

AN EFFECTIVE EVALUATION FUNCTION FOR ICA TO SEPARATE TRAIN NOISE FROM TELLURIC CURRENT DATA

Mika Koganeyama Sayuri Sawa* Hayaru Shouno† Toshiyasu Nagao◇ Kazuki Joe**

* Nara Women's University, Nara City, Japan

† Yamaguchi University, Ube City, Japan

◇ Tokai University, Earthquake Prediction Research Center, Shimizu City, Japan

ABSTRACT

Irregular changes of electric currents called Seismic Electric Signals (SESSs) are often observed in Telluric Current Data (TCD). Recently, detection of SESSs in TCD has attracted notice for short-term earthquake prediction. Since most of the TCD collected in Japan is affected by train noise, detecting SESSs in TCD itself is an extremely arduous job. The goal of our research is automatic separation of train noise and SESSs, which are considered to be independent signals, using Independent Component Analysis (ICA). In this paper, we propose an effective ICA evaluation function for train noise considering statistic analysis. We apply the evaluation function to TCD and analyze the results.

1. INTRODUCTION

Especially since the great Hanshin earthquake in 1994, short-term earthquake prediction has been investigated as an emergent and important research topic in Japan. Long-term (several to dozens of years) prediction based on past earthquakes is a standard method for earthquake prediction in conventional seismology. However, it is obviously difficult to apply the same statistical method to short-term earthquake prediction which is to predict whether earthquakes occur within several weeks or months [1]. Therefore, we have to apply different methods from conventional ones to short-term earthquake prediction. Earthquake Prediction Research Center of Tokai university has studied short-term earthquake prediction using various electromagnetic methods [2]. We have noticed Telluric Current Data (TCD) observation method in these methods. TCD is the measurement of the weak electric current flowing within the surface layers of the Earth. Irregular changes of electric currents observed in the TCD are often detected before the occurrence of strong earthquakes. We call such irregular changes of electric currents Seismic Electric Signals (SESSs). Detecting SESSs in TCD can make short-term earthquake prediction possible. Actually, Several earthquakes were successfully predicted using TCD in Greece [3].

However, the effect of train noise in TCD is the most serious problem for short-term earthquake prediction in Japan. SESSs are hidden by train noise because the amplitude of train noise is larger than the amplitude of the SESSs. So it is difficult for even experts on TCD observation method to detect SESSs hidden by train noise for short-term earthquake prediction.

Considering this background, we began research on automatic short-term earthquake prediction applying engineering methods to TCD instead of manual detection of SESSs by the experts. In this research, we apply Independent Component Analysis (ICA) which

separates each independent source signal from a mixture of independent source signals. We believe that train noise and SESSs are independent source signals because the current generating mechanism is different. We assume TCD is composed of train noise and SESSs. So we believe that train noise and SESSs can be separated by applying ICA to TCD.

In [4], we confirmed that ICA could separate train noise and an SES from an artificial mixture generated by train noise and an SES of Matsushiro, Nagano by our experiments. Although we should have taken a proper ICA evaluation function considering the data feature, we have done it for TCD by our hunch in the experiments. We applied ICA taken by our hunch to the data containing train noise with an SES and containing only train noise. As the result, we could confirm that the SES can be separated, but the wave patterns of train noise separated from each of the data are different. We assume that the reason why the ICA could not separate same train noise from each of the data is that we did not take a proper ICA evaluation function for TCD.

In this paper, we apply a more proper ICA taken by considering statistical analysis of train noise to TCD and evaluate the results.

2. TELLURIC CURRENT DATA (TCD)

2.1. The observation method

TCD measures the electric potential difference between dipoles at two points. The electrodes are $Pb - PbCl_2$ pipe non-polarizing dipoles (40cm in length and 3cm in outer diameter) and are buried at a depth of two meters.

42 observation points have been installed mainly in the Tokai and Hokuriku area since 1997. Each observation point has either 8 or 16 dipoles in different directions. We label each dipoles dp.1, dp.2, ..., dp.16.

TCD is sampled at 10 second intervals and telemetered once a day to Earthquake Prediction Research Center. So TCD is expressed by timevarying voltage data for each dipole. For example, Fig.1 shows an example of the TCD from dp.2 observed at Matsushiro, Nagano on 20th of August, 1999. The vertical axis of the graph represents potential (mV/m) and the horizontal axis represents time (hour).

2.2. Train Noise

Train noise is generated regularly and the shape of the noise is always similar. So we can find train noise from TCD using the timetable of Matsushiro station, which is near the TCD observation point.

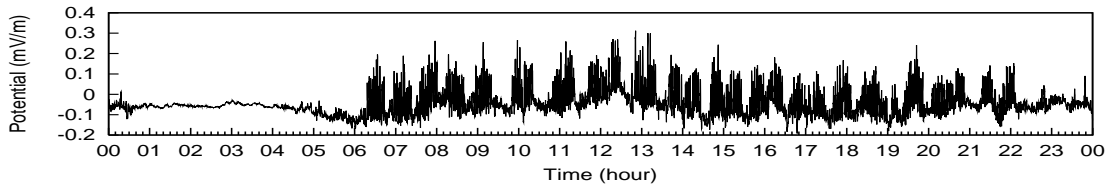


Fig. 1. Telluric current data (20th of August, 1999, dp.2 of Matsushiro, Nagano).

We can specify the train noise in TCD of Matsushiro shown in Fig.1 and the timetable. Fig.2 is enlarged in Fig.1 between 6:00 and 7:00. We can find a distinctive wave pattern in Fig.2. The wave pattern represents the train noise of the first 6:31 train. The length of a typical train noise is about 50 minutes.

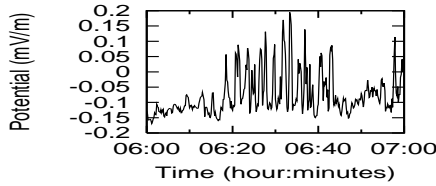


Fig. 2. A typical train noise.

2.3. Seismic Electric Signals (SESs)

It is known from laboratory experiments that electric current is generated before rocks fracture under load [5] [6]. Earthquakes are also a kind of rock fracture phenomenon, so it is known that electric currents flow within the Earth before great earthquakes. We call such irregular changes of electric currents Seismic Electric Signals (SESs). Fig.3 shows an SES as observed at dp.2 between 1:20 and 2:10. In this datum, experts on the TCD observation method could find the SES because it was observed at midnight, when no trains were running.

It is known empirically that the features of SESs are 1) the wave pattern has a positive amplitude, 2) the function consists of a rapid increase followed by a gradual decrease, 3) the duration of an SES is from about 10 sec to a few minutes or rarely up to a few hours.

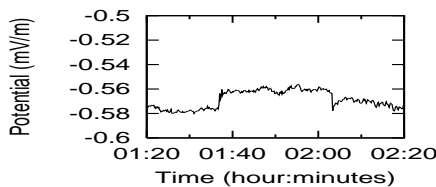


Fig. 3. A seismic electric signal.

2.4. Problems of the TCD observation method in Japan

In Japan, the most serious problem for short-term earthquake prediction using TCD is the presence of train noise in TCD. If an SES

is contained in TCD, the SES is often hidden by train noise. For example, Fig.4 shows train noise observed at dp.2 between 6:6:40 to 7:5:00 on 17th of January 1999 added to the SES of Fig.3 artificially. The vertical axis of the frame shows potential (mV/m) and horizontal axis shows frame length ($\times 10$ sec). The frame length represents the number of points whose sampling rate is 10 sec.

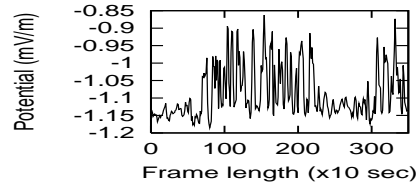


Fig. 4. A train noise with an SES.

The reason why we generate the data artificially is that it is quite difficult to classify the data which contains both train noise and SESs in real TCD. We cannot distinguish the SES also in Fig.4 manually.

We have attempted to apply ICA to TCD to separate train noise and SESs automatically because we assume that they are independent source signals. In this paper, we do not take ICA evaluation function to apply TCD by our hunch, but consider statistical analysis of train noise. Then we validate whether the ICA can separate train noise correctly.

3. DETERMINATION OF ICA EVALUATION FUNCTION

3.1. Statistical analysis of train noise

We analyze data statistically to find an efficient ICA evaluation function. In this paper, we draw histograms and normal distribution graphs of train noise because the purpose of this research is to separate train noise from TCD correctly. We use the TCD observed on January, April, August and October 1999 at Matsushiro, Nagano for statistic analysis of train noise. We collect series of 200 samples so that the middle point is the train departure time at Matsushiro station using the timetable and obtain ensemble mean of them. Fig.5 shows a histogram of the distribution of train noise at dp.2 on August. The dashed line shows the normal distribution by the mean and the variance of the train noise. The data maximum, minimum, mean and variance are $MAX = 6.25 \times 10^{-1}$ (mV/m), $MIN = -2.73 \times 10^{-1}$ (mV/m), $\mu = -1.56 \times 10^{-2}$ (mV/m) and $\sigma^2 = 6.28 \times 10^{-3}$ (mV/m), respectively. The number of classes is 50, and class interval is 1.79×10^{-2} (mV/m).

Fig.5 shows left bias characteristics of the train noise. The shapes of histograms and normal distribution graphs of the other

train noise data are almost similar to the data at dp.2 on August. Therefore, it can be assumed that train noise at Matsushiro is asymmetric data.

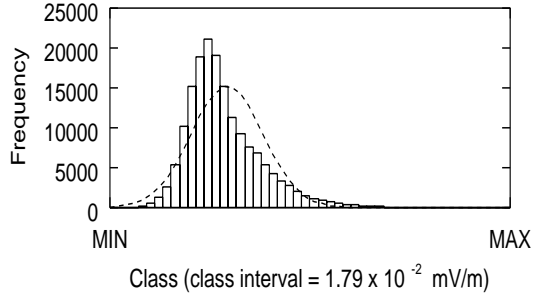


Fig. 5. Histogram and normal distribution of train noise (August, 1999, dp.2 of Matsushiro).

3.2. Effective evaluation function to separate train noise

In section 3.1, we confirmed that train noise at Matsushiro is asymmetric data. Actually, each wave pattern of train noise separated from the data without and with an SES is different using JADE [7], which is a proper ICA algorithm for symmetric data as speech signal. Therefore we can assume that a more efficient evaluation function for asymmetric data, which is expressed by odd-order cumulant, can separate train noise from TCD correctly.

In this paper, we apply ICA with third-order cumulants as follows [8].

Vectorial input signals $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$ to be observed are often a mixture of independent source signals $s_i(t)$. In this case, it is assumed that $s_i(t)$ are train noise or SESs coming from different sources. The mixing is linear, and it yields the relation

$$\mathbf{x}(t) = A\mathbf{s}(t) \quad (1)$$

with mixing matrix A and source signal $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$. $\mathbf{x}(t)$ is assumed to be sets of zero mean signals.

The input components are usually dependent, due to the mixing process, while the sources are not. If one succeeds in finding a matrix R that yields independent output components $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_n(t)]^T$, given by

$$\mathbf{u}(t) = R\mathbf{x}(t), \quad (2)$$

one can recover the original sources $s_i(t)$ up to a permutation and constant scaling of the sources. R is called the unmixing matrix and finding the matrix is referred to as independent component analysis (ICA).

We adopt third-order cumulants as our ICA evaluation function. Third-order cumulants are defined by $C_{ijk}(\mathbf{u}) := \langle u_i u_j u_k \rangle$ with $\langle \cdot \rangle$ indicating the mean over all data points. The off-diagonal elements of $C_{ijk}(\mathbf{u})$ characterize the statistical dependencies between components. Thus, obtaining R that diagonalizes the $C_{ijk}(\mathbf{u})$ is equivalent to making the output data $\mathbf{u}(t)$ independent. The second order cumulant can be diagonalized easily by whitening the input data $\mathbf{x}(t)$ with an appropriate matrix, and it yields the relation

$$\mathbf{y}(t) = W\mathbf{x}(t). \quad (3)$$

Thus, Eq(2) can be transformed as the following relation.

$$\mathbf{u}(t) = Q\mathbf{y}(t) = QW\mathbf{x}(t) \quad (4)$$

Since the goal is to diagonalize the cumulant tensors, we need to find a matrix Q that minimizes the square sum over off-diagonal elements. Because the square sum over all elements of a cumulant tensor is preserved under any orthogonal transformation, we can obtain some orthogonal transformation matrices Q . We adopt a Givens rotation matrix as Q and obtain the Q that maximizes the square sum over the diagonal elements of $C_{ijk}(\mathbf{u})$ indicating as the following relation.

$$\Psi_3(\mathbf{u}) = \frac{1}{3!} \sum_{\alpha} \left(C_{\alpha\alpha\alpha}(\mathbf{u}) \right)^2 \quad (5)$$

We represent ICA with third-order cumulants as 3cumICA in the following sections.

4. APPLICATION OF OUR ICA TO TCD

4.1. Data with train noise

At the Matsushiro observation point, typical train noise can be observed clearly. So we apply 3cumICA to TCD observed at Matsushiro to examine whether 3cumICA can separate the typical train noise.

The length of the input data and the number of input dimensions are set to 100 minutes and three, respectively. In this section, We apply 3cumICA to TCD observed at 20:00 to 21:40 on 20th August 1999 of dp.2, dp.6, dp.7, where the data is described as $\mathbf{x}(t) = [x_2(t), x_6(t), x_7(t)]^T$. Fig.6 shows the input data $x_2(t)$.

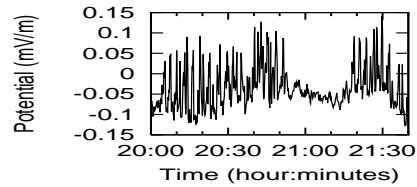


Fig. 6. Input data of 3cumICA $x_2(t)$.

Applying 3cumICA to $\mathbf{x}(t)$, we obtained the output components $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t)]^T$ shown in Fig.7. Since the vertical axis of $\mathbf{u}(t)$ cannot be determined, it is hard to evaluate the output components without rescaling. So each $u_i(t) (i = 1, 2, 3)$ is transformed back to the original signal space, and then we evaluate how each component affects $\mathbf{x}(t)$. To transform $u_1(t)$ to the original signal space, the value of $u_1(t)$ is kept and the value of the other signals is set to 0. The data is described as $\mathbf{u}_1^*(t) = [u_1(t), 0, 0]^T$. $\mathbf{x}_1^*(t) = [x_{21}^*(t), x_{61}^*(t), x_{71}^*(t)]^T$ contains only the component $u_1(t)$, given by

$$\mathbf{x}_1^*(t) = W^{-1}Q^{-1}\mathbf{u}_1^*(t)$$

with Eq(3), (4). Fig.8 shows $x_{21}^*(t)$, $x_{22}^*(t)$, and $x_{23}^*(t)$ obtained by transforming $u_2(t)$ and $u_3(t)$ back to the original space similarly.

As a result, it presumed that the component $x_{23}^*(t)$ corresponds to train noise because the amplitude of $x_{23}^*(t)$ is large and the other component appear to contain no train noise. Therefore it turned out that typical train noise can be separated from the TCD using 3cumICA.

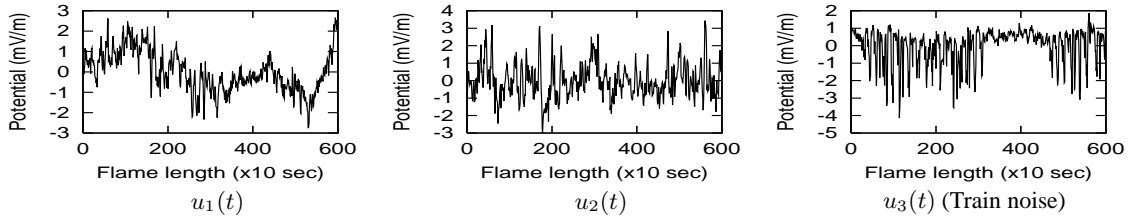


Fig. 7. Independent components $\mathbf{u}(t)$.

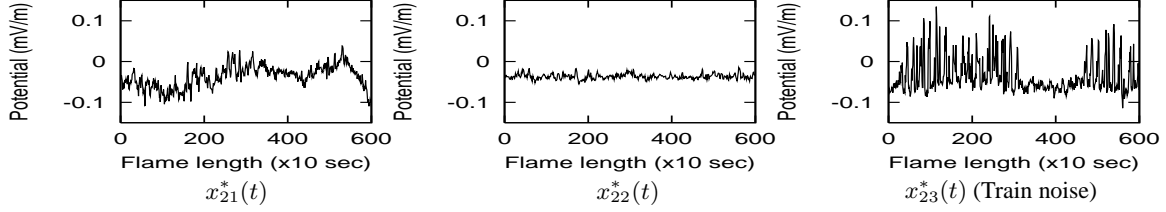


Fig. 8. Independent components in the original signal space of $x_2(t)$.

4.2. Data affected by the same train noise

We confirmed that 3cumICA can separate typical train noise in section 4.1. However, it is difficult to evaluate how correctly train noise can be separated. In this section, we apply 3cumICA to TCD observed at two different observation points, Sasadani and Ikeda in Fukui Prefecture, which contain the same train noise. We examine whether the same train noise can be separated from each observation point, namely train noise can be separate correctly using 3cumICA.

The map of Fig.9 shows the location of Sasadani and Ikeda observation points. Sasadani and Ikeda are located in about 10km of west and 18km of east from Asouzu station of Fukui railway Fukutake line, respectively.

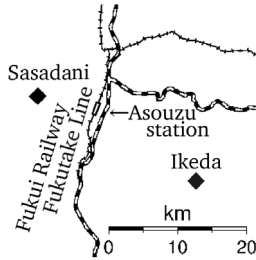


Fig. 9. The map showing the location of Sasadai and Ikeda.

In this section, we apply 3cumICA to TCD observed at 20:50 to 22:30 on 18th January 2000 of dp.1, dp.5, dp.7 at Sasadani and dp.1, dp.2-6, dp.4 at Ikeda. The data of Sasadani and Ikeda are described as $\mathbf{x}(t) = [x_1(t), x_5(t), x_7(t)]^T$, $\mathbf{x}(t) = [x_1(t), x_{2-6}(t), x_4(t)]^T$, respectively. $x_{2-6}(t)$ of Ikeda are generated by subtracting the data value of dp.6 from the data value of dp.2. Fig.10 and Fig.11 show the input data $x_1(t)$ of Sasadani and $x_4(t)$ of Ikeda, respectively.

Each component of $\mathbf{u}(t)$ of Sasadani and Ikeda is transformed back to the original signal space in the same way explained in section 4.1. Fig.12 and Fig.13 show $x_{13}^*(t)$ and $x_{42}^*(t)$ which are one of the component in the original signal space of $x_1(t)$ of Sasadani

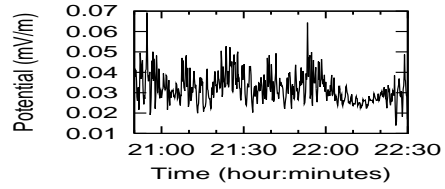


Fig. 10. Input data of 3cumICA $x_1(t)$ of Sasadani.

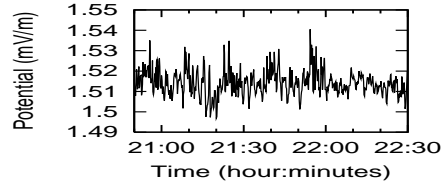


Fig. 11. Input data of 3cumICA $x_4(t)$ of Ikeda.

and $x_4(t)$ of Ikeda, respectively. It is assumed that both of the component $x_{13}^*(t)$ and $x_{42}^*(t)$ are train noise. Moreover, since the wave pattern of $x_{13}^*(t)$ and $x_{42}^*(t)$ is extremely similar, it is presumed that the components are the same train noise affecting TCD at both of Sasadani and Ikeda. Therefore, it is confirmed that train noise can be separated correctly using 3cumICA.

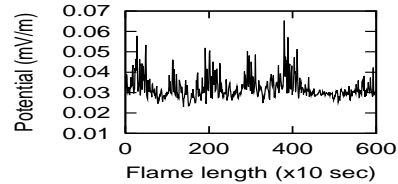


Fig. 12. x_{13}^* of Sasadani.

4.3. Data containing both train noise and an SES

We have confirmed that train noise and an SES can be separated from the data containing both same train noise and an SES using

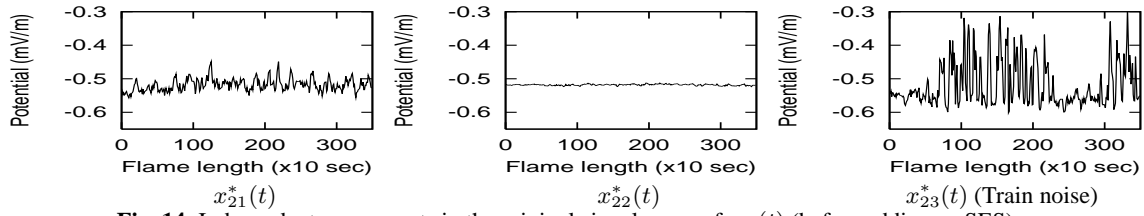


Fig. 14. Independent components in the original signal space of $x_2(t)$ (before adding an SES).

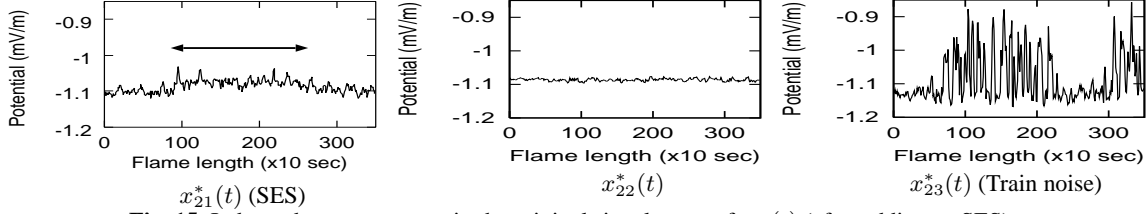


Fig. 15. Independent components in the original signal space of $x_2(t)$ (after adding an SES).

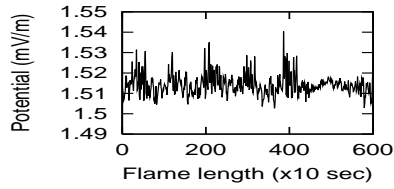


Fig. 13. x_{42}^* of Ikeda.

ICA by our experiments [4]. However, applying ICA to the data containing the same train noise without SESs, separated train noise (Fig.16) is different from one separated from the data with both train noise and an SES (Fig.17) [4]. We assume that the reason why the ICA could not separate same train noise from each of the data is that we did not take an effective ICA evaluation function for TCD.

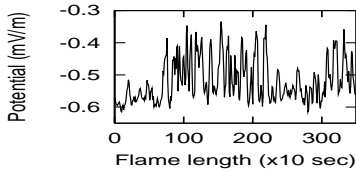


Fig. 16. Train noise separated from data without SESs.

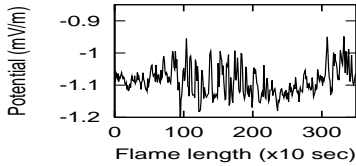


Fig. 17. Train noise separated from data with an SES.

We confirmed that 3cumICA can separate train noise correctly by our experiments in section 4.1 and 4.2. In this section, we examine whether the same train noise can be separated from the data containing train noise with and without SES using 3cumICA. In

addition, we evaluate the FFT frequency domain data of the independent components separated from the two sets of data.

4.3.1. Experiment of separating train noise and an SES

We apply 3cumICA to the data containing both train noise and an SES. However, it is difficult to find the data in real TCD as explained in section 2.4. We generated artificial data in the same manner proposed in section 2.4. We use the data without SESs which is observed at the same time as the data shown in section 2.4 (Fig.18).

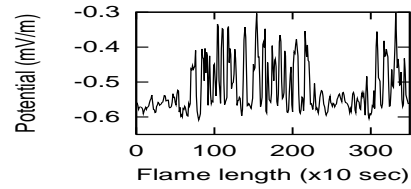


Fig. 18. Data before adding an SES of dp.2.

Both of the data before adding an SES and after adding an SES are described as $\boldsymbol{x}(t) = [x_2(t), x_6(t), x_7(t)]^T$. The length of the data is set to 3,500 sec. Each component of $\boldsymbol{u}(t)$ of the two data sets is transformed back to the original signal space in the same way explained in section 4.1. Fig.14 and Fig.15 show the independent components in the original signal space of $x_2(t)$ of the data with an SES and without SESs, respectively.

Comparing $x_{21}^*(t)$ of Fig.14 and Fig.15, it is supposed that $x_{21}^*(t)$ of Fig.15 corresponds to the SES. However, it is difficult to confirm the SES because the amplitude of the SES is extremely small. So we evaluate the FFT frequency domain data of $x_{21}^*(t)$ in the next section. It is presumed that each $x_{23}^*(t)$ of Fig.14 and Fig.15 corresponds to train noise. Moreover, comparing with $x_{23}^*(t)$ shown in Fig.16 and Fig.17, the wave pattern of each train noise is extremely similar. Therefore, it is confirmed that 3cumICA can separate train noise much more correctly than the ICA we have used in [4].

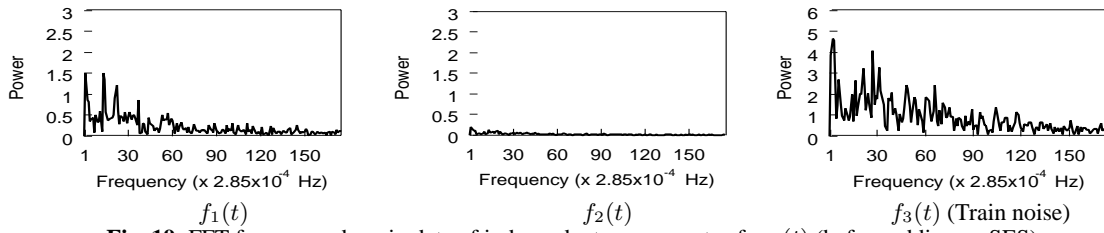


Fig. 19. FFT frequency domain data of independent components of $x_2(t)$ (before adding an SES).

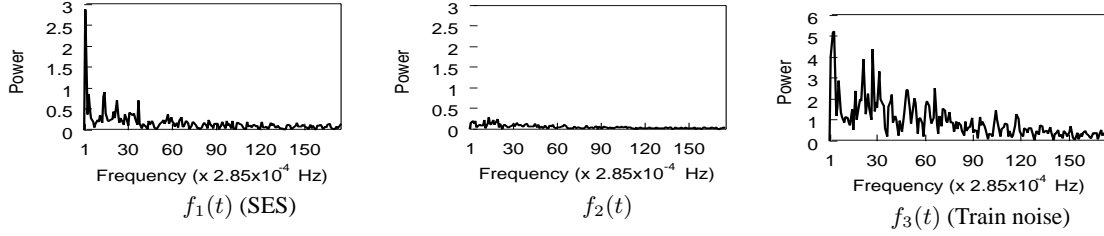


Fig. 20. FFT frequency domain data of independent components of $x_2(t)$ (after adding an SES).

4.3.2. Evaluation of FFT frequency domain data

Since the amplitude of SESs is extremely small as explained in section 4.3.1, it is difficult to find an SES despite of the independent components. So $x_{23}^*(t)$ of Fig.14 and Fig.15 are transformed into the FFT frequency domain data, and the frequency domain data corresponding the SES is compared with the data without the SES. To examine the similarity between each train noise separated from the data with and without the SES, we compare the FFT frequency domain data corresponding train noise of Fig.14 and Fig.15.

The FFT frequency domain data of $x_{21}^*(t)$, $x_{22}^*(t)$, $x_{23}^*(t)$ of Fig.14 and Fig.15 is $f_1(t)$, $f_2(t)$, $f_3(t)$ of Fig.19 and Fig.20, respectively. The power of the lowest frequency $\frac{1}{3,500} \approx 2.85 \times 10^{-4}$ Hz in $f_1(t)$ is twice as large as one of Fig.19. Moreover, the power of the lowest frequency in another $f_1(t)$ separated from the data containing another train noise and the SES is also 1.5 times as large. For the comparison with the experimental results, we examine the power of the lowest frequency in the FFT frequency domain data transformed from the SES of Fig.3 and the data without train noise nor SESs. As a result, the power of the lowest frequency in the FFT frequency domain data transformed from an SES is twice as large as one transformed from the data without train noise nor SESs as well as the experimental results. It is supposed that the SES affects the lowest frequency power. Hence, it is presumed that $x_{21}^*(t)$ corresponds to the SES.

Each frequency domain component $f_3(t)$ of Fig.20 and Fig.19 which is presumed to be train noise is almost correspondent. It is confirmed that each $x_{23}^*(t)$ of Fig.14 and Fig.15 corresponds to the same train noise.

Therefore, we confirmed that train noise can be separated correctly from the data containing both train noise and an SES, furthermore the SES can be separated using 3cumICA.

5. CONCLUSIONS

In this paper, we applied ICA with third-order cumulants (3cumICA) to TCD considering statistical analysis of train noise. As the result, it turned out that 3cumICA can separate both typical train noise at Matsushiro and train noise which affects TCD observed at two different observation points, Sasadani and Ikeda. Therefore 3cumICA can separate train noise correctly.

Applying 3cumICA to the data containing both train noise and

an SES, train noise can be separated correctly without the effect of the SES. The power of the lowest frequency in one of the independent components separated from the data containing both train noise and an SES is twice as large as one separated from the data containing only train noise. Hence it is assumed that 3cumICA can separate not only train noise but also SESs.

We intend to apply 3cumICA to more TCD sets at the same observation points as we used in this paper and different observation points, and examine whether train noise can be separated correctly in the future. We have not confirmed that the SESs separated by 3cumICA are really independent source signal because of our determination of 3cumICA considering statistic analysis of train noise. So we intend to analyze the data without train noise, and investigate methods for separating SESs correctly. We will try to detect SESs which have not been recognized yet in previous TCD.

6. REFERENCES

- [1] Nagao, T.. Evolution of earthquake prediction research, Kinmirai-sha (in Japanese) (2001).
- [2] <http://yochi.iord.u-tokai.ac.jp/>
- [3] Nagao, T.. Is Earthquake Prediction Possible or Not? - Earthquake Prediction by Telluric Current Monitoring - (in Japanese). *Japanese J. Multiphase Flow*, Vol.9, No.2, pp.98–104 (1995).
- [4] Koganezama, M., Shouno, H., Nagao, T., and Joe, K.. Separation of Train Noise and Seismic Electric Signals in Telluric Current Data by ICA (in Japanese), *IPSJ TOM*, Vol.43, No.SIG7, pp.92–104 (2002).
- [5] Yoshida, S., M. Uyeshima and M. Nakatani. Electric potential changes associated with slip failure of granite, Preseismic and coseismic signals, *J. Geophys. Res.*, 102, 14,883–14,897 (1997).
- [6] Yoshida, S., O. C. Clint, and P. R. Sammonds. Electric potential changes prior to shear fracture in dry and saturated rocks, *Geophys. Res. Lett.*, 25, 1577–1580 (1998).
- [7] Cardoso, J.F., Souloumiac, A. Blind beamforming for non Gaussian signals *IEEE Proceedings-F*, 140, 362–370 (1993)
- [8] Blaschke, T. and Wiskott, L. An Improved Cumulant Based Method for Independent Component Analysis, *Proc. Int. Conf. on Artificial Neural Networks*, (2002)