



# A Statistical Approach Towards Fast Estimates of Moderate-To-Large Earthquake Focal Mechanisms

Marisol Monterrubio-Velasco<sup>1\*</sup>, Jose C. Carrasco-Jimenez<sup>1</sup>, Otilio Rojas<sup>1,2</sup>, Juan E. Rodríguez<sup>1</sup>, Andreas Fichtner<sup>3</sup> and Josep De la Puente<sup>1</sup>

<sup>1</sup>Barcelona Supercomputing Center - Centro Nacional de Supercomputación, Barcelona, Spain, <sup>2</sup>Escuela de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela, <sup>3</sup>Institute of Geophysics, ETH Zurich, Zurich, Switzerland

## OPEN ACCESS

### Edited by:

Alice-Agnes Gabriel,  
Ludwig Maximilian University of  
Munich, Germany

### Reviewed by:

Wenhuan Kuang,  
Southern University of Science and  
Technology, China  
Stefania Tarantino,  
University of Naples Federico II, Italy

### \*Correspondence:

Marisol Monterrubio-Velasco  
marisol.monterrubio@bsc.es

### Specialty section:

This article was submitted to  
Geohazards and Georisks,  
a section of the journal  
Frontiers in Earth Science

**Received:** 19 July 2021

**Accepted:** 18 February 2022

**Published:** 07 April 2022

### Citation:

Monterrubio-Velasco M,  
Carrasco-Jimenez JC, Rojas O,  
Rodríguez JE, Fichtner A and  
Puente JDI (2022) A Statistical  
Approach Towards Fast Estimates of  
Moderate-To-Large Earthquake  
Focal Mechanisms.  
Front. Earth Sci. 10:743860.  
doi: 10.3389/feart.2022.743860

Emerging high-performance computing systems, combined with increasingly detailed 3-D Earth models and physically consistent numerical wave propagation solvers, are opening up new opportunities for urgent seismic computing. This may help, for instance, to guide emergency response teams in the wake of large earthquakes. A key component of urgent seismic computing is the early availability of source mechanism estimates, well before conventional and time-consuming moment tensor inversions are carried out and published. Here, we introduce a methodology that rapidly estimates focal mechanisms (FM) for moderate and large earthquakes ( $M_w > 4.0$ ) by means of statistical and clustering algorithms. The fundamental rationale behind the method is that events of a certain size tend to be similar to other events of similar size in similar locations. In this work, two different strategies are used to provide different FM solutions: the first is based only in spatial considerations including statistical analysis, and the other one is based on a data clustering algorithm. We exemplify our methodology with six different subsets of the open-access Global Centroid Moment Tensor (GCMT) catalog. Specifically, our study datasets include events from Japan, New Zealand, California, Mexico, Iceland, and Italy, which represent six seismically active regions, with a large FM variability. Our results show a 70–85% agreement between our fast FM estimates and inversion results, depending on the particular tectonic region, dataset size, and magnitude threshold. In addition, our FM estimation strategies only spend few seconds for processing, since they are totally independent of seismic record retrieval and inversion. Albeit not meant to be a substitute for CMT inversions, our methodologies can bridge the time gap between earthquake detection and FM inversion.

**Keywords:** focal mechanism, fast-response, urgent computing, GCMT catalog, DBSCAN clustering, source parameters

## 1 INTRODUCTION

Fast earthquake magnitude and focal mechanism (FM) estimates are key information for rapid emergency response applications, including urgent post-event seismic simulations. Few hours after the occurrence of a large-magnitude earthquake, these simulations aim to map ground shaking in the affected region to show areas of high intensity motion, which may host potential structural damages. An FM solution describes the fault-plane orientation and directions of principal stresses in the area

where the earthquake occurred (Udias and Buforn 2017; Maeda 1992). The FM is given by three angles, two of them, the strike  $\phi$  and dip  $\delta$  angles, geometrically define the fault plane, and the third angle, the rake  $\lambda$  measures the direction of fault slip. Thus, an FM is used for the mathematical description of seismic sources in terms of the moment tensor (MT) equivalent body forces. The most general approach to determine an FM solution is by computing the centroid moment tensor (CMT). For moderate-to-large magnitude earthquakes, CMTs are usually obtained from waveform or spectral data inversions (Dziewonski et al., 1981; Dziewonski and Woodhouse 1983). Although, approaches to estimate earthquake location and magnitude are consolidated and extremely fast, automatic solutions for FM estimates are not always provided by the seismological agencies, or are only available at later times after waveform inversions have been completed (Tarantino et al., 2019). However, it is worth noting that significant efforts have been made towards almost real-time determination of FM (or CMT) by using different methods and algorithms. For example, Lin et al. (2019) and Melgar et al. (2012) exploit GPS networks, while other works use source inversion algorithms based on modelling of the W-phase, a very long-period phase (100–1,000 s) arriving at the same time as the P-wave (see, e.g., Duputel et al., 2012). Another approach relies on the azimuthal distribution of early P-wave peaks in displacement, velocity and acceleration traces (Tarantino et al., 2019). Moreover, several packages have been developed for automatic MT determination, such as Scisola (open-source software developed by Triantafyllis et al., 2016) or Gisola (Triantafyllis et al., 2021). Scisola is an open source Python based software for automatic MT calculation that combines two platforms, ISOLA (Sokos and Zahradnik 2008) and SeisComp3 (Weber et al., 2007). Gisola is an evolved version of Scisola that applies enhanced algorithms for waveform data filtering. However, the response times of most aforementioned techniques are constrained by the retrieval times of their input data, i.e., seismic records of the recent earthquakes (Melgar et al., 2012; Scognamiglio et al., 2010, and ref. there in). Therefore, the development of alternative approaches for fast FM estimations, independent of seismic records, may serve as temporary replacement until inversions have finished. This work explores statistical approaches, with the assistance of clustering algorithms, to estimate FM solutions for a new earthquake based on the similarity of  $\phi$ ,  $\delta$  and  $\lambda$  with respect to past events. The historical databases are gathered by the open-access Global Centroid Moment Tensor (GCMT) catalog (GCMT, Ekström et al., 2012). As input data, our methodologies only require the hypocentral location and magnitude of the new earthquake. This information is promptly provided by different seismological agencies with a latency of few seconds after earthquake occurrence. Once triggered, our methodologies can provide FM estimates within seconds, thereby enabling real-time affectation analyses before FM inversions become available. Our results can be potentially used to add directivity information into fast shaking assessment estimates, shortly after the earthquake is recorded, because it is precise for large events which have damaging potential. The method, however, is not universal, because it behaves worse if

smaller magnitude events are used. Hence it is no substitute for CMT inversion, but a provider of fast estimates. It is important to acknowledge that the methodology results not in one single best-fitting result, but in a collection of results among which is a good-fitting result. For shaking assessment (see, e.g., Wald et al., 1999, Wald et al., 2008) this is not a strong limitation, as few scenarios can be computed using all CMT estimates and the best fit can be assessed a posteriori. Last but not least, the methodology has a potential for probabilistic seismic hazard (PSHA, see e.g. Baker 2008; Mulargia et al., 2017) studies, especially those taking into account CMT information in their attenuation relationships or seismic modelling components. When populating hypothetical future earthquakes, in the so-called earthquake rupture forecast, CMTs could be estimated using our method, thus not necessarily restricting such earthquakes to mapped faults and their prescribed tectonic regimes. In fact, using catalog information (i.e. our CMT estimate) to make assessments about a hypothetical scenario is at the very core of PSHA. Nevertheless a seismic scenario computed using our CMT estimates would lack any kind of probabilistic component: it would be a deterministic scenario with associated uncertainties inherent to the CMT estimation methodology. The proposed methodology for FM estimation exploits information on hypocenters and magnitudes of catalogued neighboring events. Specifically, we employ conventional statistical analysis and the automatic DBSCAN clustering algorithm (Ester et al., 1996). To measure distance between two FM solutions, we use the Minimum Rotated Angle (MRA) metrics Kagan (2007) that quantify similarities between two double-couple (DC) sources in absolute angle degrees. It is worth noting that non-DC components, typically associated with volcanic activity and fluid-related earthquakes, are not the focus of our work because they do not typically produce large-magnitude, i.e. highly damaging, events (Stierle 2015; Wang et al., 2018). We assess the accuracy of the methodology through statistical comparisons of our FM results with published GCMT inversion solutions. For validation, we select six seismically active regions with large variability of the rupture mechanisms for the selected historical earthquakes used as estimation basis. These regions are Japan, New Zealand, California, Mexico, Iceland, and Italy.

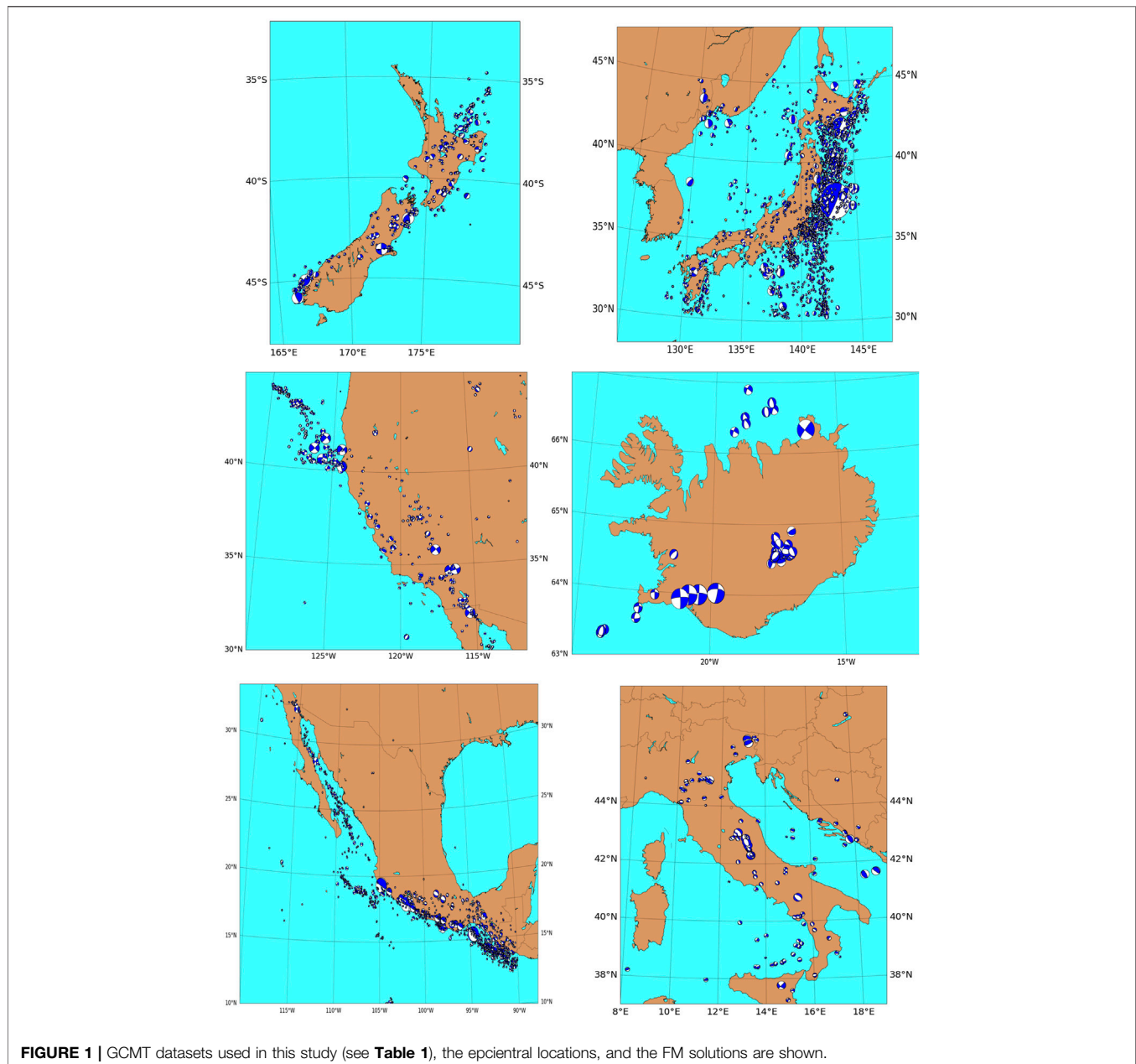
## 2 BACKGROUND

### 2.1 Global Centroid Moment Tensor Catalog and Input Datasets

We obtained the datasets used in this work from the GCMT catalog (Dziewonski et al., 1981; Ekström et al., 2012), available through the Searchable Product Depository (SPUD) of the Incorporated Research Institutions for Seismology (IRIS) (Trabant et al., 2012). Global Centroid-Moment-Tensors from the GCMT project at Lamont-Doherty Earth Observatory are available through SPUD within minutes after their publication. Initial quick-CMT solutions are shown and are later updated to GCMT solutions when updates arrive. At present, the GCMT catalog contains more than 40,000 earthquakes. From the queried catalog we use the Event information which includes the

**TABLE 1** | Filters applied to the GCMT database to extract six regional datasets. We list the total amount of events in each dataset. Databases have been queried from IRIS "https://ds.iris.edu/spud/momenttensor"

Subset	Latitude (°)	Longitude (°)	Depth (km)	Event count	Magnitude ( $M_w$ )	Date
New Zealand	[-46.02, -34.42]	[166.1, 178.71]	[0, 518]	273	[4.8, 7.8]	[1965, 2018]
Japan	[30.03, 46.3]	[128.84, 147.1]	[0, 588]	2652	[4.6, 9.1]	[1967, 2019]
California	[29.72, 44.77]	[-129.76, -110.36]	[0, 30]	460	[4.4, 7.3]	[2010, 2019]
México	[10, 33]	[-120, -90]	[0, 251.1]	1705	[4.5, 8.2]	[1967, 2020]
Iceland	[63.06, 66.94]	[-24.43, -16.58]	[0, 33]	124	[4.6, 6.5]	[1976, 2018]
Italy	[34.95, 47.94]	[5.18, 21.0]	[0, 502]	692	[3.9, 6.9]	[1976, 2015]



**FIGURE 1** | GCMT datasets used in this study (see **Table 1**), the epicentral locations, and the FM solutions are shown.

hypocenter location, earthquake magnitude, Date-Time UCT, faulting geometry, and the FM solution of moderate to large events with magnitude  $M \geq 4.5$ . It is worth to highlight that in this work we only use the hypocenter location of the Event information to give the coherence of the first information registered after an earthquake occurs.

In this study, we consider six data subsets associated with six study regions: New Zealand, Japan, California, Mexico, Iceland, and Italy. Each subset is defined by the ranges of hypocentral locations, magnitudes and event times (Table 1). Figure 1 shows each regions and FM solutions, represented as beach balls. The beach ball sizes are proportional to the event magnitude.

### 2.2 Minimum Rotated Angle Metric to Measure Similarity Between Two Focal Mechanisms

In this work, we assess the accuracy of our FM solutions by means of the MRA metric, proposed by Kagan (2007). This metric measures the distance between two double-couple (DC) solutions in absolute angular terms, and it enables a comparison between two DC solutions obtained by different methods, as well as variations of earthquake FMs in space and time. Moreover, the MRA has been used widely, thereby allowing us to compare methodologies of different authors.

MRA requires computing the matrix eigenvectors  $\mathbf{t}$ ,  $\mathbf{p}$  and  $\mathbf{b}$ , which belong to  $\mathbb{R}^3$  and represent the three orthogonal axes describing the radiation of P-waves from a DC point source (Frohlich 1996; Aki and Richards 2002). To compute the MRA, we consider the eigenvector components in terms of the strike  $\phi$ , dip  $\delta$ , and rake  $\lambda$  angles,

$$\begin{aligned}
 t_1 &= (-\sin \phi \sin \delta + \cos \phi \cos \lambda + \sin \phi \cos \delta \sin \lambda) / \sqrt{2} \\
 t_2 &= (\cos \phi \sin \delta + \sin \phi \cos \lambda - \cos \phi \cos \delta \sin \lambda) / \sqrt{2} \\
 t_3 &= (-\cos \delta - \sin \delta \sin \lambda) / \sqrt{2} \\
 p_1 &= (-\sin \phi \sin \delta - \cos \phi \cos \lambda - \sin \phi \cos \delta \sin \lambda) / \sqrt{2} \\
 p_2 &= (\cos \phi \sin \delta - \sin \phi \cos \lambda + \cos \phi \cos \delta \sin \lambda) / \sqrt{2} \\
 p_3 &= (-\cos \delta + \sin \delta \sin \lambda) / \sqrt{2} \\
 b_1 &= \cos \phi \sin \lambda - \sin \phi \cos \delta \cos \lambda \\
 b_2 &= \sin \phi \sin \lambda + \cos \phi \cos \delta \cos \lambda \\
 b_3 &= \sin \delta \cos \lambda
 \end{aligned}
 \tag{1}$$

The MRA  $\Phi$ , as defined in Kagan (2007), is given by

$$\Phi = \arccos \left[ \frac{1}{2} (|\mathbf{t}' \cdot \mathbf{t}''| + |\mathbf{p}' \cdot \mathbf{p}''| + |\mathbf{b}' \cdot \mathbf{b}''| - 1) \right]
 \tag{2}$$

Where  $\mathbf{t}'$ ,  $\mathbf{p}'$  and  $\mathbf{b}'$ , are the eigenvectors associated to one FM solution, while a second FM solution has  $\mathbf{t}''$ ,  $\mathbf{p}''$  and  $\mathbf{b}''$  as eigenvectors.

Following the formulation in Kagan (2007), Eq. 2 yields the correct value of the rotation angle for  $\Phi < 90^\circ$ . However, if Eq. 2 results in  $\Phi > 90^\circ$  then a more general equation should be applied to compute  $\Phi$ . For  $\Phi > 90^\circ$  to obtain the minimum rotation angle  $\Phi$  Eq. 3 is applied, in those cases the smallest absolute value dot product should be negative, and the other products should be positive (further details see Kagan (2007).)

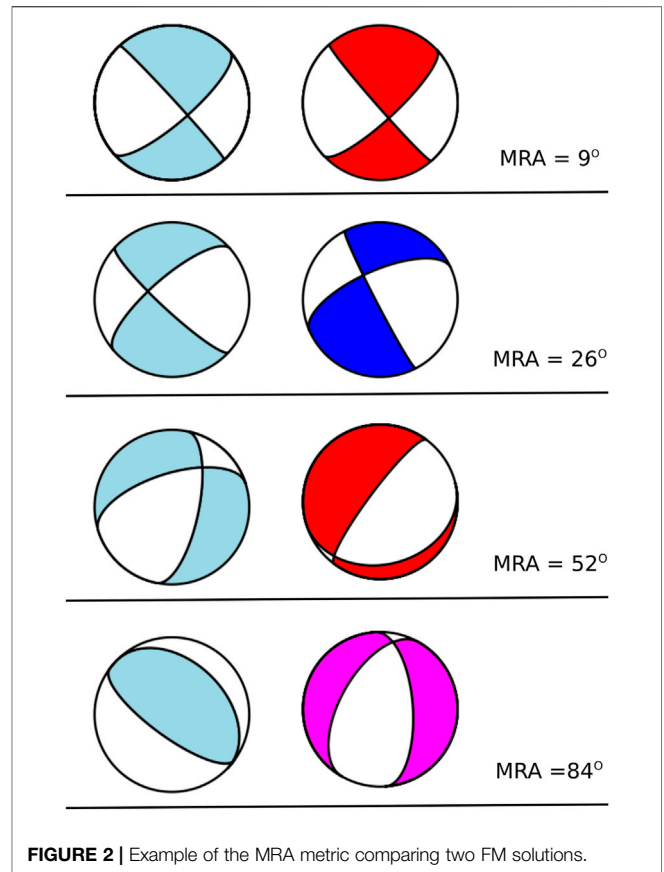


FIGURE 2 | Example of the MRA metric comparing two FM solutions.

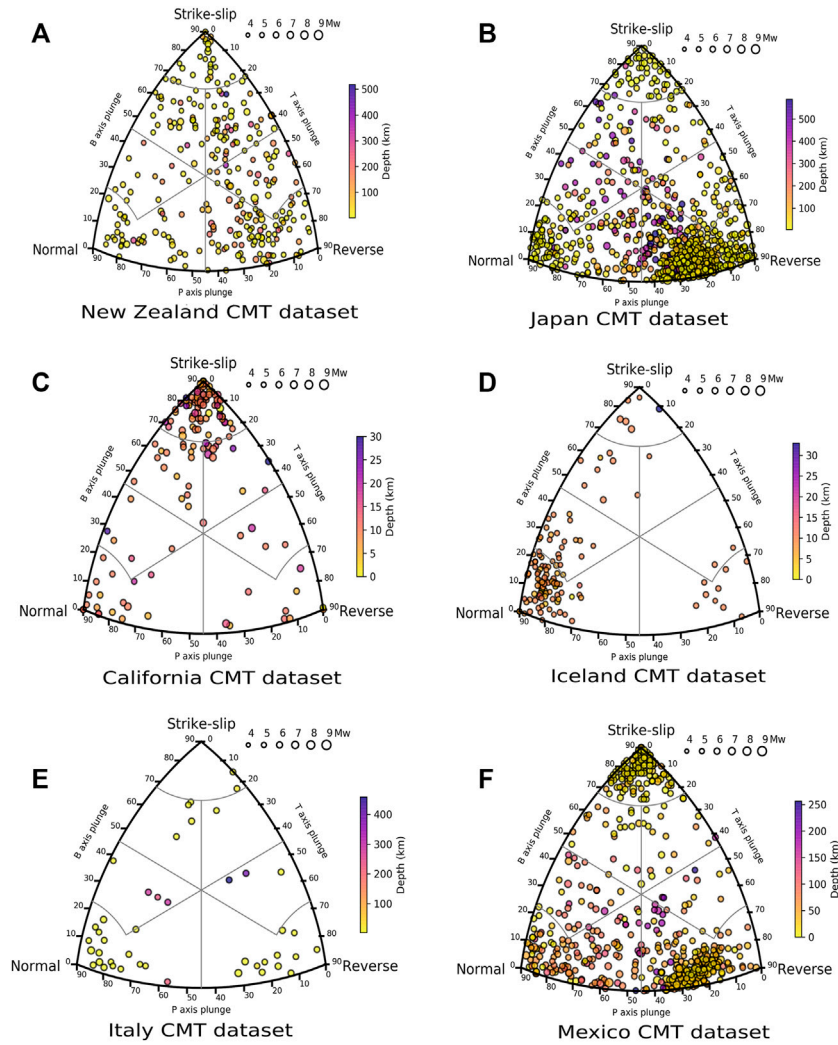
$$\Phi = \arccos \left[ \frac{1}{2} (\mathbf{t}' \cdot \mathbf{t}'' + \mathbf{p}' \cdot \mathbf{p}'' + \mathbf{b}' \cdot \mathbf{b}'' - 1) \right]
 \tag{3}$$

Figure 2 depicts some examples of different MRA values computed from the shown FM solutions.

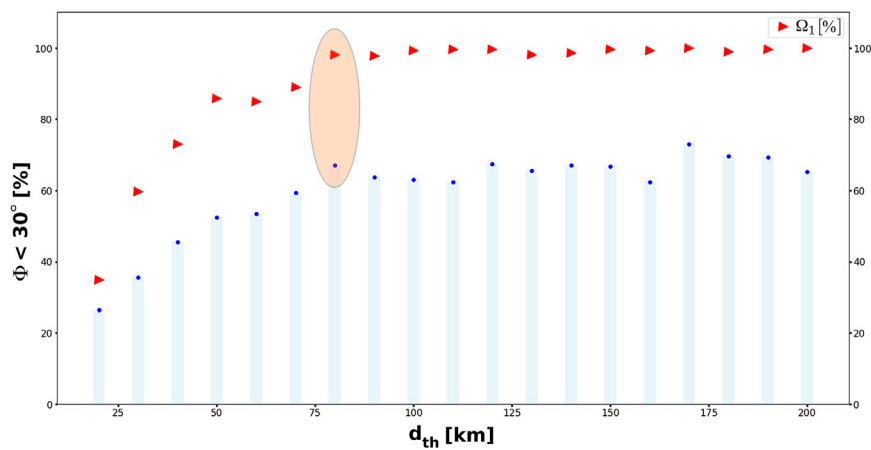
Different research results conclude that an acceptable agreement between two FM solutions is given by angle differences of the order of a few tens of degrees, while a strong variance corresponds to an angle difference greater than  $50^\circ$  (or  $60^\circ$ ) (Vannucci et al., 2004; Triantafyllis 2014; Triantafyllis et al., 2016). Triantafyllis et al. (2013) found an average error between manual and the automatic FM solutions computed using Scisola of  $\Phi \approx 37^\circ$ . Moreover, based on heuristic analyses from our statistical results, in this work we assume a similarity threshold  $\Phi_{th} < 30^\circ$ . Therefore, we suggest that two FM solutions with  $\Phi > \Phi_{th}$  are not comparable.

### 2.3 DBSCAN Clustering Algorithm

Class identification in spatial databases can be accomplished through the exploitation of clustering algorithms. Cluster analysis is developed under the assumption that spatial data has an implicit structure that can be uncovered by clustering algorithms. The clusters must comply with two characteristics: internal cohesion, also known as homogeneity, and external isolation, also known as separation. Thus, clustering techniques seek to summarize meaningfully different pattern



**FIGURE 3** | Kaverina's diagrams for selected six regions: **(A)** New Zealand  $M \geq 4.8$ , **(B)** Japan  $M \geq 5.5$ , **(C)** California  $M \geq 4.4$ , **(D)** Iceland  $M \geq 4.6$ , **(E)** Italy  $M \geq 4.6$ , and **(F)** Mexico  $M \geq 5.5$ .



**FIGURE 4** | Example of the selection process of the optimal  $d_{th}$  parameter. Red markers show the percentage of events with at least one neighbor inside the sphere ( $\Omega_1$ ). Blue marks indicate the percentage of events with at least one FM solution with  $\Phi < 30^\circ$ . Ellipse indicates the optimum  $d_{th}$  considered in the study as the inflection point in both curves.

**TABLE 2 |** Results of the methodology proposed in this work:  $d_{th}$  is the neighborhood radius size in kilometers;  $M_{th}$  the minimum magnitude considered in the study; Number of events the total number of earthquakes with a magnitude larger or equal than  $M_{th}$ ;  $\Omega_1$  and  $\Omega_3$  are the number of events with at least one or three neighbors inside the sphere of radius  $d_{th}$  respectively; and  $\Phi_{(30,\Omega_1)}$  and  $\Phi_{(30,\Omega_3)}$  depicts the percentage of events with at least one solution with a  $\Phi < 30$  computed from the Statistical method or DBSCAN algorithm respectively.

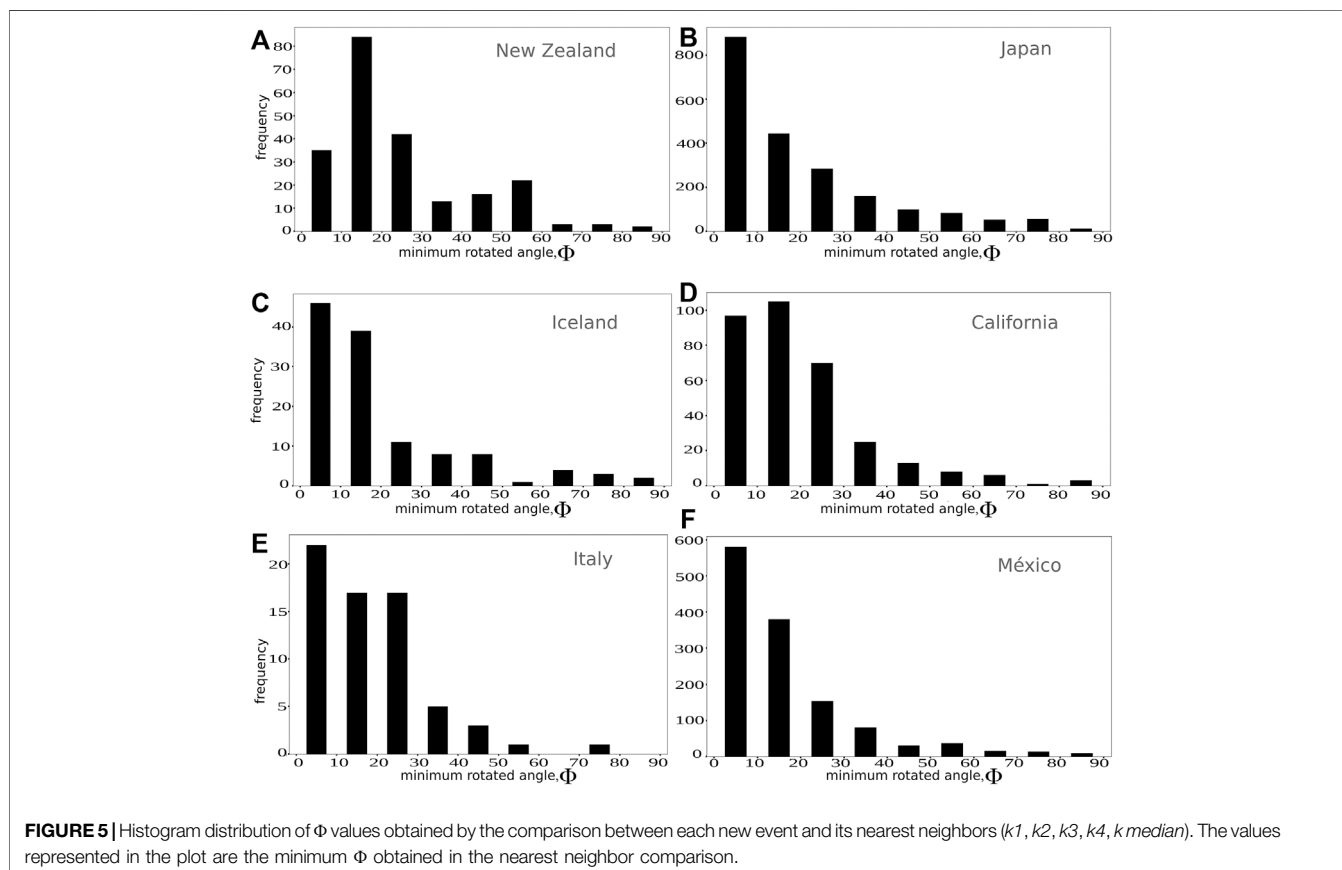
Region	$d_{th}$ [km]	$M_{th}$	Number of events	Nearest Neighbors		DBSCAN	
				$\Omega_1$	$\Phi_{(30, 1)}$	$\Omega_3$	$\Phi_{(30, 3)}$
New Zealand	80	4.8	273	220	73%	154	82%
Japan	80	5.0	2263	2085	77%	1771	79%
Japan	110	5.5	937	816	80%	621	81%
Japan	150	6.0	331	261	85%	196	88%
California	80	5.0	363	328	83%	301	79%
Mexico	60	5.0	1,426	1,302	86%	938	86%
México	80	5.5	588	542	87%	361	89%
Mexico	130	6.0	203	178	83%	101	89%
Italy	80	4.8	114	66	85%	38	92%
Iceland	70	4.6	124	122	77%	113	83%

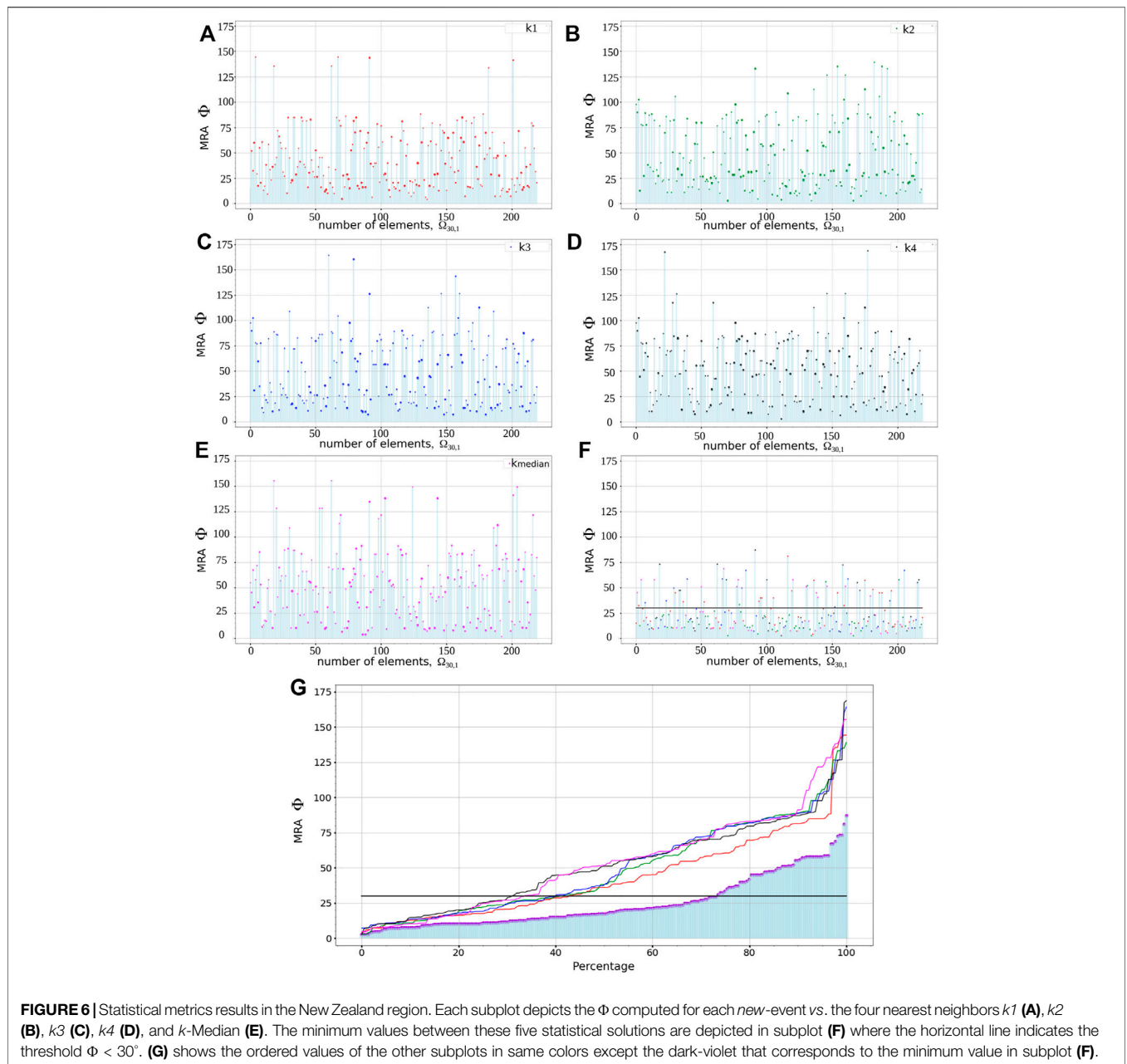
profiles by identifying segments of points in which observations within the same cluster exhibit high degrees of similarity (homogeneity) while differing in some respects from observations in other clusters (separation). In this work, a density-based method, namely Density-based Spatial Clustering

of Applications with Noise (DBSCAN) Ester et al. (1996) is used to identify geological or structural profiles, which are depicted by the clusters uncovered by the algorithm. DBSCAN relies on two parameters: a distance threshold  $\epsilon$ , which indicates the maximum distance between two observations, and the minimum number of observations  $n$  to form a cluster. The first step is to describe each observation, i.e., an earthquake with a focal mechanism, using a multi-dimensional real-valued vector representation described by the hypocentral location, strike, dip and rake. The set of observations is fed to the DBSCAN algorithm to uncover the intrinsic geological or structural profiles. In order to obtain the optimal number of profiles (i.e. clusters), the distance threshold and the minimum number of samples, both of which parameters have an effect on the number of clusters. The distance threshold  $\epsilon$  is estimated using the strategy based on knee/elbow methodology (Satopaa et al., 2011). And the minimum number of samples  $n$  are set equal to two

### 3 STATISTICAL AND CLUSTERING METHODOLOGY FOR FAST ESTIMATING FOCAL MECHANISMS

In the following subsection we provide a general description of the proposed method. Subsequently, we present its application to several regions of interest.





### 3.1 Methodology at a Glance

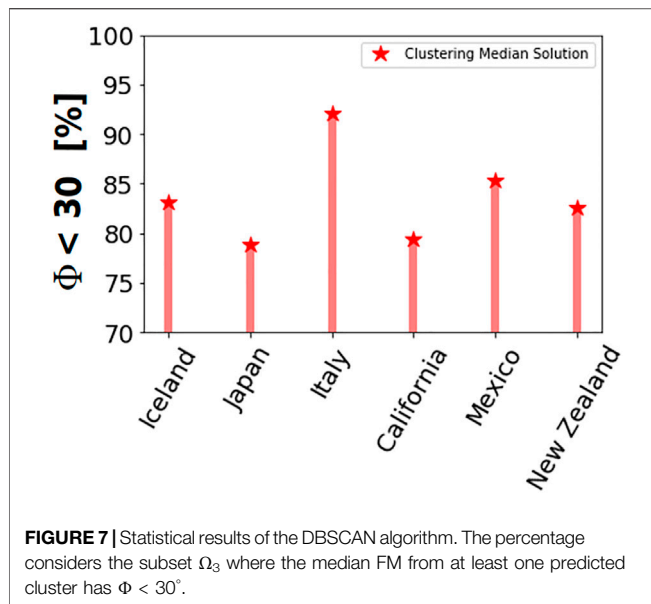
The proposed methodology to estimate an FM solution is based on spatial properties. Thus, once the hypocentral location of the new event with an unknown FM solution is available, we select a spherical neighborhood centered at its hypocentral location. Previous events inside this neighborhood will then be used to suggest different FM solutions. In this work we propose two different methods for this:

- (a) The first method is based on spatial assumptions and uses the hypocentral Euclidean distance  $\Delta$  as a metric to quantify the distance between the new event and each of its neighbors. To

measure it, we apply the Haversine formula that determines the great circle distance between two points (Sinnott 1984),

$$d = 2R \arcsin \left( \sqrt{\frac{\sin^2(\frac{\psi_2 - \psi_1}{2}) + \cos(\psi_1)\cos(\psi_2)\sin^2(\frac{\lambda_2 - \lambda_1}{2})}{2}} \right) \tag{4}$$

Where  $R$  is the radius of the Earth, and  $\psi_1, \psi_2$ , and  $\lambda_1, \lambda_2$  are the latitude and longitude coordinates of two points, respectively. Once  $d$  is computed, the Euclidean distance from the new event to



each neighbor is given by  $\Delta = \sqrt{d^2 + z^2}$ , where  $z$  is the difference in depth. Therefore, we select the four nearest neighbors (hereafter denoted as  $k_1, k_2, k_3, k_4$ ) from the closest to the farthest, respectively. Here, we assume that these four nearest earthquakes could have occurred under similar geological and tectonic conditions or under similar faulting type. Therefore, this method gives four FM solutions coming from the solutions of the four closest neighbors. The number four is an empirically motivated choice, found to produce useful results in the statistical analyses presented later. Additionally, we compute a median FM solution from all neighbors in the sphere. This median value is computed using the three angles  $\phi_i, \delta_i,$  and  $\lambda_i$ , where  $i$  is related to each the  $i$ -neighbors. This fifth FM solution, named  $k$  - median could be related with a “typical” faulting type in that region. To apply this method we consider a minimum of one neighbor ( $n = 1$ ) inside the sphere.

(b) The second method is based on the DBSCAN clustering algorithm. To apply this algorithm, we select four features, the distance from each neighbor to the hypocenter of the new event, and also the three angles of the FM of each neighbor. Once the algorithm automatically detects the clusters, we compute the position of each centroid, as well as the FM solution from the median of  $\phi_x, \delta_x,$  and  $\lambda_x$  from the events that belong to each  $x$  cluster. The number of solutions depends on the number of clusters detected in the DBSCAN algorithm. It is worth noting that a minimum number of neighbors to apply the cluster algorithm should be assumed. Hence, in this application we take a minimum of three neighbors ( $n = 3$ ). The clusters automatically detected could

**TABLE 3** | Earthquakes analyzed using the statistical methodology proposed in this work (see. section 4).

Region	$M_w$	Date	Hypocentral Location	$d_{th}$ [km]
			(lon,lat,depth)	
New Zealand	5.0	30/11/2018 4:17	166.12, -45.57, 31.5 km	80
Japan	5.2	1/6/2011 3:14	143.84, 39.79, 14.4 km	80
California	5.3	28/10/2019 11:01	-125.9, 41.91, 15.0 km	80
Iceland	4.8	29/5/2009 21:33	-22.52, 63.82, 12.7 km	70
Italy	5.2	20/5/2012 13:18	11.49, 44.81, 12 km	80
Mexico	7.4	23/6/2020 15:29	-96.06, 16.04, 20 km	60

reflect information about different regional tectonic features.

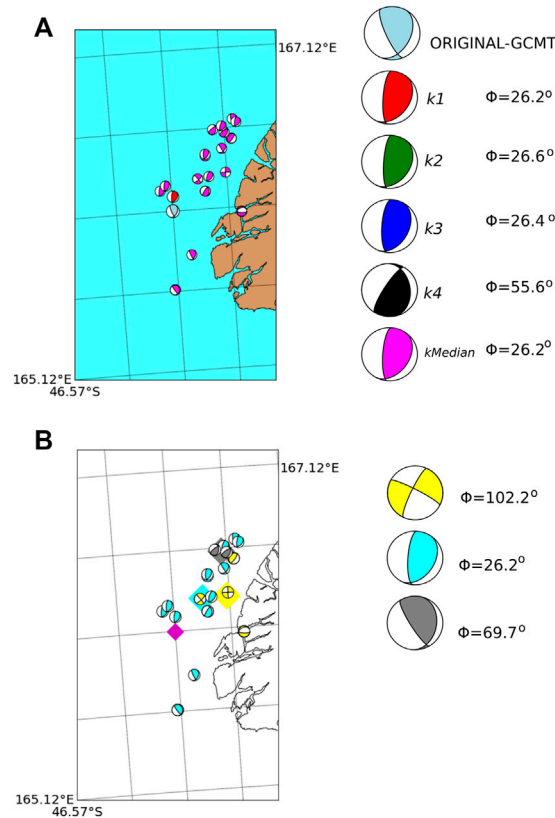
### 3.2 Methodology Application

To apply the methods described in the previous subsection we follow a step-by-step process

The first step is the compilation of suitable datasets. For this, we access moment tensor information of each selected region, summarized in Table 1, through the IRIS-SPUD website 1 (Trabant et al., 2012). We perform an exploratory study to visualise the statistical FM distribution using Kaverina diagrams (Kaverina et al., 1996) which classify events according to their double-couple (DC) rupture type. Figure 3 shows the Kaverina diagrams for each study region using the visualizing tool developed in (Álvarez-Gómez 2019). Clearly dominant rupture types are observed in some regions, for example, strike-slip events in California, or normal faulting in Iceland. Other regions, such as New Zealand, Mexico and Japan, depict a larger variety of rupture types. The main purpose of visualizing historical datasets in Kaverina diagrams is to know the rupture-type at each region. This information is useful for the interpretation of results in the following section.

In the second step, we perform a statistical analyses using the two methods described in the previous subsection. This statistical evaluation aims to test the proposed methodology using each earthquake in the dataset as an independent observation, considering them as a hypothetical new event with unknown FM. Moreover, the statistical analysis allows us to quantify the accuracy of the proposed methodology by means of the MRA computed between the proposed and the original solution. In this work, we choose an MRA threshold value of  $\Phi_{th} = 30^\circ$  (Triantafyllis 2014; Kagan 2007), below which two FM solutions are deemed sufficiently similar, suggesting that the proposed one may indeed be used as a first guess for the new earthquake with unknown FM solution. This value is motivated by earlier studies on the similarity between FM solutions (Triantafyllis et al., 2013, Triantafyllis et al., 2016; Altunel and Pinar 2021), and by the observation that differences between FM solutions for an earthquake provided by different agencies are typically below  $\Phi = 30^\circ$ . From this point of view both methods described in the previous subsection can be considered supervised





**FIGURE 8 | (A)** Nearest neighbor solutions for a particular example in New Zealand (see **Table 3**). The map shows the location and focal mechanism of the test event (light-blue) and its neighbors. From closer to farther solutions, *k1* (red beachball), *k2* (green), *k3* (dark-blue), *k4* (black), and *k - median* (magenta). The MRA values  $\Phi$ , listed to the right, measure the differences between each neighbor and the test event solution. **(B)** DBSCAN results for the same earthquake of **Figure 8A** in light-blue. The magenta diamond indicates its epicentral location. The beach balls with the same color belong to the same cluster defined by the DBSCAN algorithm. The centroids of each cluster are shown in diamond markers of the corresponding color. Those centroids are computed as the mean value of the latitude, longitude and depth from each event in the cluster. At the right side the median FM solutions computed in each cluster are shown. The MRA value  $\Phi$  relative to the test-event is also indicated.

techniques, because the FM solution for the *new*-event is already known.

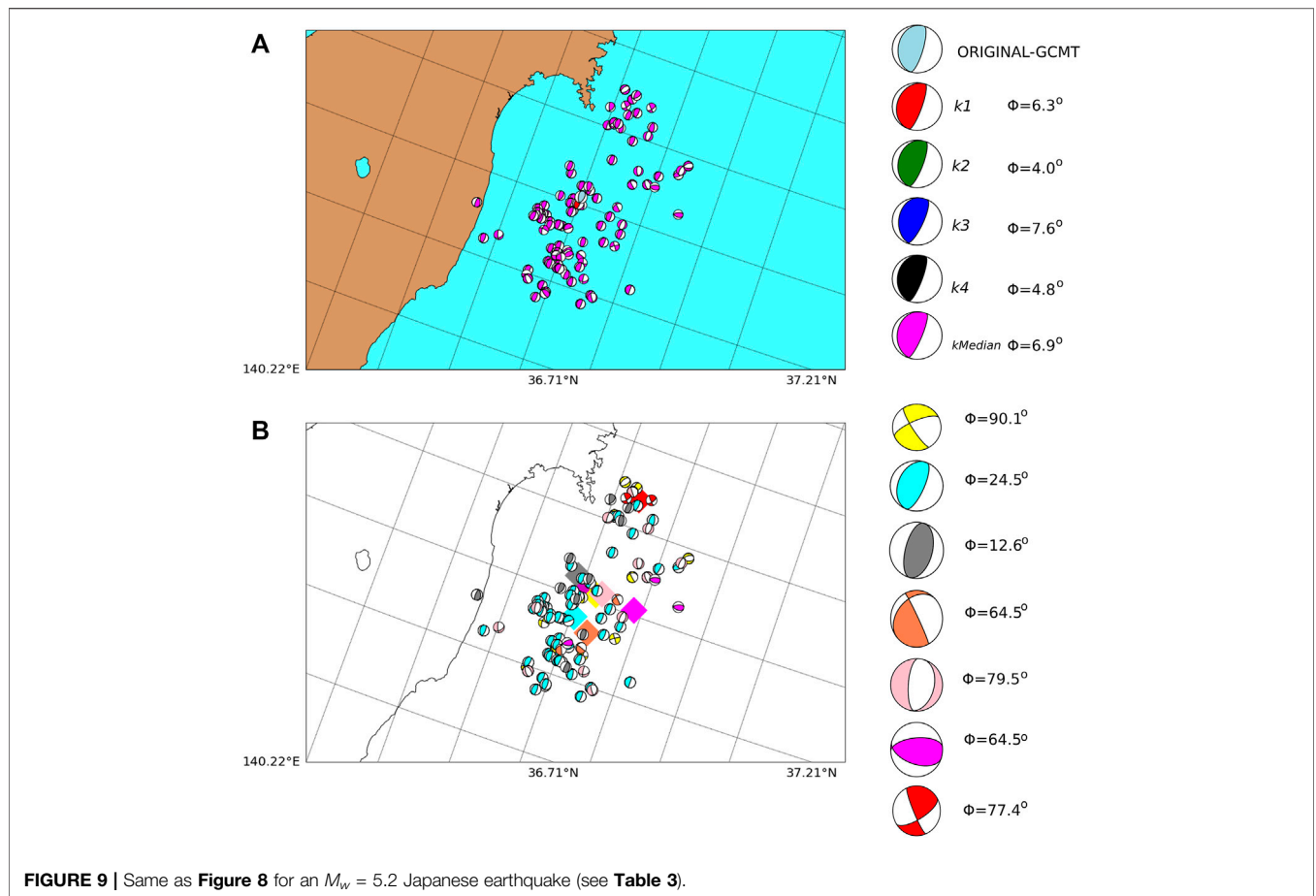
As previously mentioned, each method requires a minimum number of neighbors around the new event. At least three neighbors,  $n = 3$ , for the DBSCAN algorithm, and one neighbor,  $n = 1$ , for the hypocentral Euclidean distance. In the following, we will denote by  $\Omega_n$  the number of events in the dataset that fulfill the requirement of having a minimum of  $n$  neighbors.

In the third, step we optimize the neighborhood size. Trying to find an optimal size, we modify the radius  $d_{th}$  of the spherical neighborhood. To find the optimum  $d_{th}$ , we repeat the previous step, increasing the neighborhood radius after testing all the events in the dataset. We start with a radius of 20 km, and increase up to 200 km in 10 km steps. An optimum  $d_{th}$  value must satisfy two conditions:

- (1) a large number of events in the dataset with at least one neighbor inside the sphere, that is, a large  $\Omega_1$ ,
- (2) and a large percentage of events with at least one neighbor with  $\Phi < 30^\circ$ .

In **Figure 4**, we exemplify this search for an optimal  $d_{th}$ . Blue bars indicate the percentage of events with at least one solution with  $\Phi < 30^\circ$ . The red triangles depict the percentage of events in the dataset considered in the analysis, i.e., events with at least one neighbor inside the sphere ( $\Omega_1$ ). For all regions, as the neighborhood size increases, the percentage of spheres with at least one neighbor increases until an inflection point, beyond which it remains constant. We consider this inflection point as the threshold size  $d_{th}$ . This indicates that, although the neighborhood size increases (and thus the number neighbors), the similarities between FM solutions are not further improved. Therefore, the optimal radius of the sphere maximizes both percentages at the same time, as indicated by the orange ellipse in **Figure 4**.

To study how the minimum magnitude  $M_{th}$  in a dataset modifies the threshold radius  $d_{th}$ , we repeat the same analysis considering different  $M_{th}$  values. However, this analysis is possible only for large catalogs, such as for the Japan and Mexico datasets. **Table 2** depicts the selected  $d_{th}$  values for each region.



## 4 STATISTICAL AND CLUSTERING RESULTS

In the first part of this section we present the results of the two methods applied to the selected regions. This will be followed in the second part by specific examples for selected earthquakes in each region.

**Table 2** shows the main results of the statistical analysis for each region. It is worth noting that for a similar  $M_{th}$  the radius  $d_{th}$  is similar to with few tens of kilometers for different regions. Also, as  $M_{th}$  increases, the radius  $d_{th}$  becomes larger. The statistical results indicate that around 70–85% of the tested events have an FM solution similar to those computed in the GCMT catalog ( $\Phi < 30$ ). This percentage range is accomplished for both, the clustering algorithm and the spatial metrics. Moreover in the regions with similar rupture types, such as California, Mexico, Japan or Iceland (**Figure 3**), this percentage increases. On the other hand, this similarity decreases in regions where the faulting types are more diverse, including New Zealand and Italy. In the datasets with few events such Italy, performance of both methods is lower because the neighborhood tends to have scarce data, thereby preventing improvements of the statistics.

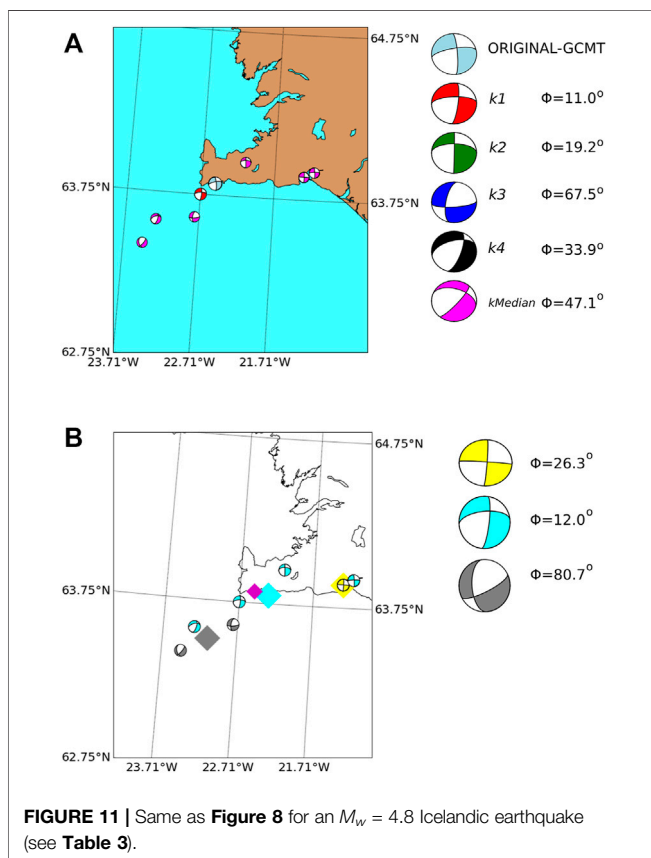
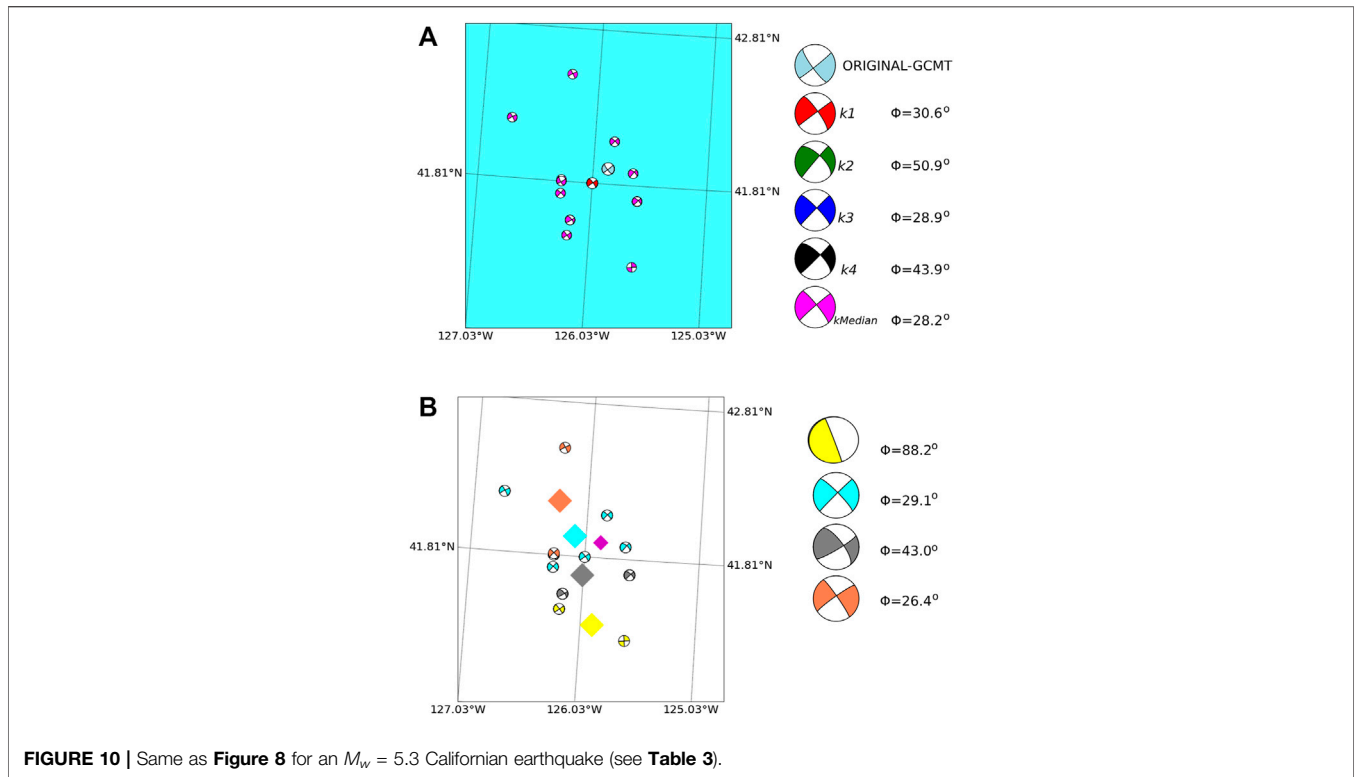
**Figures 5A–F** shows the MRA histograms for each region, obtained with the statistical metrics ( $\Omega_{th,1}$ ). We observe that the

MRA distribution is clearly skewed towards low MRA values, mostly reaching  $\Phi < 30$ .

In **Figure 6**, a detailed plot shows the behavior of  $\Phi_{(30, 1)}$  considering the nearest neighbors and the median statistical solution for New Zealand. **Figure 6G** shows the ordered  $\Omega$  values obtained in the (a-f) subplots. In dark-violet the minimum value between  $k1$ ,  $k2$ ,  $k3$ ,  $k4$  and  $k - median$  is depicted, from there is easy to observe how the FM solutions is improved.

In general, the nearest neighbor,  $k1$ , is the most likely to provide a similar FM solution. This is expected because nearby earthquakes are likely to rupture the same fault or under similar tectonic regimens. However, we also observe that the median FM solution produce MRA of  $\Phi < 30^\circ$ . In those cases, most neighbors have similar rupture types than the new-event. Moreover, in some cases, also another neighbors ( $k2$ ,  $k3$ ,  $k4$ ), show a smaller  $\Phi$  than  $k1$ . This could be related to hypocenter location uncertainties.

Statistical results of the DBSCAN algorithm are shown in **Figure 7** for each region. The percentages shown in this figure consider that at least one FM solution computed using the median values in each cluster has  $\Phi < 30^\circ$ , for events with at least three neighbors ( $\Omega_3$ ). The median FM solutions are computed by using the events belonging to each automatically detected cluster. We observe that the DBSCAN algorithm



computes similar percentages to the nearest neighbors statistics for those events that fulfill the number of neighbors  $\Omega_3$ .

### 4.1 Examples of the Methodology for Selected Regions

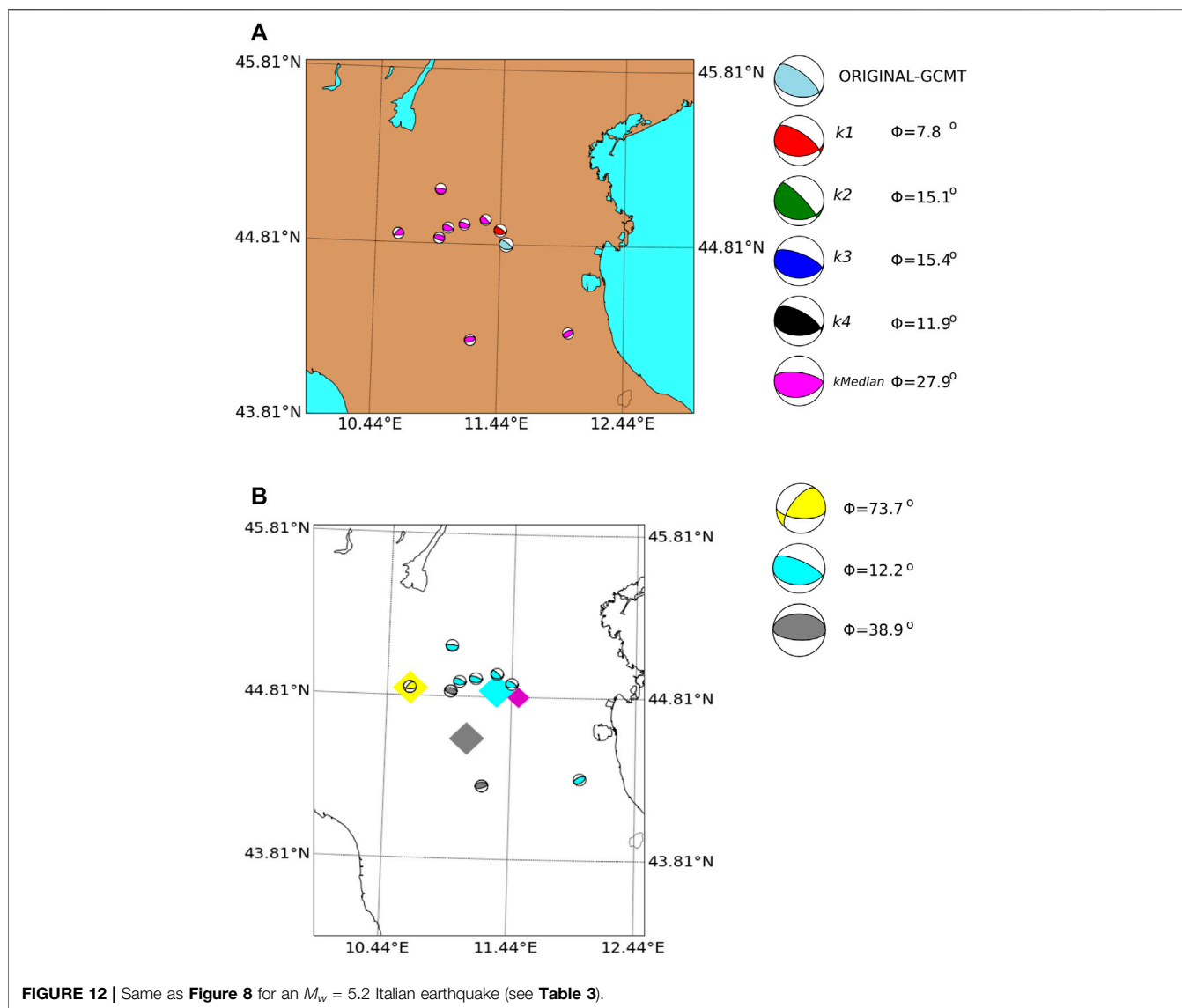
In this section, we apply the method presented above to specific earthquakes in different regions. From the databases we randomly choose one test event with at least three neighbors inside the sphere of size  $d_{th}$ . The selected earthquakes are listed in **table 3**.

New Zealand **Figure 8A** shows the spatial epicentral locations and focal mechanisms of the test event, it neighbors inside the sphere, and the median event. and their focal mechanism in magenta color is plotted. The closest neighbor is shown in red. In this example, we obtain MRA values of  $\Phi < 30^\circ$  for the three nearest events  $k1, k2$  and  $k3$ , and for the median event  $k - median$ .

The DBSCAN clustering results are presented in **Figure 8B**. The different colors indicate the clusters detected in the DBSCAN algorithm and the diamonds mark their respective cluster-centroid locations. In this particular example, the smallest cluster has the more similar FM solution to the test event.

Japan The dataset of the Japan region is the largest in our analysis. We also choose randomly an earthquake with more than three neighbors inside a sphere of radius  $d_{th} = 80 \text{ km}$  size. Our threshold magnitude in this example is 5.0.

The hypocentral locations and focal mechanisms of the test event and its four nearest neighbors are displayed in **Figure 9**. In this example, the similarity is high for all neighbors. This is also



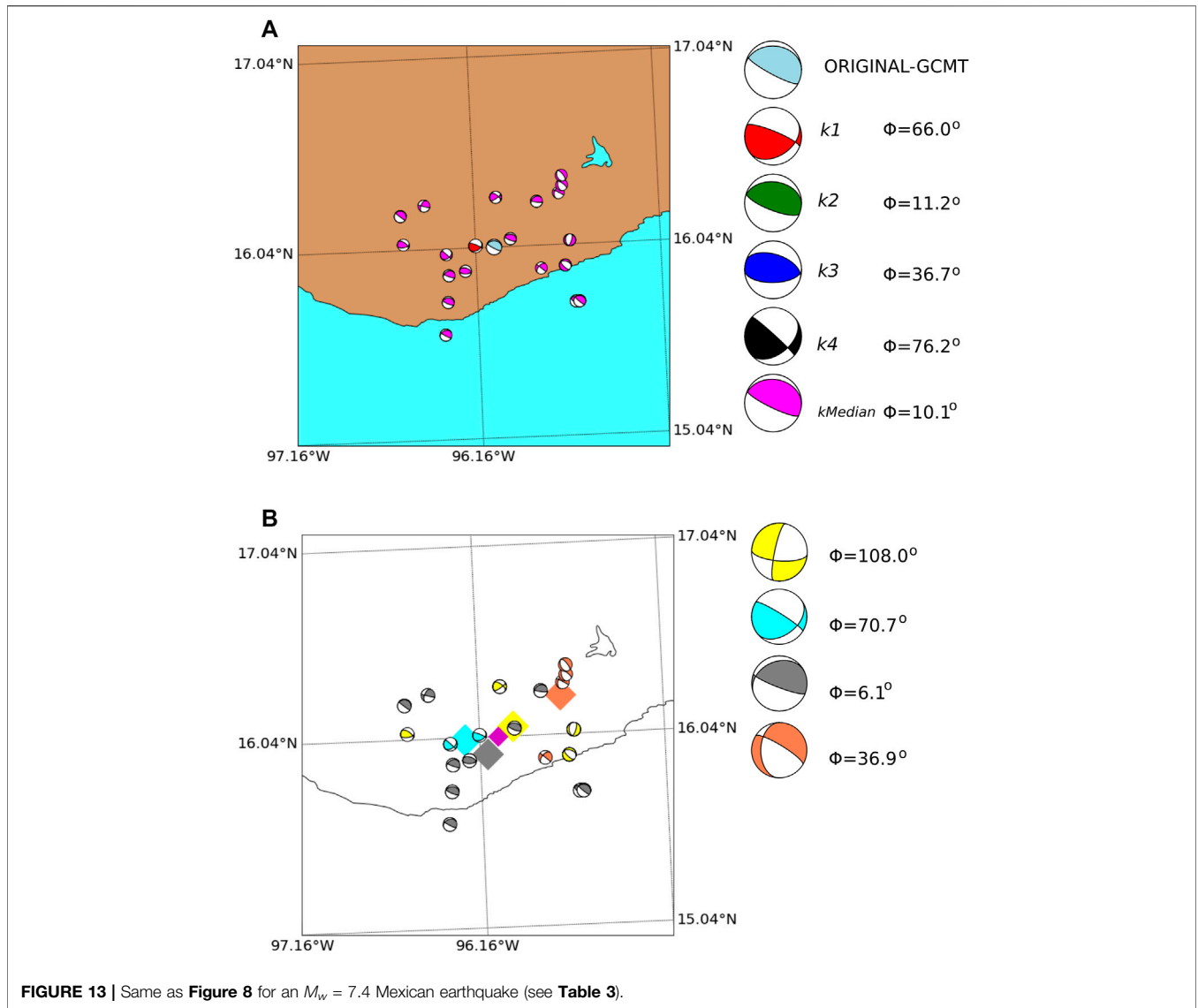
reflected in the clustering algorithm which detects three clusters. Is worth to note that the clustering algorithm joins in the yellow cluster different event solutions. Let us keep in mind that the objective of cluster analysis is to generalize the solution model, thus, some individual events may be misclassified, which does not diminish the contribution of the solution. The generalization seeks to have good classification of spatial events in the general population of events, and low error rates do not mean perfect classification. However, it is important to note that the rest of the clusters in **Figure 9** share many similarities between their elements reflecting a good clustering selection by the DBSCAN. The most similar median FM solution is for the larger cluster. These results could indicate similar geological patterns in that region, producing events with similar FM solutions.

California The California region shows a predominant strike-slip rupture style, as seen in **Figure 3**. The results of

our method, displayed in **Figure 10**, reveal a similarity close to  $\Phi < 30^\circ$ . In this example, the  $k$ -median is the event with the lowest MRA. The cluster analysis gives three groups, with the largest cluster being the one with the closest centroid and the smallest median MRA.

Iceland In this example, the nearest neighbors metric offers a better solution for the closest neighbor  $k1$  (see **Figure 11**). The clustering results also provide a good solution for the cluster with the closest centroid to the test event, thus indicating short-range similar tectonics over distances of few kilometers).

Italy The Italian region shows a predominantly normal and reverse rupture type distribution (**Figure 3**). However, this database has only 118 earthquakes, and so we consider all events provided in the GCMT catalog. In the example shown in **Figure 12**, we observe a large similarity between the selected earthquake and its neighbors, with the lowest MRA for the nearest neighbor  $k1$ . The results from the clustering algorithm



also provide a similar solution for the larger cluster with the closest centroid.

Mexico The Mexican region has one of the largest datasets. The statistical results provide one event with  $\Phi < 30^\circ$ . In this particular example, we observe that the *k* – median neighbor gives the largest similarity (**Figure 13**). Moreover, the median FM of the largest cluster provides the most similar solution to the test event.

### 5 DISCUSSION AND CONCLUSION

In this work we propose an inversion-free approach to the fast estimation of FM solutions, primarily intended to serve as input for urgent seismic simulations or similar problems with computation deadlines. Within this context, the rapid estimation of source parameters is highly relevant. Moreover, the predictive power of historical datasets represents an opportunity that is worth being exploited. In particular, in this

work we develop a methodology for the estimation of FM solutions that is based purely on past earthquakes. After statistically validating the experiments, we find some important advantages of this statistical estimation tool.

Firstly, the method is extremely fast, with time to solution at  $< 10$  s. In terms of accuracy, comparing our results to those provided in Triantafyllis (2014) by the Scisola software, we observe similar statistics in the Kagan diagrams (**Figure 5**). We have used a threshold  $\Phi \leq 30^\circ$  for MRA which is similar to the differences between various FM inversion results and published by different agencies. For example, in Scognamiglio et al. (2010), the authors estimated the FM solution for the 2009/04/06  $M_w = 6.1$  L’Aquila earthquake as  $[\theta = 139, \delta = 48, \lambda = -87]$  for nodal plane 1, and  $[\theta = 314, \delta = 42, \lambda = -94]$  for nodal plane 2. In contrast, the FM solution for the same earthquake given in the GCMT catalog was  $[\theta = 120, \delta = 54, \lambda = -113]$  for nodal plane 1, and  $[\theta = 336, \delta = 42, \lambda = -62]$  for nodal plane 2. In this case the MRA is  $\Phi \approx 21^\circ$  for both nodal planes. The threshold of  $\Phi \leq 30^\circ$

was also used in Triantafyllis (2014), where the authors mention that an acceptable agreement between FM solutions corresponds to MRA values of few tens of degrees. Nevertheless, we remark that the lowest MRA between the test event and the neighbors is not always provided by the closest event ( $k_1$ ). In fact, we observe that other neighbors, including the median event, may provide a better solution in terms of MRA.

Among the advantages of our method, it relies on catalog information about past events. This is in contrast to more traditional methods based on waveform analysis, hence our method is significantly faster. Among its disadvantages, with respect to classical CMT inversion, we can only estimate CMT from large events, which are more likely to behave similarly to other similar events in the catalog. For smaller events our approach loses precision. Moreover, non-DC components are ignored in our study, which is not very relevant for large earthquakes but definitely relevant for smaller-event catalogs. In addition, we acknowledge that the method relies in similarity among large earthquakes recorded in a particular region. If events are few or too dissimilar, our method would fail. Nevertheless, the predictive capacity of our method can be analyzed a priori for such a region and thus we can establish the method's suitability for the region beforehand.

As a conclusion, our method can provide with suitable CMT estimates, with statistically relevant accuracy, shortly after a large event is recorded. This enables the possibility or urgent computing of seismic hazard by means of physical simulations where directivity may be a significant component. We have tested our method in several regional contexts and analyzed its accuracy. The precision that can be attained depends strongly on the statistical distribution of large events in our study region's catalog but produces a starting point for analysis prior to the latter determination of the CMT by means of inversion, which supersedes our result.

In future works we will study the impact of CMT accuracy in ground motion values and intensity measures such as peak ground accelerations resulting from simulations. Similarly we will analyze means to improve the method's predictive capabilities by narrowing the choice of final CMT candidates.

## REFERENCES

- Aki, K., and Richards, P. G. (2002). *Quantitative Seismology the City*. 2nd ed. Sausalito, CA: University Science Books.
- Altunel, E., and Pinar, A. (2021). Tectonic Implications of the Mw 6.8, 30 October 2020 Kuşadası Gulf Earthquake in the Frame of Active Faults of Western Turkey. *Turkish J. Earth Sci.* 30, 436–448. doi:10.3906/yer-2011-6
- Álvarez-Gómez, J. A. (2019). FMC-earthquake Focal Mechanisms Data Management, Cluster and Classification. *SoftwareX* 9, 299–307. doi:10.1016/j.softx.2019.03.008
- Baker, J. W. (2008). An Introduction to Probabilistic Seismic hazard Analysis (Psha). *White paper*, 72. version 1.
- Duputel, Z., Rivera, L., Kanamori, H., and Hayes, G. (2012). W Phase Source Inversion for Moderate to Large Earthquakes (1990–2010). *Geophys. J. Int.* 189, 1125–1147. doi:10.1111/j.1365-246x.2012.05419.x
- Dziwowski, A. M., Chou, T.-A., and Woodhouse, J. H. (1981). Determination of Earthquake Source Parameters from Waveform Data for Studies of Global and

## 5.1 Permission to Reuse and Copyright

Figures, tables, and images will be published under a Creative Commons CC-BY licence and permission must be obtained for use of copyrighted material from other sources (including re-published/adapted/modified/partial figures and images from the internet). It is the responsibility of the authors to acquire the licenses, to follow any citation instructions requested by third-party rights holders, and cover any supplementary charges.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://ds.iris.edu/spud/momenttensor>.

## AUTHOR CONTRIBUTIONS

MM-V, JC-J, AF, and JP, contributed to conception and design of the study. MM-V, OR, and JR wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 823844, the Center of Excellence for Exascale in Solid Earth (ChEESE CoE Project).

## ACKNOWLEDGMENTS

MM-V, JR, OR, AF, JP, thank the ChEESE CoE Project.

Regional Seismicity. *J. Geophys. Res.* 86, 2825–2852. doi:10.1029/jb086ib04p02825

Dziwowski, A. M., and Woodhouse, J. H. (1983). An experiment in Systematic Study of Global Seismicity: Centroid-Moment Tensor Solutions for 201 Moderate and Large Earthquakes of 1981. *J. Geophys. Res.* 88, 3247–3271. doi:10.1029/jb088ib04p03247

Ekström, G., Nettles, M., and Dziwowski, A. M. (2012). The Global CMT Project 2004–2010: Centroid-Moment Tensors for 13,017 Earthquakes. *Phys. Earth Planet. Interiors* 200–201, 1–9. doi:10.1016/j.pepi.2012.04.002

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Kdd* 96, 226–231.

Frohlich, C. (1996). Cliff's Nodes Concerning Plotting Nodal Lines for P, Sh and Sv. *Seismological Res. Lett.* 67, 16–24. doi:10.1785/gssrl.67.1.16

Kagan, Y. Y. (2007). Simplified Algorithms for Calculating Double-Couple Rotation. *Geophys. J. Int.* 171, 411–418. doi:10.1111/j.1365-246x.2007.03538.x

Kaverina, A. N., Lander, A. V., and Prozorov, A. G. (1996). Global Creepex Distribution and its Relation to Earthquake-Source Geometry and Tectonic Origin. *Geophys. J. Int.* 125, 249–265. doi:10.1111/j.1365-246x.1996.tb06549.x

- Lin, J.-T., Chang, W.-L., Melgar, D., Thomas, A., and Chiu, C.-Y. (2019). Quick Determination of Earthquake Source Parameters from Gps Measurements: a Study of Suitability for Taiwan. *Geophys. J. Int.* 219, 1148–1162. doi:10.1093/gji/ggz359
- Maeda, N. (1992). A Method of Determining Focal Mechanisms and Quantifying the Uncertainty of the Determined Focal Mechanisms for Microearthquakes. *Bull. Seismological Soc. America* 82, 2410–2429.
- Melgar, D., Bock, Y., and Crowell, B. W. (2012). Real-time Centroid Moment Tensor Determination for Large Earthquakes from Local and Regional Displacement Records. *Geophys. J. Int.* 188, 703–718. doi:10.1111/j.1365-246x.2011.05297.x
- Mulgaria, F., Stark, P. B., and Geller, R. J. (2017). Why Is Probabilistic Seismic Hazard Analysis (Psha) Still Used? *Phys. Earth Planet. Interiors* 264, 63–75. doi:10.1016/j.pepi.2016.12.002
- Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In 2011 31st international conference on distributed computing systems workshops. IEEE, 166–171. doi:10.1109/icdcs.2011.20
- Scognamiglio, L., Tinti, E., Michelini, A., Dreger, D. S., Cirella, A., Cocco, M., et al. (2010). Fast Determination of Moment Tensors and Rupture History: What Has Been Learned from the 6 April 2009 L’Aquila Earthquake Sequence. *Seismological Res. Lett.* 81, 892–906. doi:10.1785/gssrl.81.6.892
- Sinnott, R. W. (1984). Virtues of the Haversine. *Sky Telesc* 68, 159.
- Sokos, E. N., and Zahradnik, J. (2008). Isola a Fortran Code and a Matlab Gui to Perform Multiple-point Source Inversion of Seismic Data. *Comput. Geosciences* 34, 967–977. doi:10.1016/j.jgeo.2007.07.005
- Stierle, E. (2015). *Non-Double-Couple Components in Moment Tensors of Aftershock Seismicity and Laboratory Earthquakes*. Ph.D. thesis.
- Tarantino, S., Colombelli, S., Emolo, A., and Zollo, A. (2019). Quick Determination of the Earthquake Focal Mechanism from the Azimuthal Variation of the Initial P-Wave Amplitude. *Seismological Res. Lett.*
- Trabant, C., Hutko, A. R., Bahavar, M., Karstens, R., Ahern, T., and Aster, R. (2012). Data Products at the Iris Dmc: Stepping Stones for Research and Other Applications. *Seismological Res. Lett.* 83, 846–854. doi:10.1785/0220120032
- Triantafyllis, N. (2014). *Scisola: Automatic Moment Tensor Solution for SeisComp3*. Master’s thesis. Patras: School of Engineering, Computer Engineering Informatics Departme.
- Triantafyllis, N., Sokos, E., and Ilias, A. (2013). Automatic Moment Tensor Determination for the Hellenic Unified Seismic Network. *Bull. Geol. Soc. Greece* 47, 1308–1315.
- Triantafyllis, N., Sokos, E., and Ilias, A. (2016). Scisola: Real-Time Moment Tensor Monitoring for Seiscomp3. *Bull. Geol. Soc. Greece* 50, 1120–1129.
- Triantafyllis, N., Venetis, I., Fountoulakis, I., Pikoulis, E.-V., Sokos, E., and Evangelidis, C. (2021). *Gisola: Real-Time Moment Tensor Computation Optimized for Multicore and Manycore Architectures*. Tech. Rep. Copernicus Meetings.
- Udias, A., and Buforn, E. (2017). *Principles of Seismology*. Cambridge University Press.
- Vannucci, G., Pondrelli, S., Argnani, A., Morelli, A., Gasperini, P., and Boschi, E. (2004). An Atlas of Mediterranean Seismicity. *Ann. Geophys.* 47.
- Wald, D. J., Quitoriano, V., Heaton, T. H., Kanamori, H., Scrivner, C. W., and Worden, C. B. (1999). Trinet “Shakemaps”: Rapid Generation of Peak Ground Motion and Intensity Maps for Earthquakes in Southern California. *Earthquake Spectra* 15, 537–555. doi:10.1193/1.1586057
- Wald, D., Lin, K.-W., Porter, K., and Turner, L. (2008). Shakecast: Automating and Improving the Use of Shakemap for post-earthquake Decision-Making and Response. *Earthquake Spectra* 24, 533–553. doi:10.1193/1.2923924
- Wang, R., Gu, Y. J., Schultz, R., and Chen, Y. (2018). Faults and Non-double-couple Components for Induced Earthquakes. *Geophys. Res. Lett.* 45, 8966–8975. doi:10.1029/2018gl079027
- Weber, B., Becker, J., Hanka, W., Heinloo, A., Hoffmann, M., Kraft, T., et al. (2007). Seiscomp3—automatic and Interactive Real Time Data Processing. *Geophys. Res. Abstr.* 9, 219.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Monterrubio-Velasco, Carrasco-Jimenez, Rojas, Rodríguez, Fichtner and Puente. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.