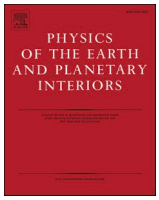
Physics of the Earth and Planetary Interiors 350 (2024) 107182

Contents lists available at ScienceDirect

Physics of the Earth and Planetary Interiors

journal homepage: www.elsevier.com/locate/pepi



Data-driven identification of earthquake clusters: Clusters before the 2010 El mayor-Cucapah earthquake M_W 7.1, Baja California, Mexico

F. Alejandro Nava [a], Lenin Ávila-Barrientos [a,b,*]

[a] CICESE, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California, División de Ciencias de la Tierra, Departamento de Sismología, Carretera Ensenada–Tijuana No. 3918, Zona Playitas, C. P. 22860 Ensenada, Baja California, Mexico
[b] CONAHCYT, Dirección Adjunta de Investigación Humanística y Científica, Av. Insurgentes Sur 1582, Col. Crédito Constructor, Alcaldía Benito Juárez, C.P. 03940 Ciudad de México, Mexico

ARTICLE INFO

Keywords:
Seismic clusters
Seismic precursors
Baja California
El mayor-Cucapah earthquake
Seismic monitoring

ABSTRACT

Seismic clusters in background seismicity have been associated with high stress levels and can be an important precursor to large earthquakes, but there is not a unanimous concept of cluster and most cluster identification methods are cumbersome and involve a priori assumptions. We propose a simple definition of seismic cluster and a straightforward method of identification involving a minimum of parameters that can be objectively determined in a data-driven way according to a principle of low random occurrence. As an illustration, definition and method were applied to the identification of cluster activity from October 1979 to March 2010 in northern Baja California, Mexico, between 118°W to 113°W and 30°N to 33°N, a tectonically complex seismic region with several fault systems. Twenty-one clusters were identified, of which 17 located around the places at the northeastern corner of the study area that would be ruptured on April 4, 2010 by the El Mayor-Cucapah M_w 7.1 earthquake, the largest recorded earthquake in Baja California, Mexico, and the four others occurred within 9 km from its epicenter. Clustering also became slightly more frequent as the time of the earthquake approached, so that if the clustering survey had been carried out before the whole northern Baja California area, the clustering might have identified the future epicentral region as a region of interest to be closely monitored (this earthquake featured foreshock activity starting some 15 days before the main event). Although the reliability of clusters as precursors to large earthquakes is still to be studied, it is certainly useful to have a reliable and simple method to identify and characterize them.

1. Introduction

Short localized seismic sequences, or clusters, are important seismic precursors (e.g. Rikitake, 1975; Cicerone et al., 2009; Lippiello et al., 2012; Hauksson et al., 2011; Sieh et al., 1993; Archuleta et al., 1982), and their study can also be useful for understanding phenomena prior to a main event, such as creep, slip, and stress concentrations (Ohnaka, 1992; Dodge et al., 1995, 1996; Ogata et al., 1995; Lippiello et al., 2012; Dominguez et al., 2016).

However, the term seismic cluster has been used in many contexts, sometimes it is used interchangeably with sequence, burst, and even swarm, and sometimes it refers to spatial only groupings (e.g. Frohlich and Davis, 1990; Chen et al., 2012; Georgoulas et al., 2013; Czecze and Bondár, 2019). Zaliapin and Ben-Zion (2013) state that the term has no formal definition. The published schemes for cluster identification are varied, but most involve complicated models and concepts that many times require multiple a priori assumptions (e.g. Gardner and Knopoff, 1974; Reasenberg, 1985; Frohlich and Davis, 1990; Ester et al., 1996; Zhuang et al., 2002; Marsan and Lengline, 2008; Zaliapin et al., 2008; Georgoulas et al., 2013; Hudyma, 2008; Jiménez et al., 2009; Konstantaras et al., 2012; Yang et al., 2019; Zaliapin and Ben-Zion, 2013, 2016).

For clustering studies to be useful and widely applicable, it is necessary to have a practical working definition of seismic cluster and a method to identify and quantify clustering in an objective way. We will propose here both a definition and a method, and we will illustrate their use by analyzing the seismicity of all northern Baja California, for some 30 years before the occurrence of the El Mayor-Cucapah earthquake.

* Corresponding author at: CICESE, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California, División de Ciencias de la Tierra, Departamento de Sismología, Carretera Ensenada–Tijuana No. 3918, Zona Playitas, C. P. 22860 Ensenada, Baja California, Mexico.
E-mail address: lenavila@cicese.mx (L. Ávila-Barrientos).

https://doi.org/10.1016/j.pepi.2024.107182
Received 1 June 2023; Received in revised form 12 March 2024; Accepted 17 March 2024
Available online 19 March 2024

2. Clusters

We will use a definition of seismic cluster as simple as possible, both to involve no a priori assumptions and to allow a simple and straightforward identification algorithm. We employ a minimum of parameters and propose a way to select them in a data-driven simple and subjective way.

Zaliapin and Ben-Zion (2016) define clustering as a partitioning of seismicity into groups closer in space and time than expected in a purely random distribution, a definition that agrees with the simplified one that will be used in this paper.

Since we will use the term background seismicity in our definition of cluster, and since this term is interpreted differently by different people, let us first state what we mean by it. For several people, background seismicity means declustered seismicity, where declustering means removing the aftershocks (e.g. Zhuang et al., 2002; Zaliapin and Ben-Zion, 2020). By background seismicity we refer to the small, and medium to medium-large seismicity that occurs in between large (significant) earthquakes. What is a large or significant earthquake depends of course on the region and the interests of the people doing the study; here, we will consider large earthquakes those with magnitudes $M \geq 6.5$. Thus, background seismicity does not include the large events or their suites of aftershocks, but includes smaller events and their aftershocks.

We define a seismic cluster as a spatio-temporal grouping of seismic activity that deviates from a spatially and temporally homogeneous Poisson process, a localized burst of background seismicity earthquakes, consisting of a set of at least ν earthquakes each one located within a radius ρ and an interval τ of at least one another member of the cluster; there are no a priori limits to the extent or duration of a cluster. Furthermore, we have in mind clusters that have a physical meaning, i.e., groupings that are indicative of activity in localized regions of high stressing rate, activity that is not large enough to significantly release the stress but rather helps in creating stress concentrations.

Thus, main earthquakes and their aftershocks are only one kind of cluster, and there can be many kinds of clusters, as recognized by Zaliapin et al. (2008); in what follows we will define four types based on the history of moment release.

The cluster identification algorithm, which we may call an associative one, is: for each successive earthquake from the data base, occurring at time t, previous clusters (if any) having no member occurring later than $t - \tau$, are considered ended and inactive; if any previous clusters are still active they are tested for having at least one member occurred later than $t - \tau$ and no farther than ρ from the current event's hypocenter and, if one is found, the new event is assigned to that cluster. If an event happens to belong to two, or more, existing clusters, all these clusters, which now have a member in common, are merged into the earliest one. If the new event does not belong to any existing cluster, it is considered to be the first one of a new cluster.

There is no limit to the number of events that a cluster can have but is there a minimum number of events, ν, that a set must have to be considered a cluster? Obviously, a set having only one event cannot be considered a cluster, but there is no theoretical lower limit to the number of events in a cluster. Hence, we postulate $\nu = 10$ as a reliable lower limit, and after each search sets having less than ν members are discarded. This choice will be discussed later in view of the results.

The number and size of clusters depend, of course, on the values assigned to the ρ and τ parameters. Extremely large values would result in all seismicity being assigned to a single cluster, while extremely small values would result in having as many clusters as earthquakes, each cluster consisting of a single event. Since the optimal set of parameters can be different for different regions, and times, it makes sense to let the data indicate which combination is best, and this search is exemplified below as applied to our data (next section).

In order to have an objective way of assigning values to the cluster parameters, we considered that clusters, to be significant, should have a small probability of occurring by chance, so that the best set of parameters is that which results in clusters having the smaller probabilities of random occurrence.

For a study region, having total area A, and study period, having total duration T, the threshold magnitude, which is the completeness magnitude M_c, results in a catalog with N events, so that the temporal and spatial (areal) occurrence densities are $\lambda_T = N/T$ and $\lambda_A = N/A$, respectively. The random occurrence probabilities of a cluster having n elements over an interval θ, and having n elements within an area α are given by Poisson distributions:

$$p_T(n,\tau) = \frac{(\lambda_T\theta)^n \, \mathrm{e}^{-\lambda_T \, \theta}}{n!}; p_A(n,\alpha) = \frac{(\lambda_A\alpha)^n \, \mathrm{e}^{-\lambda_A\alpha}}{n!}. \quad (1)$$

To determine the optimal set of parameters ρ, τ, it is necessary to know where to start the search, so, remembering that a cluster is an episode of earthquakes that are unequivocally close in time and space, there must not be more than a couple of days between successive earthquakes nor should they be more than a few kilometers apart; clusters can grow in duration and extension, so they can last for weeks and cover large areas, but always having events close in time and space to previous ones. Hence, the right combination of parameters will be looked for around $\tau \sim 1.5$ day and $\rho \sim 2.5$ km.

Next the parameter space is explored by, using the parameter combinations corresponding to points in a grid in the (ρ, τ) space, identifying clusters over the whole study region and, for clusters with at least ν events, measuring for each cluster its duration θ, and approximating its area α by the product of the cluster extents in the NS and EW directions, then estimating its random occurrence probabilities for space and time from (1), and assigning the average probability to that point in parameter space. Our preferred parameters will be those of the combination that results in the minimum average random occurrence probability.

Once clusters have been identified, each cluster is characterized by its number of events, area, and duration, and, since it is also known which events belong to each cluster, each cluster can be characterized by its time and place of occurrence, total seismic moment release, and moment release history.

3. Database

To illustrate the application of the proposed method, we chose the area between longitudes −113.5° to −118.0°, and latitudes 30.0° to 33.0°, which corresponds to northern Baja California, Mexico, and a small part of southern California, USA. The area was chosen because it is a very interesting region where the slip that accommodates the relative motion between the Pacific and North America plates is distributed among many different fault systems (e.g. Fletcher et al., 2014) (Fig. 1).

We use data from the catalog of the CICESE seismic network (*Red Sísmica del CICESE,* RSC), that spans from October 13, 1979, when the network started operating regularly, to the present, and reports magnitudes $M \geq 1.0$, but, as would be expected, coverage has changed with time and is not homogenous for small magnitudes. Fig. 2 shows the epicenters of all the 19,077 events with magnitudes $M \geq 1.0$, occurred from October 13, 1979 to April 1, 2010 in the region of study, it can be seen that all over northern Baja California seismicity is distributed mostly along SE-NW alignments, with some conjugate SW-NE alignments. This time period was chosen because no earthquakes larger than 6.4 occurred during it, so the seismicity could be considered background seismicity. Of course, the aftershocks of the largest events could have been removed ("de-clustered") but there was no reason to do so, since a main event and its aftershocks is, after all, a cluster; actually, the clusters identified by the proposed method included only magnitudes up to 5.4 (Table 1).

Fig. 3 shows the yearly number of earthquakes observed in the study area for two magnitude ranges. It shows that coverage is reasonably complete since about 1997 for magnitudes $M \geq 2.5$, and since about 2005 for $M \geq 1.5$ (data up to 2022 were included to establish reference

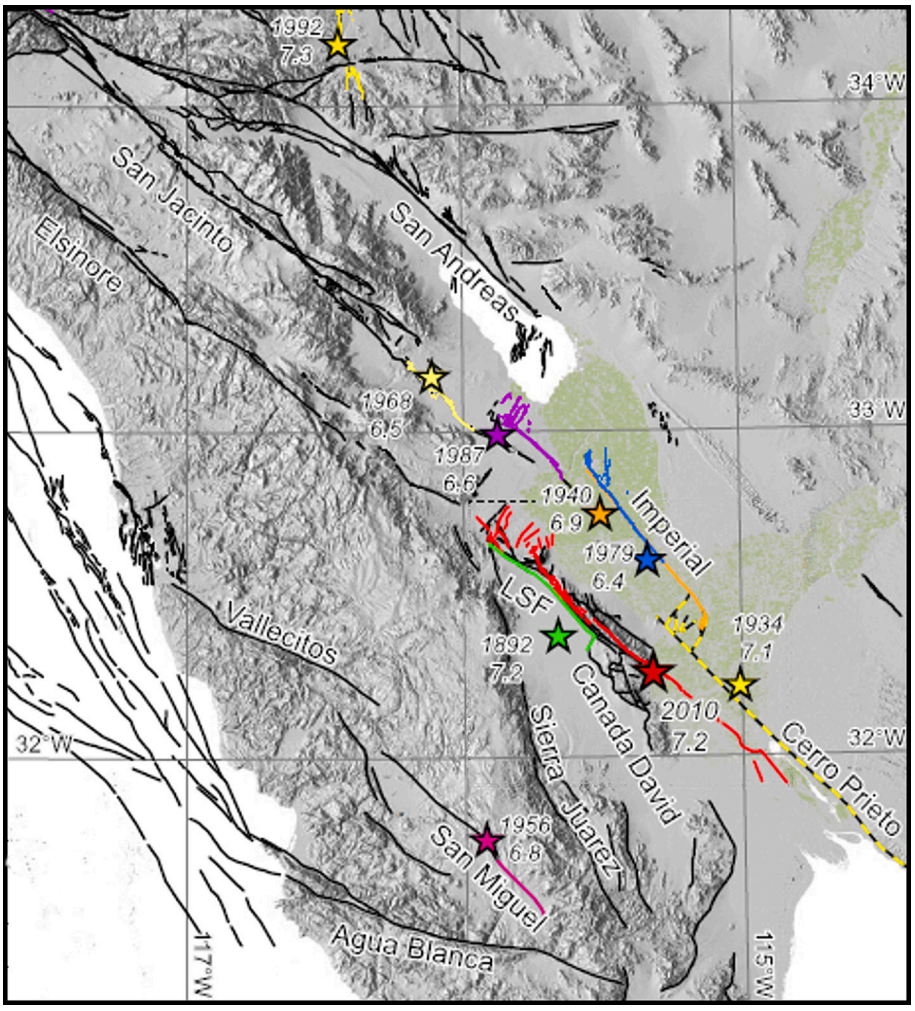


Fig. 1. Fault map of northern Baja California and southern California showing the main fault traces and the large historical earthquakes, including the 2010 El Mayor-Cucapah event (red star) (modified from Fletcher et al., 2014). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

levels). The Gutenberg-Richter histogram for 1979 to April 1, 2010 in Fig. 4 shows a reasonable linear fit for magnitudes $2.6 \leq M \leq 4.2$, and shows that magnitudes in the $M \geq 2.6$ range are not under-sampled. Actually, for our cluster identification method, it does not matter if the smaller magnitudes are under-sampled a little, as long as sampling does not change over time.

Hence, the most homogeneous database, consisting of earthquakes from January 1, 2000 to April 4, 2010 with magnitudes $M \geq 2.6$, was used for calibrating parameters, and then the cluster search was done over the whole database from 1979 to March 1, 2010.

For calibration, the total area $A = 154,202.290\ \text{km}^2$, the total time $T = 10.251904\ \text{yr}$, and $N = 2,922$ events, result in $\lambda_T = 0.780877$ events/day and $\lambda_A = 0.022040$ events/km^2. Fig. 5 shows the color-coded values of the average Poissonian probabilities after the raw values were smoothed by convolving with the unitary area, symmetric, matrix

$$\begin{pmatrix} 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix}$$

and the white diamond indicates the position of the minimum found for $\rho = 2.4\ \text{km}$ and $\tau = 1.2\ \text{day}$. The matrix was smoothed because using sharp boundaries for acceptance or rejection together with random factors in the observed seismicity, makes it possible for very small changes in the parameters to cause events to be excluded or included in clusters and for clusters to merge or not, which causes sharp changes variations in the probabilities. As will be discussed later, small changes in the parameters do not cause important changes in the results of a clustering study.

4. Results

Using the optimal parameters, $\rho = 2.4\ \text{km}$ and $\tau = 1.2\ \text{day}$, the search for clusters was carried out over the whole data set, and resulted in the identification of 21 clusters, listed with some of their characteristics in Table 1, and shown in Fig. 6. Each individual cluster will be referred to by its sequential number in time, preceded by "C".

Fig. 6 shows the epicentral location of the 21 identified clusters, with each cluster identified by a combination of symbol shape and color (Table 1). The area covered by this figure is shown by a red dashed rectangle in Fig. 2, and it is clear that clusters are concentrated in a region in the NE part of the study area. It should be emphasized that, although earthquakes with magnitudes as large as the magnitudes in the clusters occur outside the area shown in Fig. 6, no clusters at all were found outside this area.

The equivalent magnitude, M_{eq} in Table 1, is the magnitude corresponding to the sum of the seismic moments of all events in the cluster, where moments and magnitudes are related as (Hanks and Kanamori, 1979):

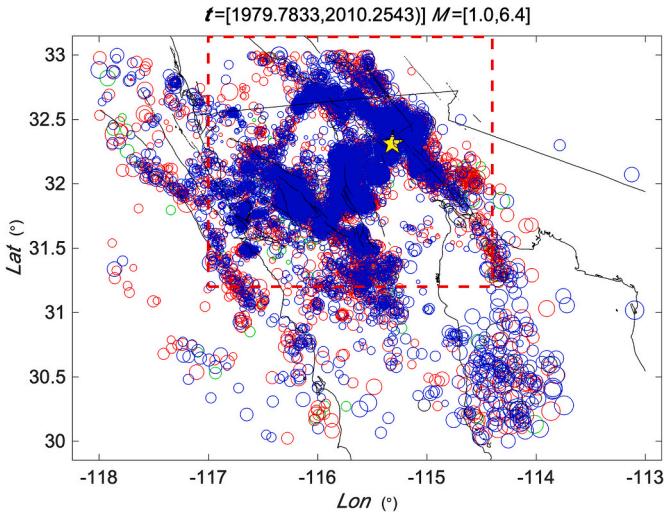


Fig. 2. Study area showing all seismicity $M \geq 1.0$ for the study area from October 13, 1979 to April 1, 2010 as 19,077 dots with sizes proportional to their magnitudes blue for depths $D \leq 10$ km, red for 10 km $< D \leq 20$ km, green for 20 km $< D \leq 30$ km, and black for $D > 30$ km.
Lines represent the coastline and the Mexico-US border, small lines are fault traces, and the red dashed rectangle is the region where clusters occurred, shown in Fig. 6, which contains 17,969 events. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$log_{10}M_0 = 1.5M + 16.05. \quad (2)$$

Fig. 7 (top) shows the logarithm of the total seismic moment released by each cluster plotted versus time, each symbol indicates the occurrence time of an earthquake, and the type of line indicates the type of moment release history described above. Clusters appear almost instantaneous at this scale, but they have durations ranging from 0.27 to 13.08 days, with a mean duration of 2.22 days; the histories for some individual clusters are shown in Figs. 8 and 9. Clusters C3, C7, C8, and C19 appear to consist each of a single event because the first event in each was much larger than all following events, but actually they are constituted by 10, 11, 13, and 20 events, respectively.

Fig. 7 (bottom) shows the cumulative seismic moment release for all clusters, with rhombs indicating the occurrence time of each cluster. The moment released in clusters from ~2002 until the time of the penultimate cluster at the end of 2009 is larger than that released over the previous 22 years, and the moment from the last cluster in early 2010 was much larger than all previously released moments.

4.1. Cluster types

In Table 1, Type indicates how seismic moment was released during the cluster duration. Type 0 corresponds to swarm-like behavior having no main event(s), where a main event is defined to be an event with magnitude larger than the average magnitude plus two standard deviations (C6 in Fig. 8). Type 1 corresponds to main event(s) with aftershocks, where the moment release after the main event is ≥ 1.5 times larger than before it (C1 in Fig. 8), as opposed to Type 2 where moment release before the main event is ≥ 1.5 times larger than after it (C17 in Fig. 8) and corresponds to foreshocks leading to a main event. Type 3 corresponds to moment release approximately equal before and after the main event (C9 in Fig. 8). Most of the identified clusters are Type 1.

The moment release histories of the last series of large clusters are shown in Fig. 9. Most large clusters in this series are Type 1, except for C16, which is Type 0.

It should be noted that, although the method was applied in the same uniform way to a large geographic area, most clusters appeared within a

Table 1
Earthquake cluster characteristics.

Cluster	n	Time Range (yr) θ (days)	D Range (km) α (km^2)	M Range M_{eq}	Type	Color Symbol
1	14	1979.79930, 1979.80316 1.41	−10.00, −5.00 3.214	2.9, 4.5 4.7	1	B O
2	16	1988.06727, 1988.06982 0.93	−18.00, −18.00 18.211	2.6, 4.8 4.8	3	R D
3	10	1991.92259, 1991.92364 0.38	−18.00, −0.00 8.213	2.6, 5.0 5.0	1	Gy S
4	25	1993.67972, 1993.68609 2.33	−11.00, −3.00 20.165	2.6, 3.9 4.3	1	Gn A
5	10	1993.77592, 1993.78098 1.85	−20.00, −9.00 26.214	2.6, 4.2 4.3	1	K X
6	16	1993.93457, 1993.93748 1.06	−5.00, −0.00 21.961	2.6, 3.9 4.2	0	M H
7	11	1994.22226, 1994.22835 2.22	−23.00, −16.00 5.713	2.6, 4.8 4.8	1	C Tl
8	13	1998.13316, 1998.14041 2.65	−7.00, −2.00 7.562	2.6, 4.4 4.4	1	Y Tr
9	29	1998.79527, 1998.80675 4.19	−16.00, −3.00 28.399	2.6, 4.2 4.4	3	B Td
10	19	1999.79405, 1999.80121 2.61	−19.00, −3.00 50.191	2.6, 4.5 4.6	3	R O
11	21	2000.33169, 2000.33903 2.68	−16.00, −5.00 65.535	2.6, 4.6 4.6	1	Gy D
12	17	2001.93700, 2001.93845 0.53	−19.00, −8.00 53.625	2.6, 3.7 3.9	1	Gn S
13	27	2002.01044, 2002.01818 2.82	−22.00, −11.00 15.943	2.6, 3.8 4.2	1	K A
14	67	2002.14471, 2002.16566 7.65	−17.00, −4.00 81.499	2.6, 4.0 4.5	1	M X
15	11	2003.16193, 2003.16267 0.27	−6.000, −4.00 21.929	2.6, 3.1 3.6	0	C H
16	38	2005.35261, 2005.37550 8.35	−10.00, −4.00 22.706	2.6, 3.4 4.1	0	Y Tl
17	15	2006.07079, 2006.07413 1.22	−9.00, −4.00 35.068	2.6, 4.1 4.3	2	B Tr
18	146	2008.10767, 2008.14351 13.08	−13.00, −2.00 198.496	2.6, 5.2 5.5	1	R Td
19	20	2009.71769, 2009.72093 1.18	−10.00, −2.00 26.802	2.6, 5.3 5.3	1	Gy O
20	25	2009.83357, 2009.84194 3.06	−8.00, −3.00 9.074	2.6, 4.3 4.4	1	Gn D
21	39	2009.99656, 2010.00161 1.84	−16.20, −5.30 98.303	2.6, 6.0 6.0	1	K S

Table lists the number of events in each cluster (n), time limits (T range) and duration in days (θ), depth (D) range and cluster area (α), magnitude (M) range and equivalent magnitude (M_{eq}), and type of cluster, Color: B = blue, R = red, Gy = gray, Gn = green, K = black, M = Magenta, C = cyan; Y = yellow; Symbol: O = circle, D = diamond, S = square, A = asterisk, X = x, H = hexagon, Tl = Triangle pointing left, Tr = Triangle pointing right, Td = Triangle pointing down.

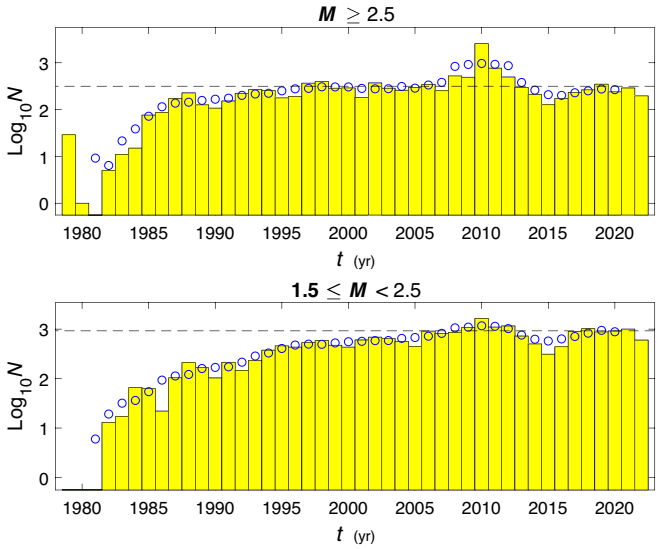


Fig. 3. Yearly number of events in two magnitude ranges for the whole region from 1979 to 2022.
The horizontal dashed lines are for reference about the approximate level of activity during 1997 to date (top) and during 2006 to date (bottom).

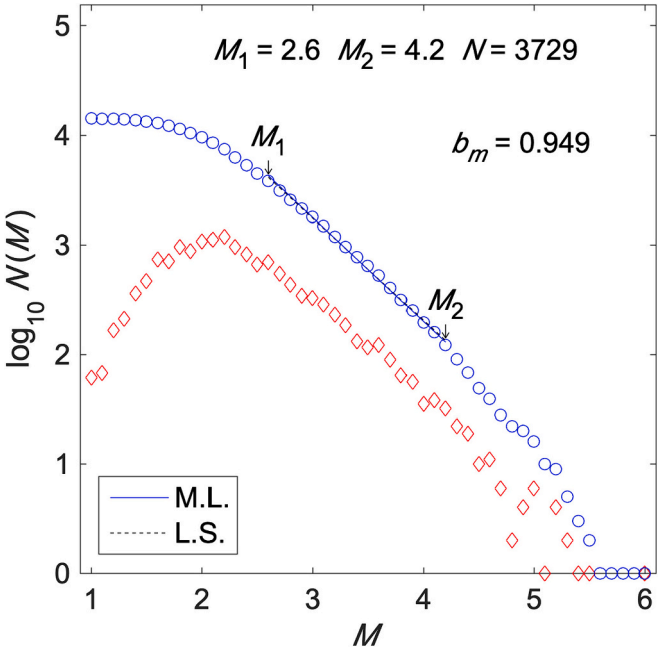


Fig. 4. Gutenberg-Richter histogram (blue) and non-cumulative histogram (red) for the RSC catalog from 1979 to April 1, 2010.
M_1 and M_2 indicate the approximate limits of the linear range. The straight lines represent fits to the data: maximum likelihood (continuous) and least squares (dashed). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

small region, and it turns out that 16 of the 21 clusters are located on what is the epicentral region of the El Mayor-Cucapah earthquake or near the extremes of the bilateral rupture associated with it.

5. The El Mayor-Cucapah earthquake and the clusters

The El Mayor-Cucapah (EMC) April 4, 2010, M_w 7.1, earthquake is the largest recorded earthquake in northern Baja California, Mexico; its mapped surface faulting and its aftershocks are located along the El

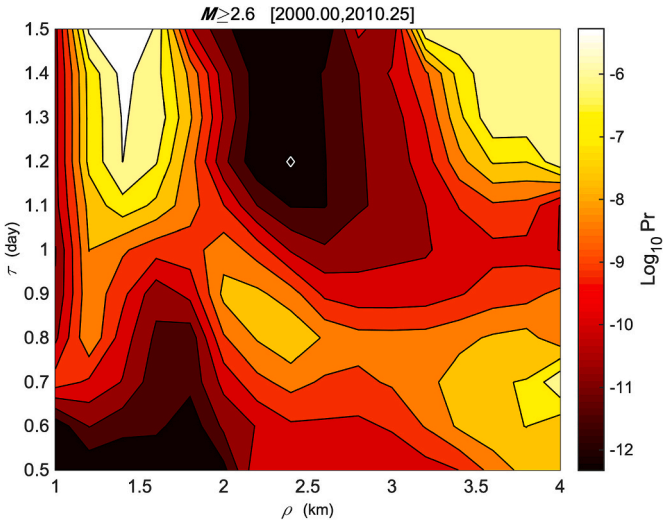


Fig. 5. Color-coded logarithm of the average Poissonian probabilities for clusters determined from each combination of the ρ and τ parameters.
The white diamond indicates the minimum probability.

Mayor and Cucapah ranges (hence its denomination) which separate the Laguna Salada and Imperial-Mexicali basins (Castro et al., 2011; Fletcher et al., 2014; Gonzalez-Ortega et al., 2014) (Fig. 1). The EMC earthquake produced a 120 km long rupture extending from the Mexico-U.S. international border to the Gulf of California (Wei et al., 2011). The rupture consisted of complex strike-slip faulting with normal components (Hauksson et al., 2011), involving several fault segments through the Cucapah and El Mayor Sierras (Sarychikhina et al., 2009; Castro et al., 2011; Fletcher et al., 2014; Gonzalez-Ortega et al., 2014). The event was felt in northwestern Mexico and southern California, Arizona, and Nevada, at distances >400 km from the epicenter (Munguía et al., 2010).

The location of the EMC earthquake was somewhat surprising, because most large earthquakes, such as the Cerro Prieto 1934 $M \sim 7.1$ and Imperial Valley 1940 $M = 6.9$ earthquakes (Fletcher et al., 2014; Munguía et al., 2010), occur farther northeast along the Imperial and Cerro Prieto faults, although there have been large earthquakes not far from the rupture of the EMC earthquake, in the eastern part of the Laguna Salada basin, including the 1892 $M = 7.2$ Laguna Salada earthquake (Hough and Elliot, 2004) (Fig. 1).

After the EMC earthquake, Hauksson et al. (2011) and Chen and Shearer (2013) identified two sequences prior to it, on March 21 and 22, 2010, and April 3 and 4, 2010; the last one just 24 h before the main event consisted of shocks with magnitudes ranging from 2 to 4.4, and found foreshock swarms with magnitudes ranging from 1.4 to 4.4, with no clear mainshock. Yao et al. (2020) analyzed an earthquake sequence starting 21 days before the EMC earthquake and identified two episodes with depths around 14 and 16 km where the mainshock started, stress drops from 3.8 to 41.7 MPa. Their results show a migration towards the EMC hypocenter within the last 8 h, and an activity burst 6 min before the main event.

The identified clusters are related to the EMC source as follows. Most clusters are located close to EMC source, except C2, C3, C7, C8, and C9, the first three located along a SE-NW seismicity alignment which points to the EMC source region (Fig. 6). Clusters C20, and C11 and C12, coincide with the extremes of the EMC rupture that propagated from the epicenter first to the northwest and later to the southeast (Hauksson et al., 2011; Castro et al., 2011). The latest clusters (C18, C19 and C21) are all very close to the EMC epicenter; the depths of the clusters that are closest in space and time to the EMC epicenter, range from 2.0 to 16.2 km (Table 1).

Clusters C2, C3, C8, and C9 belong to the conjugate seismicity

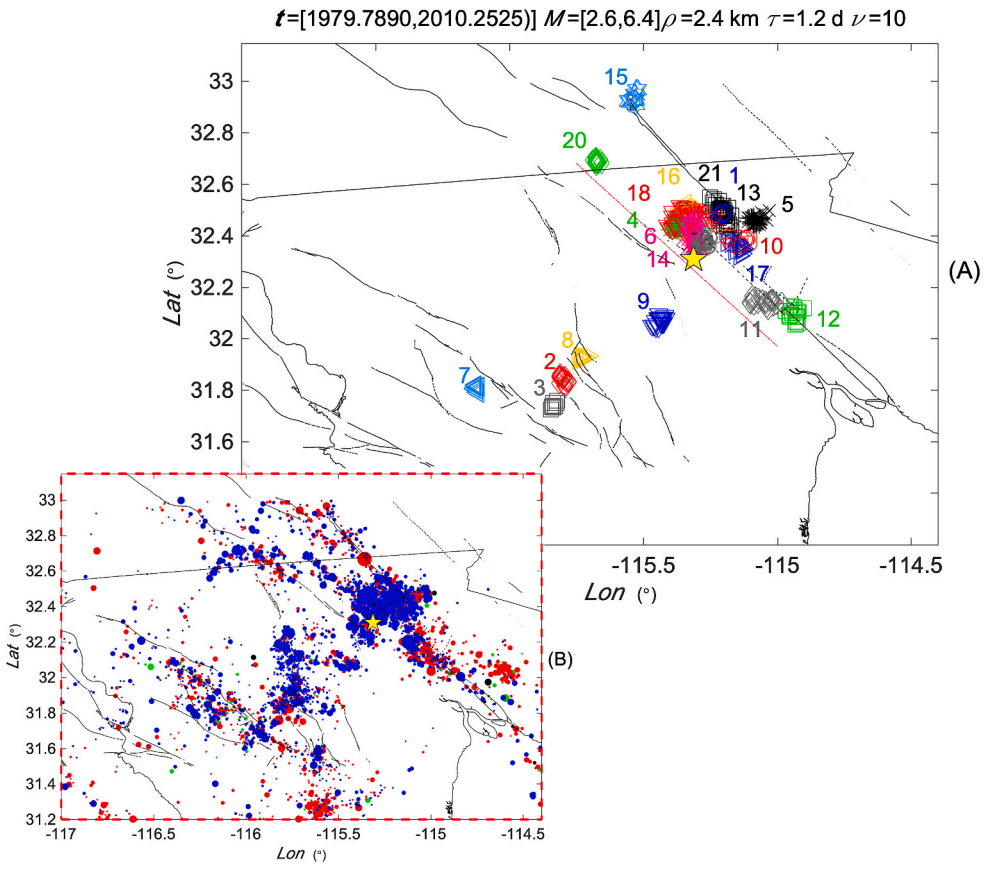


Fig. 6. (A) Epicenters of the earthquake clusters previous to the EMC earthquake (yellow star); the numbers indicate the cluster order of occurrence, and the color/symbol combination is listed in Table 1.
Dash-dot line indicates the approximate position of the rupture. Thin short lines are fault traces. (B) Total seismicity, 5560 events (same time and magnitude range), for reference. Events are color-coded for depth and sized according to magnitude as in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

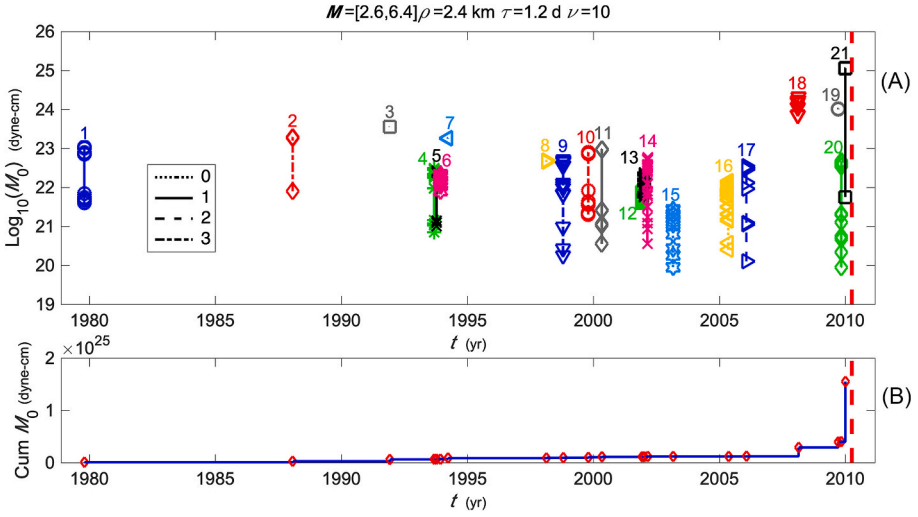


Fig. 7. (A) Total seismic moment release of each earthquake cluster vs. time.
The combinations of symbol and color (Table 1) correspond to those in Fig. 6, while the line type indicates the type of moment release history. (B) Cumulative seismic moment for all clusters vs. time (blue line), the rhombs indicate the occurrence time of each cluster. For both top and bottom, the red dashed vertical line indicates the occurrence time of the EMC earthquake. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

alignment, and their occurrence could be interpreted as a migration of the stresses causing clustering towards the EMC hypocenter. After C9, occurred in October 1988, no more clusters occurred along this alignment.

Two aspects of the cluster activity are clearly shown in Fig. 7 (top): 1) clusters are more frequent as the time of the EMC event approaches, and 2) all the latest clusters before EMC, C18 (February 2008) to C21, have equivalent magnitudes greater than the mean 4.4, except for C20 that is

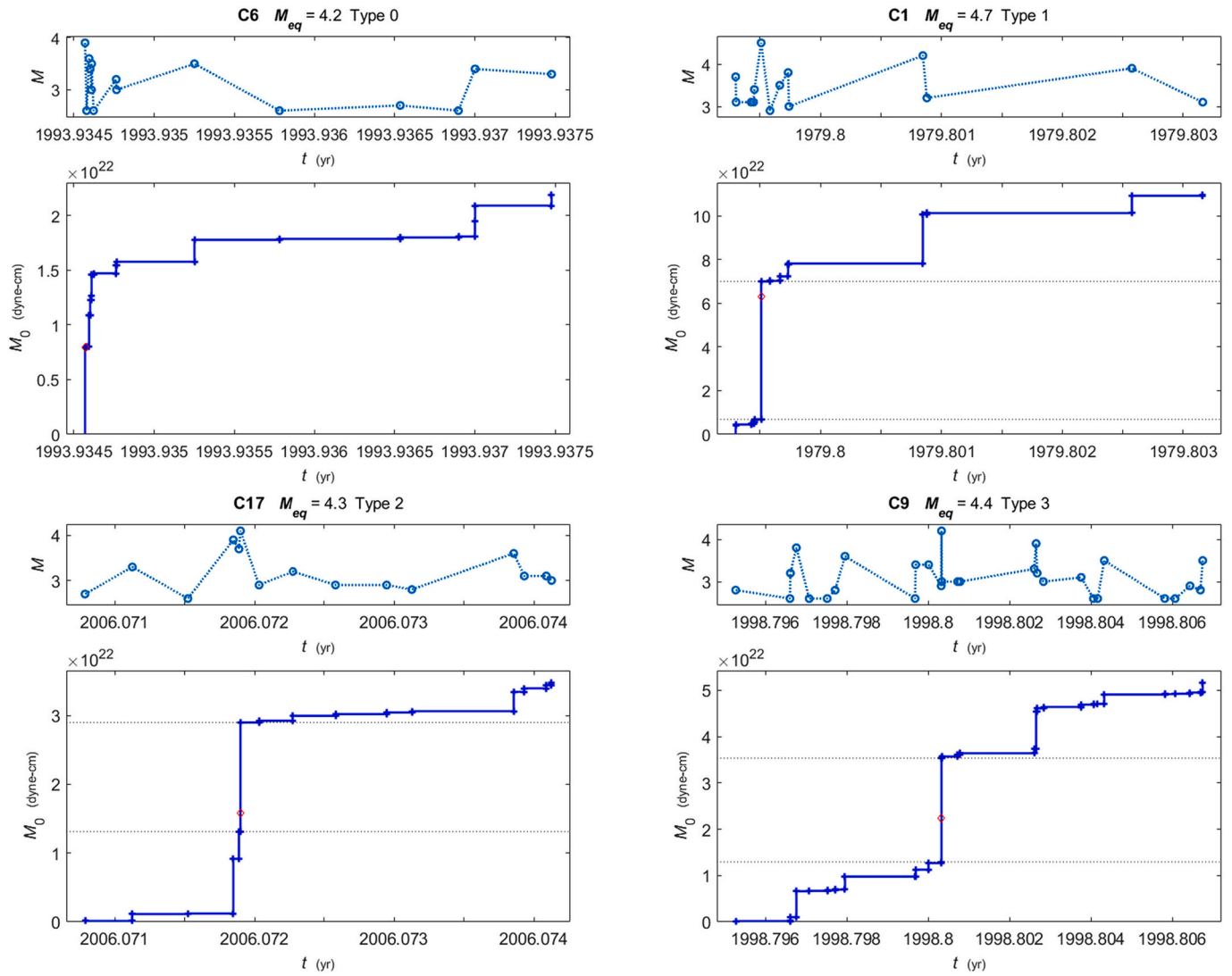


Fig. 8. Examples of cluster type, showing magnitude occurrences (top) and cumulative moment release vs. time.
C6 is Type 0; C1 is Type 1; C17 Type 2; C9 Type 3. The horizontal dotted lines, if any, indicate the contribution of the main event in the cluster to the total moment release.

located far from the EMC epicenter (Fig. 6). Cluster C21 includes the largest event recorded in the study area before the EMC earthquake, this event had $M = 6.0$ and was located about 30 km NNE of the EMC epicenter (Hauksson et al., 2011).

Even though the three last clusters before the EMC earthquake are Type 1, there is no obvious relation between cluster type and closeness in time or space to the main EMC earthquake. This may be interpreted as meaning that the overall pattern of stresses leading to the main earthquake is not a factor that determines the type of the clusters. Hence, there is not a particular cluster type to look for when trying to forecast an earthquake.

6. Discussion

In what follows, we will first discuss how results vary if we modify our threshold values or the values of our parameters, to see whether our results are stable or depend critically on a combination of parameter or threshold values.

How critical is the ν value? It is easy to see from Table 1 the effect of using a larger ν: $\nu = 11$ would mean losing events C3 and C5, the first one from the conjugate alignment and the second from around the EMC source area, and $\nu = 12$ would discard two more clusters, 7 and 15. Lowering ν results, of course, in more clusters: $\nu = 9$ results in only one cluster more, with M_{eq} 4.2 and located among those around the EMC source area, while $\nu = 8$ results in 9 clusters more, all except one located in the same EMC area. Thus, we see that, unless it is set absurdly low or high, the actual value of ν is not critical, since changing it does not change the overall distribution that points to the EMC area as being a region to watch.

Is the threshold magnitude a critical factor? As described above, we used as threshold magnitude $M_c = 2.6$, and a lower value is not recommended because it would mean inhomogeneous coverage in time, but what is the effect of a higher M_c? Setting $M_c = 2.7$ reduces the number of earthquakes from 5560 to 4484, a significant reduction which causes clusters C3 (10 events), C5 (10 events), and C15 (11 events) to be missed, while C18 (146 events) gets partitioned into two clusters, one with 109 events and another with 14 events, both occurring, naturally, within the area covered by C18. All remaining clusters, except C1, lost events and many covered smaller areas so that their random occurrence probabilities were higher than for $M_c = 2.6$. A look at Fig. 6 shows that only the loss of C15 changes minimally the cluster spatial distribution.

A further increase to $M_c = 2.8$, reduces the number of events to 3677 which causes 7 clusters (C3, C5, C7, C8, C11, C12, and C15) to be missed and one (C18) to be partitioned; only two clusters remain in the

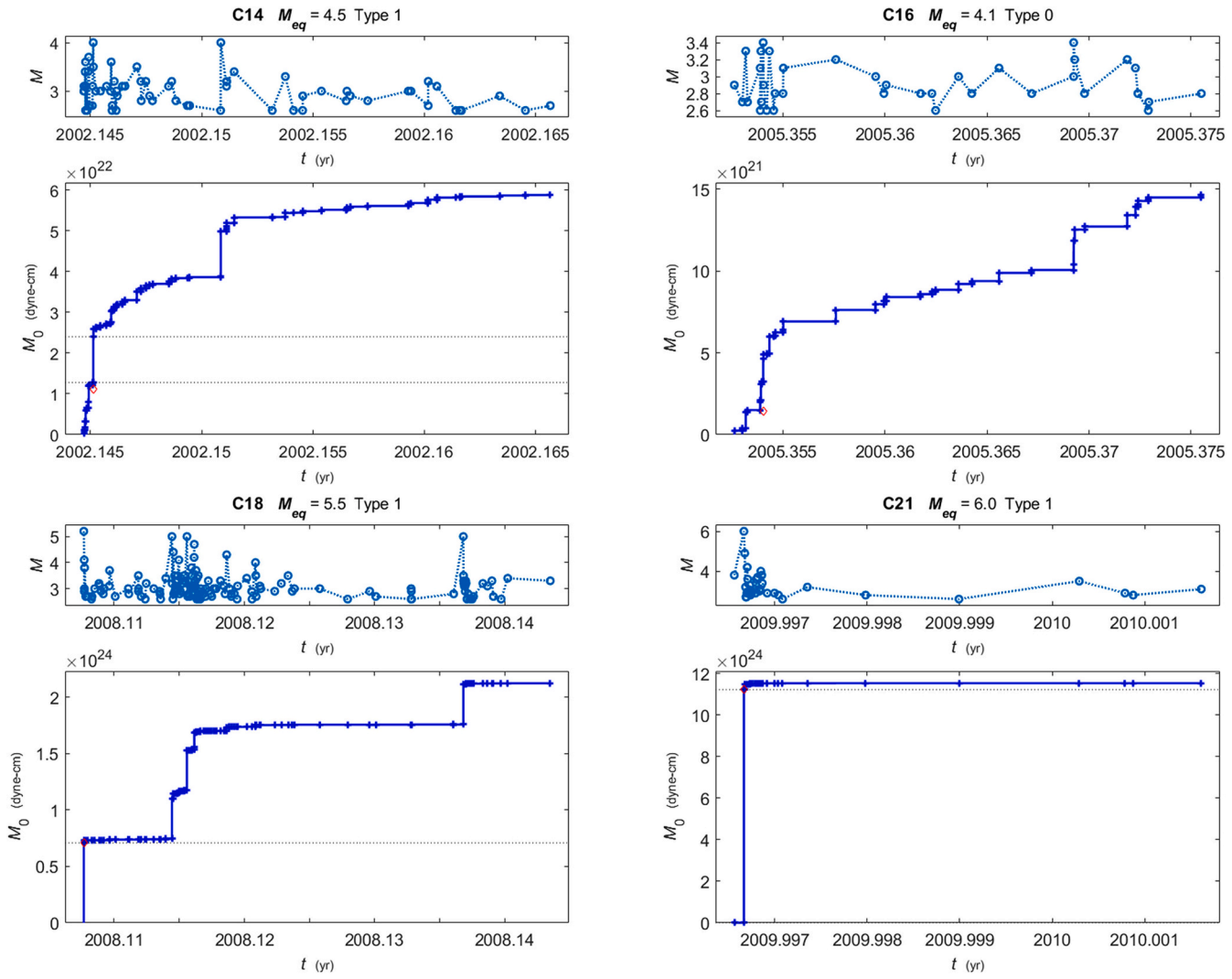


Fig. 9. Cumulative seismic moment release history for four of the large clusters occurred before the EMC earthquake. Cluster 18 is the largest cluster and the last to occur before the EMC earthquake.

orthogonal alignment, and the future SE end of the EMC shows no clusters. Yet the EMC source region is well identified by clusters.

We chose our parameters so that clusters would be improbable if seismicity was uniform, so, to see whether they worked correctly we generated synthetic catalogs having the same geographical extension, same time interval, and same number of events, with hypocenters and occurrence times occurring randomly with uniform probabilities. There were absolutely no clusters for our chosen parameter values in ten different synthetic catalogs, and these catalogs did not present clusters until both ρ and τ attained large values of about 25 or 30 km and days, respectively, which certainly do not correspond with the idea of concentrated activity.

Finally, what happens if we vary slightly our preferred parameter values? Fig. 5 shows that the random occurrence probabilities do not increase rapidly in the neighborhood of the minimum; hence, we can expect that small departures from the minimum will not change results drastically. Indeed, increasing ρ to $\rho = 3.0$ km results in three more clusters and an increase in random occurrence probability of 6.25×10^{-8}, but the spatial pattern does not change. Decreasing ρ to $\rho = 1.6$ km results in three less clusters and an increase in random occurrence probability of 1.32×10^{-8}, but, again, the spatial pattern does not change.

Increasing τ to $\tau = 1.4$ day results in the same number of clusters but five of them have more events, the largest cluster C18 has eight more events, and a small decrease in random occurrence probability of 4.35×10^{-9} (remember we chose our minimum from a smoothed matrix), but the spatial pattern does not change. Decreasing τ to $\tau = 1.0$ day results in one more cluster because the large C18 splits in two, and an increase in random occurrence probability of 1.49×10^{-9}, but, again, the spatial pattern does not change.

Thus, given that the general clustering and its spatial patterns are robust, we will discuss some of the characteristics of the cluster activity.

We identified 21 clusters in the source region of the EMC earthquake (Fig. 6), beginning in October 1979 and becoming more frequent as the time of the EMC earthquake approached. The first cluster, C1, occurred in the middle of the zone close to the EMC epicenter (EMCZ zone), which was active all the time up to the time of the large earthquake. In January 1988 and December 1991 two clusters, C2 and C3, occurred along the conjugate SW-NE alignment that joins the main SE-NW alignment at the site of the EMC epicenter; afterwards, in February and October 1998 two more clusters, C8 and C9, occurred along the conjugate alignment, each one closer to the EMC epicenter. Cluster C7, occurred in 1994, was the one located farthest away to the SW, probably related to activity on the San Miguel fault system. After that, all clusters except one, C15, occurred in the EMCZ or along an alignment, parallel to the EMC rupture (Fig. 6), with one cluster corresponding to each of the extremes of the rupture, C20 to the NW and C12 to the SE. This parallel alignment could speculatively be interpreted as activity occurring along the Cerro Prieto

and Imperial faults that indicated the high levels of stressing rate in the EMC region and maybe helped to cause stress concentrations that contributed to its triggering. Cluster C15 occurred at the NW tip of the Imperial fault and also represents activity on faults parallel to the EMC rupture. Although the occurrence of clustering along the conjugate alignment is suggestive, it is an open question whether these clusters were directly related to the EMC earthquake.

Of the last three clusters, occurring from February 2008 to December 2009, all those in the epicentral area had equivalent magnitudes greater than the mean for all clusters. Although there was seismic activity all over the study region of northern Baja California (Figs. 2 and 6(B)), all clusters occurred in or close to the EMC source region, and no clusters were found far from it.

The method proposed here for cluster identification is extremely simple and involves fewer assumptions and parameters than most methods currently used, including those specializing in aftershock identification for declustering (e.g. Gardner and Knopoff, 1974; Reasenberg, 1985; Frohlich and Davis, 1990; Molchan and Dmitrieva, 1992; Baiesi and Paczuski, 2004; Zaliapin et al., 2008; Luen and Stark, 2012; Zaliapin and Ben-Zion, 2020; and many others). While many of these methods use a minimum distance criterion, the use of this criterion together with the minimum time criterion, applied sequentially makes our method extremely easy to implement and use, because only relations with pertinent data need be considered, instead of trying to find relations among thousands of data at once. We propose a method to adjust the two parameters, but we have shown that the choice is not critical, not so minor changes in the parameters change details in the identified cluster population, but do not cause important changes that would change the overall information. Finally, our method identifies all types of clusters, so if a study is interested in a particular type, other types can be ignored.

7. Conclusions

We propose a simple and compact definition and model for seismic clusters, with a minimum of parameters, which leads to a simple and straightforward associative algorithm for cluster identification. We also propose a data-driven method for choosing the model parameters as those that minimize the probability of the identified clusters being due to random groupings in time and space.

The cluster model and identification scheme we propose do identify important clustering activity occurring within ten years previous to the EMC earthquake. We used data, from the catalog of the RSC seismic network, from all northern Baja California, but found clustering only within a relatively small area around the site where the rupturing associated to the EMC earthquake was to occur. Our results agree with clustering found by Nava et al. (2023), using a completely different method based on occurrence apparent velocities, before the EMC earthquake and other earthquakes in southern California.

The short-term precursory activity mentioned above might have been identified before the occurrence of the EMC earthquake if a study of clusters in northern Baja California, like the one shown here, had identified the future epicentral region as an area of interest to be looked at in detail. Such detailed observation or monitoring might have observed the foreshock activity reported by Hauksson et al. (2011), Chen and Shearer (2013), and Yao et al. (2020), which might have led to a timely forecast of the EMC earthquake.

We propose that cluster activity studies, such as the one we made here, would be useful for identifying possible zones where a large earthquake could occur, so that they could be adequately monitored, and we hope that the method presented here may be useful for this purpose.

CRediT author contributions statement

All authors contributed to the study conception, data analysis, and writing. All authors read and approved the final manuscript.

Funding

CICESE internal funds.

CRediT authorship contribution statement

F. Alejandro Nava: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lenin Ávila-Barrientos:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

This work was partially funded by project 2602 of the CONAHCyT *Investigadoras e Investigadores por México* program (formerly *Cátedras CONACyT* program). Thanks to Antonio Mendoza Camberos for technical support, and to the staff of CICESE's seismic network (RSC). Sincere thanks to Krittanon Sirorattanakul and to an anonymous reviewer for constructive criticism and suggestions that led to an improved paper.

References

Archuleta, R.J., Cranswick, E., Mueller, C., Spudich, P., 1982. Source parameters of the 1980 Mammoth Lakes, California, earthquake sequence. J. Geophys. Res. 87 (B6), 4595–4607.

Baiesi, M., Paczuski, M., 2004. Scale-free networks of earthquakes and aftershocks. Phys. Rev. E 69 (6), 066106.

Castro, R.R., Acosta, J.G., Wong, V.M., Pérez-Vertti, A., Mendoza, A., Inzunza, L., 2011. Bull. Seismol. Soc. Am. 101 (6), 3072–3080. https://doi.org/10.1785/0120110112.

Chen, X., Shearer, P.M., 2013. California foreshock sequences suggest aseismic triggering process. Geophys. Res. Lett. 40, 2602–2607. https://doi.org/10.1002/grl.50444.

Chen, X., Shearer, P.M., Abercrombie, R.E., 2012. Spatial migration of earthquakes within seismic clusters in Southern California: evidence for fluid diffusion. J. Geophys. Res. Solid Earth 117 (B4).

Cicerone, R.D., Ebel, J.E., Britton, J., 2009. A systematic compilation of earthquake precursors. Tectonophysics 476, 371–396. https://doi.org/10.1016/j.tecto.2009.06.008.

Czecze, B., Bondár, I., 2019. Hierarchical cluster analysis and multiple event relocation of seismic event clusters in Hungary between 2000 and 2016. J. Seismol. 23 (6), 1313–1326.

Dodge, D.A., Beroza, G.C., Ellsworth, W.L., 1995. Foreshocks sequence of the 1992 landers, California, earthquake and its implications for earthquake nucleation. J. Geophys. Res. 100 (B7), 9865–9880.

Dodge, D.A., Beroza, G.C., Ellsworth, W.L., 1996. Detailed observations of California foreshock sequences implications for the earthquake initiation process. J. Geophys. Res. 101 (B10), 22,371–22,392.

Dominguez, L.A., Taira, T., Santoyo, M., 2016. Spatiotemporal variations of characteristic repeating earthquake sequences along the middle America trench in Mexico. J. Geophys. Res. Solid Earth 121, 8855–8870. https://doi.org/10.1002/2016JB013242.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI KDD-96 96 (34), 226–231.

Fletcher, J.M., Teran, O.J., Rockwell, T.K., Oskin, M.E., Hunut, K.W., Mueller, K.J., Spelz, R.M., Akciz, S.O., Masana, E., Faneros, G., Fielding, E.J., Leprince, S., Morelan, A.E., Stock, J., Lynch, D.K., Elliott, A.J., Gold, P., Liu-Zeng, J., González-Ortega, A., Hinojosa-Corona, A., González-García, J., 2014. Assembly of a large earthquake from a complex fault system: Surface rupture kinematics of the April 04 2010 El Mayor–Cucapah (Mexico) Mw 7.2 earthquake. Geosphere 10 (3), 1–31. https://doi.org/10.1130/GES00933.1.

Frohlich, C., Davis, S.D., 1990. Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. Geophys. J. Int. 100, 19–32.

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.

Georgoulas, G., Konstantaras, A., Katsifarakis, E., Stylios, C.D., Maravelakis, E., Vachtsevanos, G.J., 2013. "Seismic-mass" density-based algorithm for spatio-temporal clustering. Expert Syst. Appl. 40 (10), 4183–4189.

Gonzalez-Ortega, A., Fialko, Y., Sandwell, D., Nava-Pichardo, F.A., Fletcher, J., Gonzalez-Garcia, J., Lipovsky, B., Floyd, M., Funning, G., 2014. El mayor-Cucapah (mw 7.2) earthquake: early near-field postseismic deformation from InSAR and GPS observations. J. Geophys. Res. Solid Earth 119, 1482–1497. https://doi.org/10.1002/2013JB010193.

Hanks, T.C., Kanamori, H., 1979. A moment magnitude scale. J. Geophys. Res. 84, 2348–2350.

Hauksson, E., Stock, J., Hutton, K., Yang, W., Vidal-Villegas, J.A., Kanamori, H., 2011. The 2010 mw 7.2 El mayor-Cucapah earthquake sequence, Baja California, Mexico and southernmost California, USA: active seismotectonics along the Mexican Pacific margin. Pure Appl. Geophys. 168, 1255–1277. https://doi.org/10.1007/s00024-010-0209-7.

Hough, S.E., Elliot, A., 2004. Revisiting the 23 February 1892 Laguna Salada earthquake. Bull. Seismol. Soc. Am. 94, 1571–1578.

Hudyma, M.R., 2008. Analysis and Interpretation of Clusters of Seismic Events in Mines. University of Western Australia, Perth, Australia.

Jiménez, A., Tiampo, K.F., Posadas, A.M., Luzón, F., Donner, R., 2009. Analysis of complex networks associated to seismic clusters near the Itoiz reservoir dam. Eur. Phys. J. Spec. Top. 174 (1), 181–195.

Konstantaras, A.J., Katsifarakis, E., Maravelakis, E., Skounakis, E., Kokkinos, E., Karapidakis, E., 2012. Intelligent spatial-clustering of seismicity in the vicinity of the Hellenic seismic arc. Earth Sci. Res. 1 (2), 1.

Lippiello, E., Marzocchi, W., De Arcangelis, L., Godano, C., 2012. Spatial organization of foreshocks as a tool to forecast large earthquakes. Sci. Rep. 2, 846. https://doi.org/10.1038/srep00846.

Luen, B., Stark, P., 2012. Poisson tests of declustered catalogues. Geophys. J. Int. 189 (1), 691–700.

Marsan, D., Lengline, O., 2008. Extending earthquakes' reach through cascading. Science 319 (5866), 1076–1079.

Molchan, G., Dmitrieva, O., 1992. Aftershock identification: methods and new approaches. Geophys. J. Int. 109 (3), 501–516.

Munguía, L., Navarro, M., Valdez, T., Luna, M., 2010. Datos de Movimientos Fuertes Registrados en Baja California Durante el Sismo El Mayor-Cucapah del 4 de Abril de 2010 (mw 7.2): Resultados Preliminaries. CICESE, p. 49. Special report,.

Nava, F., Reynoso, H., Glowacka, E., 2023. Occurrence apparent velocities for identification and quantification of space-time clustering precursory to a large earthquake. Application to large (M>7.0) earthquakes in southern California and northern Baja California. Math. Geosci. 55 (4), 579–605. https://doi.org/10.1007/s11004-023-10047-z.

Ogata, Y., Utsu, T., Katsura, K., 1995. Statistical features of foreshocks in comparison with other earthquake clusters. Geophys. J. Int. 121, 233–254.

Ohnaka, M., 1992. Earthquake source nucleation: a physical model for short-term precursors. Tectonophysics 211, 149–178.

Reasenberg, P., 1985. Second-order moment of Central California seismicity, 1969–1982. J. Geophys. Res. Solid Earth 90 (B7), 5479–5495.

Rikitake, T., 1975. Earthquake precursors. Bull. Seismol. Soc. Am. 65 (5), 1133–1162.

Sarychikhina, O., Glowacka, E., Mellors, R., Vázquez, R., Munguía, L., Guzmán, M., 2009. Surface displacement and groundwater level changes associated with the may 24, 2006 mw 5.4 Morelia fault earthquake, Mexicali Valley, Baja California, Mexico. Bull. Seismol. Soc. Am. 99 (4), 2180–2189. https://doi.org/10.1785/0120080228.

Sieh, K., Jones, L., Hauksson, E., Hudnut, K., Eberhart-Phillips, D., Heaton, T., Hough, S., Hutton, K., Kanamori, H., Lilje, A., Lindvall, S., McGill, S.F., Mori, J., Rubin, C., Spotila, J.A., Stock, J., Thio, H.K., Treiman, J., Wernicke, B., Zachariasen, J., 1993. Near-field investigations of the landers earthquake sequence, April to July 1992. Science 260, 171–176.

Wei, S., Fielding, E., Leprince, S., Sladen, A., Avouac, J.-P., Helmberger, D., Hauksson, E., Chu, R., Simons, M., Hudnut, K., Herring, T., Briggs, R., 2011. Superficial simplicity of the 2010 El mayor-Cucapah earthquake of Baja California in Mexico. Nat. Geosci. https://doi.org/10.1038/NGEO1213. Supplementary information.

Yang, J., Cheng, C., Song, C., Shen, S., Zhang, T., Ning, L., 2019. Spatial-temporal distribution characteristics of global seismic clusters and associated spatial factors. Chin. Geogr. Sci. 29 (4), 614–625.

Yao, D., Huang, Y., Peng, Z., Castro, R.R., 2020. Detailed investigation of the foreshock sequence of the 2010 Mw7.2 El mayor-Cucapah earthquake. J. Geophys. Res. Solid Earth 124. https://doi.org/10.1029/e2019JB019076 e2019JB019076.

Zaliapin, I., Ben-Zion, Y., 2013. Earthquake clusters in southern California I: identification and stability. J. Geophys. Res. Solid Earth 118 (6), 2847–2864.

Zaliapin, I., Ben-Zion, Y., 2016. A global classification and characterization of earthquake clusters. Geophys. J. Int. 207 (1), 608–634.

Zaliapin, I., Ben-Zion, Y., 2020. Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain. J. Geophys. Res. Solid Earth 125. https://doi.org/10.1029/2018JB017120 e2018JB017120.

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., Wong, H., 2008. Clustering analysis of seismicity and aftershock identification. Phys. Rev. Lett. 101 (1), 018501.

Zhuang, J., Ogata, Y., Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrence. J. Am. Stat. Assoc. 97, 369–380.