Nonlinear Processes
in Geophysics

Open Access

A tri-stage cluster identification model for accurate analysis of seismic catalogs

S. J. Nanda[1], **K. F. Tiampo**[2], **G. Panda**[1], **L. Mansinha**[2], **N. Cho**[2], **and A. Mignan**[3]

[1]School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Orissa, 751013, India
[2]Department of Earth Sciences, Western University, London, ON, N6A 5B7, Canada
[3]Swiss Seismological Service, ETH, NO H66, Sonneggstrasse 5, 8092 Zurich, Switzerland

Correspondence to: S. J. Nanda (nanda.satyasai@gmail.com)

Abstract. In this paper we propose a tri-stage cluster identification model that is a combination of a simple single iteration distance algorithm and an iterative K-means algorithm. In this study of earthquake seismicity, the model considers event location, time and magnitude information from earthquake catalog data to efficiently classify events as either background or mainshock and aftershock sequences. Tests on a synthetic seismicity catalog demonstrate the efficiency of the proposed model in terms of accuracy percentage (94.81 % for background and 89.46 % for aftershocks). The close agreement between lambda and cumulative plots for the ideal synthetic catalog and that generated by the proposed model also supports the accuracy of the proposed technique. There is flexibility in the model design to allow for proper selection of location and magnitude ranges, depending upon the nature of the mainshocks present in the catalog. The effectiveness of the proposed model also is evaluated by the classification of events in three historic catalogs: California, Japan and Indonesia. As expected, for both synthetic and historic catalog analysis it is observed that the density of events classified as background is almost uniform throughout the region, whereas the density of aftershock events are higher near the mainshocks.

1 Introduction

Examination of the spatial and temporal distributions of regional earthquakes confirms that they primarily occur on tectonic features such as faults (spatial clustering) and subsequent to large events (temporal clustering). For example, the Omori law (Omori, 1894) is an empirical relation that describes the clustering of aftershocks in time. Utsu (2002) offers a broad review on the spatial distribution of seismicity, while Stein (1999) also incorporated the principle of spatial clustering in associating triggered events or aftershocks to regions of stress increase.

The clustered nature of earthquakes has motivated many studies to understand and model seismicity, including the Epidemic Type Aftershock Sequence (ETAS) (Ogata, 1988; Helmstetter and Sornette, 2002) and the Branching Aftershock Sequence (BASS) models (Turcotte et al., 2007). Shcherbakov et al. (2005) analyzed the interoccurrence times between events as a non-homogeneous Poisson process, while Mendoza and Hartzell (1988) studied the correlation between the location of aftershocks and regions of coseismic slip in Southern California.

Declustering seismicity data has been the subject of intensive study over the years, where the main goal is to separate a given catalog into subsets in which their elements share similar characteristics based on a particular set of criteria. For example, Gardner and Knopoff (1974) developed a declustering method for California using power laws to scale the duration and the spatial regions of aftershock sequences with respect to magnitude that yielded a Poissonian-like residual catalog. Reasenberg (1985) introduced a method in which events are analyzed in pairs and aftershock sequences are modeled as time-dependent Poisson processes. Zhuang et al. (2005) developed a method for stochastic declustering based upon the ETAS model. Baiesi and Paczuski (2004) and Zaliapin et al. (2008) developed a method in which time-space-magnitude distances between earthquakes are analyzed to determine whether two events are linked. A method to decluster seismicity based on determining the probability

of direct and indirect aftershock triggering was proposed by Marsan and Lengline (2008). Wu (2010) employed a Hidden Markov model to decluster seismicity. However, it has been shown that the choice of declustering algorithm and the variation in the related parameter values can have a significant impact on the calculation of local and regional seismicity rates (Van Stiphout et al., 2011). Recent work using the ETAS model to decluster seismicity in Indonesia, Japan and the Himalayas has resulted in the identification of background anomalies of activation and quiescence (Bansal and Ogata, 2010; Bansal et al., 2012; Ogata, 2011).

In this study, we present a different method to decluster seismicity, based on a tri-stage cluster identification model that only considers event location, time and magnitude of earthquake catalog data in an objective, mathematical formulation designed to efficiently classify events as either background or mainshock and aftershock sequences. In this model, which is relatively simple when compared to many of the currently available algorithms, each event is classified as an aftershock if it belongs to at least two cluster zones with respect to time, location or magnitude. We have incorporated a simple single-iteration distance algorithm in the first two stages and an iterative K-means algorithm (Xu and Wunsch, 2005; Nazeer and Sebastian, 2009) in the third stage to accurately identify the cluster zones. There is flexibility in the model design to allow proper selection of cluster zones depending upon the nature of the mainshocks present in the catalog. Again, the goal here is to study the efficiency of a method based solely on a distance clustering algorithm for three catalog parameters, location, time and magnitude, in order to provide an objective method that can incorporate the variation in each parameter into the algorithm itself. Note also that here there is no requirement to identify and parameterize an a priori probability distribution for the relationship between events. The method relies only on the characteristics of the seismicity itself, as controlled by the local or regional tectonics. The method is tested on a synthetic catalog and three real-world seismic catalogs: California, Japan and Indonesia. The accuracy is presented in the form of a lambda plot (number of events yr^{-1}), a cummulative seismicity plot, the percentage of accuracy and index matching with true events.

The synthetic catalog formulation is described in the next section. The tri-stage cluster identification technique is detailed in Sect. 3. Application of the proposed model to analysis of the California, Japan and Indonesian catalogs is presented in Sect. 4. The declustering performance of the proposed is compared with two benchmark methods (Gardner and Knopoff, 1974; Uhrhammer, 1986) in Sect. 4. The concluding remarks are presented in Sect. 5.

Table 1. List of Parameters used in the synthetic catalog.

Parameters		Values
Magnitude-frequency	a-value	4.2
	b-value	1
	M_{min}	1
	M_{max}	< 5.1
Background	density (events day^{-2} deg.$^{-1}$)	
	– outside stress shadow	0.008
	– inside stress shadow	0.0
Aftershocks (ETAS)	p-value	1.10
	α (magnitude^{-1})	1
	c (days)	0.001
	K	0.004

2   Synthetic catalog

The basic inputs into the synthetic catalog used for testing the declustering method, including the epidemic type aftershock sequence model (ETAS), are detailed here.

2.1   ETAS

The ETAS model was developed by Ogata (1988, 1992) to model the seismic activity of a given region. The model describes the occurrence rate $n(t)$ as a superposition of background seismicity μ (shocks day^{-1}) and weighted sum of any j-th aftershock (time t_j) occurred before time t given by

$$n(t) = \mu + \sum_{\{j:t_j \prec t\}} e^{\alpha(M_j - M_c)} v\left(t - t_j\right), \forall j = 1, 2, ..N, \quad (1)$$

where N is the number of earthquakes of magnitude M_c or higher. The term $e^{\alpha(M_j - M_c)}$ represents the aftershock productivity, where M_j is the magnitude of j-th aftershock and M_c is the cut-off magnitude of the fitted data (M_c is taken as 1.0; Helmstetter et al., 2005). The aftershock decay $v(t)$ is represented by modified Omori function (Utsu, 1961)

$$v(t) = K\left(t + c\right)^{-p}, \quad (2)$$

where K (shocks day^{-1}) represents the productivity of aftershock for a short duration after the mainshock.

The parameters of the ETAS model μ, p, α, c, and K are generally calculated by the maximum likelihood method. However, most standard parameter values can be obtained from the literature (Mignan and Tiampo, 2010).

2.2   Synthetic catalog using ETAS model

For this work we considered a synthetic seismic catalog generated by a combination of a non-stationary Poisson process (region of quiescence in the seismicity) and the ETAS model detailed above. The quiescence is considered to be the surface expression of a stress shadow generated by a large earthquake that occurred in the recent past (but is not expressed

in the catalog itself). This particular synthetic catalog is designed to replicate natural seismicity, including a particular variation in the seismic rate known as Non-Critical Precursory Accelerating Seismicity Theory (Non-Critical PAST) (Mignan et al., 2007; Mignan, 2008). In Non-Critical PAST, an advancement on the original theory of Accelerating Seismic Release (ASR) (King and Bowman, 2003; Mignan, 2011), for a fixed region in space the cumulative number of events in the background seismicity increases as a power law with respect to time before the mainshock. This acceleration results in an increase of the Gutenberg–Richter a-value in local areas (Gutenberg and Richter, 1944), while events that occur in the stress shadows tend to hide the pattern of accelerating seismicity. The result is a cycle of quiescence-activation in the seismicity, localized in space and time prior to the mainshock. The identification of these spatiotemporal variations is difficult, although the pattern has been observed in a few natural cases (e.g. Mignan and Di Giovambattista, 2008; Mignan and Tiampo, 2010). Activation here refers to medium-term accelerating seismic release and not to short-term foreshock burst activity, which is not included in this version of Non-Critical PAST. One motivation for analyzing a Non-Critical PAST synthetic catalog was to study whether the method was capable of identifying regions of variable spatio-temporal activity that, on average, are relatively close to the background levels of activity.

In this synthetic catalog formulation, the background seismicity rate and the noise ratio are adjustable parameters, depending, respectively, on the regional seismic activity and on the effect of the stress shadow. The background seismicity density is fixed to 1000 background events in the catalog per year, or 0.008 events degree^{-2} day^{-1} in a 2000 by 2000 km square grid over a 20 yr period. The noise level is set at 0.1. For each time step of one year, background events have a uniform random distribution in space and time (Mignan and Tiampo, 2010). Background events are implemented randomly in space and time from a Gutenberg–Richter relation with a b-value of 1.0 (Gutenberg and Richter, 1944). The minimum magnitude of background events is $M_{\min} = 1.0$ to ensure that, for the density of events in the catalogue, the maximum magnitude $M_{\max}$ for the region is less than 6.0. Spatiotemporal clustering is included in the simulated background seismicity by following the ETAS model (Ogata, 1992; Ogata and Zhuang, 2006). Aftershocks also obey Bath's law, in which the difference in magnitude between the mainshock and its largest aftershock Δ_m is fixed to 1.0 to ensure that all aftershocks have a magnitude $M < M_{\max}$ (Shcherbakov et al., 2005). This method is designated as "Modified ETAS", due to restriction in magnitude.

The size of the stress shadow, or quiescent region, resulting from the earthquake that occurred just prior to the start of the catalog ($t = t0$), is based on dislocation theory in an elastic half-space (Okada, 1992). Actual Coulomb stress calculations were performed using the source code AlmondX (courtesy of G. C. P. King). The stress field $\sigma(x, y, ti)$ results
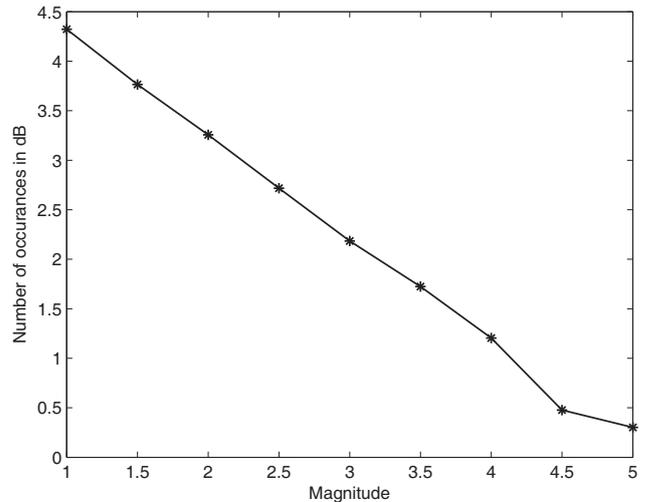

Fig. 1. Gutenberg–Richter relationship for synthetic catalog.

from a constant displacement on the fault. The stress shadow decreases through time through the linear superposition of a loading stress field by following the concept of back-slip model (Savage, 1983). The cumulative number of events inside the stress shadow increases as a power-law function through time, in agreement with Mignan et al. (2007). For more information about Non-Critical PAST simulations, the reader should refer to Mignan and Tiampo (2010).

The list of parameters used in the catalog is presented in Table 1. The Gutenberg–Richter relationship for the synthetic catalog is shown in Fig. 1.

3 Proposed tri-stage cluster identification model

The tri-stage cluster identification model is proposed to classify the background seismicity, foreshocks–aftershocks and mainshocks by considering the time, location and magnitude information as obtained from a given catalog. The block diagram of the proposed model is shown in Fig. 2. Detail about the various cluster identification stages involved in the model are outlined below.

3.1 Mainshocks

The events with large magnitude in the catalog are considered to be the mainshocks. Therefore, in the synthetic catalog events with magnitude greater than 4.5 are considered as mainshocks. The location (coordinates) and time of occurrence of the mainshocks are represented as black asterisks ($*$) in Fig. 3 (top left and top right, respectively).
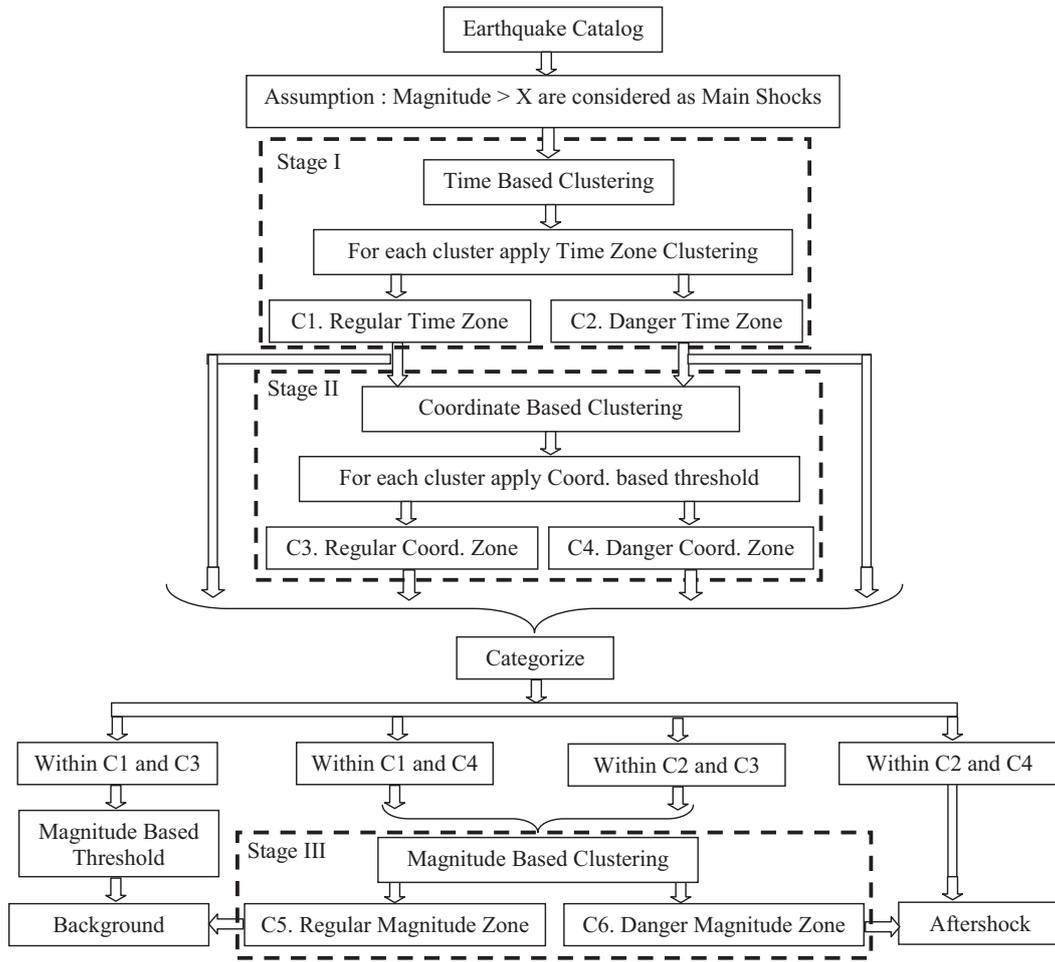
Fig. 2. Block diagram of proposed tri-stage cluster identification model.

3.2　Stage I: temporal cluster identification

The temporal cluster identification is carried out using temporal clustering and time zone based clustering as follows:

3.2.1　Temporal clustering

The events in the catalog are classified into M clusters (based on number of mainshocks) with a simple single-iteration distance algorithm. The steps involved are:

1. The time events of the mainshocks (ξ) are considered as the centers of the clusters.

$$\xi = \{\xi_1, \xi_2, .., \xi_M\}, \tag{3}$$

   where $M = 5$ is the number of mainshocks of catalog.

2. For each point $p_i \in P_{N \times D}$ (catalog having N events, D dimension) calculate the Euclidean distance (2-norm) from the cluster centers

$$d(p_i, \xi_j) = \sqrt{\sum_{d=1}^{D} \left(p_{i,d} - \xi_{j,d}\right)^2}. \tag{4}$$

   Here, $D = 1$ as clustering is based only on the time associated with each event.

3. Assign p_i to nearest cluster center ct_j for which

$$d(p_i, \xi_j) = \overset{\text{Minimum}}{\forall k \in \{1, 2..., M\}} \{d(p_i, \xi_k)\}. \tag{5}$$

From this, the synthetic catalog events are classified into five clusters (represented by red, yellow, cyan, magenta and green colors) shown in Fig. 3 (bottom left).

3.2.2　Time zone based clustering

The events that belong to each of the above five clusters are then grouped into two categories
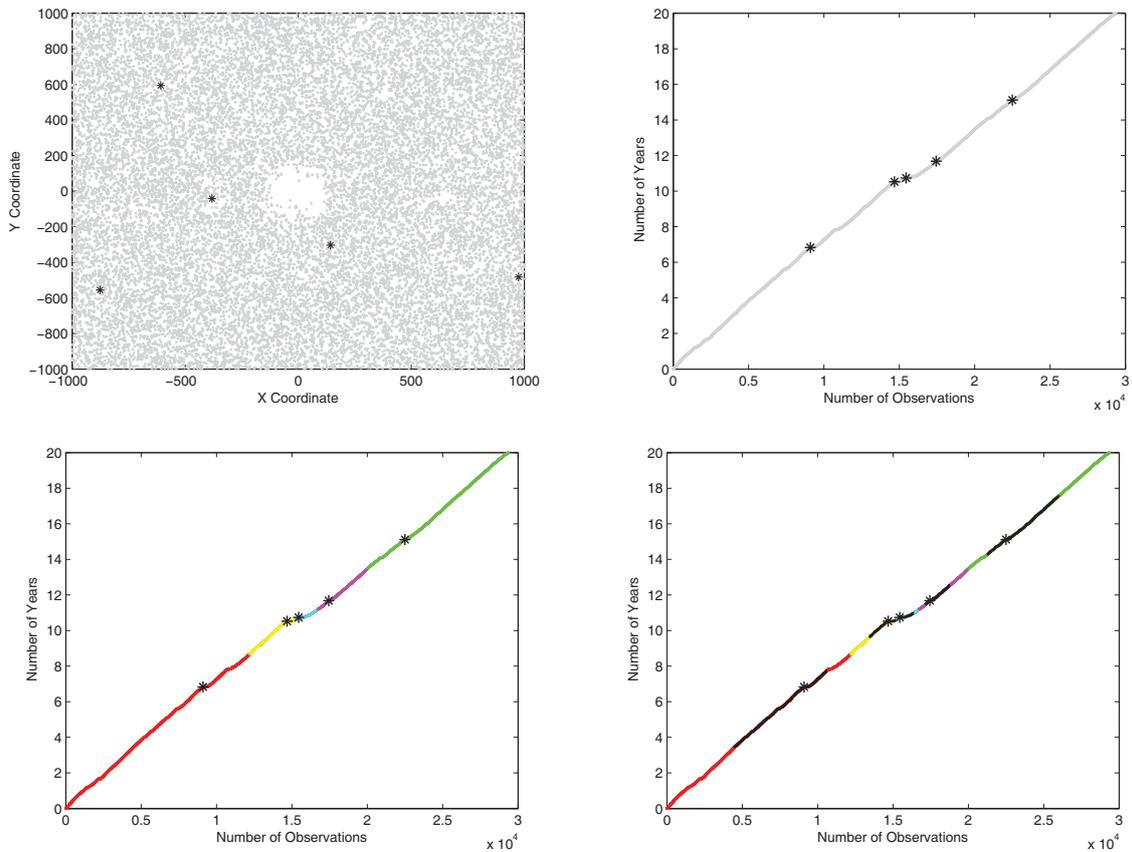
Fig. 3. Tri-stage cluster identification analysis for the synthetic seismic catalog, (top left) locations of all events coordinate in light gray and mainshocks shown in black, (top right) locations of mainshocks with respect to time, (bottom left) output of temporal clustering, (bottom right) output of time zone clustering (black solid lines representing duration of clustered time zone). Note the quiescent region at the center of the catalog (top left). This area and its surroundings (see exact extent in Mignan and Tiampo, 2010) include spatiotemporal variations in the seismicity rate, corresponding to a cycle of quiescence–activation (or accelerating seismic release pattern).

– $c1$. Regular time zone: events located far away from the occurrence time of mainshock.

– $c2$. Identified time zone (danger time zone): events located near the occurrence time of mainshock.

In order to achieve the above classification, each cluster is divided into three sub-clusters using the single-iteration distance algorithm detailed above. The mainshock (Z2) and two extreme points (at a distance from the mainshock, before occurrence (Z1) and after occurrence (Z2)) are considered as the sub-cluster centers shown in Fig. 4.
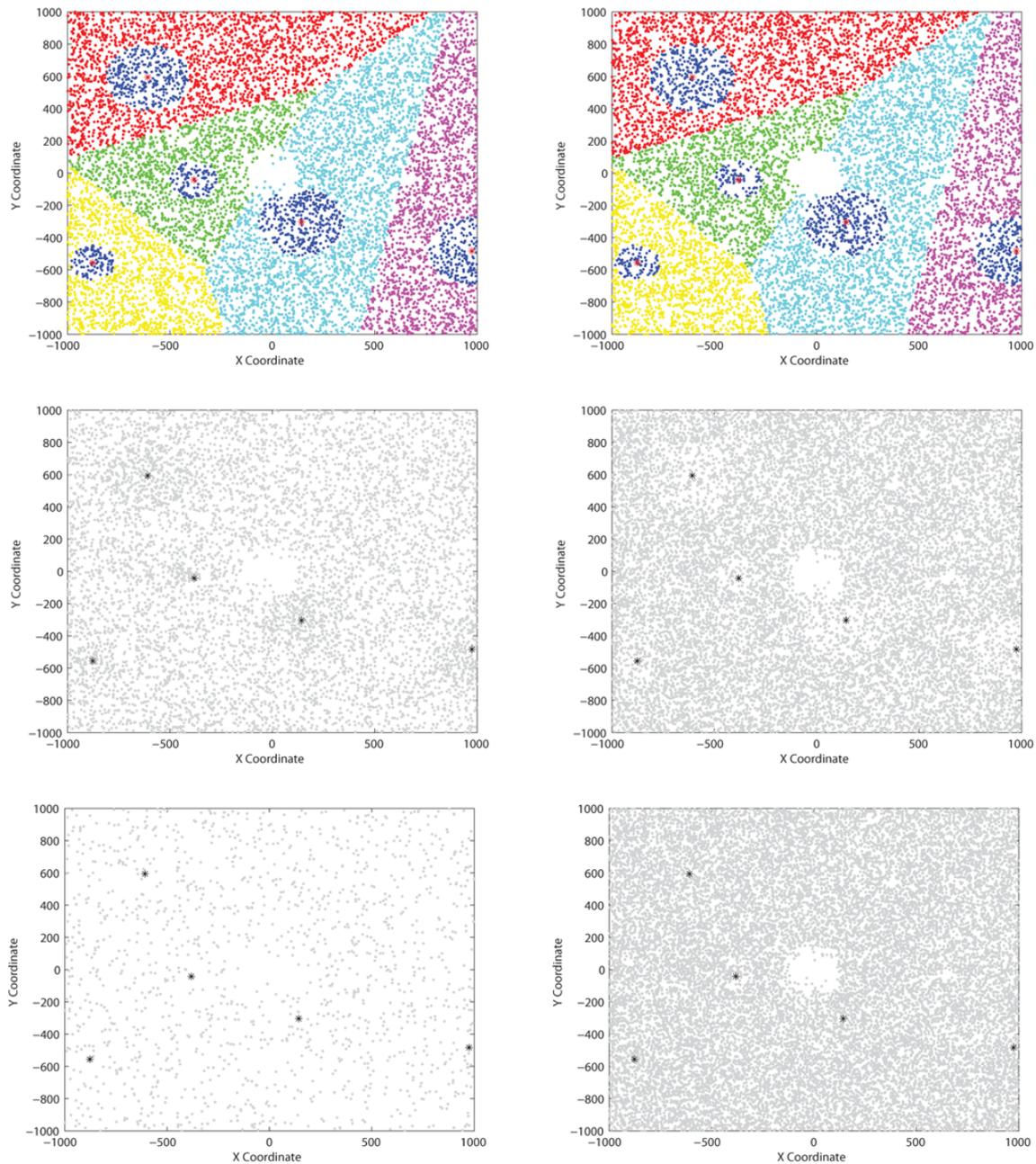
Define $C1$ and $C2$ as the cluster of events that belong to the background and earthquake sequence time zone for the entire dataset given by

$$C1 = \sum_{m=1}^{M} c1_m = \sum_{m=1}^{M} \text{Z1 Cluster}_m + \sum_{m=1}^{M} \text{Z3 Cluster}_m \quad (6)$$

$$C2 = \sum_{m=1}^{M} c2_m = \sum_{m=1}^{M} \text{Z2 Cluster}_m. \quad (7)$$
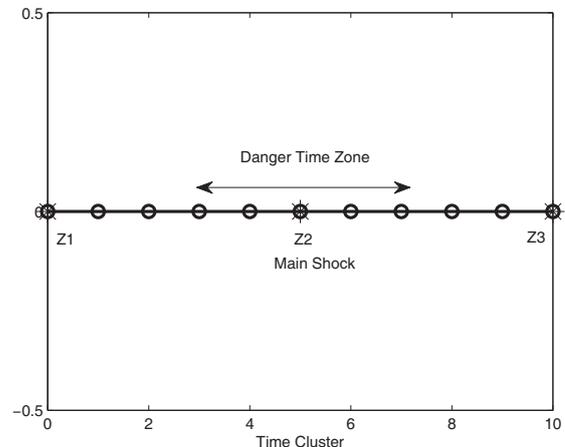
Fig. 4. Example of time zone clustering.

Fig. 5. Tri-stage cluster identification analysis for the synthetic seismic catalog based on ETAS and Non-Critical PAST, (top left) output of background and clustered coordinate zone for $C1$, (top right) output of background and clustered coordinate zone for $C2$, (middle left) model output representing events classified as aftershocks, (middle right) model output representing events classified as backgrounds, (bottom left) true aftershocks, (bottom right) true backgrounds.

The events that belong to cluster time zone $C2$ for the synthetic catalog are plotted as black dots around the main-shocks in Fig. 3 (bottom right). The addition operator in Eq. (6) represents all the events that are shown as black solid lines in Fig. 3 (bottom right). Each individual cluster, representing one danger time zone (Fig. 4), corresponds to one black line. The sum provides the collection of all the events in the black solid line time zones.

Similarly, the Eq. (7) represents the summation of all the clusters which are outside the solid lines (Fig. 3 bottom right). They represent the regular time zone events.

3.3 Stage II: coordinate based cluster identification

The location based cluster identification is carried out using coordinate clustering and thresholding as follows:

3.3.1 Coordinate clustering

Events belonging to regular zone $C1$ and cluster time zone $C2$ undergo location based cluster identification separately. The distance based algorithm discussed in Sect. 3.2 is employed for clustering of the two time zone data sets. The locations of mainshocks are considered as the cluster centers. The dimension of each event is two, which corresponds to both the x- and y- (or longitude and latitude) coordinates. Both datasets $C1$ and $C2$ are classified into 5 clusters each (represented by red, yellow, cyan, magenta and green dots) shown in Fig. 5 (top left and top right, respectively).

3.3.2 Coordinate thresholding

The events belonging to each cluster are then classified into two categories based on a threshold.

– $c3$. Background coordinate zone: Events located far away from the location of mainshock.

– $c4$. Clustered coordinate zone (Danger coordinate zone): Events located near the location of mainshock.

The threshold is determined as a portion $1/\psi$ of the maximum distance between the cluster center and the furthest point (F) inside the cluster. For all events $p_i \in C_j$ the thresholding process is as follows:

$$\text{If } \left(d\left(p_i, c_j \right) < (1/\psi) \left[d\left(p_i, F_j \right) \right] \right) \Rightarrow \begin{cases} p_i \in c4 \\ \text{else, } p_i \in c3. \end{cases} \quad (8)$$

The selection of optimal $1/\psi$ value is discussed in Sect. 3.6. The coordinate zone for both background and clustered time zone is shown in Fig. 5 (top left and top right), respectively. The red stars represent the cluster center and blue dots represent the clustered coordinate zone.

3.4 Categorization

In this model an event is classified as an aftershock if it belongs to at least two identified zones with respect to time, location or magnitude. After temporal and location based cluster identification, the events in the catalog are classified into four categories as follows:

– Category 1: Within $C1$ and $C3$ – events that belong to normal time zone and coordinate zone.

– Category 2: Within $C1$ and $C4$ – events that belong to normal time zone and clustered coordinate zone.

– Category 3: Within $C2$ and $C3$ – events that belong to clustered time zone and normal coordinate zone.

– Category 4: Within $C2$ and $C4$ – events that belong to clustered time as well as clustered coordinate zone. These events are treated as aftershocks.

The events that belong to categories 2 and 3 are either in the clustered time or in the clustered coordinate zone. In order to satisfy at least two clustered zone criteria they undergo magnitude based clustering.

The events that belong to category 1 are not located near the mainshock, in terms of geographical area, and are not associated with time immediately after the mainshock. In order to determine the aftershocks in this region the events undergo a magnitude based thresholding.

3.5 Stage III: magnitude based cluster identification

3.5.1 Magnitude clustering

Events that belong to categories 2 and 3 undergo magnitude based clustering. An event is classified as an aftershock if it belongs to the clustered magnitude zone, where it satisfies either the location (coordinate)-magnitude or the time-magnitude criteria. Therefore the data set (consisting of Category 2 and 3 events) undergoes cluster analysis with the K-means algorithm (Jain, 2010; Xu and Wunsch, 2005; Nazeer and Sebastian, 2009). The steps involved are outlined as follows:

– *Step 1*. Initialize the current iteration as $i = 1$ and let the maximum number of iterations be $i_{\max} = I$.

– *Step 2*. Let r be the number of randomly selected patterns declared as the cluster centers

$$\kappa = \{\kappa_1, .., \kappa_r\} \quad (9)$$

such that $\kappa_i \neq \kappa_j, \forall i, j = 1, ..., r$. Here r is taken as 2 as we have to divide the data set into two groups (clustered and regular magnitude zone).

For i=1 to $i_{\max}$ do

– *Step 3*. For each pattern $p_i \in P_{N \times D}$ (catalog having N events, D dimension) calculate the Euclidean distance from the cluster centers $d(p_i, \kappa_j)$ defined in Eq. (4).

– *Step 4*. Assign pattern $\boldsymbol{p_i}$ to the cluster center $\boldsymbol{\kappa_j}$ for which the distance is minimum as per Eq. (5).

– *Step 5*. Update the center by the mean of the associated patterns with the center $\boldsymbol{\kappa_j}$.

End For

– *Step 6*. Check whether the number of data points in each cluster remains constant for certain iterations. If this is not satisfied then increase the number of iteration $i_{\max}$ and continue *Step 3*.

– *Step 7*. Otherwise report the final cluster partitions κ_j obtained at iteration $I_{\max}$.

The algorithm is run for 30 iterations and the obtained partitions are described as follows:

1. Clustered magnitude zone: events that belong to this category are treated as aftershocks.

2. Regular magnitude zone: events that belong to this category are treated as background, as they failed to satisfy the two clustered zone criteria with magnitude.

3.5.2   Magnitude thresholding

The events belonging to Category 1 (within $C1$ and C3, i.e., events which are located far away with respect to time and coordinate from the mainshocks) are categorized into aftershocks or background events based upon magnitude based thresholding. For all events $p_i \in \kappa_j$ the thresholding process is as follows:

$$\text{If } (\text{Mag}[p_i] > M1) \Rightarrow \begin{cases} p_i \in \text{Aftershock} \\ \text{else, } p_i \in \text{Background,} \end{cases} \quad (10)$$

where $M1 = \text{mean}\big[\text{Mag}(\text{Category4})\big]$. This represents the average magnitude of events already classified as aftershocks. Those events that are far from the mainshock with respect to coordinate and time are considered as aftershocks if they have a higher magnitude than $M1$.

3.6   Results and discussion

The locations of events in the synthetic catalog that are categorized as either aftershock or background, as derived from the model output, are presented in Fig. 5 (middle left and right). The results are compared with the true values of aftershocks and background in Fig. 5 (bottom left and right, respectively). From these figures, it is observed that the density of background events in the synthetic catalog is almost uniform throughout the region. It is also observed that there is difference between true aftershocks (Fig. 5, bottom left) and the proposed model classified aftershocks (Fig. 5, middle left). This is likely due to the absence of spatial clusters in the true aftershocks. While the model accurately classifies the temporal clusters present in the catalog, because there is almost no spatial clustering present it effectively chooses those events randomly in the coordinate based clustering step. As a result, the model is insensitive to change in parameter value ($1/\psi$ value in Table 2) and the number of aftershocks is overestimated. Note also that the method does not identify precursory foreshocks, or NC-PAST signals. However, those events are a very small percentage of the total and as a result, other, standard declustering algorithms are known to experience similar difficulties (Mignan and Tiampo, 2010).

3.6.1   Accuracy Percentage (AP)

$$\text{AP} = \frac{\text{Correctly classified events}}{\text{Total number of events}} \times 100. \quad (11)$$

Table 2. Impact of different $1/\psi$ value on model outcome.

| Classification Result	Model Output			
Total Events	29355 (True BG-19669, True AF-9686)			
$1/\psi$ value	1/5	1/6	1/7	1/8
Number of events as background	17 860	18 361	18 648	18 796
% of Accuracy of background events	90.80	93.35	94.81	95.56
Index matching with true backgrounds	12 354	12 720	12 946	13 059
Number of events as aftershocks	11 495	10 994	10 707	10 559
% of Accuracy of aftershock events	81.32	86.50	89.46	90.99
Index matching with true aftershocks	3796	4037	4172	3941

This evaluation criteria signifies the overall performance of the model considering the time, coordinate and magnitude clustering. The events classified as aftershocks and background are compared with the ideal classification result of synthetic catalog and the results are presented in Table 2.

3.6.2   Index matching

The index matching of the obtained model output (events categorized as background and aftershocks) with the true output (available for synthetic catalog) represents the number of events that are classified correctly with the model. The results of index matching of events for the synthetic catalog are presented in Table 2.

3.6.3   Optimal coordinate based thresholding

The value for parameter $1/\psi$, discussed in coordinate based thresholding, is set in accordance with the area of coordinate based clusters. The final classification results of the model for four different $1/\psi$ values are presented in Table 2. It is observed that the model outputs provide similar results to the ideal classification obtained from the synthetic catalog with a suitable selection of the $1/\psi$ value. For $1/\psi = 1/7$, the accuracy of background prediction is 94.81 % and that of aftershock prediction is 89.46 %, both of which are better than that achieved for background and aftershock predictions with values for $1/\psi$ of 1/5 and 1/6. Further reduction of the $1/\psi$ value to 1/8 does improve the accuracy percentage of background and aftershock events, but the index matching of aftershocks reduces from 4172 to 3941 (Table 2). For $1/\psi = 1/9$, the index matching of aftershocks further reduces to 3798. This indicates that lowering the threshold ($1/\psi$) below a certain optimal value increases the
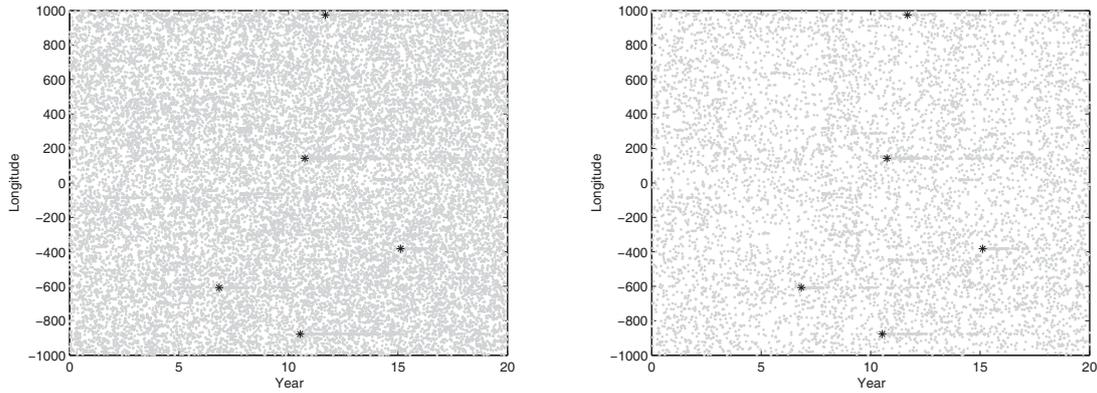
Fig. 6. Plot of space vs. time for (left) the synthetic catalog presented in Fig. 3 (top left), (right) declustered synthetic catalog obtained with the proposed tri-stage cluster identification model presented in Fig. 5 (bottom left).
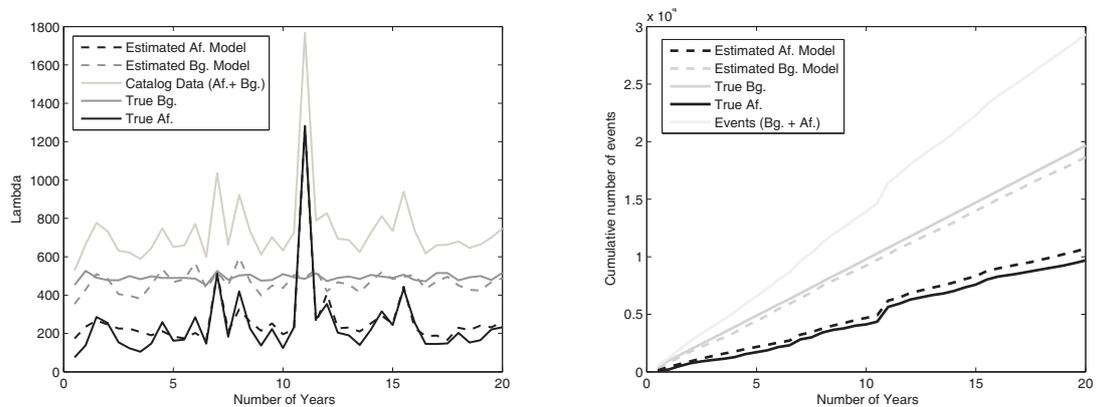
Fig. 7. Tri-stage cluster identification model result analysis for the synthetic catalog: (left) plot of Lambda values, (right) cumulative plot of events characterized as true aftershock and background for the synthetic catalog and estimated tri-stage model outputs.

misclassification rate of aftershocks. This is because aftershocks are more prominent within a certain coordinate radius of the mainshock and that this optimal radius varies with tectonic region and mainshock.

3.6.4 Space time plot

The plot of longitude vs. time for the synthetic catalog and declustered synthetic catalog obtained with the proposed tri-stage cluster identification model are presented in Fig. 6 (left and right, respectively). From both figures it is observed that there is a dense region of aftershock decay with time after the occurrence of the mainshocks.

3.6.5 Lambda and cumulative plot

The plot of the lambda (λ = number of events yr^{-1}) for the synthetic catalog is shown as the solid light-gray line in Fig. 7 (left). From Fig. 3 (top right) and the lambda plot in Fig. 7 (left), it is clear that three mainshocks occurred during the 10–12 yr and two mainshocks occurred during the 6–8 yr and 15–16 yr periods, respectively. Therefore, the number

of events (primarily aftershocks) increases sharply in these periods and are represented by peaks. The ideal λ values of backgrounds and aftershocks generated by the synthetic catalog are represented by the solid gray line and the solid black line, respectively. It is observed that the λ value of the backgrounds remains almost uniform throughout the data set, whereas the λ value of the aftershocks approach the characteristics of the total events. The λ values of backgrounds and aftershocks generated by the proposed cluster identification model are shown in Fig. 7 (left) as the dotted gray line and the dotted black line, respectively. The close agreement between lambda plots of the synthetic catalog and that generated by the proposed model for both aftershocks and backgrounds, demonstrates the accurate performance of proposed method.

The plot of cumulative number of events with respect to time for the synthetic catalog is shown in Fig. 7 (right). In this plot it can be observed that the cumulative number of aftershock events increases during the year in which the mainshocks occurred (i.e., 10–12, 6–8 and 15–16 yr). The cumulative number of background events follows its steady pattern for all the years.
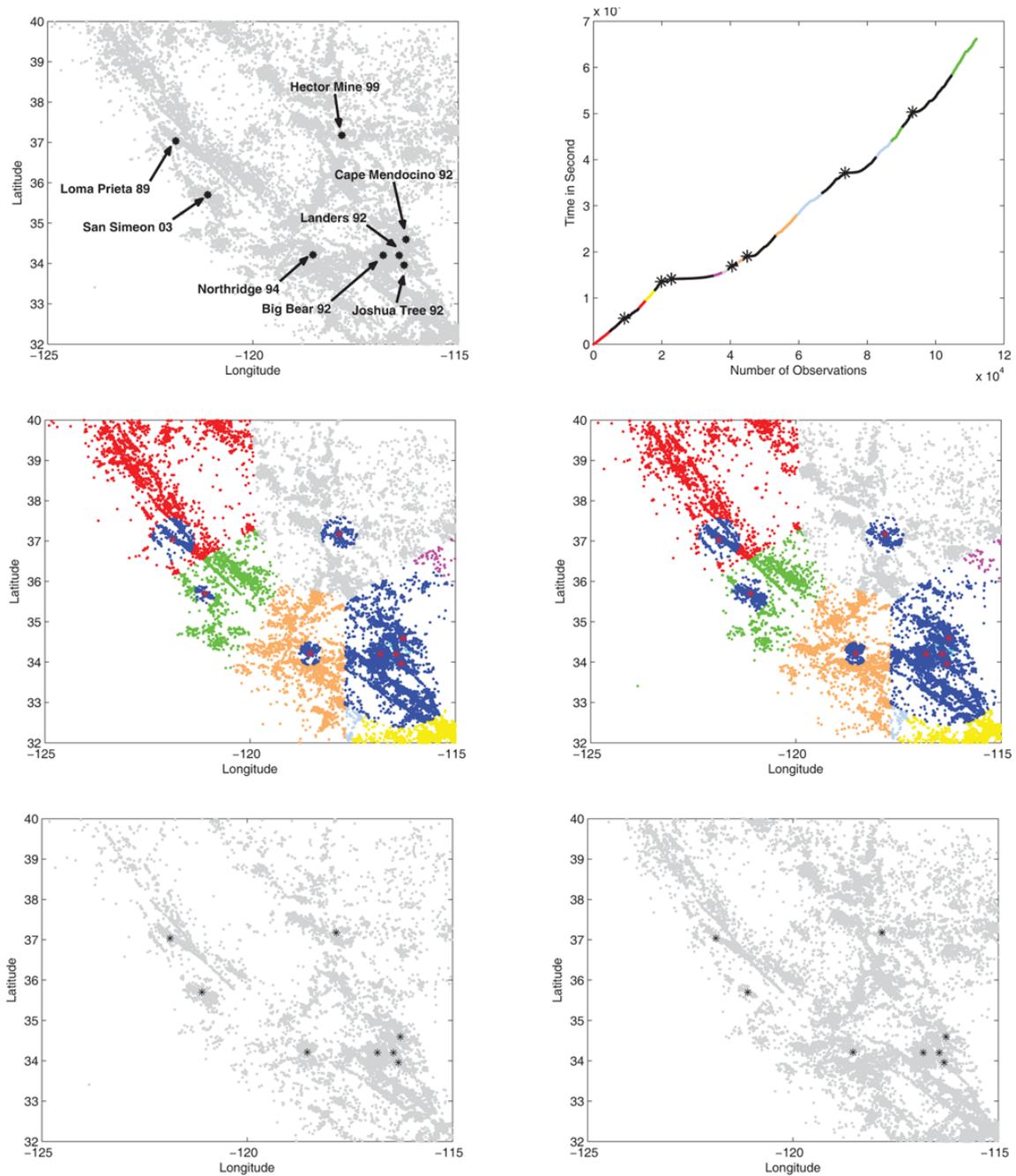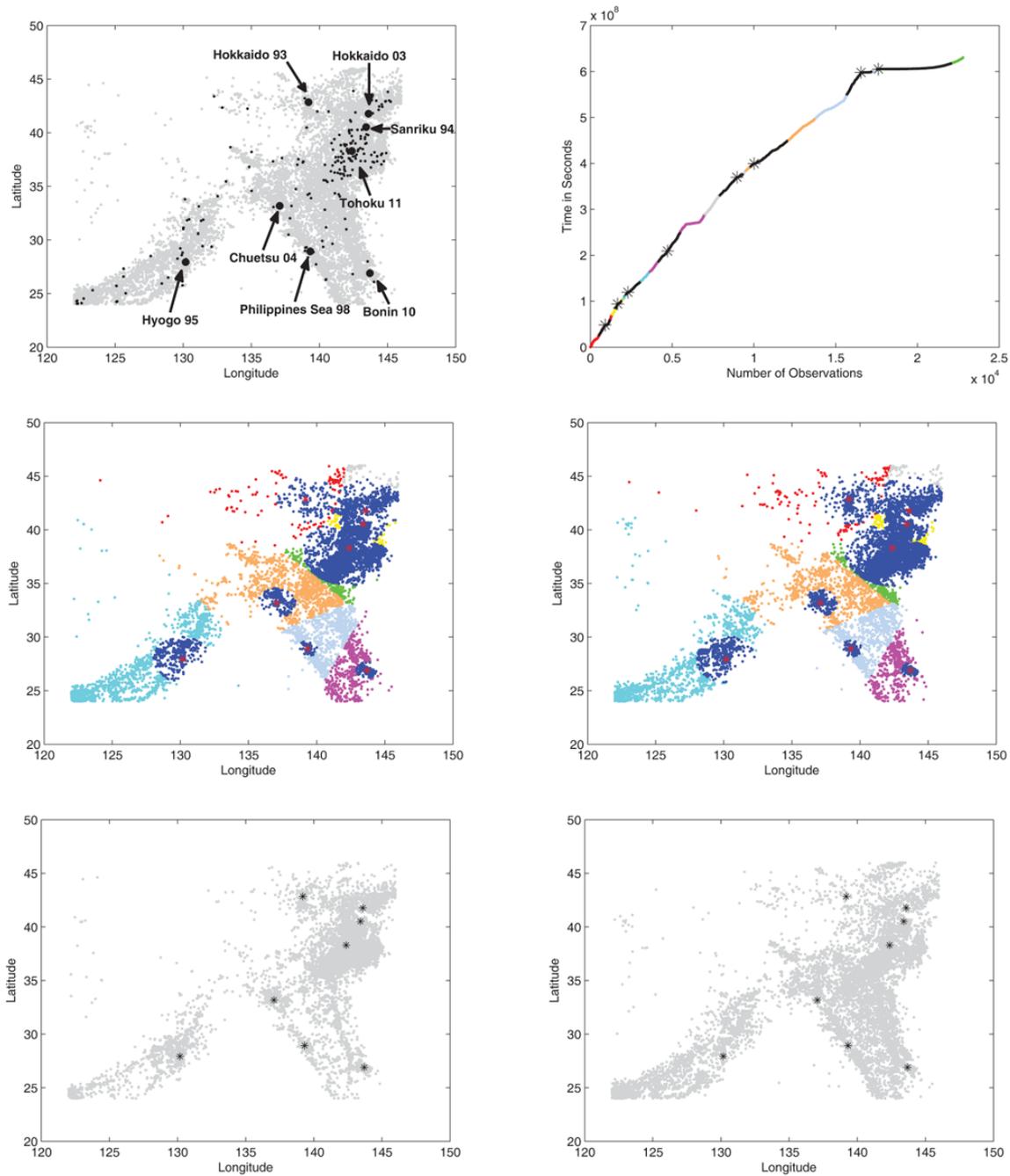
Fig. 8. Tri-stage cluster identification of the seismic catalog of California: (top left) location of mainshocks during year 1988–2008; (top right) output of time zone clustering; (middle left) output of background and clustered coordinate zone for $C1$; (middle right) output of background and clustered coordinate zone for $C2$; (bottom left) events classified as aftershocks; (bottom right) events classified as background.

4   Analysis of natural seismic catalogs using proposed model

In this section the proposed tri-stage cluster identification model is applied to the California, Japan and Indonesian catalogs obtained from the Advanced National Seismic System (ANSS, 2012) with the parameter settings as presented in Table 3. The details of the model analysis are outlined below.

4.1   California catalog

This catalog consists of seismic events occurred between January 1988 to December 2008. Locations of the eight major events, $M > 6$, are represented by black stars and their location and year of occurrence are indicated in Fig. 8 (top left). We converted the time of occurrence of the events from year, day, hour and seconds scale to seconds. The occurrence
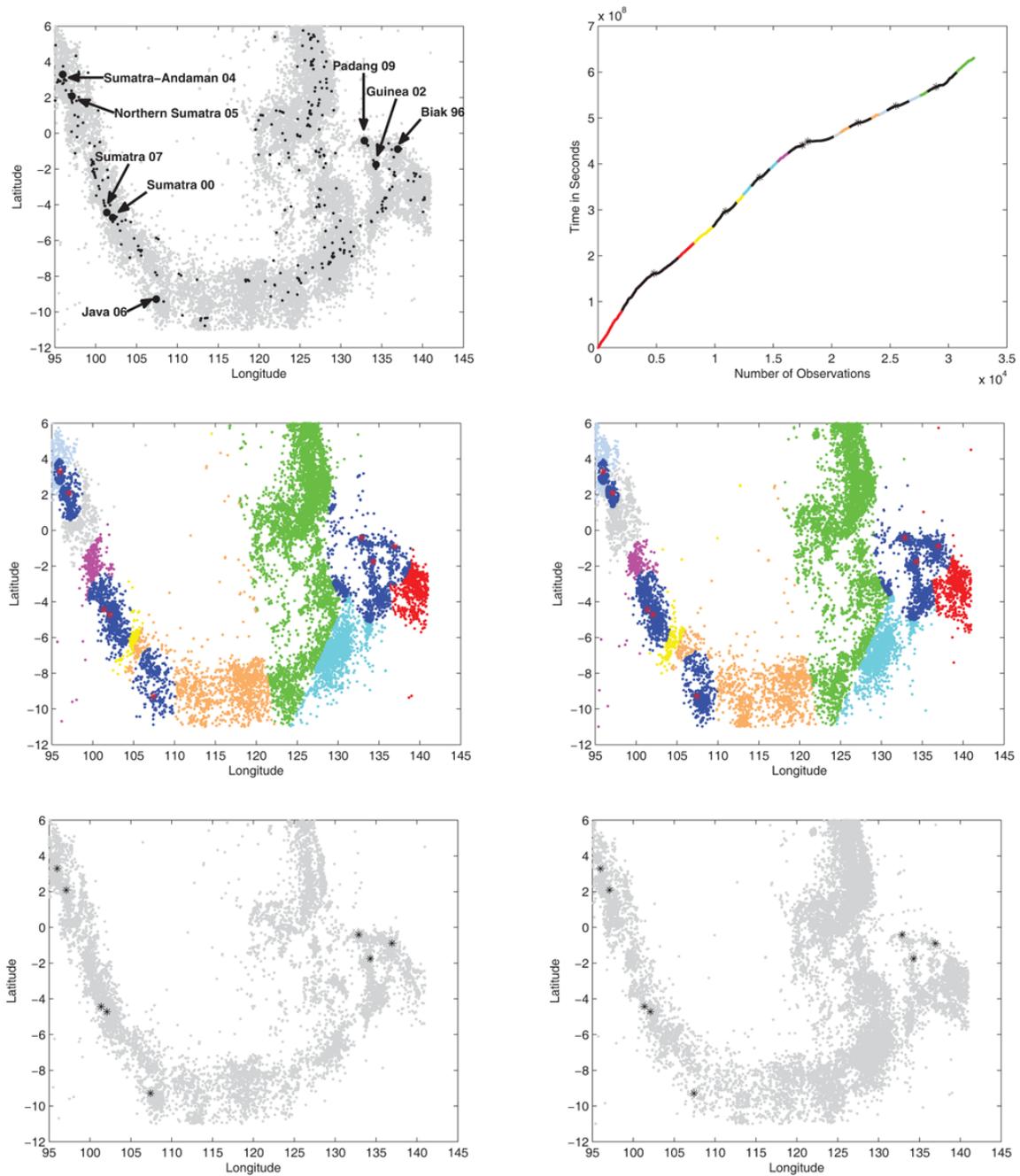
Fig. 9. Tri-stage cluster identification of the seismic catalog of Japan: (top left) location of mainshocks, designated by black asterisks, and location of events $M > 6$, designated as black dots, 1992–2011; (top right) output of time zone clustering; (middle left) output of background and clustered coordinate zone for $C1$; (middle right) output of background and clustered coordinate zone for $C2$; (bottom left) events classified as aftershocks; (bottom right) events classified as background.

of mainshocks with respect to seconds is presented by black asterisks in Fig. 8 (top right). Four mainshocks occurred during the year 1992, and two event occurrences coincide (represented by seven black asterisks in Fig. 8 (top right) instead of eight. The catalog events then are classified into eight clusters with respect to the time based clustering, represented by

eight dots in Fig. 8 (top left). The events from each cluster then undergo clustering for each time zone. Events that belong to the clustered time zone are represented by black solid lines around the mainshocks in Fig. 8 (top right).

The events belonging to the normal time zone and clustered time zone both undergo coordinate based clustering
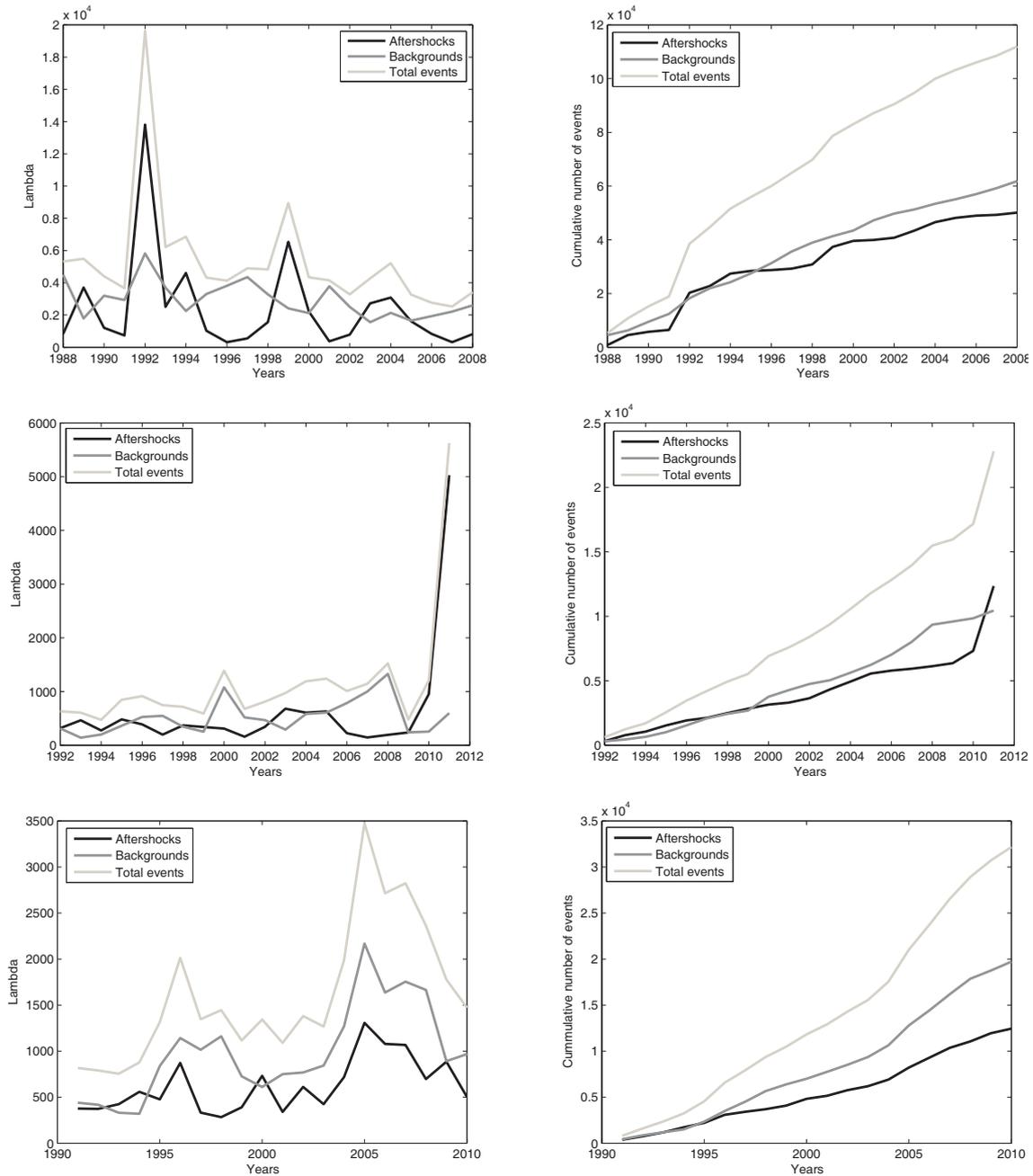
Fig. 10. Tri-stage cluster identification of the seismic catalog of Indonesia: (top left) location of mainshocks, designated as black asterisks, and location of events $M > 6$, designated as black dots, 1991–2010; (top right) output of time zone clustering; (middle left) output of background and clustered coordinate zone for $C1$; (middle right) output of background and clustered coordinate zone for $C2$; (bottom left) events classified as aftershocks; (bottom right) events classified as backgrounds.

and thresholding as described in Sect. 3.3. The results obtained after coordinate thresholding are represented in Fig. 8 (middle left and right, respectively). In both figures the entire region is divided into eight clusters (each cluster event is represented by a unique colour). Red asterisks represent the coordinates of the mainshocks surrounded by the associated coordinate zones (represented by blue dots).

4.1.1 Variable coordinate based thresholding

In this catalog four large earthquakes occurred in 1992 and the locations of their associated events are represented by four adjacent clusters (pink, cyan, sky blue and yellow) in Fig. 8 (middle left and right, respectively). These events are located near each other and the number of aftershock events

Fig. 11. Plot of lambda values and cumulative number of events characterized as aftershocks and background using the proposed tri-stage cluster identification model: (top left-right) lambda and cumulative plot for California catalog, respectively; (middle left-right) lambda and cumulative plot for Japan catalog, respectively; (bottom left-right) lambda and cumulative plot for Indonesian catalog, respectively.

in the regions surrounding these mainshocks is greater than the other four individual mainshocks. Therefore, instead of considering a fixed optimal coordinate thresholding value for all mainshocks (as in the synthetic data) the zone radius associated with these four events is increased. For the 1992 earthquakes, the $1/\psi$ value is taken as 2/3 and for the other four this value is taken as 1/8. The events are then classified into four categories as described in Sect. 3.4. Events that belong

to category 4 are taken as aftershocks. The events belong to categories 2–3 undergo a magnitude clustering and those of category 1 undergo magnitude thresholding.

4.1.2 Variable magnitude based thresholding

We have defined two different magnitude based thresholds. The magnitude threshold for the 1992 related
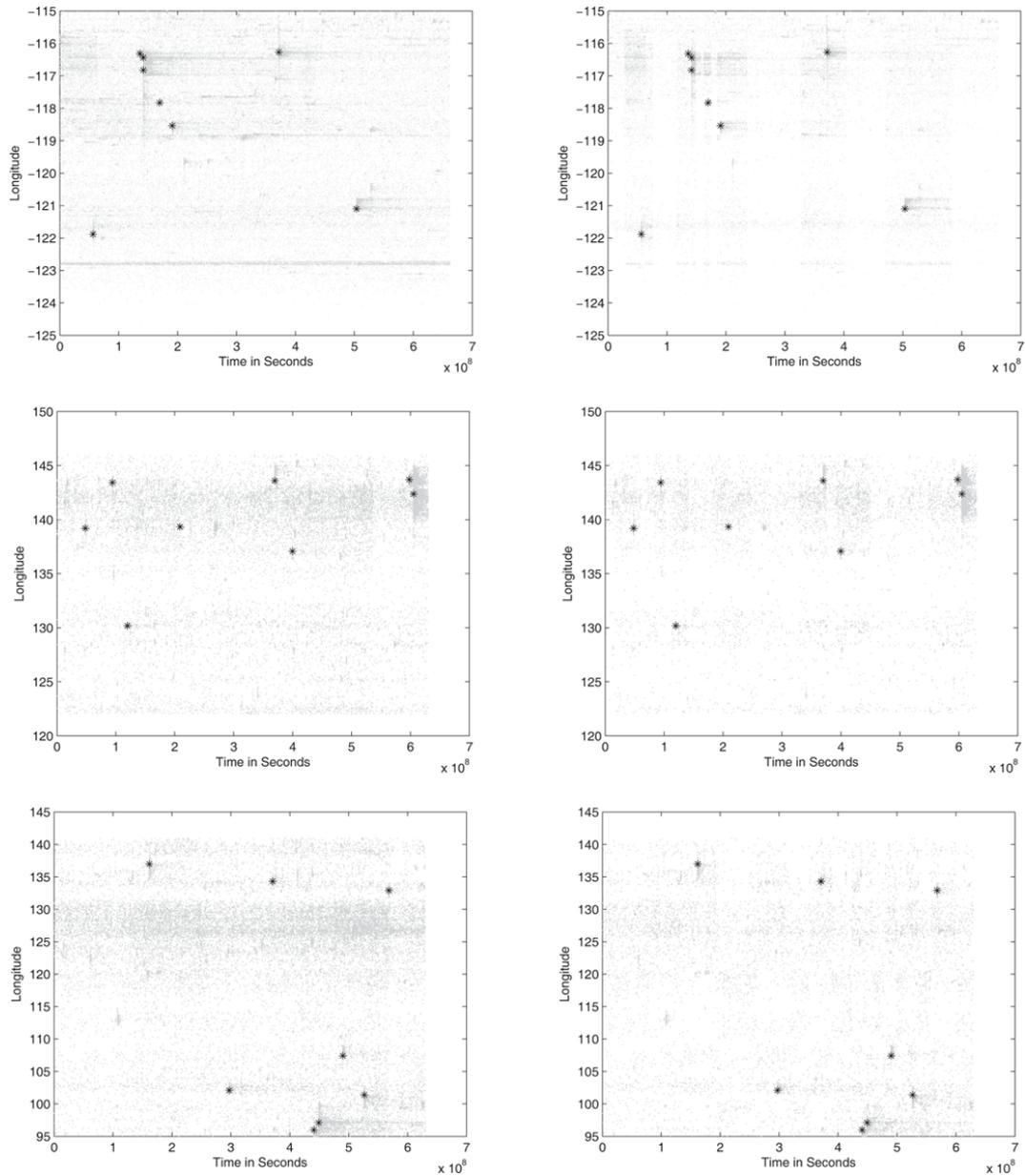
Fig. 12. Plot of longitude vs. time for true and declustered catalogs obtained with the proposed tri-stage cluster identification model: (top left–right) true and declustered catalog plots for California catalog, respectively; (middle left–right) true and declustered catalog plots for Japan catalog, respectively; (bottom left–right) true and declustered catalog plots for Indonesian catalog, respectively.

clusters is taken as $\left[\text{mean}\,(\text{Mag}(\text{Category4})) - 0.25\right]$, as four major earthquakes happened in the same year. For the remaining four mainshocks the value is taken as $\left[\text{mean}\,(\text{Mag}(\text{Category4})) + 1.25\right]$. The locations of those events identified as aftershock or background in the tri-stage model output are presented in Fig. 8 (bottom left and right, respectively). It is observed that the density of background events is almost uniform throughout the region (Fig. 8, bottom right), while the aftershock density is higher near the mainshocks (Fig. 8, bottom left).

4.2 Japan catalog

The Japan catalog consists of seismic events that occurred between January 1992 and December 2011, which includes the 2011 Tohoku earthquake and tsunami. Unlike the California catalog, there are many events which are magnitude > 6 (represented by black dots in Fig. 9, top left). As a result, here the mainshocks are identified as events of $M > 6.8$ and isolated from each other with respect to time of occurrence. Their location and year of occurrence are indicated as black dots in Fig. 9 (top left). The rest of

**Table 3.** List of parameters used to acquire the California, Japan and Indonesian catalogs (ANSS, 2012).

Parameters	California	Japan	Indonesian
Start time	1988/01/01	1992/01/01	1991/01/01
End time	2009/01/01	2012/01/01	2011/01/01
Minimum latitude	32	24	−11
Maximum latitude	40	46	6
Minimum longitude	−125	122	95
Maximum longitude	−115	146	141
Minimum magnitude	2.0	2.0	2.0
Maximum magnitude	8.0	10.0	10.0
Minimum depth	0	0	0
Maximum depth	25	500	500
Event type	E	E	E

the analysis is carried out in a similar manner as in the California catalog. The $1/\psi$ value is taken as 2/3 for Tohoku (2011), Hokkaido (2003) and Sanriku (1994) quakes, and for the other events it is taken as 1/7. Similarly, for the three events near Tohoku region, the magnitude based threshold value is taken as $\left[\text{mean}\left(\text{Mag}(\text{Category4})\right) - 0.25\right]$ and for the remaining events the threshold is set as $\left[\text{mean}\left(\text{Mag}(\text{Category4})\right) + 1.5\right]$.

4.3   Indonesian catalog

The Indonesian catalog consists of events occurring between January 1991 and December 2010, including the Sumatran earthquake and tsunami of 2004. Like the Japan catalog, there are many events of $M > 6$ (shown as black dots in Fig. 10, top left). In this case, mainshocks are identified as events of $M > 7.5$ and isolated from each other with respect to time of occurrence (black asterisks in Fig. 10, top right). The proposed cluster analysis is implemented in a similar manner as in those of California and Japan. The $1/\psi$ value is taken as 1/4 for the Sumatran earthquakes (2000, 2004, 2005 and 2007) and 1/6 for remaining events. Similarly, the magnitude based thresholding value is taken as $\left[\text{mean}\left(\text{Mag}(\text{Category4})\right) - 0.25\right]$ for the Sumatran earthquakes and for the other events the threshold is set as $\left[\text{mean}\left(\text{Mag}(\text{Category4})\right) + 1.5\right]$.

It is observed that the $\text{mean}\left(\text{Mag}(\text{Category4})\right)$ is used in all three catalogs to determine the threshold. The $\text{mean}\left(\text{Mag}(\text{Category4})\right)$ represent the average value of aftershocks in the associated regions (i.e., regions which are close in time and coordinate that of the mainshocks). Therefore, in this study the threshold value for special regions such as for 1992 in California, the Tohoku region of Japan and the Sumatra earthquake region for Indonesia, the threshold value is 0.25 lower than the average threshold value. For the other regions the value is more than 1.25 units greater than the average value, as the events which are far apart in time and coordinate should not be considered as aftershocks, unless
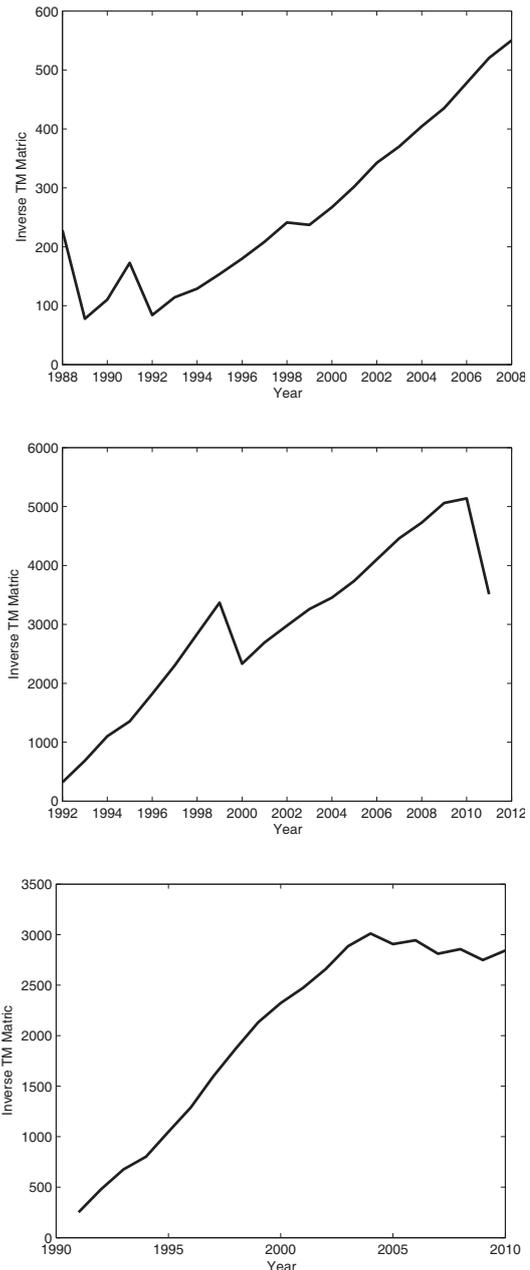
Fig. 13. Plot of inverse TM matric to measure ergodicity present in true catalogs, (top) California catalog with magnitude threshold $M \geq 4$, (middle) Japan catalog with $M \geq 5$, (bottom) Indonesia catalog with $M \geq 5$.

they are a higher magnitude and a relatively rare triggered event. These values are set after a number of repeated runs of the code, considering the optimal performance as represented by the lambda plot. Similarly, two different values of the $1/\psi$ value also are chosen to handle special regions more effectively than that of the regular events.
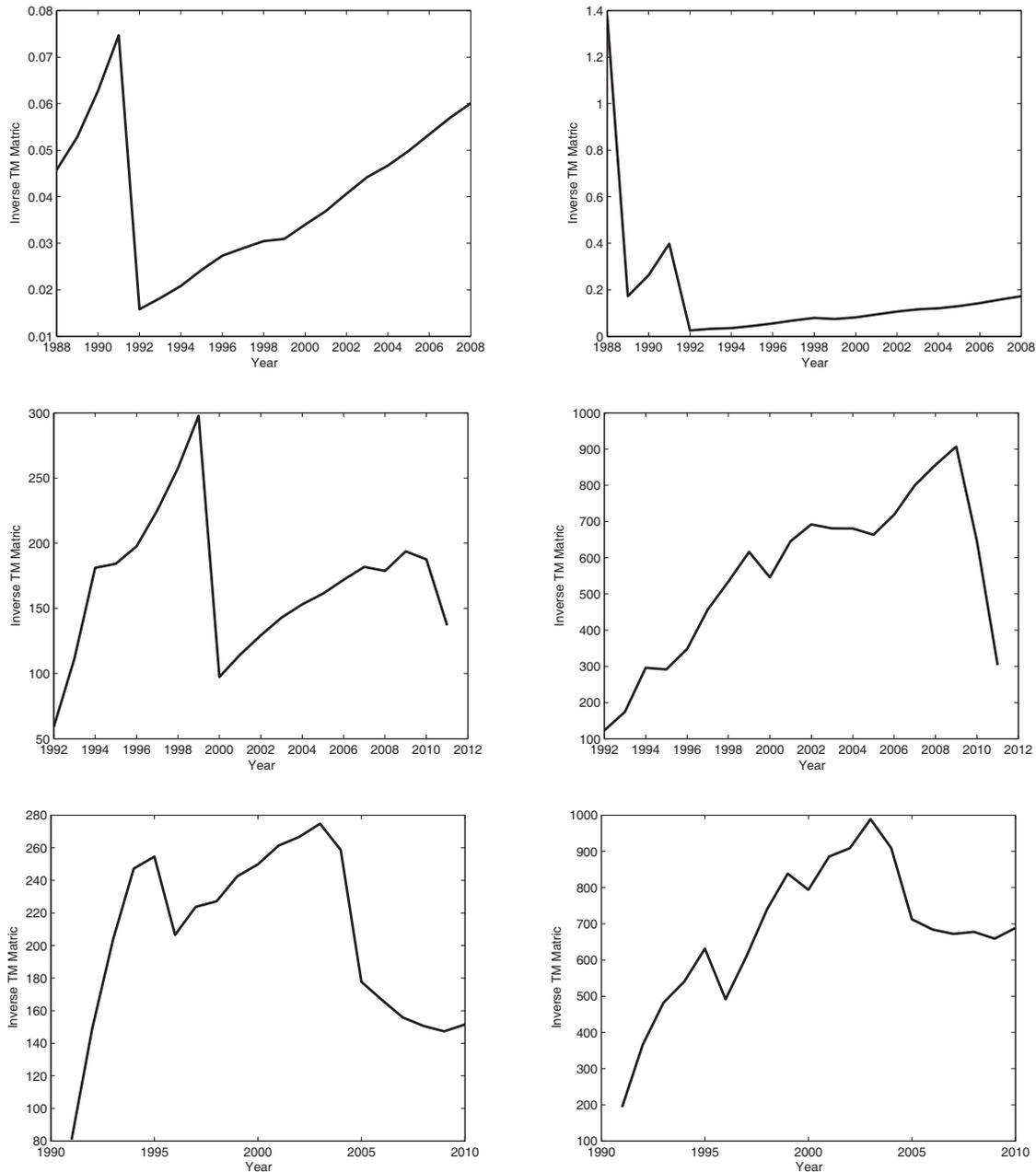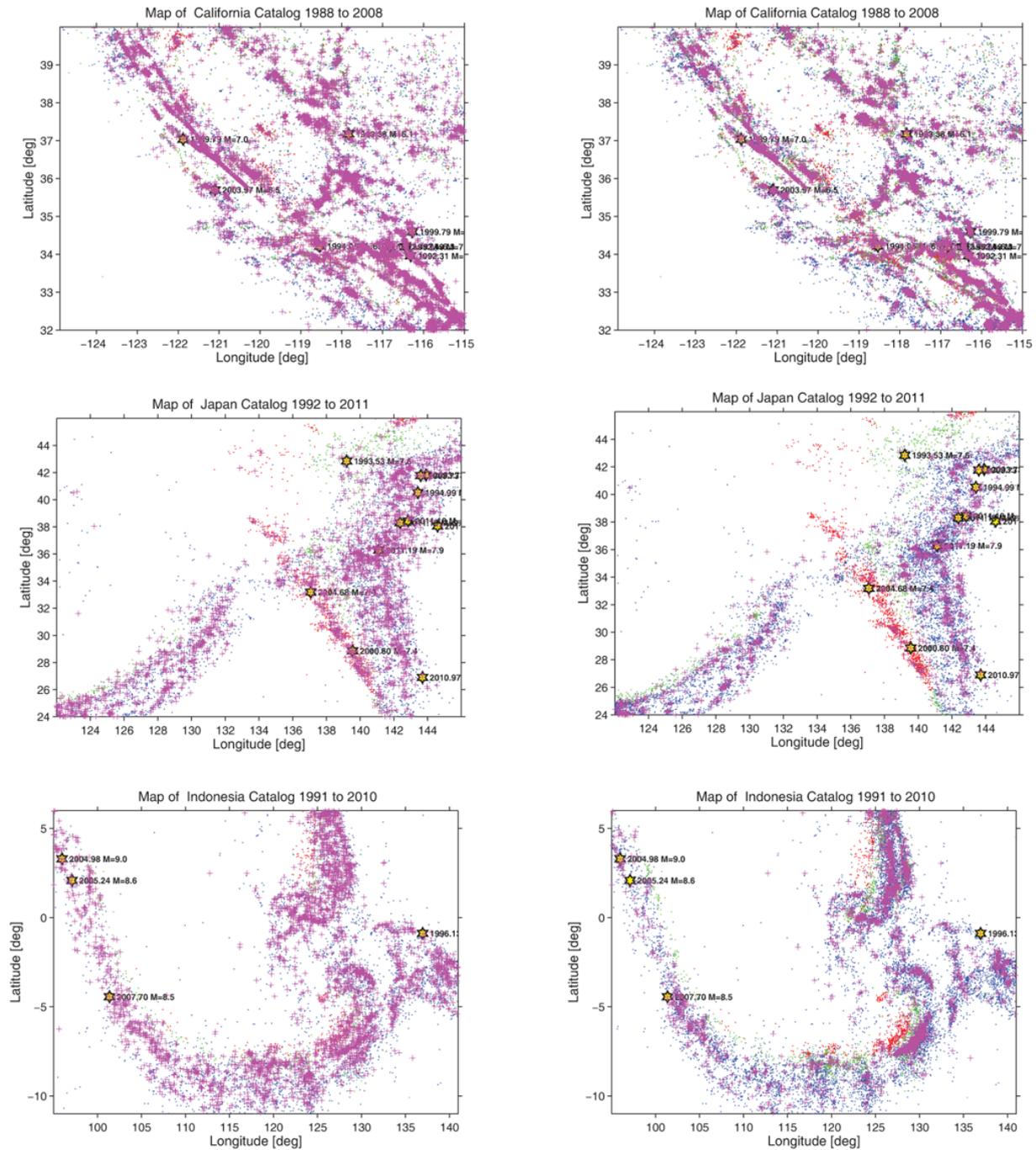
Fig. 14. Plot of the inverse TM matric for true and declustered catalogs obtained with the proposed tri-stage cluster identification model: (top left–right) true and declustered catalog plots for the California catalog, respectively; (middle left–right) true and declustered catalog plots for the Japan catalog, respectively; (bottom left–right) true and declustered catalog plots for Indonesian catalog, respectively.

4.4 Classification results

4.4.1 λ values and cumulative events plots

The plot of λ values for the total number of events, background and aftershocks are presented in Fig. 11 by solid light-gray, gray and black lines, respectively. It is observed that the λ value of the background remains almost uniform throughout the data set, with the exception of the denser

regions, such as California (1992), in the Tohoku (2011) region of Japan, and in the Sumatran region of Indonesia (2004, 2005 and 2007). The corresponding cumulative number of events plotted with respect to time in Fig. 11 supports this observation. In all three natural catalogs the cumulative number of aftershocks increases during the year in which the mainshocks take place. The cumulative number of background

Fig. 15. Comparative plot of declustered catalogs obtained with benchmark methods (Gardner and Knopoff, 1974; Uhrhammer, 1986): (top left–right) GK and Uh declustered catalog plots for the California catalog, respectively; (middle left–right) GK and Uh declustered catalog plots for the Japan catalog, respectively; (bottom left–right) GK and Uh declustered catalog plots for the Indonesian catalog, respectively.

events follow the steady pattern for all years except the special regions mentioned above.

4.4.2 Longitude vs. time plots

The plots of longitude vs. time for the ideal and declustered catalogs obtained with the proposed tri-stage cluster identification model are shown in Fig. 12. Here, it is also observed that there is a dense region of aftershock decay with time after the occurrence of the mainshocks. In general, the absence of mainshock regions is classified as background, reflected by the absence of points when compared with the true mainshock regions.

4.4.3   Inverse Thirumalai-Mountain (TM) metric plots

The inverse TM matric plot represents the ergodicity of the catalog and its associated stationarity, which requires that the ensemble average of events in time and space are equal. Ideally, the inverse TM matric plot should be linearly increasing for stationarity (Thirumalai and Mountain, 1993). In the present paper, for calculating the TM matric the box size is taken as 0.10 in the latitude and longitude directions as described by Tiampo et al. (2007). Figure 13 represents the ergodicity and resulting stationarity of the true catalogs for higher magnitude thresholds(for California catalog $M \geq 4$ shown in Fig. 13, top, for Japan $M \geq 5$ in Fig. 13, middle, for the Indonesian catalog $M \geq 5$ in Fig. 13, bottom). All the three catalogs display piecewise linear behavior associated with stationarity, but simultaneously the catalogs contain only few elements of higher magnitude (only 1019 events are present in the California catalog for $M \geq 4$, 2704 events in the Japan catalog for $M \geq 5$ and 5056 events in the Indonesian catalog for $M \geq 5$).

The minimum magnitude of completeness for California and Japan catalogs are described in Wiemer and Wyss (2000). In accordance with their results, in the current analysis the minimum magnitude of threshold is set as 2.0 for California and Japan. A minimum magnitude of 2.0 also is set for the Indonesian catalog, although this is lower than that determined by Woessner and Wiemer (2005). The computation of inverse TM matric is carried out in similar manner as given in Tiampo et al. (2007). The plots obtained for the true and declustered catalogs obtained with the proposed model are presented in Fig. 14. It is observed that while the true catalogs Fig. 14 (left panels) for California and Japan do show ergodic periods, although they are not as long as those in Fig. 13, the true Indonesian catalog is no longer ergodic. The declustered catalogs obtained with the proposed method shown in Fig. 14 (right panels) reflect similar characteristics as that of the true catalogs (for California, Japan and Indonesian catalogs). Therefore it is concluded that while the true and declustered catalogs for California and Japan are stationary for some periods, the Indonesian catalog is only intermittently stationary. However, the declustered catalogs do contain most of the significant characteristics of the true catalog.

4.4.4   Comparison with standard declustering methods

The proposed method has been compared with two standard declustering methods available in the ZMAP package (Gardner and Knopoff, 1974; Uhrhammer, 1986). The Gardner and Knopoff method (Gardner and Knopoff, 1974) declusters a catalog using a time–space windowing technique. For every predefined earthquake the subsequent shocks are identified as aftershocks if they occur within a specified interval in time and distance. Here the time–space window is chosen in accordance with the magnitude of the largest event present

Table 4. Comparative results for the proposed model and that of Gardner and Knopoff (1974) and Uhrhammer (1986) for declustering of California, Japan and Indonesian catalogs.

Methods	Catalogs	California	Japan	Indonesian
	Total events	111 941	22 794	32 168
Gardner and Knopoff	BG events	22 556	6810	7539
	AF Clusters	7780	2040	2912
	AF events	89 385	15 984	24 629
Uhrhammer	BG events	51 034	10 807	16 085
	AF Clusters	7688	1004	1757
	AF events	60 907	11 987	16 083
Proposed model	BG events	61 830	12 347	19 718
	AF events	50 111	10 447	12 450

in a sequence. The Uhrhammer (Uhrhammer, 1986) method is relatively simple and uses a lower complex window (Van Stiphout et al., 2012) than the Gardner and Knopoff method. The results of the proposed method, in the form of declustered catalogs, are compared with the two benchmark methods in Fig. 15 and Table 4. Comparing the plots in Fig. 8 (bottom right) and Fig. 15 (top left and right) for the California catalog, it is observed that the proposed method does not overclassify events as aftershocks, as occurs in both the standard methods. Similar observations were made for Japan (Fig. 9 bottom right, Fig. 15, middle left and right) and Indonesia (Fig. 10 bottom right, Fig. 15, bottom left and right). Table 4 reflects the numerical values, confirming that this method identifies fewer events as aftershocks for all three catalogs, when compared to the two benchmark methods.

5   Conclusions

In this paper we have proposed a tri-stage cluster identification model considering time, location and magnitude information of earthquake catalogs to efficiently classify the events of the catalogs as aftershocks and backgrounds. Tests on a synthetic Non-Critical PAST catalog demonstrate the efficiency of the proposed model in terms of percentage of accuracy (94.81 % for background and 89.46 % for aftershocks). The close agreement between curves generated by true events and that of the proposed model for the synthetic catalog in the form of the lambda plot and the cumulative plot supports the accuracy of the proposed method. However, this method does not successfully identify the precursory foreshocks, or NC-PAST signals. There is flexibility in the design of the model for the proper selection of location and magnitude zones depending upon the nature of the mainshocks present in the catalog. In this method, there is no requirement to identify and parameterize an a priori probability distribution for the relationship between events, but relies only on the characteristics of the seismicity itself, as controlled by the local or regional tectonics. For example, the variable $1/\psi$

provides for proper selection of danger zones and enables the model to adapt based upon the tectonic region.

The adaptability and effectiveness of the proposed model is evaluated through the classification of events from three historic catalogs from different tectonic regions: California, Japan and Indonesia. In both the synthetic and real catalog modeling, it is observed that the density of events classified as background is almost uniform throughout the entire region of each data set except where there is an unusually dense occurrence of mainshocks, whereas the density of aftershock events is higher adjacent to the mainshocks. It is observed that in both synthetic and historic catalogs the model does a good job in identifying the temporal clusters. In addition, the method effectively determines the spatial clusters that are present in the natural catalogs. The superior performance of the proposed method is demonstrated by comparison with two benchmark methods (Gardner and Knopoff, 1974; Uhrhammer, 1986) in terms of number of events classified and their distribution in the declustered catalog. While further studies are required to determine the extent of its applicability to finer spatial and temporal scales, this relatively simple three-stage technique, as compared to other more complicated declustering algorithms currently available, shows promise for the application of relatively new clustering methods to seismicity studies.

References

Advanced National Seismic System: available at: http://www.ncedc.org/anss/catalog-search.html, last access: 15 August 2012.

Baiesi, M. and Paczuski M.: Scale-free networks of earthquakes and aftershocks, Phys. Rev., 1–7, doi:10.1103/PhysRevE.69.066106, 2004.

Bansal, A. R. and Ogata, Y.: Seismicity activation before the mega-earthquake of 26 December 2004 based on Epidemic type aftershock sequence (ETAS) model, AGU Fall Meeting Abstracts, Vol. 1, 2084, 2010.

Bansal, A. R., Dimri, V. P., and Babu, K. K.: Epidemic type aftershock sequence (ETAS) modeling of northeastern Himalayan seismicity, J. Seismol., 1–10, doi:10.1007/s10950-012-9314-7, 2012.

Gardner, J. K. and Knopoff, L.: Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, Bull. Seis. Soc. Am., 64, 1363–1367, 1974.

Gutenberg, B. and Richter, C. F.: Frequency of earthquakes in California, Bull. Seismol. Soc. Am., 34, 185–188, 1944.

Helmstetter, A. and Sornette, D.: Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, J. Geophys. Res., 107, 1–29, doi:10.1029/2001JB001580, 2002.

Helmstetter, A., Kagan, Y. Y., and Jackson, D. D.: Importance of small earthquakes for stress transfers and earthquake triggering, J. Geophys. Res., 110, 1–15, 2005.

Jain, A. K.: Data clustering: 50 years beyond K-means, Pattern Recognition Lett., 31, 651–666, 2010.

King, G. C. P. and Bowman, D. D.: The evolution of regional seismicity between large earthquakes, J. Geophys. Res., 108, 1–16, doi:10.1029/2001JB000783, 2003.

Marsan, D. and Lengline, O.: Extending Earthquakes Reach Through Cascading, Science, 319, 1076–1079, 2008.

Mendoza, C. and Hartzell S. H.: Aftershock patterns and mainshock faulting, Bull. Seismol. Soc. Am., 78, 1438–1449, 1988.

Mignan, A.: Retrospective on the Accelerating Seismic Release (ASR) Hypothesis: Controversy and New Horizons, Tectonophysics, 505, 1–16, doi:10.1016/j.tecto.2011.03.010, 2011.

Mignan, A.: Non-Critical Precursory Accelerating Seismicity Theory (NC PAST) and limits of the power-law fit methodology, Tectonophys., 452, 42–50, doi:10.1016/j.tecto.2008.02.010, 2008.

Mignan, A. and Di Giovambattista, R.: Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes, Geophys. Res. Lett., 35, L15306, doi:10.1029/2008GL035024, 2008.

Mignan, A. and Tiampo, K.: Testing the Pattern Informatics index on synthetic seismicity catalogs based on the Non-Critical PAST, Tectonophysics, 483, 255–268, doi:10.1016/j.tecto.2009.10.023, 2010.

Mignan, A., King, G. C. P., and Bowman, D. D.: A mathematical formulation of accelerating moment release based on the stress accumulation model, J. Geophys. Res., 112, B07308, doi:10.1029/2006JB004671, 2007.

Nazeer, K. A. A. and Sebastian, M. P.: Improving the accuracy and efficiency of the k-means clustering algorithm: Proceedings of the World Congress on Engineering, Vol. 1, July 2009.

Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes, J. Am. Stat. Assoc., 83, 9–27, 1988.

Ogata, Y.: Detection of precursory seismic quiescence before major earthquakes through a statistical model, J. Geophys. Res., 97, 19845–19871, 1992.

Ogata, Y.: Pre-seismic anomalies in seismicity and crustal deformation: case studies of the 2007 Noto Hanto earthquake of M6.9 and the 2007 Chuetsu–Oku earthquake of M6.8 after the 2004 Chuetsu earthquake of M 6.8, Geophys. J. Int., 186, 331–348, doi:10.1111/j.1365-246X.201.05033.x2, 2011.

Ogata, Y. and Zhuang, J.: Space-time ETAS models and an improved extension, Tectonophysics, 413, 13–23, 2006.

Okada, Y.: Internal deformation due to shear and tensile faults in a half-space, Bull. Seismol. Soc. Am., 82, 1018–1040, 1992.

Omori, F.: On the aftershocks of earthquakes, J. Coll. Sci., Univ. Tokyo, 7, 111–200, 1894.

Reasenberg, P.: Second-order moment of central California seismicity, 1969–1982, J. Geophys. Res., 90, 5479–5495, 1985.

Savage, J. C.: A dislocation model of strain accumulation and release at a subduction zone, J. Geophys. Res. 88, 4984–4996, 1983.

Shcherbakov, R., Turcotte, D. L., and Rundle, J. B.: Aftershock statistics, Pure Appl. Geophys., 162, 1051–1076, doi:0.1007/s00024-004-2661-8, 2005.

Stein M. L.: Interpolation of Spatial Data : Some Theory for Kriging, Springer, 1999.

Thirumalai, D. and Mountain, R. D.: Activated Dynamics, Loss of Ergodicity, and Transport in Supercooled Liquids, Phys. Rev., E 47, 479–489, 1993.

Tiampo, K. F., Rundle, J. B., Klein, W., Holliday, J., Sa Martins, J. S., and Ferguson, C. D.: Ergodicity in natural earthquake fault networks, Phys. Rev., E 75, 1–16, doi:10.1103/PhysRevE.75.066107, 2007.

Turcotte, D. L., Holliday, J. R., and Rundle J. B.: BASS, an alternative to ETAS, Geophy. Res. Lett., 34, 1–5, doi:10.1029/2007GL029696, 2007.

Uhrhammer, R.: Characteristics of northern and central California seismicity, Earthquake Notes, 1, 57, 1986.

Utsu, T.: Statistical study on the occurrence of aftershocks, Geophys. Mag., 30, 521–605, 1961.

Utsu, T.: A list of deadly earthquakes in the World: 1500–2000, in: International Handbook of Earthquake and Engineering Seismology Part A, Academic Press, San Diego, 691–717, 2002.

Van Stiphout, T., Schorlemmer, D., and Wiemer, S.: The Effect of Uncertainties on Estimates of Background Seismicity Rate, BSSA, 101, 482–494, 2011.

Van Stiphout, T., Zhuang, J., and Marsan, D.: Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-52382934, available at: http://www.corssa.org, 2012.

Wiemer, S. and Wyss, M. : Minimum magnitude of completeness in earthquake catalogs : Examples from Alaska, the Western United States, and Japan, Bull. Seismol. Soc. Am., 90, 859–869, 2000.

Woessner, J. and Wiemer, S.: Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, Bull. Seismol. Soc. Am., 95, 684–698, 2005.

Wu, Z.: A hidden Markov model for earthquake declustering, J. Geophys. Res., 115, 1–11, doi:10.1029/2008JB005997, 2010.

Xu, R. and Wunsch, D.: Survey of Clustering Algorithms, IEEE Trans. on Neural Networks, 13, 645–678, 2005.

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., and Wong, H.: Clustering Analysis of Seismicity and Aftershock Identification, Phys. Rev. Lett., 101, 1–4, doi:10.1103/PhysRevLett.101.018501, 2008.

Zhuang, J. C., Chang, C. P., Ogata, Y., and Chen, Y. I.: A study on the background and clustering seismicity in the Taiwan region by using a point process model, J. Geophys. Res., 110, B05S18, doi:10.1029/2004JB003157, 2005.