



Identification and spatio-temporal analysis of earthquake clusters using SOM-DBSCAN model

Ashish Sharma¹ · Rahul Kumar Vijay² · Satyasai Jagannath Nanda¹

Received: 1 April 2022 / Accepted: 22 November 2022 / Published online: 8 December 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd, part of Springer Nature 2022

Abstract

Seismic catalogs are vital to understanding and analyzing the progress of active fault systems. The background seismicity rate in a seismic catalog, strongly associated with stressing rate, is the critical parameter in seismic hazard analysis. Estimating background seismicity is a complex task due to the high correlation with aftershock sequences which may dominate the background seismicity rate. In this paper, identification of the significant earthquake aftershocks and independent background events is performed using a two-stage clustering approach. It works in two phases: Self-Organized Map and Density-based Temporal Clustering. The event's location and depth information in the earthquake catalog is used to identify the major hot spots (SOM prototypes) in the region (Spatial domain). Later, density-based temporal clustering is applied to decipher the neighborhood events of each SOM prototype. The proposed two-level clustering approach performs effective spatio-temporal analysis and identifies the aftershock clusters and background. The experimental study is carried out on the prominent earthquake catalogs of Taiwan, Afghanistan, California, the Himalayas, Indonesia, Chile, and Japan. The statistical parameters: Coefficient of Variation (time-domain) and m -Morisita index (spatial domain) justify and validate the accuracy of the presented approach. The proposed model is compared with benchmark de-clustering algorithms for mainshock and background detection.

Keywords Earthquake clusters · Self-organized map · DBSCAN algorithm · Coefficient of variance · m -Morisita index

1 Introduction

Earthquakes are one of the most dreadful natural catastrophes known to humanity. It is essential to understand earthquakes' complexity, a pattern of occurrence, and behavior for successful hazard mitigation. Earthquakes are

closely connected to various clusters in the spatio-temporal domain with different sizes that generate non-stationary complex patterns of events [1].

Any earthquake that occurs near (in space and time) and after a larger earthquake (mainshock) is called aftershocks (AFs). The seismic study and analysis mainly deal in aftershock-mainshock-foreshocks clusters rather than regular or background (BGs) events. The seismic clusters are AFs sequences comprising of numerous smaller events following the higher magnitude event (mainshock) in the space-time domain [2–4]. Foreshocks are a type of cluster comprised of pre-mainshocks events associated with a smaller number and smaller areas than those for AFs [5]. Swarms are another type of seismic cluster that contains a group of neighboring earthquakes with approximate size. They are mostly related to the fluid flow and aseismic deformation [6, 7].

Earthquake clusters are frequently generated classes of events triggered within a short time interval. It leads to additional redistribution of stress in the earth's crust that causes the development of correlated AFs. The clustering

Rahul Kumar Vijay and Satyasai Jagannath Nanda have contributed equally to this work.

✉ Ashish Sharma
2018rec9054@mnit.ac.in

Rahul Kumar Vijay
rahulvijay@banasthali.in

Satyasai Jagannath Nanda
sjnanda.ece@mnit.ac.in

¹ Department of Electronics and Communication Engineering, Malaviya National Institute of Technology Jaipur, Jaipur, Rajasthan 302017, India

² Department of Computer Science, Banasthali Vidyapith, Tonk, Rajasthan 304022, India

properties are determined by examining the distribution of events in space and time (spatio-temporal domain), and relevant information is extracted through comprehensive statistical analysis. The discrimination of short-term clustered events, such as AFs, from the BGs events, is known as seismicity declustering. Earthquake catalogs are frequently used to understand seismicity patterns, generate source models, quantitatively analyze source parameters and create empirical models of ground motions. In many applications, a declustered earthquake catalog is required that does not include AFs, foreshocks, swarms, and other forms of clustered seismic patterns. The various kinds of studies which need declustered earthquake catalog include understanding the seismicity occurrences [8], seismic hazard estimation [9], detection of seasonal, tidal and climate triggering of seismicity [10, 11], focal mechanism inversion for background stress [12].

Recently, Dawood et al. [13] specified the importance of seismic declustering for the generation of ground motion prediction equations. Schaefer et al. [14] indicated that declustering is a necessary step for statistical analysis of spatial-temporal properties of seismic events. This study reveals that seismic declustering is vital for seismic studies and probabilistic hazard assessment. The past few decades have witnessed an upsurge in the use of several declustering techniques to cluster seismic data. However, there is no standard technique or rule to identify earthquake clusters. Accordingly, declustered catalogs significantly differ due to subjective assumptions in the declustering techniques. Over the years, declustering seismic activity data has always been the focus of careful research. The very first attempt to separate the earthquake catalog into aftershock-foreshocks and BGs was made by Gardner and Knopoff [8]. They proposed a declustering method based on a space-time window (where the space-time distance criterion is a function of the mainshock magnitude). This approach needs the optimum choice of window parameters in the spatio-temporal domain, which are difficult to determine in a clustering framework. The drawback of this approach is that the seismic event generated by the secondary triggering event was not considered during the analysis. Gardner and Knopoff hypothesized that events in a properly declustered catalog follow a stationary Poisson process. Reasenberg [15] proposed a method for AFs identification based on the cluster-link technique in which AFs are observed as per their spatial and temporal interaction zones. They also determined that the BGs event is distributed as a space Poisson process. These two methods defined the standard of seismicity declustering (and related quality statistics) for many years and are still used in most

of the cases. Later, the same authors, Reasenberg and Jones [16, 17] modeled the AFs events as non-homogeneous Poisson process in time, based on Omori empirical formula, and also estimated the probabilities of AFs occurrence.

Tibi et al. [18] proposed the modified cluster-link technique by replacing the Poisson model-based time link by a magnitude-dependent temporal window and eliminating the need for completeness of catalog. The clustering feature of the earthquakes attracted many researchers to investigate and develop deterministic and stochastic model for seismicity pattern analysis like Epidemic Type Aftershock-Sequences (ETAS) model [19–21] and branching aftershock sequence (BASS) model by Turcotte et al. [22] and Holliday et al. [23], Tri-stage clustering model by Nanda et al. [24] and Tetra- stage clustering identification model by Vijay et al. [25].

Zhuang et al. [26] used space-time parametric maximum likelihood estimate using ETAS model for the declustering and variable weighted kernel estimation for the BGs seismicity. Zaliapin et al. [27] used space-time-magnitude nearest-neighbor distance to identify the non-parametric AFs and cluster analysis of seismicity in the time-space-energy domain. Bottiglieri et al. [28] introduced a methodology for evaluating variability coefficient to discriminate between Poissonian independent earthquakes and AFs (based on interevent time and interevent distances). This method is efficient for AFs selection and allows a reasonable examination of the statistical properties of earthquake clustering.

Batac and Kantz [29] used a very robust approach to separate spatio-temporal clusters of earthquakes using pairwise distances and time without any prior assumptions regarding earthquake relationships and regional factors. This approach is strongly dependent on the magnitude of completeness of the catalogs. Later, Cho et al. [30] used Thirumalai metric and Gutenberg–Richter distribution as constraints for earthquake fault systems in the space-time region. They have shown that effective ergodic periods are disrupted due to seismic clustering. Davidsen et al. [31] analyzed the decay of AFs activity with time using the modified Omori formula.

Recent research work in the field of seismicity analysis provided the breakthrough by modeling the phenomenon in the spatio-temporal domain with machine learning approaches such as K-means algorithm [32] in which source of seismic events were identified, Fuzzy-based method [33], and Nearest-Neighbor approach [34], etc. Nanda et al. [35] proposed a stochastic partitional algorithm using Pearson correlation coefficient for cluster

analysis. Morales et al. [36] proposed an efficient algorithm for the identification of globally optimal partition using the adaptive Mahalanobis distance function. The separation of correlated AFs and BGs events is also modeled as an optimization framework and solved effectively using meta-heuristics-based algorithms. Nanda et al. [37] compared all nature-inspired metaheuristic algorithms for the partitional clustering and showed their application in the field of seismic signal processing. Cho et al. [38] used the concept of TM metric for quantifying the earthquake clustering and further proposed a Particle Swarm Optimization based declustering model. Recently, Vijay et al. [39] same authors introduced a Quantum Gray wolf optimization (QGWO) for declustering earthquake catalogs.

Ester et al. [40] proposed a famous density-based algorithm for discovering clusters in large spatial databases with noise (DBSCAN). It needs two input parameters and discovers clusters of arbitrary shapes even in large spatial databases. The effectiveness of density-based clustering methods motivated the researchers to find the arbitrary shape and varying size earthquake clusters in a seismic prone region. Birant et al. [41] proposed a spatio-temporal DBSCAN algorithm that is capable of discovering clusters under non-spatial, spatial, and temporal values of the dataset. Georgoulas et al. [42] described a ‘seismic-mass’ density-based clustering approach to determine potential spatio-temporal seismic sources in the vicinity of the Hellenic arc. Nanda et al. [43] proposed a merging criterion for clusters to reduce the computational complexity associated with the DBSCAN and identified the hazard regions present in the seismic catalog of Japan. Recently, Scitovski [44] used Rough-DBSCAN for earthquake zoning where SepDIRECT algorithm was introduced for the selection of radius parameter ϵ . In SepDirect, parameters of DBSCAN algorithms are modified, and clusters are formed using Least Squares distance-like function. Schaefer et al. [45] introduced a Smart clustering algorithm for the identification of spatio-temporal clusters using adaptive point process. Magnitude-dependent spatio-temporal earthquake densities are utilized to adjust the searching behavior and determine directional properties to update the search space.

In year 2020, Cesca et al. [46] developed Python open-source software for density-based clustering of seismicity named seiscloud. Seiscloud can find seismicity clusters, characterized by similar features, such as epicentral or hypocentral locations, origin times, focal mechanisms, or moment tensors from the catalog. Tanzim et al. [47] predicted the magnitude of seismic zones using hierarchical k-means clustering along with spatial, temporal density-based approach in the Bangladesh catalog.

Unsupervised neural network-based technique popularly known as self-organizing map (SOM) proposed by Kohonen [48] has been applied in various research fields such as

meteorology [49], geomorphology [50], hydrology [51] and ecology [52]. SOM is an efficient technique for the study and analysis of spatio-temporal clustering. This paper introduces a clustering approach based on the SOM and density-based framework for identifying the correlated AFs and independent BGs. Du et al. [53] did seismic facies analysis based on the SOM and empirical mode decomposition. Allamehzadeh et al. [54] used a SOM for pattern recognition of seismogenic nodes to predict the location of the earthquake and identified seismogenic nodes prone to earthquake. A hybrid approach of the SOM and K -means algorithm applied by Yaghmaei et al. [55] for classification of earthquake ground-motion records.

In the past, machine learning algorithms for earthquake clustering have relied on the K -means algorithm [56] and fuzzy-based algorithm [57] due to ease of implementation and less complexity. However, they are not effective for detecting the irregular shape clusters with outliers. These partitional clustering algorithms also initially needed the cluster’s number as input. The tri-stage clustering model and further its extension by developing the tetra-stage clustering model based on the K-means algorithm is proposed by Vijay and Nanda [24, 25]. The extended model considers significant large events (magnitude-based) as cluster prototypes to identify the clustering and BG activities present in a region.

In the proposed manuscript, the inherent characteristics of a SOM and DBSCAN are utilized to detect the complex multi-dimensional clustering structure (having variable densities and arbitrary shapes that fit the earthquake fault geometry of a given seismic region). Here, SOM provides the seismic zoning based on the event similarity by considering their location and depth in Kms. The ability of the SOM is to determine the potential seismic sources in a region in the form of a 2D weight vector associated with its neurons. DBSCAN algorithm detects the chain of correlated AFs and BGs based on the time information in each seismic zone. Furthermore, correlated AFs and BGs events obtained with the proposed method are analyzed in the spatio-temporal domain. Their accuracy is tested by the Coefficient of Variation and m -Morisita index parameters.

This paper applies a two-stage clustering method based on SOM and Temporal DBSCAN (T-DBSCAN) algorithms on seven instrumental earthquake datasets. The proposed method considers prominent features like (time, space, magnitude, and depth) present in the catalog. The seven earthquake catalogs are Taiwan, Afghanistan, California, the Himalayas, Indonesia, Japan, and Chile. This method is also compared with benchmark declustering techniques Gardner-Knopoff (GK) window method [8], Uhrhammer method [58], Gruenthal window method [59], Reasenberg method [15] and tetra stage clustering model [25] in terms of aftershock-mainshock and BGs

identification. The major contribution of the proposed manuscript is

- A new methodology based on SOM–DBSCAN algorithm has been proposed to address the real-life problem of seismicity declustering.
- This methodology has been applied on earthquake catalogs, including seismic activities of Afghanistan, California, Himalayas, Indonesia, Japan, Chile, and the Taiwan region.
- The better performance of the proposed model is assessed by comparing it with benchmark declustering techniques.
- Validation of proposed methodology has been analyzed using Cumulative and Lambda plot, Coefficient of variance, Logarithmic plot between Inter-event time and Inter-event distance, and m -Morisita index.

The rest of the paper is organized in the following manner: Sect. 2 describes the step-wise procedure of —DBSCAN model to decluster the seismicity of the earthquake-prone region. Section 3 details the instrumental catalog used in the analysis and validation measures. The results obtained are elaborated in Sect. 4 with the help of statistical and visualization-based parameters. The obtained results are compared with benchmark declustering techniques in Sect. 5, followed by a conclusion in Sect. 6.

2 Proposed SOM–DBSCAN based spatio-temporal declustering model

The proposed model based on the SOM (spatial domain) and density-based clustering (DBSCAN) in the time domain require some preprocessing before being fed to the model. The flowchart of the proposed algorithm is shown in Fig. 1, and the step-by-step procedure is outlined in the following subsections:

2.1 Input data

Initially, the data set used in this study is an earthquake catalog containing the time, location, magnitude, and depth information about each event. Mathematically, it is written as

$$E_{N \times D} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} t_1 & \theta_1 & \phi_1 & m_1 & d_1 \\ t_2 & \theta_2 & \phi_2 & m_2 & d_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_N & \theta_N & \phi_N & m_N & d_N \end{bmatrix} \quad (1)$$

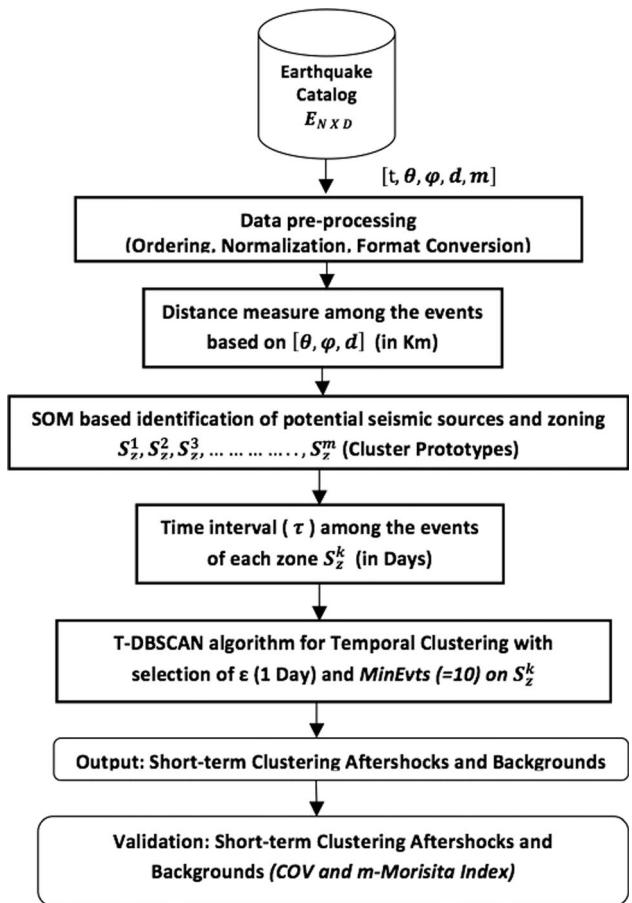


Fig. 1 Proposed SOM–DBSCAN model for declustering earthquake catalogs

where the dataset E comprise of N number of D dimensional events. Any i th event e_i is represented by occurrence time of the event (t_i), location of the event in form of coordinates (with longitude θ_i and latitude ϕ_i), magnitude of the event in Richter scale (m_i) and depth of occurrence (d_i) in kilometers.

2.2 Data pre-processing: magnitude homogeneity using magnitude of completeness

Earthquake catalogs do not cover the full spectrum of earthquake magnitudes. Before any seismic research and hazard analysis, it is essential to examine seismic catalog's accuracy, consistency, and completeness. This can be done by determining the threshold value of the magnitude known as the **Magnitude of completeness (M_c)**, which is the minimum value of magnitude above which all the event magnitudes satisfy the Gutenberg–Richter (G–R) law [60].

The value of (M_c) is measured by applying the G–R law to the frequency magnitude distribution (FMD). The point at which FMD most significantly deviates from the Gutenberg–Richter law reflects the Magnitude of completeness (M_c) value. It is given as

$$\log_{10}N = a - b(m - M_c) \quad (2)$$

Here N represents the total number of events, m is the minimum value of the magnitude, a is constant and b defines the relative number of earthquakes of different magnitude. In the proposed study, magnitude $m < M_c$ are discarded. Figs. 2 and 3 represent the plot of cumulative FMD along with non-cumulative FMD [61]. According to Gutenberg–Richter law [62], the value of M_c is the magnitude increment at which the frequency magnitude distribution deviates from the linear trend in a log-linear plot. The magnitude of completeness M_c for each catalog used in the analysis is calculated using the maximum curvature method [63]. This technique calculated M_c by determining

the maximum value of the first derivative of the FMD curve. The potential advantage of this method is that it requires less number of events to achieve stable point. Seismic catalogs are the most significant and crucial products of the geophysical sciences. These seismic catalogs are generally used as an input source of information regarding events' spatial and temporal distribution. Because of advances in technology and a better understanding of the earth's structure, the seismic network changes throughout time. The seismic network's spatial and temporal properties significantly impact the frequency of earthquake detection and provide uneven (in-homogeneous) seismic catalogs. The less detection of smaller events is due to the following reasons [64] : (1) inability to distinguish the small magnitude events from background noise on the seismograph, (2) To begin the localization procedure and ensure an accurate event detection, a minimum number of recording stations must have picked up the seismic signal. (3) Network operators also have the option

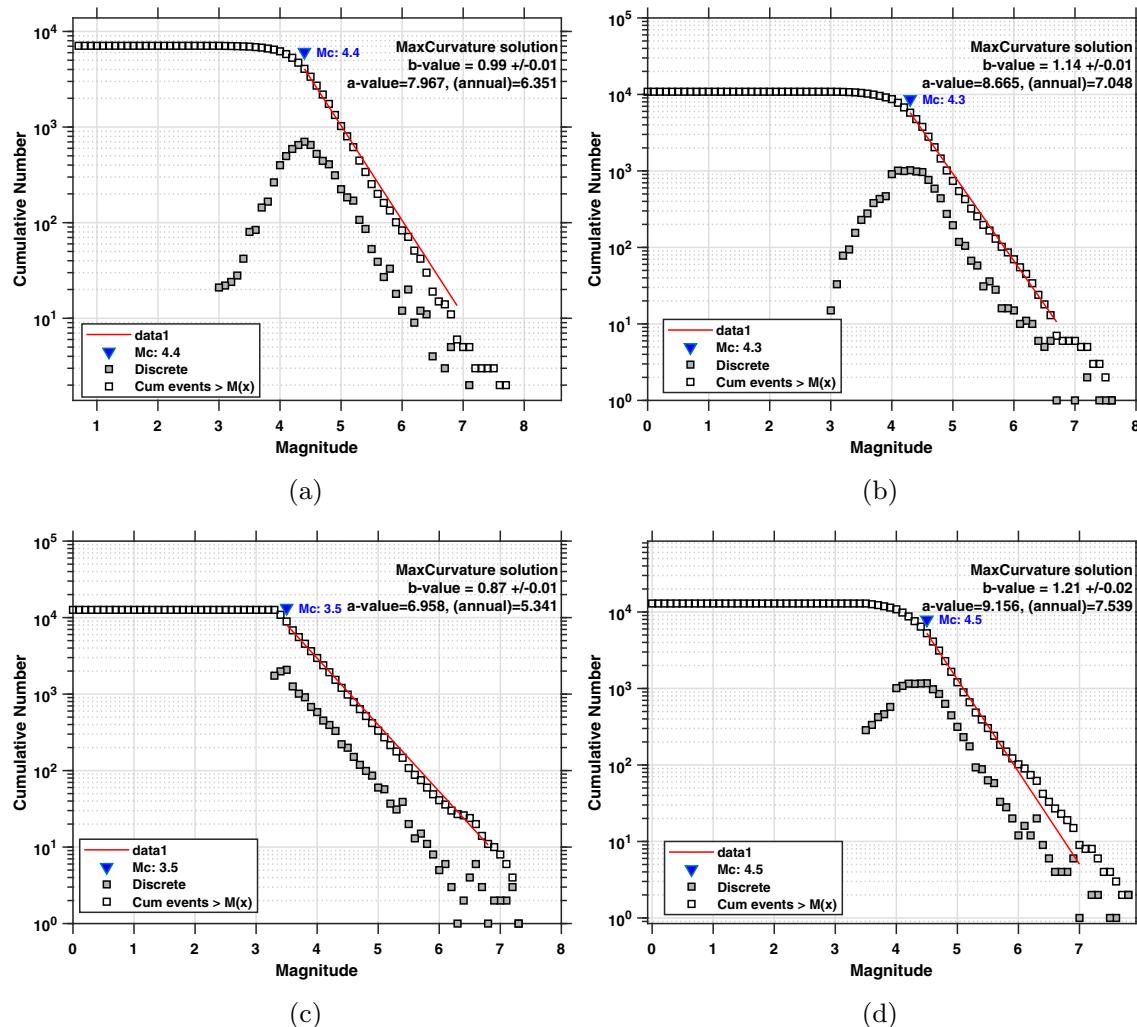


Fig. 2 Frequency Magnitude distribution (FMD) and estimated M_c for **a** Taiwan **b** Afghanistan **c** California **d** Himalaya

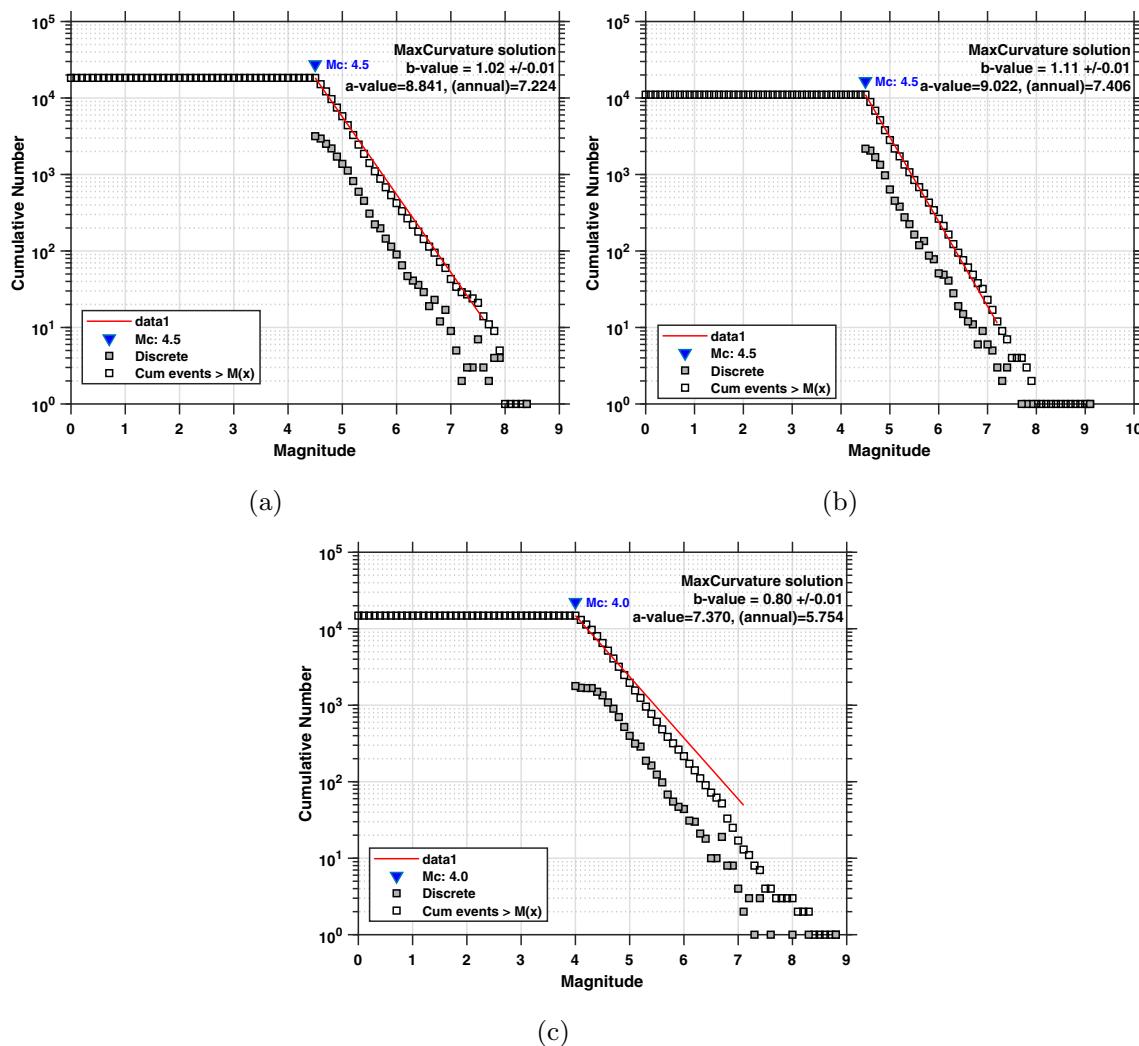


Fig. 3 Frequency Magnitude distribution (FMD) and estimated M_c for **a** Indonesia **b** Japan and **c** Chile

to select a lower bound and remove any events that fall below it. Due to the above reasons, the catalogs are only complete for events of a specific magnitude or higher than it (the magnitude of completeness, M_c). Using events with magnitude less than the magnitude of completeness, or incomplete data, results in faulty calculations of the parameters of the Gutenberg–Richter law (GR law) and incorrect interpretations of seismicity. So seismic events smaller than the magnitude of completeness are rejected in the analysis.

2.3 Phase I: SOM training algorithm for seismotectonic zoning

Seismotectonic zoning is the major classification problem in earth science where the objective is to divide the country or region into different zones that possess a similar tectonic characteristic and the seismicity pattern [65]. This research helps develop seismic hazard maps and cluster analysis. These potential seismic zones are further utilized in the identification of AFs and BGs seismicity. In the present work, the SOM training algorithm is used to find the zones (major seismic sources) based on the coordinate and depth information present in the catalog. The SOM is an artificial neural network model reported by Teuvo Kohonen [66].

Table 1 Parameters setting for the SOM training phase

| Parameters | Taiwan | Afghanistan | California | Himalaya | Indonesia | Japan | Chile |
|-----------------------------------|---------|-------------|------------|----------|-----------|--------|---------|
| Number of iteration | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Quantization error | 0.03231 | 0.06891 | 0.1261 | 0.1031 | 0.1028 | 0.0103 | 0.07073 |
| SOM Grid | 6 × 6 | 6×6 | 5×5 | 7×7 | 6×6 | 7×7 | 5×5 |
| Initial learning rate $\alpha(0)$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Neighborhood size $\sigma(0)$ | 14 | 14 | 14 | 18 | 16 | 16 | 10 |

The SOM algorithm utilizes an unsupervised, iterative procedure to model an input space with a fixed number of nodes. Given X , a set of n d -dimensional input vectors $x_i \in R^d; i = 1, 2, \dots, n$. Let M be a 2-dimensional grid of m neurons with $m = x \times y$, the dimensions of the grid. Each neuron in M has a weight vector $w_l \in R^d$ with index $l = 1, 2, \dots, m$. Then, following steps are carried out to implement SOM training algorithm.

- *Step 1* The weight vectors (SOM prototypes) w_l of all the neurons on map (i.e., $x \times y$ neurons) are initialized. The dimension of weight vector should be same as input vector.
- *Step 2* An input $x_i \in X$ is randomly selected and given to all the neurons of SOM map.
- *Step 3* The winner neuron (w_c) known as Best matching Units (BMUs) and its index c is determined which prototype is closest to the input (x_i) with $x_i \in X$, Mathematically

$$c = \arg \min_l [||x_i(t) - w_l(t)||] \quad \forall 1, 2, \dots, l \dots, M \quad (3)$$

where

$$||x_i(t) - w_l(t)|| = \left[\sum_{j=1}^d (x_{ij} - w_{lj})^2 \right]^{1/2} \quad (4)$$

here d represents the dimension of input dataset.

- *Step 4* An input x_i is used to update the BMU (winning neuron) and its neighboring nodes for all $l = 1, 2, 3, \dots, m$ as follows

$$w_l(t+1) = w_l(t) + \alpha(t) \times h_{c,i}(t) [x_i(t) - w_l(t)] \quad (5)$$

where $\alpha(t)$ is a time varying (decaying) learning rate with respect to the iteration t . The $h_{c,i}(t)$ is loss function, (also known as neighborhood function). The value of $h_{c,i}(t)$ is maximum for the winner neuron and it decreases with the increasing distance from BMU. The Gaussian function is widely used for neighborhood selection which is given by

$$h_{c,i}(t) = \exp \left[-\frac{(||r_c - r_i||)^2}{2\sigma^2(t)} \right] \quad (6)$$

where $||r_c - r_i||$ is the distance between winner neuron c and i th neuron on the map. The $\sigma(t)$ neighborhood radius which is decreased after each iteration as per the following relation.

$$\sigma(t) = \sigma(0) \left(\frac{\sigma(T)}{\sigma(0)} \right)^{t/T} \quad (7)$$

Similarly, learning rate parameter $\alpha(t)$ also decreases linearly after each iteration using following equation:

$$\alpha(t) = \alpha(0) \left(\frac{\alpha(T)}{\alpha(0)} \right)^{t/T} \quad (8)$$

where T is the number of iterations used to train the SOM algorithm. The $\sigma(0)$ and $\alpha(0)$ are initial neighborhood radius and initial learning rate.

- *Step 5* These steps (2–4) are repeated for all input data and then a fixed iteration number until the convergence criterion is achieved.

The convergence of SOM is achieved by minimizing the error between the input and corresponding neuron. This procedure effectively represents the given input space (spatial region). The entire input space is confined into the predefined node weights (SOM prototypes) arranged in a 2-D lattice. The actual output of SOM gives the topology (Hexagonal is used here) arranged feature map, which is automatically structured according to the procedure mentioned above.

2.3.1 Distance measure

This model works in the spatio-temporal domain, which provides a different set of similarity/dissimilarity measurements among the events. A spatial measure of separation between the events is determined by utilizing the longitude (θ), latitude (ϕ) along with depth (d) information between earthquake i and j as

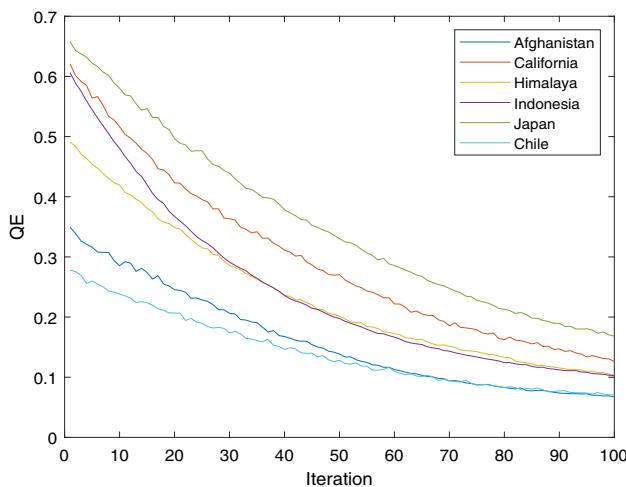


Fig. 4 Obtained quantization error after training of SOM model on earthquake catalogs

$$\eta_{ij} = \sqrt{R_{ij}^2 + (d_j - d_i)^2} \quad (9)$$

where

$$R_{ij} = R_E \times \cos^{-1} [\sin(\phi_i) \times \sin(\phi_j) + \cos(\phi_i) \times \cos(\phi_j) \times \cos(|\theta_j - \theta_i|)] \quad (10)$$

The coordinates latitude and longitude are taken in radians, $R_E = 6371$ km is the radius of the earth. This definition of η_{ij} is based on the assumption that the earth's surface is spherical using epicenter and general hypocenter distance separation used by Kagan [67].

2.3.2 Quality criterion for SOM

The quality of SOM is measured by the structure of SOM obtained after the algorithm's training. If many neurons are fed, input data will be adjusted. However, the SOM map becomes too complicated, increases computational costs, and distorts the original dataset's new classification. Also, less number of neurons results in a reasonable training cost. However, information of feature space is going to be lost, which means defining MAP architecture is a complex problem. Another way to measure the SOM Quality is to evaluate its learning quality [68]. To obtain the optimal configuration of the SOM, the learning quality parameter is measured using the Quantization error (Q_i). It is the mean distance

between the final winning neuron for each input data. The optimal configuration must be kept so that the quantization error becomes minimum. Quantization error Q_i for all the earthquake catalogs are given in Table 1. The quantization error is given as

$$Q_i = \frac{1}{n} \sum_{i=1}^n d(X_i - W_i) \quad (11)$$

Figure 4 shows the minimum quantization error achieved

during the SOM training process. However, in a number of practical applications, neurons obtained from SOM are much higher than the desired number of clusters. Hence, the DBSCAN algorithm minimizes the number of clusters after taking the edge of pre-clustering performed via SOM.

2.3.3 SOM algorithm on earthquake catalogs

The SOM training algorithm (mentioned above) identifies potential earthquake hot spots in a region by considering the event's location depth information present in a catalog. For this, longitude θ , latitude ϕ , and depth (in Kms) are fed as input in the algorithm. The SOM node weights w (neurons) are initialized, and parameters used in the algorithm are selected before iterating the algorithm. The parameters used for the SOM training phase are reported in Table 1. The size of SOM (number of node weights $m = x \times y$) cannot be arbitrarily determined; it is obtained according to the distribution of earthquakes in the spatial region. The SOM size is chosen at which the quantization error becomes minimum. The distance function η_{ij} is used for determining the similarity as given by Eq. (9). The index of winning neuron c among the nodes for each of the inputs is determined according to the given criterion (Known as Best Matching Unit). This winning neuron c is the location in a region that describes a good measure of the concentration of earthquake events. Here we used the latitude and longitude of earthquakes from the last 42 years for each catalog to find the effective 2-d grid of output space in terms of the potential seismic sources. The SOM has significant visualization abilities, allowing it to be an effective tool for analyzing the seismic activities in a region.

Algorithm 1 SOM algorithm for identifying potential Seismic sources in a region

Require: Input: Earthquake Catalog $E_{N \times D} = [f_2^i, f_3^i, f_4^i]$, $N, D = 3$.

- 1: **Initialization:** SOM 2D grid size ($m = x \times y$), Iteration T , Initial neighborhood radius and learning rate ($\sigma(0), \alpha(0)$), Initial 2D seismic sources $w^p \forall p = 1, \dots, m$.
 - 2: **Output:** Seismic Zones $S_z^1, S_z^2, \dots, S_z^m$
 - 3: **Start:** Iteration $t = 1, 2, \dots, T$, Random input selection $f^i \forall i = 1, \dots, N$, Initial SOM prototypes $w^p \forall p = 1, \dots, m$
 - 4: **Competitive Process:** $V = f^i - w^p, \eta(x, y) = (V \times V')$, Eq.(9) and Eq.(10) is used in distance calculation
 - 5: **BMU:** Determine the winner neuron $c = \min_p \eta(x, y)$
 - 6: **Determine:** Define the size of the neighborhood according to Eq.(7)
 - 7: **Determine:** Define the learning rate according to the Eq.(8)
 - 8: **Cooperative process:** Determine the Neighborhood function for each neuron according to the Eq.(6)
 - 9: **Update:** Update the SOM prototypes according to the Eq.(5)
 - 10: **Repeat:** $i = i + 1$ till $N, t = t + 1$ till T
 - 11: **Final:** Obtain the Optimum SOM prototypes (Neurons) $w_*^p \forall p = 1, 2, 3, \dots, m$.
 - 12: **Potential Zones:** Neighborhood events to each SOM prototypes are selected into distinct zones based on the distance criterion and make $S_z^1, S_z^2, \dots, S_z^m$.
-

For example, Fig. 5a represents the SOM layer (Hits), with each neuron showing the number of input vectors that it classifies for the Taiwan earthquake catalog. Also, Fig. 8a–f shows the SOM layer (Hits) characteristics for the benchmark earthquake dataset. The relative number of vectors for each neuron is demonstrated via the size of a colored patch. It is reflected from the figures that some of the nodes are triggered many times for the random inputs due to having high

density in the surrounding region. The adjacent neurons also have a high probability of hitting, as shown in Figs. 5b and 8a–f. These high probable nodes are considered to be potential seismic sources, and the surrounded zones are declared as $S_z^1, S_z^2, \dots, S_z^m$. The distribution of earthquake epicenters in a region is represented by node weights (SOM prototypes), as shown in Figs. 5b and 8g–l. The node weights

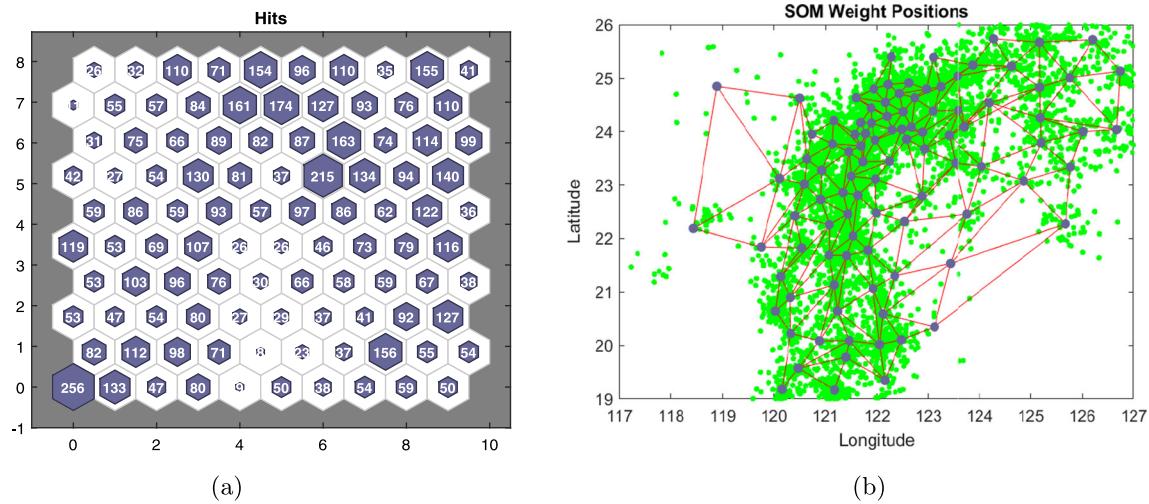


Fig. 5 Results obtained from proposed model for Taiwan region **a** SOM layer (Hits) **b** SOM prototypes

are the location of the potential seismic source for the Taiwan catalog. These are represented by the big blue circle in identical Figs. 5b and 8g–l. The green color events are original input space. These are the earthquake hot spots (location of the SOM prototypes), and further active seismic zones of a region are presented in the next section.

detecting both primary and secondary AFs generated due to the single big mainshock. Another capability of the algorithm is that it works well even in outliers in the data sets.

Algorithm 2 T-DBSCAN algortihm to identify the clusters in Seismic zones

```

1: Input: Earthquake Seismic Zones ( $E_z$ )
2: Input:  $\epsilon$ : Radius
3: Input: MinPts (Threshold)
4: Input: Dist: Distance Function
5: Input Data: Earthquake Catalogs (unlabelled)
6: for Every point in earthquake seismic zone do
7:   if label(p) = not defined then
8:     Neighbours N  $\leftarrow$  RangeQuery(Data,Dist, $\epsilon$ ,p)
9:     if N < MinPts then
10:      label(p)  $\leftarrow$  Clusters(BGs)
11:    end if
12:    c  $\leftarrow$  next cluster labelled
13:    label(p)  $\leftarrow$  c
14:    seed set (S)  $\leftarrow$  (p)
15:  end if
16:  for each q in S do
17:    if label(q) = Clusters (BGs) then label(q) $\leftarrow$  c
18:    end if
19:    if label(q) = undefined then Continue
20:      Neighbors N  $\leftarrow$  RangeQuery( $E_z$ ,dist,p, $\epsilon$ )
21:      label(q) $\leftarrow$  c
22:      if N < minPts then continue
23:        S  $\leftarrow$  S  $\cup$  N
24:      end if
25:    end if
26:  end for
27: end for
  
```

2.4 Phase-II: density-based temporal clustering (T-DBSCAN) on seismic zones obtained from SOM prototypes

The events that occurred nearby in time are essential to be in the same group and be isolated from others in each spatial cluster (seismic zones obtained from SOM). It leads to effective identification and discrimination of correlated AFs and BGs by applying the temporal density-based criterion in each potential zone. For this purpose, a well-known density-based algorithm: DBSCAN [40] is used to detect the AFs cluster and singular BGs events. The advantage of applying this algorithm is

Time-based DBSCAN (T-DBSCAN) algorithm on earthquake events work in the following manner:

- Let i th event of S_z^k zone and their time information t_k^i (relatively in days).
- Set: Epsilon neighborhood (N_ϵ). Here, N_ϵ is taken as one day.
- Core Events: An i th event that has at least $MinEts$ (including itself) events within its N_ϵ .
- Primary AFs: An event j is considered primary AFs including event i if j is core point and $j \in N_\epsilon$.
- Secondary AFs: Those events are called secondary AFs if a chain of primary AFs links these events. This expends the cluster by merging them.

- Border Events: Events that are not a core event.
- BGs: Event that does not belong to any event's N_e .

This algorithm finds the time-based primary and secondary parts of the single cluster and BGs in seismic zones. The pseudocode of T-DBSCAN is given in algorithm 2. The results obtained for Taiwan are shown in Fig. 6b. Corresponding identified AFs, BGs, and a number of clustered events are reported in Table 3. The behavior of these obtained AFs and BGs over time is represented by the Lambda plot (events per year), and Cumulative plots are discussed in the next section in detail.

2.5 Case study of the proposed SOM-DBSCAN on Taiwan catalog

The proposed SOM-DBSCAN algorithm is initially tested by taking a small regional earthquake catalog of the Taiwan region. The information about the catalog is discussed in Sect. 3 of the manuscript. The epicenters of earthquakes in a

region are demonstrated in Fig. 6a, along with depth in different colors. First, information about the earthquake coordinate and depth present in the catalog is fed into the SOM algorithm to determine the significant seismic and potential zones. The SOM map is initialized to obtain winning neurons called Best Matching Units (BMUs) according to the spatial distance criterion given in Eq. 3. It helps to find the similarity between the events and neurons. These obtained BMUs represent the high thickness of earthquake events in that region. Fig. 5a shows the SOM layer (Hits), where each cell represents the number of input neurons classified for earthquake catalogs. The size of the colored hexagonal cell reveals the number of vectors associated with that particular event. It also represents the number of times a specific node is triggered due to excessive events in the neighborhood region. These high triggered nodes are probable seismic zones (S_Z). Fig. 5b represents SOM prototypes obtained after SOM training. The big blue dots correspond to the node weights, which are the source of earthquake fault structure of the prone earthquake region and green color events are original

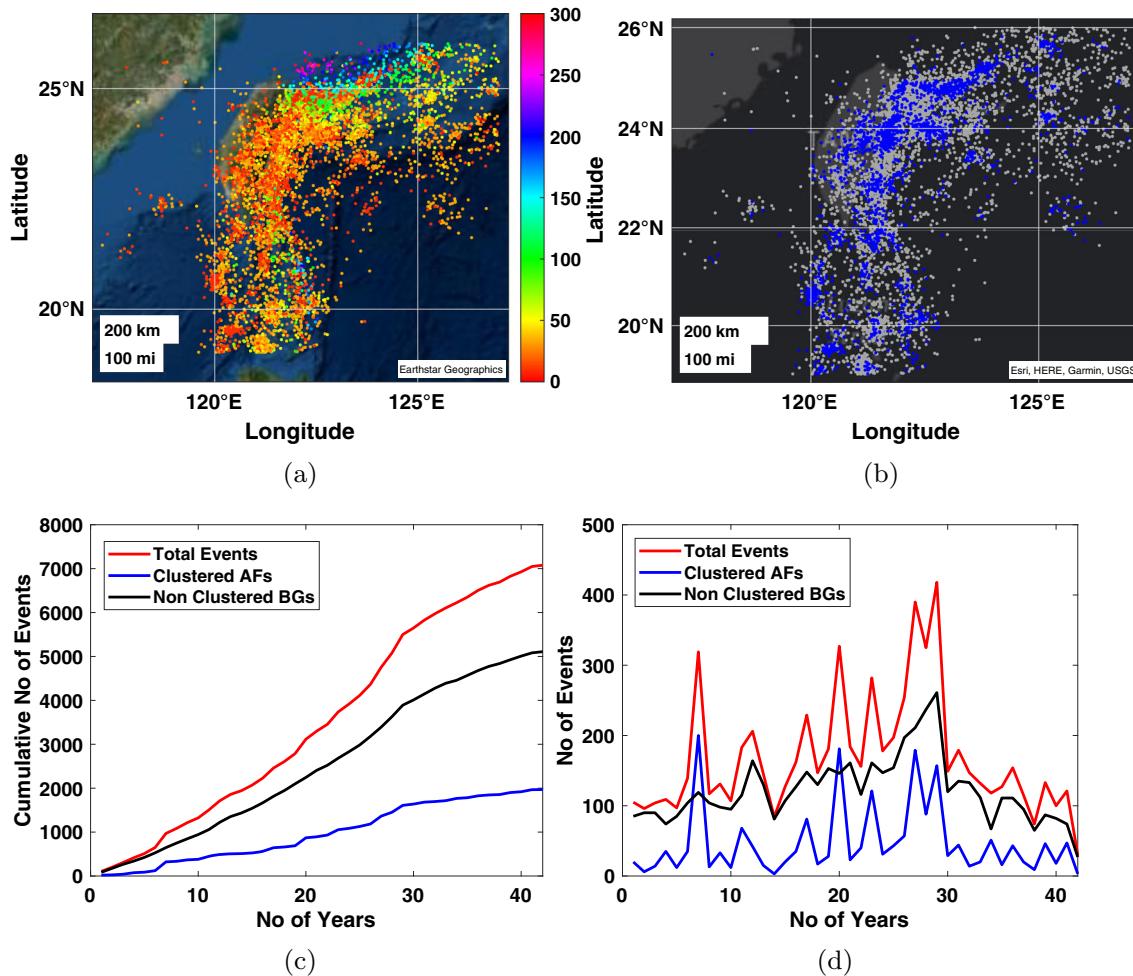


Fig. 6 Results obtained from Proposed model for Taiwan region **a** Spatial distribution of Total events **b** Spatial-temporal distribution clustered AFs epicenters (blue dots) and Declustered BGs epicenters (grey dots) **c** Cumulative plot and **d** λ plot for Total, clustered and declustered events

input space (Total events). These are the SOM prototypes which are further analyzed in terms of the foreshock, AFs, and BGs identification of seismic earthquake zones. After obtaining the seismic zones, declustering of seismic events and identification of foreshock-aftershock and BGs events are carried out based on the Time-based DBSCAN algorithm (T-DBSCAN).

The T-DBSCAN is used with only two parameters, radius ϵ and a minimum number of points ($MinPts$) within the radius ϵ . The short-term correlated AFs and BGs are identified and separated by T-DBSCAN in the seismic zone according to Algorithm 2. Epicenter plots of clustered events (foreshock-aftershock) and declustered events (BGs) are as shown in Fig. 6b. Discrete spatial clusters have been identified in the entire region with unique cluster IDs. Spatial clusters of the entire Taiwan region based on Time (Temporal clustering) are represented as blue dots. It is observed from the figure that few clusters are very close to each other because of the mainshocks. The grey color dots represent the epicenters of background events uniformly distributed throughout the region.

Figure 6c and d represents the characteristics of the AFs and BGs events concerning time for the Taiwan catalog as cumulative plot and Lambda (λ) plot, respectively. It also signifies the clustering efficiency of the proposed method by separating the clustered AFs and declustered BGs well apart. The cumulative seismic rate (red curve) follows a linear trend from the origin time of events where small AFs are stimulated and fall off concerning time. Blue and black curves represent the cumulative sum of AFs and BGs, respectively. Slight divergence in seismic rate results in significant changes in the linearity,

indicating effective identification and separation among the events. The linearity curve of BGs also reveals the events are homogeneously distributed in time. The cumulative plot shows the density of events with time. The number of events increases abruptly with the large seismic activity, as shown in Fig. 6d. The characteristics of the cumulative and λ plot are further discussed in the next section obtained from the benchmark earthquake catalog. Figures 7 and 8 represents the SOM hits and SOM weight positions for optimum number of nodes for Afghanistan, California, Himalaya, Indonesia, Japan and Chile.

3 Analysis of instrumental seismic catalogs using proposed SOM-DBSCAN model

In this section, the efficiency proposed SOM-DBSCAN model is tested on seismic catalog of Taiwan, Afghanistan, California, the Himalayas, Indonesia, Japan, and Chile. Many devastating earthquakes have struck these seismically active regions over the years. These seismic catalogs

are extracted from United States Geological Survey (USGS) [69] website by setting the appropriate parameters as mentioned in Table 2. The catalogs contain information of about the origin time (t), events location in terms of latitude(θ) and longitude (ϕ), magnitude (m), and depth (d). The Magnitude of completeness M_c for each catalog is determined, and events with magnitude (m) less than M_c are not considered for the analysis.

3.1 Seismic catalogs used in the analysis

3.1.1 Taiwan

Taiwan has very complex regions of faults and a highly tectonic active region under stress due to the collision of the Eurasian plate with the Philippines sea plate. Western Taiwan is a highly seismic active region due to shallow focal depths. In September 1999, one of the devastating earthquakes of the century struck a central island near town Chi-chi with a magnitude of 7.6 on the Richter scale, having 2300 causalities and 8700 injuries. The seismic catalog of 42 years from January 1980 till February 2021 is used. The events below M_c value of 4.4 are discarded, and the rest 7788 events are used in the analysis. Figure 6a represents the epicenter plot of earthquake catalogs.

3.1.2 Afghanistan

Afghanistan is the second most seismically active region located at the Alpide belt and near the southern extent of the Eurasian plate. Here, 42 years of earthquake data from January 1980 till February 2021 has been used for the analysis. In this duration, a number of devastating earthquakes occurred in the Hindu-Kush mountain range of Afghanistan. A total of 9809 events having a magnitude higher than $M_c = 4.3$ are used in the analysis. The epicenter distribution of seismic events during the studied time interval is illustrated in Fig. 9a.

3.1.3 California

California is one of the most seismically active regions in the world, located at the San Andreas Fault, which extends roughly 800 miles through the United State. At this fault, the tectonic boundary between the Pacific Plate and the North American Plate meets, making it a highly earthquake-prone region. The spatial distribution of 9998 events with M_c value of 3.5, as shown in Fig. 9b, is used for the analysis.

3.1.4 Himalaya

The Himalayas are highly seismic active regions due to the continuous collision between the Indian and Eurasian

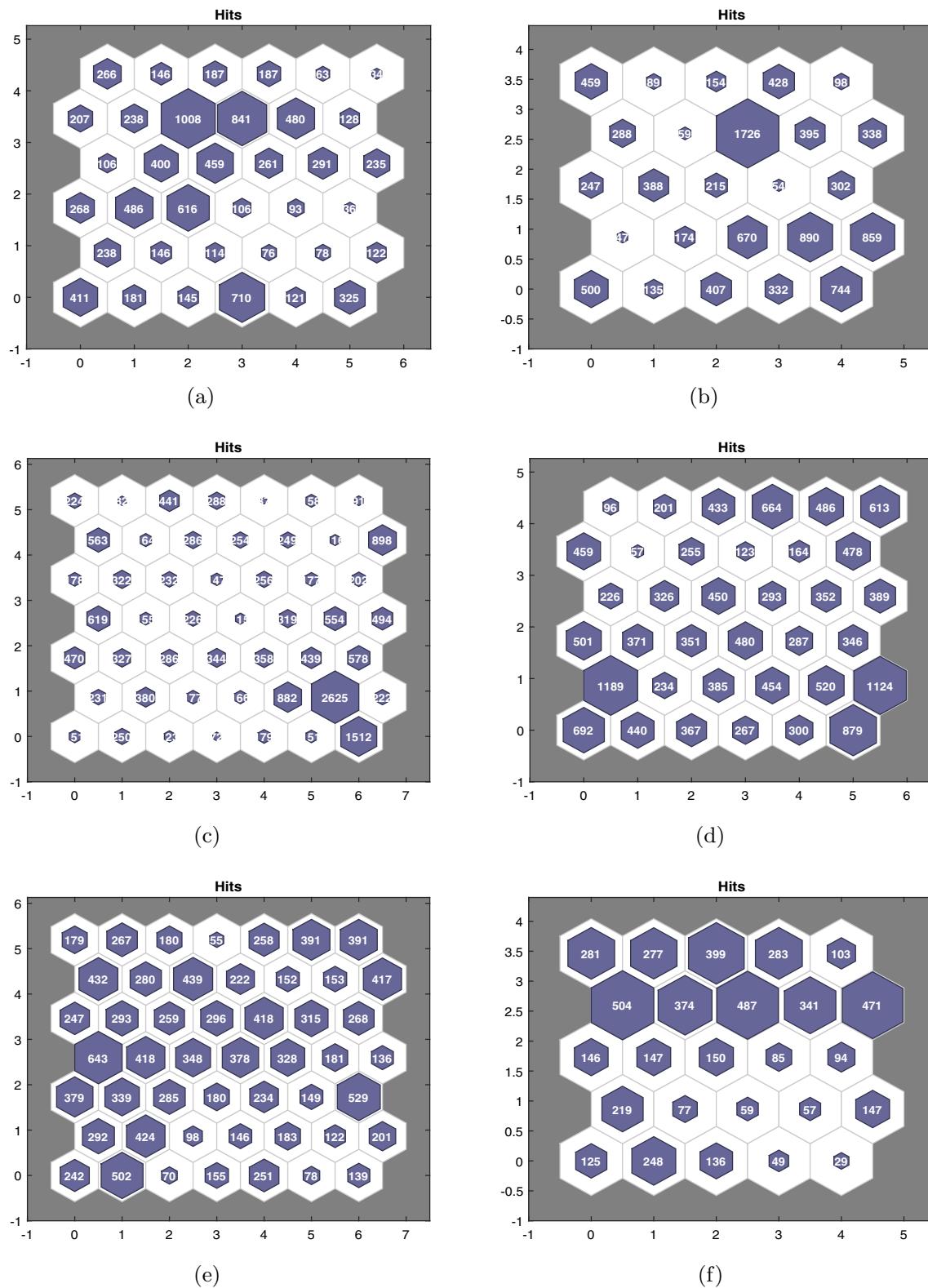


Fig. 7 SOM Hits for optimum number of nodes for **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan **f** Chile

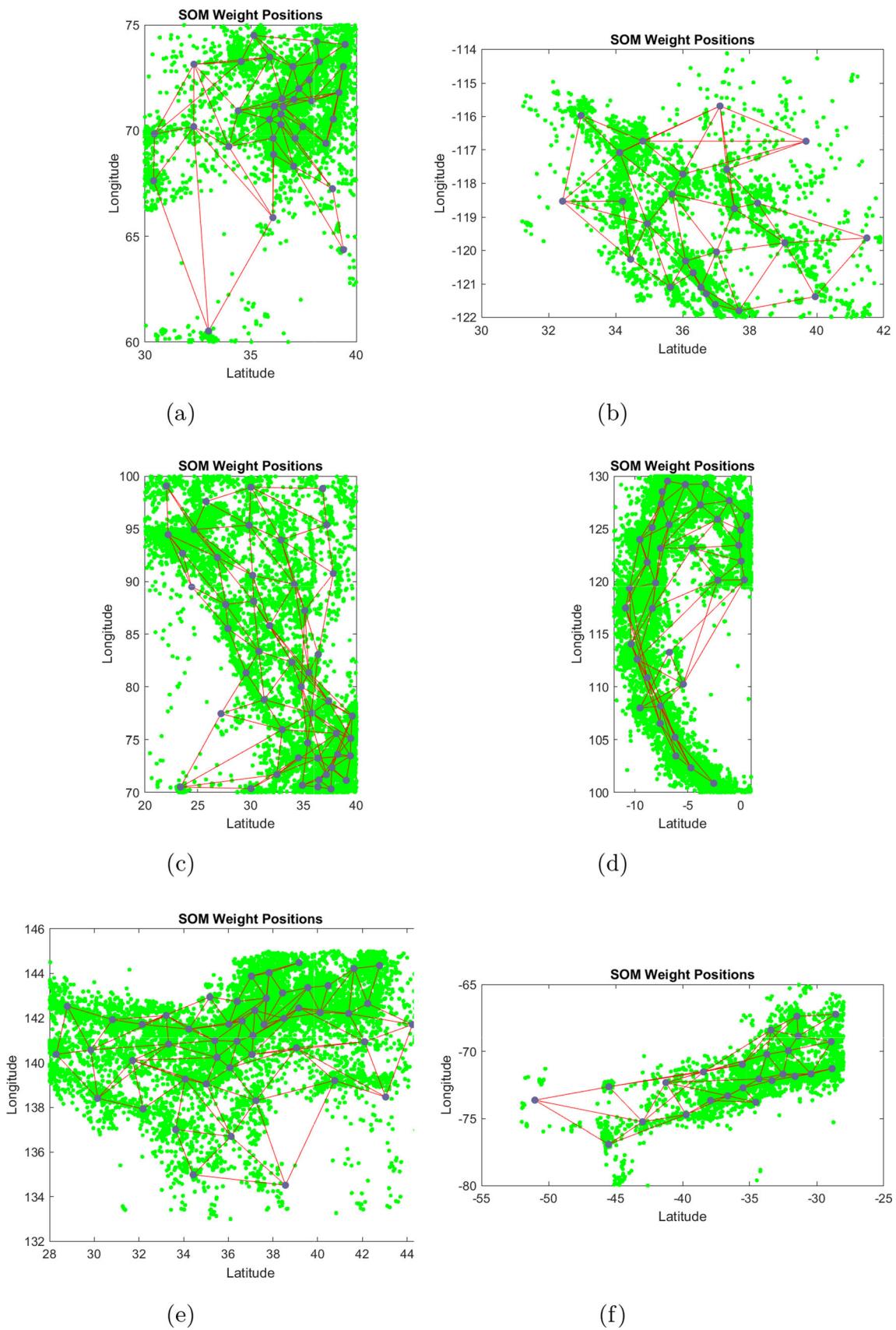


Fig. 8 SOM weight position for optimum number of nodes for **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan **f** Chile

Table 2 Information of seismic parameter to extract earthquake catalogs from USGS [69] used in the study

| Parameters | Catalogs | | | | | | |
|---------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | Taiwan | Afghan | California | Himalaya | Indonesia | Japan | Chile |
| Start time | 1980/01/01 00:00:00 |
| End time | 2021/02/01 23:59:59 |
| Min latitude | 19 | 30 | 30 | 20 | -12 | 28 | -51 |
| Max latitude | 26 | 40 | 42 | 40 | -1 | 45 | -25 |
| Min longitude | 117 | 60 | -122 | 70 | 100 | 133 | -80 |
| Max longitude | 126 | 75 | -114 | 100 | 130 | 145 | -65 |
| Min depth | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Max depth | 300 | 383 | 103.3 | 383 | 675 | 584 | 258 |
| Min magnitude | 3 | 2.9 | 3 | 3 | 3 | 3 | 3 |
| Max magnitude | 7.6 | 7.8 | 7.5 | 7.7 | 8.4 | 9.2 | 8.8 |
| Magnitude of completeness | 4.4 | 4.3 | 3.5 | 4.5 | 4.5 | 4.5 | 4 |
| Event type | E | E | E | E | E | E | E |

continental plates. The spatial distribution of 17770 earthquake events from January 1980 to February 2021, as shown in Fig. 9c, is analyzed with minimum magnitude $M_c = 4.5$.

3.1.5 Indonesia

Indonesia's tectonic structure is extremely complex because it is the meeting point of various tectonic plates. Indonesia is highly vulnerable to earthquakes due to its location on the Ring of Fire. For the study purpose, the 15252 number of earthquake events spanning the period of January 1980 to February 2021, with $M_c = 4.5$. are used spatially distributed as shown in Fig. 9d.

3.1.6 Japan

Japan is another highly seismic sensitive region located at the Ring of fire. At the Ring of fire, several tectonic plates, including the Pacific Plate beneath the Pacific Ocean and the Philippine Sea Plate, mesh and collide. Japan was struck by the most devastating earthquake of the century in the Tohoku region on 11 March 2011. From January 1980 to February 2021, many deadly earthquakes occurred, making it an exciting part of interest among geologists. For the analysis, 13342 events with $M_c = 4.5$ are used, and their spatial distribution is shown in Fig. 9e.

3.1.7 Chile

The geographical location of Chile is highly diverse. The Atacama Fault Zone (AFZ) has widespread in Chile across the Chilean Coastal Cordillera between the Pacific Ocean and Andean Mountain range. Chile suffered several devastating earthquakes over the decades, like February 2010, with many casualties. Here several events are taken in the clustering analysis from January 1980 to February 2021 with a magnitude of completeness $M_c = 4.5$. The spatial distribution of epicenters of 5288 events is depicted in Fig. 9f.

3.2 Validation measure

The following five validation measure have been employed for clustering validity.

3.2.1 Epicenter plot

Epicenter plot represents the clustered events (AFs) and regular events (BGs). It shows the coordinated information of every individual AFs and BGs event. The clustered events AFs represent a more dense region around the mainshocks and exhibit a nonuniform distribution of events around the whole tectonic zone. The de-clustered events (BGs) follow the uniform distribution of events. The results obtained by the proposed model on the benchmark dataset are explained in detail in the result and discussion section.

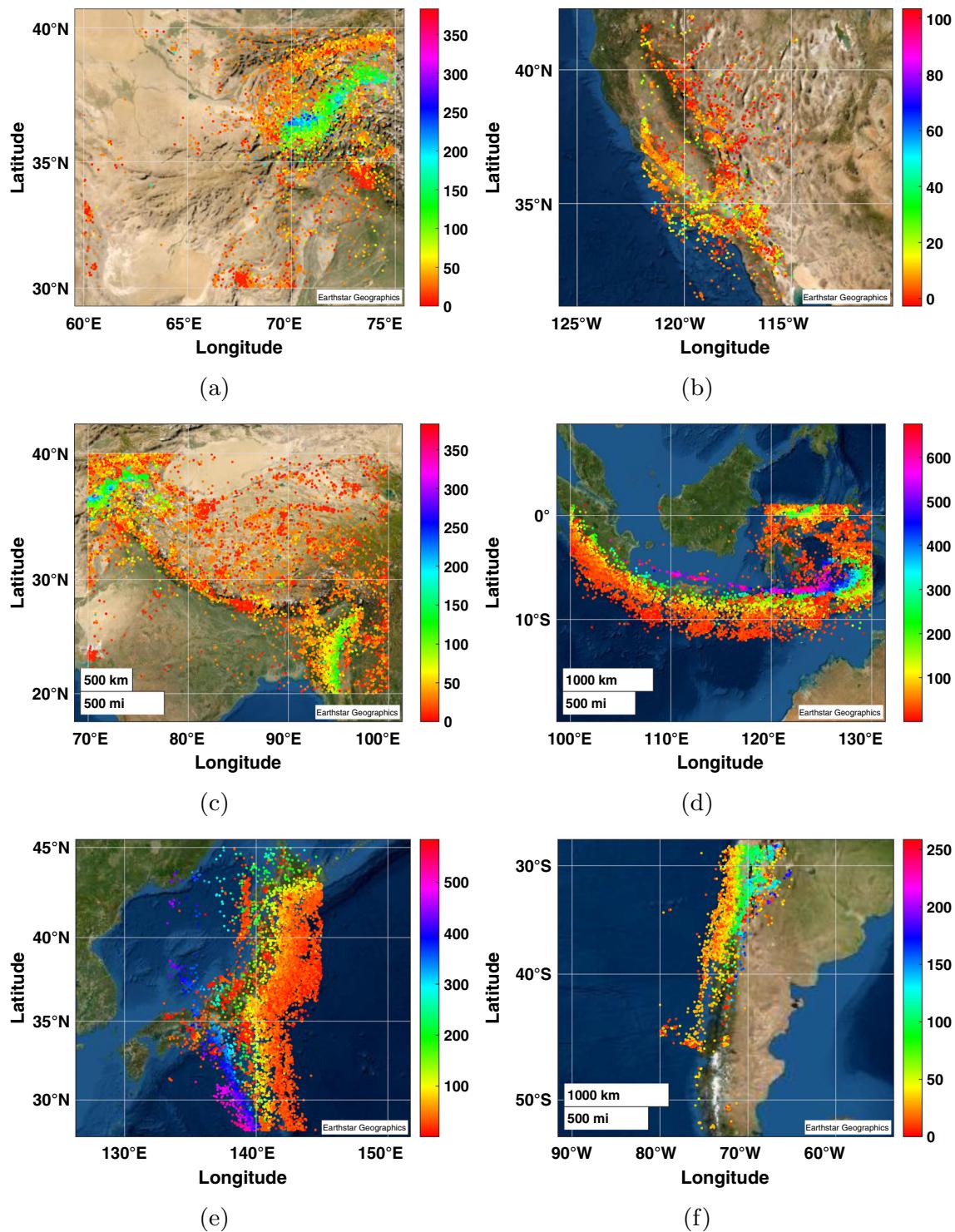


Fig. 9 Spatial distribution of earthquake epicenters occurred in **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan **f** Chile

3.2.2 Cumulative plot and lambda (λ) plot

The characteristics of AFs and BGs concerning time are represented by the cumulative and lambda (λ) plots. A cumulative plot calculates the number of events with time.

Lambda plot shows the variation in seismicity. It shows that the abrupt change in the slope represents the high seismic activity in that duration.

3.2.3 Coefficient of variance (COV)

The inter-event times (IETs) of an earthquake time series comprised of N number of events and duration T is given by

$$\Delta t = t_j - t_i \dots \forall j = 2, 3 \dots N \quad \text{and} \quad i = 1, 2, \dots, N-1 \quad (12)$$

where t_j and t_i are the occurrence time of the two successive events of a time series having duration T .

Based on IETs, further, the Coefficient of Variance (COV) is used to define the statistics (characteristics) of an earthquake time series. Coefficient of Variance (COV) is the ratio of the standard deviation and average of inter-event time (Δt). Mathematically, it is represented as

$$\text{COV}(T) = \frac{\sqrt{E[(\Delta t)^2] - [(E(\Delta t))^2]}}{E(\delta t)} \quad (13)$$

The temporal distribution of the earthquake sequences are classified into three categories based on the approximate value of $\text{COV}(T)$ for the time series of an earthquake [28]:

- If $\text{COV} < 1$, it shows that time series reflects less variability.
- If $\text{COV} \approx 1$, it represents that mean and variance of inter-event time Δt are equal and IETs Δt is exponentially distributed.
- If $\text{COV}(T) = 0$, then standard deviation is zero and IETs $\Delta t = \text{constant}$, then time series is periodic in nature.
- If $\text{COV} > 1$, then IETs Δt has high variance and inter-event time sequence follows the power law distribution which represents the clustering pattern (aftershocks).

3.2.4 IET versus IED plot

The scatter logarithmic plot between Inter-event time (Δt) vs Inter-event distance (Δd) segregates the seismicity into AFs and BGs. The plot obtained shows the dense concentration of the seismic events into two separate clusters representing AFs and BGs events. Δd versus Δt scatter plot is another crucial parameter to validate the efficiency of the proposed declustering model [70].

3.2.5 m -Morisita index

The m -Morisita index proposed by Goley et al. [71] is a statistical measure based on the grid of Q cells with changing size δ . The m -Morisita index is calculated by superimposing the grid Q on the studied data set and

determining whether the randomly selected two points within the same quadrant follow the random distribution generated from a Poisson process. The m -Morisita index (I_m) identifies whether the point distribution is random, dispersed, or clustered. The multi-point Morisita index represents the m points with $m \geq 2$. The I_m value is unity in case the point pattern is generated by the Poisson process. The point patterns are dispersed if I_m is less than one and reduces to zero as δ decreases. For the clustered pattern, I_m is greater than 1.

For the m number of points ($m \geq 2$) with quadrant Q having size δ (diagonal length), m -Morisita index ($I_{m,\delta}$) is given as

$$I_{m,\delta} = Q^{m-1} \frac{\sum_{i=1}^Q n_i(n_i-1)(n_i-2)\dots(n_i-m+1)}{N(N-1)(N-2)\dots(N-m+1)} \quad (14)$$

where n_i is the number of points in the i th quadrat and N is total number of points (sample size). Initially, $I_{m,\delta}$ is measured for the small quadrat size then increases until the chosen threshold. A plot (on a logarithmic scale) is drawn between the $\log(I_{m,\delta})$ and it is corresponding higher quadrat $\log(\delta)$ for the given value of m which results in continuous fall in the slope until a minimum value is reached.

for the study of the multifractal patterns of the for the order of $q = m$, the m -Morisita index follows the power law given as

$$I_{m,\delta} \propto \delta^{(m-1)(D_m-E)} \quad (15)$$

and

$$\lim_{\delta \rightarrow \infty} \frac{\log(I_{m,\delta})}{l} \log(\delta) \approx (m-1)(D_m-E) \approx -(m-1)(C_m) \approx -S_m \quad (16)$$

where

$$D_m \approx E - \left(\frac{S_m}{m-1} \right), m \in 2, 3, 4, 5 \quad (17)$$

Here E represents the dimension of space comprise of the coordinate location where seismic events held (here, $E = 2$ for the $X - Y$ coordinate plane). C_m is the co-dimension of order $q = m$. S_m is the m -Morisita slope and the relationship between m and S_m is referred as m -Morisita slope spectrum. S_m is also termed as the slope obtained by fitting the linear regression on seismic data points on $\log(I_{m,\delta}) - \log(\delta)$. S_m also measures of the degree of clustering. It extends between the $0 \leq S_m \leq (m-1)E$ where the lower bound depicts the regular pattern and higher bound represent the highly cluster patterns.

4 Results and discussions

The proposed SOM-based methodology is applied to the real earthquake catalogs of Afghanistan, California, the Himalayas, Indonesia, Japan, and Chile. The results and their analysis are presented in this section. The method identifies earthquake AF clusters present in the catalog and categorizes the total events into clustered AFs and non-clustered BGs. The comprehensive description of results are carried out in the following subsections:

4.1 Epicenter plots of AFs and BGs

The distribution of AFs and BGs, which are identified from the proposed method, are represented by their epicenters (coordinate information) as depicted in Fig. 10a–f for Afghanistan, California,

Himalaya, Indonesia, Japan, and Chile, respectively. The clustered events obtained from the proposed model for all the catalogs are demonstrated in Fig. 10a–f. In Fig. 10a–f, blue color dots highlight the number of clustered events, and the corresponding declustered catalog is shown with light grey colors as BGs. It is evident from Fig. 10a–f event epicenters have high density and accumulated at the active fault locations. This represents the clustering phenomenon in space shown by the correlated AFs in the surrounding area of the large event. The non-clustered events (independent BGs) depicted with grey dots are smoothly distributed along the fault boundaries due to the absence of triggering effect (occur only due to the strain accumulated by tectonic loading) as shown in Fig. 10a–f for all the catalogs.

A total 210 number of clusters have been identified from the proposed technique for the Afghanistan region. Similarly, for the California region, a total of 9998 events were analyzed, out of which 2124 events are classified as AFs, and 196 clusters are formed in the region, as shown in Fig. 10b. The highest number of earthquake events were analyzed for the Himalayan region. Here out of 17770 Total events, 247 clusters are formed, as shown in Fig. 10c. The identified clusters for Indonesia, Japan, and Chile are 236, 216, and 165, respectively. It is evident from Fig. 10d–f that AFs are very closely related to mainshocks in space, and the BGs events are homogeneously distributed over the region.

4.2 Cumulative and λ plots

The observation in the change of seismicity rate is examined by visualizing the cumulative distribution of

the number of events with respect to time effectively. The number of event occurs during the given time interval (λ) and their cumulative sum represented by their respective plots indicate the formulation of earthquake clusters in time. These plots also justify the effective separation of the clustered and non-clustered events as shown in Fig. 11 (cumulative plot) and Fig. 12 (λ plot) for each catalog.

In both figures, red represents Total seismic rates, blue for AFs, and black for BGs. Here, in Fig. 11a–f cumulative seismic rate shown by the total catalog deviates from a linear trend at the occurrence time of a significant event where correlated AFs (short-term clustered events) are triggered and decayed over time (see the red curve in Fig. 11 for each of the catalogs). The proposed method identifies these correlated AFs, and the blue line represents their cumulative sum in Fig. 11a–f. It is observed that cumulative rate changes abruptly at the time of mainshock activity, indicating the effective identification and separation by the proposed methodology. Here, the blue and red curves also have close trends that signify that the proposed model accurately detects the aftershocks from the Total catalog. In Fig. 11a–f, the black curve shows the cumulative seismic rate of BGs. Their linear nature also reflects the effective and accurate identification of the events. It is also observed from Fig. 11a–f, The cumulative seismicity rate of BGs events is linear, constant, and homogeneous for the entire duration, which is the necessary characteristic of BGs.

Similarly, in Fig. 12a–f, the λ rate (event occurrence rate in a year) of total events is represented by a red curve with large sequences with peaks in the rate. The abrupt surge and subsequent decrease of events around the time of the mainshock event show the presence of correlated aftershock clusters. In Fig. 12a–f, it is evident that all the major AFs sequences are identified, and the blue curve shows their arrival rate in a year, where all peaks are retained. A close blue and red curve similarity justifies the accurate identification of correlated AFs (see Fig. 12a–f). The absence of peaks in the black curve (see Fig. 12a–f) and almost uniform arrival rate highlight the characteristics of BGs seismicity for each of the catalogs. The nature of BGs obtained here yielded a homogeneous Poisson process with a stationary (constant) λ rate in a given period (Fig. 12).

4.3 Temporal cluster analysis of earthquake catalogs using coefficient of variance (COV)

The Coefficient of variance (COV) is determined for each of the original catalogs, which is presented in

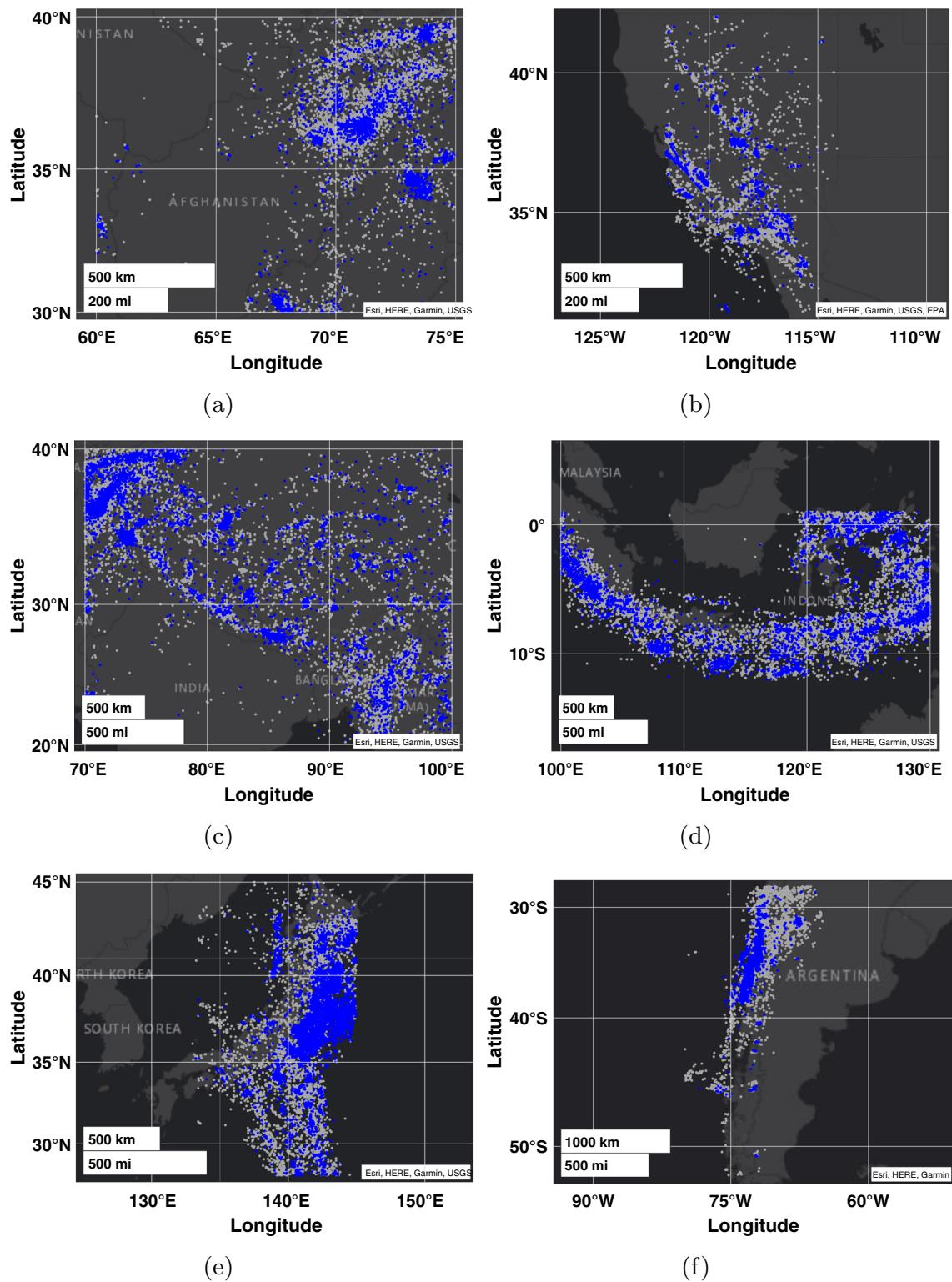


Fig. 10 Resulting Clusters AFs (blue dots) and BGs (grey dots) events obtained using SOM-DBSCAN algorithm for **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan and **f** Chile regions

Table 3. Due to the presence of both AFs (Clustered in the time domain; short IETs and BGs (Non-clustered in the time domain; large IETs) in catalogs, the time series

has a medium variance. After separating both types of events from the proposed methodology, the Coefficient of variance COV_A for obtained clustered AFs is greater

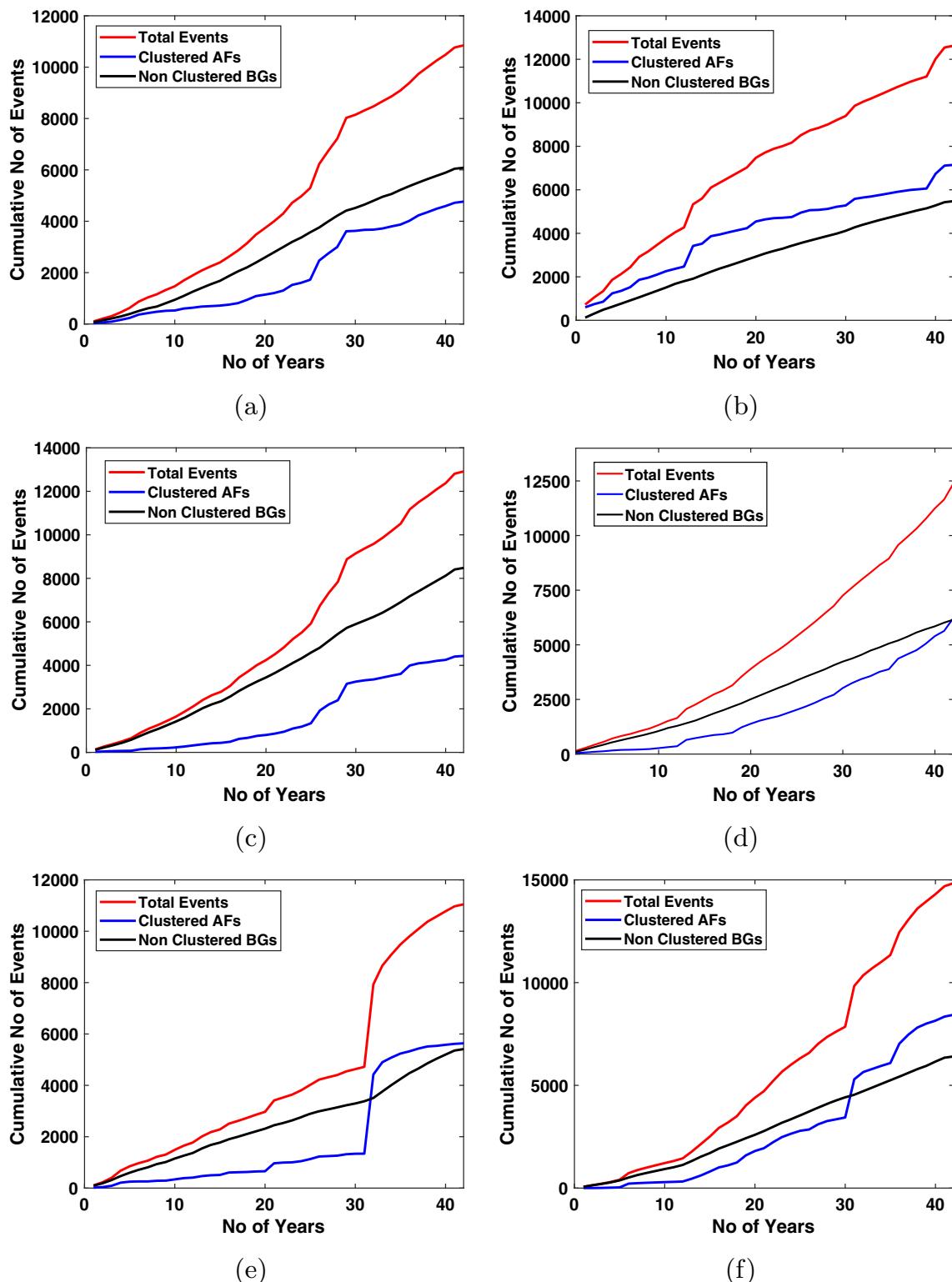


Fig. 11 Cumulative events obtained using SOM–DBSCAN for **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan **f** Chile

than 1, i.e., the IETs of AFs have high variance. The abrupt increase in the number of seismic events at the occurrence time of high magnitude events generates the highly variable temporal clusters in time, which causes

significant changes in the Coefficient of variance. The large values of COV_A in Table 3 reflect the presence of aftershocks clusters.

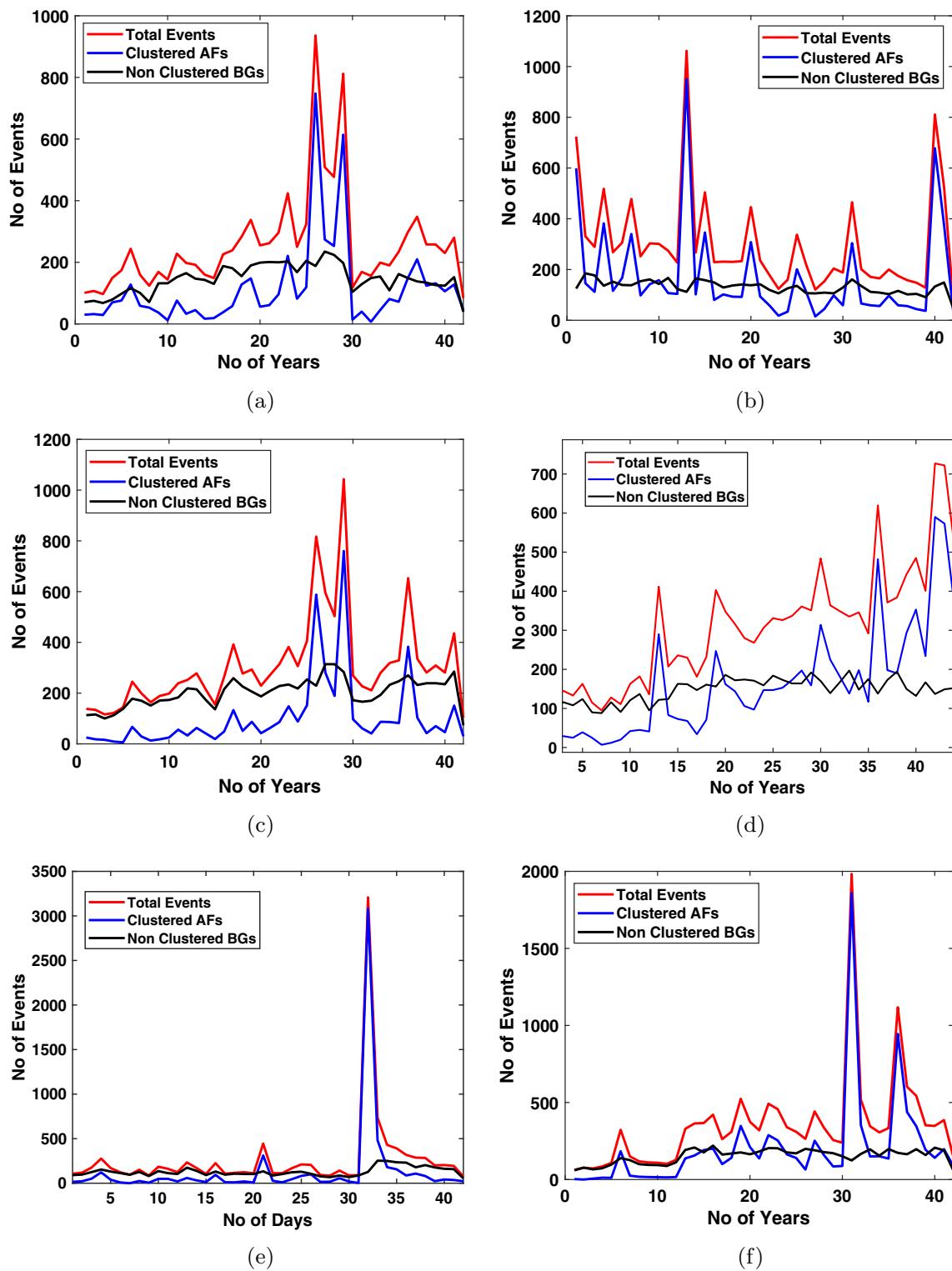


Fig. 12 Lambda (λ) plot obtained using SOM–DBSCAN for **a** Afghanistan **b** California **c** Himalaya **d** Indonesia **e** Japan and **f** Chile

The remaining events with COV_B approximately equal to 1 for all the catalogs are treated as random occurring BGs. COV_B indicates the exponential distribution of the

inter-event times of BGs. Further, the COV_B value mentioned in Table 3 justifies that BGs is approximately equal to unity. It reveals that the BGs are generated

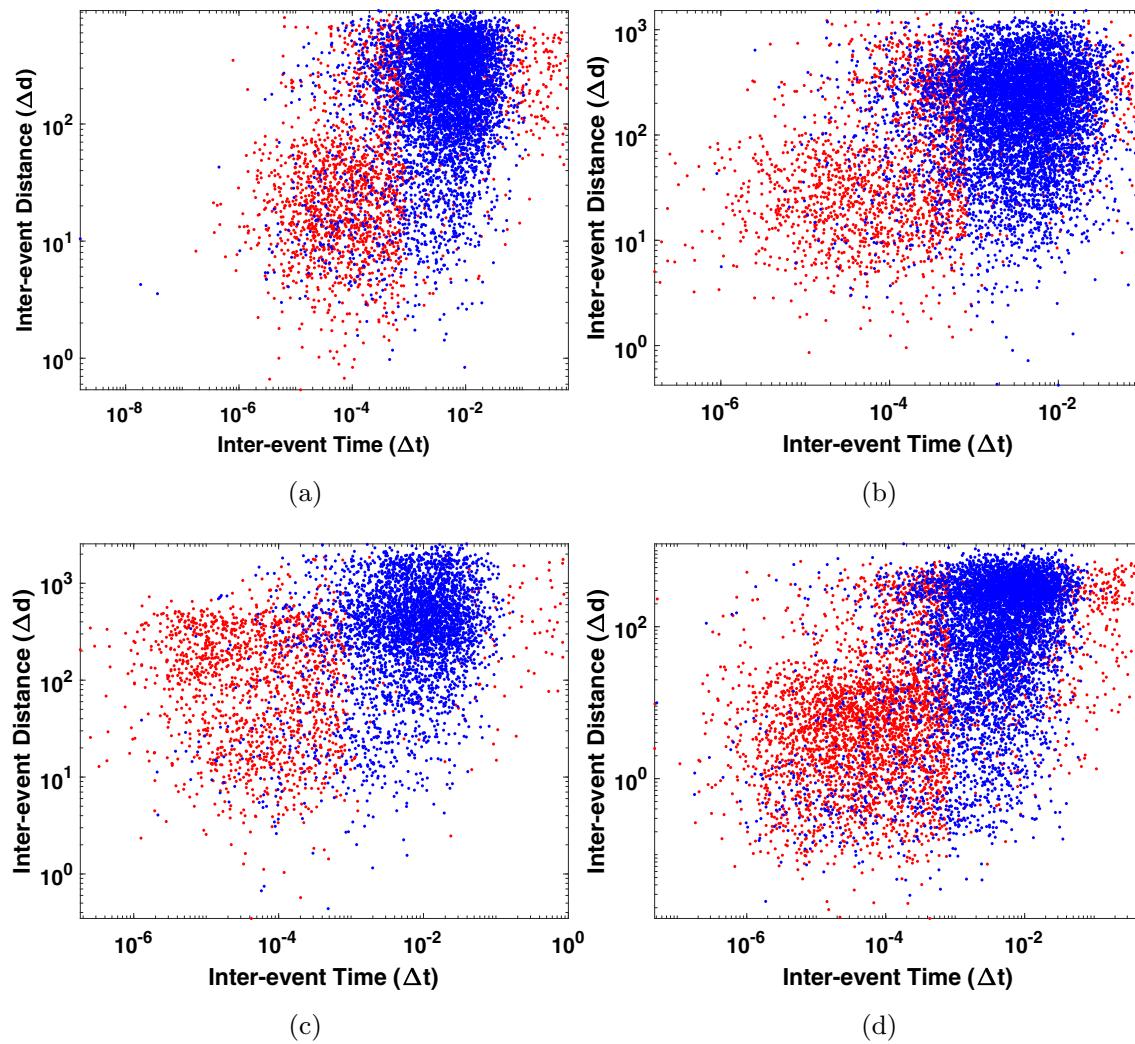


Fig. 13 IET vs IED plot on Logarithmic scale for **a** Taiwan **b** Afghanistan **c** California and **d** Himalaya

Table 3 Classification results obtained from the proposed SOM-DBSCAN

| Results | Catalogs | | | | | | |
|-------------------|----------|-------------|------------|----------|-----------|--------|--------|
| | Taiwan | Afghanistan | California | Himalaya | Indonesia | Japan | Chile |
| Total events (N) | 7788 | 9809 | 9998 | 17770 | 15252 | 13342 | 5288 |
| aftershocks (AFs) | 2545 | 4120 | 5112 | 8698 | 6287 | 6940 | 1842 |
| background (BG)s | 5243 | 5689 | 4886 | 9072 | 8965 | 6402 | 3446 |
| Clusters | 230 | 210 | 196 | 247 | 236 | 216 | 165 |
| COV_T | 1.4145 | 1.5147 | 1.7173 | 1.5469 | 1.3674 | 1.6525 | 1.6106 |
| COV_B | 1.0021 | 1.1561 | 1.0519 | 1.1354 | 1.0908 | 1.0755 | 1.0962 |
| COV_A | 5.6993 | 5.3763 | 5.0338 | 6.0210 | 4.1117 | 5.5895 | 6.2667 |

randomly in time domain and characterized by a homogeneous Poisson process generated by the exponential distribution.

4.4 Interevent time versus interevent distance plot (IET vs. IET plot)

The interevent time (Δt) and interevent distance (ΔD) between the two consecutive earthquake events for time

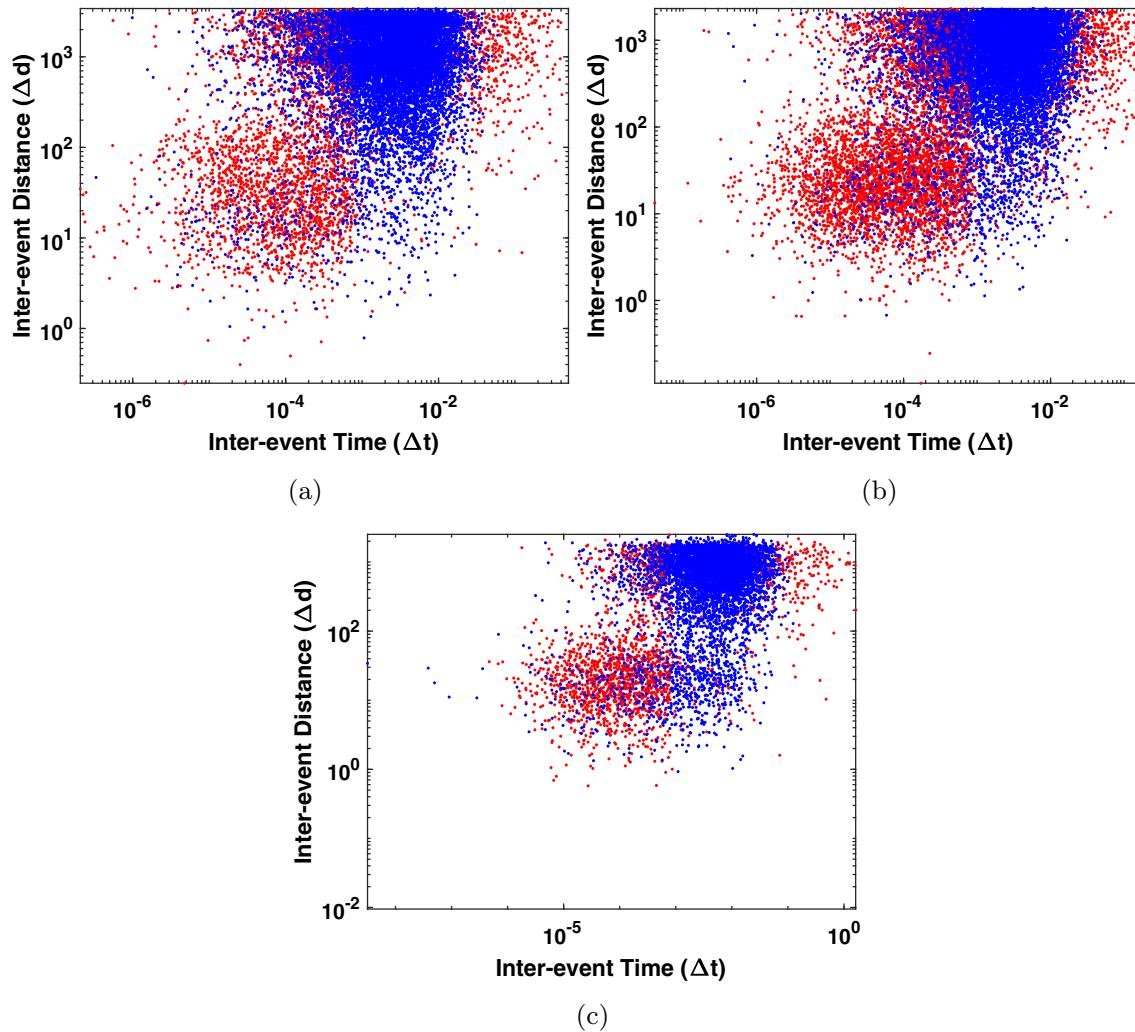


Fig. 14 IET vs IED plot on Logarithmic scale for **a** Indonesia **b** Japan and **c** Chile

interval (T) [70] provides the efficiency of declustering of the proposed model. A logarithmic scatter plot between (Δt) and (ΔD) for all seven catalogs used in the analysis is shown in Figs. 13 and 14.

This Figs. 13 and 14 represents the existence of clustered and background seismicity in the catalogs. A high dense region located at the lower part of the figure (red dots) shows the presence of the clustered seismicity (AFs) in the catalog. The BGs are represented at the upper part with blue dots having high (Δt) . The clear segregation of AFs and BGs concerning the IET and IED plot shows the potential of the proposed model.

4.5 Spatial clustering analysis of earthquake catalogs using m -Morisita index

This section presents an approach for the assessment and investigation of spatial clustering with the m -Morisita index. Here the comparative analysis of all seismic events

of Afghanistan, California, Himalaya, Indonesia, Japan, Chile and Taiwan regions is done using $\log(I_{m,\delta}) - \log(\delta)$ plots for $m = 2, 3, 4, 5$. The results obtained for the total events (TE), AFs, and BGs, are displayed in Figs. 15, 16 and 17. It is observed from the results that with the m -Morisita index is highly sensitive to m value and increases rapidly to the change in features of the seismic pattern. It is also revealed that for every seismic catalog, Morisita index ($I_{m,\delta}$) values for AFs are higher than the Morisita index ($I_{m,\delta}$) values of BGs with a fixed number of point (m). It shows that the presence of AFs increases spatial clusters in the seismic catalog. All the AFs sequences follow this same pattern, as shown in Figs. 15, 16 and 17. It is obvious from Figs. 15, 16 and 17 that the value of m -Morisita index for AFs, BGs and Total catalog follows $(I_{m,\delta})^{AFs} > (I_{m,\delta})^{TE} > (I_{m,\delta})^{BGs}$. In addition, It is visible from the figure that a least-square line well fits curves in the spatial scales. Thus

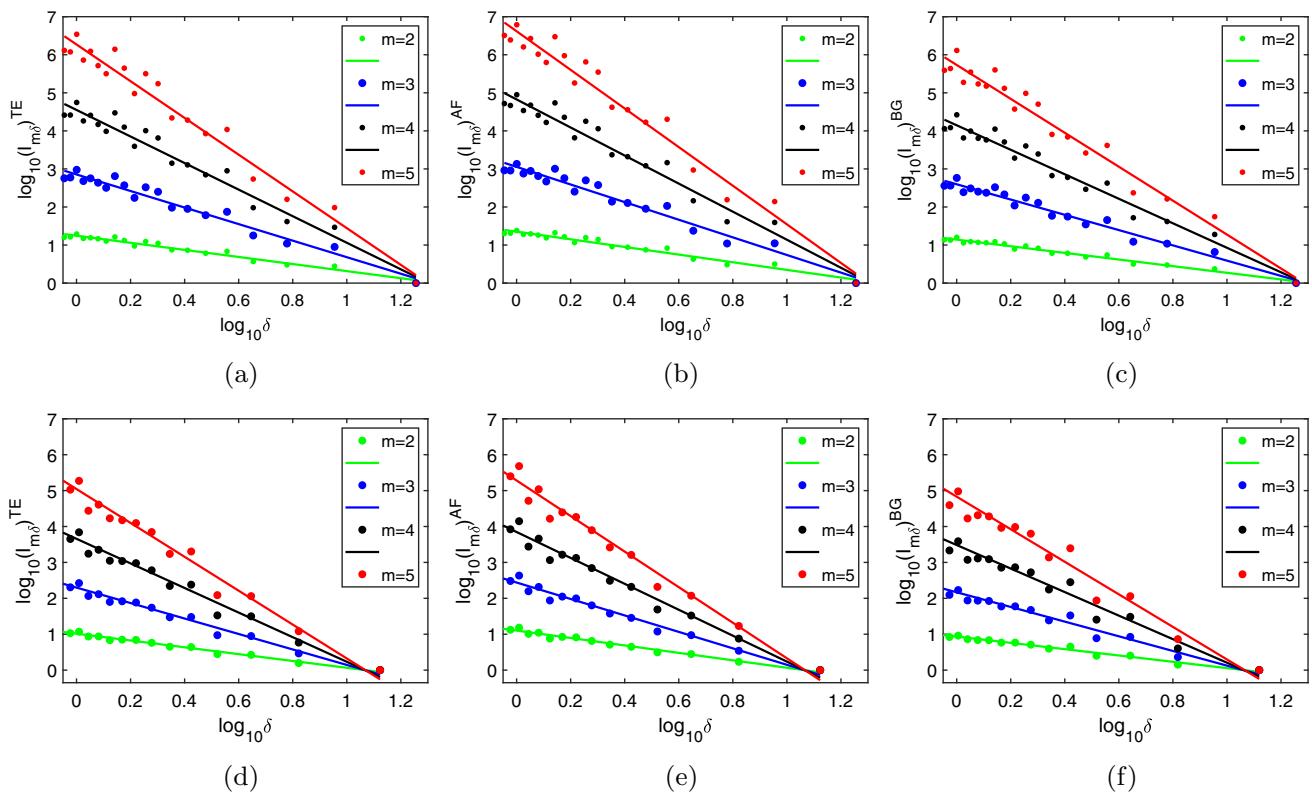


Fig. 15 m -Morisita index for total events, AFs and BGs on log-log scale for Afghanistan **a–c** and California **d–f** respectively

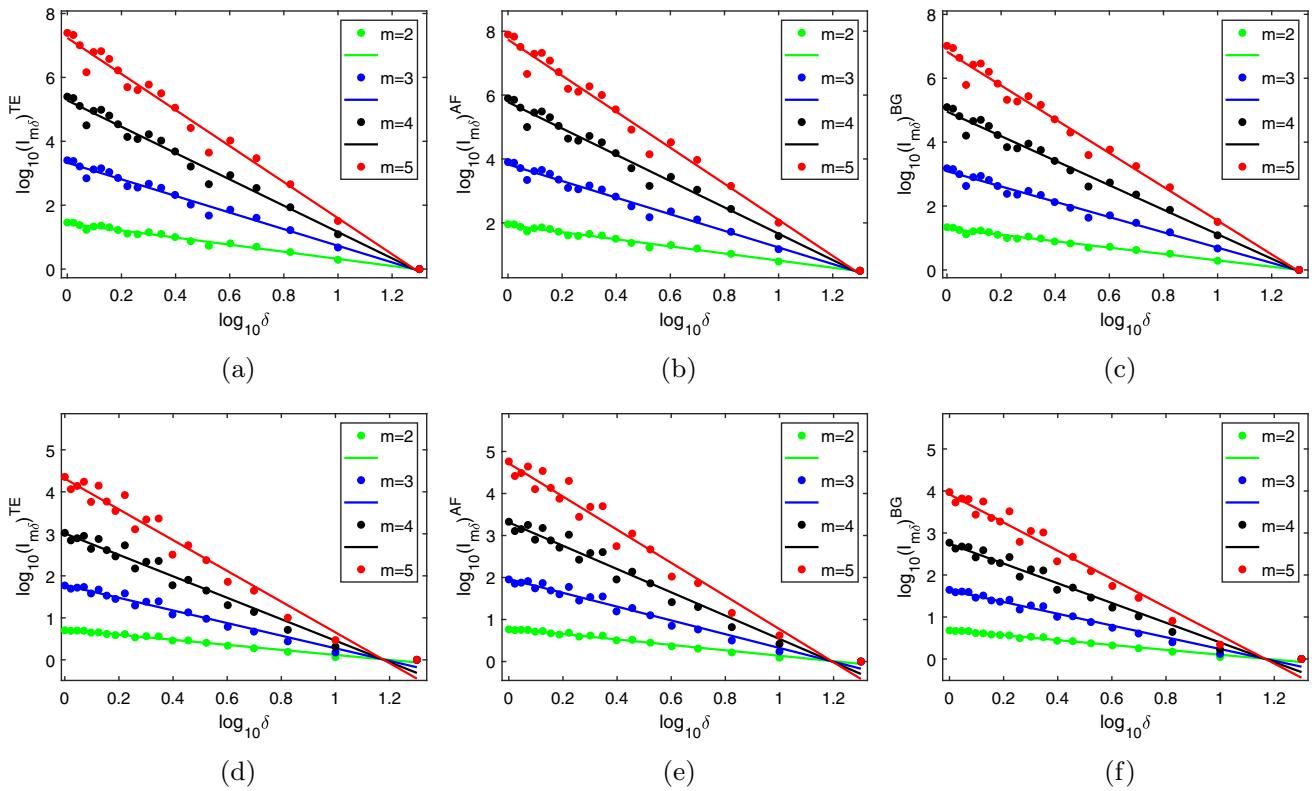


Fig. 16 m -Morisita index for total events, AFs and BGs on log-log scale for Himalaya **a–c** and Indonesia **d–f** respectively

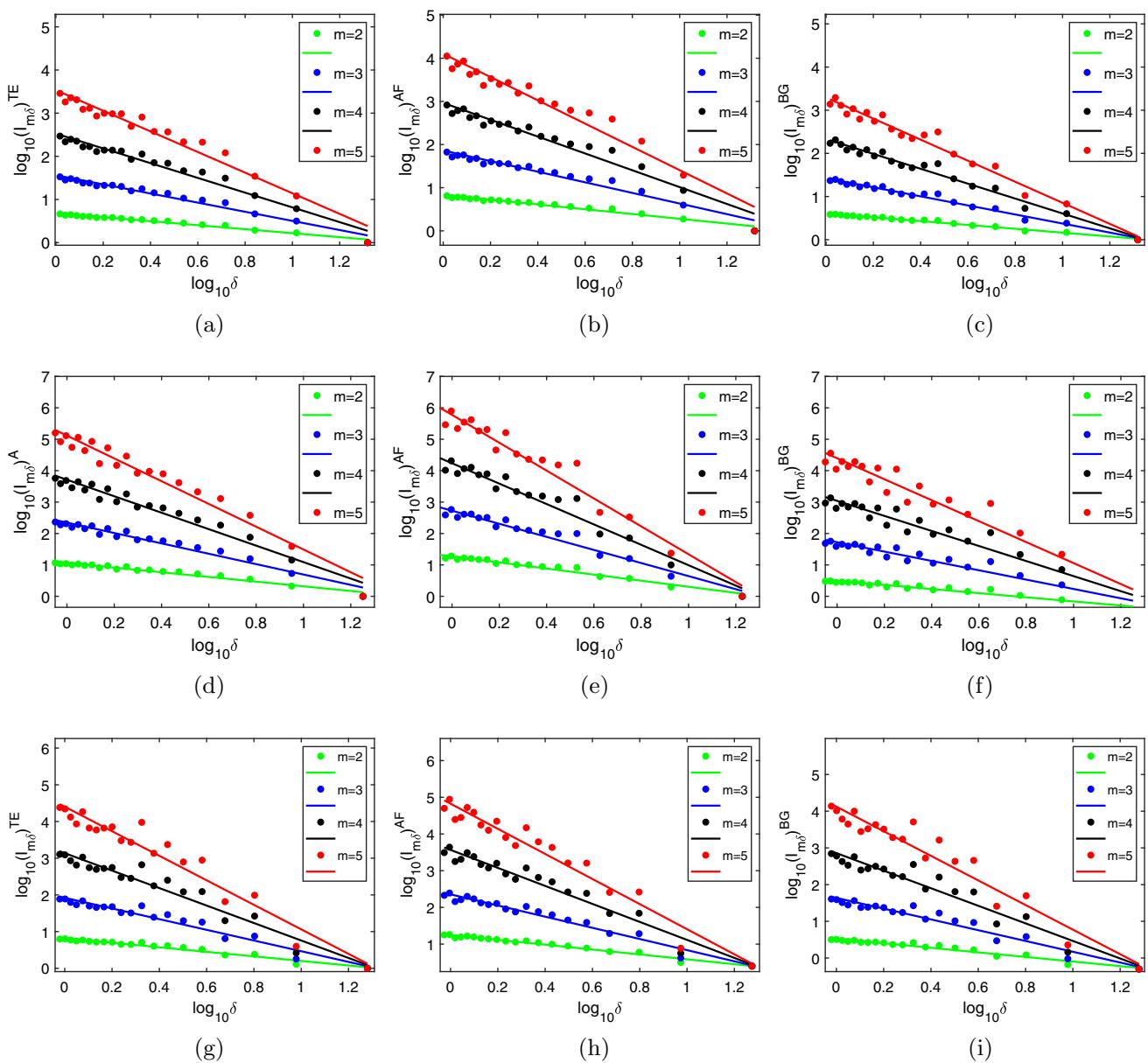


Fig. 17 m -Morisita index for total events, AFs and BGs on log-log scale for Japan **a–c** and Chile **d–f** and Taiwan **g–i** respectively

the plots demonstrate that AFs play a major contribution to spatial clustering

5 Performance analysis of proposed SOM-DBSCAN method with benchmark declustering techniques

In this section, the performance analysis of the proposed SOM-DBSCAN model has been compared with benchmark declustering methods. The benchmark declustering methods, Gardner and Knopoff [8], Gruenthal [59], and Urhammer methods [58] are windowing techniques that

identify the clusters of events by calculating the space-time distance between them. For the given window, an event with the highest magnitude is considered a mainshock, and events within the space-time window of the mainshock are removed. For each event in the earthquake catalog with magnitude m , the subsequent event is determined if it falls within a specific distance interval. In the case of window techniques proposed by Gardner and Knopoff, Gruenthal and Urhammer, the number of AFs obtained largely depends on the approximation of window size parameters. These algorithms are straightforward to implement. Usually, these algorithms do not distinguish between direct and indirect AFs, i.e., 1st-generation of AFs and AFs of AFs.

Table 4 Comparative analysis of proposed SOM–DBSCAN method with Benchmark declustering techniques

| | Methods | Catalog Total events | Taiwan 7788 | Afghanistan 9809 | California 9998 | Himalaya 17770 | Indonesia 15252 | Japan 13342 | Chile 5288 |
|---------------------|----------------|----------------------|-------------|------------------|-----------------|----------------|-----------------|-------------|------------|
| Gardner and Knopoff | AFs | 6505 | 8444 | 8112 | 14900 | 12227 | 10715 | 4229 | |
| | BGs | 1283 | 1364 | 1885 | 2870 | 3025 | 2626 | 1059 | |
| | Clusters | 749 | 846 | 834 | 1613 | 2097 | 1280 | 503 | |
| | Gruenthal | AFs | 7284 | 9142 | 8914 | 16250 | 13745 | 11914 | 4723 |
| | Window | BGs | 504 | 666 | 1083 | 1520 | 1506 | 1427 | 565 |
| | | Clusters | 621 | 703 | 725 | 1436 | 1802 | 1209 | 493 |
| | | Uhrhammer | AFs | 3999 | 5018 | 5891 | 7922 | 6853 | 8367 |
| | | BGs | 3789 | 4790 | 4106 | 9848 | 8398 | 4974 | 2220 |
| | | clusters | 605 | 813 | 833 | 1290 | 1239 | 731 | 1603 |
| Reasenberg | AFs | 1488 | 1141 | 4342 | 2230 | 1934 | 3523 | 1169 | |
| | BGs | 6300 | 8667 | 5655 | 15540 | 13317 | 9818 | 4118 | |
| | clusters | 310 | 178 | 724 | 398 | 424 | 612 | 210 | |
| Tetra Stage Model | AFs | 2885 | 5670 | 6078 | 10452 | 7212 | 6958 | 2257 | |
| | BGs | 4903 | 4139 | 3920 | 7318 | 8040 | 6384 | 3031 | |
| | Clusters | 256 | 226 | 280 | 315 | 356 | 278 | 219 | |
| | Proposed Model | AFs | 2545 | 4120 | 5112 | 8698 | 6287 | 6940 | 1842 |
| | | BGs | 5243 | 5689 | 4886 | 9072 | 8965 | 6402 | 3446 |
| | | Clusters | 230 | 210 | 196 | 247 | 236 | 216 | 165 |

Reasenberg [15] developed a linked method that identifies the seismic clusters according to spatio-temporal distances, known as interaction zones. The event within the previous event's interaction zone is considered an aftershock. The size of space-time interaction zones is defined according to Omori's law and stress redistribution among the earthquakes. Reasenberg method also proved that BGs are generated with the Poisson process. The major limitation of benchmark declustering techniques is that they are highly dependent on the default parameter values of the proposed method. The Tetra stage model [25] gives better performance but identifies fewer clusters. In this manuscript, benchmark declustering techniques with a pre-defined set of parameters of space-time window sizes have been applied to instrumental catalogs of Afghanistan, California, Himalaya, Indonesia, Japan, and Chile regions and compared with the proposed SOM–DBSACN method of classified AFs, BGs, and number of clustered obtained. The GK, Gruenthal, and Uhrhammer methods exaggerate the number of AFs, BGs and clusters obtained, as shown in Table 4.

The GK and Gruenthal window methods consider the finite space-time range of AFs and estimate the low background seismicity rate for shorter inter-event time and strong time clustering. If the AFs sequence follows the Omori type decay, Gardner and Knopoff only remove events close in space and time, overestimate the clusters in the short-range, and underestimate the cluster in the long

range. The Raesenberg method follows the Omori-type AFs decay and presents fast decay in time clusters as a power-law decay at short inter-event times and saturates at long inter-event times. The significant difference in identifying the number of clusters and AFs and BGs, reported in Table 4, is due to the AFs model's assumptions.

The proposed model in the paper comprises two stages. In the first stage, the prime objective of SOM is to understand the topological view of the partitioning of the data in the spatial domain. SOM also reduces the input space and establishes the DBSCAN in the temporal domain. DBSCAN is difficult to apply in a large dataset. After obtaining the BMU in the first phase, DBSCAN is applied to find the Temporal clusters in the catalog. The proposed model finds fewer clusters, as shown in Table 4 because it does not depend on any prior assumptions. All these results show the effectiveness of the proposed SOM–DBSCAN model compared to other benchmark techniques.

6 Conclusion

This paper achieved SOM-based spatial zoning for all the catalogs that identified major seismic sources in the regions listed in the catalog. After that, density-based clustering is applied effectively according to the time information of each event. This clustering phase discriminates the present AF clusters and BG distribution across locations and

timescales. The spatio-temporal analysis of total AFs and BGs are carried out using COV and m -Morisita index. The performance of the proposed model is illustrated in terms of epicenter plot, space-time plot, λ , and cumulative plot, SOM Hits, and Neurons spatial distribution in a spatial region. The relation $\text{COV}_A > \text{COV}_T > \text{COV}_B$ and $I_{m,\delta}^{\text{AF}} > I_{m,\delta}^T > I_{m,\delta}^{\text{BG}}$ obtained from the proposed approach for the catalog of Afghanistan, California, Himalaya, Indonesia, Japan and Chili depicting the clustering efficiency of the used approach. Thus, this method has systematically identified BG clusters with homogeneous time distribution and AF clusters with distinct spatio-temporal clusters.

Acknowledgements This research work is funded by Ministry of Electronics and Information and IT under Visvesvaraya PhD Scheme with Grant Number-1000110674 and unique awardee number MEITY-PHD-2952.

Data availability The datasets (seismic catalogs) analysed during the current study are available at the official website of United State Geological Survey [<https://earthquake.usgs.gov/earthquakes/search/>].

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ben-Zion Y (2008) Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Rev Geophys* 46:RG4006. <https://doi.org/10.1029/2008RG000260>
- Kisslinger C (1996) Aftershocks and fault-zone properties. *Adv Geophys* 38:1–36
- Utsu T, Ogata Y et al (1995) The centenary of the Omori formula for a decay law of aftershock activity. *J Phys Earth* 43(1):1–33
- Shcherbakov R, Turcotte DL, Rundle JB (2005) Aftershock statistics. *Pure Appl Geophy* 162(6):1051–1076
- Ellsworth W (2019) From foreshocks to mainshocks: mechanisms and implications for earthquake nucleation and rupture propagation. *Mech Earthq Faulting* 202:95
- Ellsworth WL, Giardini D, Townend J, Ge S, Shimamoto T (2019) Triggering of the Pohang, Korea, earthquake (mw 5.5) by enhanced geothermal system stimulation. *Seismol Res Lett* 90(5):1844–1858
- Seif S, Zechar JD, Mignan A, Nandan S, Wiemer S (2019) Foreshocks and their potential deviation from general seismicityforeshocks and their potential deviation from general seismicity. *Bull Seismol Soc Am* 109(1):1–18
- Gardner J, Knopoff L (1974) Is the sequence of earthquakes in southern California, with aftershocks removed, poissonian? *Bull Seismol Soc Am* 64(5):1363–1367
- Ellsworth WL, Llenos AL, McGarr AF, Michael AJ, Rubinstein JL, Mueller CS, Petersen MD, Calais E (2015) Increasing seismicity in the us midcontinent: implications for earthquake hazard. *Lead Edge* 34(6):618–626
- Hammond W, Kreemer C, Zaliapin I, Blewitt G (2019) Drought-triggered magmatic inflation, crustal strain, and seismicity near the long valley caldera, central walker lane. *J Geophys Res Solid Earth* 124(6):6072–6091
- Johnson CW, Fu Y, Burgmann R (2017) Stress models of the annual hydrologic, atmospheric, thermal, and tidal loading cycles on California faults: Perturbation of background stress and changes in seismicity. *J Geophys Res Solid Earth* 122(12):10–605
- Abolfathian N, Martinez-Garzn P, Ben-Zion Y (2019) Spatiotemporal variations of stress and strain parameters in the San Jacinto fault zone. *Pure Appl Geophys* 176(3):1145–1168
- Dawood HM, Rodriguez-Marek A, Bayless J, Goulet C, Thompson E (2016) A flatfile for the KiK-net database processed using an automated protocol. *Earthq Spectra* 32(2):1281–1302
- Schaefer A, Daniell J, Wenzel F (2016) EGU general assembly conference abstracts, pp EPSC2016–7830
- Reasenberg P (1985) Second-order moment of central California seismicity, 1969–1982. *J Geophys Res Solid Earth* 90(B7):5479–5495
- Reasenberg PA, Jones LM (1989) Earthquake hazard after a mainshock in California. *Science* 243(4895):1173–1176
- Reasenberg P, Jones L (1994) Earthquake aftershocks: update. *Science* 265(5176):1251–1253
- Tibi R, Blanco J, Fatehi A (2011) An alternative and efficient cluster-link approach for declustering of earthquake catalogs. *Seismol Res Lett* 82(4):509–518
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83(401):9–27
- Helmstetter A, Sornette D (2003) Predictability in the Epidemic-Type Aftershock Sequence model of interacting triggered seismicity. *J Geophys Res* 108:2482. <https://doi.org/10.1029/2003JB002485>
- Sornette D, Werner MJ (2005) Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, båth's law, and observed aftershock sequences. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2004JB003535>
- Turcotte DL, Holliday JR, Rundle JB (2007) BASS, an alternative to ETAS. *Geophys Res Lett* 34:L12303. <https://doi.org/10.1029/2007GL029696>
- Holliday JR, Turcotte DL, Rundle JB (2008) Self-similar branching of aftershock sequences. *Phys A Stat Mech Appl* 387(4):933–943
- Nanda SJ, Tiampo KF, Panda G, Mansinha L, Cho N, Mignan A (2013) A tri-stage cluster identification model for accurate analysis of seismic catalogs. *Nonlinear Process Geophys* 20(1):143–162
- Vijay RK, Nanda SJ (2017) Tetra-stage cluster identification model to analyse the seismic activities of Japan, Himalaya and Taiwan. *IET Sig Process* 12(1):95–103
- Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. *J Am Stat Assoc* 97(458):369–380
- Zaliapin I, Gabrielov A, Keilis-Borok V, Wong H (2008) Clustering analysis of seismicity and aftershock identification. *Phys Rev Lett* 101(1):018501
- Bottiglieri M, Lippiello E, Godano C, de Arcangelis L (2009) Identification and spatiotemporal organization of aftershocks. *J Geophys Res* 114:B03303. <https://doi.org/10.1029/2008JB005941>
- Batac R, Kantz H (2014) Observing spatio-temporal clustering and separation using interevent distributions of regional earthquakes. *Nonlinear Process Geophys* 21(4):735–744
- Cho N, Tiampo KF, Bhattacharya PK, Shcherbakov R, Chen C, Li H, Klein W (2010) Declustering seismicity using the Thirumalai-Mountain metric. *4400 NONLINEAR GEOPHYSICS* 2010:NG51A-1195

31. Davidsen J, Gu C, Baiesi M (2015) Generalized Omori–Utsu law for aftershock sequences in southern California. *Geophys J Int* 201(2):965–978
32. Weatherill G, Burton PW (2009) Delineation of shallow seismic source zones using k-means cluster analysis, with application to the Aegean region. *Geophys J Int* 176(2):565–588
33. Zheng YJ, Ling HF, Chen SY, Xue JY (2014) A hybrid neuro-fuzzy network based on differential biogeography-based optimization for online population classification in earthquakes. *IEEE Trans Fuzzy Syst* 23(4):1070–1083
34. Zaliapin I, Ben-Zion Y (2016) A global classification and characterization of earthquake clusters. *Geophys J Int* 207(1):608–634
35. Nanda SJ, Pradhan PM, Panda G, Mansinha L, Tiampo KF (2013) A correlation based stochastic partitional algorithm for accurate cluster analysis. *Int J Sig Imaging Syst Eng* 6(1):52–58
36. Morales-Esteban A, Martinez-Alvarez F, Scitovski S, Scitovski R (2014) A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Comput Geosci* 73:132–141
37. Nanda SJ, Panda G (2014) A survey on nature inspired meta-heuristic algorithms for partitional clustering. *Swarm Evolut Comput* 16:1–18
38. Cho N, Tiampo KF, Mckinnon S, Vallejos J, Klein W, Dominguez R (2010) A simple metric to quantify seismicity clustering. *Nonlinear Process Geophys* 17(4):293
39. Vijay RK, Nanda SJ (2019) A quantum grey wolf optimizer based declustering model for analysis of earthquake catalogs in an ergodic framework. *J Comput Sci* 36:101019
40. Ester M, Kriegel H.P., Sander J, Xu X (1996) In: Proceedings of the second international conference on knowledge discovery and data mining. AAAI Press, KDD'96, pp 226–231
41. Birant D, Kut A (2007) ST-DBSCAN: an algorithm for clustering spatial-temporal data. *Data Knowl Eng* 60(1):208–221
42. Georgoulas G, Konstantaras A, Katsifarakis E, Stylios CD, Maravelakis E, Vachtseos GJ (2013) Seismic-mass density-based algorithm for spatio-temporal clustering. *Expert Syst Appl* 40(10):4183–4189
43. Nanda SJ, Panda G (2015) Design of computationally efficient density-based clustering algorithms. *Data Knowl Eng* 95:23–38
44. Scitovski S (2018) A density based clustering algorithm for earthquake zoning. *Comput Geosci* 110:90–95
45. Schaefer AM, Daniell JE, Wenzel F (2017) The smart cluster method. *J Seismol* 21(4):965–985
46. Cesca S (2020) Seiscloud, a tool for density-based seismicity clustering and visualization. *J Seismol* 24(3):443–457
47. Tanzim SM, Yeasin S, Hussain MA, Tamal TR, Hasan R, Rahman T, Rahman RM (2018) In: Computer science on-line conference. Springer, pp 364–373
48. Kohonen T (1990) The self-organizing map. *Proc IEEE* 78(9):1464–1480
49. Oettli P, Tozuka T, Izumo T, Engelbrecht FA, Yamagata T (2014) The self-organizing map, a new approach to apprehend the Madden–Julian oscillation influence on the intraseasonal variability of rainfall in the southern African region. *Clim Dyn* 43(5–6):1557–1573
50. Huang F, Yin K, Huang J, Gui L, Wang P (2017) Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng Geol* 223:11–22
51. Lopes-Mazzetto JM, Schellekens J, Vidal-Torrado P, Buurman P (2018) Impact of drainage and soil hydrology on sources and degradation of organic matter in tropical coastal podzols. *Geoderma* 330:79–90
52. Roige M, McGeoch MA, Hui C, Worner SP, Kurle C et al (2017) Cluster validity and uncertainty assessment for self-organizing map pest profile analysis. *Methods Ecol Evolut* 8(3):349–357
53. Du H-k, Cao J-x, Xue Y-j, Wang X-j, (2015) Seismic facies analysis based on self-organizing map and empirical mode decomposition. *J Appl Geophys* 112:52–61
54. Allamehzadeh M, Durudi S, Mahshadnia L (2017) Pattern recognition of seismogenic nodes using Kohonen self-organizing map: example in west and south west of Alborz region in Iran. *Earthq Sci* 30(3):145–155
55. Yaghmaei-Sabegh S (2017) A novel approach for classification of earthquake ground-motion records. *J Seismol* 21(4):885–907
56. Rehman K, Burton PW, Weatherill GA (2014) K-means cluster analysis and seismicity partitioning for Pakistan. *J Seismol* 18(3):401–419
57. Konstantaras A, Katsifarakis E, Maravelakis E, Skounakis E, Kokkinos E, Karapidakis E (2012) Intelligent spatial-clustering of seismicity in the vicinity of the hellenic seismic arc. *Earth Sci Res* 1(2):1
58. Uhrhammer R (1986) Characteristics of northern and central California seismicity. *Earthquake Notes* 57(1):21
59. Van Stiphout T, Zhuang J, Marsan D (2012) Seismicity declustering, community online resource for statistical seismicity analysis. <https://doi.org/10.5078/corssa-52382934>
60. Rydelek PA, Sacks IS (1989) Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature* 337(6204):251–253
61. Zuniga FR, Wyss M (1995) Inadvertent changes in magnitude reported in earthquake catalogs: their evaluation through b-value estimates. *Bull Seismol Soc Am* 85(6):1858–1866
62. Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34(4):185–188. <https://doi.org/10.1785/BSSA0340040185>
63. Wiemer S, Wyss M (2000) Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. *Bull Seismol Soc Am* 90(4):859–869
64. Mignan A, Woessner J (2012) Estimating the magnitude of completeness for earthquake catalogs. Community online resource for statistical seismicity analysis, pp. 1–45
65. Zamani A, Nedaei M, Boostani R (2009) Tectonic zoning of Iran based on selforganizing map. *J Appl Sci* 9(23):4099–4114
66. Kohonen T, Oja E, Simula O, Visa A, Kangas J (1996) Engineering applications of the self-organizing map. *Proc IEEE* 84(10):1358–1384
67. Kagan Y, Knopoff L (1980) Spatial distribution of earthquakes: the two-point correlation function. *Geophys J Int* 62(2):303–320
68. Polani D (2002) Self-organizing neural networks. Springer, pp 13–44
69. U.S. Geological Survey. Earthquake lists, maps, and statistic. <https://earthquake.usgs.gov/earthquakes/search/>
70. Vijay RK, Nanda SJ (2019) Shared nearest neighborhood intensity based declustering model for analysis of spatio-temporal seismicity. *IEEE J Sel Top Appl Earth Obs Remote Sens* 12(5):1619–1627
71. Golay J, Kanevski M, Orozco CDV, Leuenberger M (2014) The multipoint Morisita index for the analysis of spatial patterns. *Phys A Stat Mech Appl* 406:191–202

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”). Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com