Negentropy anomaly analysis of the borehole strain associated with the Ms 8.0 Wenchuan earthquake

Kaiguang Zhu\textsuperscript{1,2}, Zining Yu\textsuperscript{1,2}, Chengquan Chi\textsuperscript{1,2}, Mengxuan Fan\textsuperscript{1,2} and Kaiyan Li\textsuperscript{1,2}

\textsuperscript{1} College of Instrumentation and Electrical Engineering, Jilin University, China
\textsuperscript{2} Key Laboratory of Geo-Exploration Instrumentation, Ministry of Education, Jilin University, China

Abstract: A large earthquake of 8.0 magnitude occurred on 12 May 2012, 14:28 UTC, with the epicenter in Wenchuan. To investigate the pre-earthquake anomalous strain changes, negentropy is introduced to borehole strain data at Guza station, approximated by skewness and kurtosis revealing the non-Gaussianity of recorded fluctuations. We separate the negentropy anomalies from the background by Otsu’s method and accumulate the anomaly frequency in different scales. The results show the long-scale cumulative frequency of negentropy anomalies follows a sigmoid behavior, while the inflection point of the fitting curve is close to the occurrence of the earthquake. For the short-scale analysis before the earthquake, there are two cumulative acceleration phases corresponding to the two crustal stress releases, indicating the preparation process of the Wenchuan earthquake. We consider that negentropy exhibits potential for the analysis of earthquake precursor anomalies.

Keywords: Negentropy anomaly, Otsu’s thresholding, Cumulative frequency, Wenchuan earthquake

1. Introduction

Changes in crustal deformation fields over time have preceded at least some large earthquakes (Thatcher, W. et al., 1981), such as the 2011 Tohoku earthquake (Hitoshi Hirose, 2011) and the Ruisui earthquake in Taiwan in 2013 (Canitano A. et al., 2015). Borehole strain data, which record the direct crustal changes, provide an opportunity to investigate preparation process prior to earthquakes (Linde et al., 1996, Hsu, Y.-J. et al. 2015).

Various methods are used in identifying borehole strain anomalies based on large amount of monitoring data. Experienced scholars extract borehole strain anomalies by discriminating patterns of waveform behaviors compared to those during the normal stage (Johnston M.J.S. Johnston et al., 2006, Chi S. L. et al., 2014). In the time domain, Qiu Z. H. et al. (2010) identified abnormal strain
changes by overrun rate and wavelet decomposition for the Wenchuan earthquake. While in the
frequency domain, Qi L. et al. (2011) thought the signal with a period of 10 to 60 minutes might be
anomalies through S-transform compared with the background signal. In addition, statistical methods
are proved effective in distinguishing borehole strain anomalies with regard to large earthquakes,
such as principal component analysis (Zhu K. G. et al., 2018) and correlation coefficients along with
the consistency relation (Kong X. et al., 2018).

The probability distribution function (PDF) of observation data is also an informative way of
extracting potential anomalies contained in earthquake generation processes. P. Manshour et al.
(2009) extracted variance anomalies of the probability density of the Earth’s vertical velocity
increments, and successfully found a pronounced transition from Gaussian to non-Gaussian prior to
12 moderate and large earthquakes. Before the Wenchuan earthquake, the high-frequency fluid
observational data deviated from Gaussian distributions at 16 water level and 14 water temperature
stations (Sun X. L. et al., 2016).

Rather than the whole PDF, often its moments are utilized, moments may be estimated quite
reliably from relatively small amounts of data (Sattin, F. et al., 2009). In 2016, Chen H. J. et al.
applied skewness and kurtosis (the third- and fourth-order moments) of the geoelectric data to pick
up non-Gaussian anomalies to predict impending large earthquakes in Taiwan. On the other hand, for
turbulent or disordered systems, the non-Gaussian distribution of time series in skewness-kurtosis
domain attracts attention. Observation data series from various fields of geophysics indicate that a
parabolic relation between skewness and kurtosis holds in fields such as seismology (M. Cristelli, et
al., 2012), oceanography (Sura, P. et al., 2008) and atmospheric science (A. Maurizi, 2006).

Hence, it is implied that possible precursor anomalies lead to an increase in disordered
components of observation data deviate from Gaussian distribution during earthquake preparation
processes. K. Eftaxias et al. (2008) proved that the pre-catastrophic stage could break the persistency
and high organization of the electromagnetic field through studying fractional-Brownian-motion-type
model using laboratory and field experimental electromagnetic data. In view of Lévy flight and
Gaussian processes, Lévy flight mechanism prevents the organization of the critical state to be
completed before earthquakes, since the long scales are cut-off due to the Gaussian mechanism (S.M.
Potirakis et al., 2019).
Entropy can serve as a measure of the unknown external energy flow into the seismic system (Akopian, S. T., 2014). K. Karamanos et al. (2006, 2005) quantified and visualized temporal changes of the complexity by approximate entropy, they claimed significant complexity decrease and accession at the tail of the preseismic electromagnetic emission could be diagnostic tools for the impending earthquake. Yukio Ohsawa (2018) detected earthquake activation precursors by studying the regional seismic information entropy on earthquake catalog. Angelo De Santis et al. (2011) recalled the Gutenberg - Richter law and considered the negative logarithm of b-value is the entropy of the magnitude frequency of earthquake occurrence associated with two earthquakes in Italy.

Negentropy definition is based on the entropy and it is also widely used to detect non-Gaussian features. Yue Li et al. (2018) proposed an arrival-time picking method based on negentropy for microseismic data. In this study, the negentropy is applied to borehole strain at Guza station associated with the Wenchuan earthquake, approximated by skewness and kurtosis revealing the non-Gaussianity of borehole fluctuations. Subsequently we study the extracted negentropy anomalies in different scales to investigate correlations with crustal deformation.

2. Observation

YRY-4 borehole strainmeters, which are designed to record continuous deformation occurring over periods of minutes to years, have been deployed at depths of dozens of metres at more than 40 terrain-sensitive locations within China. These strainmeters are capable of resolving strain changes of less than one-billionth. The data sampling rate is once per minute.

The study period is from January 1, 2007, to June 30, 2009. The object of the study is the Wenchuan Ms 8.0 earthquake in the region, which is shown in Fig. 1.
Fig. 1. Location map showing the epicentre of the Wenchuan earthquake epicentre and the Guza station. The Wenchuan earthquake occurred at 14:28:04 on May 12, 2008 (UTC+8). The magnitude of the earthquake was Ms 8.0, and the focal depth was approximately 14 km. The epicentre was located in Wenchuan County, Sichuan Province, at 31.01°N, 103.42°E.

Because the four gauges of the YRY-4 borehole strainmeter are arranged at 45° intervals, this design has improved its self-consistency. This arrangement produces four observation values: \(S_i\), \(i=1, 2, 3, 4\) (Qiu et al., 2013a). The self-consistency as shown in equation (1), which can be used to test the reliability of the data among the four gauges.

\[
S_1 + S_3 = S_2 + S_4
\]  

(1)

In practical application, the higher the correlation between both sides of the equation (1), the more reliable the data. Generally, we use \(S_a\) for the areal strain in describing the subsurface strain state of the observation area as shown in equation 2 (Qiu et al., 2013a):

\[
S_a = (S_1 + S_2 + S_3 + S_4) / 2
\]  

(2)

The borehole strain is highly consistent among the four gauges at the Guza station (Qiu, et al., 2009), as shown in Fig. 2.

Fig. 2. Self-consistency of the borehole strain at Guza from January 1, 2007, to June 30, 2009.

We first calculate the difference in the data because in the borehole strain data, the change in the strain is a concern (Li Jinwu et al., 2014). We then remove the components associated with the solid tide frequencies through a daily harmonic analysis. The remaining high-frequency signals are shown in Fig. 3.
Fig. 3. High-frequency areal strain after removing harmonic information at Guza from January 1, 2007 to June 30, 2009.

3. Methodology

3.1 Negentropy and non-Gaussianity

The entropy-based negentropy is a statistically justified measure of non-Gaussianity (A. Hyvarinen, et al., 2000). The entropy of a random variable \( X = \{x_1, x_2, \cdots, x_i, \cdots \} \) is defined as

\[
H(X) = - \sum_i P(X = x_i) \log P(X = x_i)
\]

where \( P \) is the probability density function. Entropy measures the randomness of a random variable.

The Gaussian random variable has the largest entropy of all other random variables with equal variance (T.M. Cover et al., 1991). The definition of negentropy is given by

\[
J(X) = H(X_{gauss}) - H(X)
\]

in which \( X_{gauss} \) is a Gaussian random variable with the same mean and covariance matrix as \( X \). The entropy of a Gaussian random variable can be estimated by

\[
H(X_{gauss}) = \frac{1}{2} \log \left| \det \Sigma \right| + \frac{n}{2} (1 + \log 2\pi)
\]

where \( n \) is the dimension of the variable, and \( \Sigma \) is its covariance matrix.

However, the theoretical calculation of negentropy also depends on the prior probability density of random variables and other information that is difficult to determine accurately. In practical applications, higher order statistics (HOS) and density polynomial expansion are usually used to approximate one-dimensional negentropy (Jones and Sibson, 1987). The approximation results are as follows:

\[
J(x) \approx \frac{1}{12} skewness^2(X) + \frac{1}{48} kurtosis^2(X)
\]
This definition suggests that any deviation from a Gaussian distribution will increase the negentropy $J(x)$. The skewness and kurtosis are the third- and fourth-order statistics, respectively, which are defined as

\[
\text{skewness}(X) = \frac{\mu_3}{\sigma^3} = \frac{E[(X - \mu)^3]}{E[(X - \mu)^3]}^{1/3}
\]  

and

\[
\text{kurtosis}(X) = \frac{\mu_4}{\sigma^4} = \frac{E[(X - \mu)^4]}{E[(X - \mu)^4]} - 3
\]

where $\mu$ is the mean of $X$ and $\sigma$ is the standard deviation of $X$. Skewness is a measure of asymmetry in a PDF. A symmetric distribution has zero skewness. Kurtosis is a measure of the heaviness of the tails. Distributions that are more outlier-prone than the normal distribution have kurtosis values greater than zero.

Moreover, the relation between the skewness and kurtosis is universal, they approximately align along a quadratic curve (Sattin, F., et al., 2009):

\[
\text{kurtosis}(X) = A \cdot \text{skewness}^2(X) + B.
\]  

Here we calculate the normalized skewness and kurtosis in the study period, so equation (9) can be derived into

\[
\text{kurtosis}(X) = A \cdot (\text{skewness}^2(X) - 1)
\]

indicating the test day is super-Gaussian when the skewness is outside the range (-1,1).

This relation is trivial in a Gaussian fluctuating system; it reduces to a fixed mass around zero (skewness= 0 and kurtosis= 0). In a turbulent environment where fluctuating quantities obey non-Gaussian statistics, the moments obey the above relation.

3.2 Otsu’s thresholding method

To solve the negentropy anomaly detection problem, we designed a simple thresholding hypothesis test using the Otsu method (Otsu, 1979) that provides an optimal separation between background and seismic-related activities. For any given value $k$, we can separate the previously calculated $J(x)$, as shown in equation (6), into the following two classes:
\[ C_0(k) = \{J(x) \leq k\}, \]
\[ C_1(k) = \{J(x) > k\}. \]  

(10)

Using these classes, the weighted average value \( \mu_r(x) \) of \( J(x) \) can be expressed as follows:

\[ \mu_r(x) = \lambda_o(k)\mu_o(x;k) + \lambda_i(k)\mu_i(x;k), \]
\[ \lambda_o(k) + \lambda_i(k) = 1. \]  

(11)

where \( \mu_o(x;k), \mu_i(x;k) \) are the mean values of the class \( C_i(k) \), \( i=1, 2 \), and \( \lambda_i(k) \) is the percentage of points belonging into each class. Following the thresholding scheme of Otsu (1979), we define the following cost function:

\[ \sigma_g^2(k) = \lambda_o(k)(\mu_o(x;k) - \mu_r(x;k))^2 + \lambda_i(k)(\mu_i(x;k) - \mu_r(x;k))^2 \]
\[ = \lambda_o(k)\lambda_i(k)(\mu_i(x;k) - \mu_o(x;k))^2 \]  

(12)

where \( \sigma_g^2 \) is the within-class variance of negentropy. Then, by finding the \( k^* \) value searching for \( k \) when \( \sigma_g^2 \) becomes the maximum

\[ k^* = \arg \max_k \sigma_g^2(k) \]  

(13)

the optimal value \( k^* \) here separates the background set and anomaly set.

In this test, our initial assumption is that the sliding window is composed of a Gaussian signal of non-seismic-related activities. When our test negentropy exceeds the critical value \( k^* \), this initial hypothesis is not valid, and the alternative is true, indicating the presence of a negentropy anomaly within the window.

4. Results

According to the empirical hypothesis that geophysical signals deviate from the Gaussian distribution when they record seismic-related activities, and based on the results of previous studies, we conduct the following investigation.

4.1 Extracting negentropy anomalies

As the negentropy is calculated using a 2-hour long sliding window, we assume that it reaches the maximum values when the time window contains anomalies from seismic-related activities. The negentropy during the study period is shown in Fig. 4.
Fig. 4. Negentropy at Guza from January 1, 2007, to June 30, 2009. The red dotted, horizontal line is the optimal threshold \( k^* \) calculated by Otsu’s method.

The within-class variance \( \sigma^2_W \) and negentropy value distribution are compared in Fig. 5. According to equations (10) to (13), when \( k^*=1.1130 \), \( \sigma^2_W \) reaches its maximum. Therefore, the negentropies were separated by \( k^* \) into the quasi-Gaussian background and non-Gaussian anomalies from 2007 to 2009. Otsu threshold \( k^* \) here is consistent with the accuracy of the negentropy and the strain data. The YRY-4 borehole strainmeter has a measurement accuracy of \( 10^{-9} \), however, we usually cutoff four digits after the decimal point in practical calculations.

In the skewness-kurtosis domain, the statistical relationship of the borehole areal strain is consistent with parabolic behaviour as described in equation (9)(Fig. 6(a)), verifying that the turbulent system of borehole strain is significantly non-Gaussian during the study period. However, the extracted negentropy anomalies are clustered strongly on the left side of the parabola, which could be a signature of crustal deformation related to the earthquake. Here, there are four points on the right side; one occurred in early 2007, and the others occurred after the earthquake. Therefore, we will not discuss them in the following.
Fig. 6. Negentropy distributions in the skewness-kurtosis domain in (a) January 1, 2007, to June 30, 2009, and (b) four shorter periods before the earthquake. Red denotes that the negentropy is greater than $k^*$, and grey indicates that the negentropy less than $k^*$. The blue curve is the quadratic fit with the a 95% confidence.

In addition, as shown in Fig. 6(b), at times far from the earthquake, the negentropy distribution is basically Gaussian in the skewness-kurtosis domain. However, at times closer to the earthquake, the distribution gradually begins to show non-Gaussianity, with more negentropy anomalies appearing on the left side of the parabola. While in 2008, almost all of the negentropy present left-skewed.

These phenomena prompt us to study the origin of this left-skewed distribution and its possible correspondence with the seismogenic process.

4.2 Negentropy anomaly frequency accumulation

The transition of negentropy anomalies in the skewness-kurtosis domain is quantified as the change of the anomaly frequency per unit time through a logarithmic-linear model. Logarithmic-linear models are often used of interest to estimate the expected frequency of the response variable at the original scale for a new set of covariate values, such as in the famous Gutenberg-Richer law, in which a linear relationship exists between the logarithm of the cumulative number of seismic events of magnitude $M$ or greater versus the magnitude $M$ (Gutenberg and Richter, 1954).

The logarithmic-linear regression model is proposed as
\[ \log N = \beta_1 \times k_j + \beta_0 + \varepsilon \]  

(14)

where \( k_j \) takes different threshold values according to the \( J(x) \) values, \( N \) is the number of occurrences in which \( J \) is greater than or equal to the threshold \( k_j \), \( \beta_0 \) and \( \beta_1 \) are the regression coefficients, where a lower slope \( \beta_1 \) indicates that there are more higher \( J \) values, implying there are more anomalies at that moment, and \( \varepsilon \) is the random error that represents the model uncertainty.

We use the logarithmic-linear model to solve the relationship between the negentropy anomaly frequency and different thresholds each day using the ordinary least squares method (OLS) method. Afterwards, an optimal threshold \( k^* \), calculated by the Ostu method, is chosen for all models, where

\[ N_j(t) = \exp(\beta_1(t) \times k^* + \beta_0(t) + \varepsilon(t)) \]  

(15)

and the \( N_j(t) \) under the threshold \( k^* \) is shown in Fig. 7. The model theoretically solves the problem of selecting the length of the time window. In addition, the estimated \( N_j(t) \) is considered as the expected frequency of anomalies.

The goodness of fit for each logarithmic-linear model was evaluated using analysis of

\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{N}_i - \tilde{N}_i)^2}{\sum_{i=1}^{n} (N_i - \bar{N})^2} \]  

(16)

and the root-mean-squared error

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{N}_i - \tilde{N}_i)^2}{n}} \]  

(17)

The \( R^2 \) and \( RMSE \) values in the study period (912 days) show that the logarithmic-linear relationship can explain the relationship between the negentropy anomaly frequency and different thresholds. The mean of \( R^2 \) is 0.9695, which is close to 1, and the variance of \( R^2 \) is 0.0435. The mean and variance of the \( RMSE \) are also small (0.1098 and 0.1301, respectively).

In general, accumulated value of a typical random process usually has a linear increase. The negentropy cumulative frequency of the study period is shown in Fig. 8. We not only do a long-scale
analysis of the whole earthquake process, but also carry out a short-scale analysis of the
pre-earthquake process.

For the entire earthquake process, a two-month long sliding window is selected for
accumulation as shown in Fig. 8(a). Beginning in July 2007, the negentropy anomalies gradually
accumulated. In particular, we find more frequent negentropy anomalies in 2008 as the earthquake
approaches, and less frequent anomalies after, so a sigmoid function is used to fit the acceleration,
before the earthquake and the deceleration after. According to De Santis, A. et al. (2017), inflection
point in this function is a reasonable estimation of the time of the significant change in the critical
dynamical system. Our calculation shows that the inflection point x0 of the optimal fitting is 8.3337,
which is surprisingly close to the actual time (8.3871) of the Wenchuan earthquake after conversion.
Thus, the earthquake moment is proved to be a critical time during the whole Wenchuan earthquake
process.

Fig. 8. (a) Results of the long-scale negentropy anomaly frequency for the Wenchuan earthquake at Guza station
from January 1, 2007, to June 30, 2009. Each circle represents an anomaly negentropy for 2 months. The
cumulative frequency of negentropy anomaly is represented. The earthquake day is represented as a vertical solid
line. The red line is a sigmoidal fit that underlines a critical point (vertical dashed line) is close to the occurrence of
the earthquake. (b) Results of the short-scale negentropy anomaly frequency prior to the earthquake, every grey
point is an anomaly for one day, the blue and red lines are two segment sigmoid fit results. Two green lines
represent the liner regression for the two phases, the first phase slope is 0.00147, the second one is 0.0021.
When we narrowed the accumulated window to one day, we observed two negentropy anomalies before the earthquake as shown in Fig. 8(b). The first anomaly frequency increase occurred from August to October 2007. In March 2008, there was a second phase of anomaly increase, and the cumulative frequency then slowly increased to a plateau period near the time of the earthquake. These two phases prior to the earthquake are also approximated with sigmoid functions. In order to further compare the anomalies of the two phases, we use linear regression to fit the central part of the two sigmoid curves. We find that the second acceleration is greater than the first acceleration. We think the accumulations of these two negentropy anomalies may be an indication of crustal activity before the Wenchuan earthquake.

5. Discussion and Conclusion

In the skewness-kurtosis domain, we observed the evolution of the negentropy distribution prior to the earthquake. Negentropy gradually transformed its distribution to a parabolic relation since July 2007, indicating a relatively stable state was broken due to the non-Gaussian mechanism.

Previous studies for the Wenchuan earthquake are consistent with our findings. Wang (2018) concluded that an apparent stress change occurred after June 2007 based on multiple focal mechanisms. Likewise, we did not find negentropy anomalies in the first six months of 2007. In the large-scale analysis, we show that the cumulative frequency of negentropy anomalies follows the a power-law behaviour approaching a critical time that is close to the earthquake time, and then recovers as a typical recovery phase after the earthquake. This process is consistent with the empirical phenomena before and after earthquakes, which is also similar with a potential earthquake precursory pattern in magnetic data from Swarm satellites by A. De Santis (2017) for the 2015 Nepal event.

Qiu (2009) and Chi (2014) speculated that the areal strain indicates that the integrity of the medium around the borehole at the Guza station began to change significantly after August 2007, because the continuity of the medium in the source region of the Wenchuan earthquake was gradually deformed during the nucleation process. In our short-scale analysis, negentropy anomalies also present a acceleration in August 2007. The second acceleration in March 2008 is also consistent with the occurrence of a phase measured by GPS (Jiang Z. S., 2009), in which the elastic deformation of the crust reaches its limit, and the deformation is resisted in the seismogenic region before the earthquake.
More importantly, we speculate that the two accelerations of the cumulative negentropy anomaly corresponding to two stress releases. Ma Jin (2014) proposed the deformation characteristics in the sub-instability stage of faults before earthquakes based on different experiments and several earthquakes. Fault zones contain relatively weak and relatively strong parts. The former is the area where strain release begins, while the latter is the stress locking part and the beginning of rapid instability (Noda et al., 2013). She thinks accelerated expansion of the strain release zone in earthquakes, the former is related to the release of weak parts, and the latter is related to the rapid release of strong parts of the fault during strong earthquakes. The accelerated expansion of the former promotes the occurrence of the latter.

We speculate that the two negentropy anomaly accelerations may represent the two instabilities associated with the Wenchuan earthquake. The first corresponds to the start of strain release and the second larger one corresponds to the acceleration of instability, indicating that strong earthquakes are likely to occur.

In our work, the extracted negentropy anomalies of the short-period signal of borehole areal strain based on Otsu’s thresholding associated with the Wenchuan earthquake are analyzed. A logarithmic-linear model is proposed to estimate the expected frequency of the left-skewed negentropy anomaly for everyday, and the evolution processes of the negentropy anomaly frequency are studied in both long- and short-scale during the study period. We consider the negentropy anomalies corresponding to crustal stress changes, which may be a reflection of the subsurface medium and faults activities in the focal area associate with the Wenchuan earthquake. Moreover, we may be able to ensure that the negentropy has great potential in the study of earthquake precursors.

Acknowledgements
The authors are appreciative of the Key Laboratory of the Geo-Exploration Instrumentation of Ministry of Education in Jilin University of China. Moreover, the authors are grateful to Professor Zehua Qiu for his guidance and helpful suggestions. This research was supported by the Institute of Crustal Dynamics, China Earthquake Administration (Grant No.3R216N620537).

References


Hirose, H. Tilt records prior to the 2011 off the Pacific coast of Tohoku Earthquake, Earth Planet and Space, 63(7): 665-658, 2011.


Ma Jin, Guo Yan-shuang. Accelerated synergism prior to fault instability evidence from laboratory experiments and an earthquake case, Seismology and Geology, 36(3): 547-560, 2014.


