## 8

#### **Research Article**

Sayed S. R. Moustafa, Gad-Elkareem A. Mohamed, and Mohamed Metwaly\*

# Production of a homogeneous seismic catalog based on machine learning for northeast Egypt

https://doi.org/10.1515/geo-2020-0295 received August 05, 2020; accepted September 09, 2021

**Abstract:** This research presents a new approach which addresses the conversion of earthquake magnitude as a supervised machine-learning problem through a multistage approach. First, the moment magnitude  $(M_w)$  calculations were extended to lower magnitude earthquakes using the spectral P-wave analyses of the vertical component seismograms to improve the scaling relation of  $M_{\rm w}$  and the local magnitude ( $M_{\rm L}$ ) of 138 earthquakes in northeastern Egypt. Second, using unsupervised clustering and regression analysis, we applied the k-means clustering technique to subdivide the mapped area into multiple seismic activity zones. This clustering phase created five spatially close seismic areas for training regression algorithms. Supervised regression analysis of each seismic area was simpler and more accurate. Conversion relations between  $M_{\rm w}$  and  $M_{\rm L}$  were calculated by linear regression, general orthogonal regression (GOR), and random sample consensus (RANSAC) regression techniques. RANSAC and GOR produced better results than linear regression, which provides evidence for the effects of outliers on regression accuracy. Moreover, the overall multistage hybrid approach produced substantial improvements in the measured-predicted dataset residuals when individual seismic zones rather than all datasets were considered. In 90% of the analyzed cases,  $M_{\rm w}$  values could be regarded as  $M_{\rm L}$  values within 0.2 magnitude units. Moreover, predicted magnitude conversion relations in the current study corresponded well to magnitude conversion relations in other seismogenic areas of Egypt.

**Keywords:** regression, source parameters, RANSAC, Earthquake magnitude

#### 1 Introduction

Earthquakes are natural disasters that significantly impact culture, politics, and health in the regions affected by them [1]. The magnitude of an earthquake, which is the energy that an earthquake releases [2], is a crucial variable for assessing the human impact of an earthquake [3]. Largemagnitude earthquakes can cause considerable damage [1].

Located in the northeastern region of the African continent, Egypt is bound by three major tectonic elements (see panel (a) in Figure 1): the African-Eurasian plate margin, Red Sea plate margin, and Levant transforming fault. Tectonic distortion is related to contact and relative movement of the Egypt's crust below Egypt's territories and its remote effects [4,5]. Egypt's seismicity is low to moderate relative to that of North-West Africa, and much more so relative to the extreme seismicity of neighboring Hellenic and Cyprus subduction arcs and the Dead Sea transformation zone (Figure 1). Local moderate earthquakes, however, pose a major threat to the population, as demonstrated by the 1992 case in Cairo that caused 561 fatalities and by the historical 1847 event [5]. Regarding the transfensional stress regime triggered by the expanding Red Sea margin, most earthquakes are characterized by a normal fault with variable strike-slip components, which increases towards the edge of the Sinai subplate. A few cases, which are mainly inland, have reverse focal mechanisms [4].

Earthquake catalogs contain data on the time of occurrence, hypocentric coordinates, magnitude of the event, and other related information about the earthquake [6]. These details are extracted from observations of various types of seismic waves, usually from heterogeneous seismological networks in the space-time domain. The data obtained from seismic catalogs provide valuable information for the study of seismotectonics and seismic hazard distribution in space and time [6]. However, accurate

<sup>\*</sup> Corresponding author: Mohamed Metwaly, Archaeology
Department, College of Tourism and Archaeology, King Saud
University, Saudi Arabia; Geomagnetic and Geoelectric Department,
National Research Institute of Astronomy and Geophysics (NRIAG),
Helwan, Cairo, Egypt, e-mail: mmetwaly@ksu.edu.sa
Sayed S. R. Moustafa, Gad-Elkareem A. Mohamed: Seismology
Department, National Research Institute of Astronomy and
Geophysics (NRIAG), Helwan, Cairo, Egypt

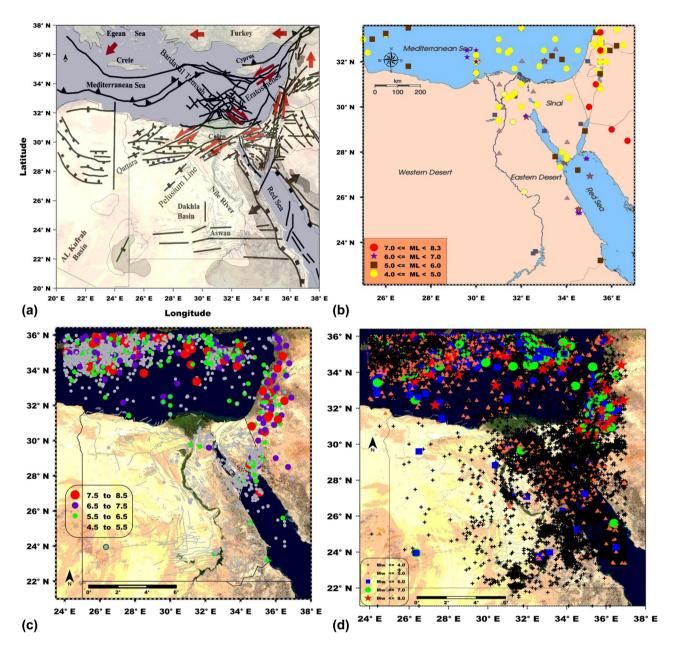


Figure 1: (a) Regional tectonic setting of Egypt. (b) Historical seismicity of Egypt based on the compiled earthquake catalog of ref. [53]. (c) Significant historical and instrumental events affecting the area (adopted from ref. [27]). (d) Instrumental seismicity of Egypt based on the compiled ENSN catalog.

results require the use of a homogenized seismicity catalog. A crucial point in the process of homogenization is the identification of all possible earthquake occurrences by using a single target magnitude scale [7].

Due to the occurrence of the catastrophic earthquake, in Cairo in October 1992, the Egyptian National Seismological Network (ENSN) project was implemented with the aim to cover and report the entire Egyptian territory (Figure 1) seismic activities. The ENSN started operating in August 1997, and since then, accurate parameters of relevant seismic

activity have been determined. Despite the establishment of ENSN, rapid changes in the method used for estimating the magnitude of seismic incidents, such as seismometers, and data reduction procedures have caused variances in the reported magnitude; thus, a consistent magnitude scale is not always available [8]. There have been several scales of magnitude utilized, and each is specified in terms of amplitudes measured with a specific type of seismometer over a very narrow spectral band; thus, variable measurement errors are manifested. For this reason, different types of

magnitudes have been reported for the same earthquake, such as local magnitude [9], surface-wave magnitude [10], body-wave magnitude [11], and moment magnitude [12]. This will have tremendous effects on the precision and homogeneity of seismic hazard, as well as on the calculation of seismic deformation [13,14]. Various empirical relations have been utilized to overcome that problem. Different types of magnitudes are transferred and unified to the same scale, prior to inclusion in the catalog for seismicity and seismic hazard studies. The majority of researchers prefer to utilize the moment magnitude  $(M_w)$ , as it is no suffer saturation for earthquakes of higher magnitude, as opposed to other types of magnitude. This is because the size of an earthquake is related to the ruptured field, average slip volume, and force necessary to overcome friction, which hold the rock together from offset by failures that are directly derived from the seismic moment  $(M_0)$  [12].

Traditionally, magnitude conversation relationships are utilized to produce homogenized earthquake catalogs that implement several regression procedures, including simple linear regression (SLR), inverted standard regression (ISR), orthogonal regression (OR), and general orthogonal regression (GOR), to produce various magnitude conversion relations [15-21]. SLR assumes that either the independent variable is error-free, or its error order is significantly small in comparison with the dependent variable measurement error. In this regression type, vertical residues are minimized. By comparison, as the locations of dependent and independent variables are reversed, ISR is accomplished by reducing horizontal offsets to the bestfit rows [22]. A comparison among regression methods applied to earthquake magnitude conversions is given in refs [23,24].

Empirical conversions between different types of magnitude are typically utilized in Egyptian catalogs to homogenize the data for statistical analysis and hazard assessment. The standard least-square linear regression method is the most common method used to determine the relationship between different magnitude scales [25–27]. These empirical relationships may be improved if we understand the spatial clustering of earthquakes, as earthquakes are known to cluster in space and time due to stress redistribution in the Earth's crust [28]. Hence, current research presents a new clustered linear regression approach, which is utilized to overcome the shortages of multiple regression. The methodology we describe in this paper uses a multistage machine-learning approach to deal with traditional magnitude conversions methods' limitations. Typical multistage learners combine two or more inferential approaches so that more complex learning problems can be solved. Our approach is a

hybrid combination of unsupervised and supervised machine-learning methods. The three cluster algorithms involved in most studies are the partitional algorithm, hierarchical algorithm, and density-based algorithm [29,30]. Because of its applicability and reliability, the partitional algorithm is most commonly used. As with most common cluster techniques, *k*-means (hard clustering) and Fuzzy *C*-means (soft clustering) are categorized in terms of the partitional algorithm. *k*-means vary from Fuzzy *C*-means in that a point belongs only to one cluster in *k*-means, whereas the point belongs to a cluster to some extent depending on its Fuzzy Membership, in Fuzzy *C*-means.

In this research, *k*-means was utilized as a convenience to prevent the additional computational costs that are incurred by fuzzy inclusion in fuzzy *C*-means. *k*-means clustering is used before the application of a multiple regression method which is a supervised analysis procedure that can be described as a model of relationships between a target continuous variable and a collection of other input variables. This multistage approach has been adopted with the objective to improve the accuracy of learned models.

## 2 Methods and data

This section explains the proposed multistage seismological and machine-learning approaches that are used to predict the conversion relation between reported local magnitude  $M_{\rm L}$  and estimated moment magnitude  $M_{\rm w}$  with the intention of enhancing our learned models and predictive accuracy. In the first phase of our strategy, we acquired the frequency-independent, long-period spectral level  $\Omega_0$  below the corner frequency of a displacement spectrum [31,32]. The long-period spectral amplitude of the source spectrum can be related to seismic moment  $M_0$  [31] from which moment magnitude  $M_{\rm w}$  can be computed [12].

For seismic moment estimation, the earthquake signals from the three component seismometers were inverted by assuming a Brune's model for the source and connecting it to a rupture area with a slip value. According to Brune's model [31,32], the displacement spectrum specifies that the frequency response of the seismic signal is considered to be flat until it reaches the corner frequency  $\omega$ ; then, the amplitude of the displacement spectrum drops to  $\omega^{-2}$ . By fitting such a model to the displacement spectrum, the low-frequency plateau W can be estimated and used directly in the calculation of the seismic moment because it relates to the area of seismic slip on the fracture

plane by adapting the model to the displacement spectrum. Geometric spreading R, seismic wave velocity c, rock density  $\rho$ , and a radiation pattern imposed by the seismic source moment tensor  $F_c$  were included in the calculation to estimate the seismic moment as given in the following equation:

$$M_0 = \frac{4\pi\rho c^3 W_c R}{F_c} = \mu dA \tag{1}$$

where u is the rock's shear modulus, d is the average fault displacement, and A is the fault area.

For P-waves  $W_c = W_p$ . According to ref. [33], F = 0.52for P-waves could be adopted as an approximate value for  $F_c$  if the radiation pattern of the moment tensor is not known. With seismic moment, measured in Nm, the moment magnitude is estimated following the ref. [12] formula as:

$$M_{\rm w} = \frac{2}{3} \log_{10} M_0 - 6.0 \tag{2}$$

In the second step of our strategy, unsupervised machine-learning data clustering utilizing k-means algorithm is carried out to find a data partition such that points in the same subgroup are close to the similar Euclidean distance  $(d(x, y) = ||x - y||_2)$  between two earthquakes x and y [34]. The k-means algorithm is a heuristic technique that often produces effective clustering outcomes. In that algorithm, the global minimizing partition S\* is usually intractable because the group of partitions is very large. The algorithm specifies an initial cluster mean  $\mu^{(1)}$  for each i = 1,..., N. Then, given a set of cluster means (t), the k-means algorithm searches for a partition S(t) of the earthquake epicenters such that [34]:

$$S^{(t)} = \{x_i : ||x_i - \mu^{(t)}||^2 \le ||x_i - \mu^{(t)}||^2, \quad l = 1, ..., N\}$$
 (3)

The algorithm then updates the group of epicenters for each i = 1, N and continues to iterate until the partition ceases shifts [35]. The within-cluster sum of squares (WCSS) is used as an indicator of the performance of clustering efficiency (lower value for WCSS is better).

For the final phase of our strategy, linear regressions between the local magnitude and moment magnitude ranges will be performed. Using the most consistent linear relationship, we convert the local magnitude  $M_{\rm L}$ into moment magnitude  $M_{\rm w}$ , for each available pair of magnitudes. In this approach, the relationship between the predictor variable and the response variable with SLR can be described as a one-to-one relationship, while multiple linear regression is a many-to-one relationship. For multiple linear regression, the relationship between Xand *Y* is expressed as:

$$Y \approx \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k$$
 (4)

Similar to SLR, the regression coefficients are estimated using the least-square approach so that  $\beta_0$ ,  $\beta_1$ , ...,  $\beta_p$  is selected to minimize the residual sum of squares (RSS). The residual is the difference between the expected magnitude value and the true magnitude value;  $y - y^2$ ,  $y^2$  can be replaced with  $\beta_0 + \beta_1 x$  and the residual sum of the squares can then be defined as:

RSS = 
$$(y_1 - (\hat{\beta}_0 + \hat{\beta}_1 X_1))^2 + (y_2 - (\hat{\beta}_0 + \hat{\beta}_1 X_2))^2 + \dots + (y_n - (\hat{\beta}_0 + \hat{\beta}_1 X_n))^2$$
 (5)

We used earthquake data collected from the ENSN to test the efficiency of our proposed method and to establish a reliable relation between the reported local magnitude  $(M_L)$  and the moment magnitude  $(M_w)$ . To estimate  $M_{\rm w}$  for the region between 27-31°N and 31-36°E (see panel (a) in Figure 2), displacement spectra of 138 earthguakes that occurred from 1998 and 2020 were used. Such events occurred in various seismic areas of Egypt, including the Red Sea, the Agaba, and Suez Gulfs, and localized seismicity in the northern part of the Eastern Desert. In addition, only stations with epicentral distances greater than 10 km were considered in order to eliminate near-field effects and obtain a point-source approximation [36]. The earthquake's hypocentral parameters were taken from the annual ENSN bulletin. Selected waveforms were captured by broadband stations with good events-stations azimuth coverage and display a higher signal-to-noise ratio and clear P-wave phase. Lowpass filters with corner frequencies of 35 Hz were used to prevent an aliasing effect. The analyzed seismograms were corrected using an instrumental response for all stations. In addition, the local site effect was ignored since all stations were in hard rock. The source parameters were estimated by utilizing the Fourier spectra of the pure P-phase displacement spectra of the integrated velocity data and after applying instrumental and baseline corrections. Before marking the corner frequency and long-period spectral level, attenuation correction was applied, assuming the well-accepted Q-value for region as suggested by ref. [37].

Next, the source term of displacement was processed based on the refs [31,32] model, which utilizes circular faults for small seismic events. Rock densities and Pwave velocities were taken from the ref. [38] velocity model. Following ref. [39] correction of the radiation pattern for P-waves, a 0.52 value was utilized. Following the procedure of ref. [40], the corner frequency  $(f_c)$  and spectral level  $(\Omega_0)$  were determined for each P-wave source spectra and used to calculate the seismic moment  $(M_0)$ by fitting each P-wave displacement spectrum to the

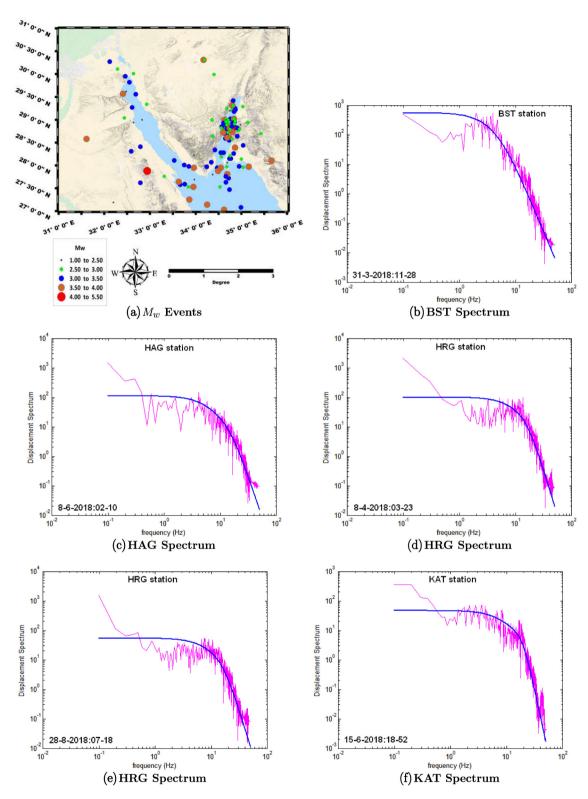


Figure 2: (a) Selected events utilized to estimate  $M_w$ . (b-f) Examples of earthquake spectra for some ENSN stations.

theoretical displacement spectra model as expressed by equation (1). The nonlinear Levenberg–Marquard fitting algorithm [41] was used to minimize the discrepancy

from the theoretical to the observed range of displacement spectra. The approach will match the spectra effectively without having to compensate for the loss of the

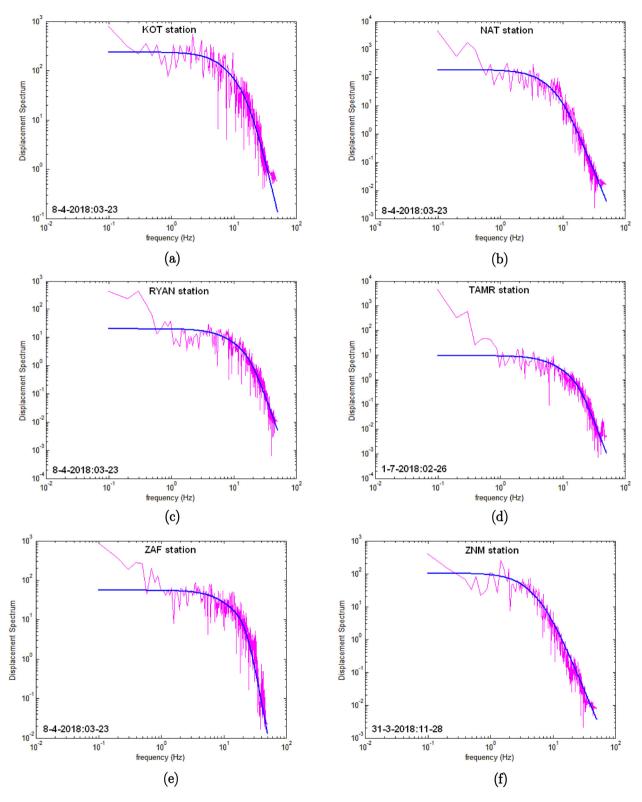


Figure 3: (a-f) Another examples of earthquake spectra for some ENSN stations. (a) KOT spectrum, (b) NAT spectrum, (c) RYAN spectrum, (d) TAMR spectrum, (e) ZAF spectrum, and (f) ZNM spectrum.

Figures 2 and 3, the best matches for the observed number of stations are plotted. Following the definition

frequency of the corner due to the flat component. In displacement spectra to the theoretical spectra at a

Table 1: Descriptive statist	ics of all machine-	learning attributes
------------------------------	---------------------	---------------------

Variables	Count	min	25%	50%	75%	max
Year	138	1998	2002.25	2013.5	2018	2020
Latitude °N	138	27.0	27.51188	27.91885	28.77	31.0
Longitude °E	138	31.0	32.86065	34.1091	36.0	35.34
Depth (km)	138	2.0	8.14	14.675	19.7225	30.35
$M_{L}$	138	2.81	3.1725	3.9	4.2	6.12
$M_{\rm w}$	138	1.94	2.5985	2.851	3.17	4.86

given in equation (6), the average seismic moment  $< M_0 >$  was calculated from the average logarithmic values obtained at different stations as proposed by ref. [42].

$$< M_0 > = \operatorname{antilog} \frac{1}{N} \sum_{i=1}^{N} \log M_{0_i}$$
 (6)

where N is the number of stations used, and  $M_0$  is the seismic moment determined from equation (1) for the ith event record. Then, the average moment of magnitude ( $M_w$ ) is calculated using equation (2) for that event. Table 1 shows all variables in the description of the analysis.

Most studies developed scaling relationships between various magnitudes (see for example, refs [25,26]), but did not consider the spatial distribution of the events that were utilized in the analysis. This sometimes causes large fitting errors. Therefore, we need to cluster these earthquake data into active seismic regions as outlined in the second step of our proposed approach before establishing the scaling relationships. In this step, every earthquake epicenter, with coordinates being their latitude and longitude, is considered to be a point source in  $\mathbb{R}^2$ . However, since the Earth is not flat, that assumption is not accurate. It is important to view the latitude and longitude values in  $\mathbb{R}^3$  as a modification of spherical coordinates.

Based on the Euclidean distance of our k-means algorithm, we need to transform our database into a 3D Euclidean coordinate. A convenient way to achieve this transformation is to convert latitude and longitude values into spherical coordinates, then into Euclidean coordinates. Recall that a spherical coordinate in  $R^3$  is a triple  $(r, \theta, \phi)$ , where r is the distance from the origin,  $\theta$  is the radial angle in the xy-plane from the x-axis, and  $\phi$  is the angle from the z-axis. In our earthquake data, the longitude is already the appropriate  $\theta$  value and the  $\phi$  value (in degrees) is simply 90° minus the latitude.

Based on the data shown in Figure 2(a) and Table 1, the spatial distribution of the data utilized to estimate  $M_{\rm w}$  is sparse and does not have enough events to obtain good clustering results. To solve the problem, we incorporated the data published in refs [17,18,43,44] in our database

for utilization in the second and third phases of our hybrid clustered-regression approach. Spatial distribution of our collected data and incorporated data is given in Figure 4. A manual check for duplicates was performed. Table 2 provides descriptive statistics for our final database attributes.

The application of the k-means algorithm requires the number of clusters k to be known. Since our seismic data are an unlabeled dataset, finding the optimal value of k is a challenge. In this case, it is not certain how many clusters are required. Therefore, we used a heuristic approach called the Elbow method [18] to identify the optimal k value from the relation between the number of clusters k and the WCSS. For the elbow method, the distance from each earthquake data point and the cluster centroid to which it was assigned is determined by

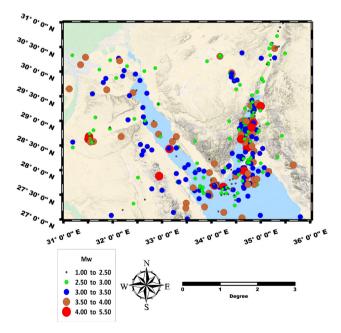
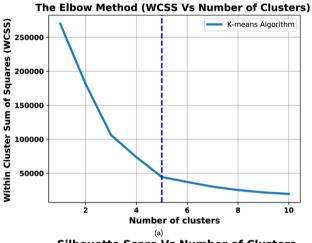


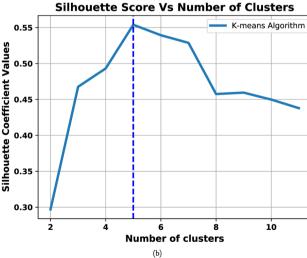
Figure 4: Final spatial distribution of seismic events utilized for machine-learning clustering and regression analyses.  $M_{\rm w}$  data are incorporated from the current study and the data published in refs [4,26,44,45].

Variables	Count	Mean	min	25%	50%	75%	max
Year	390	2005.428	1998	2001	2003	2010.75	2020
Month	390	7.210256	1	4	8	11	12
Day	390	16.66923	1	10	17	24	31
Hour	390	10.61539	0	5	10	16	23
Minute	390	28.98205	0	13.25	29	43.75	59
Second	362	28.9863	0	13.125	29.055	45.4	59.7
Latitude °N	390	28.62951	27.01	27.81	28.59	29.2	30.98
Longitude °E	390	34.13487	31.08	33.8325	34.55	34.8	35.8
Depth (km)	379	10.40557	2	5	9	15.2	35
ML	390	2.969592	1.45	2.4375	2.89	3.285	6.12
$M_{\rm w}$	390	2.984205	1.07	2.7	3	3.3	5.5

Table 2: Descriptive statistics of the final machine-learning attributes

the within-cluster sum-of-squares (inertia) using the following equation:





**Figure 5:** Comparison between WCSS (a) and Silhouette score (b) methods for finding the optimal number of clusters k-value for k-means algorithm.

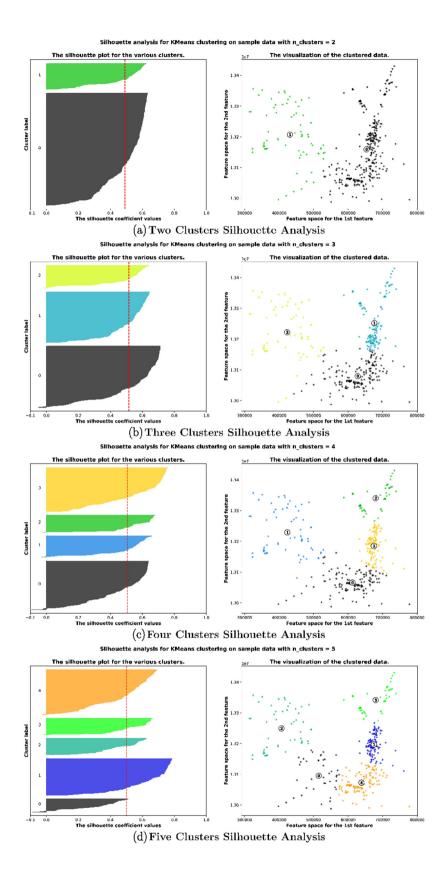
$$WCSS = \sum_{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2$$
 (7)

where  $C_k$  represents cluster k,  $x_i$  is a data point,  $\mu_k$  is the centroid for cluster k.

A plot of *k* vs the sum of squares reveals an elbow point in which the inertia starts descending much more slowly and can be selected as the best value for k. Figure 5(a) displays the approximate value of the sum of the k-means squared distance for different values of k. In the plot we can see a strong elbow at k = 5, which is the best value for k. However, the elbow method is limited. since the solution is heuristic, so we used the Silhouette method [18]. That method measures how well each datapoint  $x_i$  "fits" into its assigned cluster and how poorly it fits into other clusters. This is a different way of looking at the same objective. Denote  $a_{xi}$  as the average distance from  $x_i$ to all other points within its own cluster k; the lower the value, the better. On the other hand,  $b_{xi}$  is the minimum average distance from  $x_i$  to points in a different cluster, which has been minimized for all clusters. That is, compute separately for each cluster the average distance from  $x_i$  to the points within that cluster, and then take the minimum. The silhouette  $s(x_i)$  is defined as ref. [18]:

$$s(x_i) = \frac{b_{x_i} - a_{x_i}}{\max(a_{x_i} \cdot b_{x_i})}$$
(8)

Figure 5b shows the average silhouette score compared with results obtained by the elbow method for k-means clustering of the sample data, while Figure 6a and b show the results of the silhouette analysis with the corresponding scores computed for every earthquake data point in every cluster. The obtained silhouette scores depicted in Figure 6(a-d) and (e-h) confirmed the number of clusters that were obtained from the elbow method. Accordingly, we chose five as the optimal value of k to feed into the k-means clustering algorithm.



(Continued)

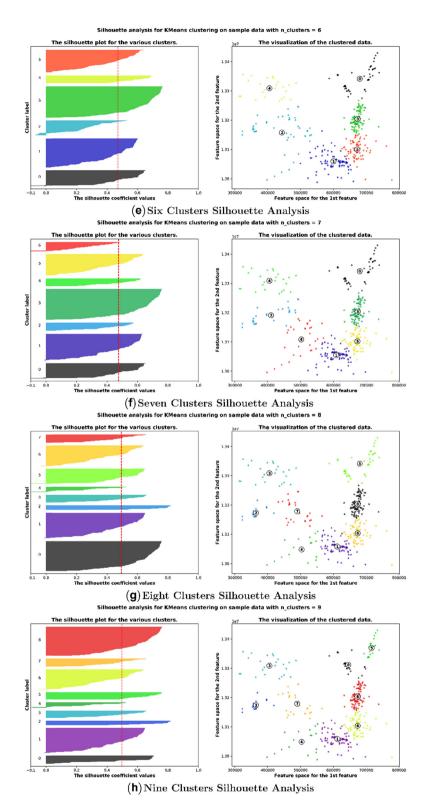


Figure 6: Silhouette analysis for k-means clustering on sample data with number of clusters ranging from two clusters to five clusters (a-d) and from six clusters to nine clusters (e-h).

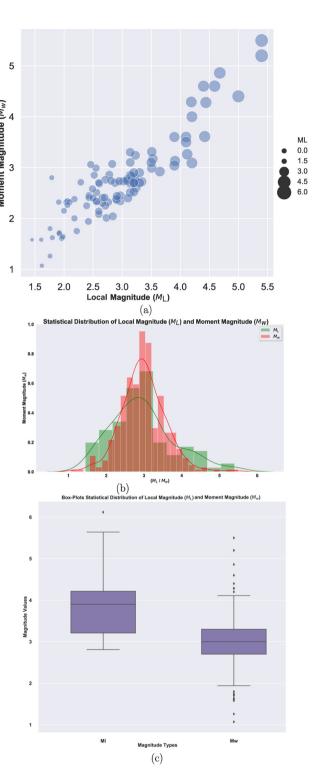
In earthquake catalogs,  $M_{\rm w}$  is not commonly considered the main magnitude, so understanding how certain magnitudes respond to  $M_{\rm w}$  in each area is a matter of considerable concern due to seismic hazard estimation. Usually regression analysis is utilized for that purpose. However, Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points [18]. Before we begin supervised machine-learning regression analysis and modeling, it is recommended that we perform an exploratory data analysis and plot the estimated and collected magnitude data, as shown in Figure 7.

By examining these initial plots, we can quickly determine that the data have a strong linear relationship as shown in Figure 7(a). From the histogram plots shown in Figure 7(b), we can clearly observe that both local and moment magnitudes are normally distributed. An example of this pattern is that the data are unimodal, or the data have a single mode, which is marked by the curve's peak. On the other hand, the local magnitude values are skewed to the right, as the histogram has a middle peak that steadily tapers toward the right side of the graph. This dataset is unimodal, as the mode closest to the left of the graph is smaller than the mean or the median. The mean for the local magnitude data is located on the right side of the diagram and is higher than the median or the norm. This graph shows that there are few data points that are larger than the mode, which may be representative of outliers in our database. Finally, boxplots are used to test the presence of outliers in our data. Box-plots (Figure 7c) are used to show the overall response patterns for each magnitude type. They provide a useful way to visualize the magnitude value ranges.

Figure 7(c) shows a variety of different box plot shapes and positions were obtained for each magnitude type. It is clear that both  $M_{\rm L}$  and  $M_{\rm w}$  suffer from the existence of outliers. The observed outliers exerted a large influence on the overall outcome of the regression model parameter's estimate.

## 3 Results and discussions

Through an analysis of earthquake source parameters, local tectonics and basic earthquake mechanics can be understood. To improve scaling between moment magnitude and local magnitude for earthquakes in the northeastern Egypt, the moment magnitude ( $M_{\rm w}$ ) estimates were extended to lower magnitudes by using the spectral analysis of P-waves from vertical component seismograms. This dataset has been processed, analyzed, and



**Figure 7:** Comparison of the statistical distributions of local magnitude  $(M_L)$  with the moment magnitude  $(M_w)$  in terms of scatter plot (a), histogram plot (b), and box-plot (c).

interpreted to study the source parameters of local earthquakes occurring around northeastern Egypt. Only events from at least three ENSN stations at different distances with reliable trace amplitude readings were selected. In this way, a total of 138 events with reliable  $M_{\rm L}$  values (Figure 2a) were chosen for  $M_{\rm w}$  determination. The long-period spectral amplitude of the source spectrum, which could be related to the seismic moment, was utilized to compute  $M_{\rm w}$  following the definition of ref. [16]. Figures 2 and 3 display examples of P-wave Fast Fourier Transform spectra fitted by the smoothed spectral [17] model at several ENSN stations.

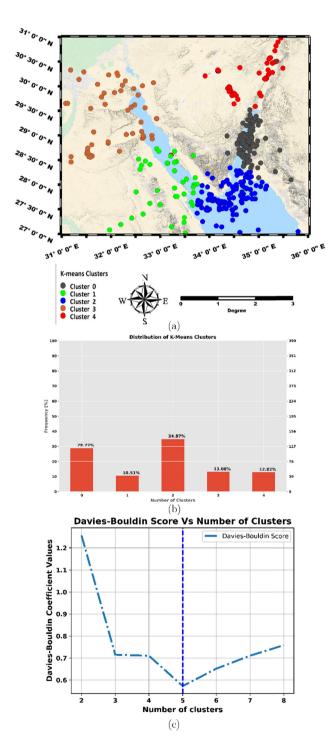
Displacement of the P-wave portion is considered to be an approximate source time function and could be integrated to obtain scalar moments at each station. Multiple records are combined to give a good estimation of the moment magnitude without correcting for the source mechanism [18]. Table 1 summarizes the average value of the obtained source parameters for small earthquakes in northeast Egypt.

Values of the  $M_{\rm w}$  calculated using the method described in this paper were compared with those obtained from automatic moment tensor solutions and published in refs [43–45]. The average deviation between the determined  $M_{\rm w}$  and that of the aforementioned studies is  $\pm 0.085$ , meaning that, on average, our moment magnitude was slightly overestimated.

After finalizing the first phase and estimating the moment magnitude  $M_{\rm w}$ , the datasets were utilized to compare the partitions with the distribution of recognized active sources and were applied to areas that may not be adequately captured by existing source models. In addition, k-means cluster analysis was applied to the hypocentral distribution of observed earthquakes in the mapped region. Since the moment magnitudes of the utilized events were lower than six ( $M_{\rm w}$  < 6.0), we can consider them to be point-source attributes. The obtained clusters comprise sets of earthquake data that are similar to other data in the same group, but dissimilar from data in other groups, according to the Euclidean distance. Based on results of the Elbow and Silhouette methods shown in Figures 5 and 6, we picked five (k = 5) to be the final number of clusters, as it was considered to be the optimum number of clusters for which the highest average silhouette score was taken from the value of the silhouette of each cluster.

Our implementation of the k-means algorithm began by assigning each seismic occurrence randomly to a cluster, and then the mean center of each cluster was determined. At this point, the Euclidean distance between each occurrence and the initial clusters was determined and reassigned to a new cluster centered on the closest mean center. Then, the mean centers were recalculated, which proceeds until the cluster elements stop changing.

Using the optimal number (*k*) as suggested by the Elbow and Silhouette score methods (Figures 5 and 6), we were able to group the earthquakes from the dataset into five different clusters. We grouped the seismic activities according to their geographic location. The latitude and



**Figure 8:** Northeastern Egypt cluster zones (a), frequency percentage derived from the *k*-means clustering algorithm (b), and Davies-Bouldin score verses number of clusters in dataset (c).

Table 3: Descriptive statistics of database attributes with respect to number of clusters

Clusters		Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Latitude	Count	112	41	136	51	50
	Min	27.95	27.01	27.03	27.76	27.33
	Max	30.50	28.80	30.31	30.33	30.98
Longitude	Min	32.14	31.97	32.34	31.08	32.64
	Max	35.43	34.96	35.80	35.00	35.49
Year	Min	1999	1999	1998	1998	1999
	Max	2019	2019	2019	2020	2019
Month	Min	1	1	1	1	1
	Max	12	12	12	12	12
Day	Min	1	1	1	2	1
	Max	31	31	31	30	31
Depth	Min	2.00	3.00	2.00	2.00	2.00
	Max	25.00	35.00	29.90	28.40	20.00
$M_{L}$	Min	1.79	2.43	1.45	2.42	2.32
	Max	5.40	4.44	4.59	5.00	4.00
$M_{\rm w}$	Min	2.10	2.25	1.07	1.94	2.00
	Max	5.50	4.28	4.60	4.40	3.90

longitude values were converted into Universal Transverse Mercator coordinates for correct estimation of the Euclidean distance and cluster centers or cluster centroid identification. Distribution of delineated clusters is shown in Figure 8(a) and the frequency of each cluster is given in Figure 8(b).

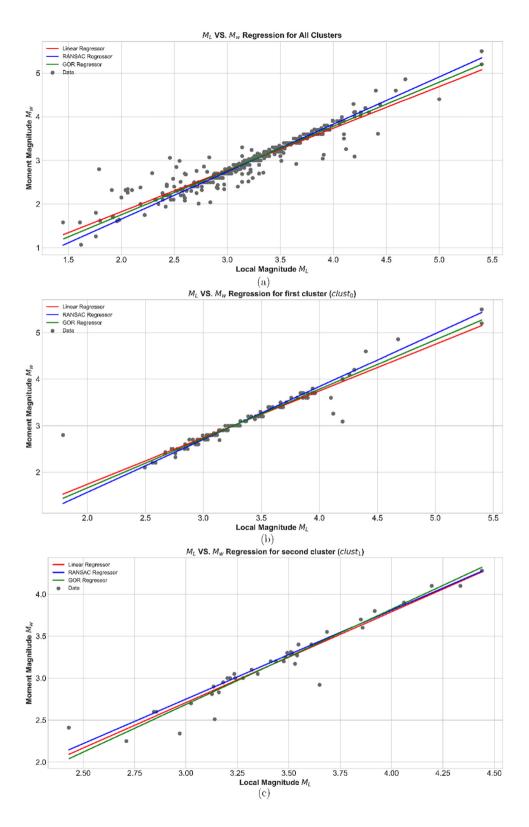
The Davies-Bouldin index [46] was used to test the clustering performance and to evaluate the k-means model, where a lower Davies-Bouldin (DB) index was correlated to a model with a better separation between the clusters. This index signifies the average "similarity" between clusters, where the similarity is a measure that compares the distance between clusters with the size of the clusters. As shown in Figure 8(c), the lowest (DB) index obtained corresponded to the five clusters, which confirmed the appropriateness of the utilized value of number of clusters for the k-means model and indicated a better partition of seismic activity zones affecting the mapped area.

The obtained clusters are in accord with various seismotectonic studies of Egypt (see for example, refs [27,38,47,48]). The clusters that were found may be linked to three main seismotectonic areas. The first region is the *Gulf Suez*, which comprises the second (*cluster*<sub>1</sub>) and fourth (*cluster*<sub>3</sub>) clusters. The area is the location of the March 31, 1969, Shadwan Island event with a surface-wave magnitude  $M_{\rm S}=6.8$  in the mouth of the Gulf of Suez. The majority of the focal mechanism solutions of reported events indicate that the current tectonic activity had a normal fault mechanism in this region. Events of this region represent 23% of our database. The second region,

which is known as the Gulf Aqaba, is an area of extreme activity situated along the main boundaries of the tectonic plate. Its major earthquakes were triggered by the displacement of this transform boundary. In this area, the largest recorded earthquake ( $M_{\rm w}=7.2$ ) occurred on November 22, 1995. This area's events represent 42% of the data in our database, and they were part of the first (*cluster*<sub>0</sub>) and fifth clusters (*cluster*<sub>4</sub>). The last area contains the largest percentage of events in our database and is known as the Northern Red Sea source region. The fault plane solutions have a normal fault mechanism with a minor strike-slip component that is compatible with the Gulf's extensional rhomb-shaped grabens system. The number of seismic activities available in each region varies considerably. The minimum number is observed in the second clustered zone with 41 delineated events, while the maximum number of 136 delineated events was obtained for the fourth zone. Table 3 provides concise statistics for each cluster.

The last stage of our multistrategy approach consists of a supervised learning process following previous clustering-based partitioning of the magnitude data. The  $M_{\rm w}$  versus  $M_{\rm L}$  relationships for the delineated five cluster zones were determined.

The most common approach to perform homogenization of different magnitude scales into a single magnitude scale and to compare various measures of magnitude is through a regression analysis that utilizes the least-squares method [49]. However, its basic assumptions are rarely satisfied in practice [15]. Hence, refs [15,22] recommend



(Continued)

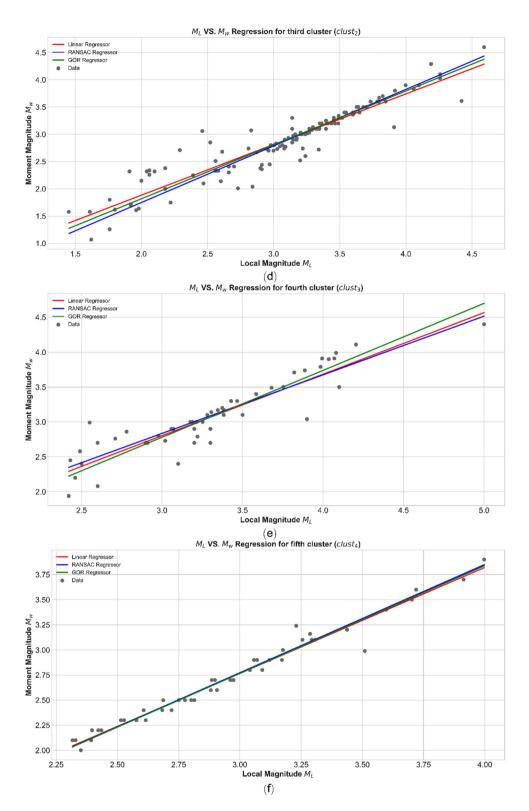


Figure 9: (a)  $M_{\rm w}$   $M_{\rm L}$  relationships for northeastern Egypt cluster zones derived from SLR, GOR, and RANSAC regression algorithms for all clusters, cluster 0 and cluster 1. (b) Same  $M_{\rm w}$   $M_{\rm L}$  relationships for cluster 2, cluster 3, and cluster 4.

that if the absolute variance of the data values is not understood, which is always the case in magnitude estimates, various regression forms should be performed and compared. Therefore, the  $M_{\rm L}$  magnitude scale can be converted into an  $M_{\rm w}$  magnitude scale by building empirical relationships between the correlated earthquake events.

Uncertainties related to various forms of magnitude play a critical role during transformation of magnitudes into a single magnitude scale. The utilized dataset also suffers from the presence of outliers (see box-plots in Figure 7c). Possible causes of observed outliers in magnitude values were not fully identified. The estimated local magnitude values that were extracted from the ENSN bulletin essentially depend on amplitude pickings, which occur over a long period of time at varying levels of seismic noise at a given station, with different levels of noise occurring at different stations. Moreover, the moment magnitude estimate also relies upon the impulse response of the medium, which was never perfectly evaluated and may be highly sparse from one seismic event to another, depending on the event's location versus the seismic stations and earthquake source depth.

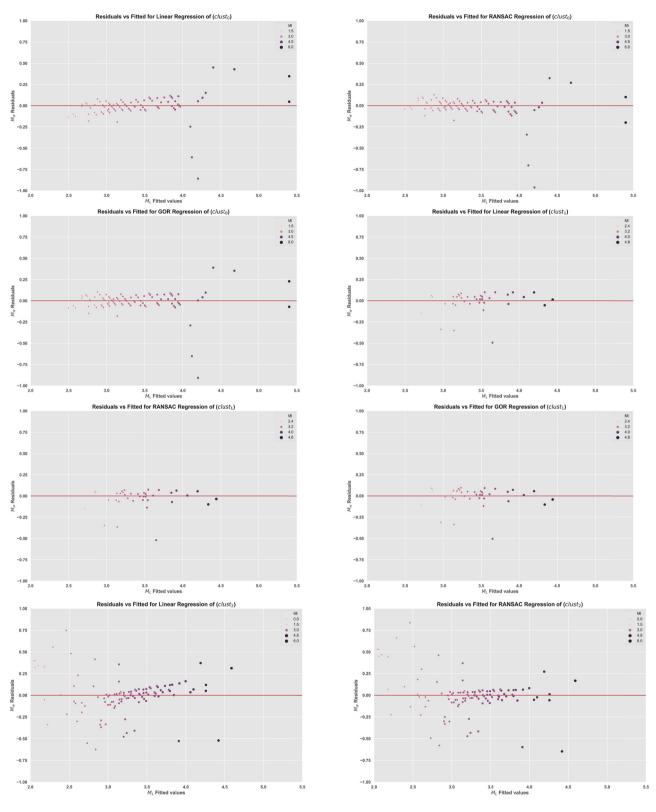
Although there is no clear description or definition of the outlier, outliers are often considered as observations that do not fit the trend of the other observations. It is not usually a concern if the outlier is merely an unusual finding made from the tail of a regular distribution; however, if the outlier comes from a non-usual calculation error or any other breach of typical ordinary least-square rules, then it undermines the integrity of the regression if a non-robust regression method is used [50].

To study the effect of the observed outliers and to develop a robust regression method, the RANdom SAmple Consensust (RANSAC) method [51] was applied to our study. RANSAC is a non-deterministic algorithm that partitions the whole input sample data into a collection of noise-prone inliers and outliers, e.g., due to incorrect measurements or invalid data hypotheses. Therefore, the resulting model is derived from established inliers only. The algorithm creates and validates a linear model based on the Minimum Least-Square (LMS) approach by extracting noisy samples (i.e., outliers), choosing the appropriate features (i.e., descriptors), deriving an LMS model from the training set samples, and finally, predicting the behavior of the test set samples while using the principle of applicability domain [51].

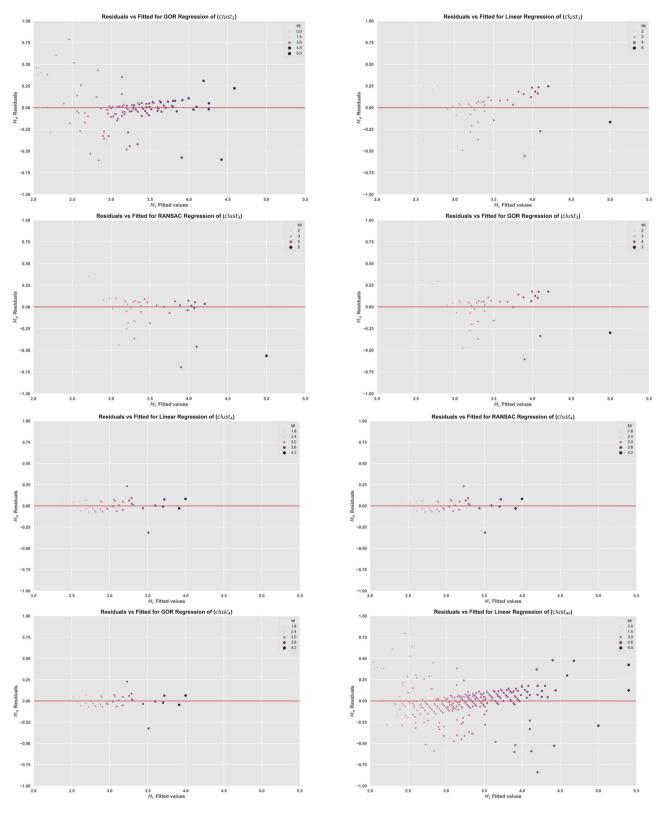
To develop the  $M_{\rm w}$  versus  $M_{\rm L}$  empirical relation, the RANSAC algorithm is utilized to estimate regression coefficients of the linear model by randomly sampling the observed earthquake magnitude data in each cluster to find the optimal fit result and implementing a voting scheme [51]. The implementation of this voting scheme is based on two assumptions: the noisy features will not vote consistently for any single model (few outliers), and there are enough features to agree on a good model (few missing data). Regression fitting is performed in two stepwise manners that are iteratively repeated. In the first step, a subset containing minimal magnitude data is randomly selected from the input earthquake dataset. A linear LMS fitting model and the corresponding model coefficient parameters are computed using only the elements of this magnitude subset. In the second step, the algorithm checks which elements of the entire magnitude

**Table 4:** Regression parameters for the  $M_w-M_1$  relationship

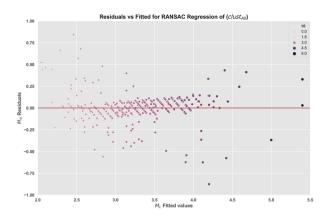
Regression Zone		Slope	Intercept	R <sup>2</sup>	rms error	
Linear	1	1.0043(±0.063)	-0.27026(±0.028)	0.78	0.26	
RANSAC	1	$1.1379(\pm0.047)$	$-0.70964(\pm0.020)$	0.99	0.07	
GOR	1	$1.0618(\pm0.063)$	$-0.46231(\pm0.003)$	0.97	0.11	
Linear	2	$1.0820(\pm0.034)$	$-0.53715(\pm0.096)$	0.60	0.26	
RANSAC	2	$1.0592(\pm0.094)$	$-0.42787(\pm0.076)$	0.97	0.07	
GOR	2	1.1354(±0.037)	$-0.71943(\pm0.014)$	0.97	0.08	
Linear	3	$0.9283(\pm0.053)$	$0.02777(\pm0.013)$	0.90	0.20	
RANSAC	3	$1.0368(\pm0.092)$	$-0.32325(\pm0.089)$	0.96	0.12	
GOR	3	$0.9880(\pm0.013)$	$-0.15752(\pm0.018)$	0.95	0.10	
Linear	4	$0.8811(\pm 0.095)$	$-0.16015(\pm0.088)$	0.80	0.24	
RANSAC	4	$0.8399(\pm 0.012)$	$-0.31575(\pm0.058)$	0.91	0.14	
GOR	4	$0.9594(\pm0.066)$	$-0.09798(\pm0.093)$	0.92	0.17	
Linear	5	$1.0554(\pm0.620)$	$-0.40013(\pm0.321)$	0.66	0.24	
RANSAC	5	$1.0776(\pm0.039)$	$-0.45901(\pm0.022)$	0.98	0.06	
GOR	5	$1.0722(\pm 0.075)$	$-0.44981(\pm0.091)$	0.96	0.05	
Linear	NE-Egypt	$0.9559(\pm0.903)$	$-0.08721(\pm0.162)$	0.84	0.23	
RANSAC	NE-Egypt	$1.0854(\pm0.039)$	$-0.51351(\pm0.082)$	0.98	0.08	
GOR	NE-Egypt	$1.0122(\pm0.0930)$	$-0.26823(\pm0.163)$	0.97	0.10	



(Continued)



(Continued)



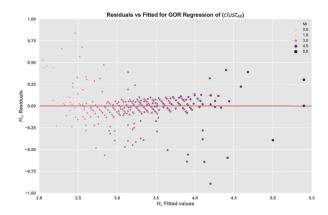


Figure 10: (a) Residuals for each  $M_{\rm w}$   $M_{\rm L}$  relationship for northeastern Egypt cluster zones derived from SLR, GOR, and RANSAC regression algorithms for all clusters and first zone. (b) Same Residuals for each  $M_{\rm w}$   $M_{\rm L}$  relationship for second and third cluster zones. (c) Same Residuals for each  $M_{\rm w}$   $M_{\rm L}$  relationship for fourth and fifth cluster zones.

dataset are consistent with the model instantiated by the estimated model parameters that were obtained from the first step. A data element will be considered an outlier if it does not fit the fitting model instantiated by the set of estimated model parameters within some error threshold that defines the maximum deviation attributable to the effect of noise [52].

The RANSAC regression relationships established between  $M_{\rm w}$  and  $M_{\rm L}$  for the various cluster zones are shown as blue regression lines in Figure 9a–f and Table 4. To check the rationality of the derived conversion relations between magnitude types, we applied ordinary least squares (OLS) and the GOR (see the red and green regression lines in Figure 9a–f). We assumed that the ratio between the variable variance  $(\eta)$ , which was not known for our dataset, was equal to one, based on the regulations described in ref. [22], assuming that the error variances of different magnitudes were approximately equal.

When comparing the different regression metrics, we can see that the RANSAC algorithm can boost our model scores when the dataset includes a significant number of outliers. In this specific case, all the regression metrics scores improved significantly. The difference is particularly observable in the  $R^2$  score, which is a statistical indicator reflecting the percentage of deviation for a dependent variable described by an independent variable. It reached almost 99% with the RANSAC regression model. The good agreement between the observed and estimated magnitudes provides greater confirmation of the results and could be easily used to validate the accuracy of the proposed relations for the earthquake early warning system in the northeastern portion of Egypt. It was also observed that the relations derived

from this RANSAC technique had significantly lower errors in their regression parameters than those of the corresponding relations established from other techniques, as shown in Table 4, for all of the computed regressions.

However, the  $R^2$  score is not always enough to determine which model is a better fit for the results. Additional error analyses are required, given the distribution of the residuals along the fitted line. We added more measurements, including the scattering residual analysis, to complete the predictive analysis on the goodness-of-fit of our regression models. Residuals will show a random scatter, as was already mentioned. Prediction error analysis for the three regression models is illustrated in Figure 10a–c.

It is very clear that most of the points in the residual plot are distributed randomly along the horizontal axis (shown with a red line) for RANSAC and GOR; hence, those two regression models are suitable for predicting the conversion relations between the two types of earthquake magnitude. This also suggests that for any possible regression analysis, grouping the used collection by considering the spatial distribution of earthquakes will vastly enhance the creation of a suitable scaling relationship that can be applied to homogenize the magnitude of the earthquake.

Finally, the proposed relationships were compared to those provided by other authors in order to analyze the validity of the established relations. The results of this study are generally consistent with the observations reported by other authors, although some differences in value were revealed. Meanwhile, there were some differences in the values among the relationships. As geographic classification of seismic events was not discussed by other researchers, this could preclude the direct use of existing scaling relationships by other writers in a similar

area such as the Gulf of Agaba. This should be used to highlight the fact that the scaling relationships in region cannot be used explicitly with that of other regions and need to be investigated further.

### 4 Conclusion

**DE GRUYTER** 

The purpose behind the conversion of different magnitude scales into a single moment magnitude  $M_{\rm w}$  lies in the fact that it remains unsaturated throughout the entire magnitude scale and achieves an accuracy that is two to three times higher than that of other magnitude types. Linear regression models are heavily impacted by the presence of outliers, which can come from extreme values of noise or erroneous measurements. The RANSAC algorithm can be used to remove sets of points that do not follow the dominant pattern of the data in the models, and as a robust machine-learning algorithm, RANSAC improves the model's performance by estimating the parameters with a high degree of accuracy, even when a significant number of outliers are present in the dataset. RANSAC usually performs poorly when the number of inliers in the dataset is less than 50%. In addition, when the noise threshold is too small, the estimated parameters tend to be unstable. Finally, this algorithm removes data from the model, but loss of data should be avoided when developing models.

**Acknowledgment:** This work would not have been possible without the help and support of the Egyptian National Seismological Network (ENSN), Seismology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Cairo, Egypt. The authors would like to thank the Research Supporting Project number (RSP-2020/89), King Saud University, Riyadh, Saudi Arabia for funding this work.

Funding information: This work is financially supported by the Research Supporting Project number (RSP-2020/89), King Saud University, Riyadh, Saudi Arabia.

Authors contributions: S.M. and G.A.M., conceived, planned, and carried out the data analysis. S.M. and G.A.M., contributed to the interpretation of the results. All authors took the lead in writing the manuscript and contributed in the preparation of the figures and finalizing the manuscript.

**Conflict of interest:** Authors state no conflict of interest.

#### References

- [1] Marino GG. Earthquake damage: inspection, evaluation, and repair. Arizona, USA: Lawyers & Judges Publishing Company, Incorporated; 1997.
- Richter CF. An instrumental earthquake magnitude scale. [2] Bull Seismol Soc Am. 1935;25(1):1-32.
- McGuire RK. Seismic hazard and risk analysis. Oakland, USA: Earthquake Engineering Research Institute; 2004.
- [4] Badawy A. Present-day seismicity, stress field and crustal deformation of Egypt. J Seismol. 2005;9(3):267-76.
- Ambrasevs NN, Melville CP, Adams RD. The seismicity of Egypt. Arabia and the Red Sea: a historical review. Cambridge, England: Cambridge University Press; 2005.
- Kagan YY. Accuracy of modern global earthquake catalogs. [6] Phys Earth Planet Inter. 2003;135(2-3):173-209.
- Naraghiaraghi N, Nawawi M, Rahman SM, Beitollahi A, Saad R, Ioneidi S. Homogenization of earthquake catalogue in terms of magnitude in Iran and adjoining region. AJGeo. 2016;6(1):65-70.
- [8] Hafiez HA. Magnitude scales regression for Egyptian seismological network. Arab J Geosci. 2015;8(10):7941-54.
- Gutenberg B, Richter CF. Earthquake magnitude, intensity, energy, and acceleration. Bull Seismol Soc Am. 1942;32(3):163-91.
- [10] Gutenberg B. Magnitude determination for deep-focus earthquakes. Bull Seismol Soc Am. 1945;35(3):117-30.
- Nuttli OW. Seismic wave attenuation and magnitude relations for eastern North America. J Geophys Res. 1973:78(5):876-85.
- [12] Hanks TC, Kanamori H. A moment magnitude scale. J Geophys Res Solid Earth. 1979;84(B5):2348-50.
- [13] McGuire RK. Computations of seismic hazard. Ann Geophysics. 1993;36(3-4):181-200.
- [14] McGuire RK. Probabilistic seismic hazard analysis: aarly history. Earthq Eng Struct Dyn. 2008;37(3):329-38.
- Castellaro S, Mulargia F, Kagan YY. Regression problems for magnitudes. Geophys J Int. 2006;165(3):913-30.
- Bayrak Y, Öztürk S, Çınar H, Kalafat D, Tsapanos TM, [16] Koravos GC, et al. Estimating earthquake hazard parameters from instrumental data for different regions in and around Turkey. Eng Geol. 2009;105(3-4):200-10.
- Das R, Wason H, Sharma M. Magnitude conversion to unified moment magnitude using orthogonal regression relation. J Asian Earth Sci. 2012;50:44-51.
- [18] Das R, Wason H, Sharma M. General orthogonal regression relations between body-wave and moment magnitudes. Seismol Res Lett. 2013;84(2):219-24.
- [19] Öztürk S. A new empirical relation between surface wave magnitude and rupture length for Turkey earthquakes. Earth Sci Res J. 2014;18(1):15-26.
- [20] Öztürk S, Ghassemi MR, Sarı M. A comparison of alternative curve fitting techniques for different earthquake fault parameters of Iranian earthquakes. Earth Sci Res J. 2020;24(4):459-72.
- [21] Antony T. Predicting earthquake magnitude error with regression, decision tree & random forest. Nanyang, Singapore: Nanyang Technological University Press; 2020.

- [22] Castellaro S, Bormann P. Performance of different regression procedures on the magnitude conversion problem. Bull Seismol Soc Am. 2007;97(4):1167–75.
- [23] Lolli B, Gasperini P. A comparison among general orthogonal regression methods applied to earthquake magnitude conversions. Geophys J Int. 2012;190(2):1135–51.
- [24] Wason H, Das R, Sharma M. Regression relations for magnitude conversion for the Indian region. Advances in indian earthquake engineering and seismology. Cham, Switzerland: Springer; 2018. p. 55–66.
- [25] Hussein H, Abou Elenean K, Marzouk I, Peresan A, Korrat I, El-Nader EA, et al. Integration and magnitude homogenization of the Egyptian earthquake catalogue. Nat Hazards. 2008;47(3):525–46.
- [26] Hafiez HA, El-Hussain I, Khalil A, Deif A. Determination of a local earthquake magnitude scale for the Sultanate of Oman. Arab J Geosci. 2015;8(4):1921–30.
- [27] Sawires R, Peláez JA, Fat-Helbary RE, Ibrahim HA. An earthquake catalogue (2200 BC to 2013) for seismotectonic and seismic hazard assessment studies in Egypt. Earthquakes and their impact on society. Cham, Switzerland: Springer; 2016. p. 97–136.
- [28] Pondard N, Armijo R, King GC, Meyer B, Flerit F. Fault interactions in the sea of Marmara pull-apart (North Anatolian Fault): earthquake clustering and propagating earthquake sequences. Geophys J Int. 2007;171(3):1185–97.
- [29] Bock HH. Origins and extensions of the *k*-means algorithm in cluster analysis. Electron J History Probabil Stat. 2008;4(2):1–18.
- [30] Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa LD, et al. Clustering algorithms: a comparative approach. PLoS One. 2019;14(1):e0210236.
- [31] Brune JN. Tectonic stress and the spectra of seismic shear waves from earthquakes. J Geophys Res. 1970;75(26):4997–5009.
- [32] Brune JN. Correction to tectonic stress and the spectra of seismic shear waves from earthquakes. J Geophys Res. 1971;76(20):5002.
- [33] Boore DM, Boatwright J. Average body-wave radiation coefficients. Bull Seismol Soc Am. 1984;74(5):1615-21.
- [34] Bock HH. Clustering methods: a history of k-means algorithms. Selected contributions in data analysis and classification. New York, USA: Springer; 2007. p. 161–72.
- [35] Naldi MC, Fontana A, Campello RJ. Comparison among methods for k estimation in k-means. 2009 Ninth International Conference on Intelligent Systems Design and Applications. Cham, Switzerland: IEEE; 2009. p. 1006–13.
- [36] Bora DK, Baruah S, Biswas R, Gogoi NK. Estimation of source parameters of local earthquakes originated in Shillong-Mikir plateau and its adjoining region of northeastern India. Bull Seismol Soc Am. 2013;103(1):437-46.
- [37] Moustafa SS, Takenaka H, Mamada Y. Estimation of coda-wave attenuation in the vicinity of metropolitan Cairo, Egypt. Bull Faculty Sci Kyushu Univ. 2002;31(2):71–9.

- [38] El Hadidy S. Crustal structure and its related causative tectonics in northern Egypt using geophysical data. Ph.D. thesis. Cairo, Egypt: Ain Shams University; 1995.
- [39] Aki K, Richards PG. Quantitative seismology. New York, USA: Freeman; 2002.
- [40] Andrews D. Objective determination of source parameters and similarity of earthquakes of different size. Earthq Source Mech. 1986;37:259-67.
- [41] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. In: Numerical analysis. Cham, Switzerland: Springer; 1978. p. 105-16.
- [42] Archuleta RJ, Cranswick E, Mueller C, Spudich P. Source parameters of the 1980 Mammoth Lakes, California, earthquake sequence. J Geophys Res Solid Earth. 1982;87(B6):4595–607.
- [43] Badawy A, Abdel-Fattah A, Ali SM, Farid W. Source parameters of the ML 4.1 earthquake of November 08, 2006, southeast Beni-Suef, northern Egypt. J Afr Earth Sci. 2008;51(3):151-9.
- [44] Mohamed GEA, Omar K. Source parameters and moment tensor of the ML 4.6 earthquake of November 19, 2011, southwest Sharm El-Sheikh, Egypt. NRIAG J Astron Geophys. 2014;3(1):27-36.
- [45] Saadalla H, Abdel-Aal AAK, Mohamed A, El-Faragawy K. Characteristics of Earthquakes recorded around the high dam lake with comparison to natural earthquakes using waveform inversion and source spectra. Pure Appl Geophys. 2020;177:3667-95.
- [46] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;2:224-7.
- [47] Abdelazim M, Samir A, El-Nader IA, Badawy A, Hussein H. Seismicity and focal mechanisms of earthquakes in Egypt from 2004 to 2011. NRIAG J Astron Geophys. 2016;5(2):393-402.
- [48] Abdalzaher MS, El-Hadidy M, Gaber H, Badawy A. Seismic hazard maps of Egypt based on spatially smoothed seismicity model and recent seismotectonic models. J Afr Earth Sci. 2020;170:103894.
- [49] Ristau J. Comparison of magnitude estimates for New Zealand earthquakes: moment magnitude, local magnitude, and teleseismic body-wave magnitude. Bull Seismol Soc Am. 2009;99(3):1841–52.
- [50] Marasinghe MG. A multistage procedure for detecting several outliers in linear regression. Technometrics. 1985;27(4):395-9.
- [51] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 1981;24(6):381–95.
- [52] Choi S, Kim T, Yu W. Performance evaluation of RANSAC family. J Comput Vis. 1997;24(3):271–300.
- [53] Badawy A. Seismicity of Egypt. Seismol Res Lett. 2005;76(2):149-60.