IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 26, 2019, accepted October 16, 2019, date of publication October 29, 2019,
date of current version November 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950011

Detecting Earthquake-Related Borehole Strain Data Anomalies With Variational Mode Decomposition and Principal Component Analysis: A Case Study of the Wenchuan Earthquake

CHENGQUAN CHI, KAIGUANG ZHU, ZINING YU, MENGXUAN FAN, KAIYAN LI, AND HUIHUI SUN

Key Laboratory of Geo-Exploration Instrumentation, Ministry of Education, Jilin University, Changchun 130061, China

Corresponding author: Kaiguang Zhu (zhukaiguang@jlu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 41974084.

ABSTRACT Borehole strain monitoring has high sensitivity and is therefore widely used to study slow earthquakes, volcanic activity, earthquake precursors, and other nature phenomena. However, environmental factors seriously affect the identification of strain changes caused by crustal deformation. This paper proposes a method of anomaly detection based on variational mode decomposition (VMD) and principal component analysis (PCA). The borehole strain signal is decomposed into a number of modes simultaneously using VMD, and a new state-space model used to determine the number of the modes those are decomposed by the VMD algorithm. The influencing factors of each component are determined by spectrum analysis and comparative analysis. An example of the separation process of borehole strain data by the VMD method is presented. Then, we use PCA to calculate eigenvalues, which are used to detect anomalies associated with an earthquake, and eigenvectors, which are applied to show the spatial distribution characteristics of the data. Our method has been applied to detect borehole strain data anomalies associated with the Wenchuan earthquake; the VMD demonstrates excellent separation performance for borehole strain signals, and eigenvalues and eigenvectors together reflect the accelerated deformation of focal faults and adjacent areas before earthquake in time and space.

INDEX TERMS Borehole strain, principal component analysis, state-space model, variational mode decomposition, Wenchuan earthquake.

I. INTRODUCTION

Multi-component borehole strain monitoring has the advantages of high resolution, high sensitivity and long term stability [1], [2]. Consequently, it is it is widely used to study slow earthquakes [3], [13]–[15], volcanic activity [18], [19], and earthquake precursors [16], [17]. Thanks to the U.S. Plate Boundary Observatory project, the development of the four-gauge borehole strainmeters (FGBSs) in China was accelerated. The YRY-4 strainmeter manufactured by Chi *et al.* is now widely installed in China. To detect anomalies associated with earthquakes, researchers have studied borehole strain data for large earthquakes [5]–[8], and extracted credible anomalies. However, because of high resolution and high sensitivity, borehole strain monitoring records not only tectonic signals but also signals from environmental disturbances. Environmental factors seriously affect the identification of strain changes caused by crustal deformation. Environmental disturbances mainly include the Earth tides [11], [12], air pressure changes, and rainfall [3]. To remove interference and correct borehole data, many scholars have analyzed these complex signals. Hart *et al.* [10] incorporated cross coupling into the strain meter calibration to investigate the estimation of Earth strain from borehole strain meter data in a

The associate editor coordinating the review of this manuscript and approving it for publication was Mithun Mukherjee.

study of tidal calibration; this yielded good agreement with the measured strain tides from a colocated laser strainmete. Langbein [20] used least squares and data covariance to describe the temporal correlation of strainmeter data to adjust strain data to recover small signals that might be related to crustal deformation; the results provide accurate estimates of the various parameters. Scholars mostly use modeling and fitting methods to analyze and correct data in the time domain. However, time domain methods can only extract a portion of the useful time domain information and the frequency domain information is lost.

Recently, Dragomiretskiy and Zosso [21] proposed an adaptive time-frequency analysis algorithm, variational mode decomposition(VMD), which can decompose a signal into an ensemble of band-limited intrinsic mode functions, where their center frequencies are estimated online and all modes are extracted synchronously. VMD is widely used in non-stationary signal processing. Xue *et al.* [22] used VMD to decompose a seismic signal and compare it with the short-time Fourier transform (STFT) or wavelet transform (WT); the instantaneous spectrum after VMD promises higher spectral and spatial resolution. Sharma [23] proposed a new technique based on VMD for estimating heart rate (HR) from a photoplethysmography (PPG) signal to provide accurate and reliable HR information using the PPG signal recorded from patients suffering from dissimilar problems. Zhang *et al.* [24] used VMD to design a blind source separation method based on variational mode decomposition (VMD), which is applied to the separation of the wind turbine aeroacoustics signals acquired by single acoustic sensors; it has excellent separation performance. These methods testify that VMD is an effective signal processing method for complex non-stationary signals; therefore, VMD could have excellent separation performance for borehole strain signals. However, it is difficult to select a reasonable mode number for VMD. Dey and Satija [25] proposed a method based on VMD and principal component analysis (PCA) for single channel blind source separation, where PCA is used to select the corresponding source components from the decomposed modes. Lian*et al.* [27] proposed a method called adaptive VMD to automatically determine the mode number based on the characteristic of intrinsic mode function. Zhang *et al.* [28] used the maximum weighted kurtosis index constructed by kurtosis and the correlation coefficient to optimize VMD parameters. When there is no data information, it is necessary to select the number of modes adaptively. However, when there is data information, it is reasonable to select the number of modes according to the characteristics of data. In this paper, we propose a new state-space model to select the number of VMD.

In the study of multi-component and multi-station earthquake precursor data processing, many researchers have applied PCA to detect the anomalies for many types of non-stationary geophysical signals, and obtained meaningful results. Hattori *et al.* [29] applied PCA to detect the anomalies of Ultra-Low Frequency (ULF) geomagnetic from three stations; the anomalies of temporary variations of eigenvalues and eigenvectors are likely to be correlated with large earthquakes. Lin [30] used PCA to determine the spatial pattern of total electron content (TEC) anomalies in the ionosphere post the Wenchuan Earthquake; the result is remarkably similar to that reported for a TEC anomaly previously identified as a precursor anomaly on 9 May, 2008 for the same time period. Zhu *et al.* [9] applied PCA to analyze YRY-4 four-gauge borehole strain data from Guza Station, and extracted credible anomalies associated with the Lushan earthquake. Therefore, PCA is promising for monitoring the crustal activity in studies of multi-dimensional earthquake precursor data.

In this work, a method of anomaly detection based on VMD and PCA is proposed. The VMD method is applied to adaptively decompose borehole strain signal completely from the data itself. We proposed a new state-space model that has been used to select a reasonable mode number for VMD. Unlike [25], PCA is used to detect the anomalies related to an earthquake.

The rest of this paper is organized as follows: Section II introduces the observation principle and strain conversion, and we propose a new state-space model. Section III introduces the principle of VMD-PCA. In Section IV, we present an example illustrating the separation process of borehole strain data by the VMD method. In Section V, we describe the detection of borehole strain anomalies related to the Wenchuan earthquake using the propose method. In Section VI, we analyze and discuss the results of the eigenvalues and eigenvectors. Conclusions are drawn in Section VII.

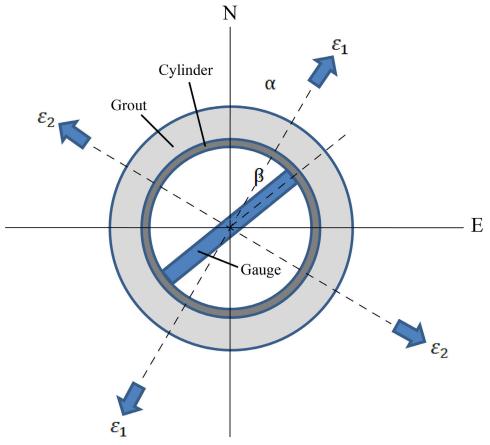


FIGURE 1. Sketch of the two-ring system for measuring strain in boreholes.

II. BOREHOLE STRAIN DATA

A. OBSERVATION PRINCIPLE AND STRAIN CONVERSION

YRY-4 FGBSs have four gauges arranged at 45° intervals in a cylindrical case, and can measure the changes of diameter in the corresponding azimuths. As shown in Fig. 1, assuming linear elasticity and isotropy of the medium, the two-ring model [31] is suitable for observing the horizontal strain state in rocks with a borehole strainmeter.

Gauge i in the cylinder directly measures the change in diameter in the corresponding azimuth β_i that results from changes in strain state. Although the solutions are complex, the resulting formula of the relationship between the measurement S_i and the strain changes (ε_1, ε_2, α) is straight forward [2]:

$$S_i = A(\varepsilon_1 + \varepsilon_2) + B(\varepsilon_1 - \varepsilon_2)cos2(\beta_i - \alpha). \quad (1)$$

The YRY-4 borehole strainmeter contains four horizontally emplaced sensors to measure changes in the borehole diameter. Self-consistency is crucial in FGBS design. According to the theoretical model (1), the relationship can be obtained as follows:

$$\begin{cases} S_1 = S_{\beta_1} = A(\varepsilon_1 + \varepsilon_2) + B(\varepsilon_1 - \varepsilon_2)cos2(\beta_i - \alpha) \\ S_2 = S_{\beta_1+\pi/4} = A(\varepsilon_1 + \varepsilon_2) - B(\varepsilon_1 - \varepsilon_2)sin2(\beta_i - \alpha) \\ S_3 = S_{\beta_1+\pi/2} = A(\varepsilon_1 + \varepsilon_2) - B(\varepsilon_1 - \varepsilon_2)cos2(\beta_i - \alpha) \\ S_4 = S_{\beta_1+3\pi/4} = A(\varepsilon_1 + \varepsilon_2) + B(\varepsilon_1 - \varepsilon_2)sin2(\beta_i - \alpha), \end{cases} \quad (2)$$

where S_i (i=1,2,3,4) is the measurement obtained from each of the fourgauges. With one additional measurement, a simple relationship among the four measurements can be obtained using the equation (2):

$$S_1 + S_3 = k(S_2 + S_4), \quad (3)$$

which is the self-consistency equation of the YRY-4 borehole strainmeter. This equation can be employed to estimate the credibility of the data.The k is the self-consistent coefficient, and $k = 1$ in ideally circumstances. We believe that the data are reliable when $k \geq 0.95$. There are only three independent variables under plain strain conditions at or near the Earth's surface. We can therefore derive various strains from the Guza recordings. The formulas used are as follows:

$$\begin{cases} S_{13} = S_1 - S_3 \\ S_{24} = S_2 - S_4 \\ S_a = (S_1 + S_2 + S_3 + S_4)/2, \end{cases} \quad (4)$$

where S_a represents the areal strain, and S_{13} and S_{24} represent the two independent shear strains.

B. DECOMPOSE DATA WITH A NEW STATE-SPACE MODEL

Strainmeters record not only tectonic-origin signals but also signals from environmental disturbances. To separate tectonic-original signals from other noise, borehole strain observation data can be indicated by:

$$X(t) = A(t) + B(t) + E(t), \quad t = 1, 2, 3, ..., \quad (5)$$

where $X(t)$ is strain observation data; $A(t)$ is the trend term for changes in strain data; $B(t)$ is a periodic term, which mainly includes the solid tidal response; $E(t)$ is short-period anomalous changes caused by crustal deformation and other factors causing tectonic earthquakes [32].

Hsu *et al.* [3] applied a state-space model to remove the strain response to rainfall, in addition to those due to air pressure changes and Earth tides, and investigated whether corrected strain changes are related to environmental disturbances or tectonic-original motions. The observed strain S_n^0 on can be represented by:

$$\begin{aligned} S_n^0 &= S_n^c + E_n + P_n + R_n + \varepsilon_n \\ \varepsilon_n &\sim N(0, \sigma^2), \quad n = 1, 2, ..., N, \end{aligned} \quad (6)$$

where S_n^c is the corrected strain data; P_n, E_n, and R_n are induced strain by barometric pressure changes, Earth tides, and rainfall, respectively; ε_n is Gaussian white noise; and N is the number of observations.

Based on the above two models, we propose a new state-space model. The principle of the effect of rainfall on borehole strain is very complex. Although it has time lags, borehole water level can still reflect the influence of rainfall [33]. Therefore, we use borehole water level data instead of rainfall data. The new state-space model is described as follow:

$$\begin{aligned} S_n^0 &= T_n + S_n^c + E_n + P_n + L_n + \varepsilon_n \\ \varepsilon_n &\sim N(0, \sigma^2), \quad n = 1, 2, ..., N, \end{aligned} \quad (7)$$

where S_n^0 is the raw borehole strain data, T_n is the trend term for changes in strain data; S_n^c is short-period anomalous changes caused by crustal deformation; E_n, P_n, and L_n are induced strain by borehole pressure changes, the Earth tides changes, and borehole water-lever changes, respectively; ε_n is Gaussian white noise; and N is the number of observations.

III. THE PRINCIPLE OF VMD-PCA

The VMD of nonlinear and nonstationary signals based on the data itself is adaptive. The VMD generalize the classic Wiener filter into multiple and adaptive bands, which can realize signal adaptive decomposition by finding the optimal solution of the constraint variational model [21]. VMD is a novel signal decomposition method that is theoretically well founded and can deal with nonlinear and non-stationary signals. PCA is a widely used technique in data analysis; it is a non-parametric method that is capable of extracting relevant information from complex data sets [26], and which often reveals relationships that were not previously suspected, thereby allowing for an otherwise unordinary interpretation. Therefore, this work combines the best features of VMD and PCA to construct an approach for signal processing and anomalies extraction.

We apply VMD to three independent strain converters (S_{13}, S_{24}, S_a), respectively. First, we determine the mode number of VMD by the new state-space model. Borehole strain data can be expressed as follow:

$$S = \left\{u_1(\omega_1), u_2(\omega_2), ..., u_k(\omega_k)\right\}, \quad k = 1, 2, ..., n, \quad (8)$$

where u_k and ω_k are shorthand notations for the set of all modes and their center frequencies, respectively. To assess the bandwidth of a mode, we propose the following VMD scheme: 1) for each mode, compute the associated analytic signal by means of the Hilbert transform in order to obtain

a unilateral frequency spectrum; 2) for each mode, shift the mode's frequency spectrum to 'baseband', by mixing with an exponential tuned to the respective estimated center frequency; 3) the bandwidth is now estimated through the H^1 Gaussian smoothness of the demodulated signal,(i.e. the squared L^2-norm of the gradient). The resulting constrained variational problem is as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t [(\delta_t + \frac{j}{2\pi}) * u_k(t)] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \sum_k u_k = f, \tag{9}$$

To render the problem unconstrained, a quadratic penalty term and Lagrangian multipliers are employed and a new solution expression can be obtained as follows:

$$\begin{aligned} &L(\{u_k\}, \{w_k\}, \lambda) \\ &= \alpha \sum \left\| \partial_t \left[(\delta(t) + \frac{j}{2\pi}) * u_k(t) \right] e^{-jw_0 t} \right\|_2^2 \\ &\quad + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle + \left\| f(t) - \sum_k u_k(t) \right\|_2^2, \end{aligned} \tag{10}$$

where α is the data-fidelity constraint parameter and λ is the Lagrangian multiplier.

The alternate direction method of multipliers (ADMM) approach is used to produce different decomposed modes and the center frequency during each shifting operation. Then, the modes u_k and their corresponding center frequency ω_k can be updated as:

$$u_k^{n+1} \leftarrow argmin L_{u_k}(u_{i<k}^{n+1}, u_{i\geq k}^{n+1}, \omega_i^n, \lambda^n), \tag{11}$$

and

$$\omega_k^{n+1} \leftarrow argmin L_{\omega_k}(u_i^{n+1}, u_{i<k}^{n+1}, \omega_{i\geq k}^n, \lambda^n). \tag{12}$$

Each mode obtained from solutions in the spectral domain can be represented as:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i\neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{13}$$

Compare the data of influencing factors and extract the components related to crustal activity $\hat{u}_{S_{13}}, \hat{u}_{S_{24}}, \hat{u}_{S_a}$.

Then, $\hat{u}_{S_{13}}, \hat{u}_{S_{24}}, \hat{u}_{S_a}$ are presented in a matrix Y of m rows and n columns:

$$Y = \begin{bmatrix} \hat{u}_{S_{13}} \\ \hat{u}_{S_{24}} \\ \hat{u}_{S_a} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}, \tag{14}$$

where m is the number of samples, and n is the dimension of the sample.

First, we calculate the co-variance matrix $C_Y(m \times m)$ of the dataset $Y(m \times n)$, and the element γ_{pq} in the co-variance matrix $C_Y(m \times m)$ can be calculated using the formula:

$$\gamma_{pq} = [1/(N-1)] \sum_{i-1}^{N} (x_p^i - \bar{X}_p)(x_q^i - \bar{X}_q), \tag{15}$$

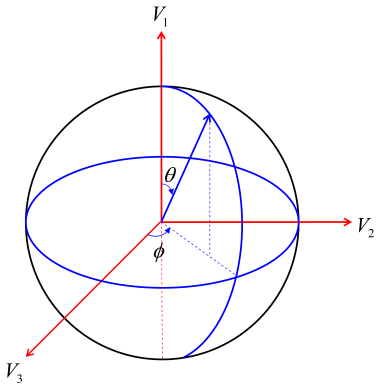


FIGURE 2. Unit spherical coordinate system of eigenvectors.

where x_p^i and x_q^i are the pth and qth columns of the ith row of data, respectively; and $\bar{X}_p$ and $\bar{X}_q$ are the averages of the pth and qth columns of data, respectively. Here N is the number of samples.

We perform the eigenvalue decomposition using the co-variance matrix:

$$C_Y = V \Lambda V^T, \tag{16}$$

where Λ is the eigenvalue matrix with λ_1, λ_2, λ_3 ($\lambda_1 > \lambda_2 > \lambda_3$) and V is the eigenvector matrix whose columns are v_1, v_2, v_3. The first principal component eigenvalue and eigenvector are λ_1 and v_1, respectively, which represent the principal characteristics of the signals [29].In this paper, the variations in the eigenvalue and eigenvector of the first principal component are investigated.

The obtained eigenvector v_1 has three dimensions in the vector space because the number of PCA dimensions is three and it is a unit vector. To present the changes in the eigenvector more intuitively, we transform the eigenvectors to the unit spherical coordinate system, as shown in Fig. 2 The eigenvector can be represented by θ and ϕ exclusively.

IV. VMD DECOMPOSITION OF THE BOREHOLE STRAIN DATA

Here, we will present the process of removing influencing factors from borehole strain data by VMD using areal strain data S_a for a period of 1 month. The raw data of S_a, borehole water-level and borehole pressure are shown in Fig. 3.

First, we apply VMD to S_a, borehole water-level and borehole pressure, respectively. As shown in Fig. 4,we decompose S_a into five components on the basis of the new state space model. It is noticeable that the removal of Gauss noise is achieved with a Wiener filter on each mode during the decomposition progress. From this figure, we can see that S_a is successfully decomposed into c_1, c_2, c_3, c_4,and c_5. And we compared the result of VMD with that of EMD (Fig. 5 shows the decomposition results of S_a by using EMD). The mode mixing problem of EMD is seen to be serious when it is used to deal with composite borehole strain data, there is hardly any completed sub-signal has been decomposed successfully.

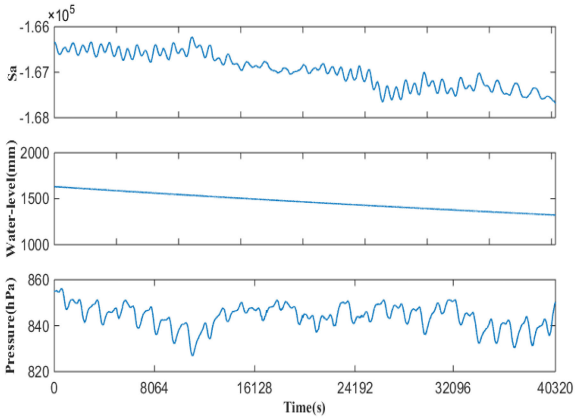


FIGURE 3. The raw data of S_a, borehole water-level and borehole pressure.

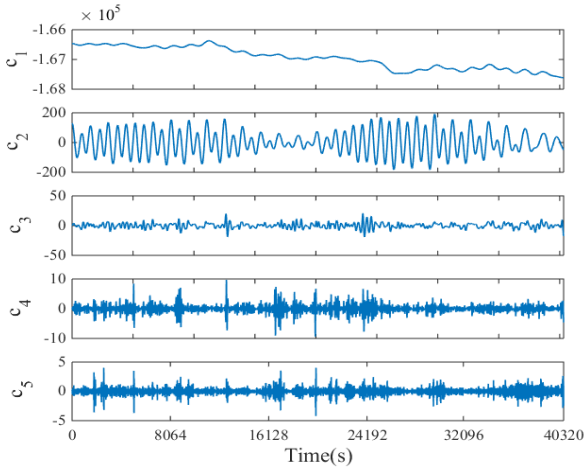


FIGURE 4. The decomposition results of S_a by using VMD.

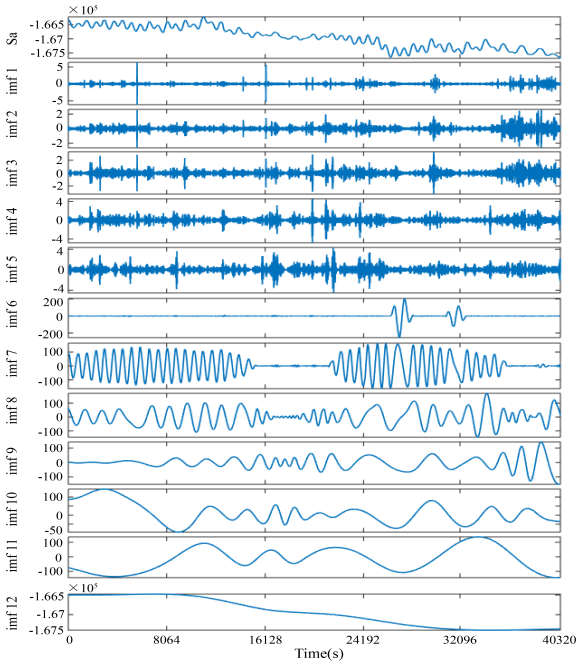


FIGURE 5. The decomposition results of S_a by using EMD.

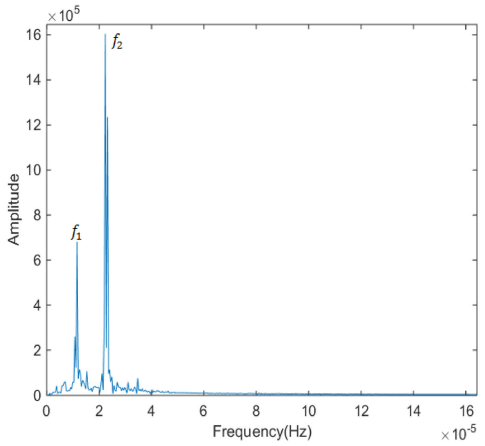


FIGURE 6. The FFT periodogram of component c_2.

In contrast to EMD, VMD can adaptively decompose borehole strain data into an ensemble of band-limited intrinsic mode functions, and VMD is suitable for decomposing non-linear and non-stationary signals. We have attempted to identify physically significant of the components of VMD. It is obvious that the first component c_1 represent the trend term. The FFT periodogram of component c_2 shows that the frequency of the signal is mainly concentrated in $f_1 = 1.157 \times 10^{-5} Hz$ and $f_1 = 2.232 \times 10^{-5} Hz$, as shown in Fig. 6; these two frequencies correspond to the semidiurnal wave and diurnal wave frequencies of the Earth tides, respectively. Therefore c_2 may be considered as the influence of the Earth tides.

To study the components of c_3, c_4 and c_5, we apply VMD to borehole water level and pressure. As borehole pressure data are also affected by the Earth tides, we decompose them into three components. As shown in Fig. 7, the first component p_1 represents the trend term, p_2 represents the Earth tides and p_3 represents the short-period change component of borehole pressure. Fig. 8 shows the result of decomposing borehole water-level, where w_1 represents the trend term and w_2 represents the short-period change component.

Comparative analysis is applied to signals c_4, c_5, p_3 and w_2, and the result is shown in Fig. 8. To contrast waveform shape, we found that c_4, c_5, and p_3 have similar fluctuations in the same period of time(as shown in the red boxes), and c_5 shows changes similar to those of w_2 (as shown in the black box). Fig. 9 indicates that c_4 and c_5 mainly correspond to the response of borehole pressure and borehole water level. At the same time, it is also verified that pressure has a significant influence on borehole strain data [34].

From the above analysis, we can see that the $1st$ component c_1 represent the trend term, the $2nd$ component c_2 is the influence of the Earth tides, and the $4th$ component c_4 and the 5^{th} component c_5 correspond to the response of borehole

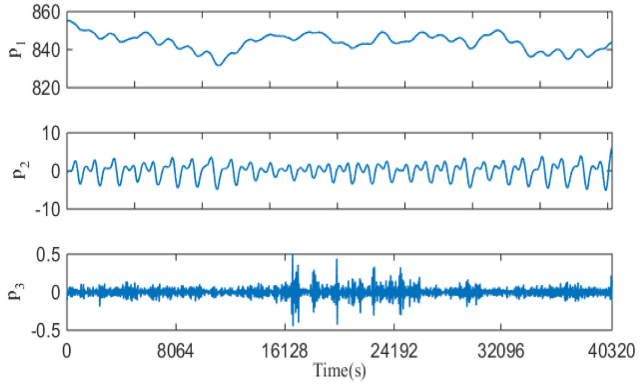


FIGURE 7. The decomposition results of borehole pressure by using VMD.

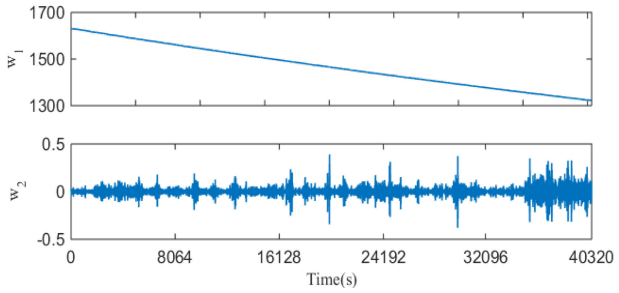


FIGURE 8. The decomposition results of borehole water level by using VMD.

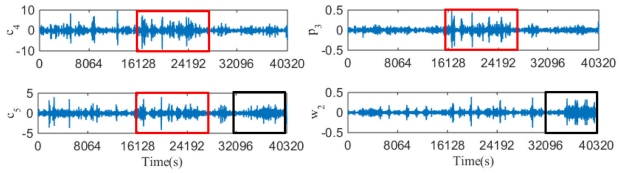


FIGURE 9. The comparative analysis result of c_4 and c_5 with borehole pressure and borehole water level.

pressure and borehole water-level. Hence, we suggest that the 3*th* component c_3 may be considered as short-period changes caused by crustal deformation in the borehole strain data; as such, we use only the 3*th* component in the next step of analysis.

The same treatment process is applied to S_{13} and S_{24}, and the results are similar. Then, short-period changes caused by crustal deformation components of S_{13}, S_{24}, and S_a are used to extract anomalies by PCA.

V. EXTRACTING ANOMALIES ASSOCIATED WITH THE WENCHUAN EARTHQUAKE

We extract anomalies of borehole strain data associated with the Wenchuan earthquake from Guza Station.

A. OBSERVATIONS AND EARTHQUAKE

The Guza Station is at the southwestern extent of the Longmenshan fault zone. The deformation observation instrumentation includes a very-long-period vertical pendulum

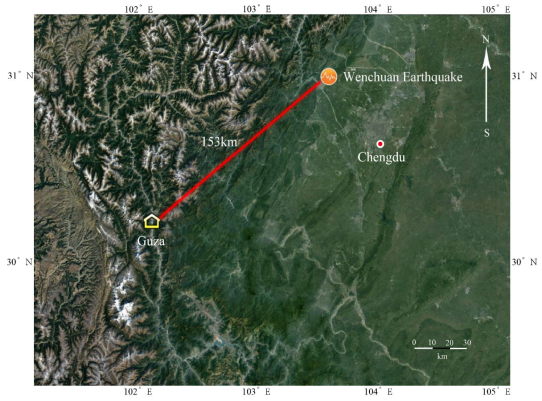


FIGURE 10. Location map showing Guza Station and the Wenchuan earthquake epicenter.

tiltmeter, YRY-4 borehole strainmeter, DSQ water pipe inclinometer, SS-Y body piercing extensometer, and DZW digitalized gravity meter. The YRY-4 borehole strainmeter was installed in October 2006; continuous recordings have been collected at a sampling rate of one sample per minute since 1 December, 2006.

At 14:28 (UTC+8) on 12 May 2008, an Ms8.0 earthquake occurred in Wenchuan County, Sichuan. The epicenter was at 31.01° N and 103.42° E. According to the data published by the China Earthquake Networks Center of the China Earthquake Administration, and the focal depth was approximately 14 km.

The distance between the epicenter of the Wenchuan earthquake and the receiver at Guza Station is 153 km (Fig. 10). The Wenchuan earthquake had a wide range of influence, and pre-earthquake physical phenomena were complicated; consequently, we cannot exclude a-priori the possibility that the YRY-4 borehole strainmeter installed at Guza Station recorded strain anomaly phenomena occurring before the earthquake.

B. DATA PROCESSING

We analyzed the borehole strain data from 1 January 2007 to 31 December 2008. First, we calculate the self-consistent coefficient k to confirm the reliability of borehole data. We found that $k = 0.9964$, which confirms that the data are reliable. Then, strain conversion is applied to the raw borehole strain data, (Fig. 11, where the black dotted line indicates the time of the earthquake).

Using the VMD method, we decompose the strain conversion data(S_{13}, S_{24}, and S_a) into five components to search for signals related to crustal deformation, as per the methodology discussed in the previous section. Fig. 12 shows the short-period changes caused by crustal deformation components. Finally, PCA is applied to extract the anomalies associated with Wenchuan earthquake. To avoid time-domain aliasing, and to distinguish anomalous days more easily, we perform PCA on daily data(i.e., 1 day is equal to 1440 points). We consider that the three components data (p_{13}, p_{24}, p_a) are

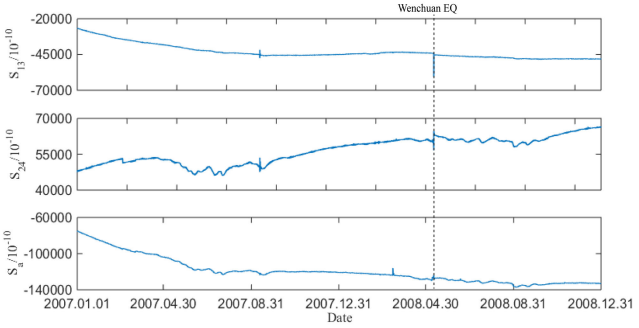


FIGURE 11. A time series plot of borehole strain data after the strain conversion.

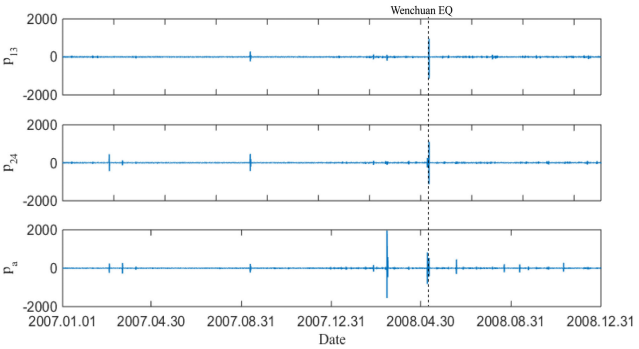


FIGURE 12. The short-period changes caused by crustal deformation components.

arranged in the form of time series data vectors.

$$\begin{cases} p_{13} = [p_{13}(t_1), p_{13}(t_2), \cdots, p_{13}(t_{1440})]^T \\ p_{24} = [p_{24}(t_1), p_{24}(t_2), \cdots, p_{24}(t_{1440})]^T \\ p_a = [p_a(t_1), p_a(t_2), \cdots, p_a(t_{1440})]^T, \end{cases} \tag{17}$$

where T indicates a transpose. Then, the data matrix $Y = [p_{13}, p_{24}, p_a]^T$ is prepared to calculate the first eigenvalues and the first eigenvectors by PCA.

VI. RESULTS AND DISCUSSION

We apply PCA to the data which is the short-period change caused by crustal deformation, and calculate the first principal component eigenvalue λ and eigenvector v. We calculate the average and standard deviation σ using all the values to recognize anomalous λ values. Those anomalous values are then defined as values that exceed the average by more than 1σ, and the result is shown in Fig. 13, where the average value is delimited by a red horizontal dashed line, and the average of more than 1σ is delimited by a red horizontal dotted line.

The variations in λ (Fig.13) illustrate that there are few anomalous values before January 2008. After January 2008, the number of anomaly eigenvalues increased, and this growth continued until a few months after the earthquake.

To express the variation characteristics of the number of anomaly eigenvalues more intuitively, we calculate the cumulative number of anomaly eigenvalues and fit it using a sigmoidal function. Fig. 14 illustrates the temporal behavior

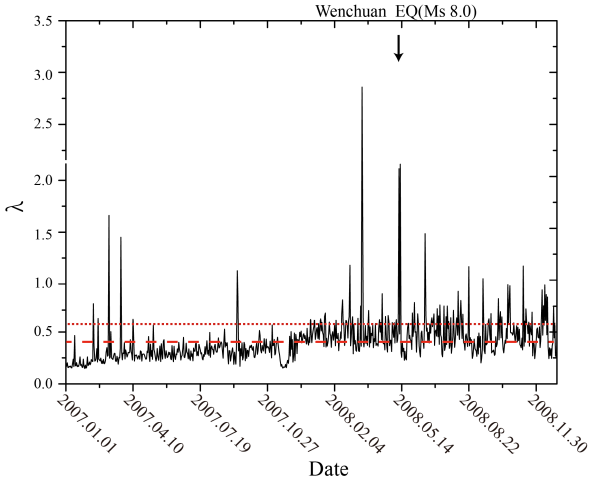


FIGURE 13. Daily variations in the first principal component eigenvalue.

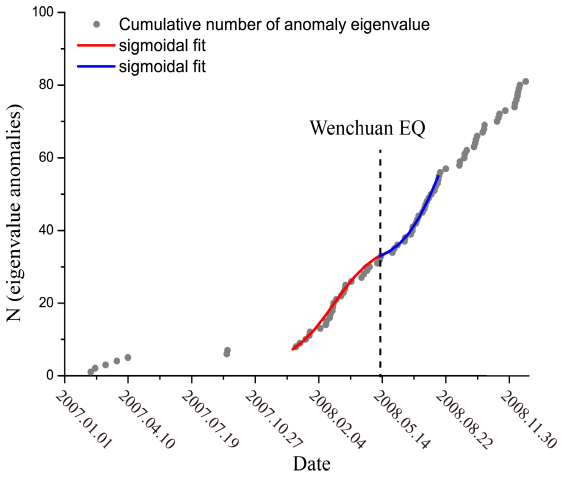


FIGURE 14. Cumulative number of anomaly eigenvalue calculate by the results of the daily variations in the first principal component eigenvalue.

of $N(t)$, denoted here as N (eigenvalue anomalies). The day of the earthquake is represented as a vertical dotted line. The red and blue curves are the sigmoidal fits before and after the earthquake. The results show sigmoidal temporal behavior before the earthquake with lower concavity, and sigmoidal behavior after the earthquake, with an opposite concavity. There is an acceleration about 4 months before the earthquake (from January 2008). Similar unusual variations are also seen after the earthquake. After such a large-magnitude earthquake, rearrangement of stresses in the crust commonly leads to a large number of anomaly events [34], [35]. As shown in Fig.14, there was a steep rise in the number of anomaly eigenvalues after the earthquake. Kong *et al.* [6] detected stress changes and the anomalies in the outgoing long wave radiation(OLR)data before the Wenchuan earthquake by using CD method, and indicated that there are large CD values three months before the Wenchuan earthquake (from January 2008), and this is consistent with our research. We consider that the eigenvalues and eigenvectors are both

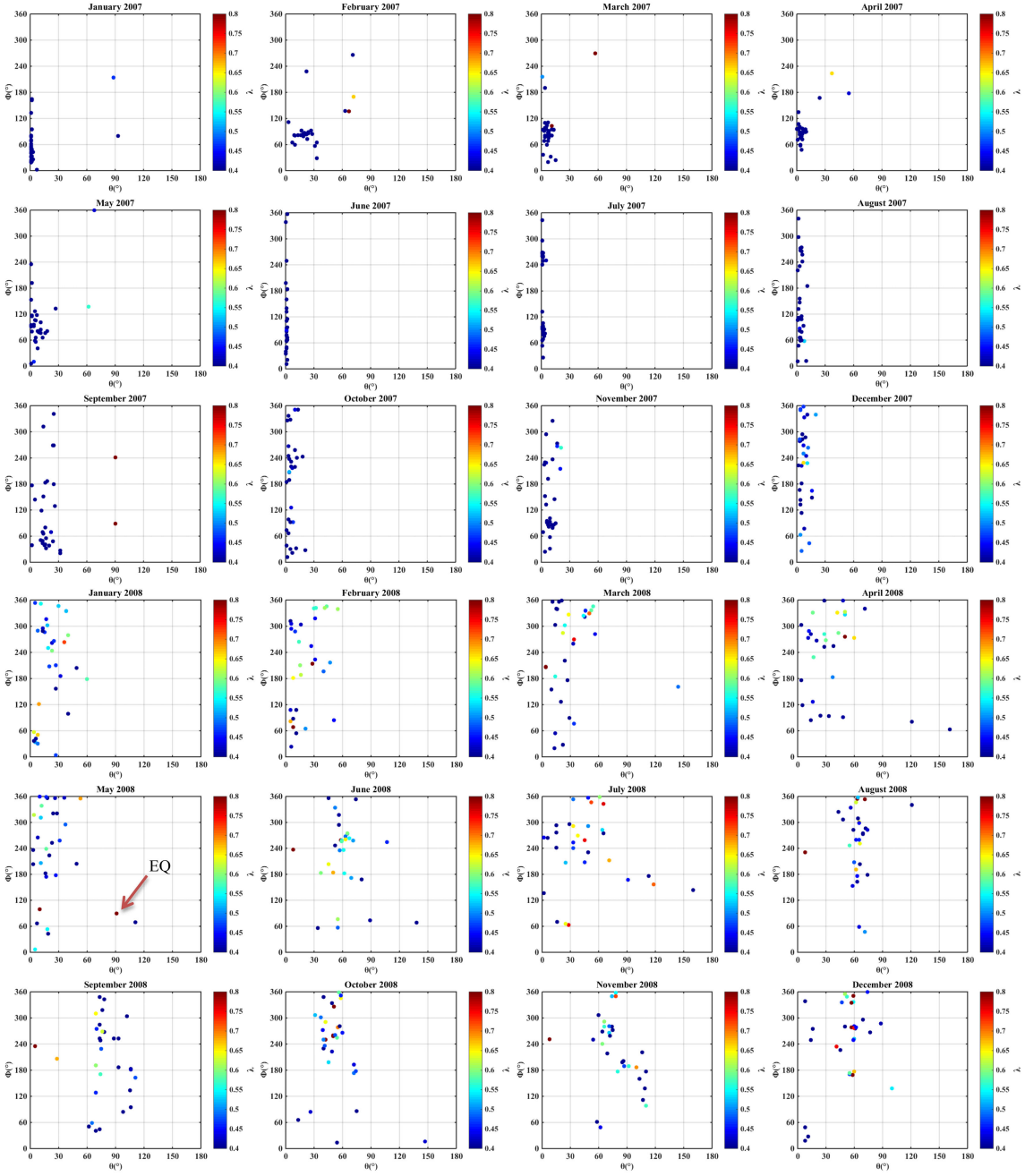


FIGURE 15. Spatial distributions of eigenvalues and eigenvectors.

obtained from the decomposition of the co-variance matrix of the data and that both contain parts of information of the data. We perform combined eigenvalue and eigenvector analysis. As shown in Fig. 15, the x-axis and y-axis represent the θ and ϕ angles, respectively, and the color indicates the changes in the eigenvalue. The red arrow denotes the day of the Wenchuan earthquake.

Fig. 15 shows that eigenvectors present aggregated steady-state phenomena before June 2007. To some extent, this reflects that the spatial distribution of the stress level in different periods is uniform and stable with time [36]. From June 2007 to December 2007, eigenvectors exhibit longitudinal diffusion phenomena, possibly because of isolated areas of strain release increase and extend steadily [38]. After January 2008, eigenvectors present transverse diffusion, and eigenvalues become abnormal, the reason is that during the nucleation of the Wenchuan earthquake, the continuity of the medium in the focal area was destroyed, and the integrity of the medium around the borehole began to change significantly [37]. It reflects the acceleration of crustal instability. After the earthquake, the eigenvectors still show diffusion phenomena, and abnormal eigenvalues appear in large

numbers. We infer that this phenomenon is due to a large number of aftershocks and the extremely unstable state of the fault.

Ma *et al.*[38] performed a laboratory modeling study of instability on a planar strike-slip fault. They indicate that the occurrence of an earthquake is related to the synergism process of a fault, which includes three stages. In the first stage, the stress curve deviates from linearity and in the second stage, the isolated areas of strain release increase and extend steadily. In the third stage, the sections of strain release on the fault accelerate and expand, and the strain levels of strain-accumulation areas accelerate and rise. As shown in Fig.14, the spatial distribution of eigenvalues and eigenvectors show a consistent result. Between January 2007 and May 2007, the spatial distribution of eigenvectors present a steady state, and strain variations of different portions of the fault begin to diverge(i.e., the first stage). The second stage begins in June 2007, when the strains of isolated areas of strain release extend steadily, and the spatial distribution of eigenvectors presents steady diffusion. As strain levels of strain-accumulation areas accelerate, eigenvectors present transverse diffusion (from January 2008) and the number of anomaly eigenvalues begins to increase, indicating signs the third stage. Eigenvalues and eigenvectors together reflect the accelerated deformation of focal faults and adjacent areas before the Wenchuan earthquake in time and space. It is presumed that they are directly related to the meta-instability process of the seismogenic area before the Wenchuan earthquake.

VII. CONCLUSION

In this paper, a method of anomaly detection based on VMD and PCA is proposed; a new state-space model was built to determin the number of VMD components. VMD has excellent separation performance for borehole strain signals, and each component corresponds to a clear physical meaning. The proposed method is applied to detect anomalies associated with the Wenchuan earthquake. After removing influence factors, eigenvalues and eigenvectors are calculated by PCA to extract anomaly information. The eigenvalues and eigenvectors clearly show the seismogenic process before the Wenchuan earthquake. In future work, we will apply the proposed method to multi-station borehole strain data and other earthquake data.

REFERENCES

[1] M. T. Gladwin, "High-precision multicomponent borehole deformation monitoring," *Rev. Sci. Instruct.*, vol. 55, no. 12, pp. 2011–2016, 1984.

[2] Z. Qiu, L. Tang, Y. Guo, and B. Zhang, "*In situ* calibration of and algorithm for strain monitoring using four-gauge borehole strainmeters (FGBS)," *J. Geophys. Res., Solid Earth*, vol. 118, no. 4, pp. 1609–1618, 2013.

[3] Y.-J. Hsu, Y.-S. Chang, C.-C. Liu, H.-M. Lee, A. T. Linde, S. I. Sacks, G. Kitagawa, and Y.-G. Chen, "Revisiting borehole strain, typhoons, and slow earthquakes using quantitative estimates of precipitation-induced strain changes," *J. Geophys. Res., Solid Earth*, vol. 120, no. 6, pp. 4556–4571, 2015.

[4] S. L. Chi *et al.*, "The necessity of building national strain-observation network from the strain abnormality before Wenchuan earthquake," *Recent Develop. World Seismol.*, vol. 1, no. 1, pp. 1–13, 2009.

[5] Z. Qiu, B. Zhang, S. Chi, L. Tang, and M. Song, "Abnormal strain changes observed at Guza before the Wenchuan earthquake," *Sci. China Earth Sci.*, vol. 54, no. 2, pp. 233–240, 2011.

[6] X. Kong, N. Li, L. Lin, P. Xiong, and J. Qi, "Relationship of Stress changes and anomalies in OLR data of the Wenchuan and Lushan earthquakes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2966–2976, Aug. 2018.

[7] Z. Ouyang, H. Zhang, Z. Fu, B. Gou, and W. Jiang, "Abnormal phenomena recorded by several earthquake precursor observation instruments before the M_s 8.0 Wenchuan, Sichuan earthquake," *Acta Geologica Sinica-English Ed.*, vol. 83, no. 4, p. 834, 2009.

[8] S. L. Chi, Q. Liu, Y. Chi, T. Deng, C. W. Liao, G. Yang, G. P. Zhang, and J. Chen "Borehole strain anomalies before the 20 April 2013 Lushan Ms 7.0 earthquake," *Acta Seismol Sin.*, vol. 35, no. 3, pp. 296–303, 2013.

[9] K. Zhu, C. Chi, Z. Yu, W. Zhang, M. Fan, K. Li, and Q. Zhang, "Extracting borehole strain precursors associated with the Lushan earthquake through principal component analysis," *Ann. Geophys.*, vol. 61, no. 4, p. 447, 2018.

[10] R. H. G. Hart, M. T. Gladwin, R. L. Gwyther, D. C. Agnew, and F. K. Wyatt, "Tidal calibration of borehole strain meters: Removing the effects of small-scale inhomogeneity," *J. Geophys. Res., Solid Earth*, vol. 101, no. B11, pp. 25553–25571, 1996.

[11] J. Langbein, "Effect of error in theoretical Earth tide on calibration of borehole strainmeters," *Geophys. Res. Lett.*, vol. 37, no. 21, pp. 1–5, Nov. 2010.

[12] K. Kathleen, J. Langbein, B. Henderson, D. Mencin, and A. Borsa, "Tidal calibration of Plate Boundary Observatory borehole strainmeters," *J. Geophys. Res., Solid Earth*, vol. 118, no. 1, pp. 447–458, 2013.

[13] I. S. Sacks, A. T. Linde, S. Suyehiro, and J. A. Snoke, "Slow earthquakes and stress redistribution," *Nature*, vol. 275, no. 5681, pp. 599–602, 1978.

[14] A. T. Linde, M. T. Gladwin, M. J. S. Johnston, R. L. Gwyther, and R. G. Bilham, "A slow earthquake sequence on the San Andreas fault," *Nature*, vol. 383, no. 6595, pp. 65–68, 1996.

[15] I. S. Sacks, A. T. Linde, J. A. Snoke, and S. Suyehiro, "A slow earthquake sequence following the Izu-Oshima earthquake of 1978," *Earthq. Predict., Int. Rev.*, vol. 4, pp. 617–628, Jan. 1981.

[16] S. Sakata and H. Sato, "Borehole-type tiltmeter and three-component strainmeter for earthquake prediction," *J. Phys. Earth*, vol. 34, pp. S129–S140, 1986.

[17] M. J. S. Johnston, A. T. Linde, and D. C. Agnew, "Continuous borehole strain in the San Andreas fault zone before, during, and after the 28 June 1992, M_W 7.3 Landers, California, earthquake," *Bull. Seismol. Soc. Amer*, vol. 84, no. 3, pp. 799–805, 1994.

[18] B. Voight, A. T. Linde, I. S. Sacks, G. S. Mattioli, R. S. J. Sparks, D. Elsworth, D. Hidayat, P. E. Malin, E. Shalev, C. Widiwijayanti, S. R. Young, V. Bass, A. Clarke, P. Dunkley, W. Johnston, N. McWhorter, J. Neuberg, and P. Williams, "Unprecedented pressure increase in deep magma reservoir triggered by lava-dome collapse," *Geophys. Res. Lett*, vol. 33, no. 3, pp. 1–4, 2006.

[19] S. Hautmann, D. Hidayat, N. Fournier, A. T. Linde, I. S. Sacks, and C. P. Williams, "Pressure changes in the magmatic system during the December 2008/January 2009 extrusion event at Soufrière Hills Volcano, Montserrat (W.I.), derived from strain data analysis," *J. Volcanol. Geothermal Res.*, vol. 250, pp. 34–41, Jan. 2013.

[20] J. Langbein, "Computer algorithm for analyzing and processing borehole strainmeter data," *Comput. Geosci.*, vol. 36, no. 5, pp. 611–619, 2010.

[21] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2010.

[22] Y.-J. Xue, J.-X. Cao, D.-X. Wang, H.-K. Du, and Y. Yao, "Application of the variational-mode decomposition for seismic time–frequency analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3821–3831, Aug. 2016.

[23] H. Sharma, "Heart rate extraction from PPG signals using variational mode decomposition," *Biocybernetics Biomed. Eng.*, vol. 39, no. 1, pp. 75–86, 2019.

[24] Y. N. Zhang, S. Qi, and L. Zhou, "Single channel blind source separation for wind turbine aeroacoustics signals based on variational mode decomposition," *IEEE Access*, vol. 6, pp. 73952–73964, 2018.

[25] P. Dey, B. Ramkumar, and U. Satija, "Single channel blind source separation based on variational mode decomposition and PCA," in *Proc. Annu. IEEE India Conf. (INDICON)*, New Delhi, India, Dec. 2018, pp. 1–5.

[26] E. G. Londoño, L. C. López, and T. de Souza Kazmierczak, "Using the Karhunen-Loève transform to suppress ground roll in seismic data," *Earth Sci. Res. J.*, vol. 9, no. 2, pp. 139–147, 2005.

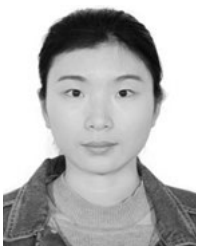
[27] J. Lian, Z. Liu, H. Wang, and X. Dong, "Adaptive variational mode decomposition method for signal processing based on mode characteristic," *Mech. Syst. Signal Process.*, vol. 107, pp. 53–77, Jul. 2018.
[28] X. Zhang, Q. Miao, H. Zhang, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mech. Syst. Signal Process.*, vol. 108, pp. 58–72, Aug. 2018.
[29] K. Hattori, A. Serita, K. Gotoh, C. Yoshino, M. Harada, N. Isezaki, and M. Hayakawa, "ULF geomagnetic anomaly associated with 2000 Izu islands earthquake swarm, Japan," *Phys. Chem. Earth, A/B/C*, vol. 29, nos. 4–9, pp. 425–435, 2004.
[30] L. Jyh-Woei, "Use of principal component analysis in the identification of the spatial pattern of an ionospheric total electron content anomalies after China's May 12, 2008, M = 7.9 Wenchuan earthquake," *Adv. Space Res.*, vol. 47, no. 11, pp. 1983–1989, 2011.
[31] Z. X. Ouyang and Z. R. Zhang, "Studies of method for coupling strain-gauge with borehole wall," *Selections Crustal Struct. Ground Stress*, vol. 2, pp. 1–10, Nov. 1988.
[32] Q. Z. Sun *et al.*, "Borehole strain observation work," in *Earthquake and Crustal Deformation Observation Technology*. Beijing, China: Seismological Press, 1995, pp. 355–356.
[33] L. J. E. Lee, D. S. L. Lawrence, and M. Price, "Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England," *J. Hydrol.*, vol. 330, nos. 3–4, pp. 604–620, 2006.
[34] T. Parsons, "Global Omori law decay of triggered earthquakes: Large aftershocks outside the classical aftershock zone," *J. Geophys. Res., Solid Earth*, vol. 107, no. B9, pp. 1–20, 2002.
[35] J. McCloskey, S. S. Nalbant, and S. Steacy, "Earthquake risk from co-seismic stress," *Nature*, vol. 434, no. 7031, p. 291, 2005.
[36] K. Y. Wang, G. YanShuang, and F. XiangDong, "Sub-instability stress state prior to the 2008 Wenchuan earthquake from temporal and spatial stress evolution," *Chin. J. Geophys.-Chin. Ed.*, vol. 61, no. 5, pp. 1883–1890, 2018.
[37] S. L. Chi, J. Zhang, and Y. Chi, "Failure of self-consistent strain data before Wenchuan, Ludian and Kangding earthquakes and its relation with earthquake nucleation," *Recent Develop. World Seismol.*, vol. 12, pp. 3–13, Jan. 2014.
[38] J. Ma and Y. S. Guo, "Accelerated synergism prior to fault instability: Evidnce from laboratory experiments and an earthquake case," *Seismol. Geol.*, vol. 36, no. 3, pp. 547–561, Sep. 2014.

CHENGQUAN CHI received the B.S. degree in automation and the M.S. degree in control theory and control engineering from Northeast Forestry University, Harbin, China, in 2009 and 2013, respectively. He is currently pursuing the Ph.D. degree in measurement technology and instrumentation with Jilin University, Changchun, China. His current research interests include blind source separation, geophysics, and signal processing especially earthquake precursor data.

KAIGUANG ZHU received the Ph.D. degree in earth exploration and information technology from Jilin University, Changchun, China, in 2002. Her current research interests include signal processing involved with airborne electromagnetic survey and earthquake precursor analysis from both ground-based and satellite observations.

ZINING YU received the bachelor's degree from the College of Instrumentation and Electrical Engineering, Jilin University, Changchun, China, in 2016, where she is currently pursuing the Ph.D. degree. Her current research interests include signal processing, seismic observation data analysis, and seismic precursor analysis.

MENGXUAN FAN received the bachelor's degree from the College of Instrumentation and Electrical Engineering, Jilin University, Changchun, China, in 2017, where she is currently pursuing the master's degree. Her current research interests include signal processing, seismic precursor anomaly detection, and satellites data processing.



KAIYAN LI received the bachelor's degree from the College of Instrumentation and Electrical Engineering, Jilin University, Changchun, China, in 2017, where she is currently pursuing the master's degree. Her current research interests include seismic precursor anomaly detection and satellite data processing.

HUIHUI SUN received the M.S. degree from the College of Instrumentation and Electrical Engineering, Jilin University, Changchun, China, where she is currently pursuing the Ph.D. degree. Her current research interests include seismic precursor anomaly detection and borehole strain data processing.