

RESEARCH ARTICLE

Open Access



# Cluster-based foreshock discrimination model with flexible time horizon and mainshock magnitudes

Shunichi Nomura<sup>1\*</sup>  and Yosihiko Ogata<sup>2</sup>

## Abstract

Foreshock detection before mainshock occurrence is an important challenge limiting the short-term forecasts of large earthquakes. Various models for predicting mainshocks based on discrimination of foreshocks activity have been proposed, but many of them work in restricted scenarios and neglect foreshocks and mainshocks out of their scope. In disaster prevention, it is often necessary to change the forecast period and the magnitude of target mainshocks. This paper presents a cluster-based statistical discrimination of foreshocks which is applicable all over Japan and adjustable with respect to forecasting time span and mainshock magnitudes. Using the single-link clustering method, the model updates the expanding seismic clusters and determines in real time the probabilities that larger subsequent events will occur. The foreshock clusters and the others show different trends of certain feature statistics with respect to their magnitudes and spatiotemporal distances. Based on those features and the epicentral location, a nonlinear logistic regression model is used to evaluate the probabilities that growing seismic clusters are foreshocks that will trigger mainshocks within 30 days. The log of odds is estimated between the foreshock clusters and other clusters for respective feature values as nonlinear spline functions from a Japanese hypocenter catalog for the period 1926–1999. Based on the estimated odds functions, foreshock clusters tend to have smaller differences in their two largest magnitudes, shorter time durations, and slightly longer epicentral distances within the individual clusters. Given a potential foreshock cluster, its mainshock magnitude can be predicted by the Gutenberg–Richter law over the largest foreshock magnitude. The timing of mainshock occurrences from foreshocks is also predicted by multiplying the portion of mainshocks within a shorter span from those within 30 days by the evaluated foreshock probabilities. The predictive performance of our model is validated by the holdout method using a Japanese hypocenter catalog before and after 2000. The evaluated foreshock probabilities are roughly consistent with the actual portion of foreshocks in the validation catalog and could serve as an alert for large mainshocks.

**Keywords** Earthquake prediction, JMA catalog, Foreshock discrimination, Mainshock magnitude, Seismic clustering, Logistic regression, Spline functions

\*Correspondence:

Shunichi Nomura  
snomura5@aoni.waseda.jp

<sup>1</sup> Graduate School of Accountancy, Faculty of Commerce, Waseda University, 1-6-1 Nishiwaseda, Shinjuku, Tokyo 169-8050, Japan

<sup>2</sup> The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan



© The Author(s) 2023. **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/>.

## 1 Introduction

Foreshocks are promising clues for short-term forecasting of large mainshocks. Many studies have addressed certain features of foreshocks and predictability of mainshocks by foreshock detection (e.g., Papazachos 1974, 1975; Jones and Molnar 1979; Smith 1981; von Seggern et al. 1981; Xu et al. 1982, Jones 1985; Wong and Wyss 1985; Agnew and Jones 1991; Console et al. 1993; Savage and dePolo 1993; Maeda 1996). Ogata et al. (1995, 1996) composed and classified seismic clusters into foreshock, swarm, and mainshock–aftershock types. They discussed different trends in spatiotemporal distances and magnitude increments among those cluster types. The initially proposed methods are mainly based on anomalies to find potential foreshocks (e.g., Jones 1985; Console et al. 1993; Maeda 1996). Such anomaly-type methods provide high performances with optimized anomaly thresholds; however, they miss some portions of foreshocks that do not satisfy their anomaly classification.

Ogata et al. (1996) proposed a logistic regression model to evaluate the probabilities that seismic clusters will be the foreshock type or other clusters. This model can be applied to all earthquake clusters that may be foreshocks by defining the odds functions over all the feature spaces. Furthermore, Ogata and Katsura (2012) validated the predictive performance of the model by Ogata et al. (1996) in their catalog after its publication and proved that its probability forecast was as good as that shown in the original paper. The probability forecasts given by Ogata et al. (1996), however, were not robust for large cluster sizes, and it was recommended to use feature statistics of the first several events in the clusters (Ogata et al. 2018).

In this study, we modified certain aspects of the foreshock discrimination model proposed by Ogata et al. (1996). Specifically, (1) we do not classify foreshocks and swarms by magnitude differences; (2) we forecast mainshock magnitudes as well as their occurrences; and (3) we set the forecasting period at 30 days from the last event. Those modifications provide mainshock forecasting in more practical ways that are comparable to the Collaboratory for the Study of Earthquake Predictability (CSEP) tests (e.g., Tsuruoka et al. 2012).

## 2 Method

### 2.1 Dataset and seismicity clustering

We analyzed the Japan Meteorological Agency (JMA) catalog of  $M \geq 4$  in the region  $128\text{--}148^\circ$  E,  $30\text{--}46^\circ$  N as observed from January 1, 1926, to October 31, 2017, at a depth shallower than 100 km. To define mainshocks and their sub-events, we first compiled data of earthquake clusters from that catalog by the same procedure used

in Ogata et al. (1995, 1996); namely, the seismic clusters were constructed by the so-called single-link clustering (SLC) algorithm of Frohlich and Davis (1990). Specifically, earthquake pairs whose spatiotemporal distances were less than  $0.3^\circ$  (33.33 km or 30 days) were linked as belonging to the same cluster. The spatiotemporal distance of SLC is defined by  $\sqrt{(\Delta d)^2 + (c\Delta t)^2}$ , where  $\Delta d$  is the epicentral distance in degrees, and  $\Delta t$  is the difference of occurrence times in a day. For conversion between space and time distances, we set  $c=0.01^\circ/\text{day}$ , which is approximately equal to 1.111 km/day, as suggested by Davis and Frohlich (1991). In addition, to separate clusters between earthquakes in the shallow crustal zones and deep plate subduction zones in Japan, we only linked earthquake pairs whose depth difference was less than 70 km. The spatiotemporal distance threshold of  $0.3^\circ$ , or 30 days, for  $M \geq 4.0$  catalog, was determined by Ogata et al. (1995) to be consistent with the clusters of the algorithm based on the magnitudes of the mainshocks. These space–time parameters were deliberately determined by comparison with the Magnitude-based clustering (MBC) algorithm that determines a main shock before the formation of a cluster for several threshold magnitudes based on physically sensible empirical laws, whereas no magnitudes are used in the SLC. However, a drawback with the MBC is that it cannot be used for cluster identification until after the main shock is identified (while the SLC can). A drawback of the SLC is that it is very sensitive to the temporal change in detectability, while the MBC is not, as shown in Ogata et al. (1995). Furthermore, Ogata et al. (1995, 1996) use both algorithms to confirm the stability of the results. Another advantage of the SLC method is that cluster membership can be easily updated by adding new links between a new event and members of growing clusters.

After seismic clusters are constructed, the mainshock of each cluster is defined by the largest event of the cluster. In this study, we then defined a foreshock cluster based on an evolving seismic cluster whose mainshock occurred within 30 days from the last event. Thus, our goal was to discriminate in real time the foreshock clusters from the growing clusters.

Ogata et al. (1995) discriminated the foreshock-type cluster and swarm by the magnitude gap between the mainshock and its largest foreshock. However, this definition does not consider the case in which the largest foreshock may be preceded by smaller foreshocks; that is, the foreshock sequence itself is a foreshock- or swarm-type cluster, which may change on account of the mainshock occurrence. Therefore, our method does not distinguish foreshocks and swarms, and it forecasts mainshocks whose magnitudes exceed the largest magnitude of the current cluster.

We obtained data on 4,150 earthquake clusters, including more than two events. We divided the clusters into those ending before 1999 as the training dataset for our model, and those ending after 2000 as the validation dataset. The training dataset was thus comprised of 408 foreshock-type clusters out of 2916 clusters, and 166 foreshock-type clusters out of 1234 clusters comprised the validation dataset.

### 2.2 Feature extraction for foreshock discrimination

Many articles have reported properties of foreshocks in terms of magnitudes, spatiotemporal distances, and hypocenter locations. Some of them discuss a means of discriminating between foreshock clusters and other cluster types by using these properties. Nevertheless, Ogata et al. (1995) revealed statistics within an earthquake cluster that are useful for discriminating foreshocks. In this study, we considered the following feature statistics, which are similar to those used in the work of Ogata et al. (1996). We calculated them each time a cluster grew by adding a new event:

- $N$ : Size of the cluster
- $M_1$ : Largest magnitude of the cluster
- $\Delta M$ : Magnitude gap between the two largest magnitudes of the cluster
- $T$ : Time duration of the cluster (day)
- $D$ : Mean pairwise epicentral distance in the cluster (km)
- $(X, Y)$ : Mean longitude and latitude of the epicenter (degree)

We limited the cluster size  $N$  from 2 to 100 in our analysis. Before constructing the foreshock models, we examined empirical distributions of those feature statistics, as shown in Figs. 1, 2, and 3. The figures show the histograms and normalized cumulative distributions of  $\Delta M$ ,  $T$ , and  $D$  of foreshock and non-foreshock clusters under a fixed cluster size  $N$ , and a certain range of the largest magnitude  $M_1$ . The distributions of those three features shift as the cluster size  $N$  and largest magnitude  $M_1$  change. Thus, we can identify differences in their distributions between foreshock clusters and the others.

Figure 1 shows that the magnitude gaps  $\Delta M$  in foreshock clusters tend to be small relative to those in the other clusters. In Fig. 2, seismic clusters with shorter time spans are more likely to be foreshocks and vice versa. We should note that those trends in Figs. 1 and 2 become more apparent as the cluster sizes increase. With respect to the mean pairwise distances in the clusters, obvious trends cannot be observed in Fig. 3.

However, when we perform Wilcoxon rank sum tests (see, e.g., Section 4.1 of Hollander et al. 2013) for those features  $\Delta M$ ,  $T$ , and  $D$ , respectively, all the tests show that there are significant differences in distributions between foreshock and non-foreshock clusters with  $P$  values less than  $10^{-16}$ . The regional foreshock trend in Japan is discussed in the next section.

Because the number of clusters rapidly decreases as cluster size  $N$  increases, we transform  $N$  into  $Nc$  for use in our analysis by:  $Nc = \{\text{number of clusters whose sizes are less than } N\}$ . Then,  $Nc$  is approximately uniformly distributed through this normalization as seen from its cumulative distribution in Fig. 4. Time duration  $T$  is also transformed into its logarithm by  $Tl = \max(\log_{10} T, -4)$  as the normalization. Mean pairwise distance  $D$  is not transformed; it is used as it currently exists.

### 2.3 Learning foreshock probability by nonlinear logistic regression

In this section, we construct a statistical model to evaluate the foreshock probability of an evolving seismic cluster—that is, the probability that a mainshock will occur within 30 days from the last event in the cluster. We extract the feature statistics ( $Nc$ ,  $M_1$ ,  $\Delta M$ ,  $Tl$ ,  $D$ ,  $X$ ,  $Y$ ) that were introduced in the previous section from the evolving clusters. We evaluate their foreshock probabilities  $p(\text{foreshock} | Nc, M_1, \Delta M, Tl, D, X, Y)$  by using the following logistic regression model:

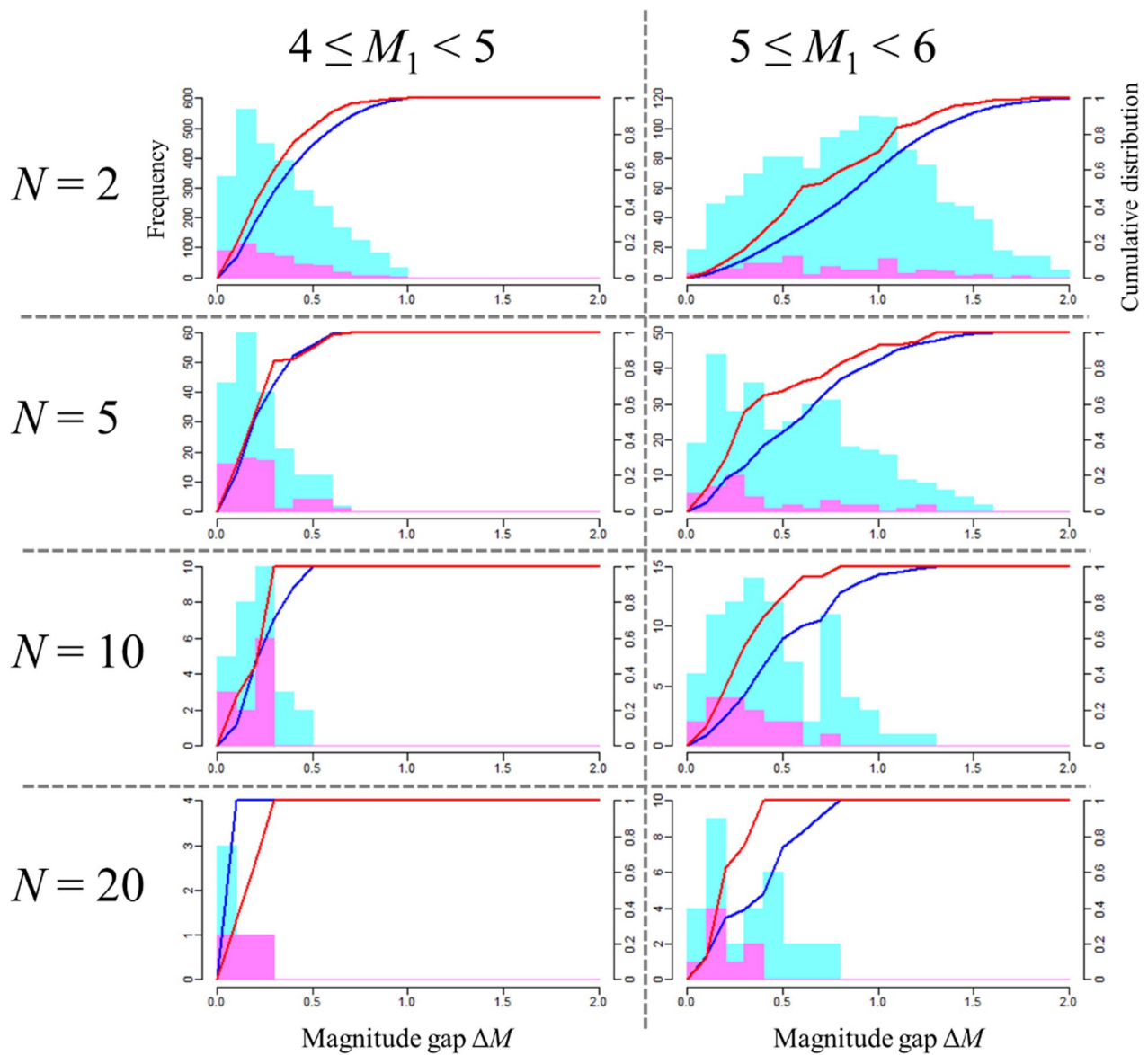
$$\begin{aligned} \text{logit } p(\text{foreshock} | Nc, M_1, \Delta M, Tl, D, X, Y) \\ = f_1(Nc, M_1, \Delta M) + f_2(Nc, M_1, Tl) \\ + f_3(Nc, M_1, D) + g(X, Y) + \varepsilon_{cl} \end{aligned} \quad (1)$$

where  $\text{logit } p = \log_e \{p/(1-p)\}$  is the logit or log odds of the probability to be a foreshock cluster. The functions  $f_1$ ,  $f_2$ ,  $f_3$  and  $g$  are nonlinear spline functions defined below and indicate the effect on log odds by their variables. The feature statistics in each function are dependent on each other and may influence the foreshock probabilities, as observed from Figs. 1, 2, and 3.

The first three functions,  $f_1, f_2, f_3$ , are the three-dimensional tensor products of cubic regression splines (see, e.g., Section 5.3.1 and 5.6.1 in Wood 2017) with 3 knots for each variable. In general, a three-dimensional tensor product  $f$  of cubic regression splines with 3 knots for each variable is defined by

$$f(x, y, z; \beta) = \sum_{i=1}^5 \sum_{j=1}^5 \sum_{k=1}^5 \beta_{ijk} f_i^x(x) f_j^y(y) f_k^z(z) \quad (2)$$

where  $f_i^x(x)$  is a basis function of the cubic regression spline for the variable  $x$  defined by



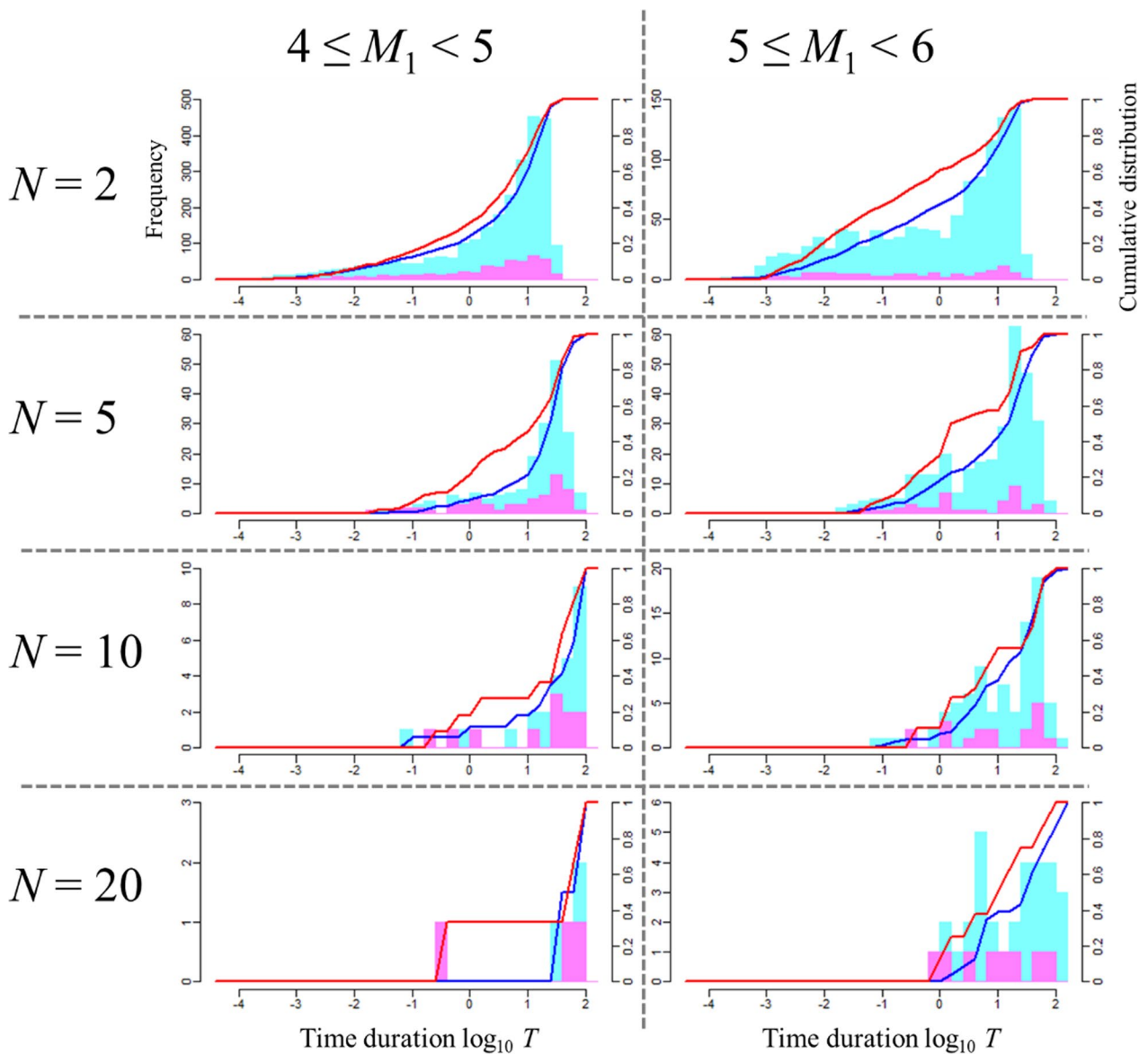
**Fig. 1** Histograms and cumulative relative frequencies of magnitude gap  $\Delta M$  of two largest events in foreshock (pink bars and red lines) and non-foreshock (light blue bars and blue lines) clusters.  $M_1$  indicates the largest foreshock magnitude

$$f_i^x(x) = \begin{cases} |x - x_i|^3 & i = 1, 2, 3 \\ x & i = 4 \\ 1 & i = 5 \end{cases} \quad (3)$$

$$\sum_{i=1}^3 \beta_{ijk} = \sum_{i=1}^3 \beta_{ijk} x_i = 0, \quad \text{for } j, k = 1, 2, 3, 4, 5 \quad (4)$$

and  $x_1, x_2, x_3$  are the knots allocated at the minimum, median, and maximum of the data  $x$ . The other spline functions  $f_j^y(y)$  and  $f_k^z(z)$  are defined in the same way as  $f_i^x(x)$ . The coefficients  $\beta = \{\beta_{ijk}; i, j, k = 1, 2, 3, 4, 5\}$  of the tensor spline  $f$  are constrained by the following conditions:

$$\sum_{j=1}^3 \beta_{ijk} = \sum_{j=1}^3 \beta_{ijk} y_j = 0, \quad \text{for } i, k = 1, 2, 3, 4, 5 \quad (5)$$



**Fig. 2** Histograms and cumulative relative frequencies of log time duration  $\log_{10} T$  of foreshock (pink bars and red lines) and non-foreshock (light blue bars and blue lines) clusters.  $M_1$  indicates the largest foreshock magnitude

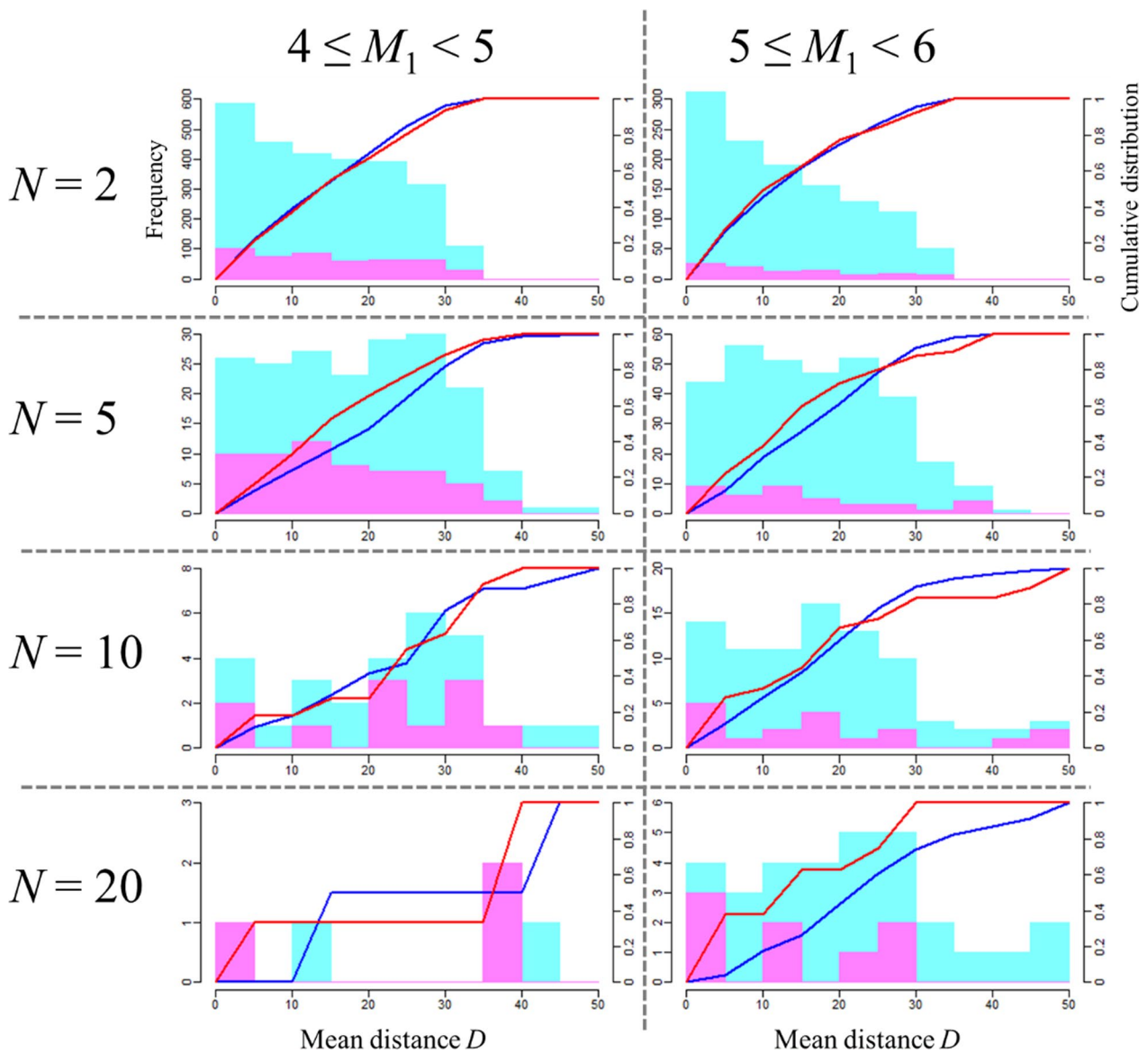
$$\sum_{k=1}^3 \beta_{ijk} = \sum_{k=1}^3 \beta_{ijk} z_k = 0, \quad \text{for } i, j = 1, 2, 3, 4, 5 \quad (6)$$

where  $y_1, y_2, y_3$  are the knots of  $f_j^y(y)$  and  $z_1, z_2, z_3$  are the knots of  $f_k^z(z)$ . These constraints contribute stable estimation of the spline function around its edge.

The fourth nonlinear function  $g$  is a thin plate regression spline function (see, e.g., Section 5.5.1 in Wood 2017) for isotropic two-dimensional features defined by

$$g(x, y) = \sum_{l=1}^L \phi_l r(x, y; x_l, y_l) \log_e r(x, y; x_l, y_l) + \phi_{L+1}x + \phi_{L+2}y + \phi_{L+3} \quad (7)$$

where  $r(x, y; x_l, y_l) = (x - x_l)^2 + (y - y_l)^2$ . The knots  $\{(x_l, y_l); l = 1, \dots, L = 20\}$  are randomly chosen from the epicenters in the training data. The coefficients  $\phi = \{\phi_l; l = 1, \dots, L + 3\}$  are constrained by



**Fig. 3** Histograms and cumulative relative frequencies of mean pairwise distance  $D$  in foreshock (pink bars and red lines) and non-foreshock (light blue bars and blue lines) clusters.  $M_1$  indicates the largest foreshock magnitude

$$\sum_{l=1}^L \phi_l = \sum_{l=1}^L \phi_l x_l = \sum_{l=1}^L \phi_l y_l = 0 \tag{8}$$

This constraint avoids the thin plate regression spline  $g$  from taking extreme values outside the distribution of epicenters in training data. This term is interpreted as the baseline log odds of foreshocks for the epicentral location.

The last term in Eq. (1),  $\varepsilon_{cl}$ , is a random effect for each cluster whose cluster size  $N$  is over ten, which has normally distributed cluster-specific values. Since there were not very many clusters whose sizes had grown over ten,

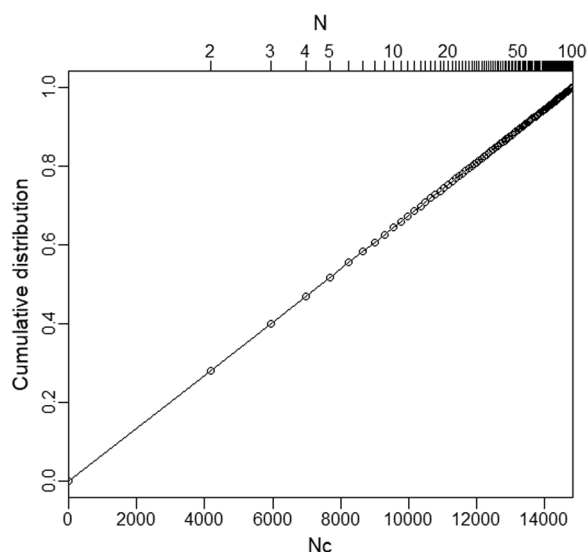
we introduced this term to avoid overfitting by easing the correlation within the individual clusters.

We estimated these terms by maximizing the penalized log-likelihood with penalties to the integral of second derivatives of the respective spline functions. The weights of these penalty functions were determined to minimize the Akaike’s information criterion (AIC; Akaike 1974).

### 3 Results and discussion

#### 3.1 Relative trends of foreshocks

We applied the proposed model to JMA catalog of  $M \geq 4$  in the region 128–148°E, 30–46°N as observed from



**Fig. 4** Cumulative distribution function of  $N$  (upper horizontal axis) and  $N_c$  (lower horizontal axis)

January 1, 1926, to December 31, 1999, at a depth shallower than 100 km and obtained relative trends of foreshocks as log odds functions shown in Figs. 5, 6, 7, and 8.

Figures 5, 6, and 7 depict the log of odds functions estimated from the JMA catalog for the period of January 1, 1926, to December 31, 1999. It is confirmed that the coincidence that the portion of foreshock clusters (red points) from others (blue points) increases as the log odds increases. As a common trend in these figures, the smaller the  $M_1$ , the higher the foreshock probability. This is because  $M_1$  is the lower limit of the mainshock magnitude to be predicted.

In Figs. 5 and 6, we can observe the same trends seen in Figs. 1 and 2, respectively. The foreshock probability is higher as the magnitude difference of the two largest events is smaller and the time interval is shorter. Figure 7 shows the trend of the log odds being slightly higher as the  $D$  is longer. When the  $N$  is larger, the trends shown in Figs. 5 and 6 become more apparent, while the trend shown in Fig. 7 becomes less apparent.

Figure 8 shows the regional change in log odds estimated from the same catalog. The log odds is relatively high in the eastern coast and offshore zones where the Pacific Plate and the Philippine Sea Plate are subducting, and it is relatively low in inland area except for the central Japan on the opposite side. The highest baseline log odds is marked off the coast of the Izu Peninsula around the middle area depicted in Fig. 8.

### 3.2 Prediction of mainshock magnitudes

When a foreshock cluster is detected, the magnitude of its mainshock may be predicted by examining the

relationship between the mainshocks and foreshocks. Figure 9 shows the cumulative counts of differences in magnitude between the mainshocks and the largest foreshocks in the JMA catalog for the period 1926–1999. The empirical cumulative distribution decreases exponentially along the gray line in Fig. 9. Therefore, we assume an exponential distribution for the conditional probability distribution of mainshock magnitude  $M_{\text{main}}$ , given the largest foreshock magnitude  $M_1$  as follows:

$$p(M_{\text{main}} > M_1 + m | \text{foreshock}, M_1) = 10^{-0.89m}, \quad m = 0.1, 0.2, \dots \quad (9)$$

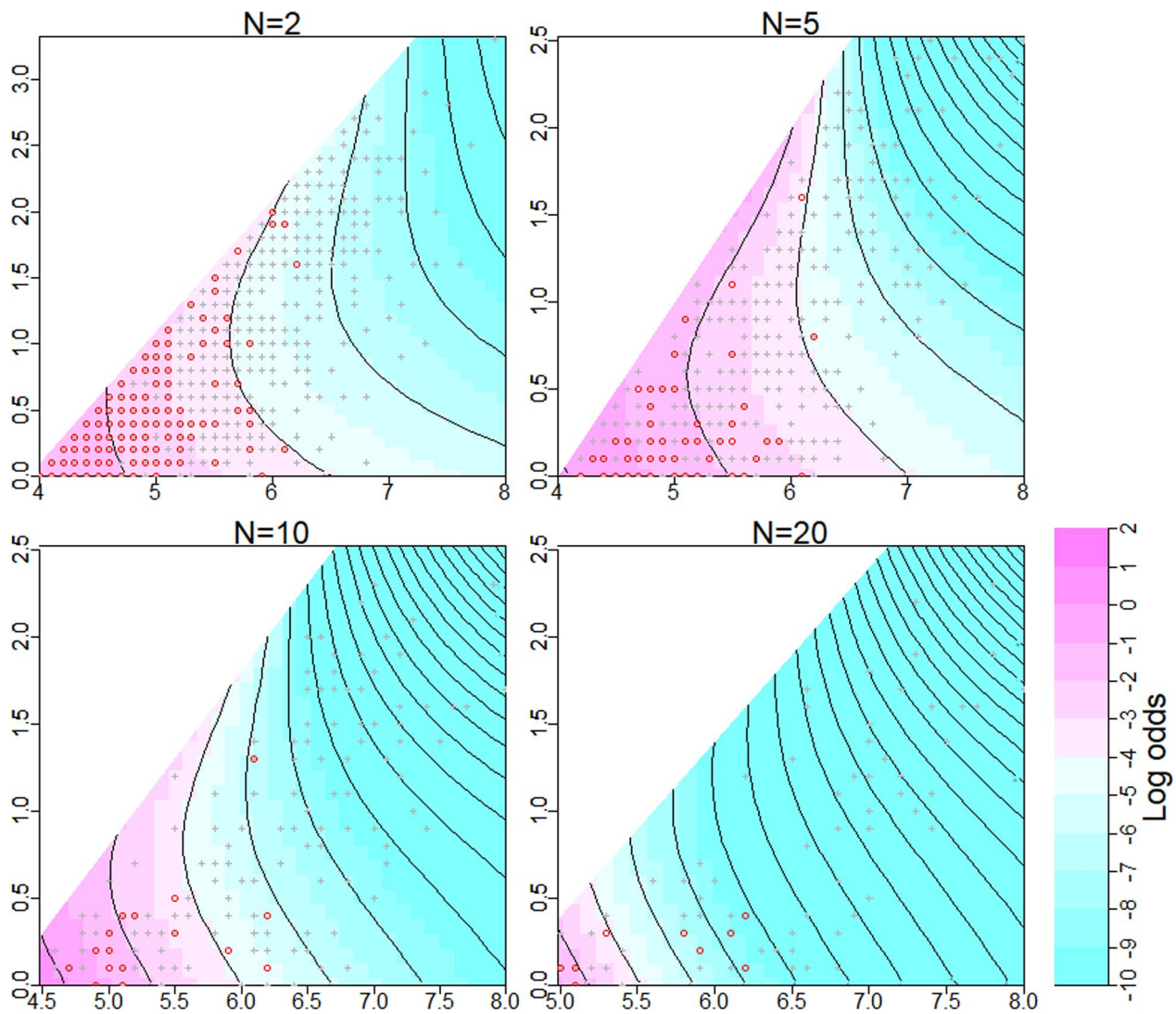
where the coefficient 0.89 is obtained by the maximum likelihood method, whereas the standard  $b$  value is 0.9 in Japan and vicinity. Ogata et al. (1995) define the foreshock-type cluster as a cluster whose magnitude gap between the mainshock and its largest foreshock is larger than 0.45 for discrimination from swarm clusters. The 0.45 magnitude gap or over is realized in less than about 20 percent of clusters, and this boundary between foreshocks and swarm pre-shocks has been heuristically set by the trade-off between the advantages of a larger magnitude gap (better discrimination) and a greater number of foreshock clusters (better statistics).

We note here that our setting generalizes Ogata et al. (1995), including their foreshock case; namely, the 0.45 magnitude gap is equivalent to 0.4 since magnitude data is given to the first decimal place, and if we limit mainshock magnitudes to be larger than the largest foreshock magnitudes + 0.4, the foreshock probability is evaluated by

$$\begin{aligned} & p(\text{foreshock}, M_{\text{main}} > M_1 + 0.4 | N_c, M_1, \Delta M, Tl, D, X, Y) \\ &= p(\text{foreshock} | N_c, M_1, \Delta M, Tl, D, X, Y) \\ &\quad \times p(M_{\text{main}} > M_1 + 0.4 | \text{foreshock}, M_1) \\ &= p(\text{foreshock} | N_c, M_1, \Delta M, Tl, D, X, Y) \times 0.44. \end{aligned} \quad (10)$$

### 3.3 Prediction of mainshock timing

Although we evaluate probabilities of foreshocks by assuming that their mainshocks occur within 30 days from the latest events, it is also important to evaluate probabilities of mainshock occurrences using the timing of mainshock occurrences from foreshocks. We demonstrate the cumulative distribution of the timing of mainshock occurrences from their foreshocks in Fig. 10. The cumulative distribution of the time lag does not vary considerably by  $T$  of the foreshock clusters. From Fig. 10, of the mainshocks within 30 days from their foreshocks, about 60% occurred within 7 days and about 30% occurred within 1 day, which are represented by vertical dashed lines, respectively. Therefore, the probabilities of mainshock occurrences



**Fig. 5** Estimated relative log odds function  $f_1(Nc, M_1, \Delta M)$  for some fixed cluster sizes. Red circles and gray pluses represent foreshock and non-foreshock clusters, respectively

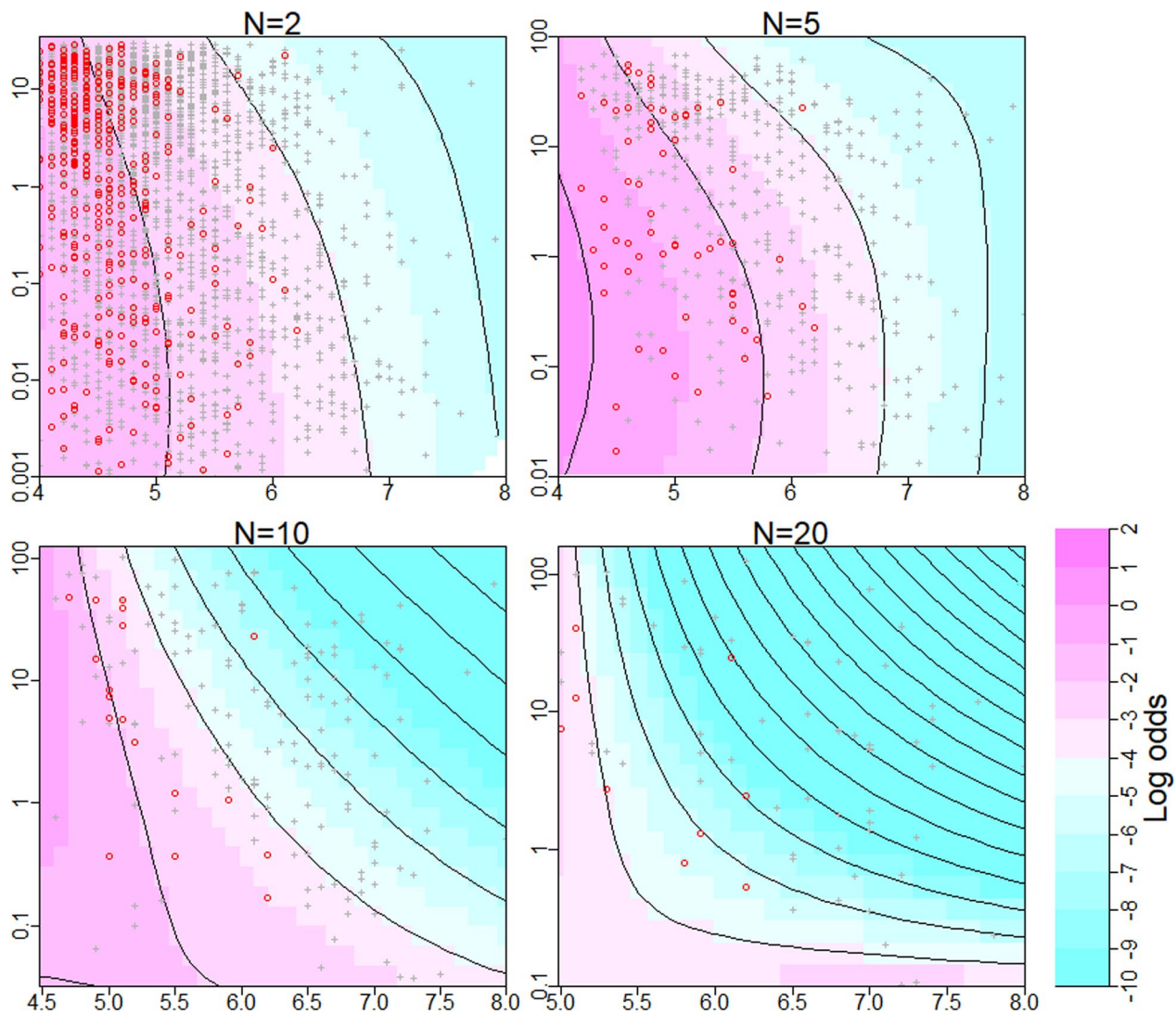
within 7 days and 1 day can be approximately evaluated by multiplying by 0.6 and 0.3, respectively, the original foreshock probabilities that mainshocks occur within 30 days.

### 3.4 Validation of foreshock discrimination

To validate our model, we evaluated the foreshock probabilities  $p(\text{foreshock} | Nc, M_1, \Delta M, Tl, D, X, Y)$  for the validation dataset for the period from January 1, 2000, to October 31, 2017, using the model estimated via the training dataset from January 1, 1926, to December 31, 1999. We compared the evaluated probabilities and the

actual proportions of foreshocks listed in Table 1. The evaluated foreshock probabilities for growing seismic clusters of fixed sizes  $N=2, 5, 10, 20$  are tabulated in 10% intervals. Since many of the actual rates of foreshocks deviate from the evaluated foreshock probabilities due to their small sample sizes, we also calculated the 90% confidence interval (CI) of the binomial proportion of foreshocks for each cell in Table 1. The 90% CI and the range of evaluated probability overlap in every cell and hence our model provided roughly consistent prediction with the actual proportion of foreshocks. The highest foreshock probabilities are marked for a seismic swarm





**Fig. 6** Estimated relative log odds function  $f_2(N_c, M_1, T_l)$  for some fixed cluster sizes. Red circles and gray pluses represent foreshock and non-foreshock clusters, respectively

off the Izu Peninsula, where the value of the baseline log odds  $g$  is the highest, and some of them are actually foreshocks. The portions of foreshocks for all clusters have similar values ranging from 13 to 16% among the different fixed cluster sizes.

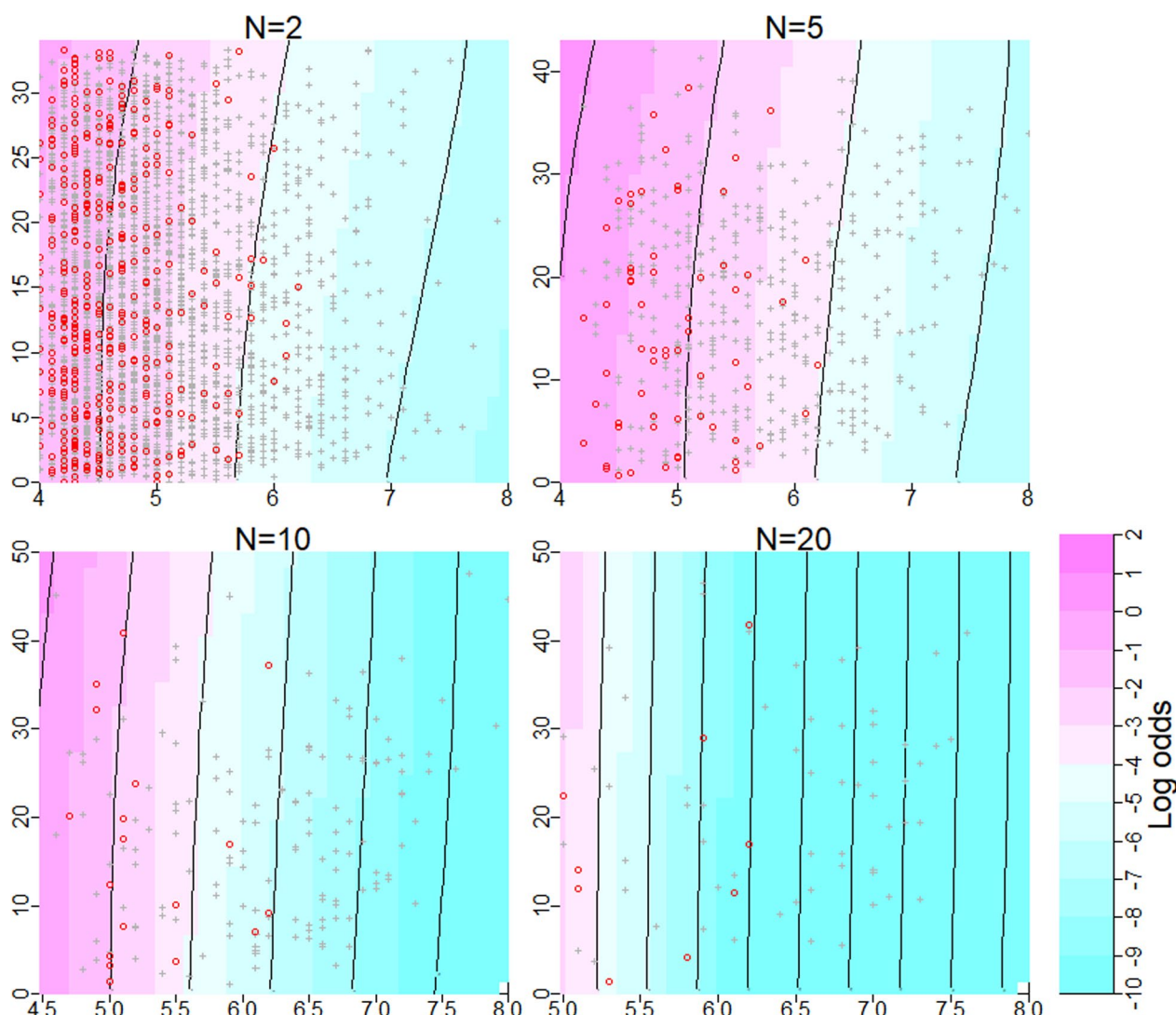
In addition, we evaluated the foreshock probabilities for mainshocks whose magnitudes are larger than their largest foreshock magnitudes + 0.4, by multiplying the probabilities evaluated in Table 1 by 0.44 as in Eq. (10). Table 2 summarizes the evaluated probabilities and the proportions of foreshocks with their 90% CIs. Most of the 90% CIs overlap with the range of evaluated probabilities. Particularly, seismic clusters with foreshock probabilities

over 10% are more likely to be actual foreshocks than those with probabilities under 10%.

Moreover, we evaluated the foreshock probabilities for M6+ mainshocks by combining Eqs. (1) and (9) as follows:

$$\begin{aligned}
 & p(\text{foreshock}, M_{\text{main}} \geq 6 | N_c, M_1, \Delta M, T_l, D, X, Y) \\
 &= p(\text{foreshock} | N_c, M_1, \Delta M, T_l, D, X, Y) \quad (11) \\
 & \quad \times p(M_{\text{main}} \geq 6 | \text{foreshock}, M_1).
 \end{aligned}$$

Table 3 shows the foreshock probabilities evaluated by Eq. (11) and the actual portion of foreshocks for

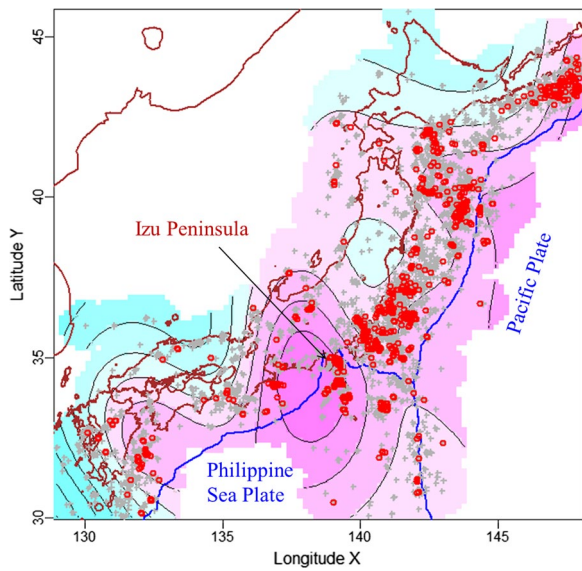


**Fig. 7** Estimated relative log odds function  $f_3(N_c, M_1, D)$  for some fixed cluster sizes. Red circles and gray pluses represent foreshock and non-foreshock clusters, respectively

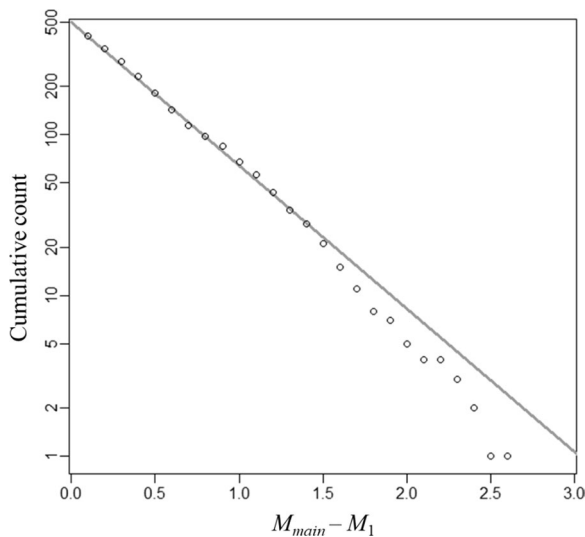
M6+ mainshocks in the contingency tables by 5% intervals. Since foreshock clusters with M6+ mainshocks are rare cases, the evaluated probabilities are mostly less than 5%. Nevertheless, seismic clusters with foreshock probabilities over 5% are more likely to be actual foreshocks than those with probabilities under 5% in Table 3. The foreshock cluster of the 2016 Kumamoto earthquake of M7.3 is evaluated as having a foreshock probability of 16% when its  $N$  reaches 20. Such a high probability is obtained because the difference between the two largest foreshock magnitudes is only 0.1 and the time duration is approximately one day in that cluster. These factors involved high odds, as shown in Figs. 5 and 6.

### 3.5 Comparison with synthetic ETAS catalogs

In the last subsection, we showed that our foreshock probability evaluation is empirically consistent with the actual foreshock rate. On the other hand, there are some studies (e.g., Helmstetter and Sornette 2003; Helmstetter et al. 2003) that criticize foreshock identification models beyond the ETAS because foreshock phenomena are also seen in the synthetic catalogs simulated from the ETAS model. In contrast, Ogata and Katsura (2014) compared probability forecasting performance between the real seismic catalog and the synthetic ETAS catalogs, and showed that the foreshock model by Ogata et al. (1996) performed better for the



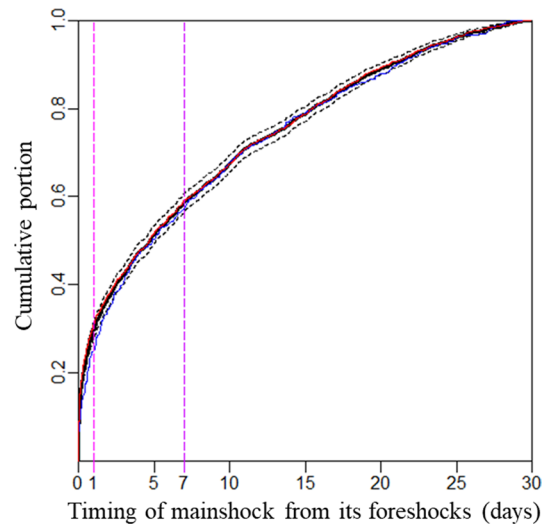
**Fig. 8** Estimated relative log odds function  $g(X, Y)$ . Red circles and gray pluses represent epicenters of foreshock and non-foreshock clusters, respectively



**Fig. 9** Cumulative counts of differences between mainshock magnitudes  $M_{\text{main}}$  and largest foreshock magnitudes  $M_1$ , which exhibits an exponential decrease as shown in gray line

real catalog than the synthetic catalogs in the log-likelihood ratio. Thus, in this subsection, we compare the predictive performance in our model between the real seismic catalog and the synthetic ETAS catalogs.

We first fitted the space–time ETAS model with the following conditional intensity function to the same catalog as that used in this study:



**Fig. 10** Empirical distribution of time lag of mainshock occurrence from its foreshock with its 95% confidence intervals (dashed curve). Those for the foreshock cluster whose time duration  $T$  is less than a day and over a day are shown in the red and blue lines, respectively

$$\lambda(t, x, y|H_t) = \mu(x, y) + \sum_{j:t_j < t} \frac{K}{(t - t_j + c)^p} \times \left[ \frac{(x - x_j)^2 + (y - y_j)^2}{e^{\alpha(m_j - m_c)}} + d \right]^{-q} \quad (12)$$

where  $H_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$  is the observation history up to time  $t$  and  $m_c = 4$  is the cutoff magnitude. The heterogeneous background rate function  $\mu(x, y)$  is estimated by the kernel method with a Gaussian kernel function. All the parameters are estimated by the maximum likelihood method.

Then, we simulated 100 synthetic catalogs from the fitted ETAS model with the following two types of magnitude sequences: (1) the same magnitude sequences as the validation catalog in the last subsection, (2) the magnitudes independently resampled from those of the validation catalog.

We applied our model to those catalogs and evaluated their predictive performances by the mean log-likelihood score per cluster introduced in Ogata and Katsura (2014). For the  $c$ th cluster, define  $p_c$  as the mean of the foreshock probabilities evaluated for its subclusters which appear during the growth of the cluster size. Then we define the log-likelihood score for the  $c$ th cluster by

$$l_c = \eta_c \log_e p_c + (1 - \eta_c) \log_e (1 - p_c) \quad (13)$$

**Table 1** Contingency table of evaluated foreshock probabilities and actual portion of foreshocks with 90% binomial proportion confidence interval in seismic clusters of size N = 2, 5, 10, 20

Foreshock probability		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	Total
N = 2	All clusters	469	489	220	59	14	2	0	0	1253
	Foreshock clusters	27	67	45	17	4	2	0	0	162
	Portion of foreshocks (90% CI)	6% (4–8%)	14% (11–17%)	20% (16–25%)	29% (19–40%)	29% (10–54%)	100% (22–100%)	N/A	N/A	N/A
N = 5	All clusters	129	57	27	13	9	3	2	2	242
	Foreshock clusters	8	12	4	3	3	1	1	2	34
	Portion of foreshocks (90% CI)	6% (3–11%)	21% (13–32%)	15% (4–34%)	23% (7–49%)	33% (10–66%)	33% (2–86%)	50% (3–97%)	100% (22–100%)	100% (22–100%)
N = 10	All clusters	68	7	3	5	4	2	2	0	91
	Foreshock clusters	5	3	1	2	3	1	0	0	15
	Portion of foreshocks (90% CI)	7% (3–15%)	43% (13–77%)	33% (2–86%)	40% (8–81%)	75% (25–99%)	50% (3–97%)	0% (0–78%)	N/A	N/A
N = 20	All clusters	34	1	2	1	0	2	0	3	43
	Foreshock clusters	1	1	1	1	0	1	0	1	6
	Portion of foreshocks (90% CI)	3% (0–13%)	100% (5–100%)	50% (3–97%)	100% (5–100%)	N/A	50% (3–97%)	N/A	33% (2–86%)	33% (2–86%)

**Table 2** Contingency table of evaluated foreshock probabilities for mainshocks with magnitudes larger than their largest foreshock magnitudes + 0.4 and actual portion of foreshocks with 90% binomial proportion confidence interval in seismic clusters of size  $N=2, 5, 10, 20$

Foreshock probability		0–10%	10–20%	20–30%	30–40%	Total
$N=2$	All clusters	1045	203	5	0	1253
	Foreshock clusters	69	28	1	0	98
	Portion of foreshocks (90% CI)	7% (5–8%)	14% (10–18%)	20% (1–66%)	N/A	8%
$N=5$	All clusters	196	36	8	2	242
	Foreshock clusters	16	5	1	2	24
	Portion of foreshocks (90% CI)	8% (5–12%)	14% (6–27%)	13% (1–47%)	100% (22–100%)	10%
$N=10$	All clusters	75	11	5	0	91
	Foreshock clusters	6	5	1	0	12
	Portion of foreshocks (90% CI)	8% (4–15%)	45% (20–73%)	20% (1–66%)	N/A	13%
$N=20$	All clusters	36	2	2	3	43
	Foreshock clusters	2	1	1	0	4
	Portion of foreshocks (90% CI)	6% (1–16%)	50% (3–97%)	50% (3–97%)	0% (0–63%)	9%

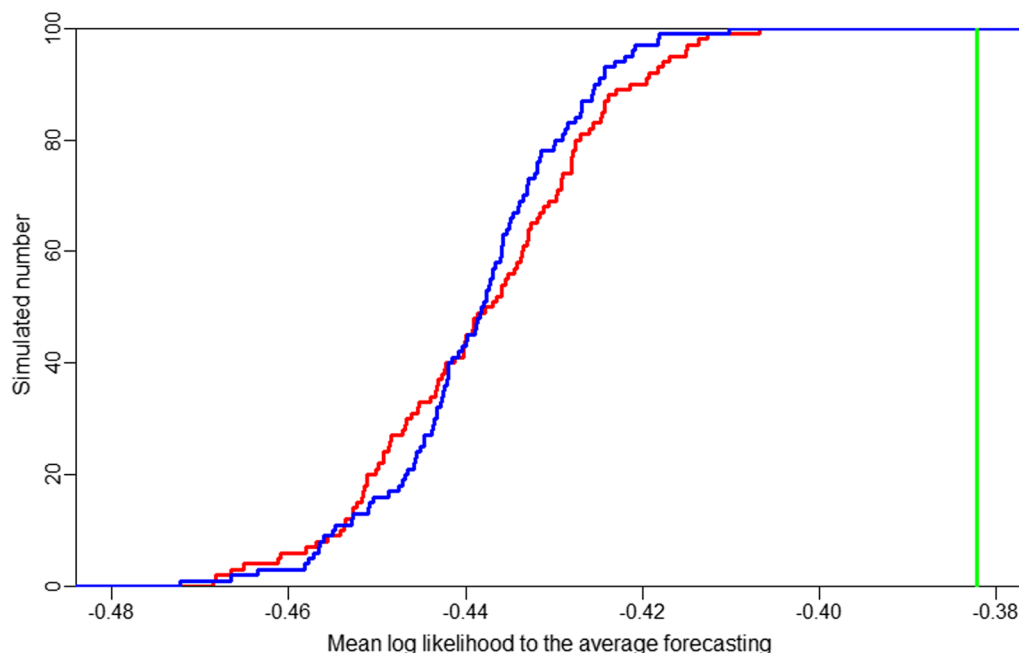
**Table 3** Contingency table of evaluated foreshock probabilities for M6+ mainshocks and actual portion of foreshocks with 90% binomial proportion confidence interval in seismic clusters of size  $N=2, 5, 10, 20$

Foreshock probability		0–5%	5–10%	10–15%	15–20%	20%–	Total
$N=2$	All clusters	1234	17	2	0	0	1253
	Foreshock clusters	28	0	1	0	0	29
	Portion of foreshocks (90% CI)	2% (2–3%)	0% (0–16%)	50% (3–97%)	N/A	N/A	2%
$N=5$	All clusters	212	24	5	1	0	242
	Foreshock clusters	12	4	0	0	0	16
	Portion of foreshocks (90% CI)	6% (3–9%)	17% (6–34%)	0% (0–45%)	0% (0–95%)	N/A	7%
$N=10$	All clusters	84	4	2	1	0	91
	Foreshock clusters	8	1	1	1	0	11
	Portion of foreshocks (90% CI)	10% (5–17%)	25% (1–75%)	50% (3–97%)	100% (5–100%)	N/A	12%
$N=20$	All clusters	36	4	0	1	2	43
	Foreshock clusters	2	1	0	1	1	5
	Portion of foreshocks (90% CI)	6% (1–16%)	25% (1–75%)	N/A	100% (5–100%)	50% (3–97%)	12%

where  $\eta_c=1$  if the  $c$ th cluster includes foreshocks and otherwise  $\eta_c=0$ . We obtained the mean log-likelihood score over all the clusters in each catalog.

Thus, for 889 clusters in the validation catalog, which is the JMA catalog from 2000 to October 31, 2017, the mean log-likelihood score was  $-0.382$ . Figure 11 shows the distribution of the mean log-likelihood scores for the 100 synthetic ETAS catalogs with each type of the magnitude sequence described above. Whereas the distribution of the scores is not

much different between the two types of the magnitude sequence, the score  $-0.382$  obtained from the real catalog was much higher than any scores from the synthetic catalogs in Fig. 11. That result implies that our model takes advantage of foreshock characteristics in the real catalog to improve the predictive performance. These results also encourage us to develop a magnitude forecast, other than the ETAS models, in order to raise the probability gain of a large earthquake forecast (see also Ogata et al. 2018).



**Fig. 11** Cumulative distribution of the mean log-likelihood scores for the 100 synthetic ETAS catalogs with the same magnitude sequence as the real catalog (red line) and the magnitudes resampled from the real catalog (blue line). The green vertical line represents the score obtained from the real catalog

#### 4 Conclusions

In this paper, we proposed a foreshock discrimination model using the information on magnitudes, space, and time in seismic clusters. The actual portion of foreshocks in the validation dataset was roughly consistent with the foreshock probability evaluated by our model. Furthermore, we provided a probabilistic evaluation of mainshock magnitudes above the foreshock magnitudes based on the Gutenberg–Richter law. Although our model showed good performance in discriminating the foreshock clusters, only approximately 10% of M6+ mainshocks were preceded by M4+ foreshock clusters. More foreshocks may be found by setting the minimum foreshock magnitudes smaller than 4 (e.g., Mignan 2014). However, if we lower the cutoff magnitude to that level, some events may be missing in the seismic clustering. Nevertheless, the features of seismic clusters defined in this study are designed to be robust to missing events and may not change significantly by them. We would like to address this issue in the future. Since our model only forecasts mainshock occurrences, in future work, we also intend to enrich our forecasts with aftershock forecasts given by the ETAS models (e.g., Ogata 2011) and submit the ensemble model to the CSEP Japan Testing Center (Tsuruoka et al. 2012).

#### Abbreviations

AIC	Akaike's information criterion
CI	Confidence interval
CSEP	Collaboratory for the Study of Earthquake Predictability
ETAS	Epidemic-type aftershock sequence
JMA	Japan Meteorological Agency
MBC	Magnitude-based clustering
SLC	Single-link clustering

#### Acknowledgements

We used the PDE Hypocenter Earthquake Catalog (as of October 2017) of the JMA. We thank the anonymous reviewers for their insightful comments, which were helpful in improving the quality of the manuscript.

#### Author contributions

SN performed data processing and statistical analysis and drafted the manuscript. YO supervised all the work of SN. Both authors read and approved the final manuscript.

#### Funding

This work was partially supported by JSPS KAKENHI (Grant No. 21H05206), by MEXT Project for Seismology toward Research Innovation with Data of Earthquake (STAR-E) Grant Number JPJ010217, and by ERI JURP 2021-B-01 in Earthquake Research Institute, the University of Tokyo.

#### Availability of data and material

We used the PDE Hypocenter Earthquake Catalog (as of October 2017) of the JMA.

#### Declarations

#### Competing interest

The authors declare that they have no competing interests.

Received: 23 November 2022 Accepted: 20 March 2023  
Published online: 12 April 2023

## Publisher's Note

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

## References

- Agnew DC, Jones LM (1991) Prediction probabilities from foreshocks. *J Geophys Res* 96:11959–11971
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723
- Console R, Murru M, Alessandrini B (1993) Foreshock statistics and their possible relationship to earthquake prediction in the Italian region. *Bull Seism Soc Am* 83:1248–1263
- Davis SD, Frohlich C (1991) Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification. *Geophys J Int* 104:289–306
- Frohlich C, Davis SD (1990) Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. *Geophys J Int* 100:19–32
- Helmstetter A, Sornette D (2003) Foreshocks explained by cascades of triggered seismicity. *J Geophys Res* 108(B10):2457
- Helmstetter A, Sornette D, Grasso JR (2003) Mainshocks are aftershocks of conditional foreshocks: how do foreshock statistical properties emerge from aftershock laws. *J Geophys Res* 108(B1):2046
- Hollander M, Wolfe DA, Chicken E (2013) *Nonparametric statistical methods*, 3rd edn. John Wiley & Sons, New York
- Jones LM (1985) Foreshocks and time-dependent earthquake hazard assessment in Southern California. *Bull Seism Soc Am* 75:1669–1679
- Jones LM, Molnar P (1979) Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J Geophys Res* 84:3596–3608
- Maeda K (1996) The use of foreshocks in probabilistic prediction along the Japan and Kuril trenches. *Bull Seism Soc Am* 86:242–254
- Mignan A (2014) The debate on the prognostic value of earthquake foreshocks: a metaanalysis. *Sci Rep* 4:4099
- Ogata Y (2011) Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth Planets Space* 63(3):217–229
- Ogata Y, Katsura K (2012) Prospective foreshock 247 forecast experiment during the last 17 years. *Geophys J Int* 191:1237–1244
- Ogata Y, Katsura K (2014) Comparing foreshock characteristics and foreshock forecasting in observed and simulated earthquake catalogs. *J Geophys Res* 119(11):8457–8477
- Ogata Y, Katsura K, Tsuruoka H, Hirata N (2018) Exploring magnitude forecasting of the next earthquake. *Seism Res Lett* 89(4):1298–1304
- Ogata Y, Utsu T, Katsura K (1995) Statistical features of foreshocks in comparison with other earthquake clusters. *Geophys J Int* 121:233–254
- Ogata Y, Utsu T, Katsura K (1996) Statistical discrimination of foreshocks from other earthquake clusters. *Geophys J Int* 127:17–30
- Papazachos BC (1974) On the time distribution of aftershocks and foreshocks in the area of Greece. *Pure Appl Geophys* 112:627–631
- Papazachos BC (1975) Foreshocks and earthquake prediction. *Technophysics* 28:213–226
- Savage MK, dePolo DM (1993) Foreshock probabilities in the western great-basin eastern Sierra Nevada. *Bull Seism Soc Am* 83:1910–1938
- Smith EGC (1981) Foreshocks of shallow New Zealand earthquakes, New Zealand. *J Geol Geophys* 24:579–584
- Tsuruoka H, Hirata N, Schorlemmer D, Euchner F, Nanjo KZ, Jordan TH (2012) CSEP testing center and the first results of the earthquake forecast testing experiment in Japan. *Earth Planets Space* 64(8):661–671
- Von Seggern D, Alexander SS, Baag CE (1981) Seismicity parameters preceding moderate to major earthquakes. *J Geophys Res* 86:9325–9351
- Wong KC, Wyss M (1985) Clustering of foreshocks and preshocks in the circum-Aegean region. *Earthq Predict Res* 1:121–140
- Wood SN (2017) *Generalized additive models: an Introduction with R*, 2nd edn. Chapman & Hall/CRC, Boca Raton
- Xu SX, Wang BQ, Jones LM, Ma XM, Shen PW (1982) The foreshock sequence of Haicheng earthquake and earthquake—the use of foreshock sequences in earthquake prediction. *Tectonophysics* 85:91–105

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)