Pattern Analysis and Applications
https://doi.org/10.1007/s10044-022-01092-1

THEORETICAL ADVANCES

# Earthquake pattern analysis using subsequence time series clustering

Rahul Kumar Vijay[1] · Satyasai Jagannath Nanda[2]

Received: 19 October 2021 / Accepted: 17 June 2022

## Abstract

In this paper, a subsequence time-series clustering algorithm is proposed to identify the strongly coupled aftershocks sequences and Poissonian background activity from earthquake catalogs of active regions. The proposed method considers the inter-event time statistics between the successive pair of events for characterizing the nature of temporal sequences and observing their relevance with earthquake epicenters and magnitude information simultaneously. This approach categorizes the long-earthquake time series into the finite meaningful temporal sequences and then applies the clustering mechanism to the selective sequences. The proposed approach is built on two phases: (1) a Gaussian kernel-based density estimation for finding the optimal subsequence of given earthquake time-series, and (2) inter-event time ($\Delta t$) and distance-based observation of each subsequence for checking the presence of highly correlated aftershock sequences (hot-spots) in it. The existence of aftershocks is determined based on the coefficient of variation (COV). A sliding temporal window on $\Delta t$ with earthquake's magnitude $M$ is applied on the selective subsequence to filter out the presence of time-correlated events and make the meaningful time stationary Poissonian subsequences. This proposed approach is applied to the regional Sumatra-Andaman (2000–2021) and worldwide ISC-GEM (2000–2016) earthquake catalog. Simulation results indicate that meaningful subsequences (background events) can be modeled by a homogeneous Poisson process after achieving a linear cumulative rate and time-independent $\lambda$ in the exponential distribution of $\Delta t$. The relations $COV_a(T) > COV_o(T) > (COV_b(T) \approx 1)$ and $COV_a(d) > COV_o(d) > COV_b(d)$ are achieved for both studied catalogs. Comparative analysis justifies the competitive performance of the proposed approach to the state-of-art approaches and recently introduced methods.

**Keywords** Earthquake time series · Subsequence clustering · Homogeneous Poisson process · Earthquake catalogs · Coefficient of Variation.

## 1 Introduction

A time-series $X = \{x_t | t = 1, 2, \dots n\}$ is a chronologically ordered sequence of values which are recorded over time. In most real-world applications, it is necessary to store and keep the data for a long time interval in the form of a time series. Its analysis helps to extract meaningful statistical information, underlying causes of trends, and identify hidden temporal patterns. Picoli et al. [20] reported a classification method for monitoring agricultural land in Brazil from 2001 to 2016 with the use of Moderate Resolution Imaging Spectro-radiometer (MODIS) time series data. Dad et al. [5] analyzed the climate variability and trends of change in precipitation and temperature on monthly, seasonal, and annual scales for Kashmir Himalaya between 1980 and 2017. Qi et al. [21] analyzed the daily counts of COVID-19 cases in 30 Chinese provinces and reported negative associations of temperature and humidity with COVID-19. They have suggested that countries and regions with low temperature and humidity should pay more attention. Bakker and Schaars [2] introduced a time-series model for solving groundwater flow problems rather than using regular ground-water models. This model measures the time series of heads in an observation well and helps to answer many groundwater queries. Recently, Khan et al. [13, 14] proposed a deep-learning-based novel hybrid architecture: 'AB-Net' and 'CL-Net' to

✉ Rahul Kumar Vijay
vijay.rahul1986@gmail.com; rahulvijay@banasthali.in

Satyasai Jagannath Nanda
nanda.satyasai@gmail.com; sjnanda.ece@mnit.ac.in

[1] Department of Computer Science, Banasthali Vidyapith, Tonk, Rajasthan 304022, India

[2] Department of Electronics and Communication Engineering, Malaviya National Institute of Technology Jaipur, Rajasthan 302017, India

Published online: 17 July 2022

forecast Renewable Energy Generation; and Batteries' State of Health and Power Consumption respectively. The first one uses an auto-encoder and bidirectional long short-term memory (BiLSTM) to intelligently match the power generation with the consumption for efficient energy management. Whereas the second architecture relies on the convolutional long short-term memory (ConvLSTM) and long short-term memory (LSTM) to prevent power shortage and oversupply by doing precise power consumption forecasts. Similarly, time-series analysis are effectively applied in marketing [34], IoT [37], seismic signal processing [4], flood detection [3] and many diversified applications [9, 29].

Nowadays, a clustering mechanism in time series analysis (Known as Time series Clustering; TSC) is an important tool where the sequential data with millions of rows is difficult to visually analyze and understand unusual hidden trends [7, 11, 23]. Similarly, time-series analyses on earthquake data are widely applied to characterize the main features of regional seismicity and to provide useful insights into earthquake dynamics in terms of self-similarity, self-organization, patterns, finite-scaling, and scale-free characteristics [12, 24]. Marsan et al. [16] analyzed the earthquake time series for monitoring the changes in fault loading rates by comparing the data with an earthquake triggering model. They have used the inter-earthquake temporal statistics for estimating time-varying forcing rates with recovery in terms of duration and intensity. Moustra et al. [18] developed an artificial neural network for earthquake prediction by utilizing the time series magnitude data or seismic electric signals. Michas and Vallianatos [17] reported a stochastic model with memory effects to reproduce the temporal scaling characteristics for regional seismicity. Vogel et al. [33] analyzed the earthquake time series based on the information theory approach to observe the mutability effects in the time interval between consecutive quakes over a predetermined magnitude. Kundu et al. [15] reported a method of determining correlations in earthquake time series using complex network analysis by considering each seismic event as a node.

Due to the space-time clustering behavior of earthquakes in terms of foreshock-aftershock (AFs) activities, it is necessary to know when (temporal) and where (spatially) these trends occur in a long earthquake time series. This problem is known as seismicity declustering where the aim is to isolate independent earthquakes (mainshocks/backgrounds) and dependent earthquakes (foreshocks and aftershocks) from the given overall earthquake catalogs of a region. The early method of earthquake declustering by [8] involved a deterministic space and time windows for different magnitude cutoffs to identify clustered AFs and BGs from the earthquake catalogs. The appropriate window sizes are very difficult to select, it varies from case to case and in most cases overestimate the aftershock population. However, several alternative window sizes were reported in later studies van Stiphout et al. [25], Uhrhammer [28]. Events within these windows are considered clustered AFs and the remaining are treated as BGs. Reasenberg [22] proposed a cluster-based approach by a pairing of earthquakes to make clusters according to the extent of spatial and temporal interaction zones. These zones are decided by the stress distribution near the mainshock for spatial bound and Omori's law for time-bound. Both these methods were developed mostly for California and are heavily dependent on the parameters which need to be optimized for better results. Later on, some probabilistic methods of declustering were proposed, but most of them are model-dependent (Epidemic-Type-Aftershock Sequence (ETAS) model) with a wide range of assumptions and techniques [1, 35, 38, 39]. Recently, Zaliapin and Ben-Zion [36] reported a declustering mechanism based on the nearest-neighbor proximity (distance metric) that describes the link between event pairs in the space-time-magnitude domain for declustering the different benchmark earthquake catalogs. The determination of distance metrics in this method is time-consuming, especially after considering smaller magnitude events. This approach is highly dependent on the size of the earthquake catalog. Vijay and Nanda also reported several statistical and swarm-intelligence-based declustering models for Spatio-temporal seismicity analysis of different seismically active regions [30–32].

Due to the limitation of the state-of-art methods and recent approaches, this research work proposes the "earthquake subsequence time series clustering (ES-TSC)" method for categorizing a large length of earthquake time series to generate the meaningful result for earthquake declustering. Here, aim to keep the simplicity of the earthquake declustering algorithm while improving its usability and applicability to various earthquake-prone regions and also minimize the user-dependent threshold or parameter tuning.

The main contributions of this work are the following:

- Temporal density of earthquake time-series are estimated with Gaussian kernel for obtaining the optimal sub-sequences of the earthquake, thus, increasing the quality of clustering approach with the formation of less size distance-metric.
- The Coefficient of Variation (COV) is determined for checking the necessity of clustering mechanism for each sub-sequences or not. It further reduces the computational time of the algorithm.
- A sliding temporal window is used on the selective sub-sequences for filtering the correlated aftershock events.
- This proposed method is applied to Sumatra-Andaman (regional) and ISC-GEM (global) earthquake catalogs to obtain the aftershock and background seismicity in the region.
- Background seismicity is found time stationary with exponential distribution.

The rest of the paper is organized as follows: Sect. 2 describes the detailed procedure of proposed earthquake sub-sequence time series clustering (ES-TSC). Section 3 presents concise information about the earthquake catalogs used in the paper. Results and discussion are carried out in Sect. 4. The comparative analysis is presented in Sect. 5. Section 6 highlights the significant aspects of the proposed work with future scope.

## 2 Earthquake sub-sequence time series clustering (ES-TSC)

The detailed step-wise procedure of the proposed ES-TSC is outlined below:

*Step 1* An earthquake catalog $\mathbf{E}_{N\times D}$ contains primarily five potential feature vectors: occurrence time (DD:MM:YYYY-HH:MN:SS), location (Longitude $\theta$ and latitude $\phi$), magnitude, and depth ($h$) in Km.

$$\mathbf{E}_{N\times D} = \begin{bmatrix} \mathbf{T} \\ \boldsymbol{\theta} \\ \boldsymbol{\phi} \\ \mathbf{M} \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} t_1 & \theta_1 & \phi_1 & m_1 & h_1 \\ t_2 & \theta_2 & \phi_2 & m_2 & h_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_N & \theta_N & \phi_N & m_N & h_N \end{bmatrix} \tag{1}$$

where $D = 5$ is number of feature vector (attributes), each having length $N$ (number of events). The catalog $\mathbf{E}_{N\times D}$ provides an information of earthquake time series $\mathbf{T}$ of length $N$, an ordered sequence of real-valued earthquake temporal data.

$$\mathbf{T} = \{t_1, t_2, \dots t_N\} \tag{2}$$

Then, an earthquake subsequence of length $n$ belongs to time series $\mathbf{T}$ is represented by

$$T_{i,n} = \{t_i, t_{i+1}, \dots t_{i+n-1}\} \text{ where } 1 \le i \le N - n + 1 \tag{3}$$

A meaningful subsequence is an arranged sequence of earthquake events that omits some events without changing the order of remaining events.

*Step 2* A temporal density of the earthquake subsequence time series $T$ of length $n$ is estimated with the following criteria:

$$\rho_n(t) = \frac{1}{n \times h} \sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right) \tag{4}$$

where $K(.)$ is called the kernel function which is a smooth, symmetric function, and $h > 0$ is termed as the smoothing bandwidth, which controls the amount of smoothing. The kernel function smoothes each time event $t_i$ into small density bumps and then sums all these small bumps together to obtain the final temporal density estimate. In this paper, Gaussian function is selected as a kernel which is written as

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \tag{5}$$

*Step 3* The $\rho_n(t)$ in (4) has multi-model characteristics due to variation in temporal density of events which leads to identifying the value and time position of local maxima (peak) and minima w.r.t change in the density level, as follows:

$$[V_{mx}, P_{mx}] = \text{ DensityPeaks } \left[\rho_n(t)\right] \tag{6}$$

$$I_n^t = 1.01 \times \max\left[\rho_n(t)\right] - \rho_n(t) \tag{7}$$

$$[V_{min}, P_{min}] = \text{ DensityPeaks } [I_n^t] \tag{8}$$

where $V_{min} = (v_{min}^1, v_{min}^2, \dots v_{min}^{g+1})$ and $P_{min} = (p_{min}^1, p_{min}^2, \dots p_{min}^{g+1})$ vector represents the density value and time information of the minima present (shown in the Fig. 1 with small triangle) after estimating the density. Similarly, $V_{max} = (v_{max}^1, v_{max}^2, \dots v_{max}^g)$ and $P_{max} = (p_{max}^1, p_{max}^2, \dots p_{max}^g)$ is the density value and time information of present local maxima in the estimated density (Shown by red filled stars in Fig. 1).

*Step 4* According to the time location of successive minima $P_{mn}$ as shown in Fig. 1 (see the small triangle), an earthquake time series $T$ is divided into successive subsequences $G_g$:

$$T = \{G_1, G_2, \dots G_g\}, \tag{9}$$

where each $j^{th}$ subsequence has length $n_j$ and $g$ represents the number of obtained subsequences (Fig. 1):

$$G_j = \{t_{k-1}, t_k, t_{k+1}, \dots, t_{n_j}\} \text{ where } j = 1, 2, 3, \dots g \tag{10}$$

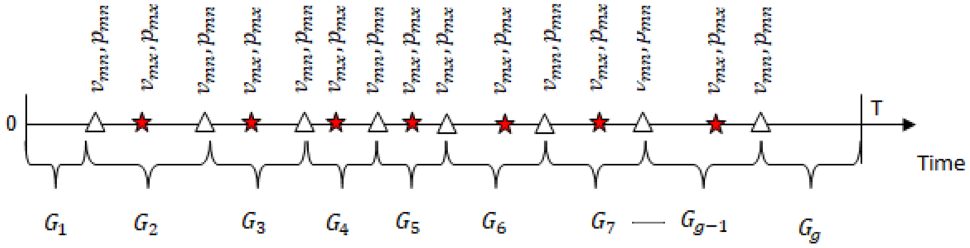


**Fig. 1** Formation of finite earthquake sub-sequences ($T_{i,n}$) from a time series $\mathbf{T}$ based on the information about local maxima and minima in the estimated temporal density $\rho_n(t)$

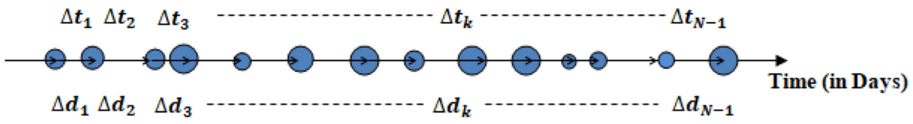

$\Delta t_1$ $\Delta t_2$ $\Delta t_3$
$\Delta t_k$
$\Delta t_{N-1}$
Time (in Days)
$\Delta d_1$ $\Delta d_2$ $\Delta d_3$
$\Delta d_k$
$\Delta d_{N-1}$

**Fig. 2** Concept of inter-event time and distance between the successive events

*Step 5* After obtaining each subsequence $G_j$ from the long time series $\mathbf{T}$ each having length $n_j$. Some of the events are removed from the selective $j^{th}$ subsequence $G_j$ to make the meaningful subsequences as per the following criterion:

- Criterion 1: A subsequence is considered meaningful which follows the homogeneous Poisson process (HPP) in time. These have a uniform arrival rate (average) with Poisson distribution and comprised of random regular (Background-BGs) earthquake events in a region.
- Criterion 2: The subsequences also have the correlated patterns (clustered foreshock-aftershocks) in time, which are generated due to the occurrence of relevant mainshock (high magnitude event). The removal of these patterns is necessary to fulfill criterion 1. These events belong to the non-homogeneous Poisson process where their average rate of arrivals is varied w.r.t. time. These are the hot spots and more hazardous compare to BGs.

```
Algorithm 1: Sliding window on $\Delta t$ for filtering of events from $G_j$
Require: Input: Earthquake subsequence: $G_j = \{t_{k-1}, t_k, t_{k+1}, ....., t_{N_j}\}$.
1: Parameter Initialization: $W_n$=10; $\Delta t_{cto}$, $M_t = 6.5$
2: Calculate ($\Delta t$): Determine the events which have short time interval $\Delta t$ by applying
   the appropriate threshold $\Delta t_{cto}$
3: Check $m_j$: Presence of the mainshock of the events based on the magnitude $m_j \geq M_t$
   for short $\Delta t$ events.
4: Consider: If events have short $\Delta t$ and also have a mainshock then remove the events
   from the sub-sequence from $G_j$
5: Categorization: short $\Delta t$ sub-sequence along with a mainshock are treated as time
   correlated AFs and remaining are considered the part of the background seismic activities
6: Determine: BGs are justified and validated using $COV(T)$, if not meet then change the
   threshold and repeat the procedure(1-4)
7: Criterion-1: Obtain the meaningful sub-sequence:- BGs
8: Criterion-2: Obtain the removed time-correlated:- AFs
```

*Step 6* A parameter called: Coefficient of Variation ($COV(T)$) in time domain is used to justify the meaningfulness of each subsequence $G_j$ and to satisfy the criterion mentioned in step-5. It is the standard deviation normalized by the mean of inter-event times (distance) of the successive events (sequence) of a given time interval (coordinates). The inter-event times ($\Delta t$) is defined as

$$\Delta t = t_{i+1} - t_i, \quad \forall i = 1, 2, \ldots N - 1 \in T \tag{11}$$

then, $COV(T)$ is calculated as follows:

$$COV(T) = \frac{\sqrt{E\left[\Delta t^2\right] - (E[\Delta t])^2}}{E[\Delta t]} \tag{12}$$

where $E(.)$ represents an average (mean) of the given quantity. The concept of inter-event time and distance are shown in Fig. 2.

The $COV(T)$ discriminates important characteristics of the subsequence (in the time domain) in three different ways: (1) If the subsequences are periodic, then $\Delta t$=constant and $COV(T)$ should be zero. (2) If the sequences follow a Poisson distribution, then $\Delta t$ has exponential behavior, and $C_v$ should be around 1 (Criterion-1). These events are said to be random in time (uniform arrival rate). (3) If $COV(T) > 1$, then the process follows a power-law distribution with non-homogeneous Poisson characteristics (Criterion-2).

Here, $COV(T)$ is determined for each subsequence to find out the temporal characteristics of the $G_j$ as mentioned in the above paragraph. Those subsequences are selected for the next phase which includes the time-correlated events as determined by the $COV(T)$. If subsequence has a $COV(T)$

nearly equal to 1, then it is not selected for subsequence clustering in the II phase due to the absence of time-correlated events in it. Otherwise, time-correlated events in each subsequence are filtered out for making the meaningful subsequences (uniform arrival rate, independent random sequence). In the next phase, the objective is to identify and remove the events from selective subsequence to make $COV(T) \approx 1$ (Criteria 1 for meaningful subsequences). The remaining sequences (which are filtered) are strongly related in time, follow Criterion-2 with $COV(T) > 1$.

*Step 7* For the next phase of subsequence clustering, normalized inter-event times $\Delta t$ between the successive events $\in G_j$ and magnitude information is used by taking the sliding window approach. An overlapping window of ten events is chosen in $\Delta t$ and observes the magnitude (presence of at least one high-intensity quake, $\geq 7$). If normalized average $\Delta t$ is less than the pre-defined threshold (i.e. short $\Delta t$) then remove events from the selective subsequences. Otherwise, remains the part of subsequence. The step-by-step procedure is outlined in Algorithm 1.

## 3 Earthquake catalog used in the simulation analysis

This proposed method is employed on the regional earthquake data of Sumatra-Andaman downloaded from Northern California Earthquake Data Center [19] and ISC-GEM global instrumental catalog (version 7.0) obtained from International Seismological Centre [10]. The parameter and their range are highlighted in Table 1. Here, for Sumatra-Andaman, magnitude completeness $M_c$ is taken 4.5 which is determined by fitting the magnitude data on Gutenberg-Richter relation. The details of both the catalog are as follows:

1. *Sumatra-Andaman* Sumatra-Andaman region is highly susceptible to tsunamis followed by earthquakes, especially from the year 2002 on-wards, which are responsible for millions of deaths, infrastructure damages, and long-lasting wounds to the civilizations. This region also got public attention worldwide, when one of the most powerful and destructive 2004 boxing day earthquakes in the Indian ocean occurred with a magnitude around 9.3 at Richter scale. Indian ocean tsunami in the year 2004 was recorded second largest on a seismograph, termed as Sumatra-Andaman earthquake. It ruptured the greatest fault length of any recorded earthquake and triggered a series of tsunamis (that was up to 30m high along the northern coast of Sumatra), killing up to 280,000 people in 14 countries, and even displaced earth's north pole by 2.5cm. Most of the events are shallow depth in this catalog. The distribution of earthquake epicenters along with the event's depth (as shown by different colors) in the region is shown in Fig. 3a. This region is highly vulnerable to seismic activities with the occurrence of a large number of high-intensity quakes as shown in Fig. 3c and they trigger the many events that result in a hike of the seismic rate at that time (see red line in Fig. 3c).
2. *Global ISC-GEM instrumental Catalog* This catalog was publicly released by International Seismological Centre on 9th April 2020 [10]. The catalog is refined and rebuild by adding the new earthquakes and improving the location and/or magnitude from the previous work [6, 26, 27]. It provides detailed information on earthquakes that occurred world wide in the period from 1904-2016. There is a total of 39,400 earthquake events and all are magnitude greater than 4.9 Mw during 1904–2016. Here, in this study, earthquakes, occurred in the period 2000–2016 are considered and their epicenter distribution is shown in Fig. 3b with depth variation by a different color. A slight change in seismicity rate at the time of a large earthquake is evident in Fig. 3d.

## 4 Results and analysis

The obtained results, their analysis and significance are presented in this section. It is discussed in the following sub-sections:

### 4.1 Density estimation and sub-sequence formation

A temporal density is estimated from the time-series earthquake data of duration $T$ according to the method described in Sect. 3 (as shown in Fig. 4) for both catalogs. From this temporal density estimation, information about local minima $P_{mn}$ and maxima $P_{mx}$ is extracted as shown in black circle and red-square respectively in Fig. 4a and b for Sumatra-Andaman and ISC-GEM world wide catalog respectively. The highest density peak due to the Sumatra-Andaman

**Table 1** Properties of the earthquake catalog

| Catalog | Long | Lat | Period | Magnitude | Depth |
|---|---|---|---|---|---|
| Sumatra-Andaman | 90°–110°E | −10°–20°N | 2000–2021 | 4.5–9.1 | 0.3–651 |
| ISC-GEM | −180°–180°E | −70°–90°N | 2000–2016 | 4.9–9.6 | 0.7–693.1 |

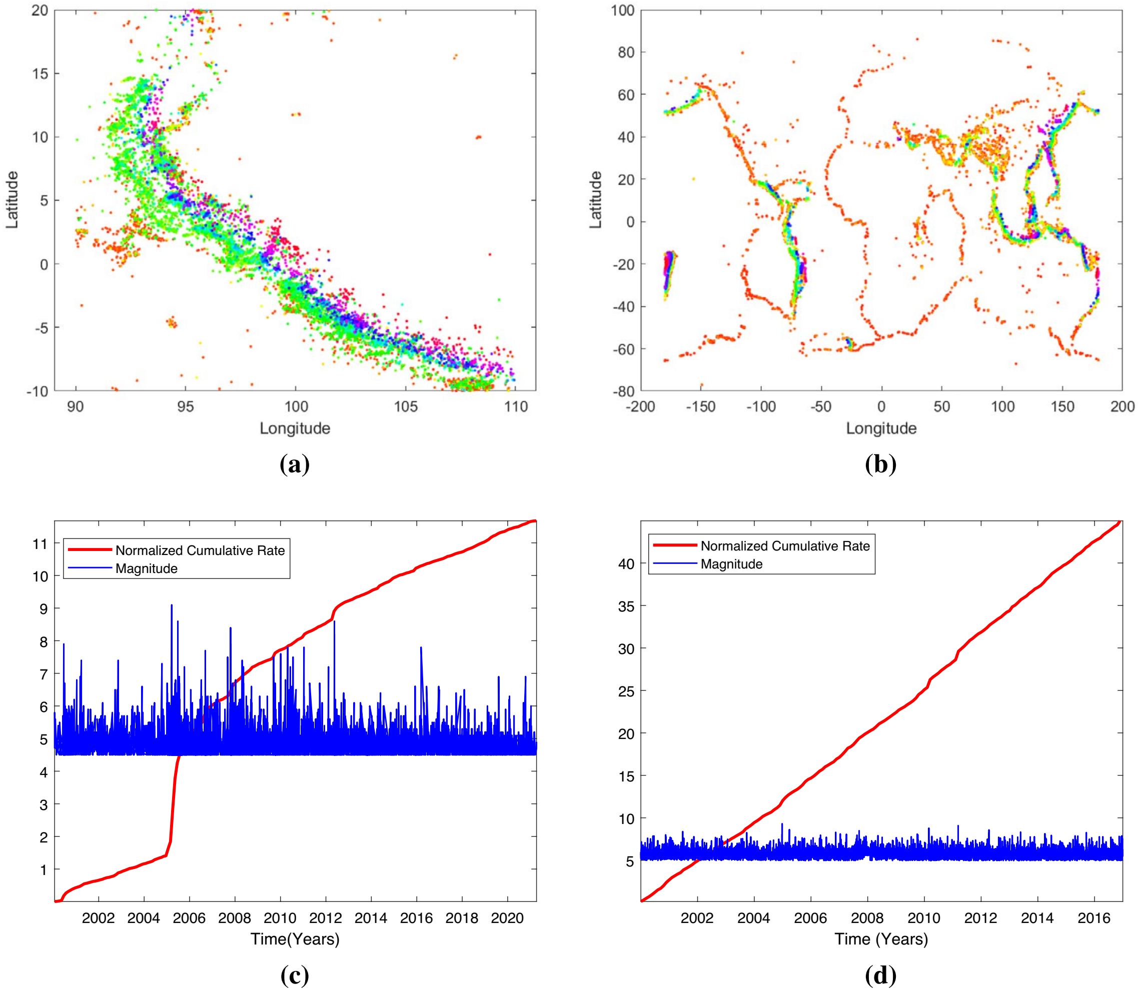


**Fig. 3** Distribution of earthquake epicenters in **a** Sumatra-Andaman region during 2000–2021 and **b** global ISC-GEM earthquakes during 2000–2016. event's depth is shown by different colors in the plots. **c** and **d** Seismicity rate deviation w.r.t. event's magnitude for both catalogs

earthquake and tsunamis (Dec 2004 at the Indian Ocean with 9.4M) is observed as shown in Fig. 4a.

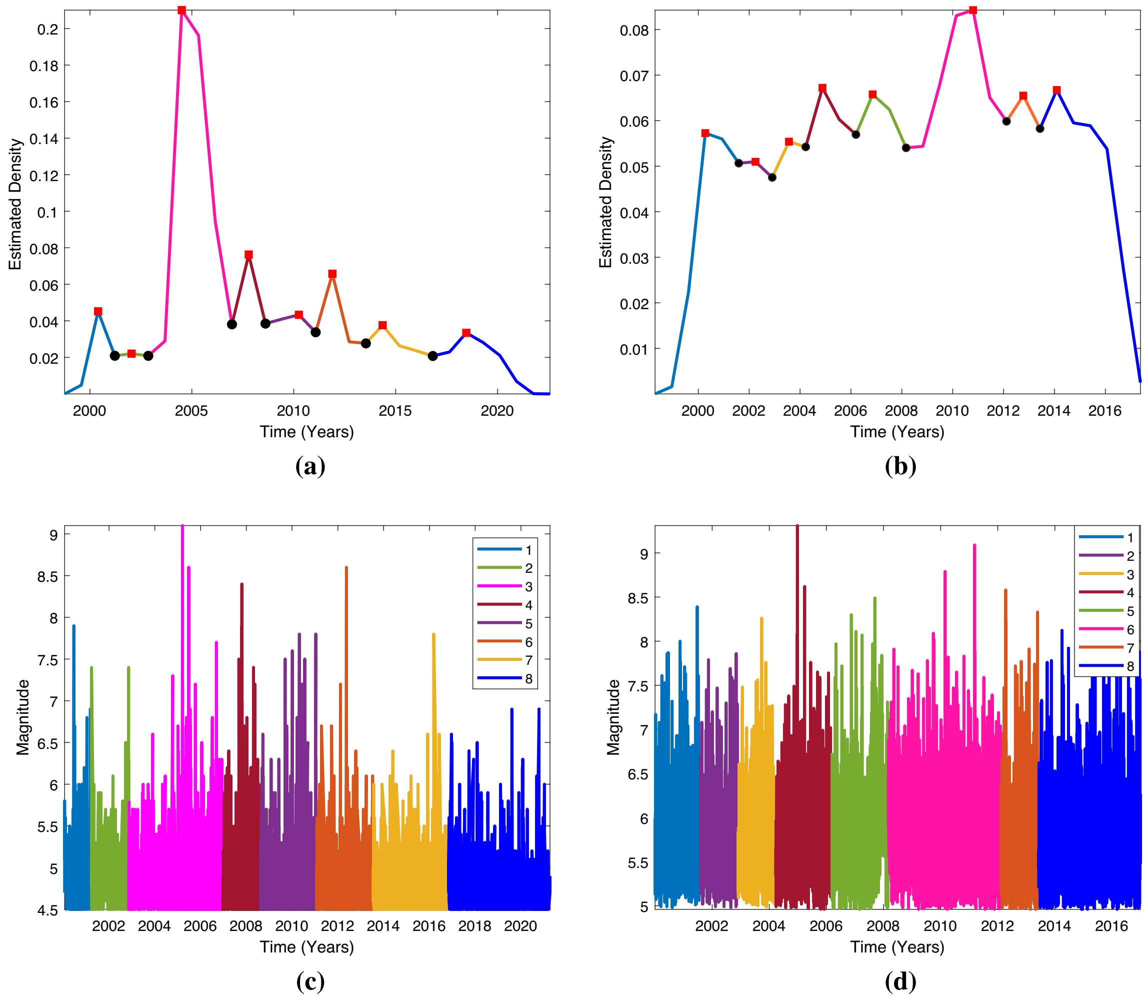
According to the time position of $P_{mn}$ obtained from Fig. 4a and b, the time series is divided into eight subsequences (for both) as shown by different colors. Each subsequence $G_j$where$j = 1...8$ has at least one high magnitude event as observed in Fig. 4c and in Fig. 4d. Tables 2 and 3 highlights the properties of each sub-sequence for Sumatra-Andaman and ISC-GEM earthquake data respectively. Both tables justify the risk factor associated in each subsequence with their duration, no. of events, $COV_o(T)$ value and event's count with magnitude greater than equal to 7. $COV_o(T)$ quantifies the presence of time-correlated events in each subsequence. The $COV_o(T)$ value and temporal density peaks of each subsequence have a strong linear correlation as evident from tables and figures. The largest $COV_o(T)$ in $G_3$ (also have highest density peak) which indicates the presence of large number of correlated events, triggered by the mainshock, occurred at the Indian ocean in 2004. Similarly, subsequence $G_6$ in Table 3 has the highest $COV_o(T)$ with a high-density peak among in all sub-sequence in Fig. 4b.

### 4.2 IET ($\Delta t$) and IED ($\Delta d$) based observation

The inter-event times $\Delta t$ and inter-event distances $\Delta d$ among the two successive pair of earthquake are the simplest way to find the time and space correlation in less computation time. The inter-event distance $\Delta d$ (in Km) is determine as follows:

$$\begin{aligned}\Delta d = R_E \times \cos^{-1}(\sin\phi_i \times \sin\phi_{i+1} + \cos\phi_i \\ \times \cos\phi_{i+1} \times \cos(|(\theta_{i+1} - \theta_i|)))\end{aligned} \tag{13}$$

The coordinates longitude ($\theta$) & latitude ($\phi$) are taken in radians, $R_E = 6371$ Km is the approximate radius of the earth. This definition of $\Delta d$ (in Kms) is based on epicenters and assuming a spherical surface. Both $\Delta t$ and $\Delta d$ avoid the frequent need of large distance matrices having size



(a)

(b)

(c)

(d)

**Fig. 4** Sub-sequence formation: **a** Estimated temporal density for Sumatra-Andaman and for **b** ISC-GEM catalog with eight sub-sequences represented by different colors. Here, black circle and red square indicates the position of local minima $P_{mn}$ and maxima $P_{mx}$ respectively. **c**, **d** Magnitude versus time plot of the formed subsequences for both catalogs

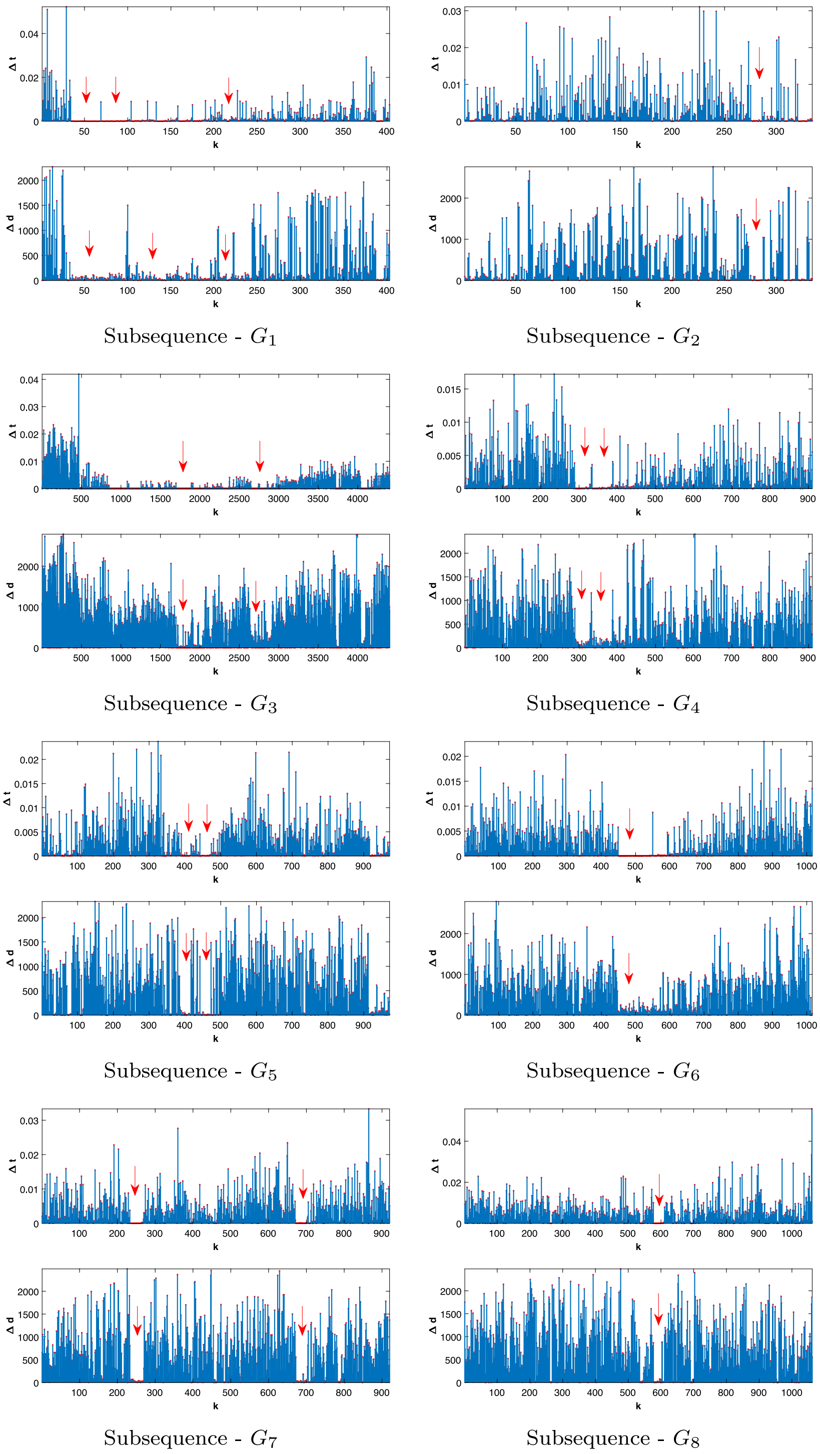
**Table 2** Sub-sequences obtained from Sumatra-Andaman earthquake time-series

| S.No. | Time Period | # Events | $COV_o(T)$ | $COV_o(d)$ | $m_j \geq 7M$ |
|---|---|---|---|---|---|
| $G_1$ | 22/01/2000–22/03/2001 | 404 | 2.01 | 1.51 | 22 |
| $G_2$ | 23/03/2012–10/11/2002 | 335 | 1.27 | 1.10 | 17 |
| $G_3$ | 13/11/2002–20/12/2007 | 4405 | 2.59 | 1.19 | 15 |
| $G_4$ | 23/12/2007–11/08/2008 | 911 | 1.46 | 1.04 | 25 |
| $G_5$ | 14/08/2008–26/01/2011 | 973 | 1.39 | 1.04 | 26 |
| $G_6$ | 28/01/2011–14/07/2013 | 1017 | 1.38 | 1.04 | 65 |
| $G_7$ | 16/07/2013–25/10/2016 | 921 | 1.20 | 0.92 | 22 |
| $G_8$ | 30/10/2016–09/04/2021 | 1063 | 1.24 | 0.90 | 48 |

**Table 3** Sub-sequences obtained from ISC-GEM world wide earthquake time series

| S.No. | Time Period | # Events | $COV_o(T)$ | $COV_o(d)$ | $m_j \geq 7M$ |
|---|---|---|---|---|---|
| $G_1$ | 01/01/2000–04/08/2001 | 931 | 1.17 | 0.85 | 1 |
| $G_2$ | 05/08/2001–27/11/2002 | 690 | 1.08 | 0.82 | 2 |
| $G_3$ | 30/11/2002–22/03/2004 | 697 | 1.10 | 0.84 | 6 |
| $G_4$ | 23/03/2004–12/03/2006 | 1224 | 1.25 | 0.95 | 6 |
| $G_5$ | 14/03/2006–02/03/2008 | 1256 | 1.21 | 0.94 | 7 |
| $G_6$ | 03/03/2008–11/02/2012 | 2761 | 1.26 | 0.97 | 3 |
| $G_7$ | 13/02/2012–07/06/2013 | 855 | 1.24 | 0.91 | 1 |
| $G_8$ | 08/06/2013–30/12/2016 | 2197 | 1.20 | 0.88 | 0 |

**Fig. 5** Space-time correlation of events observed for each subsequence ($G_1 - G_8$) from IET-IED plot for Sumatra-Andaman. Events with short $\Delta t$ and $\Delta d$ are marked with red arrow



Subsequence - $G_1$

Subsequence - $G_2$

Subsequence - $G_3$

Subsequence - $G_4$

Subsequence - $G_5$

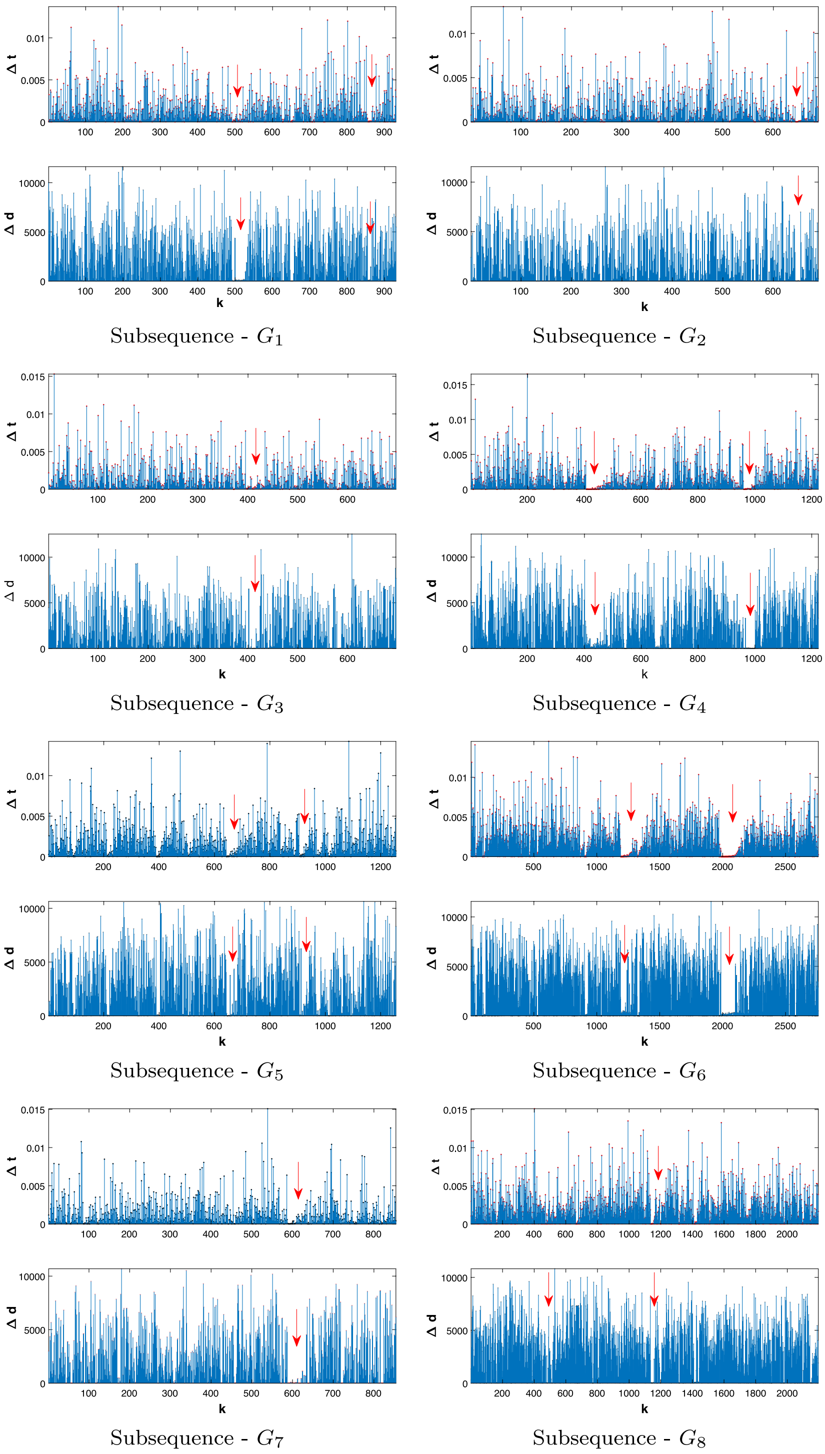
Subsequence - $G_6$

Subsequence - $G_7$

Subsequence - $G_8$

$N \times N$ where $N$ is the number of events ($N$ is large in most cases) for similarity measure. The $\Delta t$ and $\Delta d \ \epsilon \ G_j$ for both catalogs, determined from Eq.(10) and Eq.(13) are shown in Fig. 5 (for Sumatra-Andaman) and Fig. 6 (for ISC-GEM). From the figures, it is observed that some of the events $\in G_j$ have occurred in a very short interval of time (less $\Delta t$ in days) which is indicated by the red arrow. It is also evident that the same events also have short $\Delta d$ after observing both sub-figures in Fig. 5a–h. This observation reveals the consecutive triggering of events at the occurrence time of the large event and hence generation of primary-secondary AFs on the same fault and hike in the seismicity rate. From Figs. 5 and 6, it is also revealed that small or/both large dip occurs at the same $k_{th}$ position in $\Delta t$ and $\Delta d$ for almost all subsequences (see the red arrow positions in the figure). That resembles the space-time correlation theory shown by

**Fig. 6** Space-time correlation of events observed for each subsequence ($G_1 - G_8$) in IET-IED plot for ISC-GEM earthquake catalog. Events with short $\Delta t$ and $\Delta d$ are marked with red arrow



Subsequence - $G_1$

Subsequence - $G_2$

Subsequence - $G_3$

Subsequence - $G_4$

Subsequence - $G_5$

Subsequence - $G_6$

Subsequence - $G_7$

Subsequence - $G_8$

some of the events called: AFs, triggered by mainshock like $G_3$ in Fig. 5c has very less $\Delta t$ ($\Delta d$ as well) for large sample duration due to the occurrence of Indian tsunami in 2004. This strong space-time correlation, observed by $\Delta t$ and $\Delta d$ for both catalogs in Figs. 5 and 6 is filtered out for obtaining the meaningful uncorrelated subsequences in the next phase.

It is evident from the IET-IED plot and corresponding $COV(T)$ for each subsequence that they have correlated primary, secondary AFs (less $\Delta t$ and $\Delta d$ with high $COV$ value) along with independent random BGs. So, in the 2nd phase, a meaningful subsequence $G_j$ is obtained by filtering those correlated primary and secondary AFs from the subsequence according to the procedure presented in algorithm 1.

**Table 4** Outcomes after applying 2nd phase of the proposed method for Sumatra-Andaman

| S.No. | AFs | BGs | $COV_a(T)$ | $COV_b(T)$ | $COV_a(d)$ | $COV_b(d)$ | Clusters |
|---|---|---|---|---|---|---|---|
| $G_1$ | 253 (62%) | 151 | 4.10 | 1.24 | 2.05 | 1.02 | 30 |
| $G_2$ | 104 (31%) | 231 | 3.93 | 1.07 | 1.83 | 0.91 | 22 |
| $G_3$ | 3703 (84%) | 702 | 7.73 | 1.15 | 1.22 | 0.92 | 253 |
| $G_4$ | 551 (60%) | 360 | 3.59 | 1.04 | 1.23 | 0.85 | 86 |
| $G_5$ | 471 (48%) | 502 | 3.64 | 1.08 | 1.37 | 0.82 | 82 |
| $G_6$ | 525 (51%) | 492 | 3.41 | 1.05 | 1.20 | 0.91 | 87 |
| $G_7$ | 302 (32%) | 619 | 3.43 | 1.00 | 1.25 | 0.80 | 64 |
| $G_8$ | 310 (29%) | 753 | 2.58 | 1.10 | 1.16 | 0.81 | 73 |

**Table 5** Obtained results after applying 2nd phase of the proposed method for ISC-GEM

| S.No. | AFs | BGs | $COV_a(T)$ | $COV_b(T)$ | $COV_a(d)$ | $COV_b(d)$ | Clusters |
|---|---|---|---|---|---|---|---|
| $G_1$ | 271 (29%) | 660 | 3.68 | 1.07 | 1.06 | 0.77 | 50 |
| $G_2$ | 151 (21%) | 539 | 3.25 | 0.99 | 1.10 | 0.75 | 28 |
| $G_3$ | 171 (24%) | 526 | 2.89 | 1.01 | 1.00 | 0.80 | 31 |
| $G_4$ | 456 (37%) | 768 | 3.48 | 1.04 | 1.32 | 0.78 | 61 |
| $G_5$ | 450 (35%) | 806 | 3.27 | 1.07 | 1.20 | 0.80 | 68 |
| $G_6$ | 1089 (39%) | 1672 | 4.03 | 1.06 | 1.31 | 0.81 | 147 |
| $G_7$ | 325 (38%) | 530 | 3.05 | 1.02 | 1.18 | 0.81 | 56 |
| $G_8$ | 720 (32%) | 1477 | 3.17 | 1.04 | 1.11 | 0.78 | 123 |

## 4.3 Results after applying the sliding window on $\Delta t$ with mainshock magnitude

A finite length sliding window is used to obtain the meaningful sub-sequence by filtering of the correlated events. If events in a window have short IETs (and IEDs as well) and also have at least one large event (based on the magnitude) in that interval then filter out the event from given subsequence otherwise slide the temporal window further for the remaining part. This procedure achieves meaningful subsequences due to comprised of regular occurred random BG events and filtering out the time-correlated events. They are considered primary and secondary AFs according to their characteristics in soace-time-magnitude domain.

The obtained event's summery are presented in Tables 4 and 5 after applying the proposed method for both catalogs respectively. In both tables, the results of all eight subsequences are mentioned in terms of filtered AF population, meaningful subsequences comprised of BG events, $COV_a(T)$ for AFs, $COV_b(T)$ for BGs, the number of AFs cluster. In Table 4, filtered AFs population from the given sub-sequence indicates the associated risk in that zone and time-duration. The order of risk (hazard) from AFs % for Table 4: $G_3 > G_1 > G_4 > G_6 > G_5 > G_7 > G_2 > G_8$ holds true. Similarly, in Table 5, AF % and the risk order is $G_6 > G_7 > G_4 > G_5 > G_8 > G_1 > G_3 > G_2$.

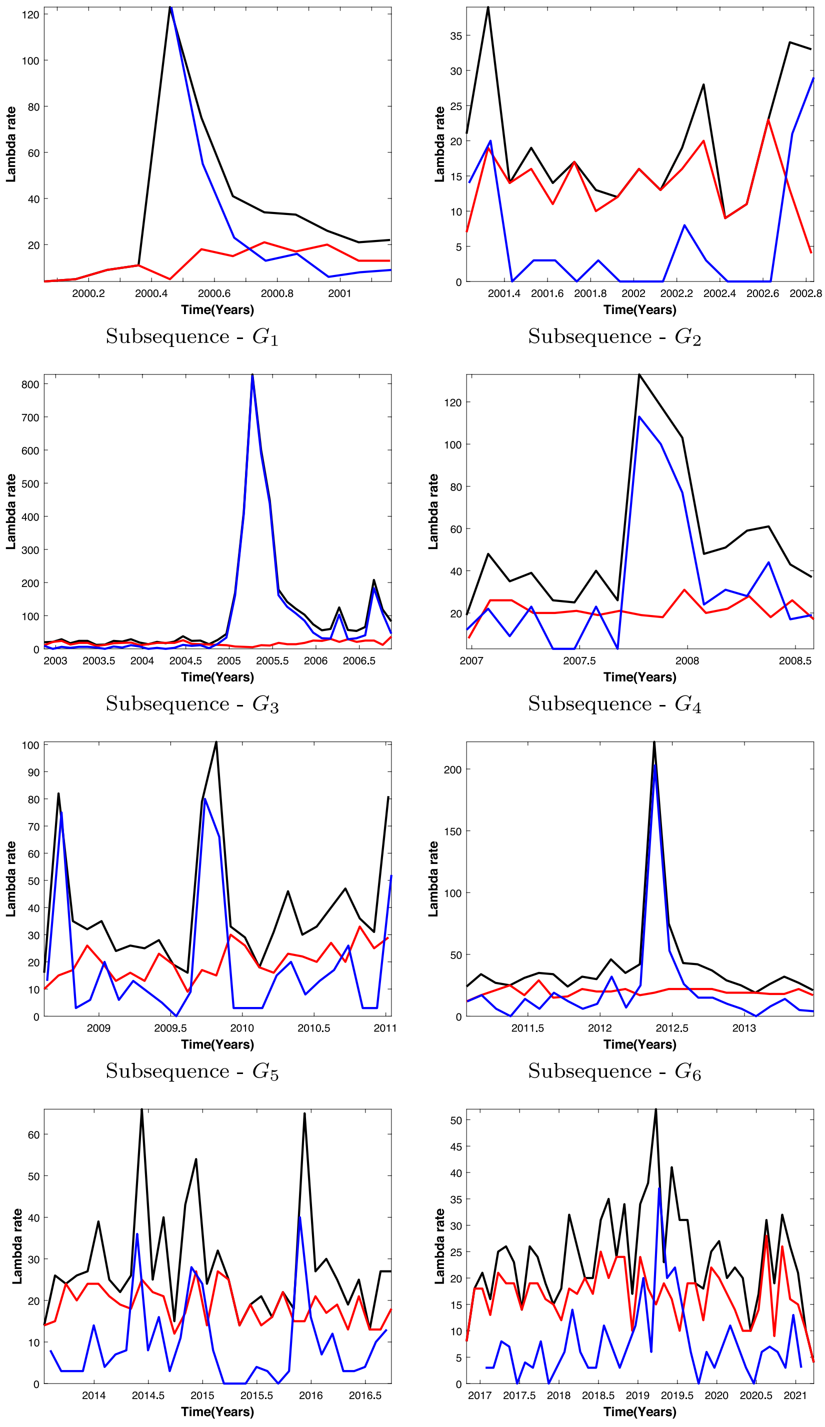
Furthermore, results obtained from the second phase are presented and analyzed in detail through following plots and statistical ways:

### 4.3.1 $\lambda$ plot

Strain accumulation and understanding the consecutive triggering activity of events is carried out by analyzing of the change in seismicity rate. The temporal relevance of events with magnitude in long term is interpreted from the seismic rate analysis. Seismicity rate variations is illustrated by observing the number of events arrival in a predefined time interval which is shown by the $\lambda$ plot in Figs. 7 and 8 for both Sumatra-Andaman and ISC-GEM respectively.

The $\lambda$ plot for original sub-sequence $\in G_j$ (in black), filtered AFs (in blue) and BGs (in red) after both phases of ES-TSC is shown in Figs. 7 and 8 for Sumatra- Andaman and ISC-GEM subsequences respectively. From Figs. 7 and 8, it is evident that the $\lambda$ rate of BGs is almost consistent w.r.t to time, there is no significant change in the arrival rate of BGs (see red line in each subplot in Figs. 7 and 8). On the other hand, when the mainshock (large event) occurs a significant change in the seismic rate is observed with a sudden hike (vertical jump) due to the post-release of seismic energy in terms of the AFs (see the peaks in black in Figs. 7 and 8) and then decays w.r.t. time. Those AFs from each sub-sequence are filtered out and their $\lambda$ rate is shown by a blue line. The $\lambda$ rate (time-varying) of filtered AFs follow the similar pattern as original sub-sequence in both figures shown by black and blue lines.

**Fig. 7** $\lambda$ plot of overall, AFs and BGs in black, blue and red color respectively for each subsequence of Sumatra-Andaman



Subsequence - $G_1$

Subsequence - $G_2$

Subsequence - $G_3$

Subsequence - $G_4$

Subsequence - $G_5$

Subsequence - $G_6$

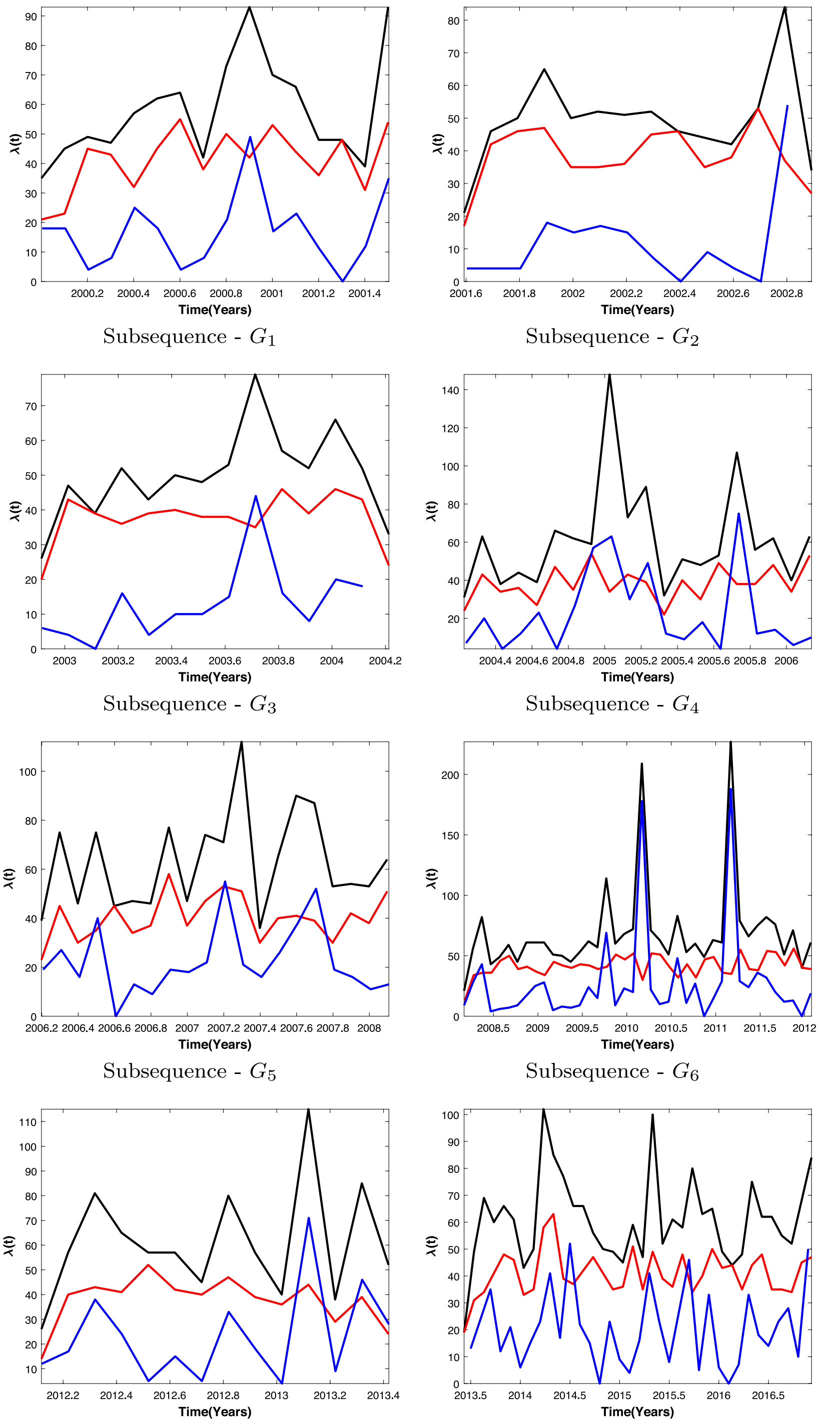
Subsequence - $G_7$

Subsequence - $G_8$

### 4.3.2 Probability density function of $\Delta t$

The statistical analysis of inter-event times between earthquakes plays an important role in the modeling of seismicity and for seismic hazards assessment. However, the inter-event time ($\Delta t$) distribution of earthquakes distinguishes the clustered AFs and non-clustered BGs with the occurrence of the mainshock. A sequence of independent non- clustered randomly occurring earthquake events (BGs) is generally described by a Poisson process with constant intensity $\lambda$ (expected number of events per unit time). This is considered as homogeneous Poisson process (HPP) with probability $P(n, \lambda, t)$ to have $n$ events in the time interval $[0, t]$. It is given by:

**Fig. 8** $\lambda$ plot of overall, AFs and BGs in black, blue and red color respectively for each subsequence of ISC-GEM



Subsequence - $G_1$

Subsequence - $G_2$

Subsequence - $G_3$

Subsequence - $G_4$

Subsequence - $G_5$

Subsequence - $G_6$

Subsequence - $G_7$

Subsequence - $G_8$

$$P(n, \lambda, t) = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \tag{14}$$

with corresponding probability density function (PDF) of $\Delta t$ (IET) between successive pair of events:

$$f(\Delta t) = \lambda \exp(-\lambda t) \tag{15}$$

Apart from negligible deviations, all histograms have the exponential distribution of $\Delta t$ (as described in Eq. 15) that follow the definition of HPP as shown in Figs. 9 and 10 for both catalogs. Alternatively, it means successively occurring BGs are not causally related to each other and have a time-independent mean arrival rate $\lambda$ (uniform in nature). Whereas the occurrence of mainshock-triggered earthquakes follows the non-homogeneous Poisson process in time, those

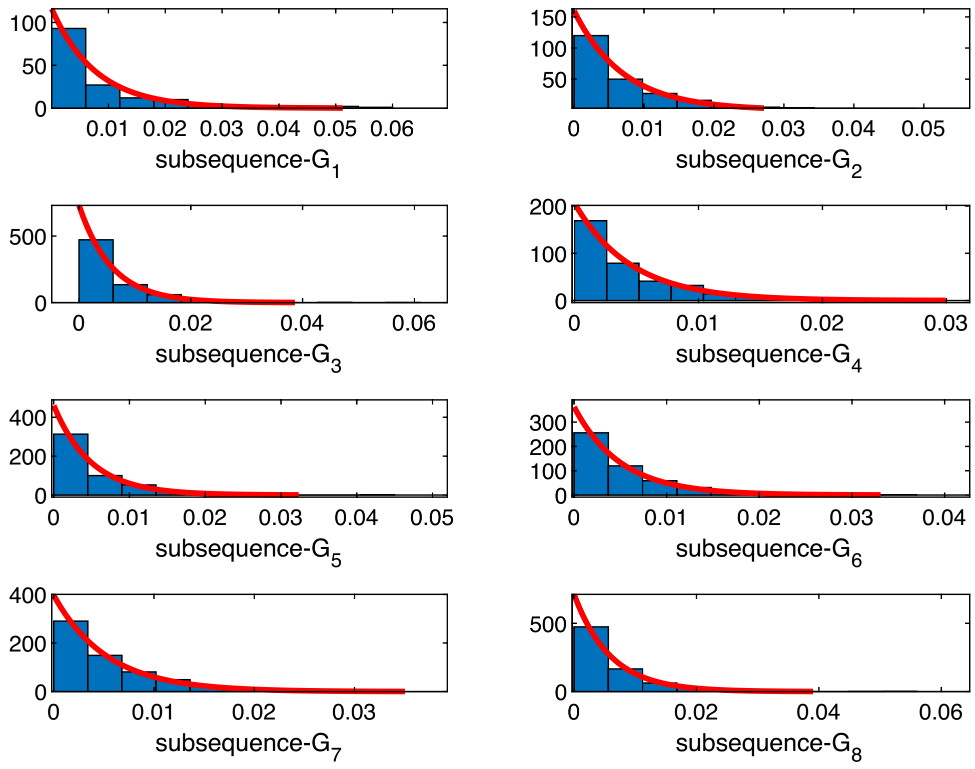


**Fig. 9** Histogram of $\Delta t$ showing exponential distribution for all the meaningful sub-sequences (BGs) obtained from proposed ES-TSC for Sumatra-Andaman

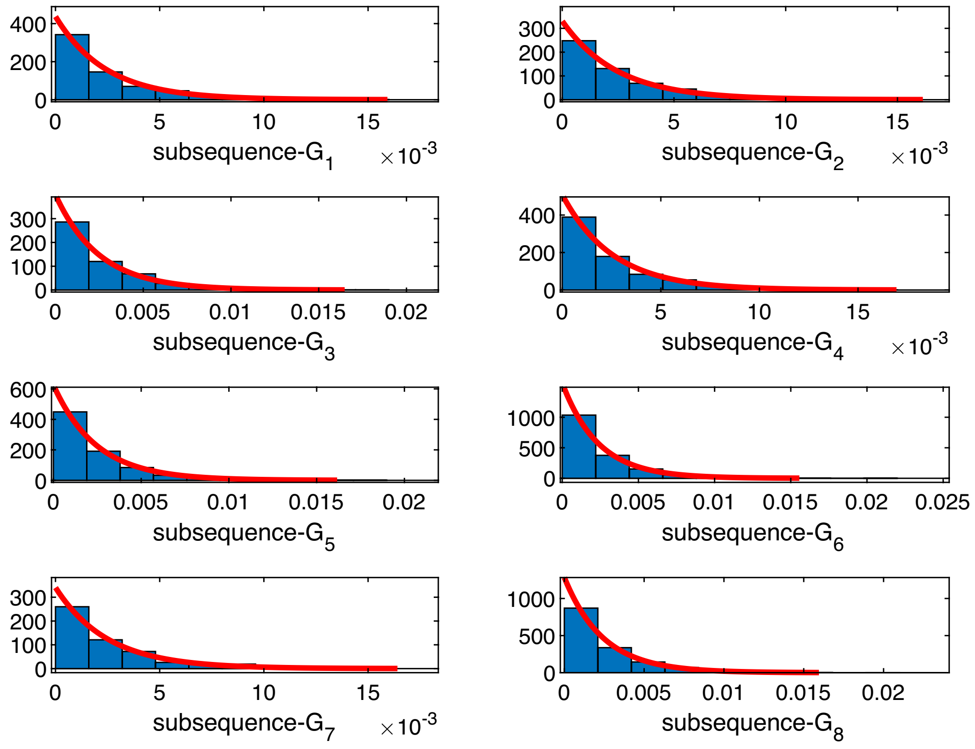


**Fig. 10** Histogram of $\Delta t$ showing exponential distribution for all the meaningful subsequences (BGs) obtained from proposed ES-TSC for ISC-GEM catalogs

are attributed to the clustering in time domain. The behavior histogram of $\Delta t$ for all the meaningful sub-sequences is exponentially fitted as shown by the red color.

The complete behavior of $\Delta t$ for overall events, total obtained AFs and BGs are shown in terms of the probability density function of $\Delta t$ in Figs. 11 and 12 on a logarithmic scale. From the figure, observed that the overall event's

**Fig. 11** PDF of $\Delta t$ for total true events, overall BGs and AFs for Sumatra-Andaman

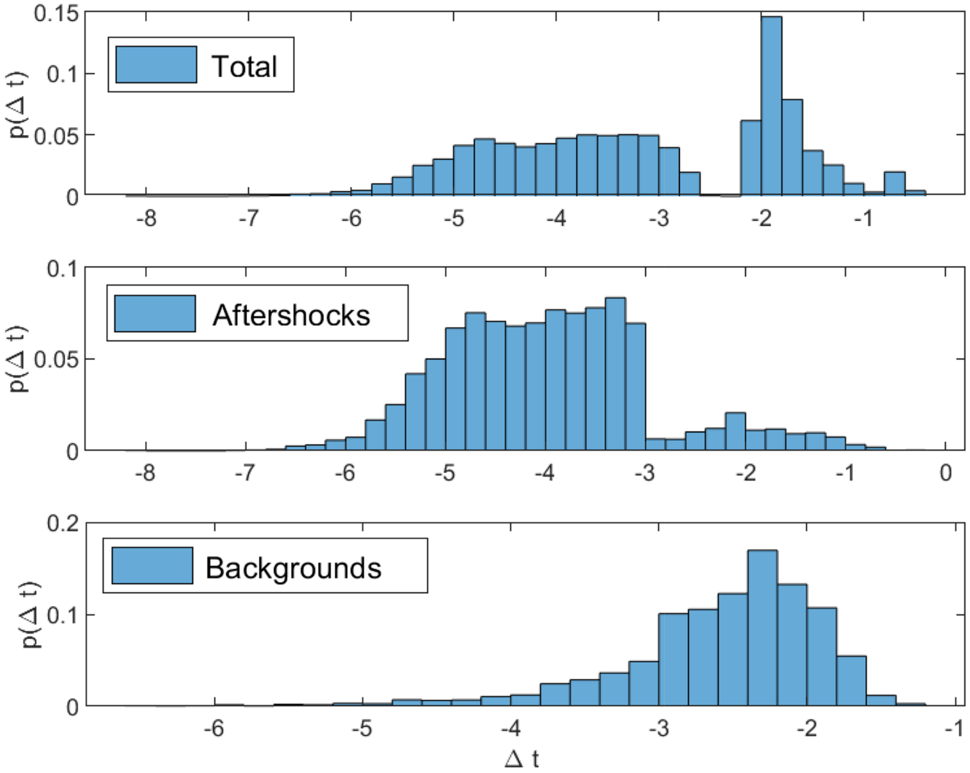
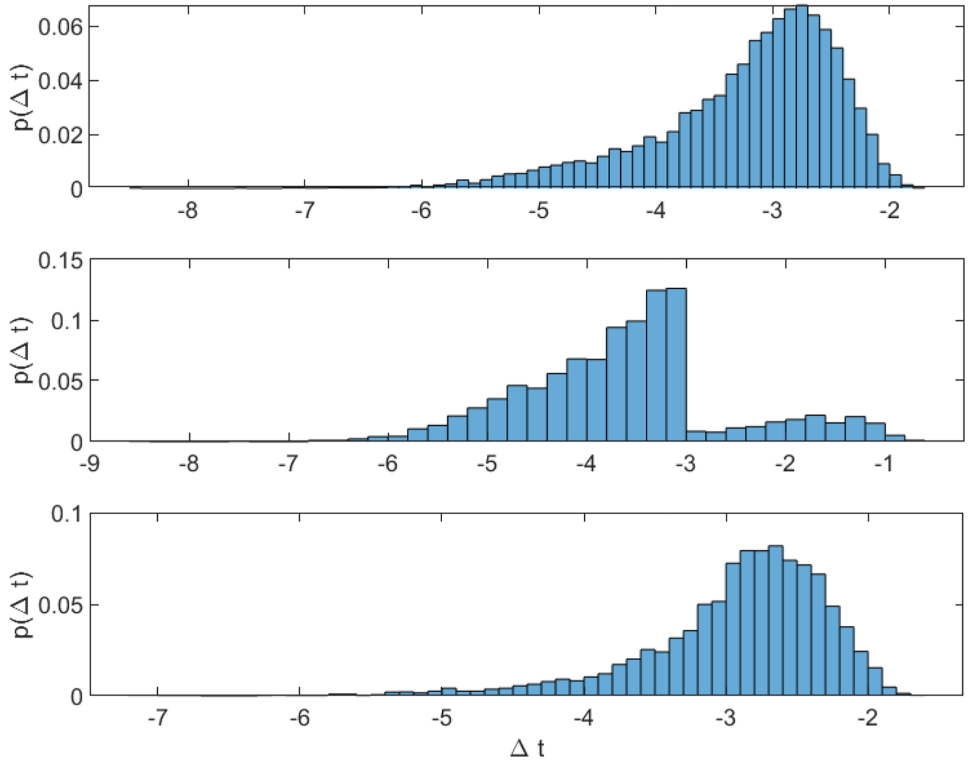


**Fig. 12** PDF of $\Delta t$ for total true events, overall BGs and AFs for ISC-GEM



characteristics in the time domain are bimodal (see Figs. 11a and 12a) and that shows the evidence of two components of seismicity. After applying the proposed ES-TSC method, these two-component are separated in terms of AFs with short $\Delta t$ and BGs with homogeneous Poisson distribution with constant $\lambda$ rate).

#### 4.3.3 Coefficient of variation ($COV(T)$, $COV(d)$)

In Tables 2, 3, 4, 5, $\Delta t$ and $\Delta d$ based $COV$ is determined to observe the characteristics of events in spatio-temporal domain for overall, meaningful (BGs), and filtered subsequences comprised of correlated AFs. Comparative analysis is carried out from obtained values (highlighted in Tables 2, 3, 4, 5) for Sumatra-Andaman and ISC-GEM catalog separately. For both catalogs, it is observed that $COV_b(T)$ (time domain) and $COV_b(d)$ (spatial domain) for each meaningful subsequences lesser than the $COV_o(T)$ and $COV_o(d)$ for true subsequences. $COV_b(T)$ values are close to 1 which justifies the HPP mentioned in criterion-1. Whereas for remaining (removed) subsequences $COV_a(T)$ and $COV_a(d)$ have slightly higher value than $COV_o$ (see Tables (2, 3, 4, 5)). The relation $COV_a(T) > COV_o(T) > COV_b(T)$ and $COV_a(d) > COV_o(d) > COV_b(d)$ reveals that the obtained patterns follow the characteristics of BGs (homogeneous Poisson process, independent and random events) and AFs (Clustered pattern in spatio-temporal domain).

#### 4.3.4 Space-time plot and hot-spot identification

The space-time plot shows the distribution of events along with one of the coordinates and time. Here, longitude vs. time information is taken to observe the characteristic of both types of events. The longitude versus time plot for filtered AFs from original sub-sequences is shown in Fig. 13a and c for Sumatra-Andaman and ISC-GEM respectively. From both figures, it is justified that filtered AFs are triggered nearby in space and time, i.e. clustered form. Their frequency of occurrence is decayed w.r.t. time. Whereas BGs have seemed smooth and non-clustered character as evident from Fig. 13b and d.

The density distribution of earthquake epicenters (longitude vs. latitude) of BGs and AFs for the Sumatra-Andaman and ISC-GEM catalog is shown in Figs. 14 and 15 respectively. From the figure, it is evident that BGs are the smooth representation of seismicity, which helps to identify the fault plane boundary (see Figs. 14b and 15b). Whereas AFs are highly dense, hazardous, and clustered forms in the spatial domain as

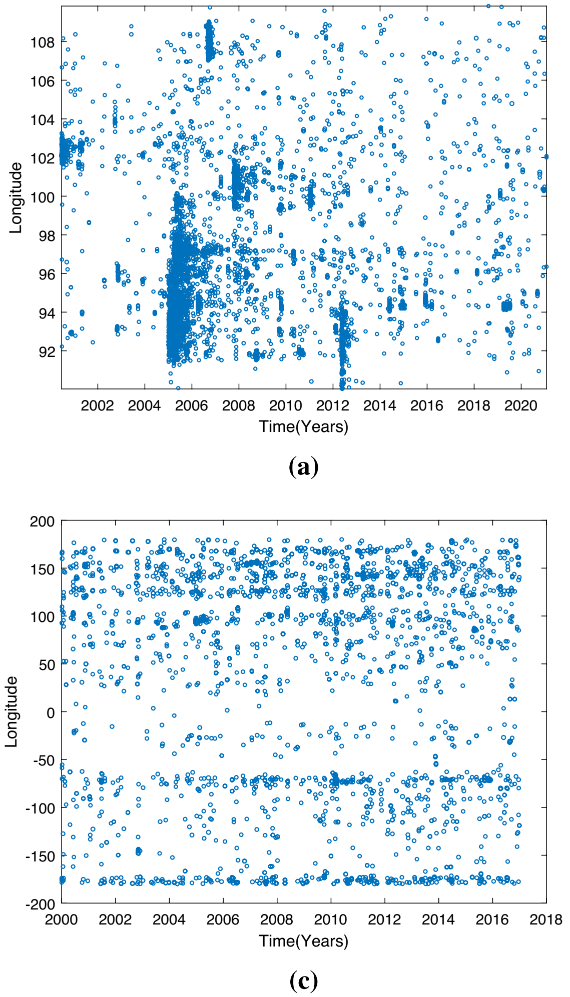
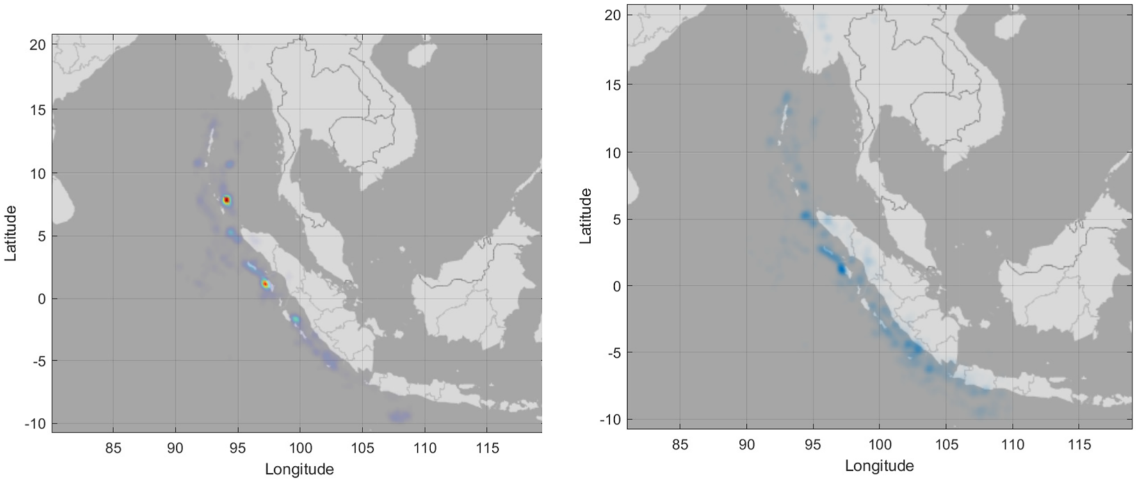


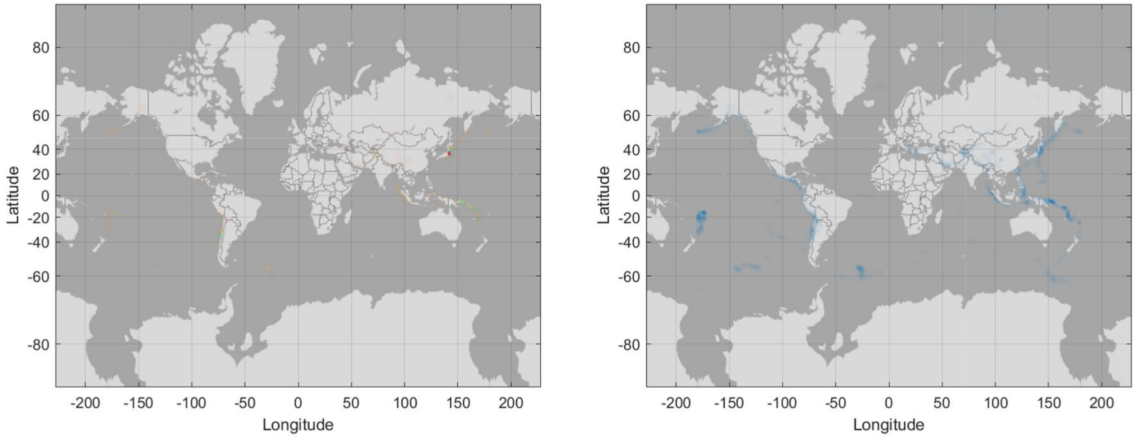
**Fig. 13** Distribution of total **a** AFs **b** BGs in Sumatra-Andaman; and total **c** AFs **d** BGs in ISC-GEM catalog w.r.t. longitude versus time



(a)

(b)

**Fig. 14** Density distribution of earthquake epicenters for **a** AFs and **b** BGs in Sumatra-Andaman



(a)

(b)

**Fig. 15** Density distribution of earthquake epicenters for **a** AFs and **b** BGs in ISC-GEM

observed in Figs. 14a and 15a. Distribution of AFs in spatial domain reveals the information of hot-spots (as shown with color map in Figs. 14a and 15a for both catalogs) and they are useful for short and long term hazard assessment.

## 5 Comparison of the proposed method with state-of-art approaches

Initially, the approach reported by Gardner and Knopoff [8] based on the magnitude-dependent space-time window and Reasenberg's cluster-based algorithm Reasenberg [22] are mostly used and considered the paradigm of earthquake declustering for decades. Firstly, the proposed approach is compared with these two state-of-art methods along with alternate space-time window size suggested by Uhrhammer [28]. From Table 6, the Gardner-Knopoff declustering approach, also Uhrhammer's window, both overestimates the aftershock (AFs) population which leads to a low value of COV for backgrounds ($COV_b$). It means a majority of background events are classified as aftershocks by both methods after setting default parameter values. The execution time of both methods is almost similar but greater than the proposed method (in minutes). In Reasenberg's declustering approach, most of the events are classified as backgrounds (BGs) which seems better as compared to G-K and Uhrhammer' method as justified from their respective COV values but not obtained the optimum $COV$.

Here, the performance of the proposed method is also compared with the recently introduced algorithms by Vijay

**Table 6** Comparison of proposed ES-TSC with benchmark and recently introduced algorithms

| Method | Catalogs | Sumatra-Andaman | ISC-GEM |
|---|---|---|---|
| | Total Events | 10029 | 10611 |
| | $COV_o(T)$ | 1.79 | 1.21 |
| Gardner and Knopoff [8] | AF | 9540 | 7140 |
| | BG | 489 | 3471 |
| | $COV_a(T)$ | 2.18 | 1.78 |
| | $COV_b(T)$ | 0.87 | 1.16 |
| | Avg. Exe. time | 0.123 | 0.395 |
| Uhrhammer [28] | AF | 8435 | 5640 |
| | BG | 1594 | 4971 |
| | $COV_a(T)$ | 2.01 | 2.89 |
| | $COV_b(T)$ | 0.91 | 1.16 |
| | Avg. Exe. time | 0.102 | 0.342 |
| Reasenberg [22] | AF | 2701 | 2667 |
| | BG | 7328 | 7944 |
| | $COV_a(T)$ | 3.14 | 2.86 |
| | $COV_b(T)$ | 1.31 | 1.16 |
| | Avg. Exe. time | 0.059 | 0.089 |
| Vijay and Nanda[30] | AF | 2747 | 1984 |
| | BG | 7282 | 8627 |
| | $COV_a(T)$ | 3.82 | 2.18 |
| | $COV_b(T)$ | 1.24 | 1.08 |
| | Avg. Exe. time | 0.0078 | 0.0082 |
| Zaliapin and Ben-Zion [36] | AF | 5489 | 3257 |
| | BG | 4540 | 7354 |
| | $COV_a(T)$ | 5.16 | 3.14 |
| | $COV_b(T)$ | 1.14 | 1.09 |
| | Avg. Exe. time | 0.65 | 0.69 |
| Proposed ES-TSC | AF | **6219** | **3633** |
| | BG | **3810** | **6978** |
| | $COV_a(T)$ | **5.36** | **3.5** |
| | $COV_b(T)$ | **1.11** | **1.04** |
| | Avg. Exe. time | **0.0137** | **0.0152** |

The best results are highlighted in bold

and Nanda [30] that are based on the K-means clustering approach. This method has very less execution time for declustering but the results are highly dependent on the selection of the cluster centroids at the beginning. The obtained BGs from this method is fair enough but they are not justified in terms of the $COV_b(T)$ (not close to 1). The results obtained from the recently reported declustering algorithm by Zaliapin and Ben-Zion [36] are also summarized in Table 6, after setting the optimum parameter values. The classification results are almost similar to the proposed method in terms of the number of AFs and BGs for both catalogs. But the execution time of the algorithm is the main concern which is highest among all the reported methods due to the determination of the 3D distance metric. Thus, comparative analysis indicates that the proposed ES-TSC is a competitive approach for categorizing the earthquake events in terms of AFs and BGs from the given overall earthquake catalogs.

## 6 Conclusion

This research work has highlighted earthquake time-series analysis with declustering capability by discriminating the clustered AFs sequences and regular background events from catalogs. The meaningful sub-sequences are comprised of random background events only and thus follow characteristics of homogeneous Poisson process in the time domain. The inter-event time statistics and $COV(T)$ are found useful in the sliding temporal window for pointing out the when and where to filter the sequence. The filtered AFs' population in each subsequence explicitly indicates the risk factor in terms of their spatio-temporal extent. Based on the graphical representation and statistical evidence, the proposed approach provides a competitive way to solve the problem of seismicity declustering in less computation time. Thus, large earthquake catalogs are used for experimental analysis in the future for declustering outcomes. The performance of the proposed approach indicates that the use of subsequence time series data is still important for improving time series data mining for discovering informative and salient sub-sequence features in different fields and research.

**Funding** There is no funding support for this research work.

**Data availability** The earthquake catalogs that support the findings of this research work are available by International Seismological Centre (2022), ISC-GEM Earthquake Catalogue, https://doi.org/10.31905/d808b825 and Northern California Earthquake Data Center (NCEDC), https://doi.org/10.7932/NCEDC.

## Declarations

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

## References

1. Aden-Antoniów F, Frank W, Seydoux L (2022) An adaptable random forest model for the declustering of earthquake catalogs. J Geophys Res Solid Earth 127(2):e2021JB023254
2. Bakker M, Schaars F (2019) Solving groundwater flow problems with time series analysis: you may not even need another model. Groundwater 57(6):826–833
3. Chitra P, Rajasekaran UM, et al (2022) Time-series analysis and flood prediction using a deep learning approach. In: 2022 international conference on wireless communications signal processing and networking (WiSPNET), IEEE, pp 139–142
4. Corsaro S, Angelis PLD, Fiore U, Marino Z, Perla F, Pietroluongo M (2021) Wavelets in multi-scale time series analysis: an application to seismic data. Dynamics of disasters. Springer, Berlin, pp 93–100
5. Dad JM, Muslim M, Rashid I, Reshi ZA (2021) Time series analysis of climate variability and trends in kashmir himalaya. Ecol Ind 126:107690
6. Di Giacomo D, Engdahl ER, Storchak DA (2018) The ISC-GEM earthquake catalogue (1904–2014): status after the extension project. Earth Syst Sci Data 10:1877–1899. https://doi.org/10.5194/essd-10-1877-2018
7. D'Urso P, De Giovanni L, Massari R, D'Ecclesia RL, Maharaj EA (2020) Cepstral-based clustering of financial time series. Expert Syst Appl, p 113705
8. 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
9. Gupta A, Gupta HP, Biswas B, Dutta T (2020) Approaches and applications of early classification of time series: a review. IEEE Trans Artif Intell 1(1):47–61
10. ISC (2020) International seismological centre (2020), ISC-GEM earthquake catalogue https://doi.org/10.31905/d808b825
11. Javed A, Lee BS, Rizzo DM (2020) A benchmark study on time series clustering. Mach Learn Appl 1:100001
12. Kagan YY, Jackson DD (1991) Long-term earthquake clustering. Geophys J Int 104(1):117–133
13. Khan N, Haq IU, Ullah FUM, Khan SU, Lee MY (2021) Cl-net: Convlstm-based hybrid architecture for batteries' state of health and power consumption forecasting. Mathematics 9(24):3326
14. Khan N, Ullah FUM, Haq IU, Khan SU, Lee MY, Baik SW (2021) Ab-net: a novel deep learning assisted framework for renewable energy generation forecasting. Mathematics 9(19):2456
15. Kundu S, Opris A, Yukutake Y, Hatano T (2020) Extracting correlations in earthquake time series using complex network analysis. arXiv preprint arXiv:2004.05415
16. Marsan D, Prono E, Helmstetter A (2013) Monitoring aseismic forcing in fault zones using earthquake time series. Bull Seismol Soc Am 103(1):169–179
17. Michas G, Vallianatos F (2018) Stochastic modeling of nonstationary earthquake time series with long-term clustering effects. Phys Rev E 98(4):042107
18. Moustra M, Avraamides M, Christodoulou C (2011) Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. Expert Syst Appl 38(12):15032–15039
19. NCEDC (2021) Northern California Earthquake Data Center. UC Berkeley Seismological Laboratory Dataset. https://doi.org/10.7932/NCEDC
20. Picoli MCA, Camara G, Sanches I, Simões R, Carvalho A, Maciel A, Coutinho A, Esquerdo J, Antunes J, Begotti RA et al (2018) Big earth observation time series analysis for monitoring brazilian agriculture. ISPRS J Photogramm Remote Sens 145:328–339
21. Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, Su Q, Wang W, Wang X, Zhang Z (2020) Covid-19 transmission in mainland china is associated with temperature and humidity: A time-series analysis. Sci Total Environ 728:138778
22. Reasenberg P (1985) Second-order moment of central california seismicity, 1969–1982. J Geophys Res Solid Earth 90(B7):5479–5495
23. Ruiz L, Pegalajar M, Arcucci R, Molina-Solana M (2020) A time-series clustering methodology for knowledge extraction in energy consumption data. Expert Syst Appl 160:113731
24. Sarlis NV, Skordas ES, Varotsos PA (2018) Natural time analysis of seismic time series. Complexity of seismic time series. Elsevier, Amsterdam, pp 199–235
25. van Stiphout T, Zhuang J, Marsan D (2012) Seismicity declustering. Commun Online Res Stat Seism Anal 10:1–25
26. Storchak EA (2013) Public release of the isc-gem global instrumental earthquake catalogue (1900–2009). Seism Res Lett 84(5):810–815. https://doi.org/10.1785/0220130034
27. Storchak EA (2015) The ISC-GEM global instrumental earthquake catalogue (1900–2009). Introd Phys Earth Planet Int 239:48–63. https://doi.org/10.1016/j.pepi.2014.06.009
28. Uhrhammer R (1986) Characteristics of northern and central california seismicity. Earthq Notes 57(1):21
29. Ullah FUM, Khan N, Hussain T, Lee MY, Baik SW (2021) Diving deep into short-term electricity load forecasting: comparative analysis and a novel framework. Mathematics 9(6):611
30. Vijay RK, Nanda SJ (2017) Tetra-stage cluster identification model to analyse the seismic activities of japan, himalaya and taiwan. IET Signal Proc 12(1):95–103
31. 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
32. Vijay RK, Nanda SJ (2019) Shared nearest neighborhood intensity based declustering model for analysis of spatio-temporal seismicity. IEEE J Select Topics Appl Earth Observ Remote Sens 12(5):1619–1627

33. Vogel E, Saravia G, Pastén D, Muñoz V (2017) Time-series analysis of earthquake sequences by means of information recognizer. Tectonophysics 712:723–728
34. Wang W, Yildirim G (2022) Applied time-series analysis in marketing. Handbook of Market Research. Springer, Berlin, pp 469–513
35. Wu Z (2010) A hidden markov model for earthquake declustering. J Geophys Res Solid Earth 115(B3)
36. Zaliapin I, Ben-Zion Y (2020) Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain. J Geophys Res Solid Earth 125(4):e2018JB017120
37. Zhu B, Hou X, Liu S, Ma W, Dong M, Wen H, Wei Q, Du S, Zhang Y (2021) Iot equipment monitoring system based on c5. 0 decision tree and time-series analysis. IEEE Access 10:36637–36648
38. Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380
39. Zhuang J, Chang CP, Ogata Y, Chen YI (2005) A study on the background and clustering seismicity in the taiwan region by using point process models. J Geophys Res Solid Earth 110(B5)

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