Design of an Integrated Database System for Short-term Earthquake Prediction

Chisato Ishikawa Hiroyasu Kamo Naoyuki Nide Kazuki Joe
Graduate School of Humanities and Sciences, Nara Women's University

Abstract
It is known that various electromagnetic methods have been investigated for the short-term earthquake prediction. To make integrated investigation of these methods and find some correlations of earthquakes and various phenomenon, an integrated database system for short-term earthquake prediction is needed. In this paper, we present the design of an integrated database system for the ICA (Independent Component Analysis) method and observed TCD (Telluric current data).

地震短期予測のための統合的データベースシステムの設計

石川千里 鴨 浩靖 新出尚之 城 和貴
奈良女子大学大学院人間文化研究科

概要
本稿は，地震短期予測のための統合的データベースシステムの設計について述べるものである．現在，様々な電磁気学的手法を用いた地震短期予測が考えられている．これらの結果を統合的に調査し，地震と電磁気学的な値のデータとの相関関係を発見するためには，統合的なデータベースが必要である．これまで，我々は地電流に対して ICA を適用することで地震短期予測を実現するという研究を行ってきたため，本稿では，統合的データベースの設計について，地電流に対して ICA を適用する部分を中心に述べる．

1. Introduction

It is known that various electromagnetic methods have been investigated for the short-term earthquake prediction. For reliable prediction of earthquakes, integrated investigation in the results by using these methods is needed. The integrated investigation may find some correlations of earthquakes and various phenomena such as solar activity that affects electromagnetic values. Therefore, an integrated database system for short-term earthquake prediction should be constructed.

We have studied the VAN method[1] for its application in Japan. The VAN method is a short-term earthquake prediction by observing telluric current data(TCD) to detect seismic electric signals(SESs) with dipoles buried in the ground. However, TCD observed in Japan contains other and stronger electric signals of train noise. The train noise makes the detection of SESs in Japan quite difficult. We have investigated the short-term earthquake prediction research by applying engineering techniques to the analysis of TCD. We have applied ICA(Independent Component Analysis) to TCD to separate train noise and SESs. In 2004, we got TCD observed at the Matsushiro observatory in Nagano prefecture that is 120km away from seismic focus of the Nigata Chuetsu earthquake (Oct. 23, 2004. M:6.8). By applying ICA to the TCD, we tried to extract rectangular signals of SESs a visual check as well as an engineering method. As a result of the visual check, we confirmed that rectangular signals of SESs are successfully detected by using ICA[6].

In this paper, to apply ICA to TCD, we describe design of an integrated database system for short-term earthquake prediction.

2. Construction of Tools and an Online System for the Application of ICA to TCD

In this section, we explain new tools and an online system which are concerned with applying ICA. These are used in the database system described in section3.

2.1. Development of tools

We develop two types of data viewers:

The viewer described in section 2.1.1 is used for detecting rectangular signals in ICA results by visual checks, and the viewer described in section 2.1.2 is used for analyzing ICA results. In order to set them up on a Web page, we use Java language for the development.

2.1.1. Data viewer for the detection of rectangular signals

We implement a data viewer which consists of several graphs for ICA results.

Fig.1 shows results of applying ICA to one-day TCD without framing[6]. Each graph in Fig.1

consists of the results of three channels in the TCD. In each graph, the intervals representing the possibility of rectangular signals, which are detected automatically as described in [6], are colored red.

Users observe the data to find some rectangular signals by visual checks with the viewer. There is the probability that a signal appears in multiple graphs of the same day simultaneously. Since users check each graph independently, it may happen that a signal is recorded as multiple signals on several channels. In order to avoid such conflicts, we implement a support function.

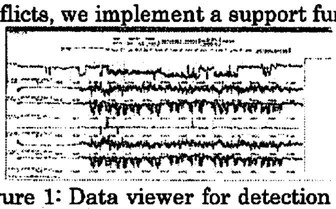

Figure 1: Data viewer for detection of rectangular signal

2.1.2. Data viewer helping for analyzing ICA results

In order to detect signals effectively, several filters such as noise reduction are needed for processing ICA result data. Hence we construct a new viewer, which provides users with only simple operations, for the data process as shown in Fig.2. This viewer generates a graph of ICA results through drag-and-drop of the data files to the viewer window.

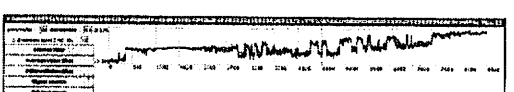

Figure 2: Data viewer helping for analyzing ICA results

As shown in Fig.3, median, average value and differentiation filters are applied to the graphs, by clicking the button labeled "Median filter", "Average value filter" and "Differentiation filter", respectively.

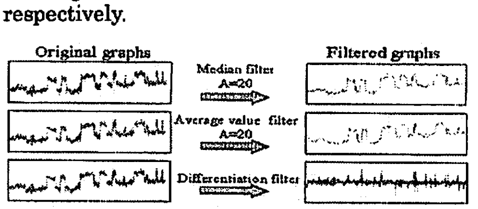

Figure 3: Original and filtered graphs

Once the button labeled "Signal search" is clicked, then an automatic detection of rectangular signals is applied to the graph on the viewer. We develop a new algorithm for the automatic detection. In the new detection algorithm, an interval covering rectangular signals is detected by using the old detection method [6], then rising and trailing edges are identified within the interval. The detected points are represented as bars on the horizontal axis of the graph as Fig.4. The bars in Fig.4 are presented with three colors according to the probability of SESs.

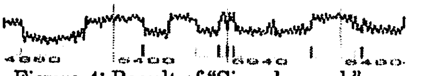

Figure 4: Result of "Signal search"

2.2. Construct of the ICA online system

We construct the ICA online system. The system executes our ICA program for the TCD observed at Matsushiro during Jan. to Nov. 2004, and displays graphs of the results.

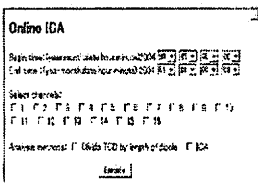

Figure 5:orm of the online ICA

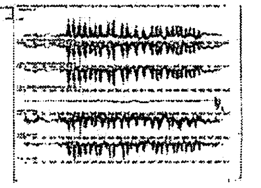

Figure 6: Display of results

Fig.5 is the Web form for specifying input files of TCD and the methods to process. In this form, users decide the interval of TCD by using the select-boxes, and choose the channels in the check-boxes. Once users choose some processes for the input TCD to click the button labeled "Execute", its result is displayed in a graph form as Fig.6.

3. Outline of Our Database

In Japan, we have forty observatories recording TCD for the past ten years. We also have other data sources such as magnetic field, radio wave, and FM wave rather than TCD. Thus, the total data file size must be huge. To deal with the huge data for analysis (e.g. ICA) efficiently, an integrated database of various observed data is required. For example, we need an integrated database for SESs and earthquake information to apply a data mining technique to obtain the knowledge for the short-term earthquake prediction. We use XML for constructing the above kind of databases. Since XML has high flexibility, it is suitable for our databases with various attributes.

We explain the multi-description database in section 4, and the signal analysis database in section 5.

4. Multi-description Database

In this paper, we explain the design of the multi-description database (MDD) for storing data obtained in multiple observatories. We know that TCD, a magnetic field, a radio wave, and an FM wave may have noise affected by earthquake hits. In this section, we focus on the design of the database for storing TCD.

4.1. Attribute of observed data

To design an effective database, we need to understand the structure of observed data. Multi-description database includes observed (measured) values as the main element. The observed data has several attributes such as data type, place and time information.

Relation among observed value and its attributes is as follows.

"Data type" means the type of observed data such as TCD or radio wave, etc.. Its attribute consists of a data type name, a measure unit, and applicable analyses.

"Observatory" is the place where the data is observed. It has the information of the name, latitude and longitude of the observatory.

For the possible earthquake prediction, each observatory should have various observation devices to support multiple data types. Dipoles for measuring TCD are a kind of observation device in an observatory. The attribute of "Dipole for TCD" includes the ID, length and direction of a dipole. Because of various observing intervals, the "measure" process requires measuring frequency as an attribute.

Observation time and data file name is attached as attributes of each "TCD".

4.2. Design of file for recording TCD

We estimate the total size of all the TCD data and the way to access the data.

Matsushiro observatory has fourteen dipoles for recording TCD every second. TCD is stored with the binary format, and its attribute is recorded in the XML files as meta-data. Since single precision (32bit) is an enough size to record the TCD, the storage cost for a dipole is 340KB par day, and the total cost is 690.5GB for all TCD in Japan.

Because ICA accepts only continuous-time data of each dipole, each file for recording TCD in the database should be generated for each dipole.

4.3. Design of XML file for TCD

The attribute of observed data is defined by its data type. If we record all attributes in an XML file, it is difficult to operate the XML file because of its complicated hierarchic structure. To avoid the problem, we use a main file, called root XML file, and several sub-files, called detail XML file. The root XML file defines the database, and has pointers to detail XML files. The detail XML files are generated for every combination of a data type and an observatory. Location information of observatories is recorded in another XML file named "Observatory XML file".

< Root XML file for multi-description database>

In this file, each data type has an element named "category". Each "category" includes a data type ID as an attribute. Sub-elements of a "category" are the data type name, the measure unit of the data type, available analysis methods, and elements, named "place", consisting of the observatory ID and the path to the detailed XML. Observatory ID is used to get location information from the observatory XML file.

<Detail XML files for observed data>

Detail XML files are generated for each combination of data type and observatory. Fig 11. describes the structure of the detail XML file for TCD as an example.

An element "detail" is defined as the root of the detail XML file. A data type ID and an observatory ID are attached as the attribute of the "detail" element.

An element has other elements as its sub-elements or has just a value. In either case, the element may have its attributes.

The "detail" has "dp_num", "interval", and "dp" elements. The "dp_num" and the "interval" elements have the values for the number of dipoles and the time interval for observing measuring, respectively. The "dp" element has other elements presenting the information of a dipole: start time of data recording, length and direction of the dipole, path to the data file. As the attribute for the "dp" element, a dipole ID is attached.

<Observatory XML file>

To avoid record duplication, we create an XML file for storing observatory information.

Each observatory has an individual "place" element. The "place" element has an observatory ID as its attribute. Sub-elements of "place" are the name, latitude and longitude of the observatory.

4.4. Basic operation tools for MDD

We prepare several operation tools for MDD as follows.

<Copy and paste TCD values tool>

By selecting an observatory, a dipole and a time range, the corresponding values are copied from the data file and pasted in a temporary file.

< Display attribute information tool>

By specifying an observatory, the information of the location, the data type and the length and direction of dipoles is displayed.

By selecting a data type, analysis methods and observatory names for the data type, are displayed.

<Observatory search tool>

This tool searches observatories with given conditions of latitude, longitude or observing data type, and display the search result.

<Statistical analysis tool>

Several statistical analyses such as computing average and variance are applied to the values in the selected range.

<Filtering tool>

This tool applies various filters, such as median and differentiation filters described in section 2.1.2, to the values in the selected range.

4.5. Using MDD in the ICA online system

The ICA online system described in section 2.2 can be modified to use MDD. We aim that TCD which the ICA online system needs is gotten from MDD through "Copy and paste TCD value" as described in the previous section. Data flow is as follows.

Information inputted with the Web form described in section 2.2 is sent to the bash script via the Perl script. In the bash script, the operation tool "Copy and Paste" for MDD is executed. It receives information of the input TCD from the bash script, and copies TCD to temporary files for each dipole according to the received information.

5. Signal Analysis Database

5.1. Outline of the signal analysis database

The purpose of the integrated database system is to explore the correlation between detected signals and earthquake records by data mining. The data mining requires a signal analysis database (SAD), which stores signal information. To find the correlation, existence or nonexistence of signals in analysis results is important. To record the both information effectively, we should divide the information into signal records and analysis operation logs. While signal records just contain detected signals, analysis operation logs include procedures for a series of analyses. By giving a log ID to each signal record, any signal record is connected to the corresponding analysis operation log. Existence or nonexistence of signals and any analysis process are obtained by using the connection information.

5.2. Design of SAD

We explain the XML files of SAD by using hierarchy diagrams, like section 4.3.

<XML file for analysis operation logs>

In this file, a "history" element is constructed for each series of analyses. For the relationship between a signal and some analyses, a history ID is attached to a "history" element as attribute. Sub-elements in "history" are "analysis" elements, which contain information of analysis, a "user ID" element including an ID of a user who executes the analyses, and "data" element containing information of data used in the series of analyses.

The "data" element has sub-elements: data type, IDs related to observatory, and start and end time of data.

In the "analysis" element, the order of the analysis in the series of analyzing is attached as attribute. Sub-elements of the "analysis" element are the analysis ID and parameters of input file names, output filenames and values.

<Signal record XML file>

Attributes of signals depend on the combination of data type and analysis. Thus, we should design XML elements for each combination. Elements named "tcd_ica_signal" are generated for each signal detected by using ICA results. The "tcd_ica_signal" element has a signal ID as attribute. Sub-elements of the "tcd_ica_signal" element are history ID which is used to refer how the signals are detected, starting time, length and amplitude of the signal.

< Analysis method XML >

The analysis method XML file is generated for the correspondence of an analysis ID to an analysis name and the version. An element "analysis" is generated for each analysis. The "analysis" element has an analysis ID as attribute, and has sub-elements of analysis name and version.

6. Conclusions

We presented the design of an integrated database system for short-term earthquake prediction in this paper.

We have future works as follows.

To implement the database designed in section 3 and 4, we will first construct a prototype of the database for TCD and ICA. Secondly, we will extend it for other observed data. Using SAD, we will develop an appropriate interface for the real detection of signals.

References

[1] S. Uyeda: Introduction to the VAN method of earthquake prediction, Critical Review of Van ,Sir Lighthill, J.(Ed.), pp.3-28, World Scientific, London, Singapore(1996).

[2] M. Koganeyama, T. Nagao, K. Joe: Reduction of Train Noise from Telluric Current Data by Neural Networks, IPSJ TOM, Vol.42, No.SIG14, pp.124-133, 2001 (in Japanese).

[3] M. Koganeyama, H.Shouno, T.Nagao, K. Joe: Separation of Train Noise and Seismic Electric Signals in Telluric Current Data by ICA, IPSJ TOM, Vol.43, No.SIG7, pp.92-104, 2002 (in Japanese).

[4] S. Sawa, M. Koganeyama, H. Shouno, T. Nagao, and K. Joe: Extraction of Train Noise affects Telluric Current Data at Two Different Observation Points by ICA and Its Statistical Evaluation, IPSJ MPS, MPS42-24, pp.95-98, 2002 (in Japanese).

[5] C. Ishikawa, C. Watanabe, T. Nagao and K. Joe: Extracting Seismic Electronic Signals from the Telluric Current Data for the Nigata Chuetsu Earthquake by ICA, IPSJ MPS, MPS04-53, p.61-65, 2005.

[6] C. Ishikawa, C. Watanabe, T. Nagao and K. Joe: Extracting Seismic Electronic Signals from the Telluric Current Data for the Nigata Chuetsu Earthquake by ICA, Vol.11,PDPTA, pp.488-494, 2005.