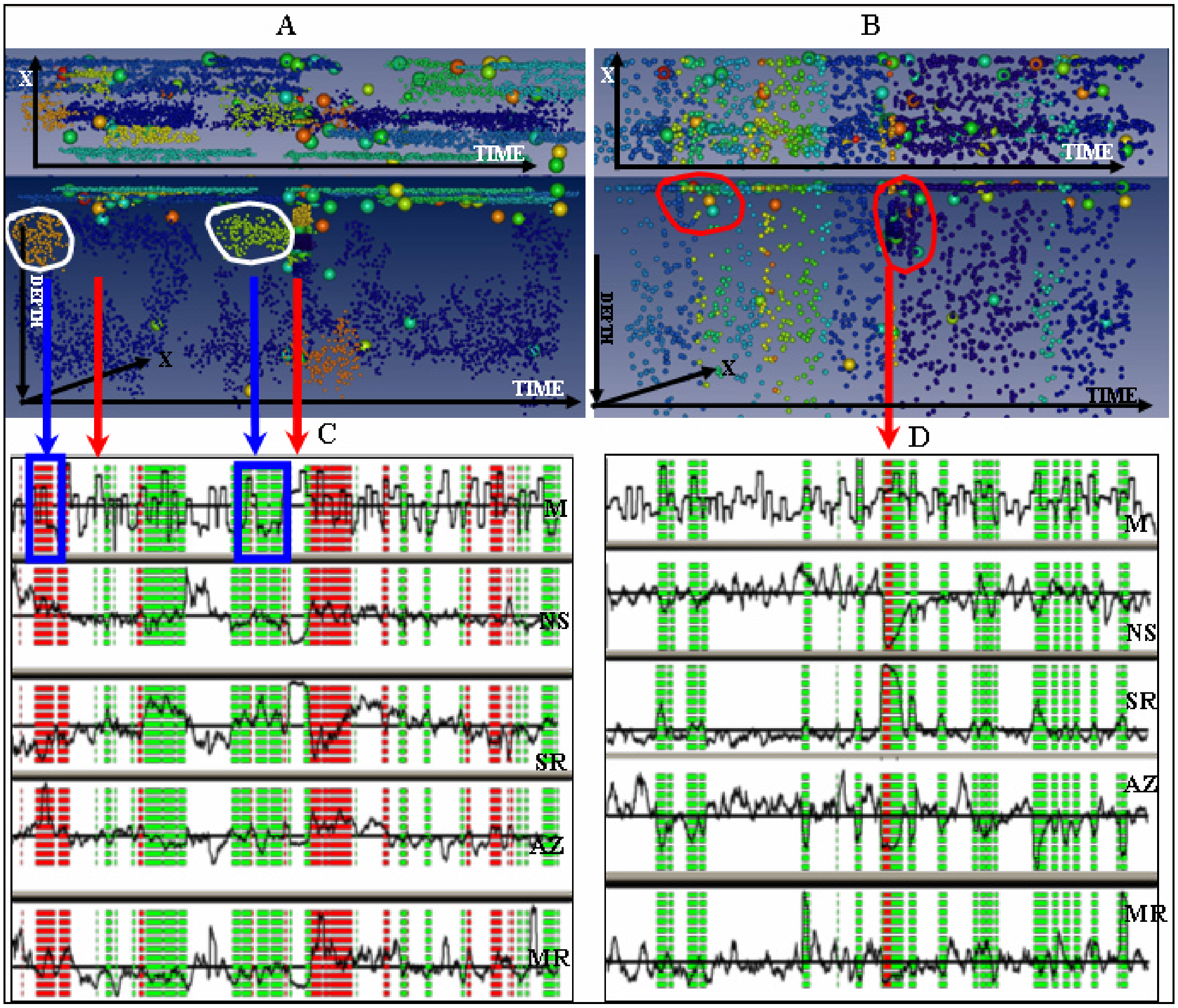

Fig.3 Actual seismic data analyzed by using multi-resolutional clustering in both the data and the feature spaces (A-C). In panels A and B one can see the results of clustering in the data space for small magnitude (3<m<4) and medium magnitude (4<m<6) events, respectively, represented by the small shaded dots. The different colors of the dots mean different clusters. Large events are visualized by the larger spheres. The shades show difference in magnitudes *m* (red – the largest, green - the smallest). The clusters in panels A-B encircled in red shows the places of the largest seismic activity, while those in white represent the clusters of precursory events. The red, white and green stripes in panels C-D representing 5 (out of 7) seismic parameters show the time events belonging to three different clusters. In panel D we show the clusters of seismic parameters obtained only from clustering the time events in the feature space. From panel C (as compared to D) we see that by combining clustering results in data space with clustering in the feature space, we can extract anomalies, preceding the two earthquake swarms (encircled in red in panel B).

From the Gutenberg-Richter self-similarity power-law relationship, the number of events of different range of magnitudes differs considerably. Therefore, we arbitrarily divide the entire set of data into three subsets comprising the small $\mathbf{S}(m_i<ma)$, medium $\mathbf{M}(ma\leq m_i<max)$ and the large magnitude events $\mathbf{L}(m_i>max)$. The last ones represent the earthquakes and are displayed in Figs.3a,b as larger spheres. The deepest earthquakes z>150 km are not displayed in the figure. The various shades represent the magnitudes of earthquakes from m=6 (green) to m=7 (red). In Figs.3a,b we present the clustering results in the data space F of the data vectors $\boldsymbol{f}_i \in \mathbf{S}(m_i<ma)$ (Fig.3a) and $\boldsymbol{f}_i \in \mathbf{M}(ma\leq m_i<max)$ (Fig.3b). We have

chosen arbitrarily that *ma=4* and *max*=6.We look for clusters of similar events as shown in [Dzwinel et. al., 2003]. The dots (data vectors), belonging to the same clusters, have the same color.

As shown in the upper part of Fig.3a, clusters made of small size events are located mainly close to the surface (0-30 km deep). They form long disparate stripes along the time axis. The stripes break-out close to the largest cluster of earthquakes – the Miyakejima event [Toda et. al., 2002; Ito and Yoshioka, 2002] - located in the middle of time interval and encircled in red in Fig.3b. The large swarm of earthquakes (June 26, 2000) occurs in the region of Miyakejima, Honshu, in central Japan [Toda et. al., 2002; Ito and Yoshioka, 2002]. The eruptions of Miyakejima and five large earthquakes with magnitudes 6.0 and above occurred together with a large number of 100,000 smaller earthquakes.

Many other compact and small clusters are strongly correlated with this cluster. The second patch of large events encircled in red in Fig.3 – representing large swarm of earthquakes in the northern part of Japan (see Fig.2a) - is shown on the left hand side of the largest one. There are clearly seen two clusters (encircled in white in Fig.3a) of similar depth preceding these two largest patches of earthquakes. Other clusters, such as the wide one spanned by a blue cluster and another smaller one, are located much deeper (larger than 100 km) and represent the seismic background for the earthquakes occurring at depth up to 200 km.

Clusters of the medium events have completely different structures. They look like bricks, which lie parallel to X-z plane. The borders between clusters roughly correspond to the borders of successive showers of the earthquakes. In Fig.3d we display 5 out of 7 time series of seismicity parameters. Due to clustering of the time events in the feature space and simultaneous inspection of the results by using multi-dimensional scaling we found 4 distinct clusters. The cluster structure in 7-D feature space reflects Fig.4a, which represents the result of its mapping into 3-D. The blue cluster forming a long thin rod in Fig.4a, corresponds to the famous Miyakejima earthquake swarm [Toda et. al., 2002] encircled in red on Fig.3b and marked by the red stripe in Fig.3d. The red and flat clusters from Fig.4a, represent both the largest earthquakes and corresponding time events, marked in Fig.3d as green. The yellow and blue clusters contain the rest of the time events. In summary, clustering of averaged time events does not detect any anomalies reflecting the precursory patterns. A cluster examination of Fig.4a shows that there exist no anomalies in the feature space. The small blue cluster from Fig.4a represents the events at the end of the time interval, which are averaged within a shrinking time window.

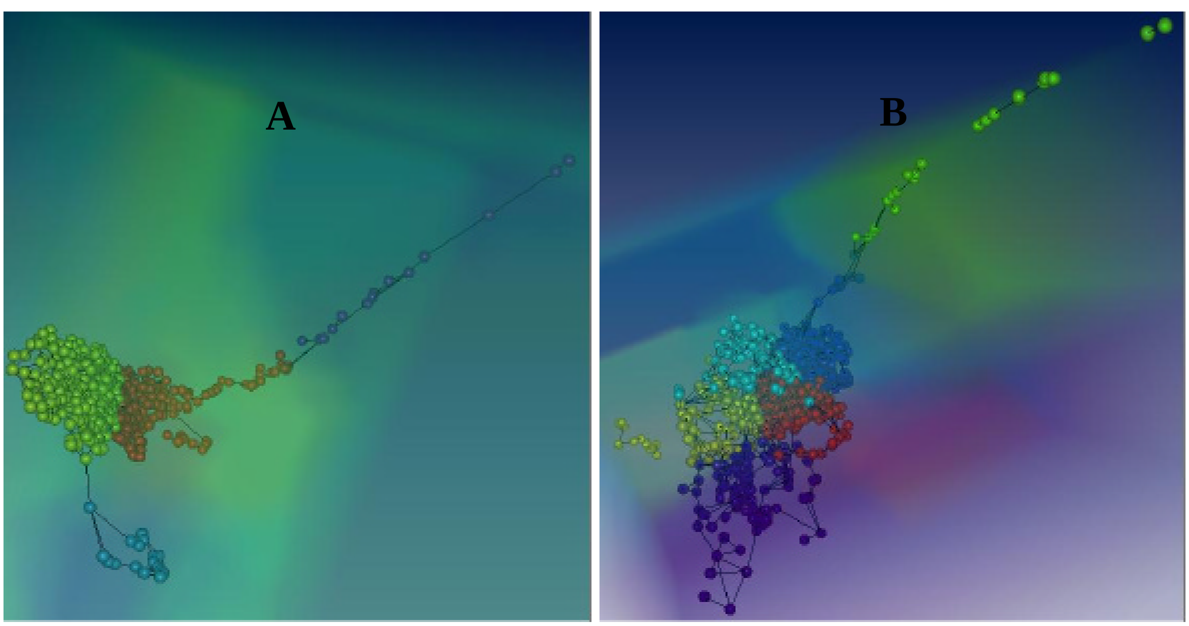

Fig.4 The results of non-linear multi-dimensional scaling of 7-D clusters from the feature spaces defined by the seismicity parameters from Fig.3D and Fig.3C, respectively, into a 3-D space.

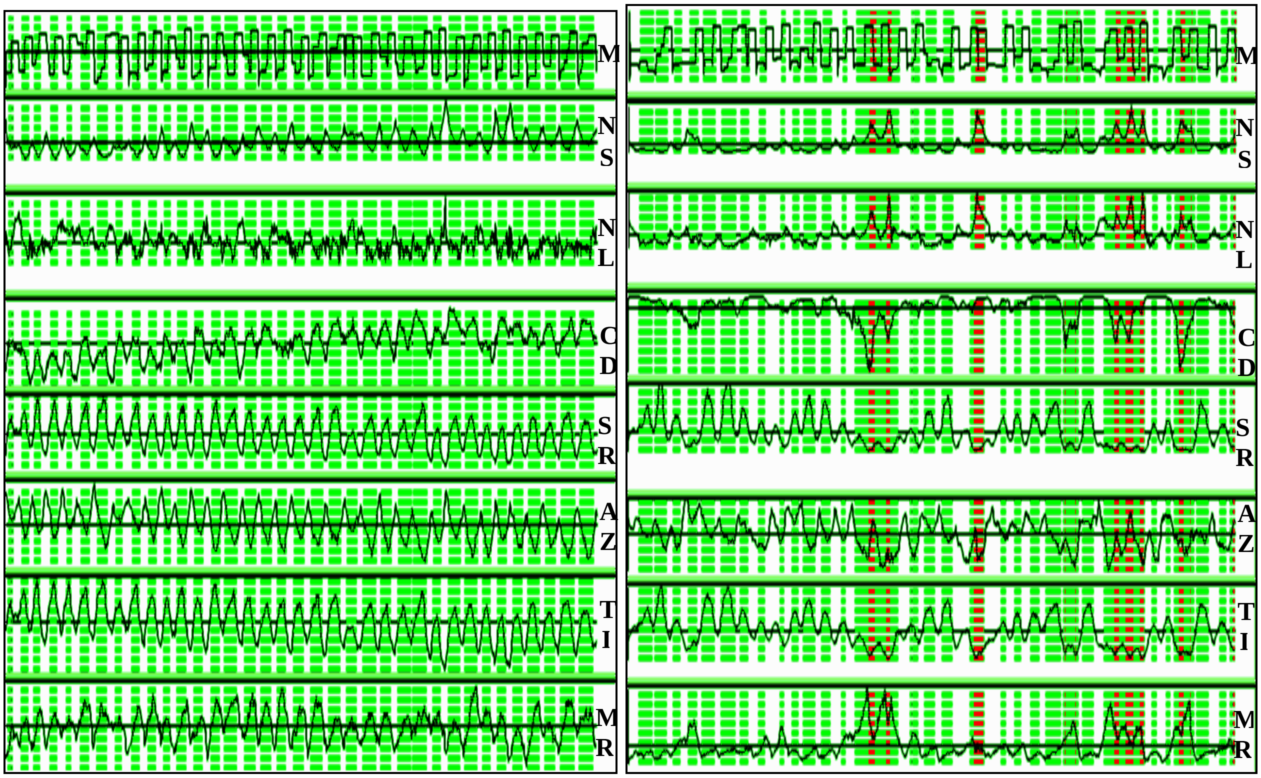

Fig.5 Temporal evolution of the seismicity parameters {M,NS,NL,CD,SR,AZ,TI,MR} (see Eqs.1-2) for synthetic data catalog A. The green stripes from figure A depict the events belonging to the green cluster from Fig.6 while the white ones to the red cluster from Fig.6. Fig.B displays the seismicity parameters computed for the same data but filtered out by using the results of clustering in the data space. In Fig.B, a new red cluster emerges.

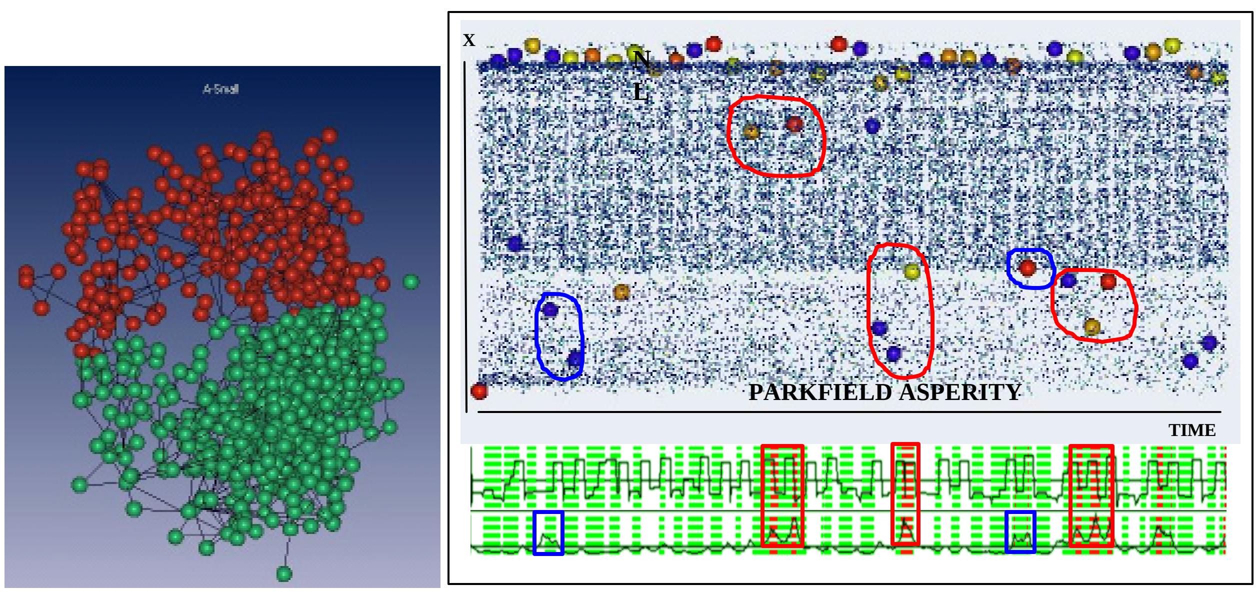

Fig.6. A) Results of non-linear, multi-dimensional scaling of 7-D clusters from the feature spaces into 3-D space. The seismicity parameters computed for the complete synthetic data catalog (catalog A). B) The clusters of the largest events, corresponding to the time events from Fig.5B, marked in red. The seismicity parameters have been computed fromthe events, which have been selected by clustering in the data space.

In Fig.3c we display the seismicity time series and in Fig.4b the 3-D image of the feature space for the data pre-selected initially by clustering in the data space. Only the largest clusters of small and medium events were used for computing the seismicity parameters. As shown in Fig.4a, the feature space became more diverse, producing several distinct clusters. In Fig.3c, we show that, by clustering in the feature space, we can extract not only the largest events but also their precursory clusters composed of small events.

A similar procedure was carried out on the synthetic data. In Fig.5a we display the seismicity parameters in time computed for the complete synthetic data catalog A. As shown in Fig.6a, the time events produce in the feature space 2 clusters. The green cluster from Fig.6a, corresponds to the time events marked in green in Fig.5a. From the top plot of Fig.5a representing the largest events in the sliding time window, we may conclude that green cluster comprises the time events, preceding the largest events. We pre-select data events by considering only these creating clusters in the data space. As a result, in the feature space we can locate more precisely in time the precursory zones for the most of earthquakes. The zones are displayed by the green stripes in Fig.5. Moreover, as shown in Fig.6b, the cluster structure of time events in the feature space of seismicity parameters reflects the changes in the spatial localization of the largest earthquakes. The events, occurring closer to the Parkfield Asperity boundary, produce a separate cluster.

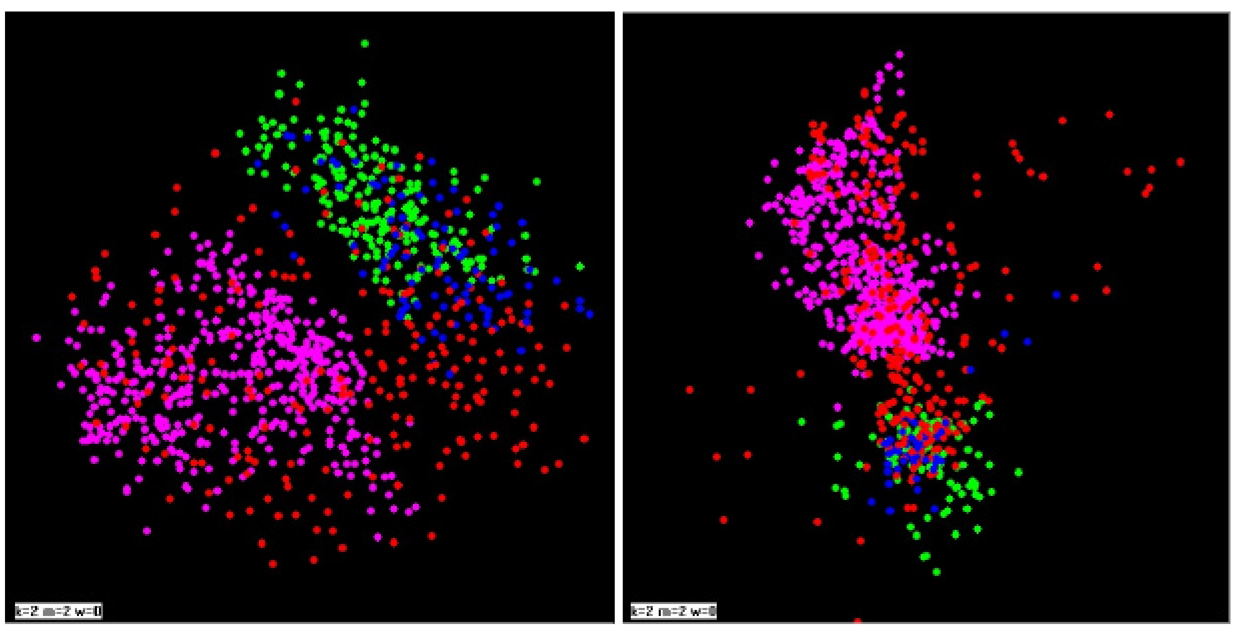

Fig.7 Results from visual classification of the synthetic data (from catalogs A and F, respectively) in the feature space transformed into 3-D space by using multi-dimensional scaling (MDS). The time events are divided onto two groups of data $\{F_i;\ i<T\}$ (violet and green) and $\{F_i;\ i>T\}$ (red and blue) representing: teaching and test sets, respectively. The classifier is constructed using teaching set of time events, which consists of two groups of events: preceding the earthquakes (green) and the others (pink). The preceding time events were taken arbitrarily as the 10 nearest events to the corresponding earthquake. In the result of training we obtain two clusters: green and violet. The same two groups of data are marked in the test set (preceding the earthquakes – blue, and the others – red). We see that all the blue points were attracted to the green cluster - taught as the cluster consisting of “precursory” events - so all of them were recognized as “precursory”.

Because the precursory events produce distinct clusters in both Fig.5a and Fig.5b, we can construct a simple visual classifier [Jain and Dubes 1988; Theodoris and Koutroumbas 1998] for recognizing of the anonymous precursory patterns. The entire time interval in the feature space has been divided into two parts. The events from the first 2/3 of the interval - approximately 700 events - represent the teaching set, the rest - about 300 events – make up the test set. From the teaching set we extract two uneven groups of events. The first one consisting of 10 successive time events preceding each of the earthquake (m>6) is shown in green in Fig.7. The rest of the events represent the second (pink) cluster from Fig.7. The clusters are visualized by using multi-dimensional scaling in 3-D. Then, the distances (see Eqs.3-4) between events belonging to the same clusters was multiplied by the factor $\lambda<1$, while the distances between events from different clusters remain the same. The vale of λ is gradually decreasing to the moment when the two clusters separates one from the another. For the situations shown in Fig.7 λ is set to 0.8. The test events from the test data are added to this teaching structure, with the determined coordinates in 3-D w space. They are "attracted" or "repelled" from the teaching clusters according to their distances to the events from the teaching set. Eventually, we obtain the situations shown in Fig.7. The precursory events from the test set marked in blue (10 events preceding the earthquake) were recognized at 100% level for synthetic data sets A, U, F. The blue points fall into the area occupied by the green cluster representing precursory events from the teaching set. We see that many other points are also situated in the area of a green cluster. However, the choice of precursory events for training was completely arbitrary. By using a more careful analysis of seismicity patterns, such as in [Eneva and Ben-Zion, 1997a], and better selection of teaching patterns, we can improve the classification.

4 Conclusions

Earthquake prediction from extracting the precursory phenomena is a very tough task. Various computational methods and tools are used for detection of the precursors and for generalization of the information. In our opinion, such a generalization is impossible, when one analyses only the data at a single level of resolution. The set of data-mining tools has to be used as a single interactive system, allowing for on-line clustering, feature extraction and visualization of the data on various levels of resolution.

We show that by using a common framework of clustering, we are able to perform multi-resolutional analysis of seismic data starting from the raw data events described only by their magnitude-spatio-temporal data space. Then we look for global cluster structure in the feature space, which is defined by using the seismicity parameters. This global view can also be divided over different levels of resolution in the feature space defined, e.g., by wavelet analysis [Holschneider, 1995; Strang and Nguyen, 1996] of the time series of the seismicity parameters [Torrence and Compo, 1998] and further classification of wavelet amplitudes by using clustering schemes.

The approach presented here is different than that of [Dzwinel et. al., 2003]. The fine-grained spatio-temporal patterns of correlated events, extracted by using agglomerative clustering schemes, can be analyzed further in the coarse grained feature space by eliminating the noisy patterns and uncorrelated events. We have developed a new software that is based on pre-clustering. This allows for the detection of precursory events with a higher accuracy (e.g., pick-up the Miyakejima event) and their generalization at the low resolution level. It also allows for constructing visual classifiers for anonymous data. We believe that a more careful extraction of the precursory events allows us to construct more accurate classifiers.

The raw seismic data contains both local and global knowledge about correlations between the seismic events. Thus, the two-level approach presented here can still be incomplete, bearing in mind that the general knowledge about the seismic background can be buried in subtle patterns of data events. Any coarse graining of the data can destroy some, if not a majority, of these patterns. Therefore, extracting global knowledge about seismic patterns corresponding to precursory events involves global clustering of data without any averaging. This is a very daunting task both methodologically and computationally. This is due to both irregular structure of seismic data, which comprise many noisy events, different accuracy of measurements, outliers, bridges, clusters of different density and the large number of data vectors (10^4+) which have to be processed. This problem can be attacked by using modern non-hierarchical clustering schemes, such as the DBSCAN [Sander et. al., 1998], CURE [Guha and Rastogi, 1998], Chameleon [Karypis, 1999] or the shared nearest neighbor clustering algorithm (SNNCA) [Ertoz, 2003].

This new methodology can be also used for the analysis of the data from other geological phenomena, e.g, we can apply this clustering method to volcanic eruptions [Amelung et. al., 2000], and astrophysical events such as dissipation phenomena, occurring in a dispersed stellar population [Briceno et.,al., 2001].

Acknowledgments

We acknowledge the Japan Meteorological Agency (JMA), which allowed us to use their hypocentral data set. Partial support for this work has come from the Polish Committee for Scientific Research (KBN) project. Support from the NSF geophysics program is acknowledged. We thank discussions with George Karypis, Y.J.B.D. Kaneko, Zack Garbow, Ben Kadlec, Evan Bolling and Gordon Erlebacher.

References

1. Amelung F., Jonsson S., Zebker H. and Segall P. Widespread uplift and 'trapdoor" faulting on Galapagos volcanoes observed with radar interferometry, *Nature*, 407, 993-996, 2000.
2. Andenberg MR, *Clusters Analysis for Applications,* New York, Academic Press, 1973.
3. Ben-Zion, Y. and J. R. Rice. Slip patterns and earthquake populations along different classes of faults in elastic solids, *J. Geophys. Res.*, **100**, 12,959-12,983, 1995.
4. Ben-Zion, Y., Stress, slip and earthquakes in models of complex single-fault systems incorporating brittle and creep deformations, *J. Geophys. Res*., **101**, 5677-5706, 1996.
5. Ben-Zion Y, Lyakhovsky V, Accelerated seismic release and related aspects of seismicity patterns on earthquake faults , *Pure and Applied Geophysics*, 159/10, 2002.
6. Boryczko K, Dzwinel W, Yuen DA, Clustering Revealed in High-Resolution Simulations and Visualization of Multi-Resolution Features in Fluid-Particle Models, *Concurrency and Computation: Practice and Experience*, 15, 101-116, 2003
7. Briceno C, Vivas AK, Calvet N, Hartmann L, Pacheco R, Herrera D, Romero L, Berlind P, Sanchez G, Snyder J.A., Andrews P, The CIDA-QUEST Large-Scale Survey of Orion OB1: Evidence for Rapid Disk Dissipation in a Dispersed Stellar Population, *Science*, 291, 93-96, 2001.
8. Dowla, F. U., Neural networks in seismic discrimination, in Monitoring a CTBT, Husebye E. S., Dainty, A.S. (eds), NATO-ASI series E 303, Kluwer Pubblishing, 777-789, 1995
9. Dzwinel W, How to Make Sammon's Mapping Useful for Multi-dimensional Data Structures Analysis?, *Pattern Recognition*, 27/7, 949-959, 1994.
10. Dzwinel W, Blasiak J, Method of particles in visual clustering of multi-dimensional and large data sets, *Future Generation Computers Systems*, 15, 365-379, 1999.
11. Dzwinel W, Yuen D.A., Kaneko Y, Boryczko K, Ben-Zion Y, Multi-Resolution Clustering Analysis and 3-D Visualization of Multitudinous Synthetic Earthquakes, *Visual Geosciences*, 8, 12-25, 2003 (e-paper) the original publication is available at http://link.springer.de/link/service/journals/10069/contents/tfirst.htm
12. Eneva, M., Ben-Zion, Y., Techniques and parameters to analyze seismicity patterns associated with large earthquakes, *Journal of Geophysical Research,* 102/B8, 785-795, 1997a
13. Eneva, M., Ben-Zion, Y., Application of pattern recognition techniques to earthquake catalogs generated by model of segmented fault systems in three-dimensional elastic solids. *Journal of Geophysical Research,* 102/B11, 513-528, 1997b
14. Enescu B, Ito K., Struzik Z., (Multi)fractality of Earthquakes by use of Wavelet Analysis. *Proceedings of American Geophysical Union Fall Meeting,* 6-10 December 2002, San Francisco, California, 2002.
15. Erlebacher, G., Yuen, D.A. 2003. A Wavelet Toolkit for Visualization and Analysis of Large Data Sets In Earthquake Research, *Pure and Applied Geophysics* in press 2003.

16. Erlebacher, G., Yuen, D. A., and Dubuffet, F.W., Current trends and demands in visualization in the geosciences, *Electronic Geosciences*, 4, 2001, http://link.springer- ny.com/link/service/journals/10069/technic/erlebach/index.htm.
17. Ertoz L, Steinbach M., Kumar V, Finding Clusters of Different Size, Shapes and Densities in Noisy, High-Dimensional Data, Army High Performance Center, technical report, April 2003
18. Freed A.M., Lin J, Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer, *Nature*, 411, 180-183, 2001.
19. Geller J.R., Jackson D.D., Kagan Y.Y., Mulargia Y.F., Earthquakes Cannot Be Predicted, *Science Online*, 275 (5306):1616, 1996.
20. Gowda C.K., Krishna G, Agglomerative clustering using the concept of nearest neighborhood, *Pattern Recogn.*, 10, 105, 1978.
21. Grossman R., L., Karnath, Ch, Kegelmeyer, P., Kumar, V., Namburu, R.,R., *Data Mining for Scientific and Engineering Applications*, Kluwer Academic Publisher, 2001
22. Guha S., Rastogi R., Cure A., An efficient clustering algorithm for large data bases, *SIGMOD 1998, ACM SIGMOD International Conference on Management of Data, Washington, USA*, 73-84.
23. Hand, D., Mannila, H. and Smyth P., Principles of Data Mining, M.I.T. Press, Cambridge, Ma, 2001.
24. Hastie, T., Tibshirani, R. and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference and Prediction, 533 pp., Springer Verlag, New York, 2001
25. Hockney RW, Eastwood JW, *Computer Simulation Using Particles*, McGraw-Hill Inc. 1981.
26. Holschneider M., *Wavelets: An Analysis Tool*, Oxford Science Publications, Clarendon Press, Oxford, 1995.
27. Ito T., Yoshioka, S., A dike intrusion model in and around Miyakejima, Niijima and Kozushima. *Tectonophysics* 359: 171– 187, 2002
28. Ismail M.A, Kamel M.S., Multi-dimensional data clustering utilizing hybrid search strategies, *Pattern Recogntion*, 22/1, 77-89, 1989
29. Jain D, Dubes R C., *Algorithms for Clustering Data*. Prentice-Hall Advanced Reference Series, 1988.
30. Joswig, M., Pattern recognition for earthquake detection. *Bull. Seism. Soc. Am.* 80, 170-186, 1990
31. Karypis G., Han, E., Kumar V., Chameleon: A hierarchical clustering algorithms using dynamic modeling, *IEEE Computer*, 32(8), 68-75, August 1999.
32. Karypis G, Kumar V, Multilevel algorithms for multi-constraint graph partitioning, technical report, University of Minnesota Supercomputing Institute, Minneapolis, MN, 1998.
33. Miller S.A., Ben-Zion Y., Burg J.P., A three-dimensional fluid-controlled earthquake model: behavior and implications, *J. Geoph. Res.*, 104/B5:10621-638, 1999
34. Rundle J.B., Klein W, Tiampo K, Gross S, Linear pattern dynamics in nonlinear threshold systems, *Phys. Rev. E*, 61/3, 2418-2143, 2000.
35. Rundle J.B., Gross S, Klein W, Ferguson C, Turcotte DL., The statistical mechanics of earthquakes, *Tectonophysics*, 277, 147- 164, 1997.
36. Rundle P.B., Rundle J.B., Tiampo, K.F. Sá Martins J.S., McGinnis S., and Klein W., Network dynamics of earthquake fault systems, *Phys. Rev. Lett.* 87, 148501(1-4), 2001
37. Sander J., Ester M., Krieger H, Density based clustering in spatial databases, The algorithm DBSCAN and its applications, *Data Mining and Knowledge Discovery*, 2(2), 169-194, 1998.
38. Siedlecki W., Siedlecka K., Sklanski, J, An overview of mapping for exploratory pattern analysis, *Pattern Recogn.*, 21/5: 411-430, 1988
39. Song, T. A. and Simons M., Large trench-parallel gravity variations predict seismogenic behavior in subduction zones, *Science*, 301, 630-633, 2003.
40. Strang, G. and Nguyen, T., *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
41. Theodoris S and Koutroumbas K, *Pattern Recognition*, Academic Press, San Diego, London, Boston, 1998.

42. Tiira T., Detecting teleseismic events using artificial neural networks, *Computers and Geosciences* 25, 929-939, 1999
43. Toda S., Stein R.S., Sagiya T., Evidence from the AD 2000 Izu Islands earthquake swarm that stressing rate governs seismicity, *Nature*, 419: 58-61, 2002
44. Wang W., Gengfeng W., A New Generation Expert System for Earthquake Prediction based on Fuzzy Neural Networks and Symbols, Selected Papers on Earthquake Precursor and Prediction in China, *Seismological Press*, 222-231, 1996
45. Wesnousky, S. G., The Gutenberg-Richter or characteristic earthquake distribution, which is it?, *Bull. Seismol. Soc. Amer.*, 84, 1940-1959, 1994
46. Xiaowei Xu, Jager J., Kriegel H-P, A fast parallel clustering algorithm for large spatial databases, *Data Mining and Knowledge Discovery*, 3/3, 263-290, 1999
47. Zhang Q. and Boyle R., A new clustering algorithm with multiple runs of iterative procedures, *Pattern Recognition*, 24/9, 835-848, 1991