Applied Network Science

RESEARCH Open Access

Spatio-temporal clustering of earthquakes based on distribution of magnitudes

Yuki Yamagishi[1,2*], Kazumi Saito[2], Kazuro Hirahara[2] and Naonori Ueda[2]

*Correspondence:
yamagishi.yuki@sist.ac.jp;
yuki.yamagishi.ks@riken.jp
[1] Faculty of Informatics, Shizuoka Institute of Science and Technology, Fukuroi, Japan
Full list of author information is available at the end of the article

Abstract

It is expected that the pronounced decrease in *b*-value of the Gutenberg–Richter law for some region during some time interval can be a promising precursor in forecasting earthquakes with large magnitudes, and thus we address the problem of automatically identifying such spatio-temporal change points as several clusters consisting of earthquakes whose *b*-values are substantially smaller than the total one. For this purpose, we propose a new method consisting of two phases: tree construction and tree separation. In the former phase, we employ one of two different declustering algorithms called single-link and correlation-metric developed in the field of seismology, while in the later phase, we employ a variant of the change-point detection algorithm, developed in the field of data mining. In the later phase, we also employ one of two different types of objective functions, i.e., the average magnitude which is inversely proportional to the *b*-value, and the likelihood function based on the Gutenberg–Richter law. Here note that since the magnitudes of most earthquakes are relatively small, we formulate our problem so as to produce one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller *b*-values. In addition, in order to characterize some properties of our proposed methods, we present a method of analyzing magnitude correlation over an earthquake network. In our empirical evaluation using earthquake catalog data covering the whole of Japan, we show that our proposed method employing the single-link strategy can produce more desirable results for our purpose in terms of the improvement of weighted sums of variances, average logarithmic likelihoods, visualization results, and magnitude correlation analyses.

Keywords: Declustering algorithm, Single-link, Correlation-metric

Introduction

Our research objective is to develop useful methods for analyzing huge earthquake catalogs as large-scale complex networks, where nodes (vertices) correspond to earthquakes, and links (edges) correspond to the interaction between them. Technically, we are not only interested in knowing what is happening now and how it develops in the future, but also we are interested in knowing what happened in the past and how it caused by some changes in the distribution of the information as studied in Kleinberg (2002) and Swan and Allan (2000). Thus, it seems worth putting some effort into attempting to find empirical regularities and develop explanatory accounts of basic properties in these complex networks. Such attempts would be valuable for understanding some structures

Springer Open

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

and trends and inspiring us to lead to the discovery of new knowledge and insights underlying these interactions.

The clustering of earthquakes is important for many applications in seismology, including seismic activity modeling, and earthquake forecasting. Here, for the sake of explaining our research purpose more specifically, we explain two important notions in seismology, i.e., the Gutenberg–Richter law and its parameter b referred to as the b-value. The Gutenberg–Richter law expresses the empirical relationship between the earthquake magnitude m and the number $N(m)$ of earthquakes with magnitudes equal to or larger than m in any given region and time period as $N(m) = 10^{a-bm}$ (Gutenberg and Richter 1954). Here a and b are the constant parameters where the value of a denotes the seismicity level in a region and the b-value is the slope parameter of the magnitude frequency curve which commonly close to 1.0 in seismically active regions. This means that when the b-value is smaller, the slope of the magnitude log-frequency curve is gradual, i.e., earthquakes with large magnitudes occur relatively frequently, while when the b-value is larger, the slope is steep, i.e., earthquakes with large magnitudes are relatively quite rare. Especially, it is expected that the pronounced decrease in the b-value of the Gutenberg–Richter law for some region during some time interval can be a promising precursor in forecasting earthquakes with large magnitudes (Nanjo et al. 2012). Thus we address the problem of automatically identifying such spatio-temporal change points as several clusters consisting of earthquakes whose b-values are substantially smaller than the total one.

In our previous work (Yamagishi et al. 2020), for this problem, we have proposed a basic idea of the method which uniquely combines some techniques developed in two different fields, i.e., declustering algorithms (Frohlich and Davis 1990; Davis and Frohlich 1991; Baiesi and Paczuski 2004) in the field of seismology and a change-point detection algorithm (Yamagishi et al. 2014; Yamagishi and Saito 2017) in the field of data mining. In this paper, we further propose to extend this method which enables to employ one of two different types of objective functions for our clustering, i.e., the average magnitude which is inversely proportional to the b-value, and the likelihood function based on the Gutenberg–Richter law. Here note that since the magnitudes of most earthquakes are relatively small, we formulate our problem so as to produce one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller b-values. Moreover, in order to characterize some properties of our proposed method, we newly present a method of analyzing magnitude correlation over an earthquake network. In our empirical evaluation using an earthquake catalog covering the whole of Japan, it was confirmed that we could generally obtain the clustering results each of which consists of one relatively large cluster and the other small clusters having substantially different average magnitude from the total one.

The paper is organized as follows. We describe related work in "Related work" section, give our problem setting and the proposed methods for clustering an earthquake catalog in "Proposed method" section, and present a method of analyzing magnitude correlation to characterize some properties of our proposed methods in "Magnitude correlation analysis" section. We report and discuss experimental results using a real catalog in "Experimental evaluation" section and conclude this paper and address the future work in "Conclusion" section.

Related work

In this paper, we propose a new method by combining declustering algorithms and a change-point detection algorithm developed in the two different fields of seismology and data mining, respectively. Thus, we describe some existing studies relating to these algorithms below.

Declustering algorithms

Seismicity declustering is the process of separating an earthquake catalog into foreshocks, mainshocks, and aftershocks, and several algorithms have been developed from various perspectives (van Stiphout et al. 2012). The window method is known as a simple way of identifying mainshocks and aftershocks. As a beginning of this method, the lengths and durations of windows were proposed by Knopoff and Gardner (1972) and Gardner and Knopoff (1974). After that, the alternative window parameter settings are proposed by Uhrhammer (1986), and comparative experiments were conducted by Molchan and Dmitrieva (1992). Meanwhile, the algorithm of Reasenberg (1985) called the cluster method assumed an interaction zone centered on each earthquake. This method based on the previous work of Savage (1972), and Molchan and Dmitrieva (1992) provide a condensed summary of the original paper of Reasenberg. As an alternative to deterministic declustering methods above, ideas of probabilistic separation appeared in the investigation of Kagan and Jackson (1991). Zhuang et al. (2002, 2004) and Zhuang (2006) suggested the stochastic declustering method also called stochastic reconstruction to bring such a probabilistic treatment into practice based on the epidemic-type aftershock sequence (ETAS) model (Ogata 1988a, 1998b). The generalization of stochastic declustering by Marsan and Lengliné (2008, 2010) has no specific underlying model and can accept any (additive) seismicity model. In other studies, Frohlich and Davis (1990) and Davis and Frohlich (1991) proposed the single-link cluster analysis based on a spatio-temporal metric between two earthquakes, and Hainzl et al. (2006) proposed the estimating background rate based on inter-event time distribution. Based on the inter-event times, the method by Bottiglieri et al. (2009) uses the coefficient of variation of the times, and Frohlich and Davis (1985) proposed the ration method which also exploits the inter-event times but without examining their distribution. As another cluster analysis with links, Baiesi and Paczuski (2004) proposed a simple spatio-temporal metric called correlation-metric to correlate earthquakes with each other and Zaliapin et al. (2008) further defined the rescaled distance and time.

Among these declustering algorithms, we focused on the single-link cluster analysis proposed by Frohlich and Davis (1990) and Davis and Frohlich (1991) and the correlation-metric proposed by Baiesi and Paczuski (2004). In our proposed method, we employ one of these two algorithms alternatively in the tree construction phase.

Change-point detection algorithms

Our research aim is in some sense the same, in the spirit, as the work by Kleinberg (2002) and Swan and Allan (2000). They noted a huge volume of the time-series data, tried to organize it, and extract structures behind it. This is done in a retrospective framework, i.e., assuming that there is a flood of abundant data already and there is a strong need to

understand it. Kleinberg's work is motivated by the fact that the appearance of a topic in a document stream is signaled by a "burst of activity" and identifying its nested structure manifests itself as a summarization of the activities over a period of time, making it possible to analyze the underlying content much easier. Kleinberg's method used a hidden Markov model in which bursts appear naturally as state transitions, and successfully identified the hierarchical structure of e-mail messages. Swan and Allan's work is motivated by the need to organize a huge amount of information in an efficient way. They used a statistical model of feature occurrence over time based on hypothesis testing and successfully generated clusters of named entities and noun phrases that capture the information corresponding to major topics in the corpus and designed a way to nicely display the summary on the screen (Overview Timelines). We also follow the same retrospective approach, i.e., we are not predicting the future, but we are trying to understand the phenomena that happened in the past.

We are interested in detecting spatio-temporal changes in the magnitude of earthquakes. For this purpose, by defining a set of links with some declustering algorithm described earlier, we construct a spatio-temporal network (spanning tree), where the nodes correspond to the observed earthquakes. After that, in order to analyze the burst of activity in an earthquake catalog and attempt to present an overview map, we employ a variant of our proposed change-point detection algorithm (Yamagishi et al. 2014; Yamagishi and Saito 2017).

Proposed method

Let $\mathcal{D} = \{(\mathbf{x}_i, t_i, m_i) \mid 1 \leq i \leq N\}$ be a set of observed earthquakes, where $\mathbf{x}_i$, t_i and m_i stand for a location vector, time and magnitude of the observed earthquake i, respectively. Here, we treat every earthquake (event) as a single point in a spatio-temporal space, as done by representative declustering methods in van Stiphout et al. (2012). Here, we assume that these earthquakes are in order from oldest to most recent, i.e., $t_i < t_j$ if $i < j$. In this paper, from the observed dataset $\mathcal{D}$, we address the problem of automatically extracting several clusters consisting of spatio-temporally similar earthquakes whose average magnitudes are substantially different from the total one. In what follows, we describe some details of our proposed algorithm consisting of two phases: tree construction and tree separation. Below, for a given number G of clusters, we describe our main algorithm, which produces a clustering result $\mathcal{R}_G$.

Phase1. Construct a spanning tree $\mathcal{T}$ of the observed earthquakes in $\mathcal{D}$ by using the single-link and correlation-metric strategies described below,

Phase2. Produce a clustering result $\mathcal{R}_G$ by separating the spanning tree $\mathcal{T}$ based on one of variance and likelihood criteria.

Figure 1 shows an illustrative example of a small tree, where we assign the time direction and the spatial distance between two earthquakes to the horizontal and vertical axes, respectively, and depicts each sample earthquake assumed to be in $\mathcal{D}$ as a circle so that a larger magnitude is indicated by a larger radius. From these earthquakes, Phase1 constructs a spanning tree $\mathcal{T}$ by producing a set of links between two earthquakes, and then Phase2 produces a set of clusters, just like a subtree $\mathcal{N}$ surrounded by the red dotted line,

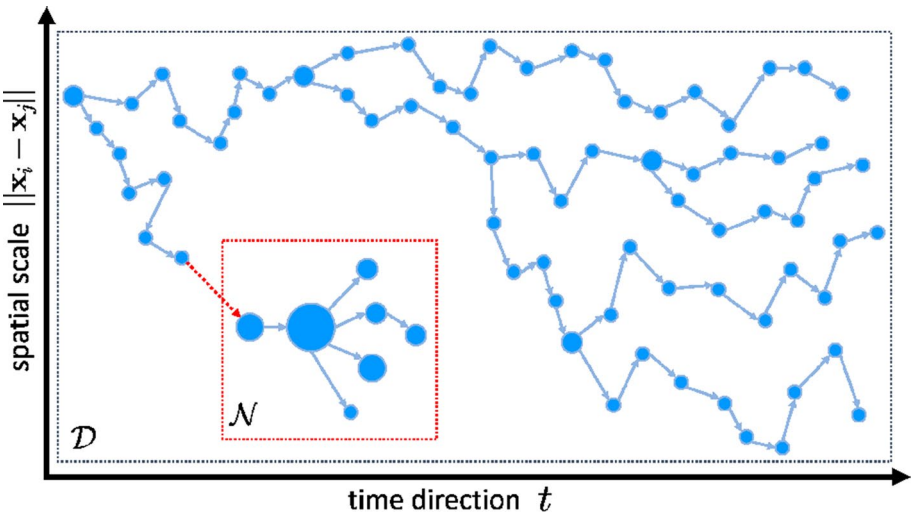


Fig. 1 An illustrative example on a small tree

by removing a subset of links, just like a link denoted by red dotted one, as shown in Fig. 1. Namely, as shown by $\mathcal{N}$ in Fig. 1, our purpose is to extract clusters having substantially larger average magnitudes or smaller b-values. In fact, although our problem setting can be categorized as event clustering according to the classification by Ansari et al. (2020), the purpose of clustering is substantially different from existing representative methods to our best knowledge.

Tree construction strategies

Among several seismicity declustering algorithms, we focus on two studies, i.e., the single-link cluster analysis proposed by Frohlich and Davis (1990) and Davis and Frohlich (1991), and the correlation-metric proposed by Baiesi and Paczuski (2004). In our experiments described later, it is shown that we obtain quite different extraction results by employing either one of these two strategies.

In the single-link strategy, with respect to two earthquakes i and j, the spatio-temporal metric $d_{i,j}$ is defined as

$$d_{i,j} = \sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|^2 + C^2(t_j - t_i)^2}. \tag{1}$$

It was found that a spatio-temporal scaling constant $C = 1$ km/day gives satisfactory results. Then, an earthquake j is regarded as the aftershock (child node) of i^{SL} if the metric $d_{i,j}$ is minimized, i.e., $i^{SL}(j) = \underset{1 \le i < j}{\arg\min}\, d_{i,j}$. Then, based on the single-link strategy, we can define a spanning tree, where the nodes correspond to the observed earthquakes, and the links are defined by $\mathcal{T}^{SL} = \{(i^{SL}(j), j) \mid 2 \le j \le N\}$.

In the correlation-metric strategy, with respect to two earthquakes i and j such that $i < j$, the spatio-temporal metric $n_{i,j}$ is defined as

$$n_{i,j} = (t_j - t_i)\|\mathbf{x}_i - \mathbf{x}_j\|^{d_f} 10^{-b\, m_i}. \tag{2}$$

Here d_f is the fractal dimension set to $d_f = 1.6$, and b the parameter of the Gutenberg–Richter law set to $b = 0.95$. Again, an earthquake j is regarded as the aftershock (child node) of $i^{CM}(j)$ if the metric $n_{i,j}$ is minimized, i.e., $i^{CM}(j) = \arg\min_{1 \le i < j} n_{i,j}$. Then, based on the correlation-metric strategy, we can define a spanning tree, where the nodes correspond to the observed earthquakes, and the links are defined by $\mathcal{T}^{CM} = \{(i^{CM}(j), j) \mid 2 \le j \le N\}$.

Tree separation algorithm

Let $\mathcal{R} \subset \mathcal{T}$ be a subset of tree links constructed by either one of the single-link and correlation-metric strategies. Here note that when $|\mathcal{R}| = G - 1$, by removing all the links in $\mathcal{R}$ from $\mathcal{T}$, we can separate the tree into G connected components. Then, the original set of observed earthquakes which correspond to nodes of the tree is also divided into G clusters as $\{\mathcal{N}_g \mid 1 \le g \le G\}$, where $\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_G = \{1, \ldots, N\}$. Now, by denoting the average magnitude of cluster $\mathcal{N}_g$ as

$$\mu_g = \frac{1}{|\mathcal{N}_g|} \sum_{i \in \mathcal{N}_g} m_i, \tag{3}$$

we can derive our first objective function $f_1(\mathcal{R})$ to be minimized as follows:

$$f_1(\mathcal{R}) = \sum_{g=1}^{G} \frac{|\mathcal{N}_g|}{N} \frac{1}{|\mathcal{N}_g|} \sum_{i \in \mathcal{N}_g} (m_i - \mu_g)^2 = \frac{1}{N} \sum_{g=1}^{G} \sum_{i \in \mathcal{N}_g} (m_i - \mu_g)^2. \tag{4}$$

Namely, we employ the definition of weighted variance in this objective function.

On the other hand, with respect to observed magnitudes for each of the G clusters $\{\mathcal{N}_g \mid 1 \le g \le G\}$, by assuming an exponential distribution with a parameter λ_g, we can consider the following logarithmic likelihood function:

$$L(\lambda_1, \ldots, \lambda_G; \mathcal{R}) = \sum_{g=1}^{G} \sum_{i \in \mathcal{N}_g} \log(\lambda_g \exp(-\lambda_g (m_i - m_c)) \tag{5}$$

where m_c denotes a cut-off magnitude defined as $m_c = \min_{1 \le i \le N}\{m_i\} - m_\Delta/2$, and m_Δ stands for the bin-width of observed magnitudes, i.e., $m_\Delta = 0.1$ in our catalog. Here we should clarify the relationship between the Gutenberg–Richter law and the above formulation, i.e., the number $N(m)$ of earthquakes with magnitudes equal to or greater than m can be formulated by using the above exponential distribution as follows:

$$N(m) = N \int_m^\infty \lambda \exp(-\lambda(x - m_c)) dx = N \exp(-\lambda(m - m_c)) = 10^{a - bm}, \tag{6}$$

where $N = N(m_c)$, $a = \log_{10} N + \lambda m_c \log_{10} e$ and $b = \lambda \log_{10} e$. Then, by substituting the maximum likelihood estimator into Eq. (5),

$$\hat{\lambda}_g = \frac{\hat{b}_g}{\log_{10} e} = \frac{1}{\mu_g - m_c}, \tag{7}$$

we can derive our second objective function $f_2(\mathcal{R})$ to be minimized as follows:

$$f_2(\mathcal{R}) = -\frac{1}{N}\sum_{g=1}^{G}\sum_{i\in\mathcal{N}_g}\log(\mu_g - m_c) = -\frac{1}{N}\sum_{g=1}^{G}|\mathcal{N}_g|\log(\mu_g - m_c). \tag{8}$$

Namely, we employ the definition of average logarithmic likelihood in this objective function.

Intuitively, we intend to produce one relatively large cluster and the other small clusters having substantially large average magnitudes from the total one. In fact, since the distribution of magnitudes in a catalog reasonably obeys the Gutenberg–Richter law (exponential distribution), i.e., the magnitudes of most earthquakes are relatively small, it is naturally expected that we can improve the objective function by separating clusters of spatio-temporally similar earthquakes with relatively large magnitudes. Here note that this objective function can be interpreted as a weighted sum of variances.

In order to compute the resultant set of separation links $\mathcal{R}$, we employ a variant of our proposed change-point detection algorithm (Yamagishi et al. 2014; Yamagishi and Saito 2017). Namely, from the observed dataset $\mathcal{D}$, the tree $\mathcal{T}$ constructed by either the single-link and correlation-metric strategies, and a given the number of clusters G, our algorithm computes $\mathcal{R}$ as follows:

Step1. Initialize $g \leftarrow 1$ and $\mathcal{R}_0 \leftarrow \emptyset$.

Step2. Compute $e_g \leftarrow \arg\min_{e\in\mathcal{T}}\{f(\mathcal{R}_{g-1}\cup\{e\})\}$, and update $\mathcal{R}_g \leftarrow \mathcal{R}_{g-1}\cup\{e_g\}$.

Step3. Set $g \leftarrow g+1$ and then return to Step2 if $g < G-1$; otherwise set $g \leftarrow 1$ and $h \leftarrow 0$,

Step4. Compute $e'_g = \arg\min_{e\in\mathcal{T}}\{f(\mathcal{R}_G \setminus \{e_g\}\cup\{e\})\}$, and update $\mathcal{R}_G \leftarrow \mathcal{R}_G \setminus \{e_g\}\cup\{e'_g\}$ and then $h \leftarrow 0$ if $e'_g \neq e_g$; otherwise set $h \leftarrow h+1$,

Step5. Output $\mathcal{R}_{G-1}$ and then terminate if $h = G-1$; otherwise set $g \leftarrow (g \bmod (G-1)) + 1$ and then return to Step4.

More specifically, after initializing the variables in Step1, we compute the optimal g-th link in e_g by fixing the already selected set of $(g-1)$ links in $\mathcal{R}_{g-1}$ and add it to $\mathcal{R}_{g-1}$ as shown in Step2. We repeat this procedure from $g = 1$ to $G-1$ as shown in Step3. After that, we start with the solution obtained as $\mathcal{R}_{G-1}$, pick up a link e_g from the already selected links, fix the rest $\mathcal{R}_{G-1}\setminus\{e_g\}$ and compute the better link e'_g of e_g as shown in Step4, where $\cdot\setminus\cdot$ represents set difference. We repeat this from $g = 1$ to $G-1$. If no replacement is possible for all g, i.e. $e'_g = e_g$ for all $g \in \{1,\ldots,G-1\}$, then no better solution is expected and the iteration stops, as shown in Step5. Here, it is not guaranteed that the above algorithm theoretically produces the optimal result, but it is confirmed that the algorithm always computes the optimal or near-optimal solutions in our empirical evaluation (Yamagishi et al. 2014; Yamagishi and Saito 2017).

Magnitude correlation analysis

In order to extract clusters consisting of earthquakes with relatively large magnitudes, by using either one of the single-link and correlation-metric strategies, we want to construct a tree where such earthquakes are connected with each other. To examine the type of properties, we present a method of analyzing magnitude correlation over a network of

these observed earthquakes. Here note that we can regard this method as an assortative mixing analysis (Newman 2003) over networks in terms of magnitudes.

Let $G = (\mathcal{V}, \mathcal{E})$ be a network obtained from the observed earthquake dataset $\mathcal{D}$, where $\mathcal{V} = \{1, \ldots, N\}$, and we assign the direction of link $(i, j) \in \mathcal{E}$ according to the time direction, i.e., $t_i < t_j$. Then, for nodes $j \in \mathcal{V}$ and $i \in \mathcal{V}$, we define the sets of in-neighbor and out-neighbor nodes as $\mathcal{IN}(j) = \{i \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$ and $\mathcal{ON}(i) = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$, respectively. Let $\mathcal{V}(m) \in \mathcal{V}$ be a set of nodes with magnitude m, and then, for each observed magnitude m, we can compute the following out-magnitude and in-magnitude correlation functions

$$C_{in}(m) = \frac{1}{|\mathcal{V}(m)|} \sum_{j \in \mathcal{V}(m)} \frac{1}{|\mathcal{IN}(j)|} \sum_{i \in \mathcal{IN}(j)} m_i \tag{9}$$

$$C_{out}(m) = \frac{1}{|\mathcal{V}(m)|} \sum_{i \in \mathcal{V}(m)} \frac{1}{|\mathcal{ON}(i)|} \sum_{j \in \mathcal{ON}(i)} m_i \tag{10}$$

where recall these observed magnitudes are quantized with the bin-width $m_\Delta = 0.1$. Actually, when earthquakes with larger magnitudes have a tendency to be connected with larger ones, we can expect that both $C_{in}(m)$ and $C_{out}(m)$ become increasing functions with respect to m.

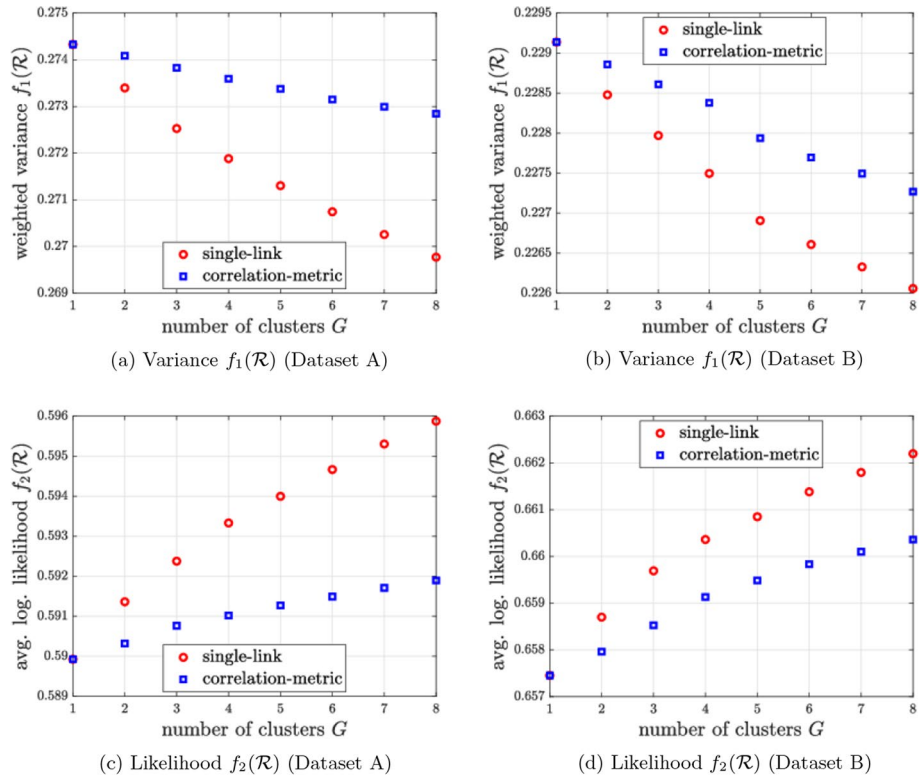
Experimental evaluation

By using an earthquake catalog which contains source parameters determined by Japan Meteorological Agency[1] in the whole of Japan, we generated two original datasets. Namely, by setting the minimum magnitude $m_{\min} = \min_{1 \le i \le N}\{m_i\}$ and the maximum depth as $m_{\min} = 3.0$ and 100 km, respectively, we selected $N = 104,343$ earthquakes during the period from Oct. 01, 1997 to Dec. 31, 2016 as dataset A, while by setting $m_{\min} = 4.0$ and 100 km, we selected $N = 27,728$ earthquakes during the period from Oct. 01, 1977 to Dec. 31, 2016 as dataset B.

Quantitative evaluation

First, we evaluate the performance of the proposed method employing our different tree construction strategies, i.e., single-link and correlation-metric. Figure 2 shows the experimental results of the datasets A and B, using the two objective functions, weighted variance $f_1(\mathcal{R})$ and average logarithmic likelihood $f_2(\mathcal{R})$, which are depicted in pairs of Fig. 2a, b and c, d, respectively, where the horizontal and vertical axes stand for the number of clusters varied from $G = 1$ to 8 and the objective function value defined in Eq. (4) or (8). Note that for each of Fig. 2a, b (or Fig. 2c, d) the value at $G = 1$ is nothing more than the total variance (or likelihood) of each dataset. From these experimental results, we can see that in the case of employing the single-link strategy, the objective function values defined by the weighted sums of variances become much smaller and those defined by the average logarithmic likelihood become much larger, in comparison to those of employing the correlation-metric strategy. This suggests that the proposed method employing the single-link strategy can

[1] https://www.data.jma.go.jp/svd/eqev/data/bulletin/hypo.html.



(a) Variance $f_1(\mathcal{R})$ (Dataset A)

(b) Variance $f_1(\mathcal{R})$ (Dataset B)

(c) Likelihood $f_2(\mathcal{R})$ (Dataset A)

(d) Likelihood $f_2(\mathcal{R})$ (Dataset B)

Fig. 2 Results of performance evaluation

produce more desirable results for our purpose of producing one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller b-values.

Next, we evaluate the similarity of the results with different numbers of clusters obtained by the proposed method. In what follows, we only show our experimental results using the dataset A, but reasonably similar results have been obtained for the dataset B. For this purpose, we employ the Adjusted Rand index (Rand 1971) for evaluating two different clustering results denoted by $\mathcal{G} = \{\mathcal{N}_g \mid 1 \le g \le G\}$ and $\mathcal{H} = \{\mathcal{N}'_h \mid 1 \le h \le H\}$, where $G \ne H$ in general and $\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_G = \mathcal{N}'_1 \cup \cdots \cup \mathcal{N}'_H = \{1, \ldots, N\}$. Let $c_{g,h} = |\mathcal{N}_g \cap \mathcal{N}'_h|$ denotes the number of objects in common between $\mathcal{N}_g$ and $\mathcal{N}'_h$, and let $cg_g = \sum_{h=1}^{H} c_{g,h}$ and $ch_h = \sum_{g=1}^{G} c_{g,h}$ denote the sums of $c_{g,h}$. Then, we can compute the original Adjusted Rand Index using the permutation model $ARI(\mathcal{G}, \mathcal{H})$ as

$$ARI(\mathcal{G},\mathcal{H}) = \frac{\sum_{g=1}^{G}\sum_{h=1}^{H}\binom{c_{g,h}}{2} - \left[\sum_{g=1}^{G}\binom{cg_g}{2}\sum_{h=1}^{H}\binom{ch_h}{2}\right]\Big/\binom{N}{2}}{\frac{1}{2}\left[\sum_{g=1}^{G}\binom{cg_g}{2} + \sum_{h=1}^{H}\binom{ch_h}{2}\right] - \left[\sum_{g=1}^{G}\binom{cg_g}{2}\sum_{h=1}^{H}\binom{ch_h}{2}\right]\Big/\binom{N}{2}}. \tag{11}$$

Figure 3 shows the similarity matrices consisting of the Adjusted Rand Indices by varying the number of clusters from $G = 2$ to 8, where Fig. 3a, b are those of the proposed method employing the single-link and correlation-metric strategies based on the weighted variance function $f_1(\mathcal{R})$, and Fig. 3c, d are those based on the average logarithmic likelihood function $f_2(\mathcal{R})$. From these experimental results, we can see that in the

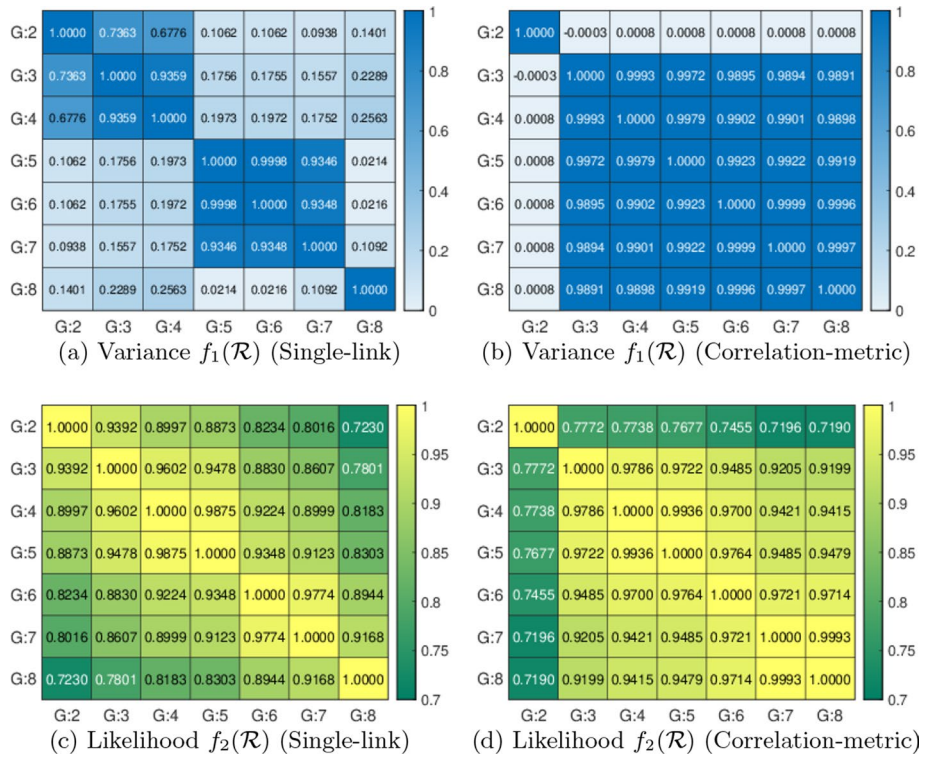


Fig. 3 Results of similarity evaluation

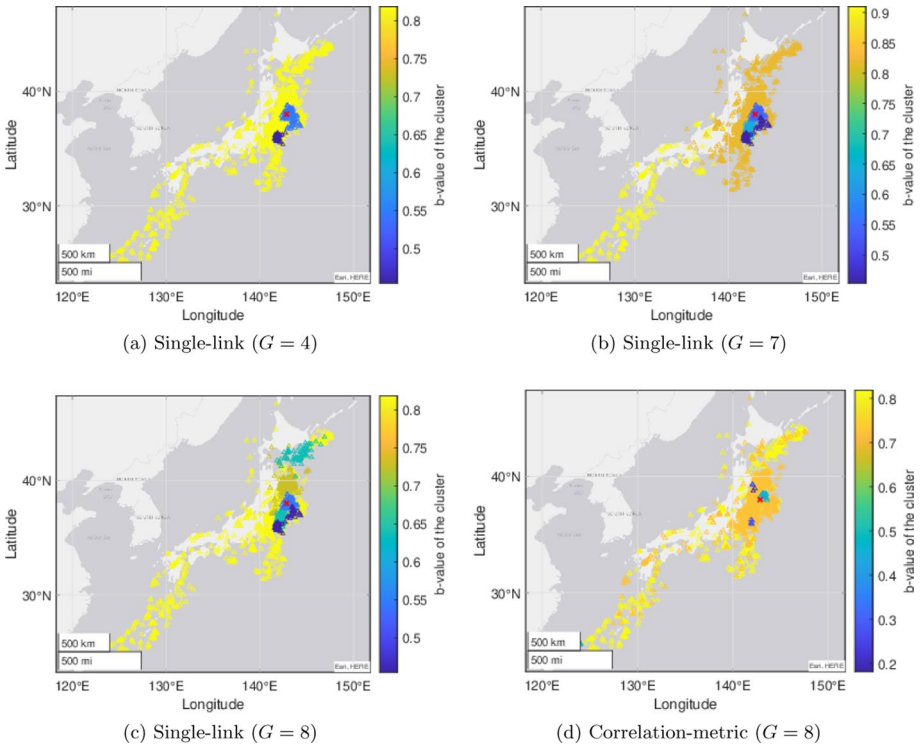
case of employing the single-link strategy, there exist three types of similar results, i.e., $2 \leq G \leq 4$, $5 \leq G \leq 7$ and $G = 8$ as for the weighted variance function, and $2 \leq G \leq 5$, $6 \leq G \leq 7$ and $G = 8$ as for the average logarithmic likelihood function, but almost a single type of similar results except for $G = 2$ in the case of employing the correlation-metric strategy, regardless of the types of objective functions. Namely, we can expect to obtain several types of results by varying G in the case of the single-link strategy. On the other hand, we can see that in the case of employing the weighted variance function, the ranges of the Adjusted Rand Indices were much wider than those employing the average logarithmic likelihood function.

Visual evaluation

First, we visually evaluate the obtained results employing the different tree construction strategies by focusing on the dataset A. Here, we characterize each cluster by using the corresponding b-value

$$b_g = \frac{\log_{10} e}{\mu_g - m_c}. \tag{12}$$

Recall that $m_c = m_{\min} - m_\Delta/2$, where the minimum magnitude $m_{\min}$ and the bin-width of observed magnitude m_Δ in the dataset A were set to $m_{\min} = 3.0$ and $m_\Delta = 0.1$. Here note that the average magnitude in dataset A is around 3.50, and the corresponding b-value is approximately amounts to 0.79. Figure 4 shows our visualization results based on the weighted variance function $f_1(\mathcal{R})$, where the numbers of clusters are 4, 7, and 8



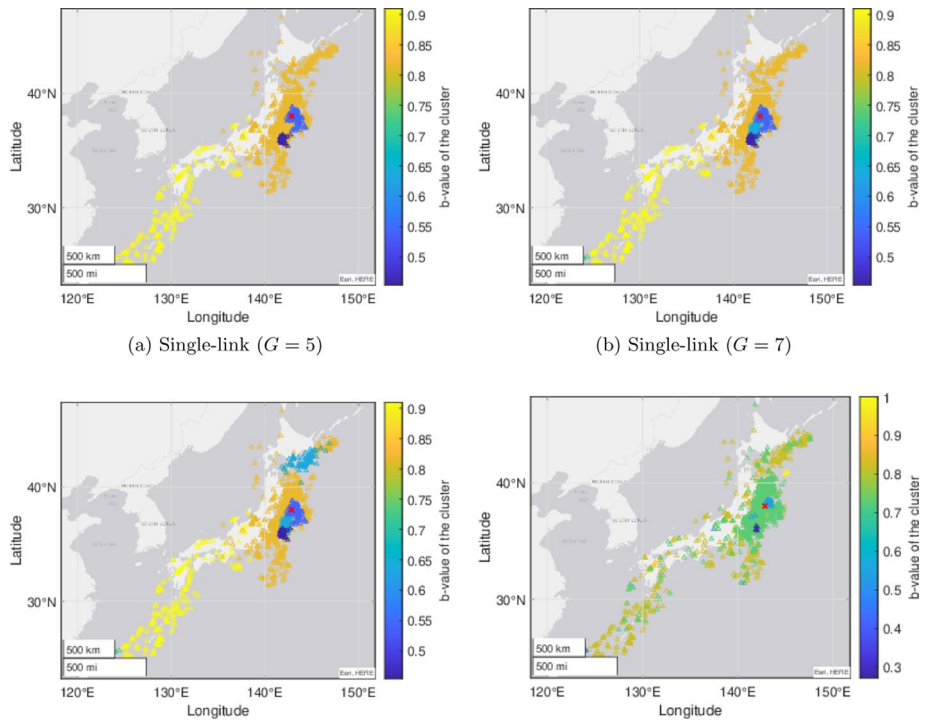
(a) Single-link ($G = 4$) (b) Single-link ($G = 7$)

(c) Single-link ($G = 8$) (d) Correlation-metric ($G = 8$)

Fig. 4 Results of visual evaluation based on Variance $f_1(\mathcal{R})$ (the largest magnitude earthquake in Japan $m_{\max} = 9.0$ is indicated by red cross)

for the single-link strategy, and 8 for the correlation-metric strategy. Here these results are selected according to the similarity matrices shown in Fig. 3a, b. Also note that by selecting earthquakes whose magnitudes are greater than or equal to 5.0, we plotted each of them as a triangle with a color shown in Fig. 4 according to its cluster's corresponding b-value. Similarly, Fig. 5 shows our visualization results based on the average logarithmic likelihood function $f_2(\mathcal{R})$, where the numbers of clusters are 5, 7, and 8 for the single-link strategy, and 8 for the correlation-metric strategy, Here these results are also selected according to the similarity matrices shown in Fig. 3c, d.

From these results, as expected, we could generally obtain the clustering results each of which consists of one or two relatively large clusters and the other small clusters having substantially different b-values from the total one. More specifically, as for the comparison between the two different strategies, single-link and correlation-metric, shown in Fig. 4c, d (or Fig. 5c, d), respectively, by employing the former strategy, we could obtain clearly visible clusters having substantially large average magnitudes around the region where the 2011 Tohoku earthquake with $m_{\max} = \max_{1 \le i \le N}\{m_i\} = 9.0$ (indicated by red cross in the figures), the largest magnitude in Japan, occurred. As for the comparison between the two different objective functions, weighted variance and average likelihood, shown in Figs. 4c and 5c (or Figs. 4d and 5d), respectively, by employing the average logarithmic likelihood function, the ranges of the b-values were somehow wider than those employing the weighted variance function. On the other hand, as for comparison among the different numbers of clusters in case of employing the single-link



(a) Single-link ($G = 5$) (b) Single-link ($G = 7$)

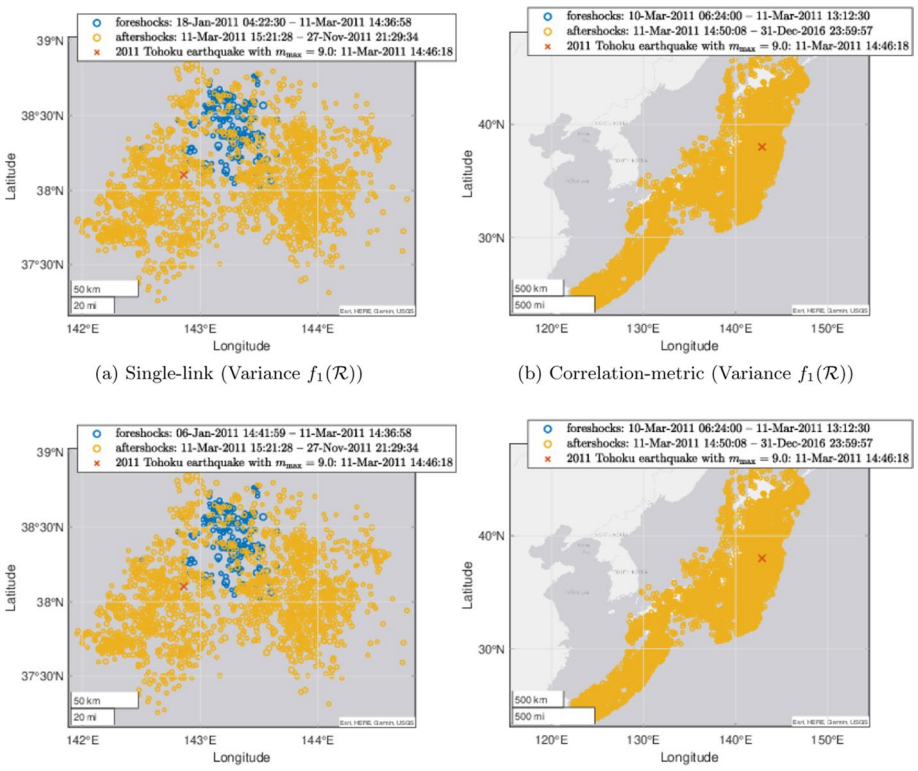
(c) Single-link ($G = 8$) (d) Correlation-metric ($G = 8$)

Fig. 5 Results of visual evaluation based on Likelihood $f_2(\mathcal{R})$ (the largest magnitude earthquake in Japan $m_{\max} = 9.0$ is indicated by red cross)

strategy, shown in Fig. 4a–c (or Fig. 5a–c), we could obtain somehow different types of clustering results, which might help to analyze the dataset from multiple viewpoints. In short, in our empirical evaluation, we can confirm that the proposed method employing the single-link strategy can produce more desirable results for our purpose of producing one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller b-values.

Next, under the setting that the number of clusters is 8 ($G = 8$), by focusing on the clusters which contain the 2011 Tohoku earthquake with $m_{\max} = 9.0$, we compare the individual clusters obtained by our methods employing different strategies for tree construction based on different objective functions for tree separation. Figure 6 shows our visualization results, where the results obtained by the single-link and correlation-metric strategies based on the weighted variance function $f_1(\mathcal{R})$ are shown in Fig. 6a, b, and those based on the average logarithmic likelihood function $f_2(\mathcal{R})$ are shown in Fig. 6c, d. In our visualization, we refer to the earthquakes that occurred before the 2011 Tohoku earthquake and those after it as foreshocks and aftershocks respectively, and depict foreshocks and aftershocks by blue and orange circles respectively, and the 2011 Tohoku earthquake by a red cross.[2] From these experimental results, we can clearly see that as for the results obtained by the single-link strategy, the clusters contain both foreshocks

[2] Of course, these terminologies, foreshocks and aftershocks, are not rigorously identical to those used in seismology, where both earthquakes refer to ones occurring on the faults of the mainshocks or in their extended regions. In this paper, we refer to earthquakes including a cluster, which occur simply before and after the largest earthquake as foreshocks and aftershocks respectively.



(a) Single-link (Variance $f_1(\mathcal{R})$) (b) Correlation-metric (Variance $f_1(\mathcal{R})$)

(c) Single-link (Likelihood $f_2(\mathcal{R})$) (d) Correlation-metric (Likelihood $f_2(\mathcal{R})$)

Fig. 6 Results of visual evaluation of the clusters which has the largest magnitude $m_{\max} = 9.0$ (when the number of clusters $G = 8$)

and aftershocks that occurred within a relatively narrow region, as shown in Fig. 6a, c. Conversely, as for the results obtained by the correlation-metric strategy, the clusters contain only aftershocks that occurred in a quite wide region, as shown in Fig. 6b, d. Moreover, we can see that the pairs of results obtained by the same tree construction strategy are quite similar, regardless of the types of objective functions, $f_1(\mathcal{R})$ and $f_2(\mathcal{R})$.

Figure 7 shows the estimated b-values over time with respect to the clusters shown in Fig. 6, where the results obtained by the single-link and correlation-metric strategies based on the weighted variance function are also shown in Fig. 7a, b, and those based on the average logarithmic likelihood function are shown in Fig. 7c, d. More specifically, we depict the b-values estimated within each cluster over time by blue lines, which are calculated from moving time windows each containing 100 earthquakes centered on the target earthquake based on Eq. (12), and depict the 90 % confidence intervals (Shi and Bolt 1982) by gray areas. Here we indicated the occurrence time of the 2011 Tohoku earthquake by a vertical red dotted line. From these experimental results, in the case of the single-link strategy, we can observe the pronounced decrease in b-value before the occurrence time of the 2011 Tohoku earthquake as shown in Fig. 6a, c. On the other hand, in case of the correlation-metric strategy, we cannot observe such a decrease in b-value because only an extremely short foreshock period is contained which is compared to the total period of the cluster, as shown in Fig. 6b, d. Again, we can see that the proposed method employing the single-link strategy can produce more desirable results

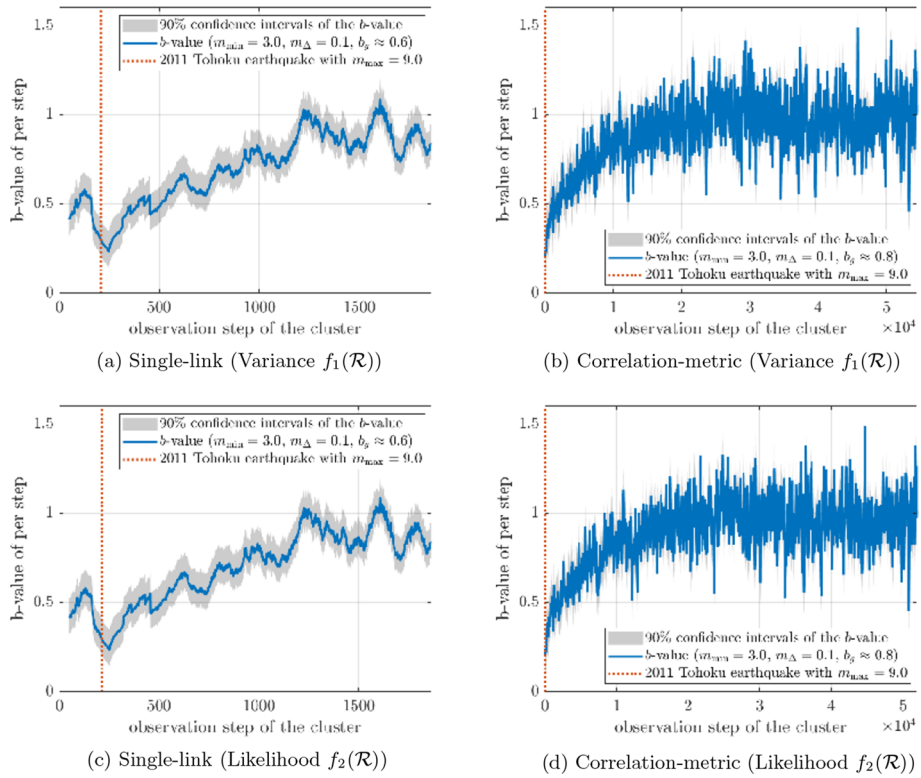


Fig. 7 Results of *b*-value of the clusters which has the largest magnitude $m_{\max} = 9.0$ (when the number of clusters $G = 8$)

for our purpose of producing one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller *b*-values.

Here recall that our research objective is also illustrated in Fig. 1, and we can confirm that the visualization results by the proposed method employing the single-link strategy are closer to that in Fig. 1 than those by the proposed method employing the correlation-metric strategy. In other words, as shown in the visualization results by the proposed method employing the single-link strategy in Fig. 6, if an earthquake of extremely large magnitude, such as the 2011 Tohoku earthquake, occurred and also many earthquakes of relatively large magnitudes occurred spatio-temporally close, the set of these earthquakes is the typical cluster such as the $\mathcal{N}$ in Fig. 1 which we aim to extract.

Magnitude correlation analysis

From our experimental results described above, we can see that the single-link strategy produces more desirable results than those by the correlation-metric strategy, regardless of the types of objective functions for tree separation. In order to clearly reveal this reason, we revisit the original trees constructed by the two strategies, and compare their properties in terms of the magnitude correlation analysis proposed in "Magnitude correlation analysis" section. Namely, we examine the properties that earthquakes with larger magnitudes have a tendency to connect each other.

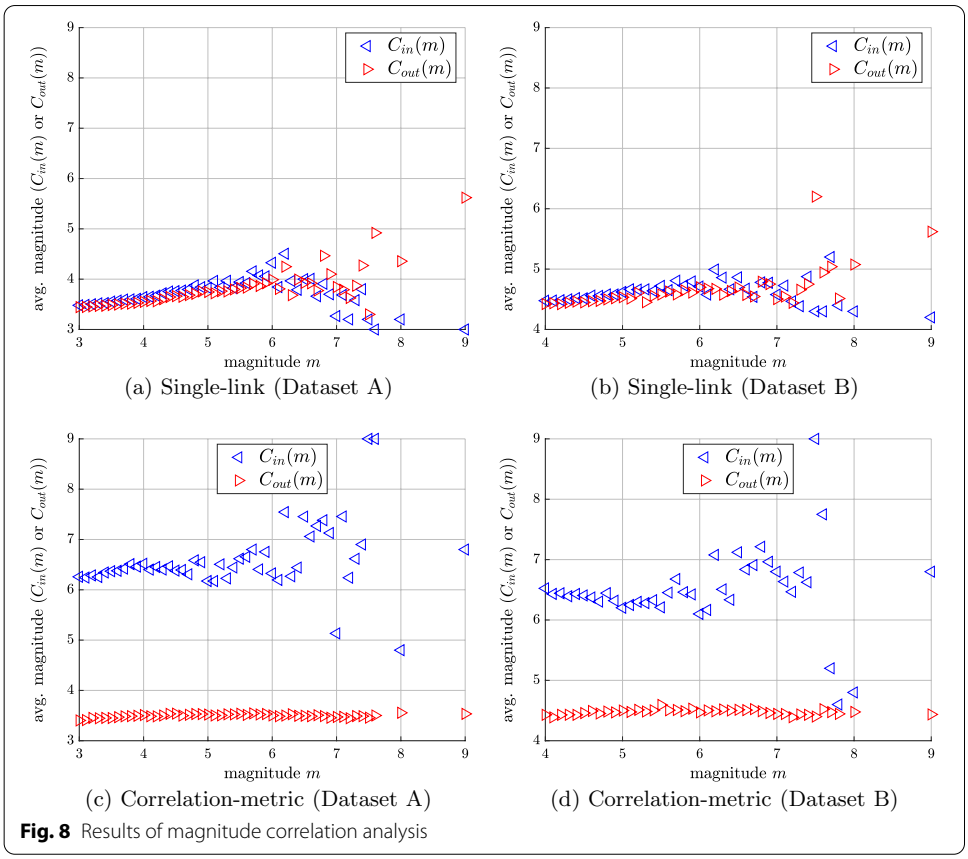


Fig. 8 Results of magnitude correlation analysis

Figure 8 shows the experimental results of the datasets A and B, where the pairs of Fig. 8a, b and c, d are those of employing the single-link and correlation-metric strategies, respectively. From these experimental results, we can observe that in the case of employing the correlation-metric strategy, the in-magnitude correlation values are much larger than the other values. This observation can be naturally explained by the fact that as for each link $(i^{CM}(j), j)$ obtained by the correlation-metric strategy, each earthquake $i^{CM}(j)$ with a quite large magnitude is likely to be selected according to the correlation-metric measure defined in Eq. (2). As another characteristic, we can observe that in the case of employing the single-link strategy, the in- and out-magnitude correlation values slightly increase as m becomes large, i.e., earthquakes with larger magnitudes have a tendency to connect each other. This also indicates that the single-link strategy has a desirable property for our purpose of producing one relatively large cluster and the other small clusters having substantially larger average magnitudes or smaller b-values.

Conclusion

In this paper, for a given dataset of observed earthquakes, we addressed the problem of automatically extracting several clusters consisting of spatio-temporally similar earthquakes whose magnitudes are substantially larger average magnitudes or smaller b-values in comparison to the total ones. Especially, we intended to produce one relatively large cluster and the other small clusters having substantially different average

magnitudes from the total one. For this purpose, we proposed a new method consisting of two phases. In the former tree construction phase, we employ one of two different declustering algorithms called single-link and correlation-metric developed in the field of seismology, while in the later tree separation phase, we employ a variant of the change detection algorithm, developed in the field of data mining, based on one of two different types of objective functions, i.e., the average magnitude which is inversely proportional to the *b*-value, and the likelihood function based on the Gutenberg–Richter law. In our empirical evaluation using earthquake catalog data covering the whole of Japan, it was confirmed that we could generally obtain the clustering results each of which consists of one relatively large cluster and the other small clusters having substantially different average magnitude from the total one. Moreover, we showed that the proposed method employing the single-link strategy can produce more desirable results, in terms of the improvement of weighted sums of variances, average logarithmic likelihoods, visualization results, and magnitude correlation analyses. As a future task, we plan to conduct more experiments to see that our clustering method can provide new findings on the earthquake statistics, the underlying earthquake dynamics, and so on, by producing one relatively large cluster and the other small clusters having substantially different average magnitudes from the total one. Further theoretical studies to find the optimal number of clusters are also future works.

Acknowledgements
Not applicable.

Author's contributions
All authors equally contributed to the technical content of the work from different perspectives. KS designed the basic algorithm. NU enhanced the algorithm. YY performed the experiments and drafted the paper. KH analyzed the experimental results from the seismology perspective. All authors read and approved the final manuscript.

Funding
The authors did not receive support from any organization for the submitted work.

Availability of data and materials
The dataset analysed during the current study is available in Japan Meteorological Agency (https://www.data.jma.go.jp/svd/eqev/data/bulletin/eqdoc_e.html).

Declarations

Competing interests
The authors have no competing interest to declare that are relevant to the content of this article.

Author details
[1]Faculty of Informatics, Shizuoka Institute of Science and Technology, Fukuroi, Japan. [2]Center for Advanced Intelligence Project, RIKEN, Kyoto, Japan.

Received: 8 March 2021 Accepted: 15 July 2021
Published online: 25 September 2021

References

Ansari MY, Ahmad A, Khan SS, Bhushan G, Mainuddin (2020) Spatiotemporal clustering: a review. Artif Intell Rev 53(4):2381–2423
Baiesi M, Paczuski M (2004) Scale-free networks of earthquakes and aftershocks. Phys Rev E Stat Nonlinear Soft Matter Phys 69:066106
Bottiglieri M, Lippiello E, Godano C, De Arcangelis L (2009) Identification and spatiotemporal organization of aftershocks. J Geophys Res 114:B03303
Davis SD, Frohlich C (1991) Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification. Geophys J Int 104(2):289–306
Frohlich C, Davis S (1985) Identification of aftershocks of deep earthquakes by a new ratios method. Geophys Res Lett 12:713–716

Frohlich C, Davis SD (1990) Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. Geophys J Int 100(1):19–32
Gardner JK, Knopoff L (1974) Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? Bull Seismol Soc Am 64(5):1363–1367
Gutenberg B, Richter C (1954) Seismicity of the earth and associated phenomena, vol 310, 2nd edn. Princeton University Press, Princeton, NJ
Hainzl S, Scherbaum F, Beauval C (2006) Estimating background activity based on interevent-time distribution. Bull Seismol Soc Am 96:313–320
Kagan Y, Jackson D (1991) Long-term earthquake clustering. Geophys J Int 104:117–133
Kleinberg J (2002) Bursty and hierarchical structure in streams. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining (KDD-2002), pp 91–101
Knopoff L, Gardner JK (1972) Higher seismic activity during local night on the raw worldwide earthquake catalogue. Geophys J Int 28:311–313
Marsan D, Lengliné O (2008) Extending earthquakes reach through cascading. Science 319(5866):1076–1079
Marsan D, Lengliné O (2010) A new estimation of the decay of aftershock density with distance to the mainshock. J Geophys Res: Solid Earth 115:09302
Molchan GM, Dmitrieva OE (1992) Aftershock identification: methods and new approaches. Geophys J Int 109:501–516
Nanjo K, Hirata N, Obara K, Kasahara K (2012) Decade-scale decrease in *b* value prior to the M9-class 2011 Tohoku and 2004 Sumatra quakes. Geophys Res Lett 39:L20304
Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256
Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27
Ogata Y (1998) Space-time point-process models for earthquake occurrences. Ann Inst Stat Math 50:379–402
Rand WM (1971) Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66(336):846–850
Reasenberg P (1985) Second-order moment of central California seismicity, 1969–1982. J Geophys Res 90:5479–5495
Savage WU (1972) Microearthquake clustering near Fairview Peak, Nevada, and in the Nevada seismic zone. J Geophys Res 77:7049–7056
Shi Y, Bolt BA (1982) The standard error of the magnitude-frequency *b* value. Bull Seismol Soc Am 72(5):1677–1687
Swan R, Allan J (2000) Automatic generation of overview timelines. In: Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval (SIGIR 2000), pp 49–56
Uhrhammer RA (1986) Characteristics of northern and central California seismicity. Earthq Notes 57(1):21–37
van Stiphout T, Zhuang J, Marsan D (2012) Seismicity declustering. Community Online Resource for Statistical Seismicity Analysis
Yamagishi Y, Okubo S, Saito K, Ohara K, Kimura M, Motoda H (2014) A method to divide stream data of scores over review sites. In: PRICAI 2014: trends in artificial intelligence—13th Pacific Rim international conference on artificial intelligence. Lecture Notes in Computer Science, vol 8862, pp 913–919. Springer
Yamagishi Y, Saito K (2017) Visualizing switching regimes based on multinomial distribution in buzz marketing sites. In: Foundations of intelligent systems—23rd international symposium, ISMIS 2017. Lecture Notes in Computer Science, vol 10352, pp 385–395. Springer
Yamagishi Y, Saito K, Hirahara K, Ueda N (2020) Spatio-temporal clustering of earthquakes based on average magnitudes. In: The 9th international conference on complex networks and their applications (ComplexNetwork2020). Lecture Notes in Computer Science, pp 627–637. Springer
Zaliapin I, Gabrielov A, Keilis-Borok V, Wong H (2008) Clustering analysis of seismicity and aftershock identification. Phys Rev Lett 101(1):1–4
Zhuang J (2006) Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data. J R Stat Soc 68(4):635–653
Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380
Zhuang J, Ogata Y, Vere-Jones D (2004) Analyzing earthquake clustering features by using stochastic reconstruction. J Geophys Res 109:B05301

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.