Analysis of Seismic Anomalies of the Jiuzhaigou Earthquake

Xiaoying Jiang, Xiangzeng Kong*, Gongde Guo
College of Mathematics and Informatics, Fujian Normal University, Fuzhou, China
xzkong_fjnu@163.com

Abstract. On August 8, 2017, a devastating mega-earthquake of magnitude 7.0 struck the Jiuzhaigou County, Sichuan Province, China, resulting in significant casualties and property damage. Actually, before an earthquake occurs, a lot of abnormal information often appears suddenly. For example, the thermal infrared radiation suddenly increased in a short time. However, it is hard to identify and analyze valuable anomalous information from large-scale data. To solve this problem, this paper proposes a random walk method to analyze the outgoing long-wave radiation (OLR) anomaly before an earthquake. The National Oceanic and Atmospheric Administration (NOAA) data are used to process and analyze the OLR data before and after the Jiuzhaigou earthquake. In order to prove the effectiveness of our algorithm, we use this method to analyze the OLR data of Jiuzhaigou region from 2006 to 2017, and compare the OLR anomaly difference between the year of earthquake occurrence and a normal year. Experimental results show that there are obvious abnormal changes in OLR data in the year of earthquake. In order to further verify the applicability of this algorithm, we also used this algorithm to analyze the OLR data of another 10 earthquakes with magnitude 5.0 or above in western China in 2017. The results show that the proposed algorithm can extract OLR anomaly information in most earthquakes. Therefore, our algorithm can effectively extract the pre-earthquake abnormal information in OLR data.

1. Introduction
On August 8, 2017, a magnitude 7.0 earthquake occurred in Jiuzhaigou, Sichuan Province, China, causing heavy casualties and property losses. What's more, China is a country with many earthquakes. In particular, in the western part of China, many earthquakes occur each year, posing a serious threat to people's lives and property in the western region. The identification and analysis of abnormal information before the earthquake has attracted the attention of more and more researchers. And satellite remote sensing technology is widely used by scientists to research the relationship between thermal infrared remote sensing and earthquakes. They hope to excavate useful pre-earthquake anomaly information and achieve more accurate prediction earthquake. They hope to improve the reliability of seismic analysis and provide more possibilities for predicting earthquakes by more effectively identifying and analyzing useful pre-earthquake anomaly information.

Kong et al. [1] analyzed seismic anomalies within outgoing long-wave radiation (OLR) data observed by satellites from 2006 to 2013 for Wenchuan and Lushan earthquakes and four comparative study areas. Zhou et al. [2] detected the OLR data and total electron content (TEC) data by using the standard deviation threshold method and the quaternary-decay method in the Nepalese earthquake in 2015. They found that in the vicinity of the epicenter within 3 days of the earthquake, satellites remote sensing OLR abnormal increase phenomenon. At the same time, the TEC in the ionosphere near the epicenter shows significant positive anomalies and magnetic conjugation. Reference [3] presented a statistical study upon the geomagnetic data observed at Kakioka station, Japan, during 2001-2010.
They found that the ultralow frequency (ULF) seismomagnetic phenomena at Kakioka clearly contain precursory information and have a possibility of improving the forecasting of large earthquakes. Xiong et al. [4] proposed a statistical analysis method based on the Robust Satellite data analysis technique to detect seismic anomalies within the NOAA OLR dataset about 3376 earthquake cases from September 01, 2007 to May 23, 2015 based on spatial/temporal continuity analysis. However, although OLR data has been used as a new index of earthquake precursors anomalies, most of the existing achievements are mainly based on the research methods of geography, and the analysis of the results is mainly based on manual observation. Therefore, the experimental results are directly influenced by experts in the field and the influence of experience has great uncertainty on the judgment of the related anomalies before the earthquake.

The application of data mining techniques in the field of earthquake anomaly analysis has been explored by more and more scholars. For example, Xiong et al. [5] used the wavelet transformations and spatial/temporal continuity analysis of wavelet maxima to analyze the continuous OLR. The experimental results show that there are singularities in OLR data that correspond to seismic precursors. Wu et al. [6] studied the surface latent heat flux (SLHF), thermal infrared radiation (TIR), outgoing longwave radiation (OLR), diurnal temperature range (DTR), atmospheric temperature, and skin temperature, for GEOSS-based earthquake anomaly recognition (EAR) by analyzing the M7.3 Yutian earthquake occurred in 2008, the M8.0 Wenchuan earthquake occurred in 2008, and the M7.1 Christchurch earthquake occurred in 2010 in New Zealand. The results showed that the obtained composite thermal anomaly has a significant effect on earthquake prediction. However, most of the methods can only find a very small decrease or increase in the signal. For example, the wavelet-based data mining technique was able to discover singularities from OLR data. However, it could not provide the measurements of change degree of OLR data. In addition, most of the experimental data in these studies are too limited. In this paper, we will analyze the OLR data from 2006 to 2017 to exploring the abnormal changes before the earthquake.

Our main contributions are as follows: we propose a data mining algorithm called random walk method, which is based on Martingale theory [7]. This article applies data mining technology to the field of seismic research. We analyze the OLR data before and after the Jiuzhaigou earthquake. The comparing experiences have been done using the OLR data from 2006 to 2017 to look for the correlation between OLR data and earthquakes. And the experimental results show that there are obvious abnormal changes in OLR data in the year of earthquake. We also used this algorithm to analyze the OLR data of another 10 earthquakes with magnitude 5.0 or above in western China in 2017. The results show that the proposed algorithm can extract OLR anomaly information in most earthquakes. Therefore, the method proposed in this paper is suitable for studying OLR anomalies before earthquakes. Meanwhile, it has an important reference for applying the computer technology in the field of geography science.

2. Data and methodology

2.1. Related data sources

The OLR data of 2.5° longitude × 2.5° latitude grids in this paper is provided by the NOAA of the United States. OLR data includes daytime data and nighttime data. In order to minimize the interference of daytime human activities and environmental factors such as sunlight, this paper uses nighttime data [8]. All the selected earthquakes (11 in total) are those with magnitudes ≥ 5.0 and occurred in western China in 2017. The details of these earthquake events are shown in Table 1.
Table 1. List of the Selected Earthquake Events

| Date       | Latitude (° N) | Longitude (° E) | Depth (km) | Magnitude (JMA) | Location     |
|------------|----------------|-----------------|------------|-----------------|--------------|
| 17-02-01   | 30.67          | 83.34           | 8          | 5.0             | Zhongba      |
| 17-03-27   | 25.89          | 99.80           | 12         | 5.1             | Yangbi       |
| 17-05-11   | 37.58          | 75.25           | 8          | 5.5             | Tashkurgan   |
| 17-08-08   | 33.20          | 103.82          | 20         | 7.0             | Jiuzhaigou   |
| 17-08-09   | 44.27          | 82.89           | 11         | 6.6             | Jinghe       |
| 17-09-16   | 42.11          | 83.43           | 6          | 5.7             | Kuche        |
| 17-09-30   | 32.27          | 105.00          | 13         | 5.4             | Qingchuan    |
| 17-11-18   | 29.75          | 95.02           | 10         | 6.9             | Milin        |
| 17-11-18   | 29.88          | 94.92           | 6          | 5.0             | Bayi         |
| 17-12-07   | 35.69          | 77.46           | 87         | 5.2             | Yecheng      |
| 17-12-20   | 29.88          | 95.08           | 6          | 5.0             | Bayi         |

2.2. Random walk method

This paper proposes a random walk algorithm to identify and analyze the thermal infrared anomaly characteristics through a certain signal amplification. And Martingale theory [7] is applied to measure the changing degree of the OLR data.

In this paper, dataset }{1, zzzZ 11− = represents historical data and \( z_i \) represents the currently calculated data point. Obviously, when the activity of the geological plate is relatively stable, the sample distribution in \( Z \) should be maintained at a relatively stable level with some similar characteristics between the samples. Therefore, the data set \( Z \cup z_i \) can be treated as a time series.

The first step is use a fixed size window to determine whether the value of the current data point is greater than or equal to the value of the previous data point. The size of this window is marked as \( ws \).

If the value of the current data point is greater than or equal to the value of the previous data point, the direction of the current data point is defined as leftward; otherwise, the direction of the current data point is defined as rightward. This is based on the principle of equal probability distribution of all data points.

Then we select \( ws \) data points before the current data point. The probability distributions of the current data point direction to the left or right are respectively obtained by the probability distribution formula (1):

\[
\begin{align*}
\text{rw}(ws) &= p^{\frac{n+i}{2}} (1-p)^{\frac{n-i}{2}} \times C(n, \frac{n+i}{2}) \\
\end{align*}
\]  

(1)

where \( n = ws, i \in [-n, n], p = 0.5, \) and \( C(n, \frac{n+i}{2}) \) represents a combination in mathematics.

Then we should calculate the Martingale value according to the formula (2):

\[
M_{n,k}^{(c)} = \prod_{i=1}^{n} (e_{i,k}^{c_{i-1}})
\]

(2)
where $\varepsilon \in [0,1]$, $\hat{p}_{i,k}$ is computed according to the formula (4).

The k-means clustering method is adopted to analyze the data, because OLR data is a time series. The strangeness value $s_i$ can be defined as

$$s_i = s(Z, z_i) = \|z_i - m\|$$

where $m$ is the average of all the data points and $\|\|$ represents the Euclidean distance.

$\hat{p}_{i,k}$ is computed according to the following formula:

$$\hat{p}_{i,k}(Z \cup \{z_n\}, \theta_n) = \frac{\#\{j | s_j > s_i\} + \theta_{i,k} \#\{j | s_j = s_i\}}{i}$$

(4)

where $\#\{\}$ is a function that returns the number of samples that satisfy a given condition. For example, $\#\{j | s_j > s_i\}$ is the number of $j$ satisfying $s_j > s_i$, where $s_j$ is the strangeness measure mentioned in function (3). $\theta_{i,k}$ is a random number which is selected from the interval $[0,1]$, $i = 1, 2, \cdots n$, and $s_j$ is the strangeness measurement for $z_j, j = 1, 2, \cdots, i$ with the initial Martingale value $M_0^\varepsilon = 1$.

We use a CD value to reflect the change degree of the current data point as follows:

$$CD_n^{(\varepsilon)} = \frac{\sum_{k=1}^{100} M_{n,k}^{(\varepsilon)}}{100}$$

(5)

where 100 is the number of Martingale value which are used to calculate the CD value. According to [9], we set $\varepsilon = 0.82$ and initialize $CD_0 = 1$. CD can reflect the average change degree of the data points before the current data point. Therefore, the proposed method can ignore some isolated false changes which may be caused by instrumental malfunctions or other faults.

If the computed $CD_n$ value is very high and exceeds the threshold $h$, this could correspond to an abnormal condition. The condition $CD_n \geq h$ can be used to determine whether a change has been detected for suitable values of $h$. Therefore, the stopping rule is as shown in expression (6):

$$CD_n \geq h$$

(6)

the proposed algorithm should be restart when a abnormal condition is found. Then continue to calculate the remaining features of random walk at each point to obtain a new feature sequence and calculate the remaining points abnormality.

3. Experimental results and analysis
The OLR data used covers the area of 28.75°N–36.25°N and 100°E–107.5°E. And the Jiuzhaigou earthquake’s epicenter is in the cell 5 which covers latitude 31.25˚N–33.75˚N, and longitude 102.5˚E–105˚E. The data are the night-time OLR data from March 1, 2017 to February 28, 2018. The red vertical line in the following figures indicates the time of the earthquake. Experimental results as shown in Fig.1. The CD value began to increase significantly two months before the earthquake, and a peak appeared during this period. A few days before and after the earthquake, the anomaly suddenly decreased, but another peak appeared again within two months after the earthquake. And the maximum value appeared around September 20. On September 30, a magnitude 5.4 earthquake occurred in Qingchuan County, Sichuan Province. The maximum peak appeared just about 10 days
before the earthquake. The author speculated that the second peak appeared related to the Qingchuan earthquake.

![Figure 1](image1.png)

**Figure 1. CD value for cell 5 from March 1, 2017 to February 28, 2018.**

![Figure 2](image2.png)

**Figure 2. CD values of cells 1, 2, 3, 5, 6, 7, 9 in the Jiuzhaigou area, from March 1, 2017 to February 28, 2018.**

Fig.2 shows the comparative analysis of cells 1, 2, 3, 5, 6, 7 and 9 in the Jiuzhaigou area from March 1, 2017 to February 28, 2018. The CD values are very small in cells 4 and 8 before the earthquake time, so they are not presented in Fig.2. It can be seen from Fig.2 that even though the maximum values of some cells did not occur at the time of the earthquake or before the earthquake occurred, the OLR of all cells increased significantly before and after the earthquake. The maximum values of the seven cells in Fig.2 are greater than 200, and even the maximums of five cells exceeds 300. For example, the maximum value of cell 5 is 556.8. The results show that the OLR data showed obvious anomalies before and after the earthquake. And our algorithm can effectively extract the abnormal information in the OLR data.

We also study the OLR changes in the Jiuzhaigou area from 2006 to 2016 to compare the OLR difference between the earthquake occurrence year and the normal year, and explore the relationship between OLR data and earthquake. The results are shown in the Table 2. In order to minimize the error caused by accidental factors, the maximum value here is the result after subtracting the average of 12 years. The average CD maximum value of the Jiuzhaigou area is 287.8 from 2006 to 2017, and the CD maximum value of 2017 when the Jiuzhaigou earthquake occurred is 468.2, so it is 1.63 times the average CD maximum value from 2006 to 2017 year. As can be seen from Table 2, the CD value in 2007 is larger than in other years, except for 2006 and 2007. A larger CD value in 2006 might be due to the drought in Sichuan Province which occurred just once in 100 years. Like the earthquakes, the drought could cause the Earth’s skin temperature to change which could in turn affect the OLR values. As for the reason why the CD value in 2007 is abnormally large, it has not yet been determined, and it might be a false positive.

![Table 2](image3.png)

**Table 2. Maximum CD values of the cell 5 in Jiuzhaigou area from 2006 to 2017.**

| Year | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 average |
|------|------|------|------|------|------|------|------|------|------|------|------|---------------|
| Max  | 540.6| 883.5| 2.1  | 169.3| 158.7| 178.7| 371.7| 287.2| 164.0| 193.4| 36.9| 468.2         |
| 287.8|  |  |  |  |  |  |  |  |  |  |  |
In order to further prove the applicability of the proposed algorithm, we apply the algorithm to another 10 earthquakes of magnitude 5 or higher in western China in 2017. We found that seven of the earthquake cases detected OLR anomalies before the earthquake, and only three earthquake cases detected almost no abnormal OLR changes. The applicability of this algorithm is 70% and the false negative rate is 30%.

4. Discussion and conclusion
This paper proposed a random walk algorithm to detect pre-earthquake OLR anomalies in Jiuzhaigou area and another 10 earthquakes which magnitude $\geq 5.0$ and occurred in western China during 2017. Generally, when an earthquake occurs, the closer the earthquake is, the more obvious the abnormality is. Although the anomalies from the epicenter are not necessarily the largest, there must be significant OLR anomalies in and around the epicenter before and after the earthquake. And there are several years where the anomalies are greater than that of the earthquake year. This might be due to other factors such as high temperatures and droughts that affect the thermal infrared values. Therefore, our method can effectively detect pre-earthquake anomalies and is suitable for studying pre-earthquake anomalies in most earthquake cases in western China. In the study, we also found that almost no OLR anomalies were detected in some earthquake years. There are also anomalies detected by earthquakes with small magnitudes that are larger than those with large magnitudes, so the relationship between magnitude and anomaly needs to be further explored.

Acknowledgments
This work was supported by the National Natural Science Foundation of China under Grant No. 41601477, the Leading Project in Fujian Province of China under Grant No.2015Y0054 and NOAA for making OLR data available for various research communities and anonymous reviewers for their constructive suggestions.

References
[1] KONG X, BI Y and GLASS D H 2015 Detecting seismic anomalies in outgoing long-wave radiation data IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing vol 8 no.2 pp 649-660.
[2] ZHOU Q B,TANG J and CNNC 2016 Analysis of Anomalies of Pre-Earthquake Based on OLR and Ionosphere Data of 2015 Mw 7.8 Earthquake in Nepal Journal of Geomatics Science and Technology vol 33 pp 458-463.
[3] Han P , Hattori K , Hirokawa M , et al. 2014 Statistical analysis of ULF seismomagnetic phenomena at Kakioka, Japan, during 2001-2010 Journal of Geophysical Research: Space Physics vol 119 no.6 pp 4998-5011.
[4] Xiong P and Shen X H 2016 Outgoing Longwave Radiation anomalies analysis associated with different types of seismic activity Advances in Space Research vol 59 no.5.
[5] Xiong P , Bi Y and Shen X 2009 Study of Outgoing Longwave Radiation Anomalies Associated with Two Earthquakes in China Using Wavelet Maxima Hybrid Artificial Intelligence Systems, International Conference, Hais, Salamanca, Spain, June. DBLP.
[6] Wu L , Qin K , Liu S 2012 GEOSS-Based Thermal Parameters Analysis for Earthquake Anomaly Recognition Proceedings of the IEEE vol 100 no.10 pp 2891-2907.
[7] HO S S and WECHSLER H 2010 A Martingale framework for detecting changes in data streams by testing exchangeability vol 32 pp 2113-2127
[8] ZHAO J 2011 Identification and Analysis of Pre-earthquake Thermal Anomalies—take six earthquakes in western China for example Intelligent Information Technology Application Association.Proceedings of 2011 International Conference on Ecological Protection of Lakes-Wetlands-Watershed and Application of 3S Technology(EPLWW3S 2011 V3)
[9] VORK V, NOURETDINOV I and GAMMERMAN A 2003 Testing exchangeability on-line Proc of the 12th International Conference on Machine Learning San Francisco: Morgan Kaufmann pp 768-775.