Possibility of short-term probabilistic forecasts for large earthquakes making good use of the limitations of existing catalogs

Yoshito Hirata, Koji Iwayama, and Kazuyuki Aihara

Phys. Rev. E **94**, 042217 — Published 20 October 2016

DOI: 10.1103/PhysRevE.94.042217

Possibility of short-terms probabilistic forecasts for large earthquakes making good use of the limitations of existing catalogues

Yoshito Hirata[1,*], Koji Iwayama[1,2], and Kazuyuki Aihara[1]

[1]Institute of Industrial Science, University of Tokyo,

4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

[2]FIRST, Aihara Innovative Mathematical Modelling Project,

Japan Science and Technology Agency,

Meguro-ku, Tokyo 153-8505, Japan

[*]yoshito@sat.t.u-tokyo.ac.jp

Abstract:

Earthquakes are quite hard to predict. One of the possible reasons can be the fact that the existing catalogues of past earthquakes are limited at most to the order of 100 years, while their characteristic time scale is sometimes greater than that time span. Here we rather use these limitations positively and characterize some large earthquake events

as abnormal events that are not included there. When we constructed probabilistic forecasts for large earthquakes in Japan based on similarity and difference to their past patterns, which we call known and unknown abnormalities respectively, our forecast achieved probabilistic gains of 5.7 and 2.4 against a time independent model for main shocks with the magnitudes of 7 or above. Moreover, the two abnormal conditions covered 70% of days whose maximum magnitude was 7 or above.

PACS number(s): 91.30.Ab, 05.45.Tp, 07.05.Kf

I. INTRODUCTION

The idea of plate tectonics [1] implies that the earthquake activity can be governed to some extent by dynamical laws. In addition, there are many mathematical models for earthquakes [2-4]. However, we cannot predict earthquakes in deterministic ways until now [5]. There also exist three empirical statistical laws related to the earthquake activity: The Omori-Utsu formula [6-8] describes the decay of aftershocks after their main shocks; The Gutenberg-Richter law [9] describes the relation between the magnitude and the number of earthquakes; it is also known that the hypocenters of earthquakes are located in a fractal manner [10]. By combining the Omori-Utsu formula with the Gutenberg-Richter law, we can construct forecasts [11, 12] for aftershocks that may follow main shocks. However, the predictability of the main shocks like the Tohoku-Oki earthquake [13-16] is highly limited partially because these gigantic events were not recorded in the existing catalogues.

In this paper, we rather use the property that gigantic events are mostly not included in the existing catalogues, for forecasting such events in the short-terms. We divided the time axis into time windows whose length is one day. Based on the similarity of the marked point process pattern of earthquakes on each day with those of the past time windows, we define two types of abnormal time windows: the known abnormal and the unknown abnormal. We call time windows as the unknown abnormal if we do not have similar time windows in the past; the detailed definitions are given below. In addition, we call time windows as the known abnormal if they are not the unknown abnormal and the similar time windows in the past are likely to have been followed by time windows containing an event with a large magnitude. To evaluate the similarity among time windows, we use the edit distance for marked point processes [17-19]. The edit distance was more powerful than the commonly used inter-event intervals [20-22] when we characterize marked point processes because the edit distance can evaluate the times, hypocenters and magnitudes of earthquakes, simultaneously. Previously, the edit distance was used to characterize the dynamics of foreign exchange markets [18, 23], classify aftershocks of earthquake activity [17], and characterize the response of foreign exchange markets to the earthquake activity [24].

This paper is organized as follows: In Section II, we introduce the datasets we analyzed. In Section III, we explain the methods we analyzed the datasets. In Section IV, we present the results. In Section V, we discuss the results and conclude this paper.

II. ANALYZED DATSETS

We prepared the dataset around Japan by selecting earthquakes whose longitudes were between 125 $^{\circ}$E and 150 $^{\circ}$E, and the latitudes were between 25 $^{\circ}$N and 48 $^{\circ}$N. The time period was from 1 January 2000 to 30 June 2011. The events whose magnitudes are greater than or equal to 4 were selected because they could be detected completely without being missed [25].

In addition, we also analyzed the dataset of earthquake activity around New Zealand. The longitudes were between 164 °E and 136 °W, the latitudes were between 15 °S and 50 °S, and the times were between 1 January 1990 and 7 February 2011. We treated the earthquakes whose magnitudes were greater than or equal to 3.5.

III. METHODS

A. Size of time window

1. Backgrounds

We sample a time window every day to reconstruct the earthquake activity. We decide the length of the time window by extending the idea of delay coordinates in dynamical systems.

First, we review delay coordinates for a time series with a fixed sampling interval. Let L be an l-dimensional manifold. Suppose that a dynamical system $f: L \to L$ is given by $u_{t+1} = f(u_t)$. We also have an observation function $s_t = g(u_t)$ through which we observe the dynamical system. Then, delay coordinates for a time series $\{s_t\}$ are defined as $G_d(u_t) = (s_t, s_{t+1}, \dots, s_{t+d-1})$. The constant d is called the embedding dimension. Takens [26] showed that when $d \geq 2l + 1$, delay coordinates $G_d(u_t)$ are generally one-to-one with a state u_t. Later, Sauer *et al.* [27] replaced it by the condition of $\mathrm{d} > 2d_o$ with the box-counting dimension d_o of the attractor.

States u_t and u_{t+1}, and their delay coordinates $G_d(u_t)$ and $G_d(u_{t+1})$ are related by the following diagram:

$$
\begin{array}{ccc}
u_t & \overset{f}{\rightarrow} & u_{t+1} \\
\downarrow G_d & & \downarrow G_d \\
G_d(u_t) & \overset{\tilde{f}}{\rightarrow} & G_d(u_{t+1}).
\end{array}
$$

When u_t and $G_d(u_t)$ are one-to-one, then $G_d(\cdot)$ has an inverse. Thus, we can write $\tilde{f}$ by $\tilde{f} = G_d \circ f \circ {G_d}^{-1}$. Because $\tilde{f}$ is written only by the observed values, we can predict $G_d(u_{t+1})$ without knowing f itself directly.

2. How to decide the size of time window

(a) Using a time window instead of delay coordinates. We can formulate the problem for deciding the length of time window by extending the idea of delay coordinates mentioned above. Suppose that we have dynamics of $u_{w(t+1)} = f^w(u_{wt})$, where $f^w\colon L \to L$ defines mapping to the state the time of w later. We define an observation function $H_w\colon L \to W_w$ that returns a point process on the time interval $[wt, w(t+1))$, whose set is written by W_w. Then, the states u_{wt} and $u_{w(t+1)}$, and their corresponding time windows $H_w(u_{wt})$ and $H_w(u_{w(t+1)})$ of point processes are related by the following diagram:

$$\begin{array}{ccc} u_{wt} & \overset{f^w}{\rightarrow} & u_{w(t+1)} \\ \downarrow H_w & & \downarrow H_w \\ H_w(u_{wt}) & \overset{\tilde{f}^w}{\rightarrow} & H_w\left(u_{w(t+1)}\right). \end{array}$$

Then, if the window size w is large enough such that H_w has a unique inverse, then we can write $\tilde{f}^w$ by $\tilde{f}^w = H_w \circ f^w \circ {H_w}^{-1}$. Therefore, we can predict $H_w(u_{w(t+1)})$ from $H_w(u_{wt})$ if w is sufficiently large. The window size for a point process is the notion similar to the embedding dimension for a time series with a fixed sampling interval. We call the method for sampling the time intervals $[wt, w(t+1))$ as the Uniformly

Sampling Window method. The Uniformly Sampling Window method was previously applied to the datasets of foreign exchange markets [18, 23, 28] and neurons [29-31].

(b) Calculation of edit distances for marked point processes. We formulate the prediction and evaluate its goodness by an edit distance for a marked point process. In this edit distance, we evaluate how much it costs to edit one marked point process to the other by deletion, insertion, and/or shift of events. We assign a cost of 1 for deleting or inserting an event. We also assign a cost proportional to the time and the values shifted when we shift an event. In all the examples, we normalized marks so that the standard deviation for each mark is the same as the standard deviation of inter-event intervals.

To calculate edit distances for marked point processes more efficiently, we employ the following method.

As mentioned above, the edit distance between two marked point processes $H_w(u)$ and $H_w(u')$ is defined as the minimum total cost to edit $H_w(u)$ to $H_w(u')$. Let $h_i(u)$ denote the ith event of the marked point process $H_w(u)$ and $|H_w(u)|$ denote the number of events contained in $H_w(u)$. Without loss of generality, we assume that the number of events in the marked point process $H_w(u)$ is smaller than or equal to that in $H_w(u')$. We consider a set of $|H_w(u)|$ pairs of events in two marked point processes $C = \{(h_i(u), h_j(u'))\} \subset H_w(u) \times H_w(u')$, where any event cannot be included in multiple different pairs. For a pair of events $(h_i(u), h_j(u'))$, if the cost of shift of $h_i(u)$ to $h_j(u')$ is larger than 2, we should choose deletion of $h_i(u)$ from and insertion of $h_j(u')$ into $H_w(u)$ to minimize the total cost to edit rather than shift of $h_i(u)$ to $h_j(u')$. Otherwise, we should shift $h_i(u)$ to $h_j(u')$. Finally, we insert all of remaining $|H_w(u')| - |H_w(u)|$ events of $H_w(u')$ into $H_w(u)$ to complete editing. Thus, the edit distance for marked point process can be calculated by

$$\delta(H_w(u), H_w(u')) = \min_C \sum_{(h_i(u), h_j(u')) \in C} \min\left(2, \sum_k \lambda_k \left|h_{ik}(u) - h_{jk}(u')\right|\right) + |H_w(u')| - |H_w(u)|, \quad (1)$$

where $h_{ik}(u)$ means the kth mark of $h_i(u)$ and λ_k denotes the coefficient of shift of the kth mark. This definition of the edit distance can be represented by the complete bipartite graph $(H_w(u) \cup D, H_w(u'), (H_w(u) \cup D) \times H_w(u'))$ (Fig. 1), where vertices are events and all events of $H_w(u)$ connect to all events of $H_w(u')$. Here, D denotes the set of dummy vertices whose edges to vertices of $H_w(u')$ mean insertion of corresponding events. Edges from the vertices of events of $H_w(u)$ have the costs of shift or insertion and deletion, that is, the first term of Eq. (1). On the other hand, those from dummy nodes have costs of 1. Then, the minimum-cost perfect matching in this bipartite graph provides us the editing which minimizes the cost and the edit distance between two marked point processes (Fig. 1B). The minimum-cost perfect matching can be solved in a polynomial-time. Thus, we can calculate the edit distance for marked point processes in a polynomial time. This minimum cost perfect matching was first used in Ref. [24] to calculate the edit distance for marked point processes. We need about 15 days to calculate all the distances for 4199 days to produce Table III, for example using a computer with 2 CPUs of 6-Core Intel Xenon (2.66GHz) and 64 GB memory.

(c) Nearest neighbor prediction. We find the closest match $H_w(u_{wc(t)})$ for the current time window $H_w(u_{wt})$ from the past part of point processes using the above edit distance $(\mathrm{c(t)} = \mathrm{argmin}_{\iota<t}\, \delta(H_w(u_{wt}), H_w(u_{w\iota})))$ and letting the following time window

$H_w(u_{w(c(t)+p)})$ as the prediction $H_w(u_{w(t+p)})$ for p windows ahead (see Fig. 2 for the illustration). The prediction error for p windows ahead can be evaluated as $\delta\left(H_w\left(u_{w(t+p)}\right), H_w(u_{w(c(t)+p)})\right)$.

(d) How to decide the window size. When we decide the window size w, we compare the above nearest neighbor prediction with the persistence prediction where we let the current time window $H_w(u_{wt})$ as the prediction for p windows ahead $H_w(u_{w(t+p)})$ between $1 \leq p \leq 5$ (See Fig. 2 for the illustration). Namely, letting $2T_w$ the total number of windows, we minimize $e_w = \frac{\sum_{t=\lfloor T_w/2 \rfloor}^{T_w-6} \sum_{p=1}^{5} \delta\left(H_w\left(u_{w(t+p)}\right), H_w\left(u_{w(c(t)+p)}\right)\right)}{\sum_{t=\lfloor T_w/2 \rfloor}^{T_w-6} \sum_{p=1}^{5} \delta\left(H_w\left(u_{w(t+p)}\right), H_w(u_{wt})\right)}$. When e_w is smaller than 1, the nearest neighbor prediction is better than the persistence prediction. We eventually use the first half of the dataset to decide the window size. The candidates of w are selected in such a way that the smaller window length is multiplied by a number between 1.5 and 3 to obtain the next window size.

We evaluate the nearest neighbor prediction by using the second half of the dataset.

3. Examples

We here show four examples for choosing the length of the time window.

The first example is an integrate-and-fire neuron [32] driven by the Lorenz model [33]. The equations can be written as

$$\dot{x} = -10(x - y), \quad (2)$$

$$\dot{y} = -xz + 28x - y, \quad (3)$$

$$\dot{z} = xy - \frac{8}{3}z, \quad (4)$$

$$X(t) = 20(0.025x(t) + 1),$$

$$\int_{t_a}^{t_{a+1}} X(t)dt = 1,$$

where t_a represents the time for the ath firing, and t_0 is arbitrarily chosen.

The result presented in Fig. 3 shows that the window size of 0.1 is optimal in this case. Thus, we used this window size to predict the following part of time series. We found that the nearest neighbor prediction was better than the persistence prediction in 1247 out of 1499 time windows and these predictions were tie in 5 out of the remaining 252 time windows (the winning rate: 0.83, p-value < 0.001). Thus, by choosing the window size optimally, we could predict the future more efficiently than the persistence prediction.

The second example is a local maxima series [34] of the Rössler model [35]. The equations for the Rössler model are written as

$$\dot{x} = -(y + z),$$

$$\dot{y} = x + 0.36y,$$

$$\dot{z} = 0.4 + z(x - 4.5).$$

We observed the series of x, extracted local maxima, and recorded their times and values to generate a series of the marked point process.

The result presented in Fig. 4 shows that the window size of 5 is optimal. When we used 5 for the window size, we could predict the next window with the nearest neighbor prediction better than with the persistence prediction in 924 out of 999 time windows and these predictions are tie in 3 time windows out of the remaining 75 time windows (the winning rate 0.92, p-value < 0.0001). Thus, in this example of the Rössler model, the window size was appropriately chosen such that the nearest neighbor prediction is effective.

In the third example, we used the Lorenz model of Eqs. (2)-(4) and generated a marked point process by extracting times and values of local maxima for the upper lobe of x as well as those of local minima for the lower lobe of x (see Fig. 5). We integrated Eqs. (2)-(4) for the duration of 10000 after throwing away the initial transient.

We set 2.5 for the minimal size of time window because we need to span the time range of length 2.21 to include at least three events in any of the time windows. When we looked for an optimal time window by the normalized prediction errors, we found that 2.5 was optimal (Fig. 6). When we used the second half to evaluate the prediction performance, we found that the nearest neighbor prediction won the persistence prediction in 1825 out of 1999 time windows. The winning rate was 0.93 (the p-value < 0.001).

As the fourth example, we even restricted ourselves to use only the events on the upper lobe (Fig. 7) and applied the same analysis using the Lorenz model of Eqs. (2)-(4). Then, still the time window of length 2.5 was chosen as the optimal (Fig. 8). This length of time window accompanied with the nearest neighbor prediction has the superior prediction skill to the persistence prediction because the nearest neighbor prediction won the persistence prediction in 1730 time windows out of 1999 time windows predicted and tied in 3 time windows. Thus, the winning rate for the nearest neighbor prediction was 0.87 (the p-value < 0.001).

When we applied the above way of choosing the length of the time window to the series of earthquakes around Japan, we chose 1 day as the optimal length for the window size (see Fig. 9).

4. Notes

For a time series with a fixed sampling interval, the false nearest neighbor method [36] is a standard technique for deciding the embedding dimension. However, we cannot extend the false nearest neighbor method for deciding the length of time window of point processes because the edit distance between neighbors jumps when we follow its change along the time axis. Instead, we used the approach of using prediction errors [37-39] to evaluate the length of time windows.

B. Converting the edit distances for the marked point process to the ones with long-term memory

Because the earthquakes may depend on their long history, we convert the edit distances for the marked point process obtained using the 1 day window above into the ones that can retain the long-term memory. For this sake, we use the Fréchet product metric [40] as follows:

$$\tilde{\delta}(u_{wg}, u_{wt}) = \sum_{\beta=0}^{99-\max[0,99-g]} \zeta^{\beta} \delta\left(H_w(u_{w(g-\beta)}), H_w(u_{w(t-\beta)})\right), \tag{5}$$

where $g < t$ and $0 \leq \zeta < 1$.

This definition is similar to the Bielecki metric [41], but the direction of the sum is the opposite: our metric takes the sum toward the past. We chose $\zeta = 0.5$.

C. Weighted average for magnitude

Suppose that $\tau_1(t), \tau_2(t), \tau_3(t), \tau_4(t)$, and $\tau_5(t)$ be a set of time indices for the 5 nearest neighbors for day t. Let $m(t)$ be the maximum magnitude for day t. Then, we took the weighted average [42] of the maximum magnitudes for the next days in the following way:

$$\frac{\sum_{\kappa=1}^{5} m(\tau_\kappa(t)+1)\exp\left(-0.01\tilde{\delta}(u_{w\tau_\kappa(t)}, u_{wt})\right)}{\sum_{\kappa=1}^{5} \exp\left(-0.01\tilde{\delta}(u_{w\tau_\kappa(t)}, u_{wt})\right)}. \qquad (6)$$

If the current time window is not classified to unknown abnormal conditions, and Expression (6) is larger than the threshold magnitude M, we declare that the current time window belongs to known abnormal conditions.

D. Optimization of forecast parameters

We optimize forecast parameters, namely the threshold magnitude M and the threshold percent tail q for the median distances of each time window with all the time windows in the database, by maximizing the product of modified odds ratios for aftershocks and main shocks with magnitude greater than or equal to 7. To define the quantity for the optimization more precisely, let $R = \{R_{\rho,\sigma}\}$ be a 3 by 3 matrix containing the numbers of days satisfying specified two classifications in the second quarter of the given dataset: The first classification is which condition a day belongs to, the normal ($\rho = 1$), the known abnormal ($\rho = 2$), or the unknown abnormal ($\rho = 3$) defined using the first quarter of the given dataset; the second classification is the outcome of the following day, whether the following day does not have an earthquake with magnitude greater than or equal to 7 ($\sigma = 1$), it has an aftershock with magnitude greater than or equal to 7 ($\sigma = 2$), or it has a main shock with magnitude greater than or equal to 7 ($\sigma = 3$). Then we define the modified odds ratio for the known abnormal conditions by

$$\frac{(R_{1,1}+R_{1,2})R_{2,3}}{(R_{2,1}+R_{2,2}+0.5)(R_{1,3}+0.5)}, \qquad (7)$$

and the modified odds ratio for the unknown abnormal conditions by

$$\frac{(R_{1,1}+R_{1,2})R_{3,3}}{(R_{3,1}+R_{3,2}+0.5)(R_{1,3}+0.5)}. \qquad (8)$$

We maximized the product of Expressions (7) and (8) in terms of M and q by the grid search. Here M was chosen from 50, 60, 70, 80, 90, 95, 98, 99, and 99.5% points of the weighted average of magnitudes for the following day in the first quarter of the dataset, and q was chosen from 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95%.

E. Time-independent model

A time-independent model is often used to evaluate the predictability for earthquake activity [43, 44]. Here, we use the following time-independent model: Suppose that we predict over N days, within which $\bar{N}_2$ and $\bar{N}_3$ days have the maximum magnitude over or equal to the used magnitude threshold $\bar{m}$ achieved by an aftershock and a main shock, respectively. Then, the probability that the maximum magnitude was greater than or equal to the used threshold is $\bar{N}/N$ where $\bar{N} = \bar{N}_2 + \bar{N}_3$. Let T be a matrix representing the results shown, for example, in Table III. Namely, $T_{1,1} = 1624, T_{1,2} = 1, T_{1,3} = 2, T_{2,1} = 135, T_{2,2} = 0, T_{2,3} = 3, T_{3,1} = 330, T_{3,2} = 1$ and $T_{3,3} = 3$. Then, the probabilistic gains for the aftershocks and the main shocks of known abnormal conditions against the time—independent model were $\left(\frac{T_{2,3}}{T_{2,1}+T_{2,2}+T_{2,3}}\right)/(\bar{N}_2/N)$ and $\left(\frac{T_{2,3}}{T_{2,1}+T_{2,2}+T_{2,3}}\right)/(\bar{N}_3/N)$, and those of unknown abnormal conditions were $\left(\frac{T_{3,3}}{T_{3,1}+T_{3,2}+T_{3,3}}\right)/(\bar{N}_2/N)$ and $\left(\frac{T_{3,3}}{T_{3,1}+T_{3,2}+T_{3,3}}\right)/(\bar{N}_3/N)$, respectively.

F. Epidemic Type Aftershock Sequence (ETAS) model

The ETAS model is a stochastic point process model for earthquake occurrences [45-47], and has been recently recognized as a standard model for probabilistic earthquake forecasting. Although various types of the ETAS models have been proposed, the central assumption of the model is that each earthquake with any magnitude can trigger its own aftershocks, the number of which depends on its magnitude. Here we consider the hierarchical space-time ETAS (Hist-ETAS) model [46], which is realistic enough to reproduce the actual seismicity. For the Hist-ETAS model, the occurrence rate $\theta(t, \mathbf{r})$ at time t and location $\mathbf{r}$ given an occurrence history $I_t = \{(t_v, M_v, \mathbf{r}_v) \mid t_v < t, M_c \leq M_v\}$ is expressed as

$$\theta(t, \mathbf{r}|I_t) = \mu(\mathbf{r}) + \sum_{t_v<t} \frac{K(\mathbf{r})}{(t-t_v+\gamma)^{\mathbf{p}(\mathbf{r})}} \left[\frac{(\mathbf{r}-\mathbf{r}_v)S_v(\mathbf{r}-\mathbf{r}_v)'}{\exp\left(\alpha(\mathbf{r})(M_v-M_c)\right)} + \Delta\right]^{-\mathbf{q}(\mathbf{r})}, \qquad (9)$$

where the first and the second terms, respectively, represent the background activity and the triggering effect from the preceding events. Some parameters $\mu(\mathbf{r})$, $K(\mathbf{r})$, $\mathbf{p}(\mathbf{r})$, $\alpha(\mathbf{r})$, and $\mathbf{q}(\mathbf{r})$ are assumed to be location-dependent, modeled as the piecewise function with the Delaunay triangulation for the set of the locations $\{\mathbf{r}_v\}$ of the past events. The Hist-ETAS model is estimated based on objective Bayesian estimation [46].

The synthetic seismicity is simulated by using the estimated Hist-ETAS model. Here the magnitude of each simulated earthquake is randomly sampled from the following Gutenberg-Richter formula $\breve{m}(M)$ of magnitude distribution:

$$\breve{m}(M) \propto 10^{-bM}, \tag{10}$$

where we set $b = 0.9$. The distribution of the simulated events is plotted in Fig. 10.

IV. RESULTS

A. Predicting earthquake patterns

First, we compared the nearest neighbor prediction with the persistence prediction to consider a possibility that if the earthquake activity of some day in the past is similar to that of the current window, the earthquake activity similar to that following the "some day" occurs for the next day, namely there is possibly an underlying deterministic law behind the earthquake activity. In the nearest neighbor prediction, we found the closest match from the past part of the series in the edit distance or the Fréchet product metric

[40] with the exponentially decaying weights (see Section IIIB) and let the next window as the prediction for the following window [23] (see Fig. 2 for the schematic illustration; see also Section IIIA). In the persistence prediction, we let the current window as the prediction for the following window [23] (see Fig. 2 for the schematic illustration; see Section IIIA as well). The prediction errors were also evaluated by using the edit distance. At the beginning, we only used the numbers of earthquakes within a day to predict the future. But, in this case, the nearest neighbor prediction is worse than the persistence prediction (see Tables I and II, respectively). On the other hand, if we represented a series of earthquakes as a point process and increased the number of marks, namely if we used additional information such as times, magnitudes, and places (longitudes, latitudes, and depths) of the earthquakes, the predictability has improved (see Tables I and II). Therefore, the earthquake activity can be better predicted by the nearest neighbor prediction because if the earthquake activity for the most recent day is similar to that in some past day, the earthquake activity for the next day becomes similar to that for the day after the “some past” day, and that the marks provide useful information for the prediction for the next day activity.

In what follows, we use the Fréchet product metric for forecasting.

B. Forecasting large earthquakes

Therefore, we have constructed a probabilistic forecast of the maximum magnitude for the next day by combining the concept of the nearest neighbor prediction with the concept of the known and unknown abnormalities. Because the dataset we have was too short compared with the characteristic time scale of the earthquake activity, we divided time windows into 3 categories: normal conditions, known abnormal conditions, and unknown abnormal conditions. If the median distance of the current time window with all the time windows in the database is larger than the q% tail for the median distance of each time window with all the time windows in the database, then we classified the current time window as an unknown abnormal condition. If the current time window is not classified into the unknown abnormal conditions and its spatio-temporal 5 nearest neighbors in the edit distance has the weighted average magnitude (Section IIIC) greater than or equal to M in their following next days, then the current time window is classified to a known abnormal condition. Otherwise the current time window is classified to a normal condition. We used the first quarter as the database and the second quarter for optimizing the parameters M and q to 4.12 and 90 so that we can achieve a large product of two modified odds ratios for forecasting main shocks and aftershocks (see Section IIID for details. We used the criteria of Ref. [47] to distinguish main shocks from aftershocks). Then we evaluated the probabilistic forecast using the third and fourth quarters.

The results presented in Table III show that the probabilistic forecast achieved the probabilistic gains of (2.17/0.38=) 5.7 and (0.90/0.38=) 2.4 for the known abnormal conditions and the unknown abnormal conditions against a time independent model under which we assume that there is no time-dependence and we evaluated the probability that the maximum magnitude was more than or equal to 7 by the empirical histogram without any conditioning (see Section IIIE). When the two abnormal conditions were combined, the probabilistic gain was 3.1. In addition, the 95% confidence interval for the odds ratio of the abnormal conditions was [1.85, 49.02]. This forecast could provide some warning by either the known abnormality or the unknown abnormality for 70% cases when the maximum magnitude for the next day was more than or equal to 7. For example, the day before the Tohoku-Oki earthquake was classified as a day for the unknown abnormal condition. The forecasted earthquakes were located widely all over Japan but the days classified to the unknown abnormal and normal conditions were concentrated around the east of Japan's main island, while the days classified to the known abnormal conditions were far from the center of Japan (see Figs. 11 and 12).

Marks such as magnitudes, longitudes, latitudes, and depths of earthquakes helped to improve the accuracy of the probabilistic forecast, especially because the lower bound for the 95% confidence interval of the odds ratio for the known and unknown abnormal conditions became highest if we added all these pieces of information as marks when obtaining the edit distances (see Tables III-VII). Therefore, it is more informative to forecast based on more information. In addition, even if we evaluate only the time period before the Tohoku-Oki earthquake, the probabilistic forecasts had the forecast skill (Table VIII).

Our probabilistic forecast did not work well in predicting the artificial earthquake series generated from the Epidemic-type aftershock sequence (ETAS) model [46] (the Hist-ETAS model, see Section IIIF and Table IX), which is a current standard statistical model for earthquake occurrence. The probabilistic gains for the known abnormal and the unknown abnormal conditions were 1.0 and 0.8 against the time independent model.

When we reduced the time periods for the database and the parameter optimization into a half, the probabilistic gain for the abnormal conditions against the time-independent model became 1.0 (see Table X), which was smaller than when we used the first and the second quarters of the whole dataset as the database and the optimization (see Table III). Thus we expect that we may increase the accuracy of the short-terms probabilistic forecast by accumulating longer-term observations of earthquakes.

C. Results on the earthquake activity around New Zealand

We obtained the similar results by analyzing the dataset of New Zealand. While we analyzed the dataset of New Zealand, we also used the time windows of 1 day for the analysis. When we predicted the pattern of the earthquake activity using the nearest neighbor prediction, the winning rate was the best when all the information of the marks was used (See Tables XI and XII). When we constructed the probabilistic forecasts for the earthquake activity around New Zealand (see Tables XIII-XVII), we found that the lower bound of the 95% confidence interval for the odds ratio for the abnormality conditions was the highest when we used all the marks, namely, the times, magnitudes, longitudes, latitudes, and depths of earthquakes (See Table XVII) compared with the cases where we only used the partial information (see Tables XIII-XVI). Here we used the different magnitude thresholds for the dataset of New Zealand from those for the dataset of Japan because the earthquake activity around Japan was more active. The forecasted days are illustrated in Figs. 13 and 14. In this case, the known abnormal days were spread along the islands of New Zealand, while the days classified to normal were concentrated on north-east of New Zealand.

V. DISCUSSIONS

The proposed method is based on the embedding theorem for non-uniformly sampled data generated from a dynamical system [48]. Actually, when we convert the distance matrices obtained for Figs. 3, 4, 6, and 8 to recurrence plots [49, 50], making them continuous by the method of Ref. [51], and convert back to evenly sampled time series by the method of Refs. [52,53], the reconstructed time series look similar to the original time series (Figs. 15-18; their correlation coefficients were 0.8865, 0.6157, 0.3369, and 0.3826, respectively). These figures mean that even if we only have a series of events, we have sufficient information for reconstructing the underlying dynamics. When we visualize the exponentially weighted distance matrix for the earthquake data around Japan by the multidimensional scaling [54] without assuming the continuity respect to the time axis, we found that the absolute values for the top three components correlated well with the maximal magnitudes for the next days (see Figure 19; the correlation coefficients were 0.1996, 0.2214, and 0.2257, respectively). Because the similar observation holds for the case of New Zealand as well (see Fig. 20; the correlation coefficients were 0.0797, 0.0682, 0.0644, respectively, while their p-values were less than 0.001).

Judging from Figures 19 and 20, the catalogs we used contain substantial information for forecasting large earthquake events. But, we are not sure whether the catalogs miss some other important pieces of information for such a purpose. Thus, we should check, in our future research, whether or not we should include other pieces of information in the catalogs to improve our forecasts further, while based on the embedding theorem by Ref. [48], the other pieces of information might not be necessary because we could reproduce them from a general series of events.

Because our prediction is based on the embedding theorem by Ref. [48], we expect that a time window should be longer than one day if we use a magnitude threshold M greater than 4. This point should be also examined in our future research.

The novel part of the proposed method is that we consider abnormal conditions that are not recorded in the detailed existing catalogues. Thus, our retrospective probabilistic forecasts could achieve high probabilistic gains as demonstrated in Table III. It should be also important, on the other hand, to keep accumulating data of earthquake catalogues for further improvement of forecasts. We hope that the proposed probabilistic forecast will help to not only start preparing countermeasures for the large earthquakes before they will actually happen, but also establish short-terms insurances for the casualties and damage that the forecasted large earthquakes might cause.

ACKNOWLEDGEMENTS

We deeply appreciate Dr. Takahiro Omi, Prof. Yosihiko Ogata and Prof. Kunihiko Shimazaki for the stimulating discussions. We also appreciate Prof. Ogata for providing the simulation of the ETAS model used in this study. In addition, Y.H. appreciates Dr. Andrea Taroni and Dr. Aditya Riadi Gusman for their helping him to revise the abstract. We thank the Japan Meteorological Agency and GeoNet Project for providing the datasets used in this study. We also thank the National Oceanic and Atmospheric

Administration for making their datasets available on the web through the National Geophysical Data Center so that we could plot Figs. 10, 12 and 14. This research was supported by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through its "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP), and CREST, JST. Y. H. was also partially supported by Grants-in-Aid for Young Scientists (B) Grant No. 23700261 from the JSPS.

REFERENCES

1. B. Isacks, J. Oliver, and L. R. Sykes, *J. Geophys. Res.* **73**, 5855 (1968).
2. J. Huang and D. L. Turcotte, *Nature* **348**, 234 (1990).
3. J. McCloskey and C. J. Bean, *Science* **266**, 410 (1994).
4. D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics,* Cambridge University Press, Cambridge UK, 1997.
5. Y. Y. Kagan, *Geophys. J. Int.* **131**, 505 (1997).
6. F. Omori, *J. Coll. Sci. Imp. Univ. Tokyo* **7**, 111 (1894).

7. T. Utsu, *Geophys. Mag.* **30**, 521 (1961).

8. T. Utsu, Y. Ogata, and R. S. Matsu'ura, *J. Phys. Earth* **43**, 1 (1995).

9. B. Gutenberg and C. F. Richter, *Bull. Seism. Soc. Am.* **34**, 185 (1994).

10. R. F. Smalley, J.-L. Chatelain, D. L. Turcotte, and . R. Prévot, *Bull. Seism. Soc. Am.* **77**, 1368 (1987).

11. P. A. Reasenberg and L. M. Jones, *Science* **243**, 1173 (1989).

12. T. Omi, Y. Ogata, Y. Hirata, and K. Aihara, *Sci. Rep.* **3**, 2218 (2013).

13. M. Sato *et al., Science* **332**, 1395 (2011).

14. M. Simons *et al., Science* **332**, 1421 (2011).

15. S. Ozawa *et al.*, *Nature* **475**, 373 (2011).

16. R. J. Geller, *Nature* **472**, 407 (2011).

17. F. P. Schoenberg and K. E. Tranbarger, *Environmetrics* **19**, 271 (2008).

18. S. Suzuki, Y. Hirata, and K. Aihara, *Int. J. Bifurcat. Chaos* **20**, 3699 (2010).

19. Y. Hirata and K. Aihara, *Chaos* **25**, 123117 (2015).

20. T. Matcharashvili, T. Chelidze, Z. Javakhishvili, and E. Ghlonti, *Computers & Geosciences* **28**, 693 (2002).

21. A. C. Iliopoulos and G. P. Pavlos, *Int. J. Bifurcat. Chaos* **20**, 2071 (2010).

22. G. Molchan and L. Romashkova, *Pure Appl. Geophys.* **171**, 2339 (2014).

23. Y. Hirata and K. Aihara, *Physica A*, **391**, 760 (2012).

24. S. Nakano, Y. Hirata, K. Iwayama, and K. Aihara, *Physica A* **419**, 203 (2015).

25. K. Z. Nanjo *et al.*, *B. Seismol. Soc. Am.* **100**, 3261 (2010).

26. F. Takens, *Springer Lect. Notes Math.* **898**, 366 (1981).

27. T. Sauer, J. A. Yorke, and M. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).

28. K. Iwayama, Y. Hirata, and K. Takahashi, K. Watanabe, K. Aihara, and H. Suzuki, *Sci. Rep.* **2**, 423 (2012).

29. Y. Hirata and K. Aihara, *J. Neurosci. Meth.* **183**, 277 (2009).

30. R. G. Andrzejak and T. Kreuz, *EPL* **96**, 50012 (2011).

31. Y. Hirata, E. J. Lang, and K. Aihara, “Analyzing multiple spike trains using distance measures and recurrence plots,” In N. Kasabov ed., *Springer Handbook of Bio- and Neruoinformatics*, Springer, Heidelberg Germary, 2013.

32. T. Sauer, *Phys. Rev. Lett.* **72**, 3811 (1994).

33. E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).

34. N. Yabuta, and T. Ikeguchi, "Prediction of high-dimensional multivariate information as an amplitude-event dynamical system," In: *Proc. 2007 Int. Symp. Nonlinear Theory and Its Applications*, Vancouver, Canada, September 16-19, 2007, pp.188-191, 2007.

35. O. E. Rössler, *Phys. Lett.* **57A**, 397 (1976).

36. M. B. Kennel, R. Brown, H. D. I. Abarbanel, *Phys. Rev. A* **45**, 3403 (1992).

37. K. Judd and A. Mees, *Physica D* **120**, 273 (1998).

38. Y. Hirata, H. Suzuki, and K. Aihara, *Phys. Rev. E* **74**, 026202 (2006).

39. L. C. Uzal, G. L. Grinblat, and P. F. Verdes, *Phys. Rev. E* **84**, 016223 (2011).

40. M. M. Deza and E. Deza, *Encyclopedia of Distances.* 2nd edition, Springer, Berlin Germany, 2013.

41. S. Shirali, *Math. Commun.* **15**, 139 (2010).

42. G. Sugihara and R. M. May, *Nature* **344**, 734 (1990).

43. R. Console, M. Murru, and G. Falcone, *Pure Appl. Geophys.* **167**, 693 (2010).

44. M. J. Werner, A. Helmstetter, D. D. Jackson, and Y. Y. Kagan, *Bull. Seism. Soc. Am.* **101**, 1630 (2011).

45. Y. Ogata, *Ann. Inst. Stat. Math.* **50**, 379 (1998).

46. Y. Ogata, K. Katsura, and M. Tanemura, *J. R. Stat. Soc. Ser. C Appl. Stat.* **52**, 499 (2003).

47. T. Omi, Y. Ogata, Y. Hirata, and K. Aihara, *J. Geophys. Res. Solid Earth* **120**, 2561 (2015).

48. J. P. Huke and D. S. Broomhead, *Nonlinearity* **20**, 2205 (2007).

49. J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, *Europhys. Lett.* **4**, 973 (1987).

50. N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, *Phys. Rep.* **438**, 237 (2007).

51. M. Tanio, Y. Hirata, and H. Suzuki, *Phys. Lett. A* **373**, 2031 (2009).

52. Y. Hirata, S. Horai, and K. Aihara, *Eur. Phys. J. Spec. Top.* **164**, 13-22 (2008).

53. Y. Hirata, M. Komuro, S. Horai, and K. Aihara, *Int. J. Bifurcat. Chaos* **25**, 1550168 (2015).

54. J. C. Gower, *Biometrika* **53**, 325 (1966).

TABLE I. Comparison of the winning rate for the nearest neighbor prediction using the edit distance against the persistence prediction in the case of Japan. We increased the amount of information we can use for the nearest neighbor prediction. Each winning rate was obtained by dividing the number of wins by the number of time windows within which the nearest neighbor prediction and the persistence prediction provided different predictions.

Used information	Winning rate
Number of events within a day only	0.461 (681/1478)
Times	0.503 (913/1817)
Times and magnitudes	0.533 (941/1764)
Times and places	0.556 (928/1670)
Times, magnitudes, and places	0.562 (930/1656)

TABLE II. Comparison of the winning rate for the nearest neighbor prediction using the Fréchet product metric against the persistence prediction in the case of Japan. We increased the number of information we can use for the nearest neighbor prediction. See the caption of Table I to find the definition for the winning rate.

Used information	Winning rate
Number of events within a day only	0.433 (698/1613)
Times	0.610 (1043/1711)
Times and magnitudes	0.811 (1326/1635)
Times and places	0.880 (1402/1593)
Times, magnitudes, and places	0.881 (1404/1594)

TABLE III. Experiment for forecasting days with large events around Japan using the times, places (longitudes, latitudes, and depths), and magnitudes of earthquakes. Max$M$ shows the maximum magnitude of earthquakes that happened within a day. Each integer shows the number of days for the corresponding classification. The 95% coincidence interval for the odds ratio of abnormal conditions was [1.85, 49.02]. The p-value obtained by the Fisher's exact test was 0.0018 when we grouped up the abnormal conditions.

	Max$M$<7	Max$M$≥7 & aftershock	Max$M$≥ 7 & main shock	Total
Normal	1624(99.82%)	1(0.06%)	2(0.12%)	1627(100.00%)
Known abnormal	135(98.31%)	0(0.00%)	3(2.17%)	138(100.00%)
Unknown abnormal	330(98.78%)	1(0.30%)	3(0.90%)	334(100.00%)
Total	2089(99.52%)	2(0.10%)	8(0.38%)	2099(100.00%)

TABLE IV. Experiment for forecasting days with large events around Japan using the number of earthquakes for each day only. See the caption of Table III to interpret this table. The probabilistic gains of known abnormal conditions and unknown abnormal conditions were 0.0 and 2.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of abnormal conditions was [1.09, 20.75]. The p-value was 0.019 when we grouped up the abnormal conditions.

	Max$M$<7	Max$M$≥7 & aftershock	Max$M$≥ 7 & main shock	Total
Normal	1726(99.71%)	0(0.00%)	5(0.29%)	1731(100.00%)
Known abnormal	42(100.00%)	0(0.00%)	0(0.00%)	42(100.00%)
Unknown abnormal	321(98.47%)	2(0.61%)	3(0.92%)	326(100.00%)
Total	2089(99.52%)	2(0.10%)	8(0.38%)	2099(100.00%)

TABLE V. Experiment for forecasting days with large events around Japan using the times of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown abnormal conditions were 1.5 and 0.9 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of abnormal conditions was [0.27, 13.05]. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M$<7	Max$M \geq 7$ & aftershock	Max$M \geq 7$ & main shock	Total
Normal	526(99.62%)	0(0.00%)	2(0.38%)	528(100.00%)
Known abnormal	357(99.44%)	0(0.00%)	2(0.56%)	359(100.00%)
Unknown abnormal	1206(99.50%)	2(0.17%)	4(0.33%)	1212(100.00%)
Total	2089(99.52%)	2(0.10%)	8(0.38%)	2099(100.00%)

TABLE VI. Experiment for forecasting days with large events around Japan using the times and magnitudes of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown abnormal conditions were 0.0 and 2.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.67, 13.81]. The p-value was 0.075 when we grouped up the abnormal conditions.

	Max$M$<7	Max$M\geq 7$ & aftershock	Max$M\geq 7$ & main shock	Total
Normal	1734(99.66%)	1(0.06%)	5(0.29%)	1740(100.00%)
Known abnormal	26(100.00%)	0(0.00%)	0(0.00%)	26(100.00%)
Unknown abnormal	329(98.80%)	1(0.30%)	3(0.90%)	333(100.00%)
Total	2089(99.52%)	2(0.10%)	8(0.38%)	2099(100.00%)

TABLE VII. Experiment for forecasting days with large events around Japan using the times and places (longitudes, latitudes and depths) of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown abnormal conditions were 0.8 and 2.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.04, ∞]. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M < 7$	Max$M \geq 7$ & aftershock	Max$M \geq 7$ & main shock	Total
Normal	35(100.00%)	0(0.00%)	0(0.00%)	35(100.00%)
Known abnormal	1724(99.62%)	1(0.06%)	5(0.29%)	1730(100.00%)
Unknown abnormal	330(98.80%)	1(0.30%)	3(0.90%)	334(100.0%)
Total	2089(99.52%)	2(0.10%)	8(0.38%)	2099(100.00%)

TABLE VIII. Experiment for forecasting days with large events around Japan evaluated up to the 4000th day after 1 January 2000 using the times, magnitudes, and places of earthquakes. This table is the same as Table III except that only 4000 days after January 2000 were considered. See the caption of Table III to interpret the results. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 5.6 and 2.5 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.94, 99.23]. The p-value was 0.029 when we grouped up the abnormal conditions.

	Max$M$ < 7	Max$M \geq 7$ & aftershock	Max$M \geq 7$ & main shock	Total
Normal	1604(99.88%)	0(0.00%)	2(0.12%)	1606(100.00%)
Known abnormal	135(98.54%)	0(0.00%)	2(1.46%)	137(100.00%)
Unknown abnormal	156(99.36%)	0(0.00%)	1(0.64%)	153(100.00%)
Total	1895(99.74%)	0(0.00%)	5(0.26%)	1900(100.00%)

TABLE IX. Forecasting experiment for dataset generated from the ETAS model. In this table, we used the times, magnitudes, and places of earthquakes. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown abnormal conditions were 1.0 and 0.8 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.08, 55.05]. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M$ < 7	Max$M$ ≥ 7 & aftershock	Max$M$ ≥ 7 & main shock	Total
Normal	863(99.88%)	0(0.00%)	1(0.12%)	864(100.00%)
Known abnormal	1102(99.91%)	0(0.00%)	1(0.09%)	1103(100.00%)
Unknown abnormal	1370(99.85%)	1(0.07%)	1(0.07%)	1372(100.00%)
Total	3335(99.88%)	1(0.03%)	3(0.09%)	3339(100.00%)

TABLE X. Experiment for forecasting days with large events around Japan using a small database. In this table, the conditions are the same as those of Table III except that the time periods for the database and the optimization were made half. Namely, we only used the second quarter to forecast the third and fourth quarters. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 0.8 and 2.8 against the time-independent model, respectively. When the two abnormal conditions were combined, the probabilistic gain was 1.0. Both the abnormal conditions covered all the days with earthquakes with the magnitudes greater than or equal to 7. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M$ < 7	Max$M$ ≥ 7 & aftershock	Max$M$ ≥ 7 & main shock	Total
Normal	21 (100.00%)	0 (0.00%)	0 (0.00%)	21 (100.00%)
Known abnormal	1882 (99.63%)	1 (0.05%)	6 (0.32%)	1889 (100.00%)
Unknown abnormal	186 (98.41%)	1 (0.53%)	2 (1.06%)	189 (100.00%)

Total	2089 (99.52%)	2 (0.10%)	8 (0.38%)	2099 (100.00%)

TABLE XI. Comparison of the winning rate for the nearest neighbor prediction using the edit distance against the persistence prediction in the case of New Zealand. We increased the number of information we can use for the nearest neighbor prediction. See the caption of Table I for the definition of the winning rate.

Used information	Winning rate
Number of events within a day only	0.467 (1410/3022)
Times	0.407 (1565/3849)
Times and magnitudes	0.428 (1648/3849)
Times and places	0.452 (1691/3743)
Times, magnitudes, and places	0.630 (2078/3300)

TABLE XII. Comparison of the winning rate for the nearest neighbor prediction using the Fréchet product metric against the persistence prediction in the case of New Zealand. We increased the number of information we can use for the nearest neighbor prediction. See the caption of Table 1 for the definition of the winning rate.

Used information	Winning rate
Number of events within a day only	0.483 (1488/3082)
Times	0.515 (1948/3786)
Times and magnitudes	0.601 (2180/3625)
Times and places	0.830 (2739/3302)
Times, magnitudes, and places	0.876 (2865/3271)

TABLE XIII. Experiment for forecasting days with large events around New Zealand using the number of earthquakes for each day only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 2.7 and 0.7 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.30, 2.43]. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M$ < 6	Max$M$ ≥ 6 & aftershock	Max$M$ ≥ 6 & main shock	Total
Normal	2856 (99.34%)	1 (0.03%)	18 (0.63%)	2875 (100.00%)
Known abnormal	60 (98.36%)	0 (0.00%)	1 (1.64%)	61 (100.00%)
Unknown abnormal	912 (99.45%)	1 (0.11%)	4 (0.44%)	917 (100.00%)
Total	3828 (99.35%)	2 (0.05%)	23 (0.60%)	3853 (100.00%)

TABLE XIV. Experiment for forecasting days with large events around New Zealand using the times of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 1.7 and 0.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.43, 3.05]. The p-value was 0.64 when we grouped up the abnormal conditions.

	Max$M < 6$	Max$M \geq 6$ & Aftershock	Max$M \geq 6$ & main shock	Total
Normal	2897 (99.38%)	1 (0.03%)	17 (0.58%)	2915 (100.00%)
Known abnormal	475 (98.96%)	0 (0.00%)	5 (1.04%)	480 (100.00%)
Unknown abnormal	456 (99.56%)	1 (0.22%)	1 (0.22%)	458 (100.00%)
Total	3828 (99.35%)	2 (0.05%)	23 (0.60%)	3853 (100.00%)

TABLE XV. Experiment for forecasting days with large events around New Zealand using the times and magnitudes of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown abnormal conditions were 2.0 and 1.3 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.81, 4.72]. The p-value was 0.10 when we grouped up the abnormal conditions.

	Max$M$ < 6	Max$M \geq 6$ & aftershock	Max$M \geq 6$ & main shock	Total
Normal	2309 (99.53%)	1 (0.04%)	10 (0.43%)	2320 (100.00%)
Known abnormal	244 (98.79%)	0 (0.00%)	3 (1.21%)	247 (100.00%)
Unknown abnormal	1275 (99.14%)	1 (0.08%)	10 (0.78%)	1286 (100.00%)
Total	3828 (99.35%)	2 (0.05%)	23 (0.60%)	3853 (100.00%)

TABLE XVI. Experiment for forecasting days with large events around New Zealand using the times and places of earthquakes only. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 1.1 and 0.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.02, ∞]. The p-value was 1 when we grouped up the abnormal conditions.

	Max$M < 6$	Max$M \geq 6$ & aftershock	Max$M \geq 6$ & main shock	Total
Normal	16 (100.00%)	0 (0.00%)	0 (0.00%)	16 (100.00%)
Known abnormal	3394 (99.33%)	1 (0.03%)	22 (0.64%)	3417 (100.00%)
Unknown abnormal	418 (99.62%)	1 (0.24%)	1 (0.24%)	420 (100.00%)
Total	3828 (99.35%)	2 (0.05%)	23 (0.60%)	3853 (100.00%)

TABLE XVII. Experiment for forecasting days with large events around New Zealand using the times, places, and magnitudes of earthquakes. See the caption of Table III to interpret this table. The probabilistic gains for the known abnormal conditions and the unknown large abnormal conditions were 1.9 and 0.4 against the time-independent model, respectively. The 95% confidence interval for the odds ratio of the abnormal conditions was [0.99, 6.13]. The p-value was 0.041 when we grouped up the abnormal conditions.

	Max$M$ < 6	Max$M \geq 6$ & aftershock	Max$M \geq 6$ & main shock	Total
Normal	2192 (99.59%)	1 (0.05%)	8 (0.36%)	2201 (100.00%)
Known abnormal	1213 (98.86%)	0 (0.00%)	14 (1.14%)	1227 (100.00%)
Unknown abnormal	423 (99.53%)	1 (0.24%)	1 (0.24%)	425 (100.00%)
Total	3828 (99.35%)	2 (0.05%)	23 (0.60%)	3853 (100.00%)

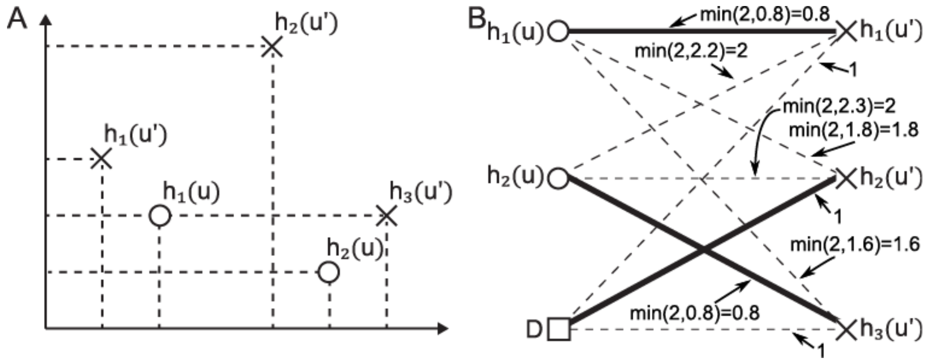

FIG. 1. Calculation of edit distances. (a) Examples of two marked point processes. Circles and crosses indicate event series of $H_w(u) = \{h_1(\mathrm{u}), h_2(\mathrm{u})\}$ and $H_w(u') = \{h_1(u'), h_2(u'), h_3(u')\}$, respectively. (b) The bipartite graph representing two marked point processes in (a). The black square indicates a dummy node. The numbers shown by the arrows represent the costs required for editing. If the numbers are 1, then it means deletion or insertion of events. If "min" is shown, then its first arguments are the costs for deletion and insertion of the corresponding events, and the second arguments are the costs required for a shift when $\lambda_1 = \lambda_2 = 0.4$. In addition, the numbers after "=" show the costs chosen after taking the minimum of the two costs. When a cost for a shift is larger than 2, which is the total cost of an insertion and a deletion, we choose the insertion and deletion of these events rather than the shift. For example, because the cost of shift from $h_2(u)$ to $h_2(u')$ is 2.3 and larger than 2, the cost of the edge between

these nodes is 2 which is required for the deletion of $h_2(u)$ and insertion of $h_2(u')$. Solid lines indicate the editing procedure which minimizes the total cost, while broken lines show the other potential edges. The edit distance between these two marked point processes is 2.6.

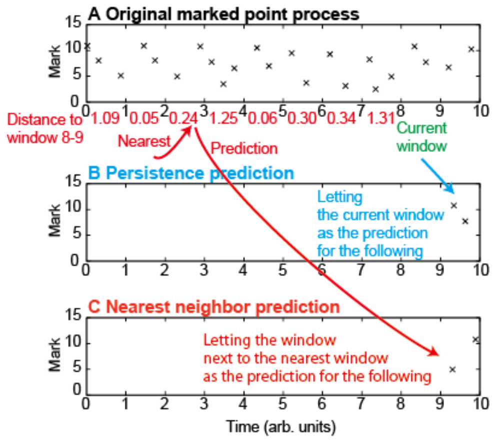

FIG. 2. (color online) Schematic diagram for persistence prediction and nearest neighbor prediction. Panel A corresponds to an original marked point process. Panel B shows how we generate persistence prediction, which is just letting the current window as the prediction for the following window. Panel C shows how we generate nearest neighbor prediction, by which we first find the closest match in the past and let the window next to the closest match as the prediction for the following window.

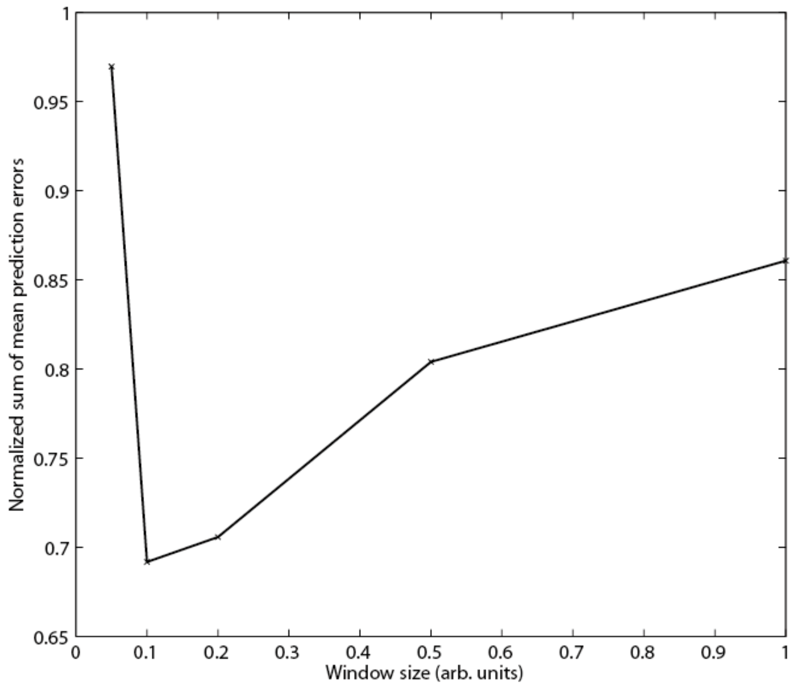

FIG. 3. The window size vs the normalized sum of mean prediction errors e_w, for the case of the integrate-and-fire neuron forced by the Lorenz model.

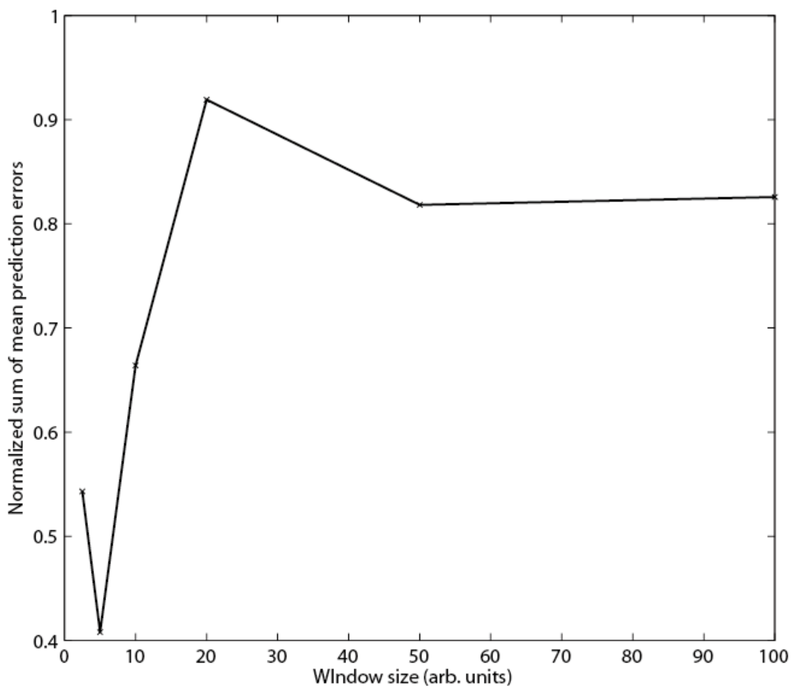

FIG. 4. The window size vs the normalized sum of mean prediction errors e_w, for the case of a local maxima series of the Rössler model.

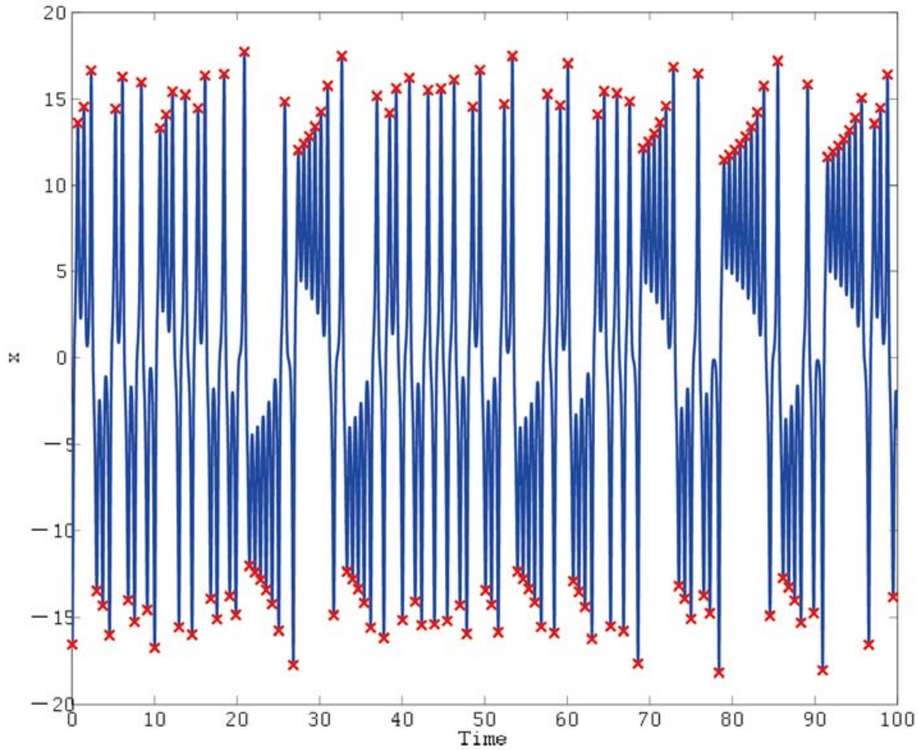

FIG. 5. The original time series (the solid line) of the Lorenz model and the extracted events (the crosses) for the third example. Here we extracted the times and values for the local maxima on the upper lobe as well as the times and values for the local minima on the lower lobe.

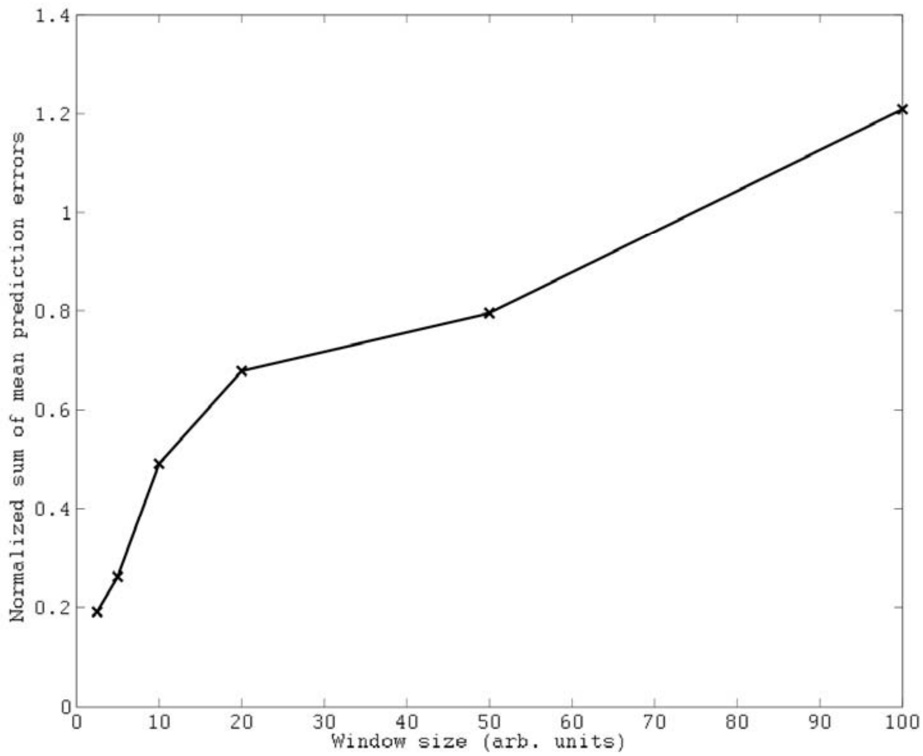

FIG. 6. The size of time window vs the normalized sum of mean prediction errors e_w, for the third example of the Lorenz model.

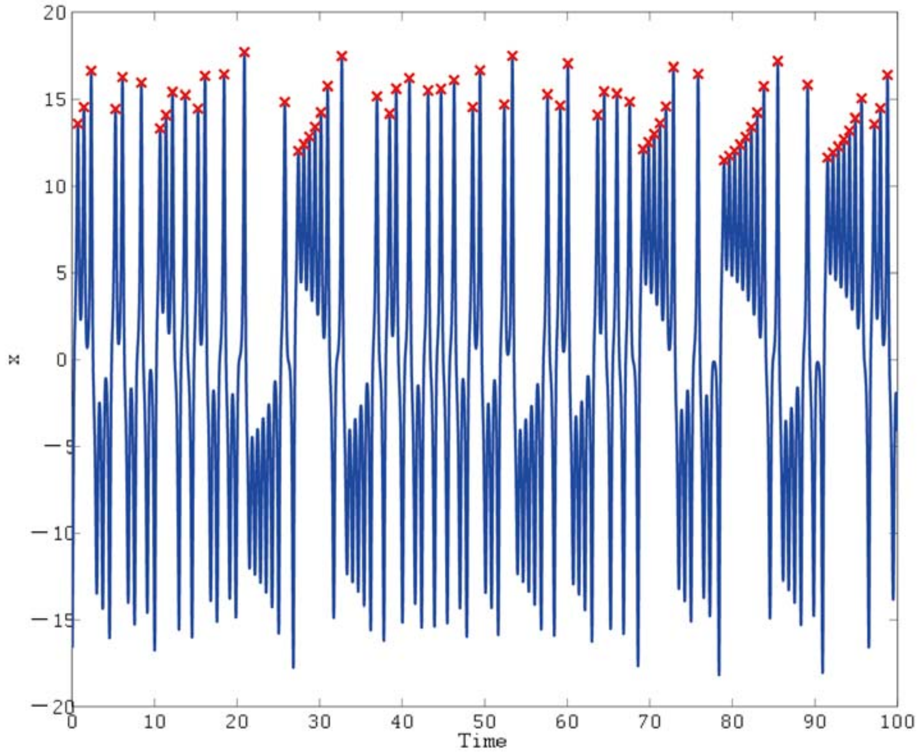

FIG. 7. The original time series (the solid line) of the Lorenz model and the extracted events (the crosses) for the fourth example. Here we extracted the times and values for the local maxima for the upper lobe.

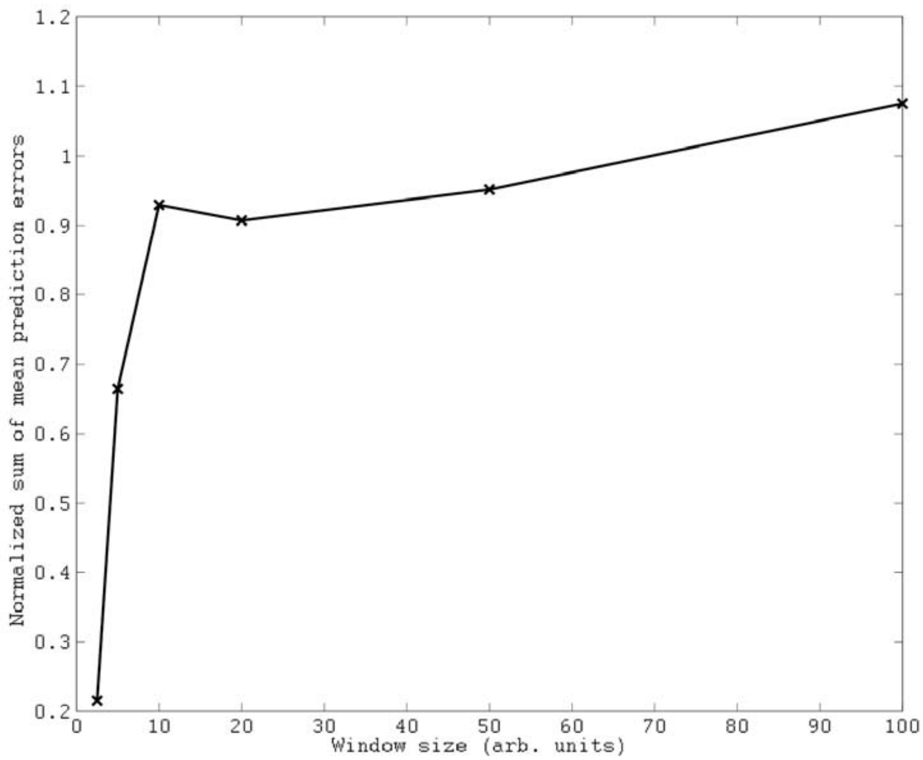

FIG. 8. The size of time window vs the normalized sum of mean prediction errors e_w, for the fourth example of the Lorenz model.

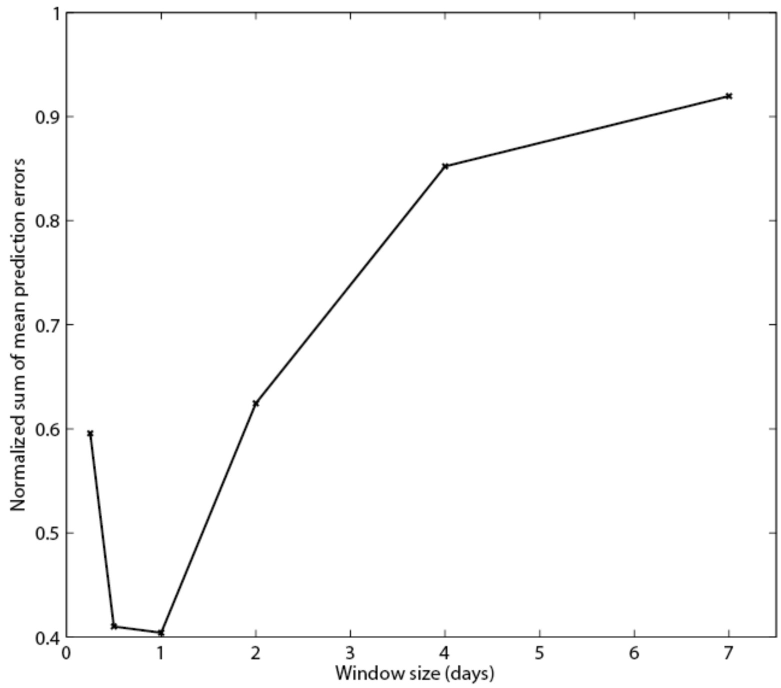

FIG. 9. The window size vs the normalized sum of mean prediction errors e_w, for the earthquake series around Japan.

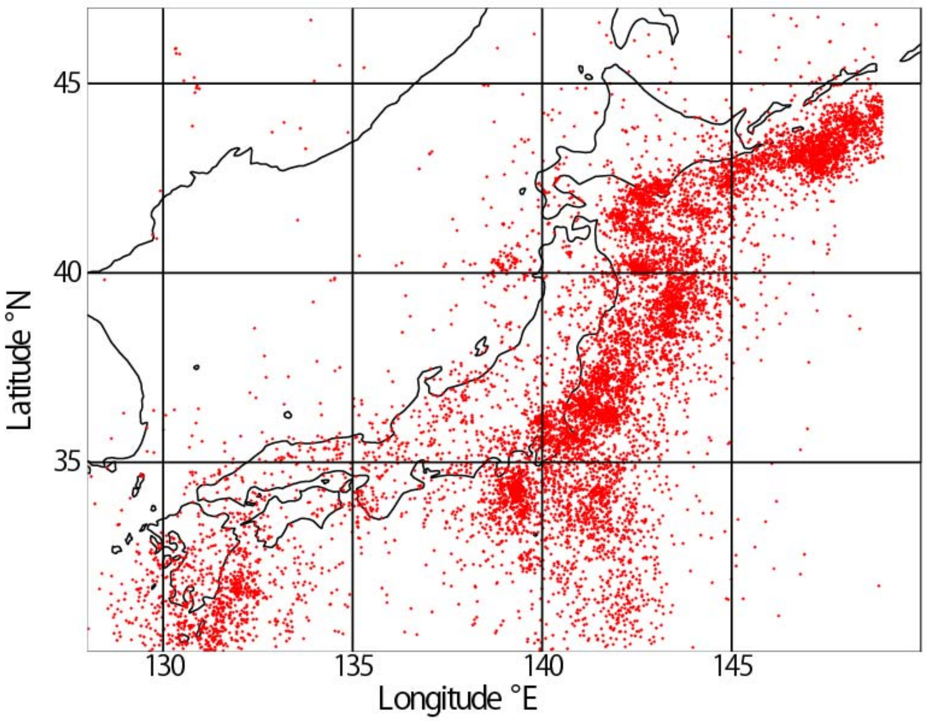

FIG. 10. (color online) Distribution of earthquakes in the ETAS model.

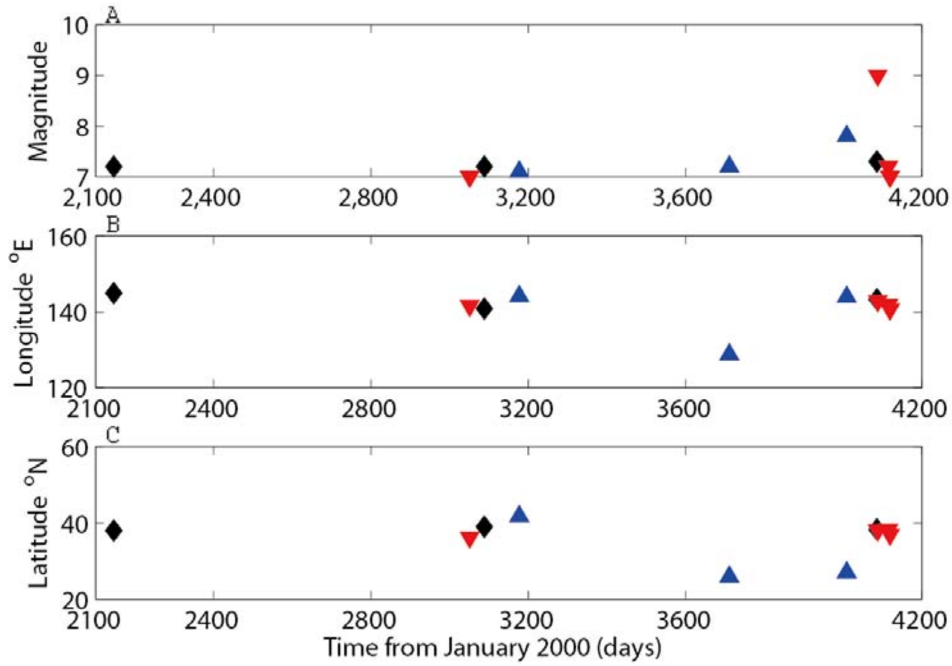

FIG. 11. (color online) Magnitudes, longitudes, and latitudes of earthquakes with the largest magnitudes that were 7 or above, depending on the probabilistic forecasts. Black diamonds, blue upper headed triangles and red down headed triangles show the days with forecasts with normal conditions, known abnormal conditions, and unknown abnormal conditions, respectively. Panels A, B, and C show magnitudes, longitudes, and latitudes, respectively.

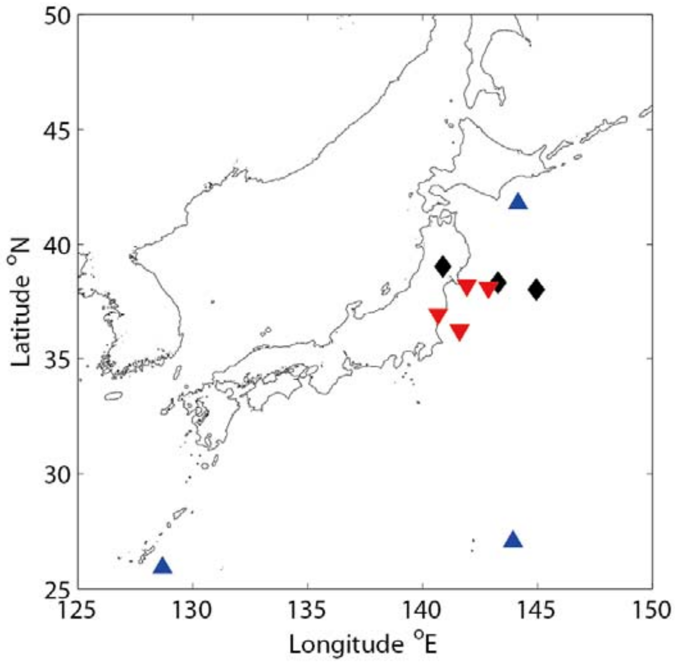

FIG. 12. (color online) Locations for earthquakes whose magnitudes were 7 or above. Black diamonds, blue upper headed triangles and red down headed triangles show the days with forecasts with normal conditions, known abnormal conditions, and unknown abnormal conditions, respectively

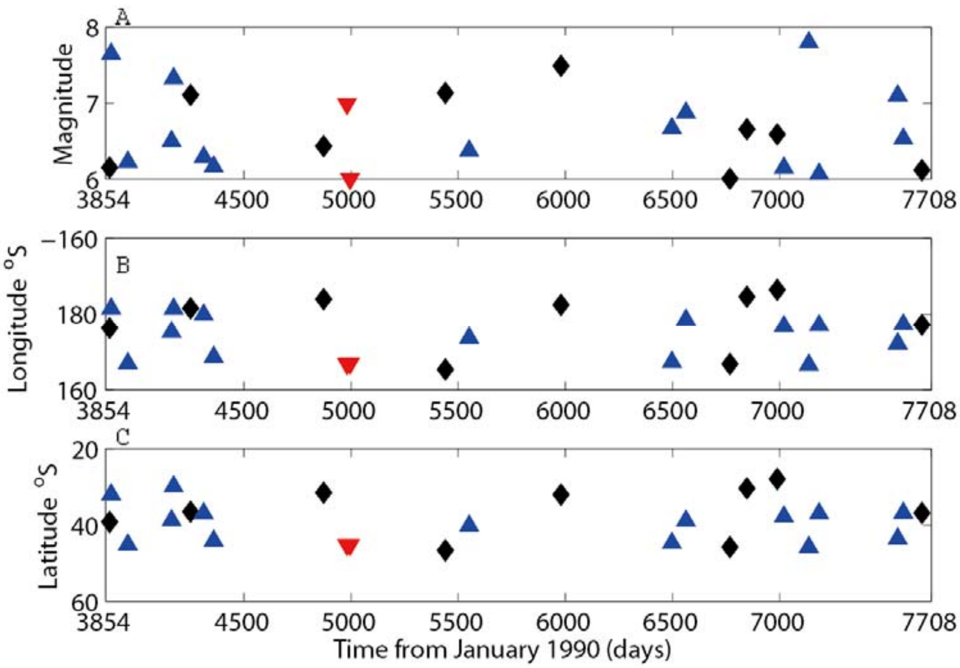

FIG. 13. (color online) Magnitudes, longitudes, and latitudes of earthquakes with the largest magnitudes that were 6 or above, depending on the probabilistic forecasts in the case of New Zealand. See the caption of Fig. 7 to interpret the results.

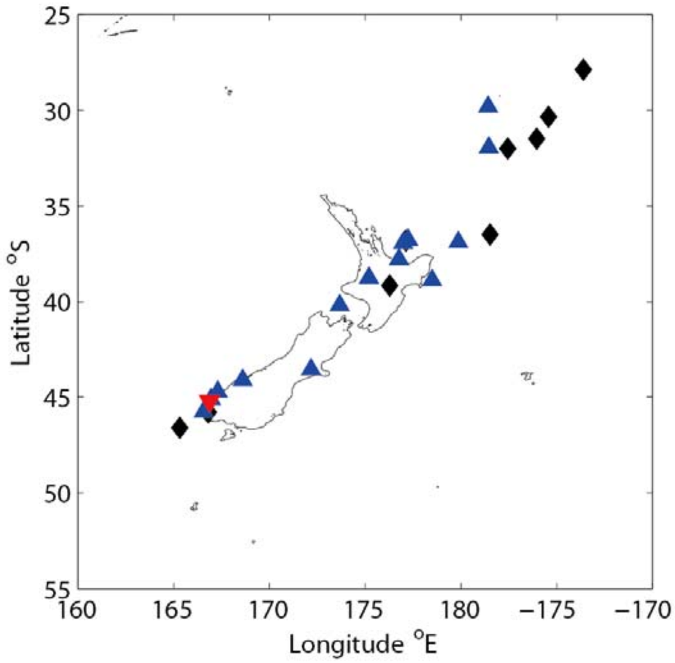

FIG. 14. (color online) Locations for earthquakes whose magnitudes were 6 or above in the case of New Zealand. See the caption of Fig. 8 to interpret the results.

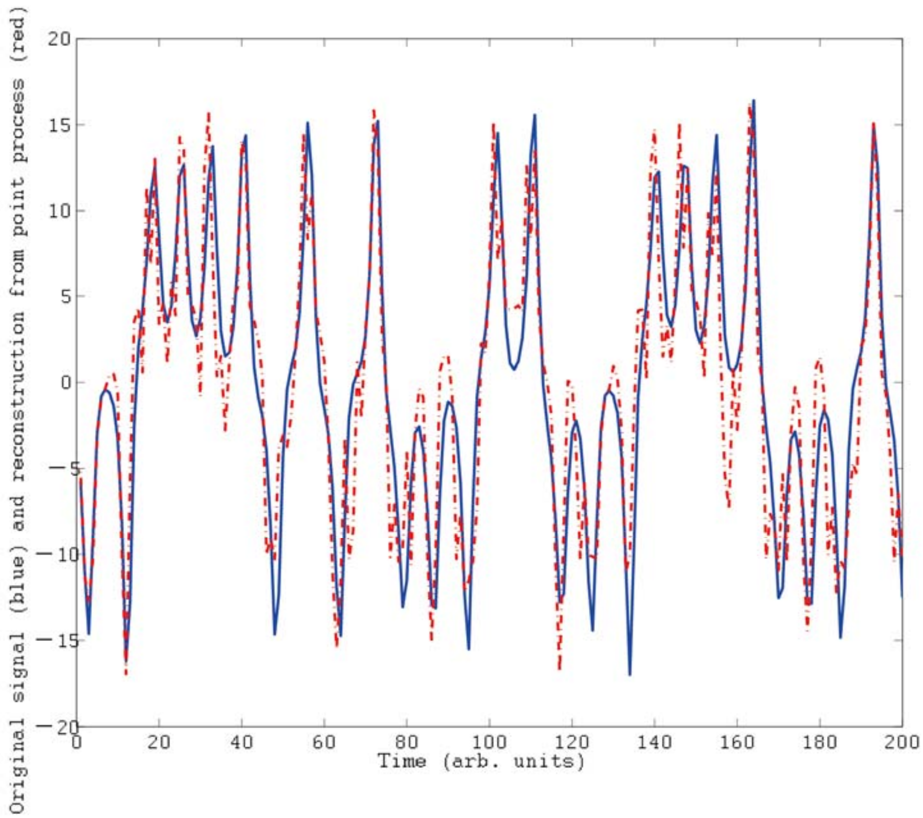

FIG. 15. The comparison between the original time series (the solid line) of the Lorenz model and its reconstruction (the dash dotted line) from the simple point process generated by the integrate-and-fire neuron. The reconstruction shown here is the first principal component obtained after drawing a recurrence plot by plotting points on 10% of places, making the plot continuous by the method of Ref. [51], and reproducing the original time series by the method of Refs. [52,53].

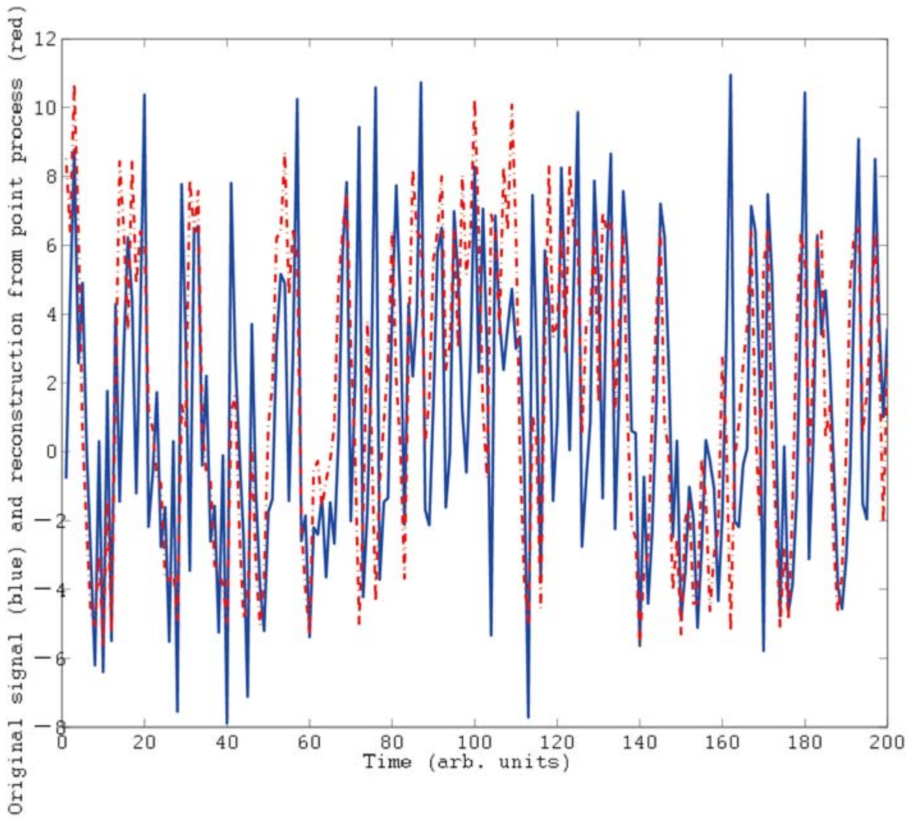

FIG. 16. The comparison between the subsampled time series (the solid line) of the Rössler model and its reconstruction (the dash dotted line) from the local maxima series of the Rössler model. Here, the reconstruction shown is the first principal component obtained after drawing a recurrence plot, making the plot continuous in time by the method of Ref. [51], and reproducing its time series by the method of Refs. [52.53].

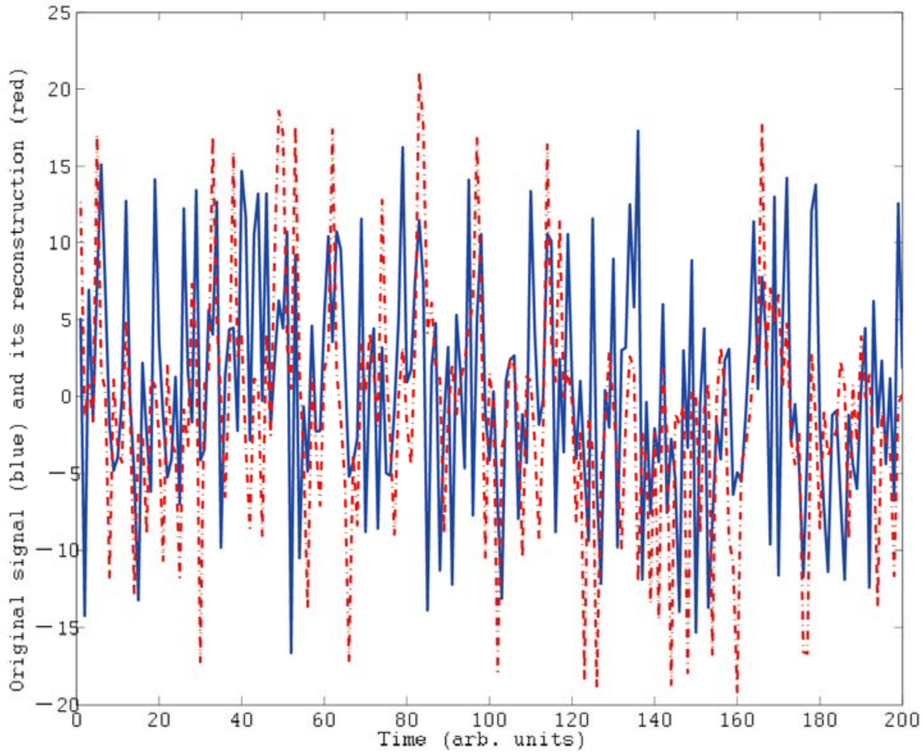

FIG. 17. The comparison between the subsampled time series (the solid line) of the Lorenz model and its reconstruction (the dash dotted line) for the third example. Here, the reconstruction shown is the third principal component after obtaining a recurrence plot, making the plot continuous by the method of Ref. [51], and reproducing the encoded time series by the method of Refs. [52,53].

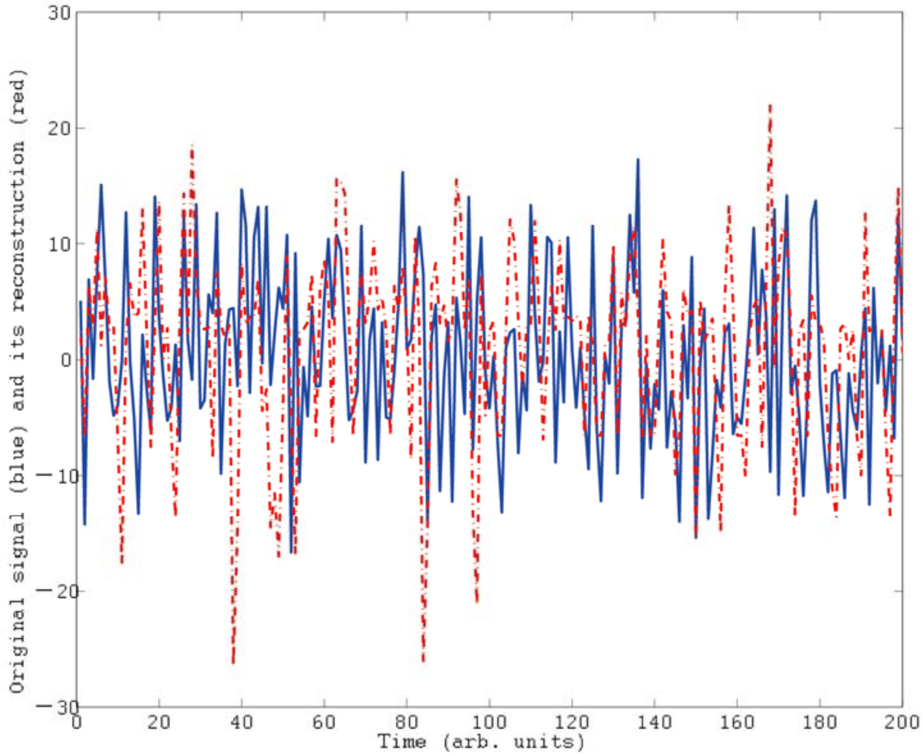

FIG. 18. The comparison between the subsampled dataset (the solid line) of the Lorenz model and its reconstruction (the dash dotted line) from the local maxima series of the upper lobe of x. Here, the reconstruction shown is the fourth principal component obtained after obtaining a recurrence plot, making the plot continuous by the method of Ref. [51], and reproducing the time series by the method of Refs. [52,53].

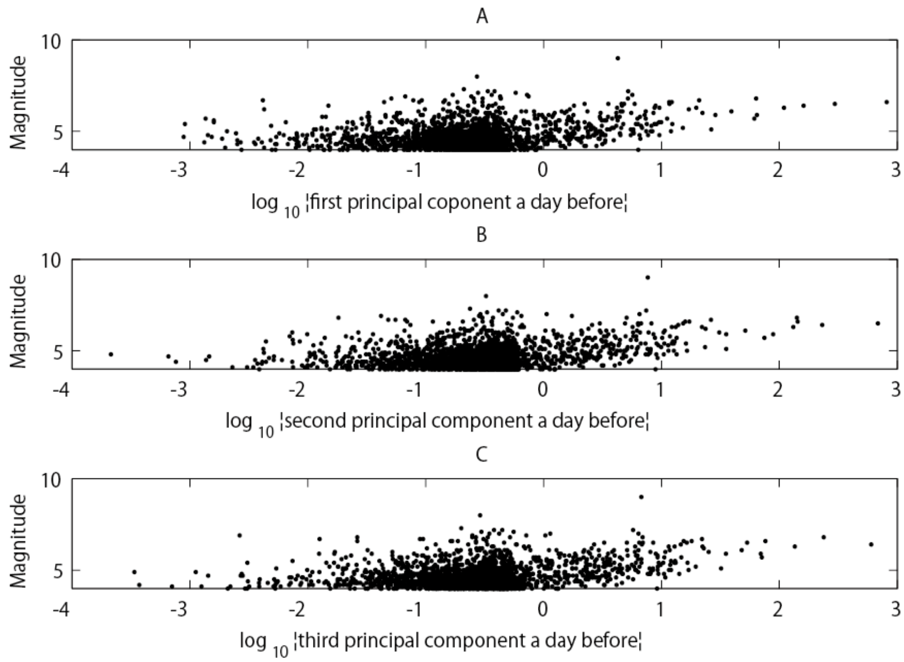

FIG. 19. The comparison between the top three components underlying the earthquake activity around Japan obtained by the edit distances and the maximum magnitude for the next days. We excluded the days when the maximum magnitudes were less than 4.

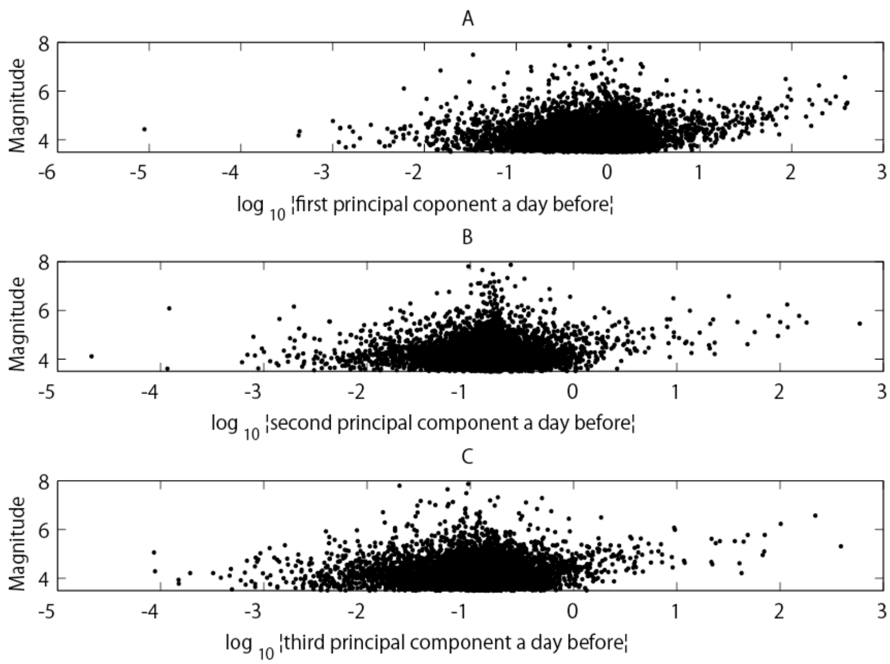

FIG. 20. The comparison between the top three components for the earthquake activity around New Zealand obtained by the edit distance, and the maximum magnitude for the next days. We excluded the days whose maximum magnitudes were less than 3.5.