APPLICATION OF AUTOENCODER TO STRAIN TIME SERIES CALCULATED FROM GNSS OBSERVATION DATA

A. Mikami[(1)], M. Kamiyama[(2)]

[(1)] *Professor, Tokai University, atsushi.mikami@tokai.ac.jp*
[(2)] *Professor emeritus, Tohoku Institute of Technology, mkamiyam@tohtech.ac.jp*

Abstract

Applications of machine learning techniques in the field of earthquake engineering are in progress. This study presents a retrospective analysis of earthquake occurrence applying an autoencoder method, which is one of machine learning techniques, to the maximum shear strain time series calculated from GNSS (Global Navigation Satellite System) observation data. The main purpose of this study is to investigate if the method is capable of detecting premonitory symptoms of an earthquake. This study deals with the 2016 Kumamoto Earthquake, which struck the Kumamoto area of Kyushu Island in Japan with a magnitude of Mw 6.2 and a magnitude of Mw 7.0, causing fatal victims of more than 250 and badly damaging various kinds of structures in the Kumamoto Prefecture district. The temporal processes leading to the earthquake were accurately monitored by GEONET which stands for GNSS Earth Observation Network System operated by the Geospatial Information Authority of Japan. It covers entire area of Japan with approximately 1,300 reference points, and the observed data are provided through the Internet. Among observed data, this study utilizes F3 solution (daily coordinate value). The strain time series during the 2016 Kumamoto earthquake are numerically obtained in the neighborhood of the seismic source as well as in distant areas from the source. The Kyushu area is divided into triangular meshes with their nodes by the GEONET observation points. The shear strains of the triangular meshes are calculated from the deformation data at the nodal points using a method used in FEM. Then, the autoencoder, which belongs to the unsupervised learning, is applied to the time series of the strain data at each triangular element. The autoencoder used in this research is a three-layer neural-network consisting of the so-called encoder and decoder with accompanying dimension reduction. Considering that GNSS data are more or less contaminated with noises, the encoder is expected to efficiently work for denoising original input to obtain compressed representation. The weights of the network are determined so that the same data as the input layer are reproduced at the output layer. Hence, if normal data are given to the input, the network reproduce data similar to the input normal data at the output. In this study, the autoencoder learns a representation of the strain data for the period between the 2011 Tohoku earthquake and several years before the 2016 Kumamoto earthquake occurrence, assuming that this period of the data includes no earthquake symptom. Then, using all the strain data after the 2011 Tohoku earthquake until just before the earthquake occurred as an input, the autoencoder reproduces the output. The difference between the output and the original data is called as reconstructed error which is assumed to be related with abnormality of the maximum shear strain data. Obtained results showed that reconstructed errors drastically increased before the earthquake occurred near the large coseismic deformation area, whereas the method is not effective in the smaller coseismic deformation area. This study conclusively indicates that the application of machine learning techniques to the temporal variations of crustal strains is useful for detecting premonitory symptoms of an inland earthquake such as the 2016 Kumamoto Earthquake.

Keywords: autoencoder, GNSS observation data, 2016 Kumamoto earthquake

1. Introduction

The GNSS Earth Observation Network System operated by the Geospatial Information Authority of Japan (GSI) is called GEONET. Here, GNSS stands for Global Navigation Satellite System. The system covers the entire area of Japan with approximately 1,300 reference stations, and the observed data are provided through the Internet. The provided data can be utilized not only for surveying work but also for other purposes including i-Construction, disaster prevention for volcano and earthquake [1].

The authors have also been utilizing the data for the observation of recovery process from crustal movement caused by the 2011 off the Pacific Coast of Tohoku Earthquake, and have discussed effects of recovery from the crustal movement on the reconstruction work of fishing port facilities [2],[3]. We are also pushing utilization of the GNSS data for the earthquake prediction. All that matters in the prediction is how to make judgement on the phenomenon that is likely to be the symptom of occurrence. Without depending on a subjective judgement, objective judgement based on the objective data analysis is quite important. Kamiyama et al. [4] have attempted to detect symptom of earthquake occurrence from the pattern recognition of an index that is believed to represent the daily activity degree of the crustal movement.

In this study, a machine learning method used in the field of anomaly detection is incorporated for the prediction of earthquake occurrence. Autoencoder [5], which belongs to one of unsupervised machine learning algorithms, is applied to the strain time series (maximum shear strain) calculated from the GEONET observation data of displacements to detect any kind of anomaly. This study utilizes strain time series of ground obtained by Kamiyama et al. [6] for the 2016 Kumamoto earthquake near the epicenter as well as distant from the epicenter.

2. Evaluation of strain time series from GEONET data

In this study, strain time series are evaluated based on the method by Kamiyama et al. [6] for the Kyushu island during the time period including the 2016 Kumamoto earthquake. An assumption is made that strain time series up to several years before the main shock does not include any symptom of earthquake occurrence.

2.1 Data used

This study utilizes the F3 solution of daily coordinate value provided by the GSI. A value averaged over the 24 hours is open to the public, as 12:00 (UTC) value.

2.2 Element division

Kamiyama et al. [6] divided the whole Kyushu island by the triangular mesh based on the Delaunay triangulation algorithm as shown in Fig.1. The vertexes of each triangle correspond to the GEONET observation stations where two horizontal and one vertical displacements are available.

2.3 Maximum shear strain

Referring Kamiyama et al. [6], maximum shear strain of a triangular mesh is evaluated based on equation

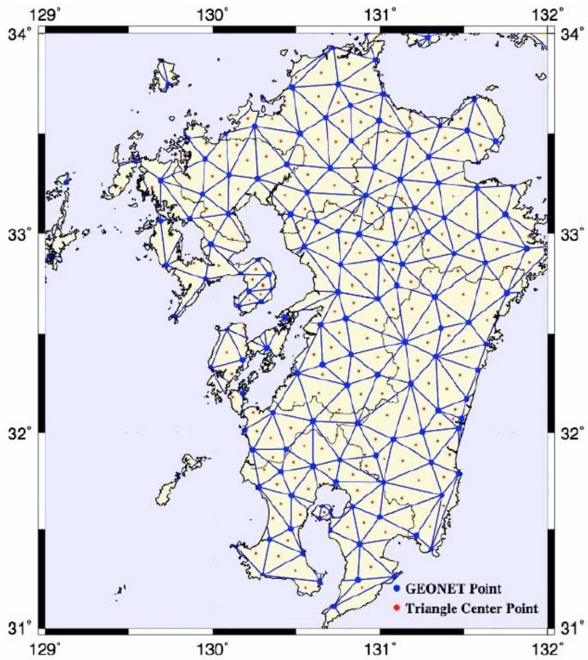

Fig.1 Delaunay triangulation

(1) through (4). Strain and displacement relationship is given by equation (1).

$$\{\varepsilon\} = [B]\{U\} \tag{1}$$

Here, $\{\varepsilon\} = \{\varepsilon_{EW}, \varepsilon_{NS}, \gamma_{NE}\}$ where ε_{EW} stands for normal strain in east-west direction, ε_{NS} is normal strain in north-south direction, γ_{NE} represents shear strain. Regarding the displacement vector, $\{U\} = \{u_i, v_i, u_j, v_j, u_k, v_k\}^T$, u_i, u_j, u_k are displacements in the east-west direction at the nodes i, j, k, v_i, v_j, v_k stands for displacements in the north-south direction at the nodes i, j, k. Maximum principle strain $\varepsilon_{\max}$, minimum principle strain $\varepsilon_{\min}$ and maximum engineering shear strain $\gamma_{\max}$ can be calculated from the following equations.

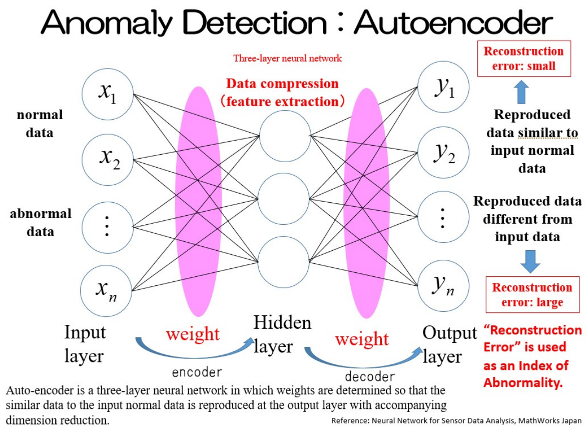

Fig.2 Autoencoder

$$\varepsilon_{\max} = \frac{\varepsilon_{EW} + \varepsilon_{NS}}{2} + \sqrt{\left(\frac{\varepsilon_{EW} - \varepsilon_{NS}}{2}\right)^2 + \left(\frac{\gamma_{NE}}{2}\right)^2} \tag{2}$$

$$\varepsilon_{\min} = \frac{\varepsilon_{EW} + \varepsilon_{NS}}{2} - \sqrt{\left(\frac{\varepsilon_{EW} - \varepsilon_{NS}}{2}\right)^2 + \left(\frac{\gamma_{NE}}{2}\right)^2} \tag{3}$$

$$\frac{\gamma_{\max}}{2} = \frac{\varepsilon_{\max} - \varepsilon_{\min}}{2} \tag{4}$$

3. Evaluation of reconstruction error by using autoencoder

3.1 Autoencoder

As shown in Fig.2, autoencoder [5] used in this research is a three-layer neural network consisting of the so-called encoder and decoder with accompanying dimension reduction. The network learns the characteristics of the input signal by weights so that output signal is reproduced as same as the input signal. Hence, if normal data is given as input, weights are determined so that normal data is reproduced as output. Considering that GNSS data are more or less contaminated with noises, the encoder is expected to efficiently work for denoising original input signal to obtain compressed representation. Note that this study used Neural Network Toolbox and Statistics and Machine Learning Toolbox of MATLAB [7] for the application of autoencoder.

3.2 Anomaly detection by reconstruction error

If the input signal is normal, its reconstruction error becomes small as almost the same signal is reproduced as output. Here, the reconstruction error is defined in this study as the square root of the mean square error between the input signal and the output signal. However, the reconstruction error becomes larger if the input signal includes some abnormal data, as the reconstructed signal becomes different from the input signal. In this study, the reconstruction error is assumed to be an index that reflects premonitory symptoms of an earthquake.

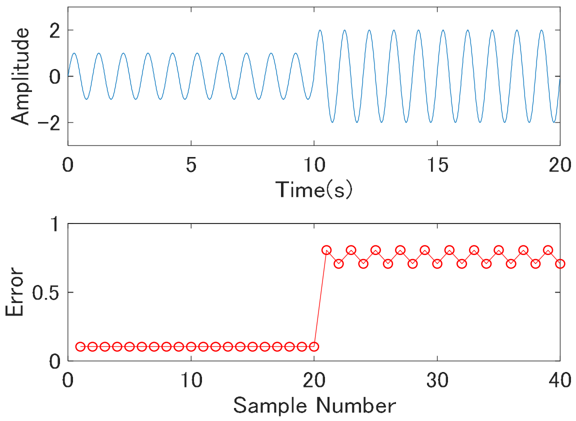

Fig.3 Change of amplitude

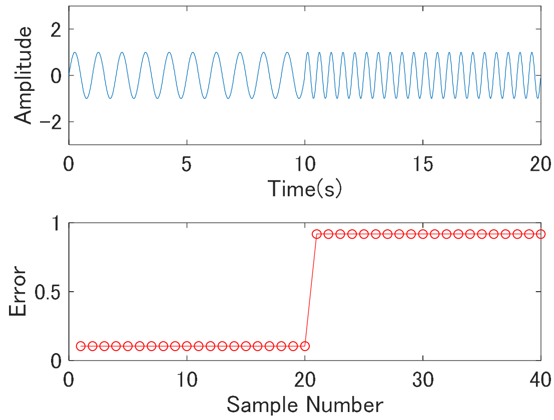

Fig.4 Change of frequency

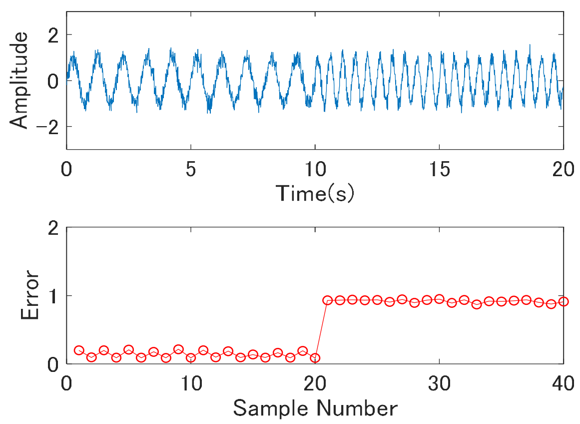

Fig.5 Change of frequency with noise

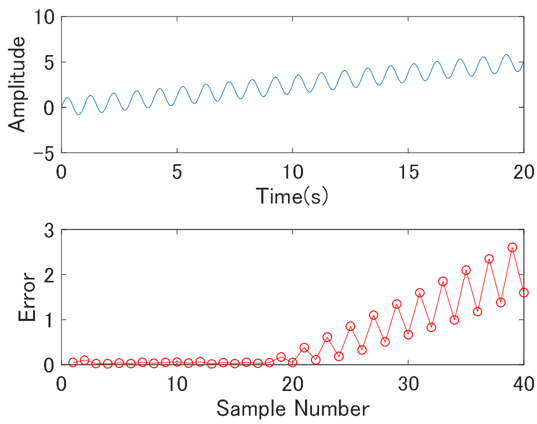

Fig.6 Effect of bias

3.3 Basic characteristics of autoencoder

Before applying the autoencoder method to the GEONET observation data, it is applied in this section to simple sinusoidal wave to deepen understanding of the algorithm. A sinusoidal wave that consists of 2000 data points with 0.01 sec time interval is prepared. The first half is used for learning and the whole data is used for the examination. One sample consists of 50 data points, hence, it follows that 2000 data consists of 40 samples.

Fig.3 shows a sinusoidal wave that changes its amplitude from 1 to 2 at the 10 sec. The bottom of Fig.3 shows calculated reconstruction error. It is understood from the results that the change of amplitude of the signal affects reconstruction error, and abnormality is detected.

Next, we consider a sinusoidal wave that changes its frequency from 1Hz to 2Hz at 10 sec. As shown in Fig.4, frequency change of a time series is recognized as a kind of anomalies. Effect of noise on the anomaly detection is also discussed in Fig.5. As previous case, a sinusoidal wave that changes its frequency from 1Hz to 2Hz is considered but the signal is contaminated with a noise generated based on normal random numbers. In this case, the frequency change is also detected as anomalies.

Some crustal movements exhibit distinguished ordinary crustal rising/descending, hence, deviation from the baseline of a sinusoidal wave of constant amplitude and frequency is considered next as shown in Fig.6. The reconstruction error shows that if the ordinary crustal movement is non-negligible compared with

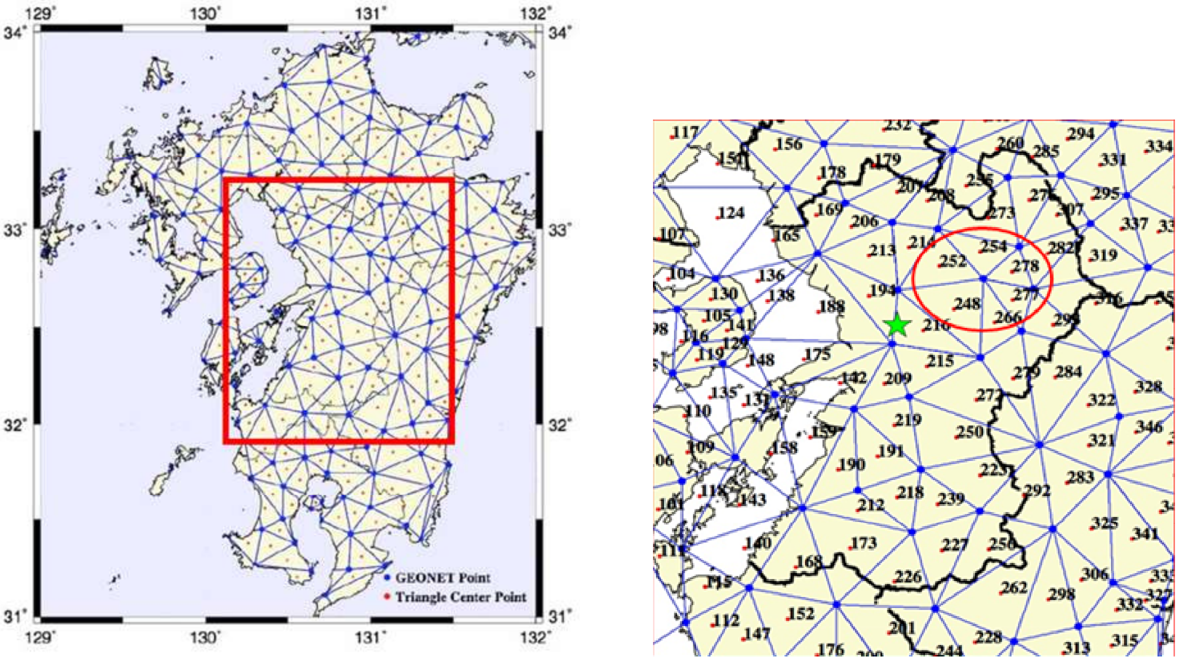

Fig.7 Triangular meshing in Kyushu island

Table 1 Available number of data

	Element	Available number of data between 2993-4850	Element	Available number of data between 2993-4850
1	252	1838	196	1848
2	248	1824	297	1815
3	254	1839	149	1819
4	266	1803	173	1829
5	277	1816	84	1849
6	278	1839	334	1850

the crustal movement by an earthquake, the ordinary movement may affect the results. Hence the baseline correction may be needed.

4. Results

Taking into account the above discussion, the autoencoder method is applied to the GEONET data. Maximum shear strain time series are evaluated for all the triangular elements in Kyushu (Fig.7) by Kamiyama, however, this study exemplarily used six time series near the large coseismic deformation area and another six time series distant from the large coseismic deformation area.

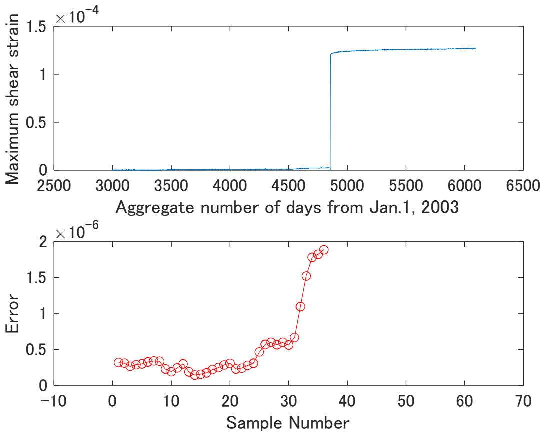

Fig.8 Element 252

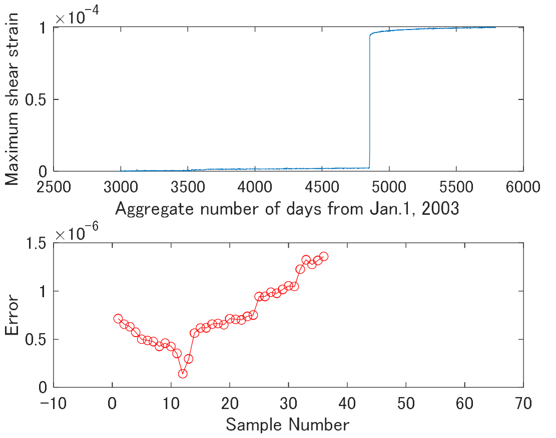

Fig.9 Element 248

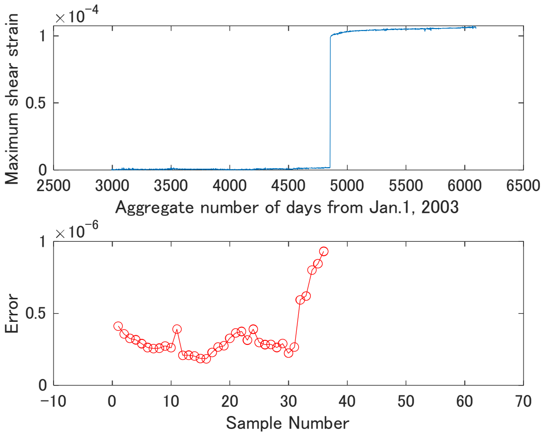

Fig.10 Element 254

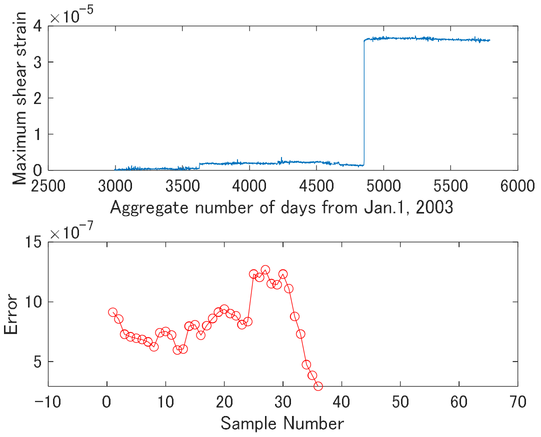

Fig.11 Element 266

4.1 Used data

Setting the starting date of observation to be Jan.1, 2003, the F3 solution data (the daily coordinate data) from March 12, 2011 (the aggregate number of days from Jan.1, 2003 is 2993) to the day several weeks before the 2016 Kumamoto earthquake were utilized. Here, March 12, 2011 is the next day after the 2011 off the Pacific Coast of Tohoku Earthquake. The biggest foreshock of the Kumamoto earthquake occurred on April 14, 2016 on which the aggregate number of days from Jan.1, 2003 is 4853. Hence, we assumed the data from 2993 to 4850 (a couple days before the foreshock) can be used at most for the purpose of this study.

Table 1 shows available number of strain data between 2993 and 4850. Because of data missing differently occurred at each observation station, the numbers of available strain data at each observation station are different from each other. This study used 1800 data, simply treating the daily coordinate values as a time series for the machine learning. Among the total data, the first half 900 data are used for the learning assuming that they include no earthquake symptom, and the total 1800 data are used for the final examination. Assuming that a single sample consists of 50 data, it follows that 36 samples were analyzed to calculate the reconstruction error for each element. Note that the dimension of the hidden layer is set as 25.

4.2 Reconstruction error near the large coseismic deformation area

We first focus on the elements near the large coseismic deformation area shown in Fig.7. The autoencoder method is applied to the maximum shear strain time series of the elements.

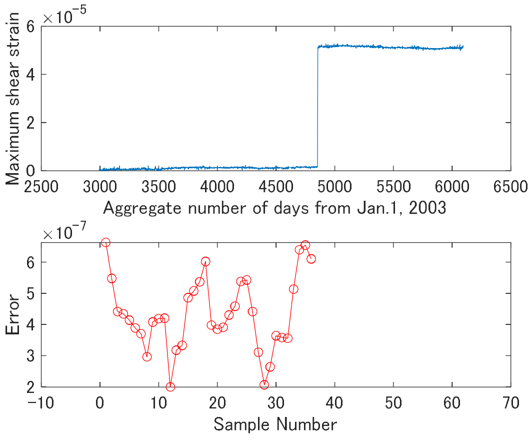

Fig.12 Element 277

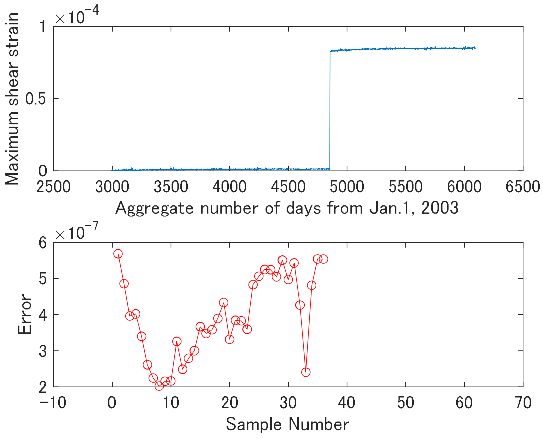

Fig.13 Element 278

As an example, explanation is given about element 252. The element 252 consists of 3 nodes (the Kikuchi, Choyo, and Kumamoto observation stations of GEONET). Fig.8 top shows maximum shear strain time series where some sudden change like a step-function can be seen at the time of the Kumamoto earthquake. The calculated reconstructed error is shown in the bottom of the figure. An drastic increase of reconstruction error can be recognized approximately 9 months before the earthquake.

An increasing tendency of the maximum shear strain also can be found for elements 248 and 254 as shown in Fig. 9 and 10, however, only element 254 showed drastic increase similarly to element 248. For other elements 266, 277 and 278, which locate in the smaller coseismic deformation area, such a drastic increase in the reconstruction error before the earthquake is unnoticeable as shown in Figs.11 through 13.

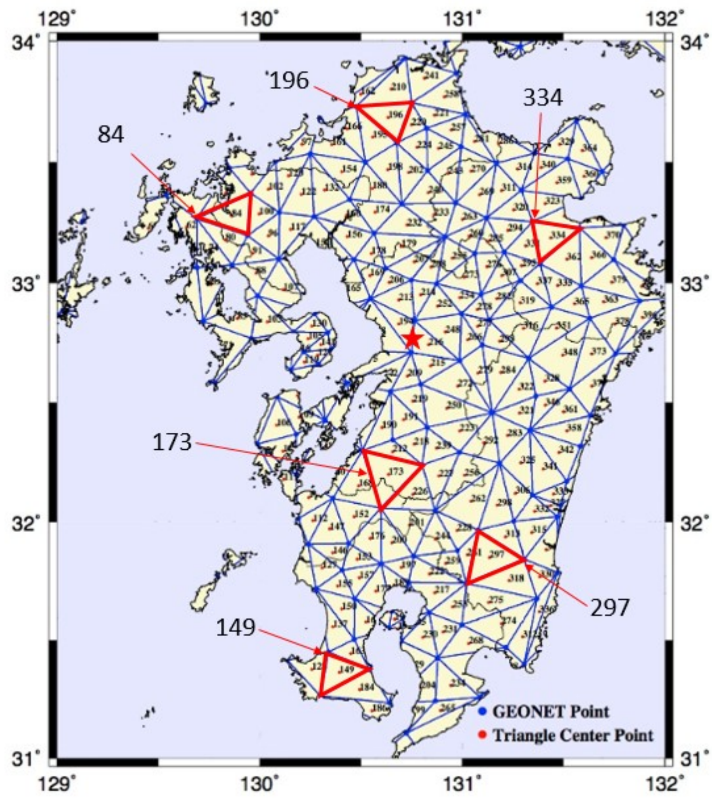

Fig. 14 Distant elements chosen

4.3 Reconstruction error distant from the large coseismic deformation area

Elements not only near the large coseismic deformation area but also distant from the large coseismic deformation area were also examined. Fig.14 shows the elements chosen for this purpose. The distances between the epicenter and the centers of each element range from approximately 65 km to 160 km in all directions. Triangular elements here were chosen so as to be regular shaped as possible.

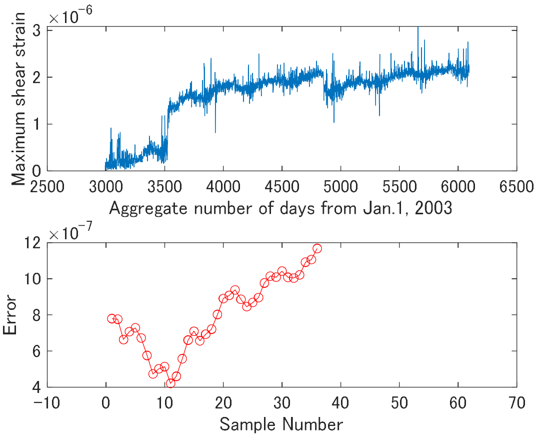

Fig.15 Element 196

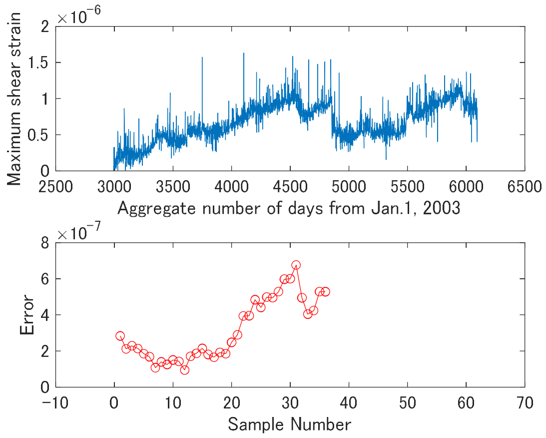

Fig.16 Element 297

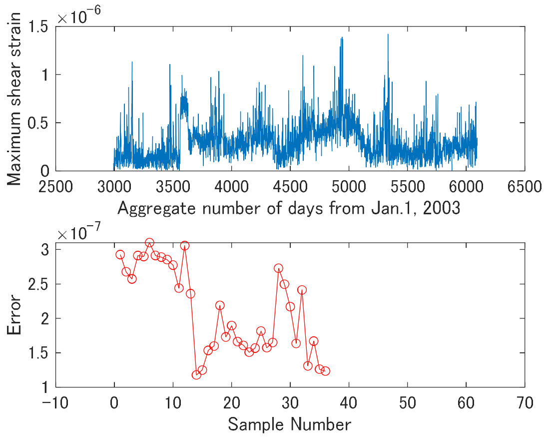

Fig.17 Element 149

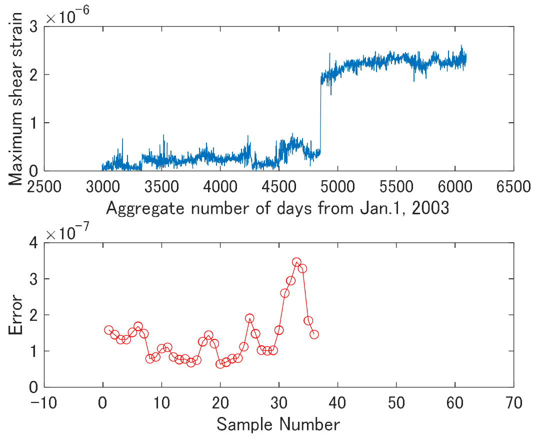

Fig.18 Element 173

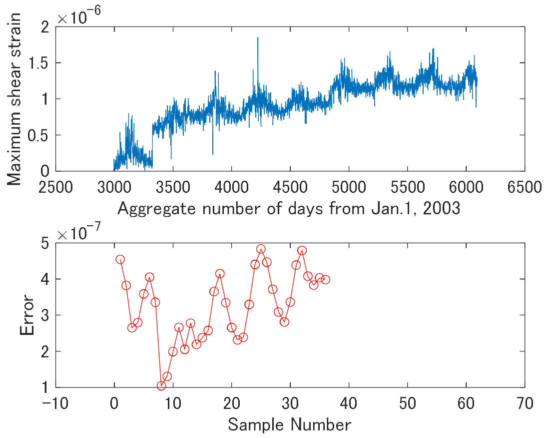

Fig.19 Element 84

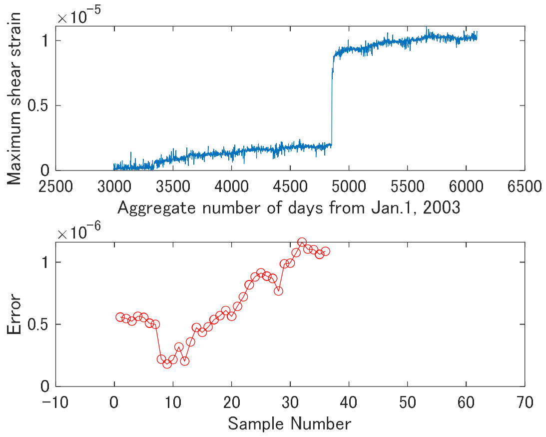

Fig.20 Element 334

Fig.15 through 20 show the maximum shear strain time series (top) and reconstruction error (bottom) for these elements. Looking at the bottom of these figures, none of them showed drastic increase of the reconstruction error.

5. Concluding remarks

This study presented a retrospective analysis of earthquake occurrence applying the autoencoder method, which is one of machine learning techniques, to the maximum shear strain time series calculated from GEONET observation data. The autoencoder learned a representation of the strain data by the weights for the time period between the 2011 Tohoku earthquake and several years before the 2016 Kumamoto earthquake occurrence, assuming that this period of the data includes no earthquake symptom. Then, using all the strain data after the 2011 Tohoku earthquake until just before the earthquake occurrence as an input, the autoencoder method reproduced the output. The difference between the output and the original input data is called as the reconstructed error which is assumed to be related with abnormality of the maximum shear strain data. Results showed that reconstructed errors showed drastic increase some months before the earthquake occurrence near the large coseismic deformation area, whereas the method is not effective in the smaller coseismic deformation area. This study conclusively indicates that the application of machine learning techniques to the temporal variations of crustal strains is useful for detecting premonitory symptoms of an inland earthquake such as the 2016 Kumamoto Earthquake.

6. Acknowledgements

This study used GEONET data provided by the Geospatial Information Authority of Japan (GIS).

7. References

[1] Geospatial Information Authority of Japan (2019): Overview of the GIS. (in Japanese)
https://www.gsi.go.jp/kikakuchousei/kikakuchousei40194.html.

[2] Mikami A, Kamiyama M (2016): Observation of recovery process from crustal movement caused by The 2011 off the Pacific Coast of Tohoku Earthquake. *Journal of JSCE, Division A1,* Vol.72, No.4, pp.I_413-I_423.

[3] Mikami A, Kamiyama M (2018): Effects of recovery from crustal movement on the reconstruction work of fishing port facilities due to The 2011 Tohoku Earthquake. *Journal of JSCE, Division B3,* Vol.74, No.2, pp.I_133-I_138.

[4] Kamiyama M, Koide, H, Sawada, Y, Akita, H, Chiba, N. (2017): Monitoring of crustal deformations and its application to mitigation of earthquake disasters. Journal of JSCE (Japan Society of Civil Engineers) Division A, Structural Engineering/Earthquake Engineering & Applied Mechanics;5:1206-225.

[5] Hinton GE, Salakhutdinov RR (2006): Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol.313 (5786), pp.504-507.

[6] Kamiyama M, Mikami A, Koide H, Sawada Y, Akita H, Chiba N (2017): Damage analysis for The 2016 Kumamoto Earthquake using coseismic strains obtained by GEONET data, *Journal of JSCE, Division A1*, Vol.73, No.4, pp.I_270-I_281.

[7] MathWorks Japan (2015): Neural Network for the analysis of sensor data. (in Japanese)

