Available online at www.sciencedirect.com
SCIENCE DIRECT
Physics and Chemistry of the Earth 31 (2006) 281–291
PHYSICS and CHEMISTRY of the EARTH
www.elsevier.com/locate/pce

Singular spectral analysis and principal component analysis for signal discrimination of ULF geomagnetic data associated with 2000 Izu Island Earthquake Swarm

Katsumi Hattori [a,*], Aya Serita [b,1], Chie Yoshino [a], Masashi Hayakawa [c], Nobuhiro Isezaki [d]

[a] Marine Biosystems Research Center, Chiba University, Inage, Chiba 263-8522, Japan
[b] Graduate School of Science and Technology, Chiba University, Inage, Chiba 263-8522, Japan
[c] The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan
[d] Department of Earth Science, Faculty of Science, Chiba University, Inage, Chiba 263-8522, Japan

Accepted 6 February 2006

Abstract

In order to extract any ULF signature associated with earthquakes, the singular spectral analysis (SSA) and the principal component analysis (PCA) have been performed to investigate a possibility of discrimination of signals from different sources (geomagnetic variation, artificial noise, and the other sources (earthquake-related ULF emissions)). We apply SSA to the time series data observed at closely spaced stations; Seikoshi (SKS), Mochikoshi (MCK), and Kamo (KAM) stations. Then, PCA is applied to the time series data sets filtered at 0.01 Hz of NS component at three stations. In order to remove the most intense signal like the first principal component, we make the differential data sets of SKS–MCK and MCK–KAM for the above data. The major findings are summarized as follows. (1) It is important to apply simultaneously SSA and PCA, because SSA gives the structure of signals and the number of sensors for PCA is estimated. This makes the results more convincing. (2) There is a significant advantage using PCA with differential data sets of filtered (0.01 Hz band) signal between SKS–KAM and MCK–KAM in NS component for removing the most intense signal like global variation (solar–terrestrial interaction). This yields that the anomalous changes in the second principal component were detected more clearly, and the contribution of the second principal component is found to be 20–40%. It is enough to prove mathematical accuracy of the signal. Further application is required to accumulate events. These facts demonstrate a possibility of monitoring the crustal activity with using the SSA and PCA.

Keywords: Singular spectral analysis (SSA); Principal component analysis (PCA); Signal discrimination; ULF magnetic data; 2000 Izu Island Earthquake Swarm; Earthquake-related ULF emissions; PCA with differential dataset

1. Introduction

Many studies have been reported on electromagnetic phenomena associated with earthquakes (e.g. Hayakawa and Fujinawa, 1994; Hayakawa, 1999; Hayakawa and Molchanov, 2002; Hattori, 2004; Hayakawa and Hattori, 2004). One of the most frequent studies among them is to investigate the relationship between ULF (ultra low frequency) geomagnetic anomalous changes and earthquake activity (Fraser-Smith et al., 1990; Bernardi et al., 1991; Hattori, 2004; Hayakawa and Hattori, 2004). In comparison with other higher frequency ranges, ULF has an advantage in low-loss propagation in the crust due to skin depth. Therefore, we investigate the anomalous ULF geomagnetic changes associated with 2000 Izu Islands

* Corresponding author. Tel.: +81 43 290 2801; fax: +81 43 290 2859.
E-mail address: hattori@earth.s.chiba-u.ac.jp (K. Hattori).
[1] Now at Japan Defense Agency, Japan.

1474-7065/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2006.02.034

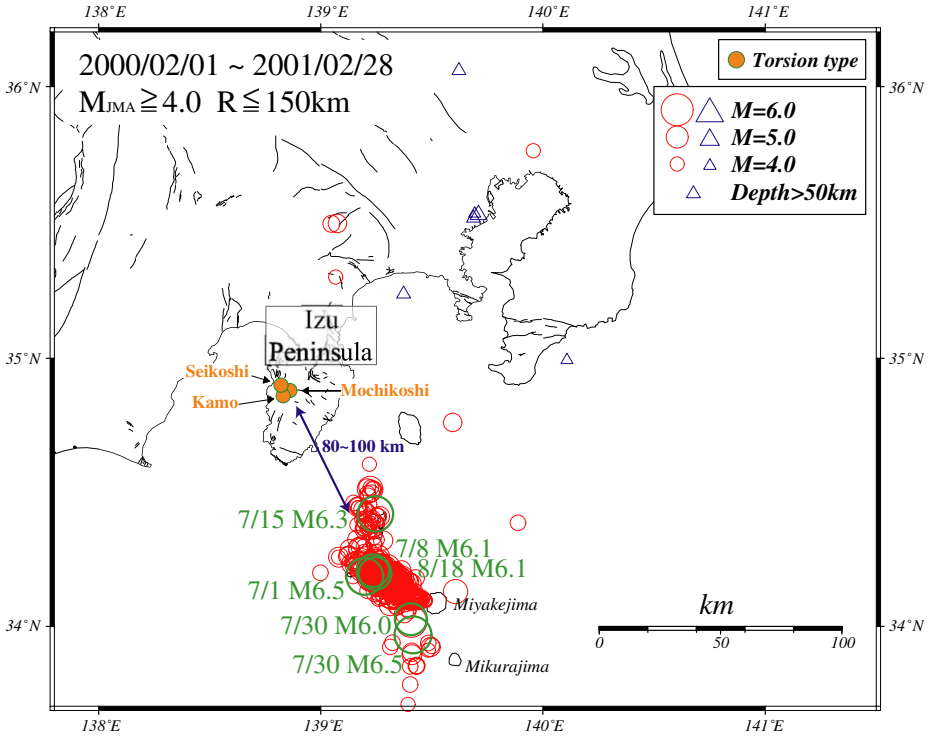

Fig. 1. Map of ULF geomagnetic stations (array station) and seismic activity of 2000 Izu Islands Earthquake.

Earthquake Swarm with using singular component analysis (SCA) and principal component analysis (PCA).

Miyake-jima Island is a basaltic volcano, about 150 km south of Tokyo, Japan. On June 26, 2000, swarm earthquakes suddenly became active beneath the summit. Then, it is observed migration of magma and a small-scale sea floor eruption were observed. After a while, earthquake epicenters also migrated from the island first westward and then northward. During this period 6 large (>M6.0) earthquakes, 7/1:M6.5, 7/8:M6.1, 7/15:M6.3, 7/30:M6.5, M6.08/18:M6.1 occurred within 150 km from Mochikoshi station, Izu as shown in Fig. 1. They occurred in shallow depth less than 20 km. On the other hand, phreatic explosion and the formation of the initial sink at the summit of the volcano occurred on July 8. The August 18, 2000 largest eruption was followed by low temperature ash flow on August 29.

We measured three component ULF geomagnetic fields and two horizontal geoelectrical potential differences at Izu Peninsula, Japan (Hattori et al., 2004b). There is an array with intersensor distance of 5 km as shown in Fig. 1. The names of stations are Mochikoshi (MCK), Seikoshi (SKS), and Kamo (KAM). Tables 1 and 2 summarize the position of stations and large earthquakes with $M > 6$. Observed ULF geomagnetic fields are considered to be a superposition of solar-origin signal, artificial noise, and other noises propagated in the crust. The signal associated with crustal activity is generally very week, so that sophisticated signal classification methods are required. In this paper, we applied singular spectral analysis (SSA) method to the observed data in order to discriminate signals associated with the crustal activity from other sources (geomagnetic variation, artificial noise, and so on).

Table 1
The geographical location of Izu array station

Name of station	Latitude (North, deg)	Longitude (East, deg)
Kamo (KAM)	34.86	138.83
Seikoshi (SKS)	34.90	138.82
Mochikoshi (MCK)	34.88	138.86

Table 2
The location of earthquakes with $M > 6$ near the Izu array station during February 2000–2001

Origin time (UT)	Magnitude (JMA)	Latitude (North, deg)	Longitude (East, deg)	Depth (km)	Distance from MCK (km)
20000701 07 h 01 min 56 s	6.5	34.187	139.197	16	83
20000708 18 h 57 min 44 s	6.1	34.209	139.234	15	82
20000715 01 h 30 min 32 s	6.3	34.420	139.245	9	62
20000730 00 h 28 min 02 s	6.0	34.027	139.405	11	107
20000730 12 h 25 min 46 s	6.5	33.968	139.414	17	113
20000818 01 h 52 min 22 s	6.1	34.198	139.244	12	84

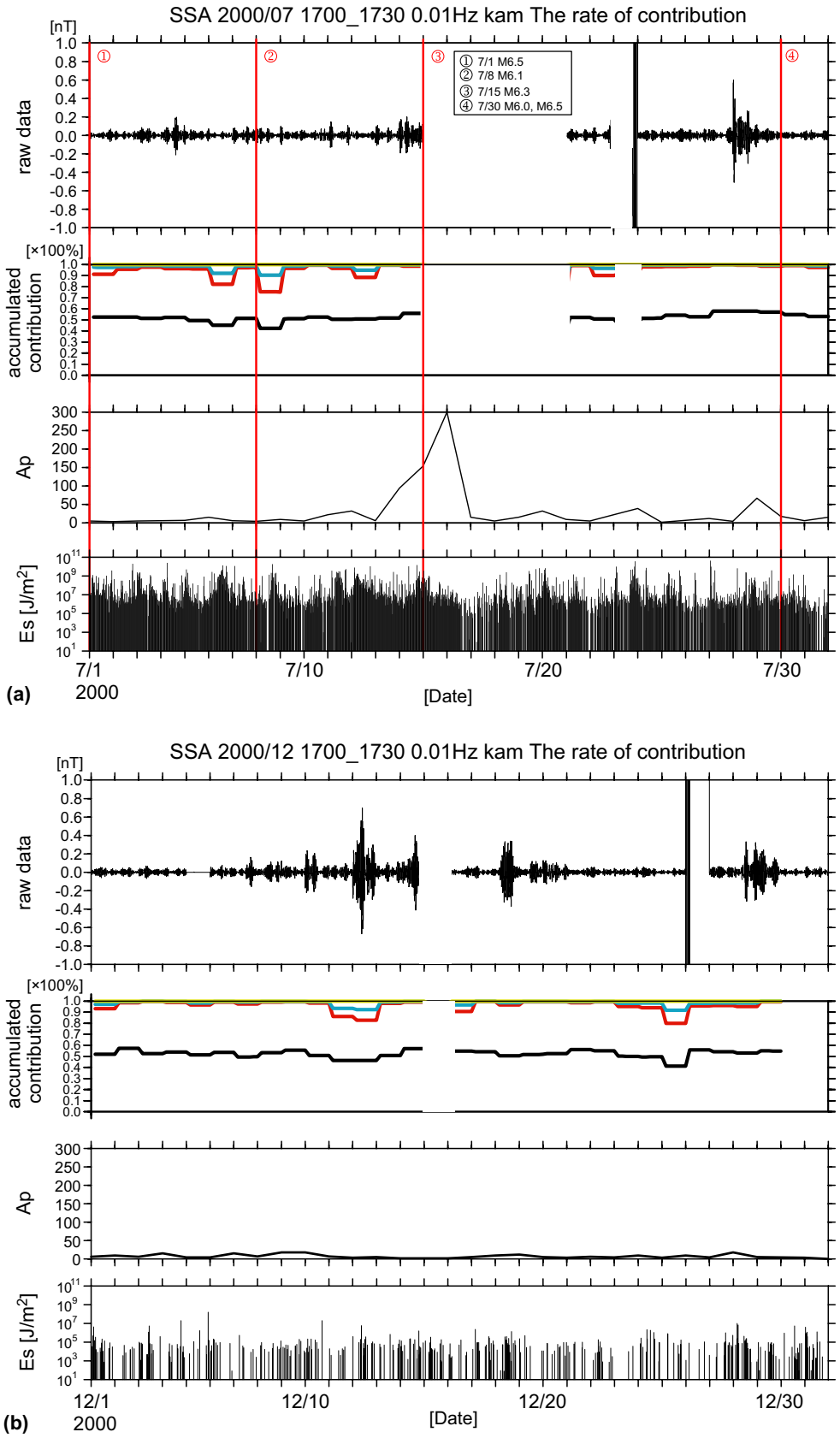

Fig. 2. (Colour online) Results of SSA at Kamo station (KAM) in July, 2000 (a) and in December, 2000 (b). The top panel indicates the ULF magnetic variation in NS component at 0.01 Hz band. The second panel indicates the accumulated contribution. The third and bottom panel indicate the variation of Ap index and seismic activity Es around Izu array station. The vertical lines indicate the day of earthquake with $M > 6$.

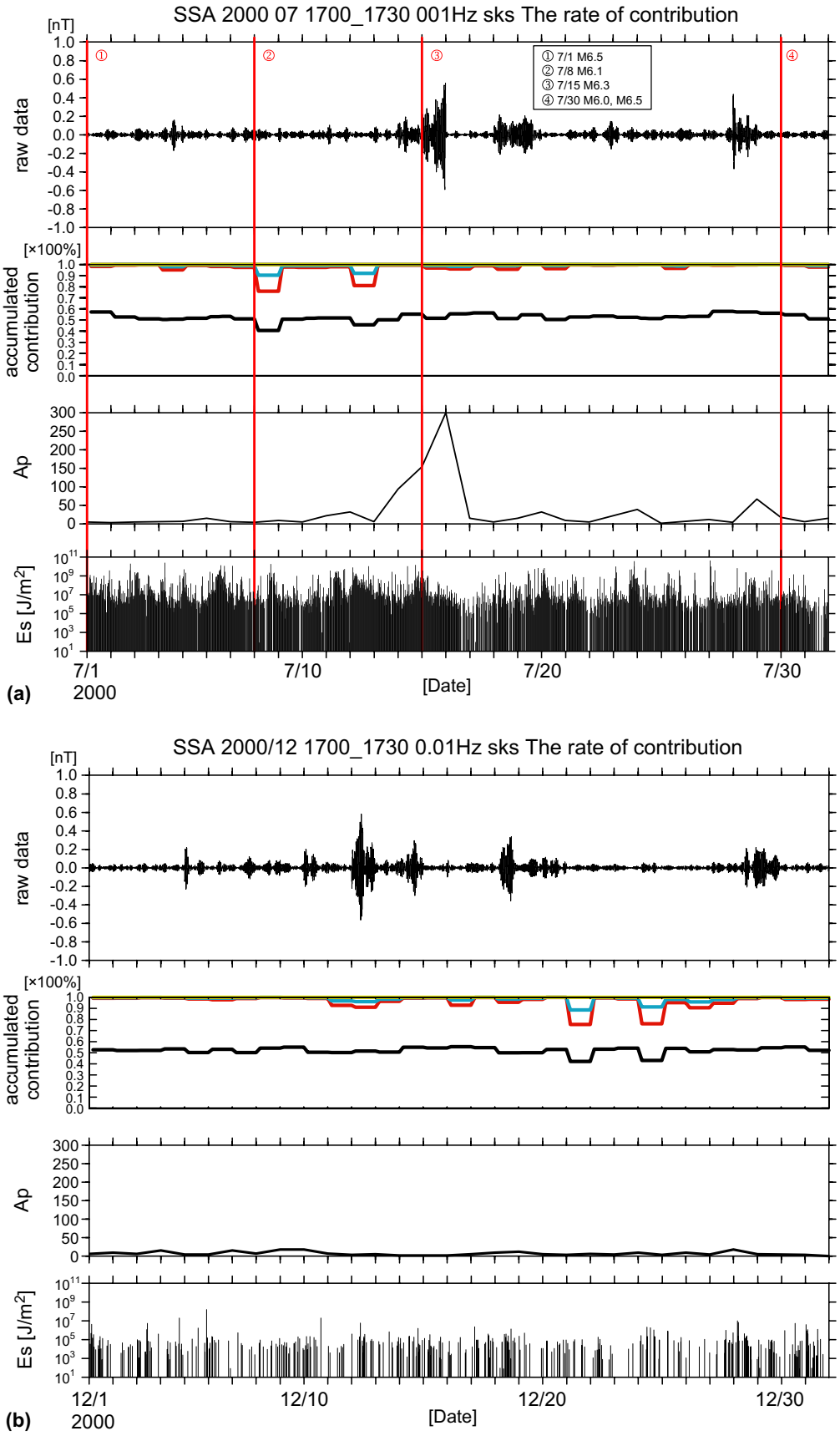

Fig. 3. (Colour online) Results of SSA at Seikosho station (SKS) in July, 2000 (a) and in December, 2000 (b). See caption of Fig. 2.

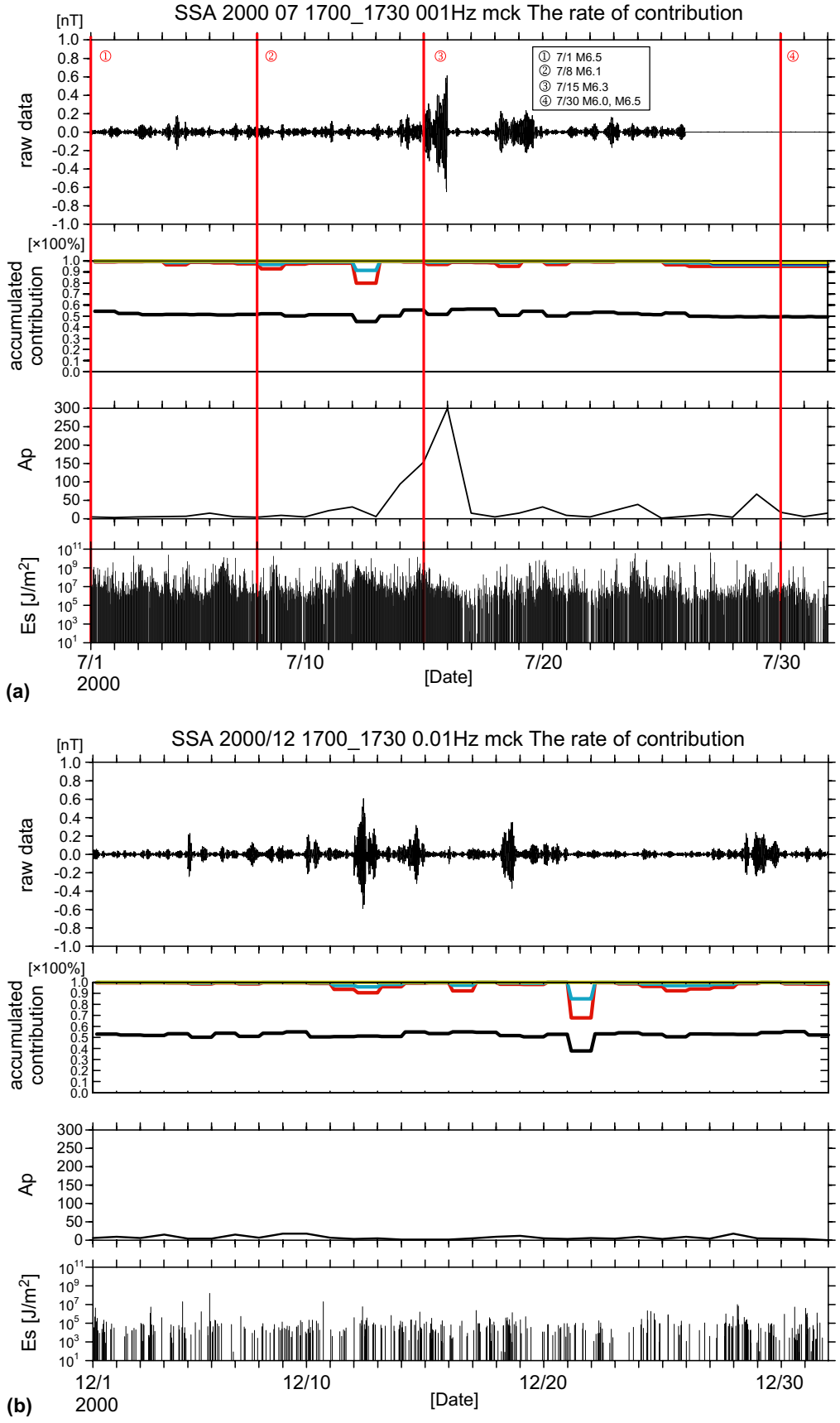

Fig. 4. (Colour online) Results of SSA at Mochikoshi station (MCK) in July, 2000 (a) and in December, 2000 (b). See caption of Fig. 2.

2. Singular spectral analysis (SSA) of ULF geomagnetic data

In order to estimate the number of signals observed at each station, singular spectral analysis (SSA) (Golyandina et al., 2001; Troyan and Hayakawa, 2002) has been applied to the time series data at 0.01 Hz. SSA is sometimes called caterpillar method. The mathematical basis is very similar to PCA. The procedure is as follows. We consider one dimensional time series data $\{x_i\}_{i=1}^{N}$. The multidimensional matrix X is built by using the evolvement of the 1-D time series,

$$X = (x_{ij})_{i,j=1}^{k,M} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_M \\ x_2 & x_3 & x_4 & \cdots & x_{M+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \cdots & x_N \end{bmatrix}.$$

Here M is caterpillar length with $M < N/2$ and $k = N - M + 1$. Then the matrix X is normalized. The components of the normalized matrix X^* are given by

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{s_j}$$

where $\bar{x}_j = \frac{1}{k}\sum_{i=1}^{k} x_{i+j-1}$ and $s_j = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(x_{i+j-1} - \bar{x}_j)^2}$.

Then, we computed the covariance matrix $R = \frac{1}{N}(X^* X^{*\mathrm{T}})$ and the singular value decomposition of $R = V \Lambda V^{\mathrm{T}}$ (Λ: eigenvalue matrix) has been performed. Here T means transpose. The procedure is very similar to that of PCA except the dimension of the matrix. In this paper, the variations of accumulated contribution of singular values have been investigated for the number of signals detected at the station. In this analysis, we have taken $M = 1800$ and $N = 3750$ (12.5 Hz × 5 min × 60 s).

3. Principal component analysis (PCA) of ULF geomagnetic data

In this paper, we applied principle component analysis (PCA) method to the observed data at the arrays to investigate the signal discrimination of different sources (geomagnetic variation, artificial noise, and the other sources

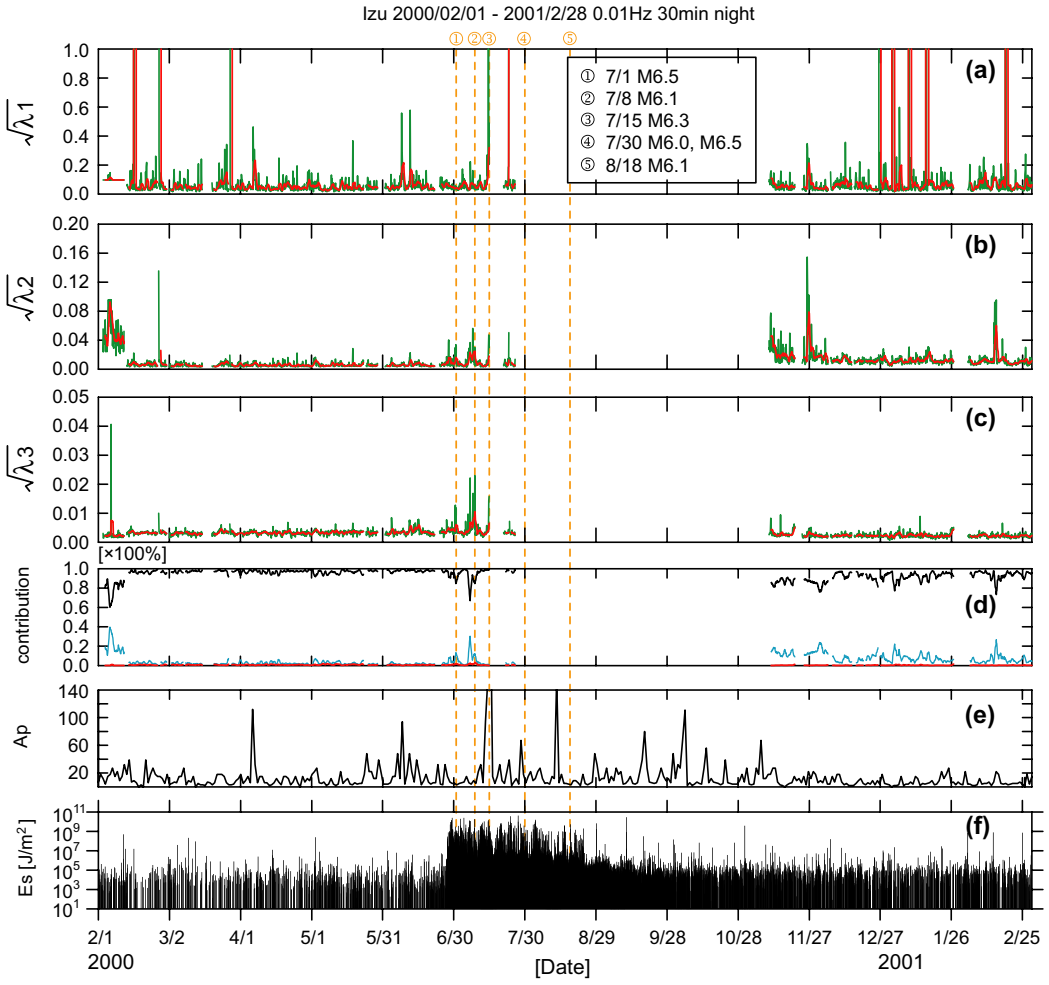

Fig. 5. (Colour online) Results of PCA using the three-station data. (a) Eigenvalue of the first principal component. (b) Eigenvalue of the second principal component. (c) Eigenvalue of the third principal component. (d) Contribution rate of principal components. Black, blue and red lines indicate the total contribution up to the first, up to the second and up to the third principal component, respectively. (e) Es, i.e. the seismic activity around the station MCK. The vertical lines indicate the day of earthquake with $M > 6$.

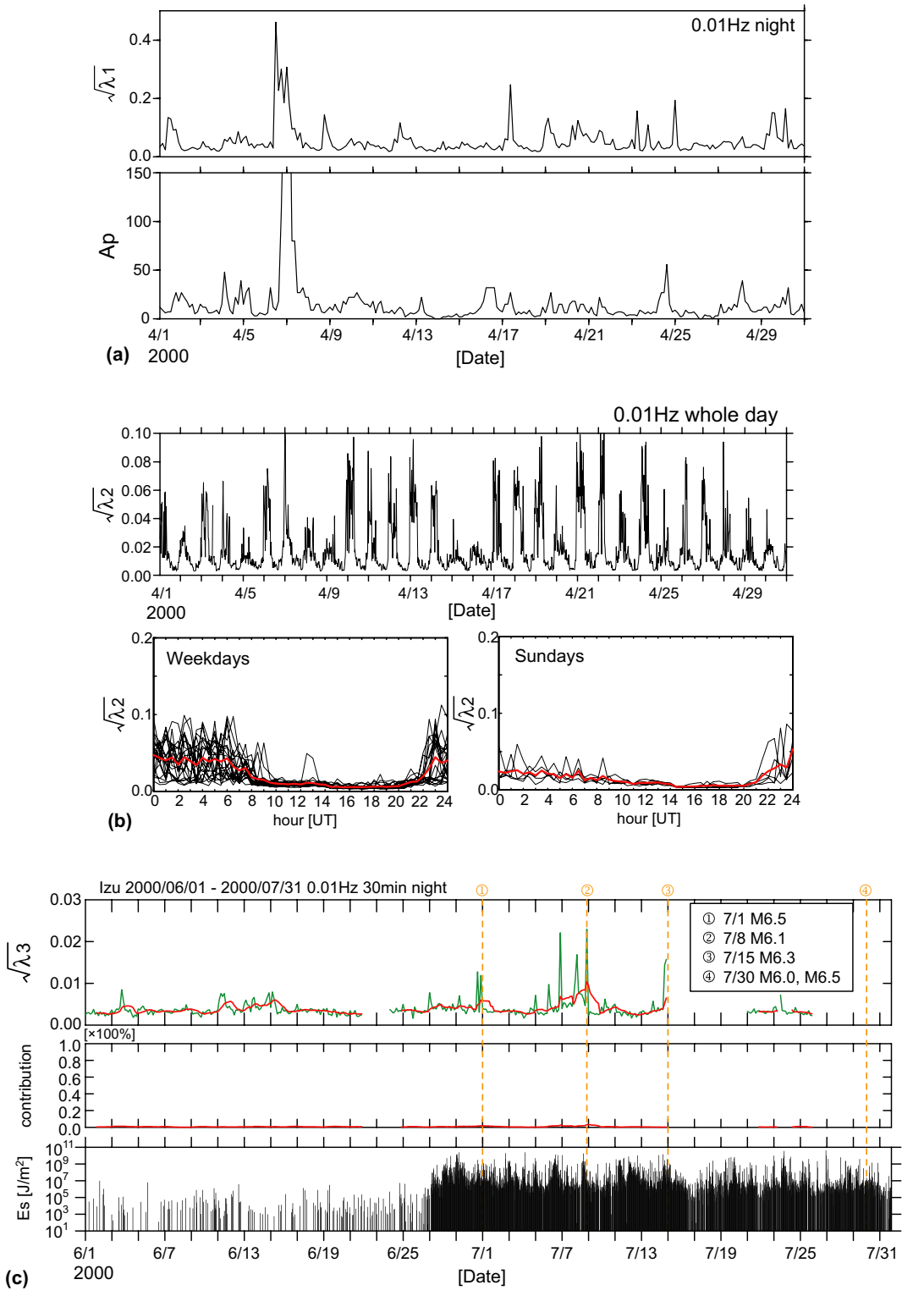

Fig. 6. (Colour online) Detailed results of PCA for the three-station data. (a) First principal component $\sqrt{\lambda_1}$ in the night time together with that of Ap index in April 2000. (b) Second principal component in all day data $\sqrt{\lambda_2}$ in April, 2000, diurnal variation of $\sqrt{\lambda_2}$ on week days, and on holidays. The red line indicates the average variation. (c) Third principal component $\sqrt{\lambda_3}$ in the interval of June–July, 2000, variation of contribution rate of the third principal component, and Es index. The vertical lines indicate the day of earthquake with $M > 6$.

(like earthquake-related ULF emissions)). We apply PCA to the time series data observed at closely spaced stations. The analyzed period is from February 2000 to 2001 (13 months). In this paper, we show the results of 0.01 Hz band with 12.5 Hz sampling rate. Let us assume that the time series data (30 min) observed at each station is given by

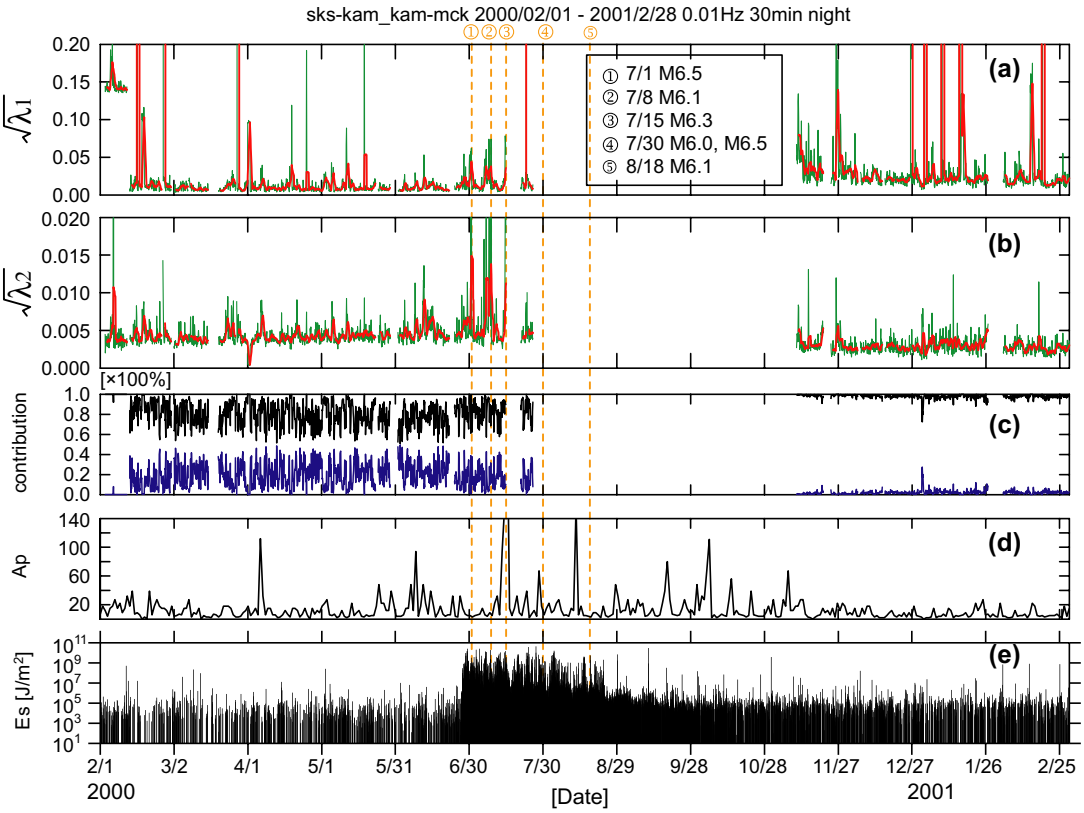

Fig. 7. (Colour online) Results of PCA with differential data. (a) Eigenvalue of the first principal component. (b) Eigenvalue of the second principal component. (c) Variation of contribution rates. Black and blue indicate first and second principal component, respectively. (d) Ap index. (e) Es index. The vertical lines indicate the day of earthquake with $M > 6$.

$X(n) = [x_n(t_1), x_n(t_2), ,. . . x_n(t_N)]$, where n is the index of the site and N is the number of the data. The data matrix $X = [x_1 x_2 x_3]^{\mathrm{T}}$ is obtained, and then we calculate the covariance matrix $R = \frac{1}{N}(X^* X^{*\mathrm{T}})$. The eigenvalue decomposition of $R = V \Lambda V^{\mathrm{T}}$ has been performed, and we investigate the temporal variations of eigenvalues (Λ) and eigenvectors (V).

In addition, we perform the PCA with differential data at Izu Peninsula; that is, we use the difference between the two stations, SKS–KAM and MCK–KAM data as **X**(1) and **X**(2) in order to remove the coherent influence in the region of the array.

4. Results of SSA and PCA

The data during 13 months from February 2000 to 2001 have been analyzed by means of the proposed SSA and PCA methods. The results are shown in this section.

4.1. SSA results at each station

Fig. 2 shows the results of SSA around 0.01 Hz band in the nighttime (17:00–17:30 UT (02:00–02:30 JST)) data observed at KAM. Fig. 2(a) and (b) show the result of SSA in July, 2000 and that in December, 2000, respectively. Top panel indicates the input data for SSA (0.01 Hz band signal) and the second panel shows the accumulation of the contribution from 1st, 2nd, 3rd and 4th components. Black line indicates the contribution of the first principal component. Red, blue and yellow line indicate the total contribution (in percentage) up to the second, up to the third and the up to the fourth principal components. The third panel is the variation of geomagnetic activity Ap. The bottom panel is seismic activity Es around the Izu array station and Es means E/r^2 computed from the 30 min sum of the released energy E around Izu array station and their epicentral distance from MCK station. In this calculation earthquakes with $r < 150$ km from MCK station are selected. Es is considered to be the seismically felt energy at the site. The signals at 0.01 Hz band observed at KAM consist of at most four components, but usually they consist of at most three components.

Similar results are obtained at the other two, SKS and MCK stations as shown in Figs. 3 and 4. There seems to be no apparent relation neither to the seismic nor to the geomagnetic activity. Therefore, it is safe to say that the structure of the observed signal at three stations is very similar and we have to consider generally less than three signal sources. It means that the PCA with using three station data observed at Izu peninsula in the frequency range of 0.01 Hz is valid, but four-station data analysis is desirable.

4.2. PCA results of three station data

In this paper, we propose the differential PCA results because normal PCA results have already been reported (Gotoh et al., 2002; Hattori et al., 2004a). However, we will mention them briefly. Fig. 5(a)–(c) show the variation during the interval of 2000 and 2001 of the first, second, and third eigenvalues, and Fig. 5(d) refers to their contribution. Fig. 5(e) indicates the temporal evolution of Ap and Fig. 5(f), Es. The corresponding eigenvectors show that of the first principal component is very stable and the other two show scattered variations during February–July 2000 (Hattori et al., 2004a). Fig. 6(a) indicates the detailed variation of eigenvalue of the first principal component (upper panel) in April 2000 and it is similar to that of Ap index (lower panel). The corresponding eigenvectors are found to be stable. These facts suggest that the dominant origin of the first principal component may be of solar–terrestrial interaction. Fig. 6(b) shows that the variation of eigenvalue of the second principal component (upper panel) seems to be related to that of electricity consumption power around the stations. Fig. 6(c) shows the variation of eigenvalue of the third principal component (residual) in June–July, 2000 and we have found the peak values a few days before the earthquakes with M greater than 6. The variation of the third principal component indicates that there is also a local maximum about two weeks before the swarm activity. However, there remains a problem because the contribution of the third principal component is too small.

4.3. PCA results of differential station data

In order to remove the most intense signal like the first principal component in the three-station data analysis, we make the differential data sets of SKS–KAM and MCK–KAM in NS component. On the other hand, the signal associated with the earthquake has a small propagation velocity and there is an arrival time difference among the

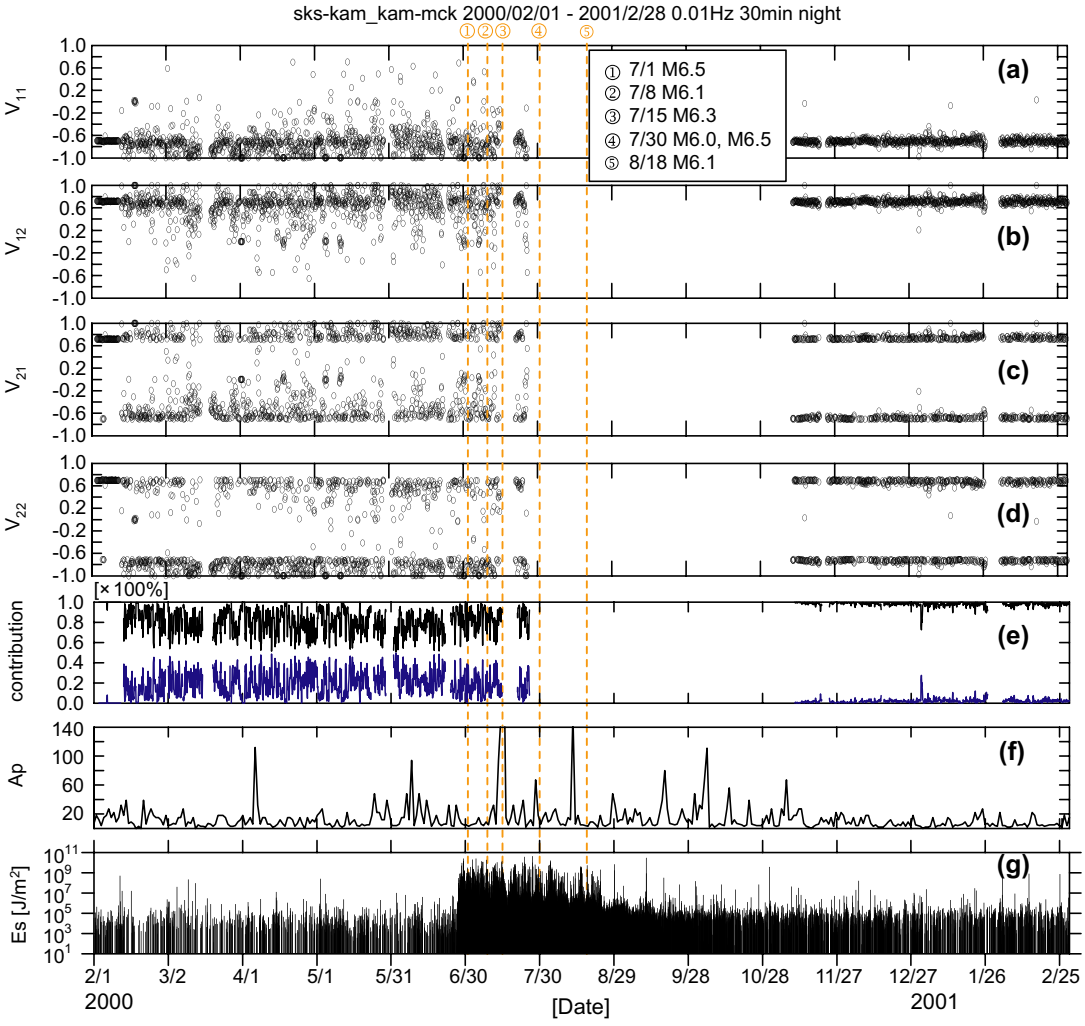

Fig. 8. (Colour online) Results of PCA using the three-station data. (a) and (b) Eigenvector of the first principal component, v_{11} and v_{12}. (c) and (d) Eigenvector of the second principal component, v_{21} and v_{22}. (e) Contribution rate of principal components. Black, blue and red lines correspond to the first, second and third principal component, respectively. (f) The variation of Ap index. (g) Es index which indicates the seismic activity around the station. E/r^2 of earthquakes indicates seismic energy felt at the observation site (MCK). The vertical lines indicate the day of earthquake with $M > 6$.

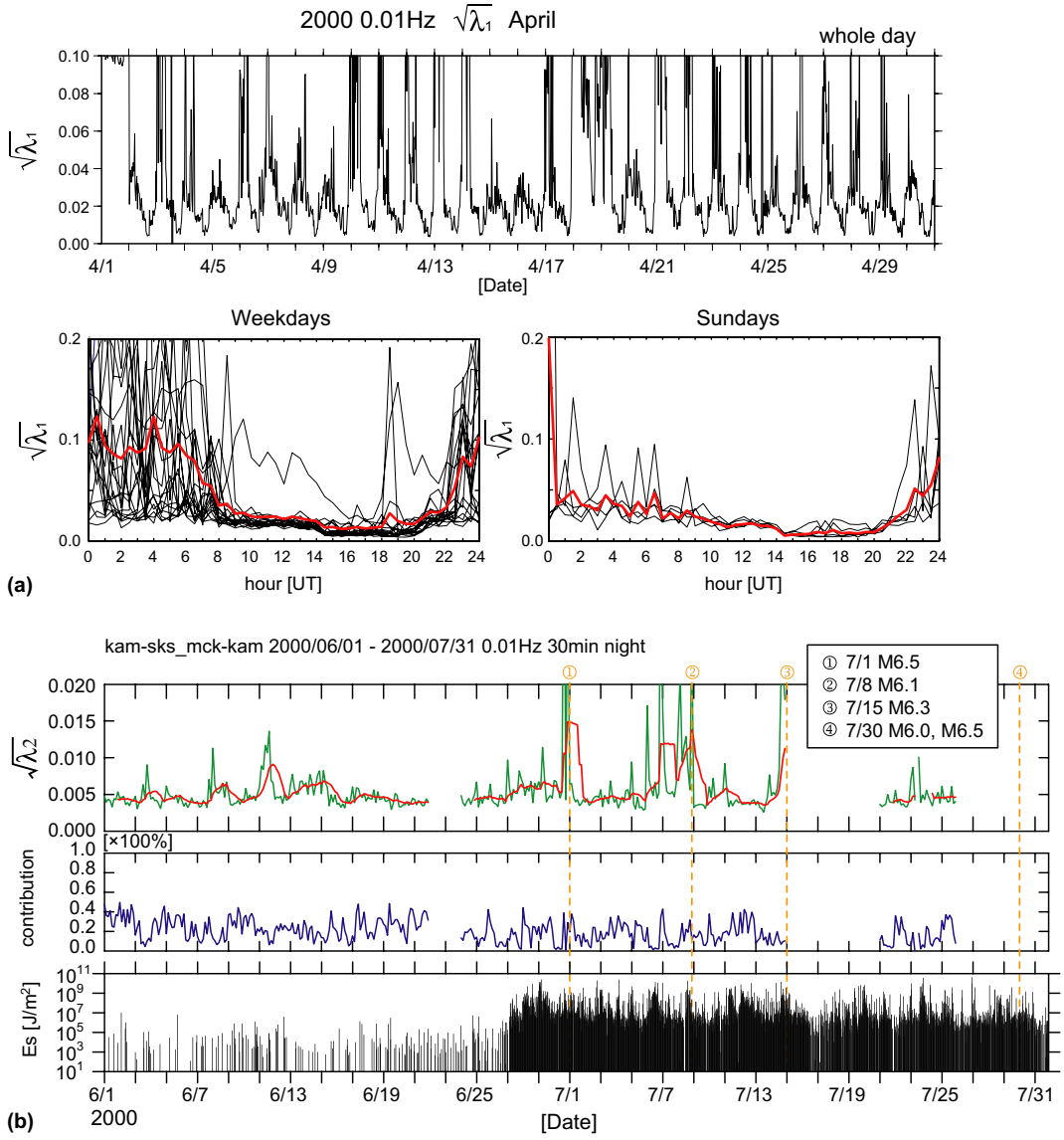

Fig. 9. (Colour online) Detailed results of PCA with differential data. (a) First principal component in all day data $\sqrt{\lambda_1}$ in April, the diurnal variation of $\sqrt{\lambda_1}$ on week days, and the diurnal variation on holidays. The red line indicates the average variation. (b) Third principal component $\sqrt{\lambda_2}$ in the interval of June–July, 2000, the variation of contribution rate of the third principal component, and the seismic activity Es around the station.

three stations with 12.5 Hz sampling rate (Ismaguilov et al., 2003). Results of differential PCA are illustrated in Figs. 7–9. The behavior of the first and second principal components in Fig. 7 resembles that of the second and third components using three station data in Fig. 5, respectively. The corresponding variations of eigenvectors are illustrated in Fig. 8. The detailed daily variation of the first principal component is plotted in Fig. 9(a) as compared with the previous results in Fig. 6(b). The tendency is very similar. The daily variation of the second principal component and contribution in the interval of June–July is plotted in Fig. 9(b). The results show that the contribution is rather significant and the variation is very close to that of the previous third component as shown in Fig. 6(c). These results suggest that the differential PCA is very effective to remove the global variation originated from solar-terrestrial interaction and to enhance the residual signals including earthquake-related one.

5. Discussions and concluding remarks

Taking account of SSA results, the observed signal at 0.01 Hz band in NS component at each element of the Izu array is generally a superposition of three waves, and at most four waves. It gives the objectivity and persuasive power for performing PCA with three station data. The results indicate that the proposed PCA method with the use of closely distributed multiple geomagnetic station data has a capability to discriminate observed signals and detect weak earthquake-related ULF geomagnetic changes or to enhance the distinctive characteristics of earthquake-related signals in a certain epicentral distance.

The results of PCA with three station data suggest that the most intense and dominant signal is that associated with solar–terrestrial interactions in comparison with variation of Ap index. The influence from solar–terrestrial interaction is universal so that the variation at three stations should occur simultaneously. When comparing the results in the three data PCA with those in differential PCA, it is strongly suggested that the first principal component is removed quite well. It is safe to say that the differential data set seems effective for closely distributed stations in order to remove this global effect. The results of signal discrimination of the residuals with PCA scheme look very similar to those using the original three station data set with larger contributions, and the contribution of the second principal component is 20–40%. It is large enough to prove mathematical significance of the signal. Further application is required to accumulate the events.

The variations of the first and second principal components of the differential PCA are found to be very similar to each other. But the variation of the first principal component sometimes exhibits some peaks without any large earthquakes, so that we conclude that the second principal component of the differential PCA is a better index for monitoring regional seismic activity.

The results of Izu array, demonstrate the possibility of monitoring the crustal activity by means of closely distributed geomagnetic stations. The major findings are as follows: (1) It is important to apply simultaneously singular spectral analysis (SSA) and principal component analysis (PCA). SSA gives the structure of signals and the number of sensors for PCA is estimated. This makes the results more convincing to the readers. (2) There is a significant advantage when using PCA with differential data sets of filtered (0.01 Hz band) SKS–KAM and MCK–KAM signal in NS component to remove the most intense signal like global variation (solar–terrestrial interaction). This procedure enabled us to detect more clearly the anomalous changes in the second principal component.

Acknowledgements

The authors thank deeply Prof. M. Taguri of Chiba University for useful discussion. They also wish to express their thanks to Japan Meteorological Agency for providing earthquake catalogue. This work was partially supported by the JSPS Grants-in-Aid for Scientific Research #13650477.

References

Bernardi, A., Fraser-Smith, A.C., McGill, P.R., Villard, O.G., 1991. ULF magnetic field measurements near the epicenter of the Ms 7.1 Loma Prieta earthquake. Phys. Earth Planet. Inter. 68, 45–63.

Fraser-Smith, A.C., Bernardi, A., McGill, P.R., Ladd, M.E., Helliwell, R.A., Villard, O.G., 1990. Low-frequency magnetic field measurements near the epicenter of the Ms 7.1 Loma Prieta earthquakes. Geophys. Res. Lett. 17, 1465–1468.

Golyandina, N., Nekrutkin, V., Zhigljavsky, A., 2001. Analysis of Time Series Structure. Chapman and HALL/CRC, p. 305.

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.

Hattori, K., 2004. ULF geomagnetic changes associated with large earthquakes. Terr. Atmos. Ocean. Sci. 15, 329–360.

Hattori, K., Serita, A., Gotoh, K., Yoshino, C., Harada, M., Isezaki, N., Hayakawa, M., 2004a. ULF geomagnetic anomaly associated with 2000 Izu islands earthquake swarm, Japan. Phys. Chem. Earth 29, 425–436.

Hattori, K., Takahashi, I., Yoshino, C., Isezaki, N., Iwasaki, H., Harada, M., Kawabata, K., Kopytenko, E., Kopytenko, Y., Maltsev, P., Korepanov, V., Molchanov, O., Hayakawa, M., Noda, Y., Nagao, T., Uyeda, S., 2004b. ULF geomagnetic field measurements in Japan and some recent results associated with Iwateken Nairiku Hokubu earthquake in 1998. Phys. Chem. Earth 29, 481–494.

Hayakawa, M., Fujinawa, Y. (Eds.), 1994. Electromagnetic Phenomena Related to Earthquake Prediction. Terra Scientific Publishing Company (TERRAPUB), Tokyo, p. 677.

Hayakawa, M. (Ed.), 1999. Atomospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes. TERRAPUB, Tokyo, p. 996.

Hayakawa, M., Molchanov, O. (Eds.), 2002. Seismo Electromagnetics, Lithosphere-Atomosphere-Ionosphere Coupling. TERRAPUB, p. 477.

Hayakawa, M., Hattori, K., 2004. Ultra-low-frequency electromagnetic emissions associated with earthquakes: a review. Inst. Elect. Engrs. Japan, Trans. Fund. Mater. 124, 1101–1108.

Ismaguilov, V., Kopytenko, Y., Hattori, K., Hayakawa, M., 2003. Variations of phase velocity and gradient values of ULF geomagnetic disturbances connected with the Izu strong earthquake. Nat. Hazards Earth System Sci. 3, 211–215.

Troyan, V., Hayakawa, M., 2002. Inverse Geophysical Problems. Terrapub, Tokyo, p. 289.