

Principal component analysis and singular spectrum analysis of ULF geomagnetic data associated with earthquakes

A. Serita^{1,*}, K. Hattori², C. Yoshino², M. Hayakawa³, and N. Isezaki⁴

¹Graduate School of Science and Technology, Chiba University, Inage, Chiba 263-8522, Japan

²Marine Biosystems Research Center, Chiba University, Inage, Chiba 263-8522, Japan

³The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan

⁴Department of Earth Science, Faculty of Science, Chiba University, Inage, Chiba 263-8522, Japan

* now at: Japan Defense Agency, Japan

Received: 7 July 2005 – Revised: 15 August 2005 – Accepted: 15 August 2005 – Published: 19 September 2005

Part of Special Issue “Seismic hazard evaluation, precursory phenomena and seismo electromagnetics”

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

nawa, 1994; Hayakawa, 1999; Hayakawa and Molcahnov, 2002; Hattori, 2004). One of the important problems is the investigation of the relationship between ULF (Ultra Low Frequency) geomagnetic anomalous changes and earthquake activity (Fraser-Smith, 1990; Bernardi, 1991; Hattori, 2004). In comparison with other higher frequency ranges, ULF has the advantage to propagate in the crust due to the skin depth. Therefore, we investigate the anomalous ULF geomagnetic changes associated with 2000 Izu Islands Earthquake Swarm by using principal component analysis.

Miyake-jima Island is a basaltic volcano, about 150 km south of Tokyo, Japan. On 26 June 2000, a seismic swarm beneath the summit suddenly occurred. Then, a migration of magma and a small-scale sea floor eruption started. Earthquake epicenters migrated from the island westward and then northward. During this period 4 large (>M6.0) earthquakes, 7/1:M6.4, 7/8:M6.1, 7/15:M6.3, 8/18:M6.0 occurred within 100 km from Mochikoshi station, Izu as shown in Fig. 1. They occurred in shallow depth 15 km, 13 km, 5 km, and 11 km, respectively. On the other hand, a phreatic explosion and the formation of the initial sink at the summit of the volcano occurred on 9 July. 18 August 2000 largest eruption followed by low temperature ash flow on 29 August occurred.

We measured three components of the ULF geomagnetic field and two horizontal geoelectrical potential differences at Izu Peninsula, Japan (Hattori et al., 2004b). The measurement equipment is constituted by an array with intersensor distance of 5 km as shown Fig. 1. We recorded the signals at stations, Mochikoshi (MCK), Seikoshi (SKS), and Kamo (KAM). Observed ULF geomagnetic fields are considered a superposition of solar origin signal, artificial noise, and other noises that propagate in the crust. The signal associated with crustal activity is very weak and sophisticated signal classification methods are required. In this paper, we applied principle component analysis (PCA) and singular spectral

1 Introduction

Many studies on electromagnetic phenomena associated with earthquakes have been reported (e.g. Hayakawa and Fuji-

Correspondence to: K. Hattori
(hattori@earth.s.chiba-u.ac.jp)

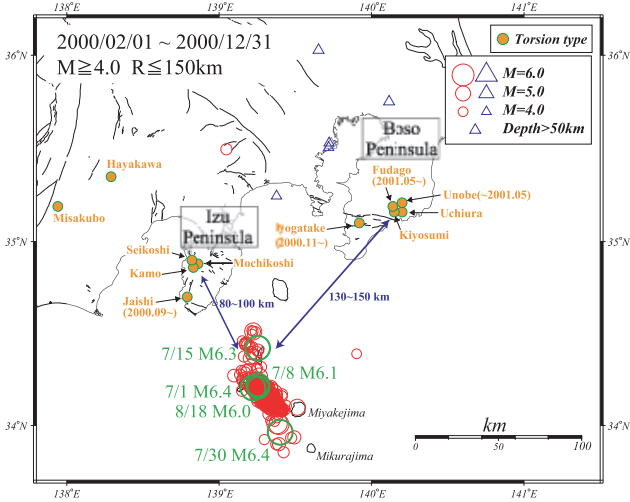


Fig. 1. The map of ULF geomagnetic station (array station) and seismic activity of 2000 Izu Islands Earthquake.

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).

2 Principal component analysis (PCA) of ULF geomagnetic data

We apply PCA to the time series data observed at closely separated stations. The analyzed period is from February 2000 to December 2000. In this paper, we show the results of 0.01 Hz band with 12.5 Hz sampling rate. Let us indicate the time series of data (30 min) observed at each station by $X(n)=[x_n(t_1), x_n(t_2), \dots, x_n(t_N)]$, where n is an index denoting the site and N is the number of the data (12.5×30×60=225 000). The data matrix $\mathbf{X}=[x_1x_2x_3]^T$ is obtained, where T means transpose, and then the covariance matrix $\mathbf{R}=\frac{1}{N}(\mathbf{X}\mathbf{X})^T$ is calculated. The eigenvalue decomposition of $\mathbf{R}=\mathbf{V}\lambda\mathbf{V}^T$ has been performed. The variations of eigenvalues and eigenvectors are investigated.

In addition, we perform the PCA with differential data at Izu Peninsula applying the methods to SKS-KAM and MCK-KAM data as $\mathbf{X}(1)$ and $\mathbf{X}(2)$ in order to remove the coherent influence in the region of the array.

3 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, 2004) has been applied to the time series of data at 0.01 Hz. SSA is sometimes called caterpillar method. The mathematical basis is very similar to PCA. The procedure is as follows. Let's consider one dimensional time series data

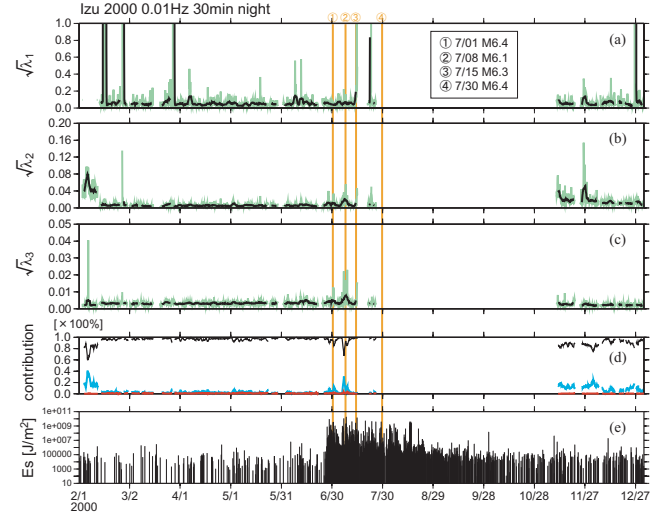


Fig. 2. The results of PCA using the three-station data. (a) The variation of eigenvalue of the first principal component. (b) The variation of eigenvalue of the second principal component. (c) The variation of eigenvalue of the third principal component. (d) Contribution rate of principal components. Black, blue and red lines correspond to the first, second and the third principal component, respectively. (e) E_s index (E/r^2) which indicates the seismic energy felt at the observation site (MCK), where E and r are the energy of earthquake and hypocentral distance during 30 min.

$\{x_i\}_{i=1}^N$. The multidimensional matrix \mathbf{X} is obtained as using the evolvement of the 1-D time series,

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

Here M is caterpillar length with $M < N/2$ and $k = N - M + 1$. Then the matrix \mathbf{X} is normalized. The element (i, j) component of normalized matrix \mathbf{X}^* is 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, the covariance matrix of $\mathbf{R} = \frac{1}{N}(\mathbf{X} * \mathbf{X}^*)^T$ and the singular value decomposition of $\mathbf{R} = \mathbf{V}\lambda\mathbf{V}^T$ are computed. 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 by the station. In this analysis, $M=1800$ and $N=3750$ (12.5 Hz×5 min×60 s).

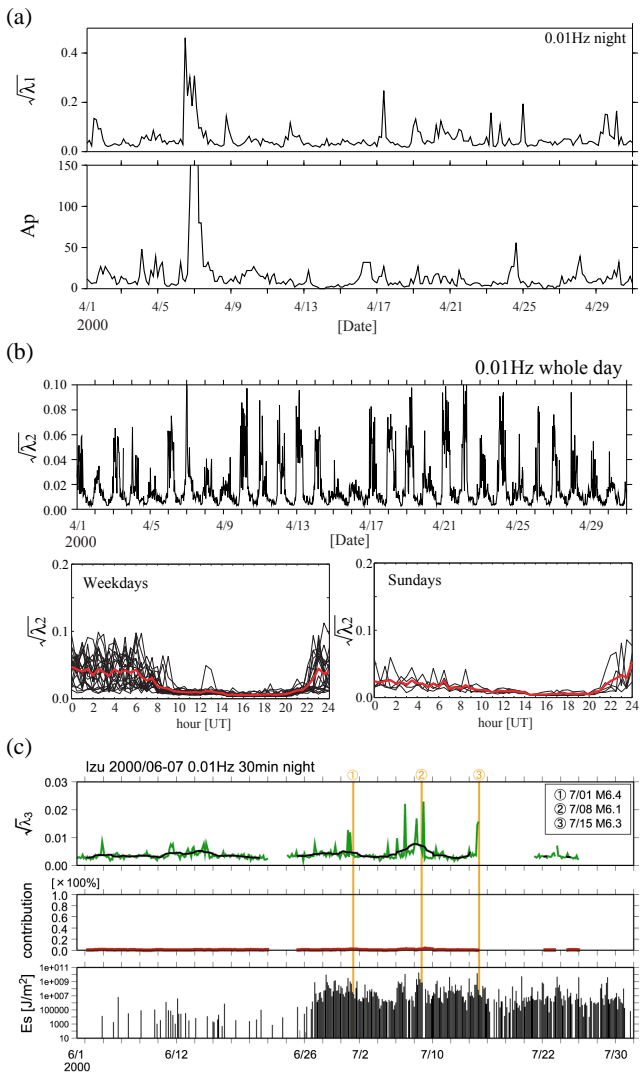


Fig. 3. The detailed results of PCA the three-station data. (a) The variation of the first principal component $\sqrt{\lambda_1}$ in the night time with that of A_p index in April 2000. (b) The variation of second principal component in all day data $\sqrt{\lambda_2}$ in April 2000, the diurnal variation of $\sqrt{\lambda_2}$ in week days, and the diurnal variation of $\sqrt{\lambda_2}$ in holidays. (c) The variation of the third principal component $\sqrt{\lambda_3}$ in the interval of June–July 2000, the variation of contribution rate of the third principal component, and E_s index (E/r^2) which indicates the seismic energy felt at the observation site (MCK), where E and r are the energy of earthquake and hypocentral distance during 30 min.

4 Results of PCA and SSA

4.1 PCA results of three station data

In this study the results of differential PCA and SSA will be shown, the results of normal PCA have been already reported (Hattori et al., 2004a). However, we will mention them briefly. Figures 2 and 3 indicate that (1) the variation of the first eigenvalue is similar to that of A_p index. The corresponding eigenvector is found to be stable. This suggests that the dominant origin of the first principle component may be

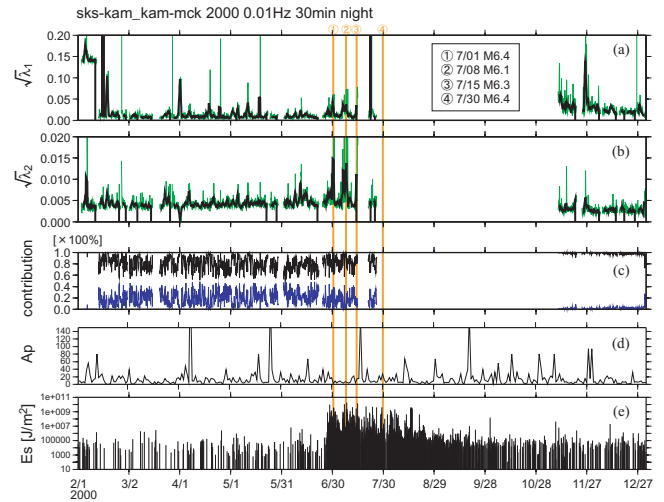


Fig. 4. The results of PCA with differential data. (a) The variation of eigenvalue of the first principal component. (b) The variation of eigenvalue of the second principal component. (c) The variation of contribution rates. Black and blue indicate first and second principal component, respectively. (d) The variation of A_p index. (e) E_s index (E/r^2) which indicates the seismic energy felt at the observation site (MCK), where E and r are the energy of earthquake and hypocentral distance during 30 min.

solar-terrestrial interaction. (2) The variation of the second eigenvalue seems to be related to that of electricity consumption power around stations. (3) The variation of the third eigenvalue shows the peak values before a few days before the earthquakes with M greater than 6, which are marked by vertical lines. The variation of the third principal component in June–July 2000 shows that there is also local maximum about two weeks before the swarm activity. However, someone may comment on (1) a small contribution of the third principal component, and (2) uncertainty of the number of signals observed at each station.

4.2 PCA results of differential station data

In order to remove the most intense noise source denoted by the first principal component in the three-station data analysis, we referred the SKS and MCK signals to KAM station, that is, we analyzed data sets of SKS-KAM and MCK-KAM in NS component. On the other hand, the signal associated with earthquakes has a small velocity and there is an arrival time difference among three station with 12.5 Hz sampling rate (Ismaguilov et al., 2003). Results of differential PCA are shown in Figs. 4 and 5. The behavior of the first and second principal component in Fig. 4 resembles to that of the second and third component using three station data in Fig. 2, respectively. The daily variation of the first principal component is plotted in Fig. 5a and compared the previous results in Fig. 3b. 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. 5b.

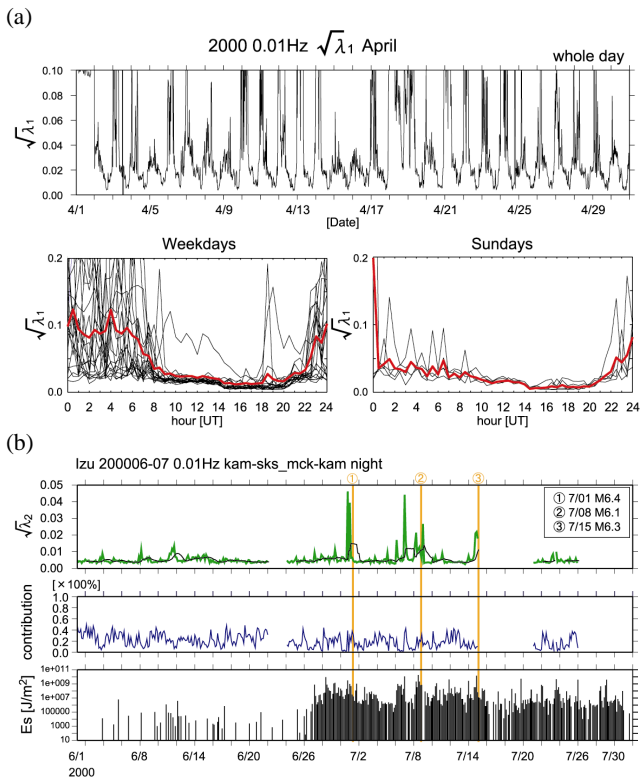


Fig. 5. The results of PCA with differential data. **(a)** The variation of first principal component in all day data $\sqrt{\lambda_1}$ in April, the diurnal variation of $\sqrt{\lambda_1}$ in week days, and the diurnal variation of in holidays. **(b)** The variation of the third principal component $\sqrt{\lambda_2}$ in the interval of June–July 2000, the variation of contribution rate of the third principal component, and the seismic activity around the station. E_s index (E/r^2) which indicates the seismic energy felt at the observation site (MCK), where E and r are the energy of earthquake and hypocentral distance during 30 min.

The results show the contribution is rather significant and the variation is very close to that of the previous third component as shown in Fig. 3c. These results suggest the differential PCA is very effective to remove the global variation due to the solar-terrestrial interaction and enhances the remaining signals including earthquake-related one.

4.3 SSA results of each station data

Figure 6 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 Kamo (KAM) station. Figures 6a and 6b 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 upper middle shows the accumulated contribution. Black line indicates the contribution of the first principal component. Red, blue and yellow line indicate accumulated contribution to the second, the third and the fourth principal component. The lower middle is seismic activity around the Izu array station and the bottom is the variation of geomagnetic activity A_p . The signals at 0.01 Hz

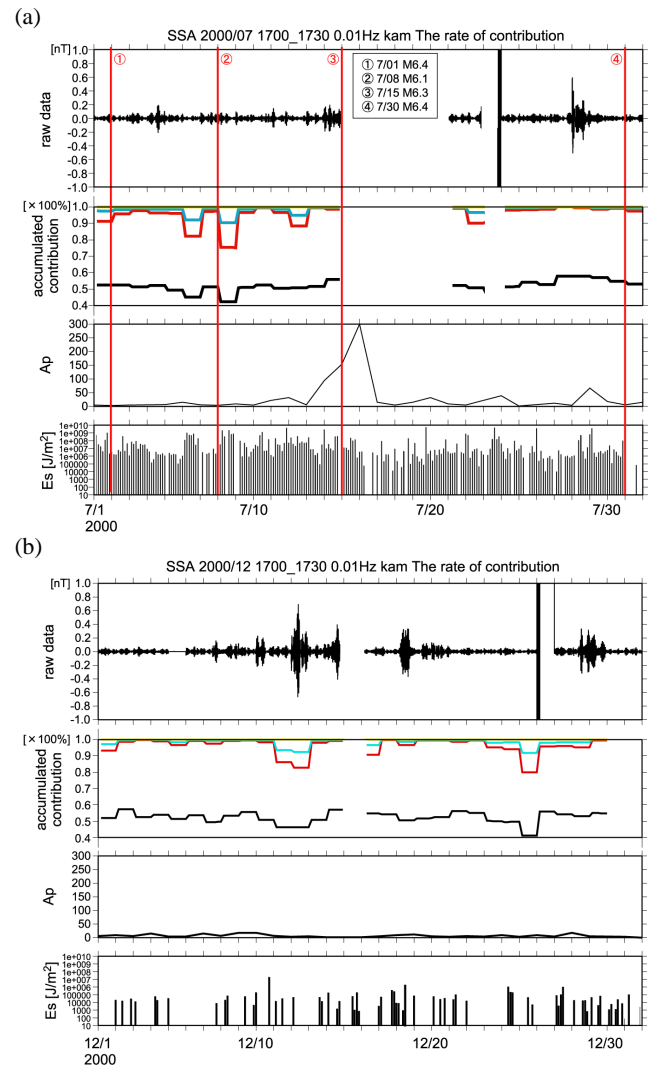


Fig. 6. The results of SSA at Kamo station in July 2000 **(a)** and in December 2000 **(b)**. The variation of A_p index and E_s index (E/r^2) which indicates the seismic energy felt at the observation site (MCK) are also shown, where E and r are the energy of earthquake and hypocentral distance during 5 min. The hatched areas indicate the day of earthquake with $M > 6.0$.

band observed at Kamo consist of at most four components. Usually, they consist of at most three components.

Similar results have been obtained at the other two stations, Seikoshi (SKS) and Mochikoshi (MCK). There seems no apparent relation with both the seismic and the geomagnetic activity. Therefore, the structure of the observed signal at three stations is very similar and generally less than three signal sources should be considered. It means that the PCA by using three station data observed at Izu peninsula in the frequency range of 0.01 Hz is valid even four-station data analysis is desirable.

5 Concluding remarks

The results show that the proposed principal component analysis method with the use of closely distributed multiple geomagnetic station data is able to detect weak earthquake-related ULF geomagnetic changes or to enhance the peculiar characteristics of earthquake-related signals in a certain hypocentral distance. The results of Izu array, demonstrate the possibility of monitoring the crustal activity. The major findings are as follows. (1) It is important to apply principal component analysis and singular spectral analysis simultaneously. SSA gives the structure of signals and the number of sensors for PCA is estimated. This makes the results convincing. (2) There is a significant advantage using PCA with differential data sets of filtered (0.01 Hz band) SKS-KAM and MCK-KAM in NS component for removing the most intense signal like global variation (solar-terrestrial interaction). This provides that the anomalous changes in the second principal component are appeared more sharply. And the contribution of the second principal component is 20–40%. It is large enough to prove mathematical accuracy of the signal. Further application is required to accumulate events.

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

Edited by: P. F. Biagi

Reviewed by: L. Telesca and another referee

References

- Bernardi, A., Fraser-Smith, A. C., McGill, P. R., and Villard, O. G.: ULF magnetic field measurements near the epicenter of the Ms 7.1 Loma Prieta earthquake, *Phys. Earth Planet. Inter.*, 68, 45–63, 1991.
- Fraser-Smith, A. C., Bernardi, A., McGill, P. R., Ladd, M. E., Helliwell, R. A., and Villard, O. G.: Low-frequency magnetic field measurements near the epicenter of the Ms 7.1 Loma Prieta earthquake, *Geophys. Res. Lett.*, 17, 1465–1468, 1990.
- Golyandina, N., Nekrutkin, V., and Solntsev, V.: Caterpillar-SSA Technique for Analysis of Time Series in Economics, <http://vega.math.spbu.ru/caterpillar/>, 2004.
- Golyandina, N., Nekrutkin, V. and Zhigljavsky, A.: *Analysis of Time Series Structure*, Chapman and HALL/CRC, pp. 305, 2001.
- Hattori, K.: ULF geomagnetic changes associated with large earthquakes, *Terrestrial, Atmos. Ocean. Sci.*, 15, 329–360, 2004.
- Hattori, K., Serita, A., Gotoh, K., Yoshino, C., Harada, M., Isezaki, N., Hayakawa, M.: ULF geomagnetic anomaly associated with 2000 Izu islands earthquake swarm, Japan, *Phys. Chem. Earth.*, 29, 425–436, 2004a.
- 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., and Uyeda, S.: ULF geomagnetic field measurements in Japan and some recent results associated with Iwateken Nairiku Hokubu earthquake in 1998, *Phys. Chem. Earth*, 29, 481–494, 2004b.
- Hayakawa, M. and Fujinawa, Y. (Eds.): *Electromagnetic Phenomena Related to Earthquake Prediction*, Terra Scientific Publishing Company (TERRAPUB), Tokyo, pp. 677, 1994.
- Hayakawa, M. (Ed.): *Atomospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes*, TERRAPUB, Tokyo, pp. 996, 1999.
- Hayakawa, M. and Molchanov, O.: *Seismo Electromagnetics Lithosphere-Atomosphere-Ionosphere Coupling*, TERRAPUB, Tokyo, pp. 477, 2002.
- Ismaguilov, V., Kopytenko, Y., Hattori, K., and Hayakawa, M.: Variations of phase velocity and gradient values of ULF geomagnetic disturbances connected with the Izu strong earthquake, *Nat. Hazards Earth Syst. Sci.*, 3, 211–215, 2003, **SRef-ID: 1684-9981/nhess/2003-3-211**.