PRINCIPAL COMPONENT ANALYSIS OF GEOMAGNETIC DATA FOR THE PANZHIHUA EARTHQUAKE (Ms 6.1) IN AUGUST 2008

Junhui Li[1, 2], Qi Li[3], Dongmei Yang[3], Xingzhou Wang[1, 2], Dequan Hong[1, 2] and Kang He[1, 2]*

[1]Seismological Bureau of Anhui, Hefei 230031, China
  Email: lijunhui@mail.ustc.edu.cn
[2] *National Geophysical Observatory at Mengcheng, Hefei 230031, China*
Email: wxz0208@sina.com
[3] *Institutes of Geophysics, CEA, Beijing 100086, China*
*Email:*darcyli@163.com

ABSTRACT

Principal component analysis is applied to analyze the horizontal component of geomagnetic data for the Panzhihua Ms 6.1 earthquake. We investigate temporal variations in eigenvalues and find that only the first principal component has good correlation with the Ap index for which the cross-correlation correlation R is larger than 0.6, which may imply solar–terrestrial activity. Both the second and third principal components show clear daily variation, being high during work hours and low at night and on weekends. The mean eigenvalue of the third component at night (00:00–04:00 LT) increased about 40 days before the earthquake and returned to normal 10 days before the earthquake. These features are likely to be correlated with the earthquake.

Keywords: Horizontal geomagnetism, Principal component analysis, Eigenvalue, Increase, Panzhihua earthquake

1   INTRODUCTION

An important task in research on earthquake prediction is to search for precursory signatures of large earthquakes. Recently, electromagnetic emissions in a wide frequency range from direct current and ultra-low frequency to very-high frequency were found to be associated with the occurrence and gestation of earthquakes (Nagao et al., 2002; Johnston, 1997). Hayakawa's groups demonstrated that electromagnetic phenomena might be promising candidates for short-term earthquake predictors, and they have published many papers on seismic magnetism in the past two decades (Hayakawa et al., 1994; Hayakawa et al., 1999; Hayakawa et al., 2000). To extract precursory magnetic signature anomalies, many new methods have been applied to analyze the variation in geomagnetism during large earthquakes in recent years, such as polarization analysis (Hayakawa et al., 1996), multifractal analysis (Ida et al., 2006), and principal component analysis (PCA) (Gotoh et al., 2002; Hattori et al., 2004; Han Peng et al., 2009).

Gotoh et al. (2002) and Hattori et al. (2004) reported the effectiveness of PCA for signal discrimination and found an ultra-low frequency geomagnetic signature associated with the Izu Islands earthquakes in 2000 (Gotoh et al., 2002; Hattori et al., 2004). They employed PCA to analyze horizontal-component (north–south) data recorded at three closely spaced stations. Both studies showed that the third eigenvalue was significant in monitoring the forthcoming large earthquake. The third principal component might include information on future earthquake activity; the component began increasing before the earthquake swarm one month earlier in the latter study than in the former study, possibly owing to the different sampling rates in the data analysis (i.e., Gotoh et al. used 1 Hz data whereas Hattori et al. used 12.5 Hz data).

Telesca et al. (2004) adopted PCA to analyze time series of the geoelectrical field measured at the Giuliano station, located in the seismically active region of Basilicata in southern Italy, from 1 March 2001 to 30 April 2004. They studied the daily time evolution of two principal components and found an interesting seismic precursory-like pattern for both components, revealing PCA to be a promising method for the monitoring of seismic areas.

Han Peng et al. (2009) employed the same method to study the geomagnetic diurnal variation associated with an M 6.1 earthquake in Japan. They investigated temporal variations in the contribution of each principal

component. The results showed that the contribution of the second principal component, which may relate to the local underground conductivity structure or local electromagnetic disturbance possibly due to a local seismogenic process, increased significantly about two weeks before the earthquake. These features are consistent with those obtained independently in previous case studies employing polarization analysis. Han's study may strengthen our understanding of seismic-electromagnetic phenomena.

The 2008 Panzhihua earthquake was an Ms 6.1 (Mw 5.7) earthquake that struck southern Sichuan province, China on 30 August 2008 at 16:30:50.5 local time. The earthquake was not an aftershock of the 2008 Sichuan earthquake since it related to different faults. With more than 400 aftershocks, it resulted in more than 40 deaths, the collapse of 10,000 homes, and damage to infrastructure in the area. The maximum intensity of the earthquake was VIII on the liedu scale.

Fortunately, there were three (Pingdi, Huili, and Nanshan) stations nearby that recorded the variation in geomagnetism during the Panzhihua earthquake. We applied PCA to analyze three components of the geomagnetic field—the horizontal component (H), vertical component (Z), and declination (D)—but found that only the horizontal component became clearer. Therefore, this paper introduces the procedure of PCA and the results for the horizontal component (H).
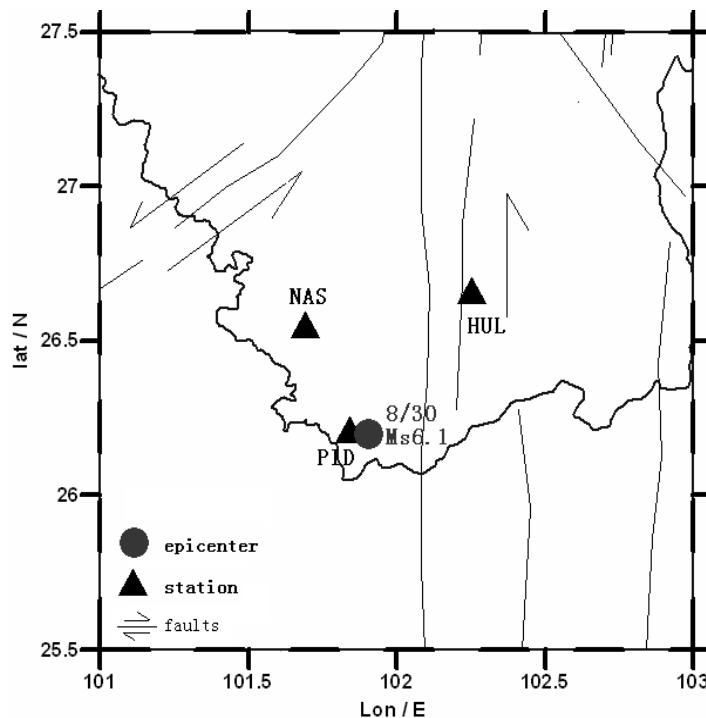

Figure 1. Relative locations of three observing stations and the epicenter of the Ms 6.1 earthquake in the Panzhihua region; thin lines are active faults.

2 OBSERVATIONS AND DATA ANALYSIS

2.1 Earthquake and Observations

The epicenter of the Panzhihua earthquake was located at 26°12′ N, 101°54′ E, 50 km southeast of the city center, and approximately in the Renhe District of Panzhihua, Sichuan, and had a depth of about 10 km. The epicenter was 60 km from Huili County in Liangshan Yi Autonomous Prefecture, Sichuan, and 30 km from Yongren County and 55 km from Yuanmou County in neighboring Yunnan province.

Figure 1 shows the relative locations of three observing stations and the earthquake epicenter; the thin lines are active faults. The three stations (HuiLi (HUL; 26.65°N, 102.25°E), NanShan (NAS; 26.54°N, 101.69°E), and PingDi (PID; 26.20°N, 101.84°E)) are in the Panzhihua region near the epicenter. The epicentral distances of the stations are roughly estimated to be about 6–65 km. The stations operate torsion-type magnetometers with three components and a sampling rate of 1 Hz. We used data from January to December 2008 to investigate the

precursor of the earthquake.

2.2   The procedure of PCA

PCA is a standard tool used in modern data analysis in diverse fields from neuroscience to computer graphics, because it is a simple, non-parametric method for extracting relevant information from confusing datasets. With minimal effort, PCA provides a roadmap for how to reduce a complex dataset to lower-dimension datasets to reveal the sometimes hidden and simple structures that often underlie a dataset (Smith, 2002).

To extract any existing geomagnetic signature of the Panzhihua earthquake, PCA was performed to investigate long-term variations due to different sources (e.g., geomagnetic variations and man-made noise).

The procedure of PCA is as follows (Smith, 2002). Step1: Collect data. The geomagnetic data recorded at the three stations are down-sampled to 1 Hz, and then all data are fed to numerical narrow band-pass filters without delay. We applied PCA to the time series data recorded at the closely separated stations. Let us consider that the time series (30 min) recorded at the three stations are given by $y_{PID} = [y_{PID}(1), y_{PID}(2),..., y_{PID}(1800)]$, $y_{NAS} = [y_{NAS}(1), y_{NAS}(2),..., y_{NAS}(1800)]$, and $y_{HUL} = [y_{HUL}(1), y_{HUL}(2),..., y_{HUL}(1800)]$, where the subscripts PID, NAS, and HUL correspond to the three stations. Step 2: Subtract the mean. For PCA to work properly we need to subtract the mean from each of the data dimensions; the mean is the average across each dimension. The data matrix $\mathbf{Y} = [y_{PID,} y_{NAS}, y_{HUL}]^{\mathbf{T}}$ is obtained, where T denotes the transpose. Step 3: Calculate the covariance matrix $\mathbf{R} = \mathbf{YY^T}$. Step 4: Calculate the eigenvectors of the covariance matrix. Since the covariance matrix is square, we can calculate the eigenvectors and eigenvalues for this matrix. The eigenvalue decomposition of $\mathbf{R}$ is $\mathbf{R} = \mathbf{VKV^T}$, where K is the eigenvalue matrix with values λ_1, λ_2, and λ_3, and V is the eigenvector matrix with columns v_1, v_2, and v_3. Here the subscripts 1, 2, and 3 indicate the order of magnitude (i.e., $\lambda_1 > \lambda_2 > \lambda_3$). The mathematical process of PCA is eigenvalue analysis of the covariance matrix of the observed signal matrix.

3   PCA EIGENVALUE RESULTS

This section describes the temporal variation of horizontal eigenvalues. Here we take the square root of eigenvalues λ_1, λ_2, and λ_3, where λ_1, λ_2, and λ_3 correspond to the first, second, and third principal components, respectively.

Figure 2 is an overall summary of the temporal evolution of the three principal components ($\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, $\sqrt{\lambda_3}$) at a frequency of 10 mHz together with the corresponding variation in horizontal geomagnetic activity expressed by the Ap index (lower panel). The upper three panels in the figure are continuous plots of the eigenvalues estimated every 30 minutes with the above procedures. The period of analysis is from January to December 2008, as seen in Figure 2, but there is a gap in the result from the end of January to the end of March because PID had technical problems in data recording, so it is not possible to use data from all three stations. Using data for the full year in Figure 2, it is difficult to find correlation of any one principal component ($\lambda_1 \sim \lambda_3$) with the Ap index.
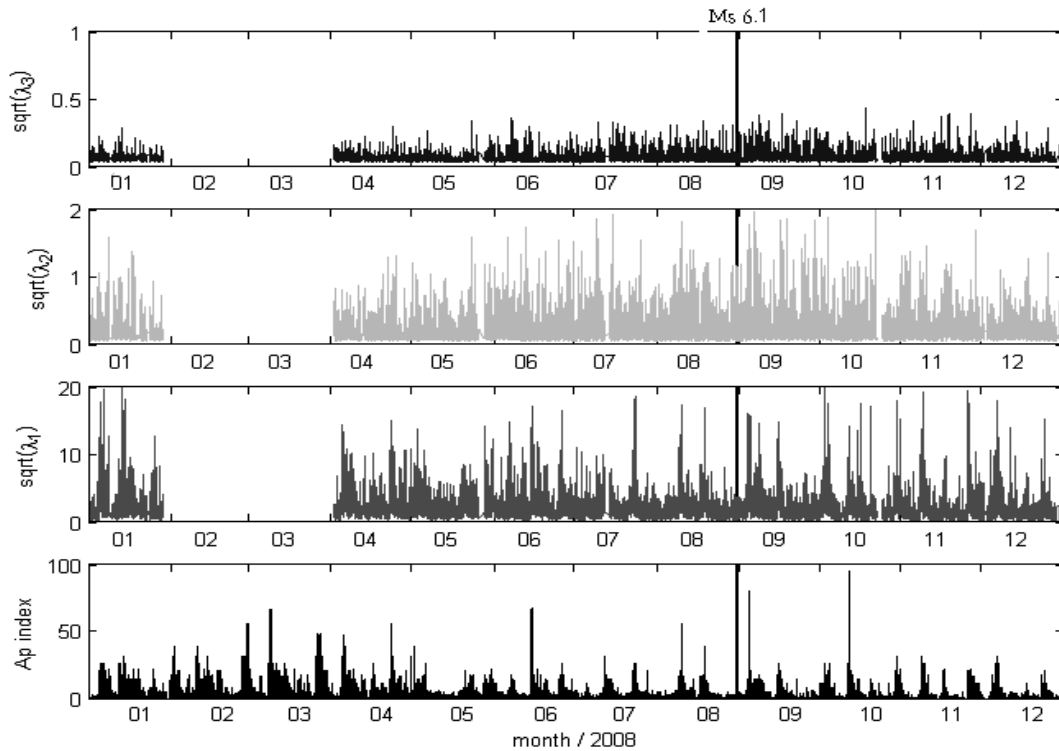
Figure 2. Overall temporal evolution of three principal components (upper three panels) compared with that of geomagnetic activity (Ap index) (lower panel) for the year 2008.

To find a relationship for these variables, we considered the data variation during April 2008. As shown in Figure 3, the variation in $\sqrt{\lambda_1}$ seems to adequately correlate with the Ap index, with there being a number of simultaneous peaks. In addition, we calculated the cross-coefficient R between each of three components and the Ap index, as shown in Figure 4. It is seen that only the variation in $\sqrt{\lambda_1}$ has good correlation with the Ap index; the coefficient R (red line) is greater than 0.6 in this case, which suggests that the first principal component may reflect signals originating from the solar activity. At the same time, the correlation coefficients for the second and third components (blue and black lines) are very low and have no regular pattern; therefore, correlations of $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$ with Ap are not so obvious.

Figure 5 shows the universal time diurnal variations in $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$ in April 2008. Both the second and third components (Figures 5(a) and 5(c)) have clear daily variation with there being high values during work hours (LT = UT + 8:00) and low values at night and on weekends (Figures 5(b) and 5(d)). Both components seem to have good correlation with human activity (i.e., work during the daytime and rest at night). Therefore, to discover precursory signatures of an earthquake, we should use nighttime data because there is less human activity at that time.
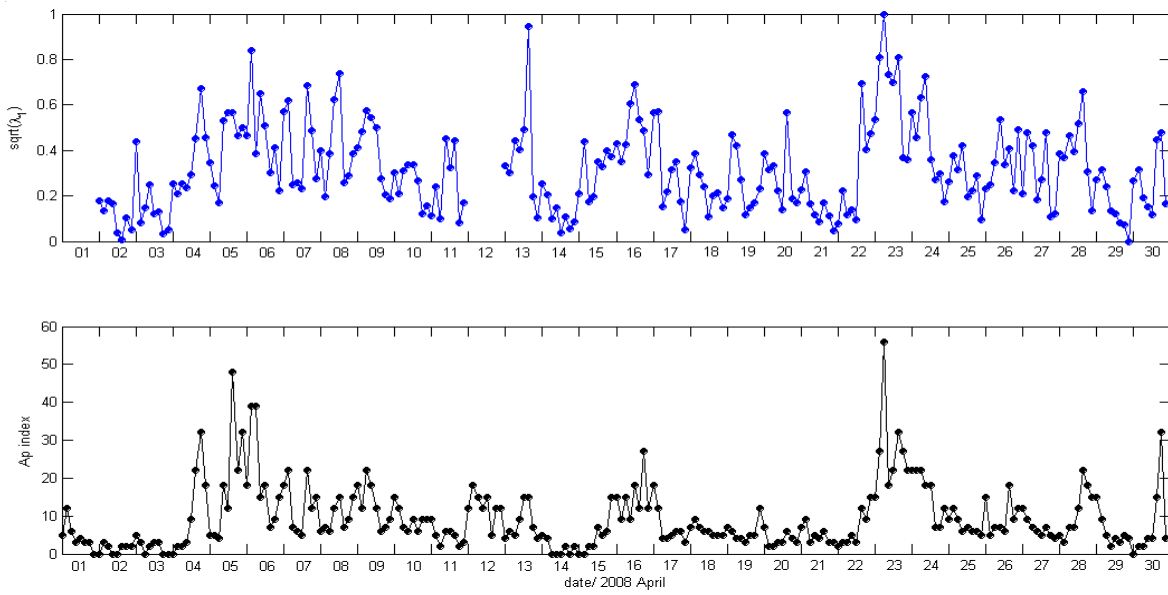
Figure 3. Comparison of the temporal evaluation of the first principal component $\sqrt{\lambda_1}$ with that of the Ap index during the short period of 1–30 April 2008.
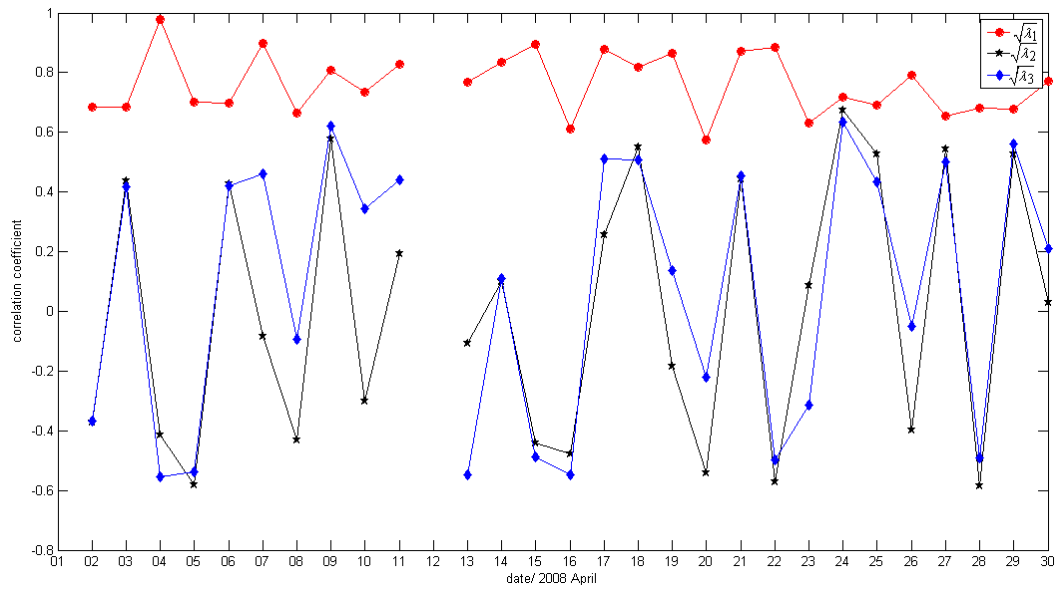

Figure 4. Coefficient of the correlation of three principal components with the Ap index during the period of 1–30 April 2008; the red line indicates $\sqrt{\lambda_1}$, the black line $\sqrt{\lambda_2}$, and the blue line $\sqrt{\lambda_3}$.
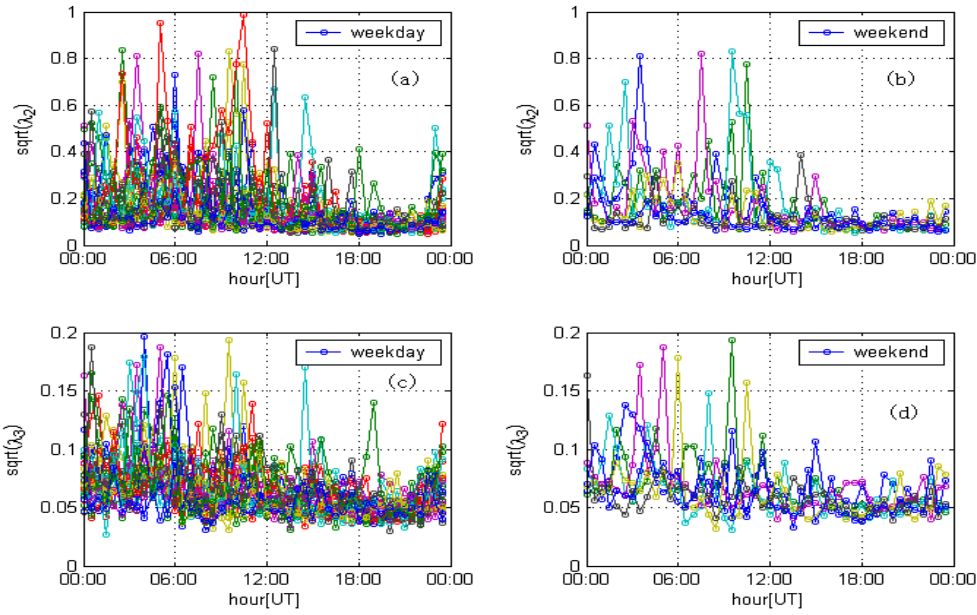
Figure 5. Variations in $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$ on each day of April 2008. Figures (a) and (c) show the diurnal variations in $\sqrt{\lambda_2}$ and $\sqrt{\lambda_3}$ on weekdays respectively, while Figures (b) and (d) show those on the weekend. LT = UT + 8:00.
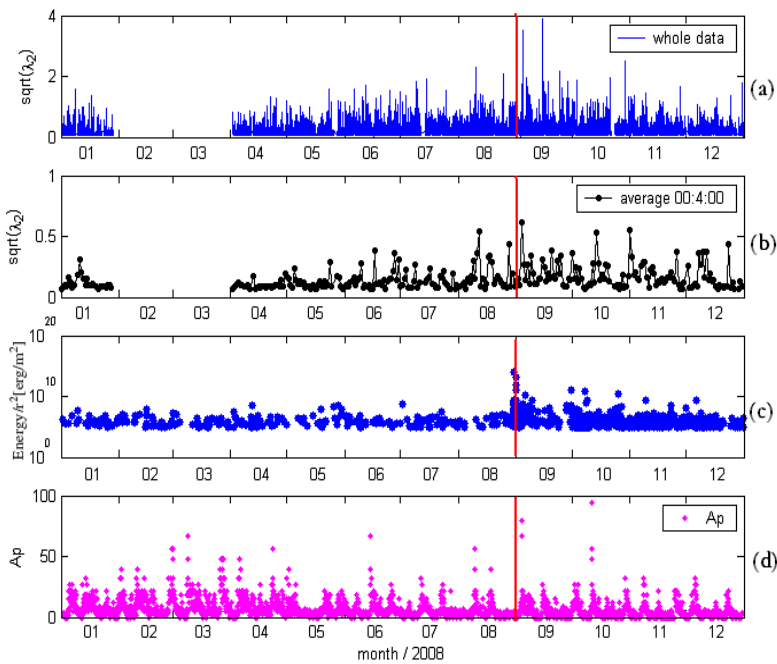

Figure 6. Variation in the second principal component $\sqrt{\lambda_2}$ during 2008. The occurrence time of the large earthquake (Ms 6.1) is indicated by a vertical line. (a) Variation in $\sqrt{\lambda_2}$ for a whole day. (b) Variation in the midnight average (00:00–04:00 LT) of $\sqrt{\lambda_2}$. (c) Released energy (E/r^2) around the array station (r < 100 km). (d) Variation in the Ap index.
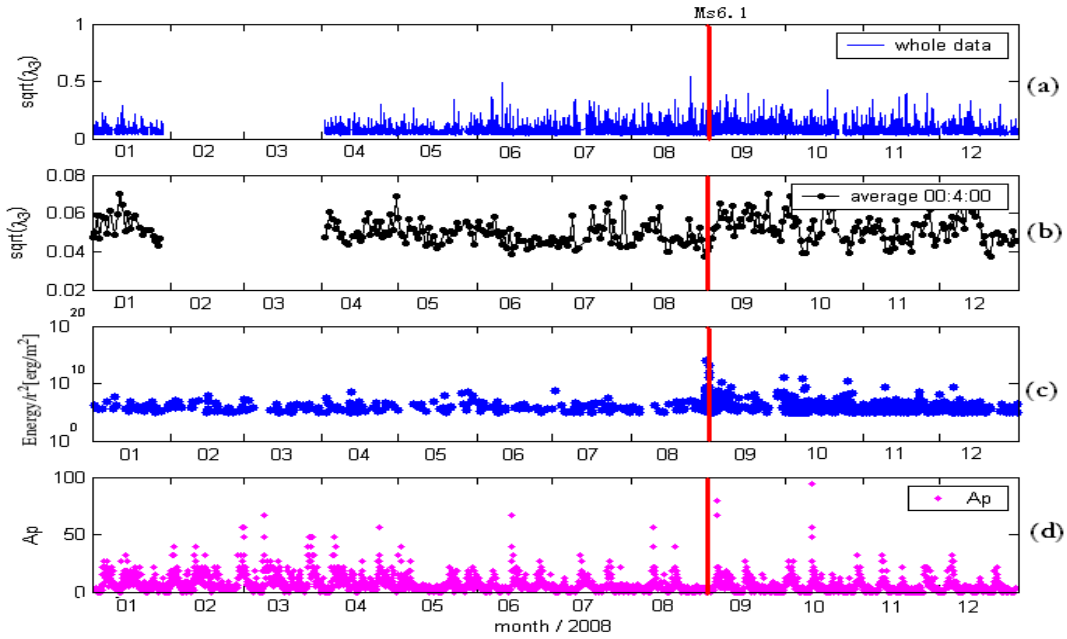
Figure 7. Variation in the third principal component $\sqrt{\lambda_3}$ during 2008. The occurrence time of the large earthquake (M > 6) is indicated by a vertical line. (a) Variation in $\sqrt{\lambda_3}$ for a whole day. (b) Variation in the midnight average (00:00–04:00 LT) of $\sqrt{\lambda_3}$. (c) Released energy (E/r^2) around the array station (r < 100 km). (d) Variation in the Ap index.
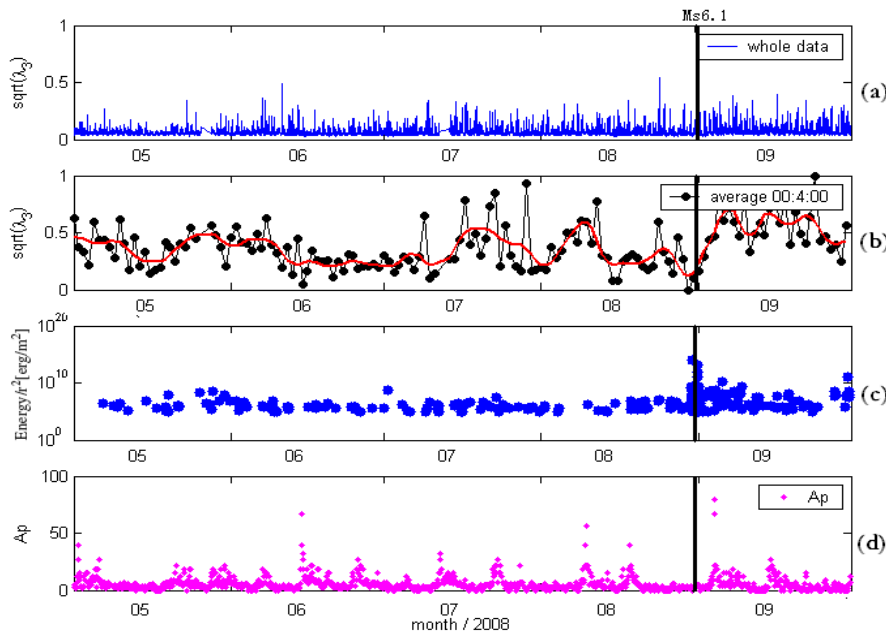

Figure 8. Variation in the third principal component $\sqrt{\lambda_3}$ from May to September 2008. The occurrence time of the large earthquake (Ms 6.1) is indicated by a vertical line. (a) Variation in $\sqrt{\lambda_3}$ for a whole day. (b) Variation in the midnight average (00:00–04:00 LT) of $\sqrt{\lambda_3}$, with the red line being the 5-day running mean. (c) Released energy (E/r^2) around the array station (r < 100 km). (d) Variation in the Ap index.

Figure 6 shows the variations in $\sqrt{\lambda_2}$ (Figure 6(a) shows the whole-day variation and Figure 6(b) the nighttime variation), the seismicity E/r^2 (Figure 6(c)), and the Ap index (Figure 6 (d)). The results show that the variation in $\sqrt{\lambda_2}$ does not change throughout the year, and it is difficult to find an anomaly before or after the earthquake.

The weakest third component $\sqrt{\lambda_3}$, which may be the residual of the first and second principal components, is the third possible signal candidate to contain hidden earthquake-related emissions (Figure 7). As shown in Figures 5(c) and 7(a), $\sqrt{\lambda_3}$ is likely to have diurnal variation, being highly affected by man-made signals during the day but little affected at night (Figure 7(b)). Therefore, the component could be used to detect weak earthquake-related emissions, if they exist. The variation in midnight $\sqrt{\lambda_3}$ in Figure 7(b) shows that the amplitude of $\sqrt{\lambda_3}$ began to increase from the middle of July 2008, about 40 days before the Ms 6.1 earthquake, and then sharply decreased about 10 days before the earthquake. This kind of variation is not seen for the Ap index (Figure 7(d)). After approximately 4 months, $\sqrt{\lambda_3}$ became rather stable and returned to the level seen at the beginning of the plot.

We need to investigate the possible causes of the increase in $\sqrt{\lambda_3}$ and any shaking of the sensor by mechanical seismic waves. Therefore, we use the China Earthquake Net Catalog to investigate the regional seismicity E/r^2, where E and r are the energy released by an earthquake and the hypocentral distance from a certain array station. The formula $\log E = 1.695ML + 3.18$ is adopted for this computation. Figure 7(c) shows the index of E/r^2, which indicates regional seismic energy. The seismic energy decays macroscopically as an exponential function, but the changes in $\sqrt{\lambda_3}$ are quite different. Figure 8 is an enlarged plot of the nighttime variation in $\sqrt{\lambda_3}$ from May to September 2008 and shows the above feature more clearly. After the strong earthquake, $\sqrt{\lambda_3}$ sharply increased from 31 August, but at the same time, the Ap index was very stable, perhaps indicating a post-seismic effect.

4   DISCUSSION AND CONCLUSIONS

The PCA results for horizontal geomagnetism for the Ms 6.1 Panzhihua earthquake indicate that the first principal component originates from solar effects since the coefficient of correlation between Ap and $\sqrt{\lambda_1}$ is larger than 0.6, whereas the other second and third principal components do not obviously correlate with Ap. The second and third components have clear daily variation of high values during work hours and low values at night and on weekends, which suggests they are a combination of artificial and possible earthquake-related signals. It is difficult to find an anomalous variation in the nighttime average $\sqrt{\lambda_2}$ at the time of the earthquake (Fig. 6). To avoid artificial signals, we used average nighttime (00:00–04:00 LT) data for $\sqrt{\lambda_3}$. The results indicate that $\sqrt{\lambda_3}$ increased about 40 days before the earthquake and returned to normal 10 days before the earthquake and sharply increased after the quake and returned to normal 3 months later. The correlation of $\sqrt{\lambda_3}$ with the earthquake is rather better than the correlation of $\sqrt{\lambda_2}$ with the earthquake.

We also rule out the Ap index and the possibility of shaking effects as possible causes of the anomalous increases in the third principal component preceding (a few days prior to) the large earthquake. It is clearly found that the variation in Ap is smooth and the released energy (E/r^2) is normal with 100 km of the epicenter during the three months before the earthquake. These observations suggest the credibility of the increase in $\sqrt{\lambda_3}$ about 40 days before the earthquake as a precursor. The $\sqrt{\lambda_3}$ increase is weaker than that found by Hattori (2004), possibly because the stations are more separated or the earthquake is less intense in our study.

5   ACKNOWLEDGEMENTS

6   REFERENCES

Gotoh, K., Akinaga, Y., Hayakawa, M., & Hattori, K. (2002) Principal component analysis of ULF geomagnetic data for Izu islands earthquakes in July 2000. *J. Atmos. Electr.* 22, 1–12.

Han, P., Huang, Q. H., & Xiu, J. G. (2009) Principal component analysis of geomagnetic diurnal variation associated with earthquakes：case study of the M6．1 1wate ken Nairiku Hokubu earthquake．*Chinese Geophysics (*in Chinese*)*52(6):556～1563.

Hattori, K., Takahashi, I., Yoshino, C., Isezaki, N., Iwasaki, H., Harada, M., et al. (2004) ULF geomagnetic

field measurements in Japan and some recent results associated with Iwateken Nairiku Hokubu Earthquake in 1998. *Phys. Chem. Earth* 29(4-9):481-494.

Hayakawa, M. (Ed.) (1999) *Atmospheric and Ionosphere Electromagnetic Phenomena Associated with Earthquakes*. Terra Scientific Publishing Co., Tokyo, p. 996.

Hayakawa, M. & Fujinawa, Y. (Eds.) (1994) *Electromagnetic Phenomena Related to Earthquake Prediction.* Terra Scientific Publishing Co., Tokyo, p. 667.

Hayakawa, M., Itoh, T., Hattori, K., & Yumoto, K. (2000) ULF electromagnetic precursors for an earthquake at Biak, Indonesia on February 17, 1996. *Geophys. Res. Lett*. 27, 1531–1534.

Hayakawa, M., Itoh, T., & Smirnova, N. (1999) Fractal analysis of ULF geomagnetic data associated with the Guam earthquake on August 8, 1993. *Geophys. Res. Lett*. 26, 2797–2800.

Ida, Y., Hayakawa, M., et al. (2006) Fractal analysis for the ULF data during the 1993 Guam earthquake to study prefracture criticality. *Nonlinear Processes in Geophysics 13*: 409～41 2.

Johnston, M. (1997) Review of electric and magnetic fields accompanying seismic and volcanic activity. *Surveys in Geophysics* 18: 441- 4 76.

Nagao T., Enomoto, Y.，Fujinawa, Y, et a1. (2002) Electromagnetic anomalies associated with 1996 Kobe earthquake. *Journal of Geodynamics* 33(4～5): 401～411.

Smith, L. (2002) *A Tutorial on Principal Component Analysis 1*:1-27.

Telesca, L., Colangelo, G., Hattori, K., & Lapenna, V. (2004) Principal component analysis of geoelectrical signals measured in the seismically active area of Basilicata Region (southern Italy). *Natural Hazards and Earth System Sciences* 4: 663–667.