SEPARATION OF TRAIN NOISE AND SEISMIC ELECTRIC SIGNALS FROM TELLURIC CURRENT DATA BY ICA

Mika Koganeyama * *Hayaru Shouno* * *Toshiyasu Nagao* † *Kazuki Joe* *

* Nara Women's University
Graduate School of Human Culture
Nara City, Japan

† Tokai University
Earthquake Prediction Research Center
Shimizu City, Japan

ABSTRACT

Recently, detection of seismic electric signals (SESs) in telluric current data (TCD) observed using the VAN method has attracted notice for short-term earthquake prediction. However, since most of the TCD collected in Japan is affected by train noise, detecting SESs in TCD itself is an extremely arduous job. The goal of this research is to derive a method for detecting SESs, which is difficult for VAN method experts because of train noise. We believe that SESs and train noise are independent signals. Therefore we attempted to apply Independent Component Analysis (ICA) to several TCD data sets which were measured at Matsushiro, Nagano. As a result, train noise and SESs were successfully separated using ICA.

1. INTRODUCTION

Since the great Hanshin earthquake in 1995, short-term earthquake prediction has been investigated as an emergent and important research topic. Although some statistical methods are used for long-term earthquake prediction in conventional seismology, it is obviously difficult to apply the same statistical methods to short-term earthquake prediction [1]. Three Greek physicists (Varotsos, Alexopoulos, Nomikos) suggested the VAN method as one useful method [2][3]. The VAN method is known to be an effective method for short-term earthquake prediction based on observations of telluric current data (TCD) in many observation points. TCD is the measurement of the weak electric current flowing within the surface layers of the Earth. In the TCD observed using the VAN method, seismic electric signals (SESs) are often detected before the occurrence of strong earthquakes. Several earthquakes were successfully predicted by the VAN method in Greece. In recent years in Japan, TCD has been recorded for research about the VAN method at the Information Frontier Program on Earthquake Research (RIKEN IFPER) [4].

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

Considering this background, we began research on automatic short-term earthquake prediction applying engineering methods to TCD instead of manually detecting SES by experts of VAN method. In TCD that contains little train noise, the neural network approach demonstrates good performance [5]. So, the most important problem is the reduction of train noise. We have already tried to detect SES buried in train noise by a Multi-Layer Perceptron with Back-Propagation [6]. It seems effective for several training data sets, but it is of doubtful application for general TCD. Thus we believe we should first separate or reduce train noise effects by another method. In this research, we applied Independent Component Analysis (ICA) [7] which separates each independent source signal from a mixture of independent source signals. We believe that train noise and SES are independent source signals because the current generating mechanism may be different. Thus we assume TCD is composed of train noise and SESs. So we believe that train noise and SESs can be separated by applying ICA to the TCD.

In this paper, we apply ICA to the TCD observed at Matsushiro, Nagano, which is one of the observation points where train noise have been very clearly observed, and then evaluate the ICA results.

2. TELLURIC CURRENT DATA (TCD)

2.1. The observation method

TCD measures the electric potential difference between dipoles at 2 points. The electrodes are $Pb{-}PbCl_2$ pipe non-polarizing dipoles (40 cm in length and 3 cm in outer diameter) and are buried at a depth of 2 meters.

42 observation points have been installed mainly in the

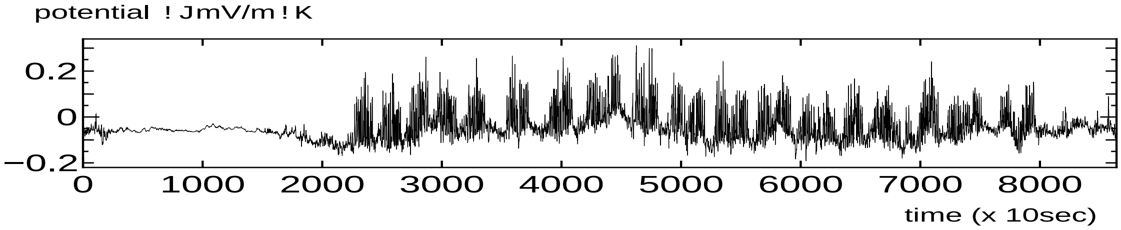


Fig. 1. Telluric Current Data! Jdp.2 of Matsushiro, Nagano on 20th of August, 1999 ! K

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
up-train	31	10 48	27	7	8	8	15	8	13	13	13	13 53	35	0 37	39	53
down-train	31	10 48	27	7	8	8 55		8 53	53	53	50	53	35	6 37	17	31

Table 1. Timetable (Nagano Railway Matsushiro Station)

Tokai and Hokuriku area since 1997. Each observation point has either 8 or 16 dipoles in different directions because the inherent noise of dipoles. Noise like SES which occurs distant from the observation point is observed on every dipole. If a noise is observed on either parallel dipole, the occurrence point of the noise is not distant from the observation point. Thus, when the measured data of either parallel dipoles is changed, the change is believed to be the effect of inherent noise. We label each dipoles dp.1, dp.2, ..., dp.16.

TCD is sampled at 10 second intervals and telemetered once a day to RIKEN IFPER. So TCD is expressed by time-varying 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 (x10sec).

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.

We can specify the train noise in TCD of Matsushiro shown in Fig.1 and the timetable of Tab.1[1]. We explain the specifying method as follows by an example of the first 6:31 train. 0 sec on horizontal axis means 0:00, and 23,460 sec represents 6:31. We can find a distinctive wave pattern between 22,200 and 24,400 sec in Fig.2 which is enlarged in Fig.1 between 21,000 and 25,000 sec. The wave pattern represents the train noise of the first 6:31 train.

[1] up-train = train which is bound to Tokyo, down-train = train which is bound from Tokyo

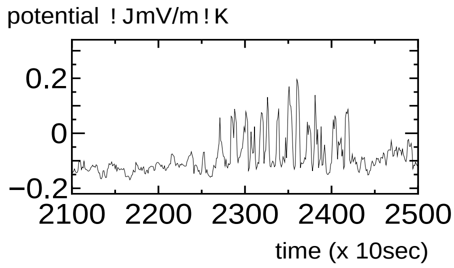


Fig. 2. Train noise (dp2 on 20th of August, 1999)

2.3. Seismic Electric Signals (SESs)

It is known from laboratory experiments that electric current is generated before rocks fracture under load [8][9]. 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. Fig.3 shows an SES as observed at dp.2 at approximately 1:30 on the 17th of January, 1999 in Matsushiro. In this datum, VAN method experts 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.

An example of short-term earthquake prediction by the VAN method is the great earthquake at Pirgos city in Greece on March 1993. Before the earthquake, SESs were observed in the TCD. As a result of acting on the earthquake prediction, although half of buildings in the city were completely or partially destroyed, there were no casualties [10]. Therefore the VAN method is available for short-term earthquake prediction.

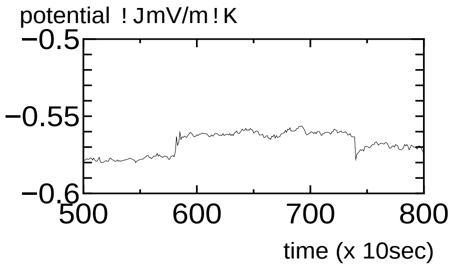


Fig. 3. Seismic electric signal (dp.2 on 17th of January, 1999)

2.4. Problems with using the VAN method in Japan

In Japan, the most serious problem for short-term earthquake prediction using the VAN method is the presence of train noise in TCD. Even if an SES is contained in the TCD, the SES is often hidden by train noise. Fig.4 shows train noise added to an SES artificially. It is difficult to classify data which contains both train noise and SESs (like Fig.4) in real TCD. On the other hand, since TCD is a direct current, we assume that the data containing SES and train noise will be equivalent to the linear addition of the train noise and the SES. For example, Fig.4 was generated by adding the train noise from 17th January 1999 and adding an SES in the range of point 50 to point 200 (Fig.5). The vertical axis of the frame shows potential (mV/m) and horizontal axis shows frame length (sampling point). A frame length represents the number of points whose sampling rate is 10 sec.

We cannot easily distinguish the SES in Fig.4. Therefore, even when TCD contains not only train noise but also actual SESs, it is quite difficult to find the SESs by hand.

We want to detect SESs by applying several engineering methods to TCD instead of using VAN method experts.

Since, train noise and SESs are generated by different sources, we assumed they are independent. TCD can be considered as the overlap of the electric potential of these signal sources. Assuming linear overlap, we can apply the standard ICA algorithm to the TCD. In this paper, we investigate the separation of train noise and SESs by applying ICA to TCD for Matsushiro, Nagano.

3. APPLICATION OF ICA TO TCD

In this chapter, we propose 3 categories of TCD data as follows :

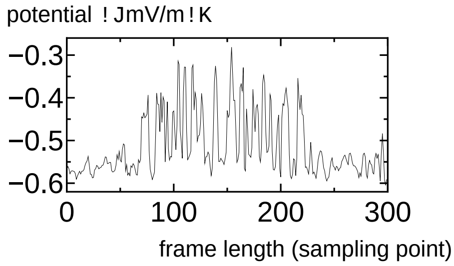


Fig. 4. Train noise added to an SES

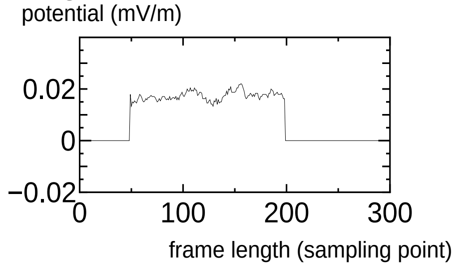


Fig. 5. an SES in the range of point 50 to point 200

1. the data contained train noise but no SESs.
2. the data contained an SES but no train noises.
3. the data contained both train noise and an SES.

These data are given to ICA as input, and we evaluate the separated output. The experimental purpose of pattern 1 and 2 is to examine whether only train noise or SES can be detected as an independent source signal. The pattern 3 is to examine whether the mixed train noise and SESs can be separated.

In this study, we apply ICA as follows.

When $\boldsymbol{\xi}(t)$ indicates input data, A denotes a mixed matrix, and $\boldsymbol{s}(t)$ denotes a source signal, the observed data can be assumed to be :

$$\boldsymbol{\xi}(t) = A\boldsymbol{s}(t).$$

We can observe only $\boldsymbol{\xi}(t)$, and ICA decomposes the observed signal into the independent signal $\boldsymbol{y}(t)$. It can be described formally as follows:

$$\boldsymbol{y}(t) = W\boldsymbol{\xi}(t).$$

The ICA algorithm estimates the independent signal $\boldsymbol{y}(t)$, and the matrix W simultaneously by enlarging the independencies of each component in $\boldsymbol{y}(t)$.

$x_i(t)$ is the observation data of the *i*-th dp.. We preprocessed $\boldsymbol{x}(t)$ to obtain the ICA input $\boldsymbol{\xi}(t)$ as follows. Applying Principle Component Analysis (PCA) to the TCD from dp.1–8 for each of the 3 categories $(x_1(t), x_2(t), ..., x_8(t))$, we can represent the 8 dimensions of the vector (corresponding to each dipole) with only three dimensions. The contribution ratio using three principal components is 99.9%. So we set the number of input dimensions to three. We use the $\boldsymbol{x}(t) = (x_2(t), x_6(t), x_7(t))^T$ because the dipoles are not parallel.

$\hat{\boldsymbol{x}}(t) = (\hat{x}_2(t), \hat{x}_6(t), \hat{x}_7(t))^T$ indicates the average normalized data. We denote the average of $\boldsymbol{x}(t)$ to be $\bar{\boldsymbol{x}} = (\bar{x}_2, \bar{x}_6, \bar{x}_7)^T$. So $\bar{\boldsymbol{x}}$ can be described by $\hat{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \bar{\boldsymbol{x}}$.

Λ denotes the diagonal matrix of eigenvalues of the covariance matrix of $\hat{\boldsymbol{x}}$, and matrix V is $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ which is the eigenvector for each eigenvalue. Then, R can be described as:

$$R = \sqrt{\Lambda}^{-1} V^T.$$

We define the input as sphered data $\boldsymbol{\xi}(t) = (\xi_1(t), \xi_2(t), \xi_3(t))$ as following :

$$\boldsymbol{\xi}(t) = R\hat{\boldsymbol{x}}(t).$$

The sampling time length of $\xi_i(t)(i = 1, 2, 3)$ should be chosen carefully because SESs are not long-term signals. If the sample is too long for a short-term signal like SES, the signal may not be able to be separated well by ICA. So we should choose the sampling time length so that an SES could be assumed not to be a short-term signal. Since the length of the SES which is used in this experiments is about 25 minutes, the length of $\xi_i(t)$ is 100 minutes in section 3.1 and 3.2, 50 minutes in section 3.3.

We adopted the sigmoid function as the non-linear function in ICA algorithm, and Kullback-Leibler divergence as the criterion for independence.

3.1. Data with train noise

In this section, the data with train noise but no SESs are applied to ICA. $x_i(t)(i = 2, 6, 7)$ is the TCD from about 6:00am to 7:00am on 20th of August 1999. Fig.6 shows the data of $x_2(t)$.

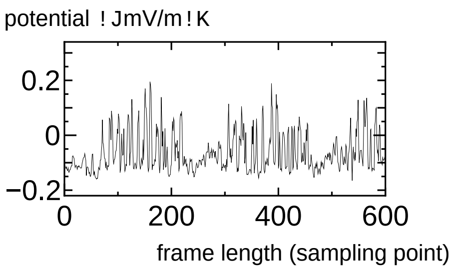


Fig. 6. $x_2(t)$

The size of the output data is equal to the size of the input data in ICA of the experiments, so the three graphs in Fig.7 are obtained. The output data $\boldsymbol{y}(t) = (y_1(t), y_2(t), y_3(t))^T$ denote independent signals.

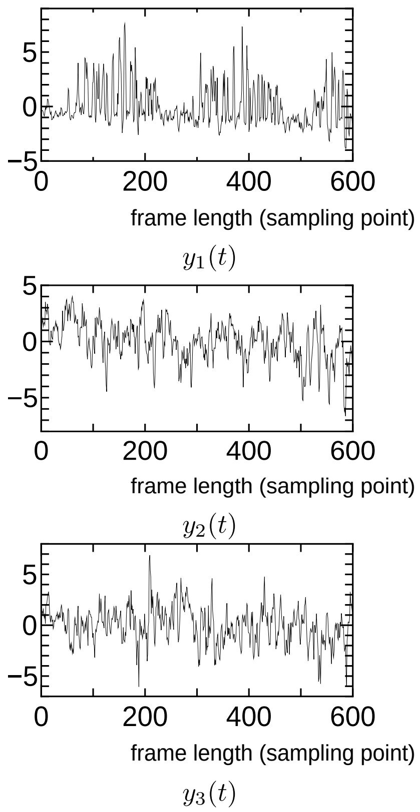


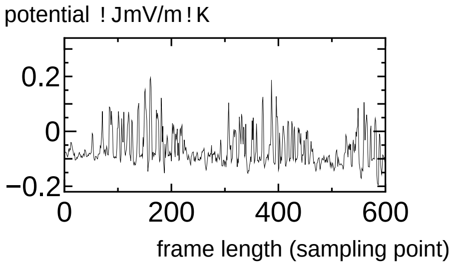
Fig. 7. Independent source signals

Since the vertical axis of $y_i(t)(i = 1, 2, 3)$ cannot be determined, it is hard to evaluate the experimental result as it is. So each $y_i(t)$ is transformed back to the original signal space, and then we evaluate how each signal affects $\boldsymbol{x}(t)$. To transform $y_1(t)$ to the original signal space, the value of $y_1(t)$ is kept and the value of other signals is set to 0. The data is described as $\boldsymbol{y}^*(t)$.

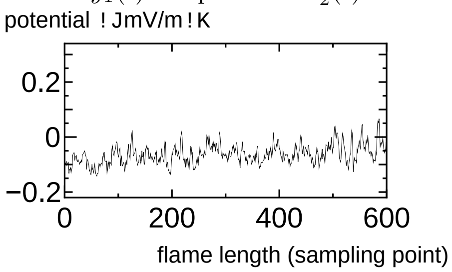
$$\boldsymbol{y}^*(t) = (y_1(t), 0, 0)^T$$
$$\boldsymbol{x}^*(t) = R^{-1}W^{-1}\boldsymbol{y}^*(t) + \bar{\boldsymbol{x}}$$

Thus the $y_1(t)$ component contained in $\boldsymbol{x}^*(t)$ represents $\boldsymbol{x}^*(t) = (x_2^*(t), x_6^*(t), x_7^*(t))$. Fig.8 shows each $x_2^*(t)$ to which $y_2(t), y_3(t)$ are transformed back to the original space similarly.

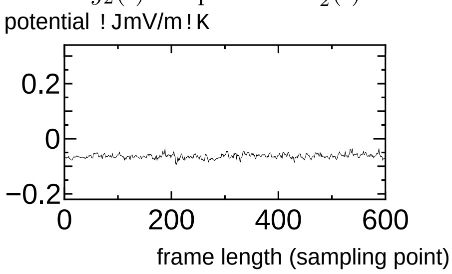
In Fig.8, it presumed that the signal $y_1(t)$ corresponds to train noise and the other signals appear to contain no train noise. Therefore it turned out that train noise can be separated from the TCD using ICA.



$y_1(t)$ component in $x_2^*(t)$



$y_2(t)$ component in $x_2^*(t)$



$y_3(t)$ component in $x_2^*(t)$

Fig. 8. Estimated independent source signals in $x_2^*(t)$

3.2. Data containing an SES

We use the TCD from 17th January 1999 as the observed data $x_i(t)(i = 2, 6, 7)$, which contains an SES but no train noise. Fig.9 shows $x_2(t)$ where an SES can clearly be observed in the data.

Fig.10 shows the presumptive independent source signals in $x_2^*(t)$ to which each signal in $\boldsymbol{y}(t)$ is transformed to the original space like in section 3.1. Seeing Fig.10, it is clear that the signal $y_3(t)$ corresponds to the SES and the other signals correspond to other noise besides the SES. Therefore, we have successfully determined that an SES can be separated from the data.

3.3. Data containing both train noise and an SES

In section 3.1, we evaluated only train noise data, and in section 3.2, we also evaluated only SES data. Each signal can be separated successfully. We use the data contained both train noise and an SES as input data. However, it is difficult to find the signals in real TCD as explained in section 2.4. We generated artificial data for dp.2, 6, 7 in the same manner proposed in section 2.4.

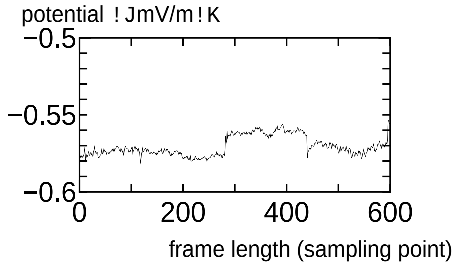


Fig. 9. $x_2(t)$

Fig.11 shows the estimated independent signals in $x_2^*(t)$. Comparing Fig.11 and Fig.5 which is an SES in $x_2(t)$, it is confirmed that the range shown by the arrow in $y_3(t)$ is identical to the range the SES is added in Fig.5. Hence, it is supposed that the signal $y_3(t)$ corresponds to the SES and the signal $y_1(t)$ corresponds to train noise. Therefore it turned out that while the experimental data is generated artificially, train noise and an SES are separated from the data contained both train noise and an SES by ICA.

4. CONCLUSION

The goal of our research is automatic short-term earthquake prediction by separating train noises and seismic electric signals from telluric current data using independent component analysis. In this paper, we applied ICA to several TCD collected at Matsushiro, Nagano. As a result, it turned out that train noise or an SES can be separated in a single channel. Moreover, it turned out that train noise and an SES are separated from the data contained both train noise and an SES, when the data was generated artificially. Therefore SESs which have not been recognized yet may be detected in old TCD using this analysis.

However, we have not confirmed that the output data of the experiments are really an independent source signal because of lack of experimental data. We intend to investigate into the independence of the signals using more output data in the future.

The input data used in the experiments covered only a short time span because SESs are not very long. On the basis of these experiments, we will propose an algorithm for non-stationary data.

Inherent dipole noise in TCD from each station are often observed, so we are going to examine methods for removing inherent dipole noise by pre-processing using ICA.

5. ACKNOWLEDGMENT

We wish to thank Steve Carroll for helpful comments on the paper.

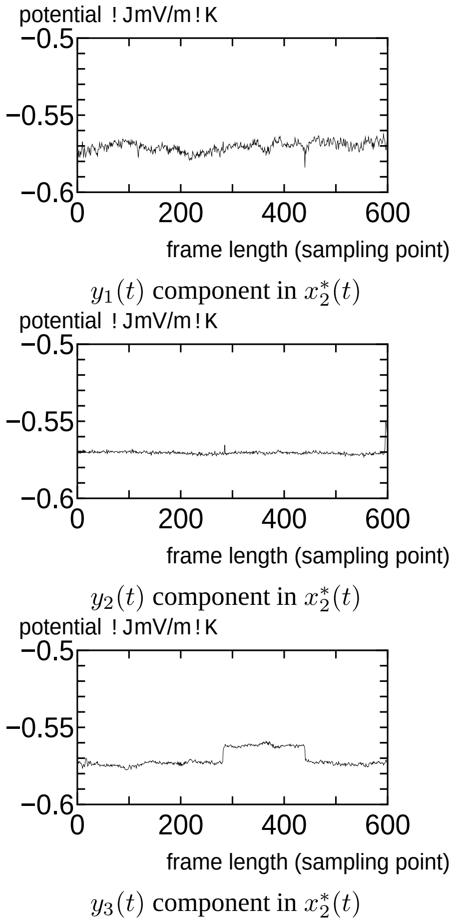


Fig. 10. Estimated independent source signals in $x_2^*(t)$

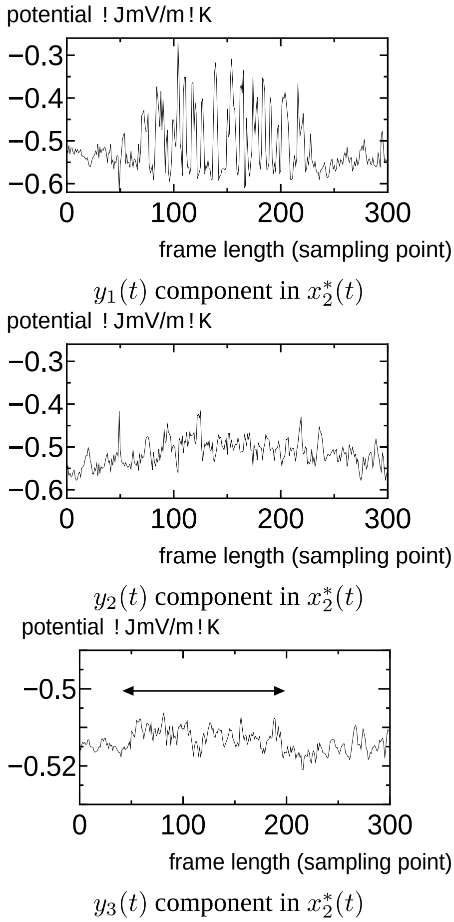


Fig. 11. Estimated independent source signals in $x_2^*(t)$

6. REFERENCES

[1] Nagao, T.. Evolution of earthquake prediction research, Kinmiraisha (in Japanese) (2001).

[2] Uyeda, S. Introduction to the VAN method of earthquake prediction!% *in Critical Review of VAN (ed. Sir James Lighthill)*, World Scientific, London, Singapore, 3-28.(1996).

[3] Nagao, T., Uyeshima, M. and Uyeda, S. it An independent check of VAN's criteria for signal recognition, *Geophys. Res. Lett.*, 23, 1441-1444(1996).

[4] http://yochi.iord.u-tokai.ac.jp/

[5] Fukuda, K., Koganeyama, M., Shouno, H., Nagao, T., and Joe, K. Detecting Seismic Electric Signals by LVQ based Clustering, *PDPTA2001*, pp.1305–1311(2001).

[6] Koganeyama, M., Nagao, T., and Joe, K.!%Reduction of Train Noise from Telluric Current Data by Neural Networks (in Japanese)!$*IPSJ TOMS*, (now printing) (2001).

[7] Murata, N.. Independent Component Analysis (in Japanese)!% *Computer Today*!$No.89, pp.55–61, SAIENSU-SHA Co,.Ltd. Publishers (1999).

[8] Yoshida, S., Uyeshima, M and Nakatani, M.!%Electric potential changes associated with slip failure of granite, Preseismic and coseismic signals, *J. Geophys. Res.*, 102, 14,883-14,897(1997).

[9] Yoshida, S., Clint. O. C., and Sammonds. P. R.. Electric potential changes prior to shear fracture in dry and saturated rocks, *Geophys. Res. Lett.*, 25, 1577-1580(1998).

[10] 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).