Seismic anomaly detection based on PSO-VMD separation of satellite electromagnetic signals

Yongming Huang[a,*], Hongbin Qiu[a], Yiheng Meng[a], Yongsheng Ma[a], Yong Lu[b], Guobao Zhang[a], Yuntian Teng[c]

[a] Southeast University, Nanjing, 210096, China
[b] Seismological Bureau of Jiangsu Province, Nanjing, 210006, China
[c] Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China

Abstract

During the process of earthquake development, minute fluctuations in the Earth's geomagnetic field strength can be affected, thereby enabling the analysis of seismic anomalies through geomagnetic data. Since the variations in the geomagnetic field are relatively small compared to the overall field strength, it is essential to decompose the geomagnetic field before performing anomaly detection. In the analysis of the seismic event that occurred in the Nicaragua region on September 15, 2016, an enhanced signal decomposition technique based on Variational Mode Decomposition (VMD), specifically Particle Swarm Optimization-VMD (PSO-VMD), was employed to mitigate the interference of the electromagnetic dynamic field and investigate the association between electromagnetic data and earthquakes. Furthermore, the Empirical-Cumulative-distribution-based Outlier Detection (ECOD) method was utilized for anomaly clustering of orbital numbers.

Compared to the raw data, the methodology presented in this study shows that the accumulation of anomalies preceding the earthquake matches a slow-fast-slow pattern. To further explore the statistical attributes, a random earthquake approach was implemented, and the results demonstrated that the level of detected anomalies during earthquakes using this methodology increased by 120% relative to non-earthquake periods. This finding compellingly substantiates a robust correlation between electromagnetic data and earthquakes.

Keywords: PSO-VMD, Seismic anomaly detection, ECOD, Feature fusion

1. Introduction

The increasing number of global satellite launches has led to a vast accumulation of electromagnetic monitoring data. Abnormal earthquake detection now relies heavily on satellite signals, particularly from DEMETER (Detection of Electro-Magnetic Emissions Transmitted from Earthquake Regions) and the Swarm satellites as well as CSES (China Seismo-

---

Electromagnetic Satellite)(Zhao et al., 2022). DEMETER, a French satellite launched on June 29, 2004, was specifically designed for detecting electromagnetic signals associated with earthquake precursors. It is equipped with electromagnetic sensors, low-energy particle detectors, and high-energy electron detectors, enabling it to observe environmental changes in the ionosphere and magnetic field. Its main objective was to identify electromagnetic signals related to earthquake precursors, enhancing earthquake prediction and early warning capabilities (Shi et al., 2023). DEMETER followed a sun-synchronous orbit, with nighttime observations during the upward half and daytime observations during the downward half.

Statistical analysis of the data obtained from the DEMETER satellite has demonstrated a certain correlation with earthquakes (Parrot, 2012) . Xiong et al. (2020) utilized the Light-GBM method based on gradient descent to analyze DEMETER data and found that certain frequency ranges of low-frequency electric and magnetic fields were key features in identifying electromagnetic precursor disturbances. Akhoondzadeh and Marchetti (2022) proposed a Mamdani fuzzy inference system (FIS) for earthquake prediction using DEMETER data, finding that the highest number of anomalies appeared approximately one month before an earthquake. Additionally, studies by Nemec et al. (2008) and Molchanov et al. (2006) showed correlations between VLF electromagnetic wave intensity and major earthquakes, but these correlations were often weak and inconsistent. Zlotnicki et al. (2006) attempted multi-scale statistical analysis to identify pre-seismic anomalies, yet their methods were limited by the complexity and variability of the data.

The Swarm (Friis-Christensen et al., 2006) satellite mission, launched by the European Space Agency (ESA) on November 22, 2013 with the aim of collecting data on Earth's magnetic field, magnetic field variations, and geomagnetic activity, consists of three satellites: Swarm A, Swarm B, and Swarm C. Unlike DEMETER, Swarm satellites are not sun-synchronous, meaning their local time measurements vary with each pass, complicating the analysis of data due to time-dependent variations.

Marchetti et al. (2022) analyzed magnetic field and electron density signals from Swarm and found a statistical correlation with M5.5+ shallow earthquakes, proposing frequency as a criterion for anomaly detection. However, geomagnetic storms and other natural phenomena often interfere with these signals, making it challenging to isolate true seismic anomalies.

In the analysis of Swarm satellite data, De Santis et al. (2017) introduced the slow-fast-slow pattern, known as the "S-shape," explaining it as a critical system nearing a mainshock. Fan et al. (2022) further developed a signal decomposition method for Swarm data analysis. Our research builds on these foundational studies.

According to the existing research content, this study obtained the vector information of electromagnetic data through the Swarm satellite. First, the CYI4 model was used to eliminate the static field of the geomagnetic vector data, and then the VMD method based on PSO optimization was used to extract the two-modal data to eliminate interference from geomagnetic activities, and then use the unsupervised outlier detection (ECOD) of the empirical cumulative distribution function to cluster the data points and judge the cumulative anomalies within a certain time span. The study mainly contains the following contributions:

1. The study presents an innovative methodology for seismic anomaly detection. By leveraging a signal decomposition technique based on Variational Mode Decomposition (VMD) and specifically Particle Swarm Optimization-VMD (PSO-VMD), the authors successfully mitigate the interference of the electromagnetic dynamic field. This novel approach significantly enhances the analysis of seismic anomalies through geomagnetic data.
2. This study marks the novel application of the Empirical-Cumulative-distribution-based Outlier Detection (ECOD) method for anomaly clustering of orbital numbers. This methodology improves the detection and clustering of seismic anomalies.
3. The research offers significant practical contributions to the field of seismic anomaly detection. It demonstrates the potential of PSO-VMD and ECOD in extracting valuable insights from satellite electromagnetic signals. The findings could be used to better complete pre-earthquake anomaly detection tasks, potentially saving lives and reducing property damage.

2. Data and Methods

Based on the analysis of existing research, this paper initially acquired vector information of electromagnetic data through the Swarm satellite. The CYI4 (Sabaka et al., 2018) model was then utilized to remove the core and lithospheric magnetic fields from the geomagnetic vector data. To eliminate the interference of geomagnetic activity, a Particle Swarm Optimization (PSO) (Wanpeng and Zhenhong, 2010) - optimized Variational Mode Decomposition (VMD) method was employed to extract two modal data, using the average sliding energy as a basis for eliminating the effects of geomagnetic activity. For anomaly extraction, an unsupervised outlier detection technique called Empirical Cumulative Distribution Function (ECOD)(Li et al., 2022) was used to cluster data points and assess accumulated anomalies within a certain time span. These accumulations were then fitted with a Boltzmann function at a given time scale to determine if they exhibited sharp increases before earthquakes and stabilized afterward. The work flow of the model is shown as Figure 1. By analyzing seismic events in the Nicaragua region, it was observed that the seismic anomalies exhibited a distinct gradual-steep-gradual characteristic. Building upon this observation, seismic events from random global locations were randomly selected, and a statistical analysis of the pre-earthquake cumulative anomaly growth trend for both non-seismic and seismic events was conducted. The results demonstrated the effectiveness of this method in extracting pre-earthquake anomaly features.

2.1. Data from Swarm

The Swarm satellite is equipped with multiple payloads to meet various observational requirements. These payloads include magnetometers, GPS receivers, star trackers, and a gravity gradient instrument, among others. The Level 1b data is a standardized format used to describe the measurement results from the Swarm satellite. This data can be obtained for free through https://swarm-diss.eo.esa.int/. The Level 1b data record the raw data from satellite observations, including satellite position, velocity, attitude, and Earth's magnetic
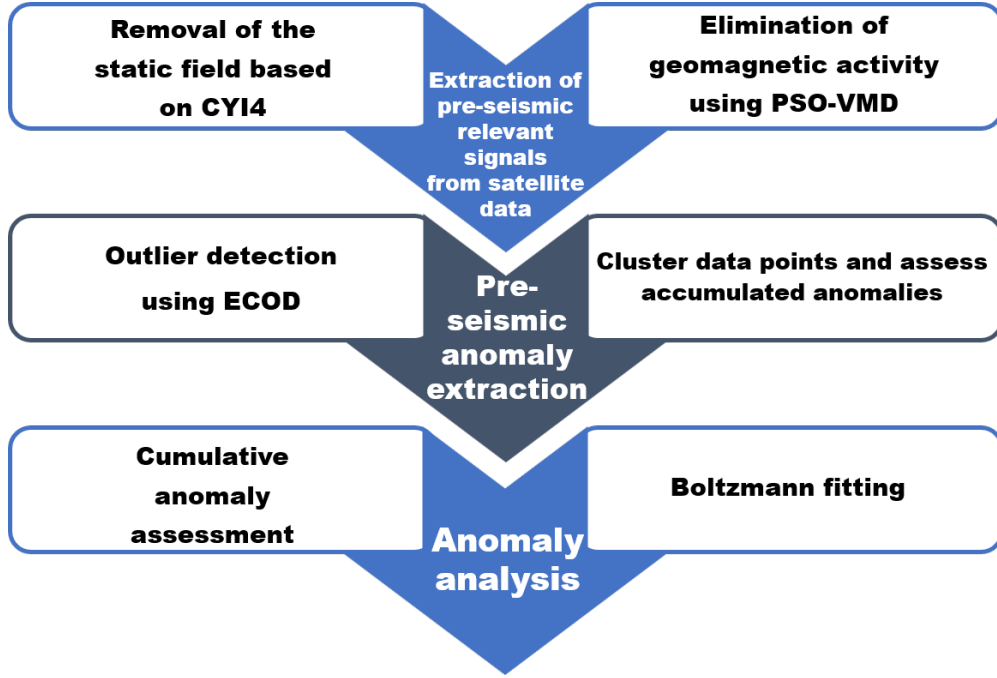
Figure 1: The workflow for pre-seismic anomaly detection based on Swarm satellite electromagnetic data

field, as well as pre-processed and corrected measurements of the magnetic and gravity fields. Specifically, the Level 1b data consists of the following components: Timestamp: Records the time and date of the data record.

- **Raw Data:** Includes satellite attitude, position, and velocity data, as well as raw data measured by various sensors.

- **Data Pre-processing:** Pre-processes and corrects the raw data to compensate for sensor errors and signal noise.

- **Earth's Magnetic Field:** Models and measures Earth's magnetic field to obtain more accurate magnetic field data.

- **Gravity Field:** Models and measures the gravity field to obtain more accurate gravity field data.

- **Correction Coefficients:** Includes various correction coefficients and correction factors to correct deviations and errors in the data.

The Level 1b data format of the Swarm satellite adopts the standard NetCDF format, facilitating data exchange and processing on different platforms. Data files in this format can be read, processed, and visualized using various scientific data processing software, greatly facilitating the analysis of magnetic field data. All the data are sourced from MAGX_LR, X represents (A, B, C) for the three satellites.

2.2. CYI4 model

In order to remove the influence of the earth's main magnetic field, lithospheric magnetic field, external source field, geomagnetic activity, etc. from the electromagnetic information obtained by the swarm satellite, and extract the part of the electromagnetic signal directly affected by earthquakes, this study uses a joint model to remove the static field. For existing joint models, such as CM5, CIY4, CHAOS, etc., since CIY4 improves the possible magnetic anomaly areas in CM5 and enriches the ionospheric inversion data compared to CHAOS, this study uses the CIY4 model to analyze the static field to remove.

The inversion process of CIY4 mainly includes the following steps: first, preprocessing the geomagnetic vector data, including removing the influence of satellite attitude; then transforming the processed data into the geomagnetic field in spherical coordinates; next, decomposing the geomagnetic field into different levels and types of spherical harmonic terms using the spherical harmonic expansion method; calculating the analytical solution of the spherical harmonic coefficients; smoothing and predicting the spherical harmonic coefficients; finally, evaluating and validating the inversion results. These steps aim to accurately invert the Earth's magnetic field information and ensure the accuracy and reliability of the inversion results.

Table 1 presents the parameter values for the given models in the Swarm product. The model corresponding to the core field is MCO_SHA_2C, the ionospheric field is represented by the model MIO_SHA_2C, the lithospheric field is associated with the model MLI_SHA_2C, and the magnetospheric field corresponds to the model MMA_SHA_2C. In these models, N_{min} denotes the minimum degree of the spherical harmonic expansion, while N_{max} represents the maximum degree of the expansion. The degree of expansion determines the accuracy and spatial resolution of the model.

Table 1: Value of Swarm Model Parameters

Swarm Models	Description	Value
MCO_SHA_2C	core field	$N_{min} = 1, N_{max} = 18$
MIO_SHA_2C	ionospheric field	$N_{min} = 1, N_{max} = 60$
MLI_SHA_2C	lithospheric field	$N_{min} = 16, N_{max} = 120$
MMA_SHA_2C	magnetospheric field	$N_{min} = 1, N_{max} = 2$

The table is a table of the values of various swarm model parameters, where the first and second columns represent different swarm models and the corresponding magnetic fields. In the third column, N_{min} represents the minimum expansion order in the parameters, and N_{max} represents the maximum expansion order.

Figures 2(a)-(d) illustrate the core field data, ionospheric field data, lithospheric field data, and magnetospheric field data obtained from the CIY4 inversion model. Among them, Figure 2(e) displays the observational data from the Swarm satellites, while Figure 2(f)

depicts the residuals between the observed data and the inversion results of the CIY4 model. Through careful examination, it is evident that the core field exhibits an exceptionally high correlation with the observed data, with a Pearson coefficient of 0.99982901. The very high value of the Pearson coefficient is probably based on periods without geomagnetic storms. Given the magnitude of the core field, reaching up to 10^4 nT, it is crucial to remove its influence from the raw observational data.
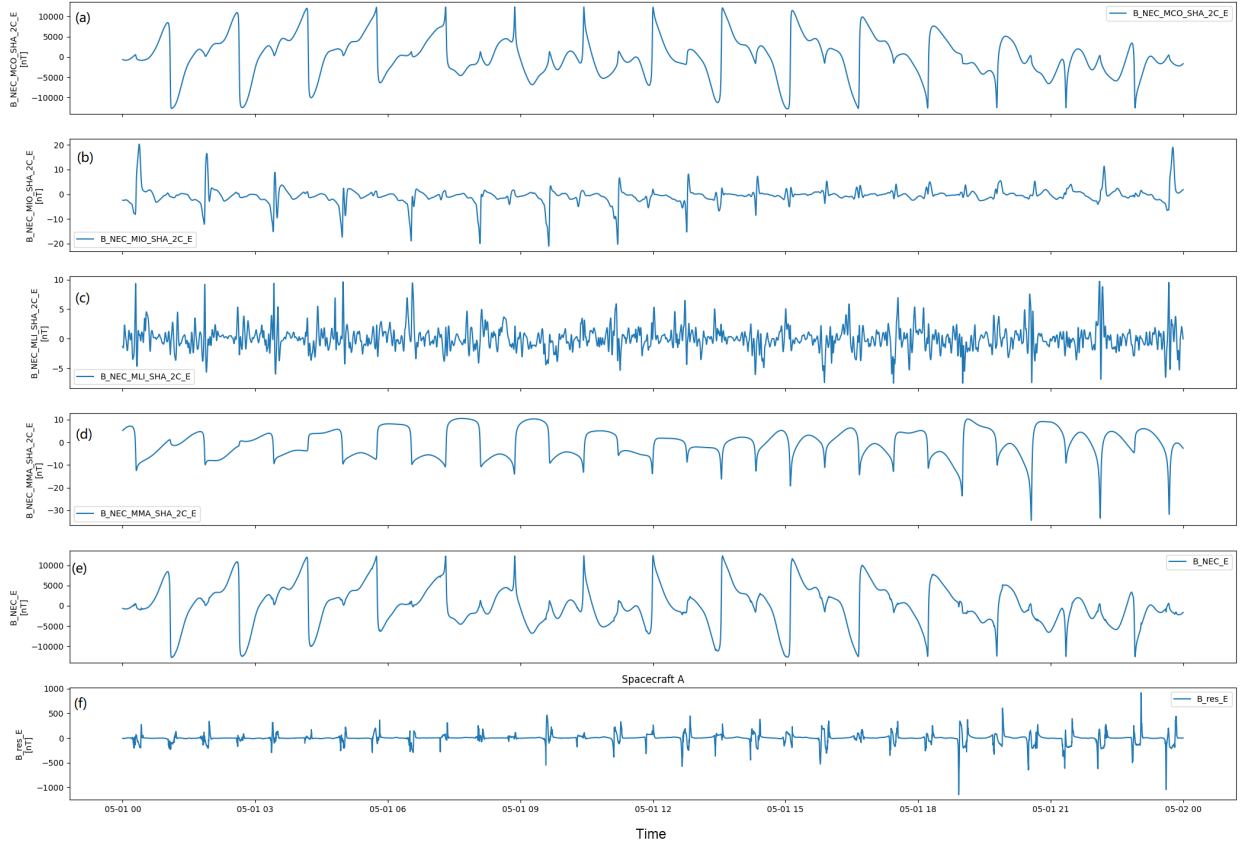

Figure 2: Data of CIY4 model. Each sub-figure represents the electromagnetic E component data and the electromagnetic E component residual graph, where (a) is the main magnetic field component data, (b) is the ionosphere magnetosphere component data, (c) is the lithosphere magnetic field component data, (d) is the magnetosphere magnetic field component data, (e) is the original observation data, and (f) is the residual of the observation data and the model inversion data.

2.3. PSO-VMD method

VMD (Variational Mode Decomposition) (Dragomiretskiy and Zosso, 2014) is a signal decomposition method based on variational principles, which can decompose any signal into a set of oscillatory modes and corresponding auxiliary signals. Unlike traditional signal decomposition methods such as wavelet decomposition and Fourier analysis, VMD does not assume the signal to be stationary or periodic, and it can adaptively extract the local characteristics of the signal.

However, before using the VMD algorithm, it is necessary to determine the penalty coefficient α and the number of modes K in advance. The selection of these two parameters directly affects the results of the VMD algorithm. If K is too large, it may lead to over-decomposition, while if it is too small, it may result in under-decomposition. The value of α affects the bandwidth of each mode. A smaller α leads to larger bandwidth for each mode, which may cause overlapping passbands among the modes. On the other hand, a larger α leads to smaller bandwidth, which may result in the loss of signal information. Therefore, to determine appropriate values for α and K , this study adopts the Particle Swarm Optimization (PSO)-based VMD algorithm (referred to as PSO-VMD hereinafter). By using the PSO algorithm to adaptively search for the optimal α and K values, the subjectivity and limitations associated with manual parameter selection can be avoided.

The Particle Swarm Optimization (PSO) algorithm treats each solution as a particle in the entire solution space. Each particle is initialized with its initial position and velocity, and it moves in the solution space at a unit time according to the direction and step size determined by its own best local history and the best history of the collective particles. By starting from random positions in the solution space, multiple particles avoid the drawback of a single particle getting trapped in local optima. Compared to other optimization algorithms, PSO algorithm has advantages such as strong global search capability, easy implementation, fast convergence speed, etc., making it suitable for solving complex optimization problems. The update equations for the particle velocity and position are shown in Equation 1.

$$
\begin{aligned}
v_i^{t+1} &= wv_i^t + c_1r_1(pbest_i - x_i^t) + c_2r_2(gbest - x_i^t) \\
x_i^{t+1} &= x_i^t + v_i^{t+1}
\end{aligned}
\tag{1}
$$

In the PSO-based VMD algorithm, v_i^t represents the velocity of particle i at the $t-th$ iteration, and x_i^t represents the position of particle i at the $t-th$ iteration. $pbest_i$ denotes the best position (local optimum) achieved by particle i throughout its history, and $gbest$ represents the best position (global optimum) obtained by any particle in the swarm. r_1 and r_2 are random numbers between 0 and 1, used to introduce randomness to the particles' movements and prevent them from being trapped in local optima.

The remaining three coefficients in the algorithm, w , c_1 and c_2 , represent the inertia weight, cognitive weight, and social weight, respectively. The inertia weight determines the influence of a particle's previous velocity on its current velocity. A higher value of w means that the particle's direction of iteration is less likely to change. The cognitive weight c_1 controls the particle's tendency to move towards its personal best position, while the social weight c_2 determines the degree of influence from the swarm's global best position.

In the context of the VMD algorithm, K and α are treated as the position parameters of the particles. The loss error is calculated as the residual of the sum of all modes, as shown in Equation 2. The L^2 norm, denoted by $\|\cdot\|_2$, is used to measure the magnitude of the residual.

$$
u_{error} = \left\| f(t) - \sum_k u_k(t) \right\|_2
\tag{2}
$$

The inputs are the electromagnetic data $f(t)$ and the number of iterations $Iter$, and the

output is the decomposed modal data u_k. In the algorithm, the particle swarm is constructed by setting the number of particles num and initializing their initial states. The parameters for the PSO algorithm are also set. In each iteration, the VMD decomposition u_k^j is computed for the current particle state (as shown in Equation **??**). The particle states are updated by calculating the residual (as shown in Equation 2) and recording the individual best solution $pbest_i$ as well as the global best solution $gbest$ (as shown in Equation 1). After each particle completes one iteration, the next iteration begins. Once all iterations are completed, the VMD decomposition values are computed based on $gbest$, and the modal data u_k is returned.

After completing all iterations, the VMD decomposition values are obtained based on the global best position $gbest$. The modal data u_k can then be extracted from the VMD decomposition.

Algorithm 1 delineates the complete procedural flow of PSO-based Variational Mode Decomposition (VMD). This algorithm takes electromagnetic data $f(t)$ and the number of iterations $Iter$ as inputs and yields the decomposed mode data u_k as output. The algorithm which is listed below begins by constructing the particle swarm through initializing the number of particles as num and specifying their initial states. Additionally, the iteration coefficient for the particle swarm algorithm is set. Within each iteration, the VMD decomposition u_k^j is computed under the current state of each particle (refer to Equation **??**). The state of each point is updated by evaluating the residual (as shown in Equation 2), while simultaneously recording the individual best solution $pbest_j$ and the global best solution $gbest$ (as depicted in Equation 1). Once all particles have completed one iteration, the algorithm proceeds to the next iteration. Upon completion of all iterations, the VMD decomposition value is computed based on $gbest$, and the resulting solution u_k is returned.

Algorithm 1 PSO-based-VMD

---

input: electromagnetic $f(t)$ and the number of iterations $Iter$
output: the decomposed mode data u_k

1: initializing the number of the particles num and specifying their initial states;
2: **for** $i = 1, 2, \cdots, Iter$ **do**
3:     **for** $j = 1, 2, \cdots, num$ **do**
4:         Perform VMD decomposition u_k^j based on the current particle state
5:         Update the state of each point using the residual calculation
6:         Update the individual best solution $pbest_j$ and the global best solution $gbest$
7:     **end for**
8: **end for**
9: Calculate the VMD decomposition value based on $gbest$
10: **return** u_k;

The PSO-based VMD algorithm offers several advantages. First, it automatically determines the optimal values of the parameters K and α through the particle swarm optimization process, eliminating the subjective and limited manual selection. Second, the PSO algorithm's global search capability helps to avoid getting trapped in local optima, leading

to better decomposition results. Third, the PSO algorithm is computationally efficient and can handle complex optimization problems.

Overall, the PSO-based VMD algorithm provides an effective and automatic approach for decomposing signals into modal components and auxiliary signals. By adaptively determining the values of K and α, it ensures accurate and reliable signal decomposition, enabling the extraction of meaningful features and improving the analysis of the original signals.

2.4. ECOD method

ECOD (Empirical Cumulative Outlier Detection) is an unsupervised outlier detection method that effectively detects outliers in data (Li et al., 2022). This method is based on the Empirical Cumulative Distribution Function (ECDF), which is a non-parametric empirical distribution function representing the proportion of data points in a dataset that are less than or equal to a given value. To ensure outlier detection in different directions, ECOD calculates the left-tail ECDF and right-tail ECDF of the data in the $j-th$ dimension, as shown in Equation 3. Here, x_i^j represents the $j-th$ dimensional value of the $i-th$ sample, n is the total number of samples, and $1(\cdot)$ is the indicator function. $F_{left}^j(z)$ represents the proportion of samples in the $j-th$ dimension that are less than or equal to z, i.e., the left-tail ECDF, while $F_{right}^j(z)$ represents the proportion of samples in the $j-th$ dimension that are greater than or equal to z, i.e., the right-tail ECDF. By calculating the left-tail ECDF and right-tail ECDF in each dimension, ECOD constructs a multi-dimensional ECDF in the multi-dimensional space and utilizes it for outlier detection of multi-dimensional data.

$$
\begin{aligned}
F_{left}^j(z) &= \frac{1}{n} \sum_{i=1}^n 1(x_i^j \leq z) \\
F_{right}^j(z) &= \frac{1}{n} \sum_{i=1}^n 1(x_i^j \geq z)
\end{aligned}
\tag{3}
$$

To calculate the skewness of the data in the $j-th$ dimension, this paper uses the sample's third central moment as a measure of skewness, as shown in Equation 4. This skewness value will serve as a reference for the final outlier score. Here, SK represents the skewness, $E[\cdot]$ denotes the mathematical expectation, X represents a random variable, μ represents the mean, σ represents the standard deviation, x_i represents the $i-th$ sample, $\bar{x}$ represents the sample mean, and n represents the sample size.

$$
\begin{aligned}
\text{SK} &= E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \\
&= \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\right)^{\frac{3}{2}}}
\end{aligned}
\tag{4}
$$

In ECOD, there are three components involved in calculating the outlier scores: left-only

value, right-only value, and auto value $O_{auto}(X_i)$, as shown in Equation 5.

$$O_{left\_only}(X_i) = -\sum_{j=1}^{d} \log \left(F_{left}^j \left(X_i^j \right) \right)$$

$$O_{right\_only}(X_i) = -\sum_{j=1}^{d} \log \left(F_{right}^j \left(X_i^j \right) \right) \tag{5}$$

$$O_{auto}(X_i) = -\sum_{j=1}^{d} [1\{SK_j < 0\} \log(F_{left}^j(X_i^j))$$
$$+ 1\{SK_j >= 0\} \log(F_{right}^j(X_i^j))]$$

In outlier detection, the larger the value of $O(\cdot)$, the farther the data point is from the collective characteristics of the entire dataset, indicating a higher likelihood of being an outlier. $O_{left\_only}(X_i)$ represents the distance of data point X_i from the left-tail outliers, while $O_{right\_only}(X_i)$ represents the distance from the right-tail outliers. $O_{auto}(X_i)$ considers the skewness of each feature in X_i to determine which direction of outliers to consider.

Specifically, if the skewness of the $j - th$ feature is less than 0, indicating a right-skewed distribution, the left-tail outliers are more likely, so the left-tail outliers are used. Conversely, if the skewness of the $j - th$ feature is greater than or equal to 0, indicating a left-skewed distribution, the right-tail outliers are more likely, so the right-tail outliers are used. The maximum value among $O_{left\_only}(X_i)$, $O_{right\_only}(X_i)$, and $O_{auto}(X_i)$ is taken as the outlier score O_i for data point X_i , as shown in Equation 6.

$$O_i = \max \{O_{left\_only}(X_i), O_{right\_only}(X_i), O_{auto}(X_i)\} \tag{6}$$

This outlier score O_i helps identify the extent to which a data point deviates from the overall distribution of the dataset, thereby indicating its likelihood of being an outlier.

3. Results and Discussion

In order to verify the effectiveness of the anomaly detection method proposed in this study, this study selected an incident that occurred on September 15, 2016, 11 kilometers north-northeast of La Paz Centro urban area in Leon Province, Nicaragua (west longitude 86°, north latitude 12°) of a shallow earthquake of magnitude 5.7. After obtaining electromagnetic data from the swarm satellite in the time range from 6 months before the earthquake to 3 months after the earthquake, anomalies were detected within this interval. After obtaining the electromagnetic data, effective earthquake-related signals are obtained by removing the static field and dynamic field of the earth. On this basis, the data is segmented. The segmentation is based on the orbit number (OrbitNumber) of the Swarm satellite. For the Swarm A satellite, it is in a near-polar orbit with an altitude of 462-524 kilometers and an orbital inclination of 87.4 degrees. run. In this orbit, the Swarm A satellite takes approximately 91 minutes to orbit the Earth. Therefore, the data time span of each piece of data is 91 minutes.

Figure 3 presents the geographic distribution of earthquake epicenters with a magnitude greater than 5.0 in the Nicaragua region from 2000 to the present (March 4, 2023). In the Figure, the size of the blue dots represents the magnitude intensity of each earthquake. Furthermore, the red star in the figure indicates the location of the seismic event selected in this survey.
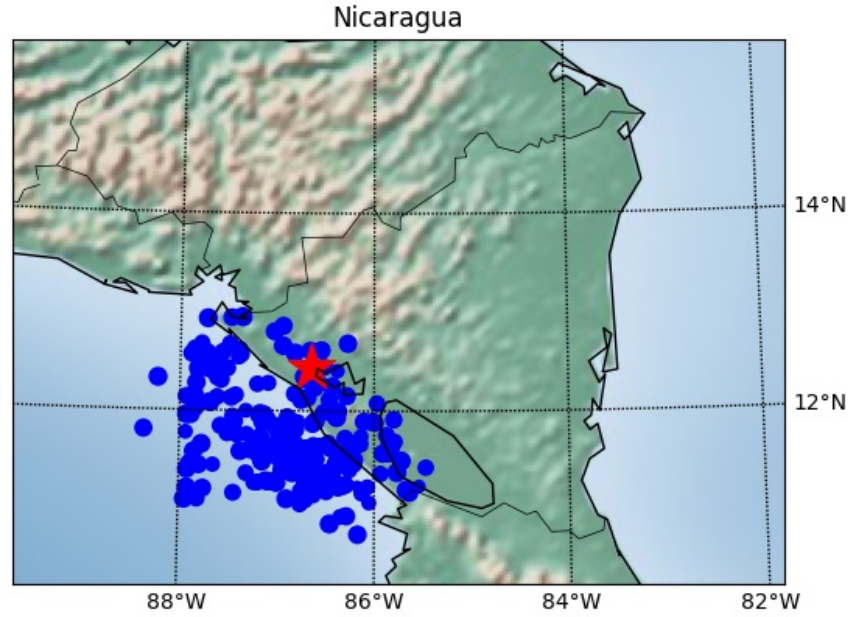

Figure 3: Nicaragua earthquake distribution map. Each point in the map represents the location of an earthquake with a magnitude greater than 5. The size of the blue point represents the magnitude intensity, and the red asterisk represents the earthquake selected in the experiment.

As shown in Figure 4, Figure 4a displays the cumulative results obtained by applying the ECOD method directly to the original data without utilizing the PSO-VMD method for preprocessing. On the other hand, Figure 4b represents the cumulative results obtained after applying the PSO-VMD method for preprocessing. Since the trend of the sigmoid function(Oksum et al., 2021) is gradual-steep-gradual, fitting the cumulative results with the sigmoid function allows for a better representation of the pre-earthquake anomaly trend. Consequently, the fitted Sigmoid function is utilized to discern the trend of the critical anomaly.

To determine whether anomalies are related to earthquakes, analysis can be conducted by calculating cumulative anomalies. If the cumulative anomalies show a linear increasing trend, it indicates a random process. Conversely, if the cumulative anomalies exhibit a sharp rise before the earthquake, especially with multiple sudden anomalies occurring in the period leading up to the earthquake, followed by a return to stability after the earthquake, it can be concluded that there is a strong correlation between the anomalies and the earthquake
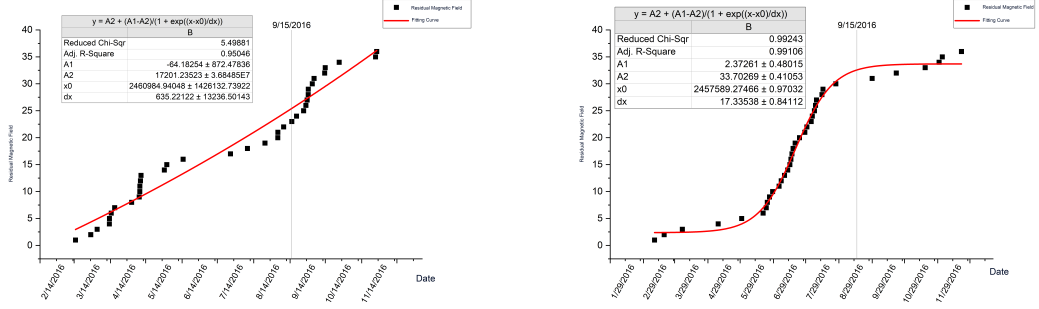
(a) Anomaly detection based on raw data   (b) Anomaly detection based on PSO-VMD

Figure 4: Pre-earthquake anomaly detection comparision of Nicaragua between raw data and preprocessed data with PSO-VMD.

event. Since the trend of the sigmoid function (Equation 7) is gradual-steep-gradual, fitting the cumulative results with a sigmoid function can better characterize the trend of pre-earthquake anomalies, allowing the identification of critical anomaly trends. Compared with other sigmoid functions (such as Logistic function, Hill function, SWeibull function), the Boltzmann function has more adjustment parameters, so this study chose it as the fitting function. The adjusted sigmoid function is Boltzmann function like Equation 7, where A_1 controls the upper bound of the function, and the A_2 control the lower bound, x_0 controls the growth inflection point, and dx controls the growth slope.

$$f(x) = A_2 + \frac{A_1 - A_2}{1 + e^{-\frac{(x-x_0)}{dx}}} \tag{7}$$

To provide an initial assessment of the fitting results, this study utilizes the Adjusted R-Squared coefficient. The main purpose of Adjusted R-Squared is to measure the difference between the fitting results and the original data. Equation 8 represents the calculation method of Adjusted R-Squared, where RSS denotes the residual sum of squares of the regression model, TSS represents the total sum of squares of the data, n is the sample size and p is the number of independent variables. Compared to R-Squared, Adjusted R-Squared adjusts R-Squared based on the sample size and the number of independent variables to account for the influence of irrelevant dimensions on the fitting results. The use of Adjusted R-Square in this study can better reflect the correlation between the abnormal accumulation trend and the fitting function, so as to better study pre-earthquake anomalies.

$$\begin{cases} R^2 = 1 - \dfrac{RSS}{TSS} = 1 - \dfrac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \\ R^2_{adj} = 1 - \dfrac{(1 - R^2)(n-1)}{n - p - 1} \end{cases} \tag{8}$$

In order to further study the generalization ability of this method, we randomly selected multiple earthquake events and non-earthquake events and applied this method to perform

2025-04-05

pre-earthquake anomaly detection. Within the year 2016, a random selection of events was made. The time range for the selected events was 180 days before and 90 days after the given earthquake occurrences. Among the selected events, there were a total of 6 earthquake events and 14 non-earthquake events (no earthquakes greater than magnitude 5 occurred within the given grid). The detailed information of the earthquake and non-earthquake events is presented in Tables 2 and Tables 3, respectively.

Table 2: random earthquake event

Index	Date	Latitude	Longitude	Magnitude Level
A	2016/3/18 16:11	43°	72°	5.6
B	2016/4/2 5:50	62°	-155°	5.9
C	2016/3/10 3:24	-55°	-66°	5.2
D	2016/5/28 9:46	-57.74°	-27.15°	7.2
E	2016/9/15 5:57	13°	-85.75°	5.7
F	2016/5/28 4:43	5°	-38°	5.2

Table 3: random non earthquake event

| Index	Date	Lat	Lon	Index	Date	Lat	Lon
A	2016/5/27 19:32	-71°	-60°	**H**	2016/8/22 16:41	-24°	-49°
B	2016/8/13 2:18	54°	89°	**I**	2016/8/21 17:40	-15°	146°
C	2016/5/6 20:24	-45°	115°	**J**	2016/6/12 10:49	38°	11°
D	2016/4/10 21:59	36°	-103°	**K**	2016/7/25 3:31	34°	114°
E	2016/8/26 6:32	67°	-173°	**L**	2016/8/21 1:22	-63°	-71°
F	2016/6/8 20:35	-35°	15°	**M**	2016/5/20 9:51	68°	138°
G	2016/6/9 21:50	65°	46°	**N**	2016/4/21 14:11	18°	69°

The earthquake events and non-earthquake events have undergone data preprocessing, signal classification, and anomaly detection, and the results have been fitted. Figure 5 shows the final results. The time scale for all events is constrained between -180 days and 90 days. In the figure, the gray line represents the occurrence of earthquake events.

Figure 5a represents the results of anomaly detection for earthquake events. It is evident that there is a significant increase in the number of outlier points at the occurrence of earthquake events. Furthermore, these outlier points follow the trend of a Sigmoid function more closely. In contrast, the data during non-earthquake events, as shown in Figure 5b and Figure 5c, exhibit smaller increases in the fitted anomalies, which are closer to linear and thus more consistent with the probability of random event occurrences.

To qualitatively study the fitting quality, we will use the given R^2_{adj} in Equation 8 as a criterion to compare the linear fitting and Boltzmann function fitting for earthquake and non-earthquake events. Additionally, we will use the ratio of the Boltzmann function-fitted R^2_{adj} to the linear-fitted R^2_{adj}, as shown in Equation 8, and take its logarithm as a
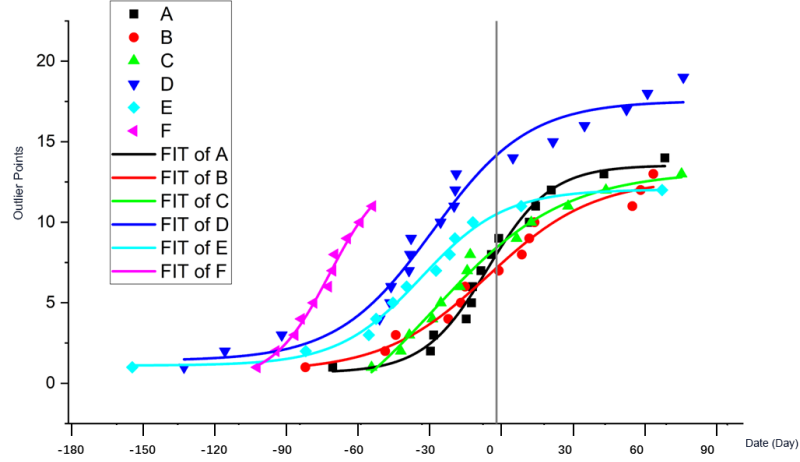
(a) Earthquake event anomaly detection
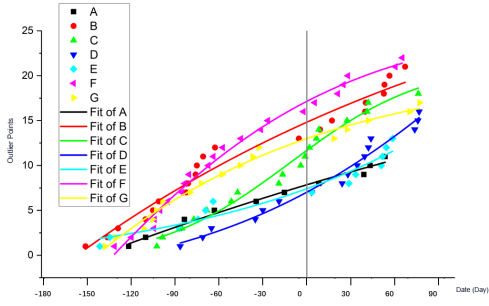
(b) Non-seismic events (part 1)
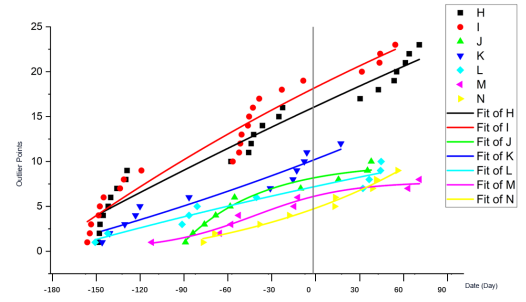
(c) Non-seismic events (part 2)

Figure 5: Random earthquake event analysis

basis. If this ratio is greater than 0, it indicates that the curve is closer to the Boltzmann function; if the ratio is less than 0, it indicates that the curve is closer to a linear function. Furthermore, we will use two parameters to assist in assessing the quality of the fitting results. x_0, which represents the time difference between the Boltzmann function-fitted center value and the occurrence of earthquake events, will be used to determine if the fitting result detects anomalies within a reasonable time. The larger the absolute value of x_0 is, the greater the difference between the detected anomaly and the earthquake occurrence time is, and the weaker the correlation. dx will characterize the growth trend of the Boltzmann function and can indicate the magnitude of anomaly amplitudes.

$$R_{ratio} = \log(\frac{R^2_{Boltzmann}}{R^2_{linear}}) \times 100 \tag{9}$$

According to Table 4 and Table 5, x_0 represents the number of days relative to the occurrence of the earthquake, where negative values indicate pre-earthquake and positive

values indicate post-earthquake. dx represents the rate of anomaly accumulation, and its value is converted to the ratio with the average value of all dx data. The closer the value is to 0, the more pronounced the rate of anomaly accumulation. Regarding earthquake events, we can observe from Table 4 that almost all x_0 values are close to the occurrence of the event, and R_{ratio} values are greater than 0, indicating that the trend of outlier points is closer to the Boltzmann function rather than linear growth. Additionally, the dx values are close to 0, indicating significant increases in anomaly rates occurring before the earthquake.

However, for the anomaly detection result indicators in the non-earthquake time period in Table 5, almost all values are unreasonable. For example, events A, E, H, I, J, and L are closer to the fitting result of random events. For events B, G, K, and N, the inflection points are too far from the events, making them meaningless in practical terms. Although the data point M may seem reasonable, it is actually due to a small value of $|A_1 - A_2|$, which is 7. This means that although there is some level of anomaly, the perturbation is small and can be ignored.

Table 4: Earthquake Event Anomaly Detection Result Metrics

Index	R_{ratio}	x_0	dx
A	5.08392	-4	0.06
B	1.034487	-4	0.10
C	3.362188	-30	0.12
D	1.896997	-30	0.09
E	8.724399	-36	0.08
F	0.664581	-79	0.05

Table 5: Non Earthquake Event Anomaly Detection Result Metrics

Index	R_{ratio}	x_0	dx	Index	R_{ratio}	x_0	dx
A	-0.14521	-2282	2.65	**H**	-0.25267	-2873	4.58
B	0.053099	-1614	1.46	**I**	-0.12263	-1995	2.19
C	0.105649	**-13**	0.22	**J**	1.982042	-537	0.25
D	0.655548	169	0.46	**K**	-0.94167	2728	2.32
E	-0.016	1232	0.74	**L**	-1.04415	-2358	2.83
F	1.071618	-207	0.44	**M**	0.679202	**-47**	**0.14**
G	1.401158	-456	0.63	**N**	0.450923	785	0.59

By utilizing the three parameters, R_{ratio} , x_0 and dx , along with the examples of earthquake events, we can compare the proposed method with other methods as shown in Table 6 based on the average values of these parameters. It can be observed that all the methods are able to fit the Sigmoid function well in terms of the fitting aspect. Compared to the EEMD (Barman et al., 2016), EMD(Yang et al., 2023), SSA (Rasheed et al., 2023), and Kalman

filtering (Kitahara et al., 2023) methods, the proposed method exhibits a noticeable rise in anomalies during the 30 days before the earthquake, while the other methods show an accumulation of anomalies that starts after a certain period following the earthquake. Even for the Kalman filtering and SSA methods, anomalies start accumulating after approximately 500 days post-earthquake, indicating that these two methods perform poorly in extracting pre-earthquake anomalies. Compared to the CEEMDAN (Chen et al., 2021) method, both the proposed method and the CEEMDAN method can detect anomalies within a certain period before the earthquake. However, only the proposed method can identify the most obvious process of anomaly accumulation. The other methods are also able to fit the Sigmoid function, but the accumulation of anomalies is less evident. Only the proposed method and the CEEMDAN method show clear anomalies in the processed waveforms, and only the proposed method and the CEEMDAN method can observe the accumulation of anomalies before the earthquake.

Table 6: Comparative Experiment

	R_{ratio}	x_0	dx (10^5)
Proposed Method	3.46	-30.58	18.25
CEEMDAN(Chen et al., 2021)	6.78	-20.73	378.73
EEMD(Barman et al., 2016)	3.72	140.20	3280.62
EMD(Yang et al., 2023)	6.90	69.48	416.32
SSA(Rasheed et al., 2023)	1.45	492.15	1843.93
Kalman(Kitahara et al., 2023)	1.145	512.85	1854.91

4. Conclusion

The study explores the detection of seismic anomalies through the analysis of geomagnetic data, acknowledging fluctuations in Earth's geomagnetic field strength reflected by processed Swarm data during earthquake development. Given the relative smallness of these variations compared to the overall field strength, it's necessary to decompose the geomagnetic field before performing anomaly detection. This approach is applied in the analysis of a seismic event that occurred in the Nicaragua region on September 15, 2016.

The study employs an enhanced signal decomposition technique based on Variational Mode Decomposition (VMD), specifically Particle Swarm Optimization-VMD (PSO-VMD), to mitigate the interference of the electromagnetic dynamic field. This method assists in investigating the association between electromagnetic data and earthquakes. Then we further use the Empirical-Cumulative-distribution-based Outlier Detection (ECOD) method for anomaly clustering of orbital numbers.

The methods presented in this study indentify a slow-fast-slow pattern in the accumulation of anomalies preceding the earthquake, a finding not easily discernable in the raw data. This research offers significant contributions to the field of seismic anomaly detection, demonstrating the potential of PSO-VMD and ECOD in unearthing valuable insights

from satellite electromagnetic signals. However, its complexity may necessitate additional time and resources when applied to large satellite datasets. Future research should aim to streamline the methodology and explore computational optimization techniques to facilitate its widespread application in earthquake precursor analysis.

5. Acknowledgement

References

Akhoondzadeh, M., Marchetti, D., 2022. Developing a Fuzzy Inference System Based on Multi-Sensor Data to Predict Powerful Earthquake Parameters. REMOTE SENSING 14, 3203. doi:`10.3390/rs14133203`.

Barman, C., Ghose, D., Sinha, B., Deb, A., 2016. Detection of earthquake induced radon precursors by Hilbert Huang Transform. JOURNAL OF APPLIED GEOPHYSICS 133, 123–131. doi:`10.1016/j.jappgeo.2016.08.004`.

Chen, W., Xiong, C., Yu, L., Lian, S., Ye, Z., 2021. Dynamic monitoring of an offshore jacket platform based on RTK-GNSS measurement by CF-CEEMDAN method. APPLIED OCEAN RESEARCH 115, 102844. doi:`10.1016/j.apor.2021.102844`.

De Santis, A., Balasis, G., Pavón-Carrasco, F.J., Cianchini, G., Mandea, M., 2017. Potential earthquake precursory pattern from space: The 2015 Nepal event as seen by magnetic Swarm satellites. Earth and Planetary Science Letters 461, 119–126. doi:`10.1016/j.epsl.2016.12.037`.

Dragomiretskiy, K., Zosso, D., 2014. Variational Mode Decomposition. IEEE Transactions on Signal Processing 62, 531–544. doi:`10.1109/TSP.2013.2288675`.

Fan, M., Zhu, K., De Santis, A., Marchetti, D., Cianchini, G., Piscini, A., He, X., Wen, J., Wang, T., Zhang, Y., Cheng, Y., 2022. Analysis of Swarm Satellite Magnetic Field Data for the 2015 Mw 7.8 Nepal Earthquake Based on Nonnegative Tensor Decomposition. IEEE Transactions on Geoscience and Remote Sensing 60, 1–19. doi:`10.1109/TGRS.2022.3195726`.

Friis-Christensen, E., Lühr, H., Hulot, G., 2006. Swarm: A constellation to study the Earth's magnetic field. Earth, Planets and Space 58, 351–358. doi:`10.1186/BF03351933`.

Kitahara, M., Kakiuchi, Y., Yang, Y., Nagayama, T., 2023. Adaptive Bayesian Filter with Data-Driven Sparse State Space Model for Seismic Response Estimation. doi:`10.2139/ssrn.4521593`.

Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., Chen, G., 2022. ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions. IEEE Transactions on Knowledge and Data Engineering , 1–1doi:`10.1109/TKDE.2022.3159580`.

Marchetti, D., De Santis, A., Campuzano, S.A., Zhu, K., Soldani, M., D'Arcangelo, S., Orlando, M., Wang, T., Cianchini, G., Di Mauro, D., Ippolito, A., Nardi, A., Sabbagh, D., Chen, W., He, X., Shen, X., Wen, J., Zhang, D., Zhang, H., Zhang, Y., Zeren, Z., 2022. Worldwide Statistical Correlation of Eight Years of Swarm Satellite Data with M5.5+ Earthquakes: New Hints about the Preseismic Phenomena from Space. Remote Sensing 14, 2649. doi:`10.3390/rs14112649`.

Molchanov, O., Rozhnoi, A., Solovieva, M., Akentieva, O., Berthelier, J.J., Parrot, M., Lefeuvre, F., Biagi, P.F., Castellana, L., Hayakawa, M., 2006. Global diagnostics of the ionospheric perturbations related to the seismic activity using the VLF radio signals collected on the DEMETER satellite. NATURAL HAZARDS AND EARTH SYSTEM SCIENCES 6, 745–753. doi:`10.5194/nhess-6-745-2006`.

Nemec, F., Santolik, O., Parrot, M., Berthelier, J.J., 2008. Spacecraft observations of electromagnetic perturbations connected with seismic activity. GEOPHYSICAL RESEARCH LETTERS 35, L05109. doi:`10.1029/2007GL032517`.

Oksum, E., Le, D., Vu, M., Hang, N., Pham, L., 2021. A novel approach based on the fast sigmoid function for interpretation of potential field data. Bollettino di Geofisica Teorica ed Applicata 62, 543–556. doi:`10.4430/bgta0348`.

Parrot, M., 2012. Statistical analysis of automatically detected ion density variations recorded by DEMETER and their relation to seismic activity. ANNALS OF GEOPHYSICS 55, 149–155. doi:`10.4401/ag-5270`.

Rasheed, A., Osama, M., Nikolopoulos, D., Rafique, M., 2023. Singular spectral and control chart analysis of soil radon and thoron time series for forecasting seismic activities. Journal of Atmospheric and Solar-Terrestrial Physics 249, 106108. doi:`10.1016/j.jastp.2023.106108`.

Sabaka, T.J., Tøffner-Clausen, L., Olsen, N., Finlay, C.C., 2018. A comprehensive model of Earth's magnetic field determined from 4 years of Swarm satellite observations. Earth, Planets and Space 70, 130. doi:`10.1186/s40623-018-0896-3`.

Shi, W., Peng, Z., Huang, Y., Zhang, G., Wang, C., 2023. An unsupervised anomaly detection approach for pre-seismic ionospheric total electron content. Measurement Science and Technology 34, 055101. doi:`10.1088/1361-6501/acb453`.

Wanpeng, F., Zhenhong, L.I., 2010. A novel hybrid PSO/simplex algorithm for determining earthquake source parameters using InSAR data. Progress in Geophysics 25, 1189–1196. doi:`10.3969/j.issn.1004-2903.2010.04.007`.

Xiong, P., Long, C., Zhou, H., Battiston, R., Zhang, X., Shen, X., 2020. Identification of Electromagnetic Pre-Earthquake Perturbations from the DEMETER Data by Machine Learning. REMOTE SENSING 12, 3643. doi:`10.3390/rs12213643`.

Yang, B.Y., Li, Z., Huang, J.P., Yang, X.M., Yin, H.C., Li, Z.Y., Lu, H.X., Li, W.J., Shen, X.H., Zeren, Z., Tan, Q., Zhou, N., 2023. EMD based statistical analysis of nighttime pre-earthquake ULF electric field disturbances observed by CSES. Frontiers in Astronomy and Space Sciences 9. doi:`10.3389/fspas.2022.1077592`.

Zhao, B., Qian, C., Yu, H., Liu, J., Maimaitusun, N., Yu, C., Zhang, X., Ma, Y., 2022. Preliminary Analysis of Ionospheric Anomalies before Strong Earthquakes in and around Mainland China. Atmosphere 13, 410. doi:`10.3390/atmos13030410`.

Zlotnicki, J., Le Mouël, J.L., Kanwar, R., Yvetot, P., Vargemezis, G., Menny, P., Fauquet, F., 2006. Ground-based electromagnetic studies combined with remote sensing based on Demeter mission: A way to monitor active faults and volcanoes. Planetary and Space Science 54, 541–557. doi:`10.1016/j.pss.2005.10.022`.