Statistical properties of inter-event times in seismic time series transformed by occurrence rate: An analysis from the viewpoint of hierarchy in the temporal nature of seismicity

Hiroki Tanaka[1]*

[1]*Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto-shi, 606-8501, Japan*

Our previous study proposed a Bayesian framework to enhance the approach by the renewal process to forecast earthquakes' timing and tested it in simulated seismicity-like time series. As a first step toward applying the Bayesian approach to actual seismic activity, it is crucial to use seismic catalogs for examination of the probability density functions in Bayes' theorem, which is in its simplest form in the Bayesian approach: the inter-event time distribution and the conditional and inverse probability between inter-event times at two cut-off magnitudes. In this study, I examined the properties of these probability density functions using time series with weak inter-event correlations extracted from three seismic catalogs: stationary time series with a nearly constant occurrence rate and aftershock sequences transformed by the Omori – Utsu law. I found a new scaling property related to the temporal hierarchy of seismic activity. Using this property, I derived the above three probability density functions. Regarding the inter-event time distribution, I discuss its approximate scaling universality from the viewpoint of temporal fluctuations of transformed seismic time series by instantaneous occurrence rate. The derived inverse probability enables probabilistic evaluation of the large earthquakes' timing in the simplest Bayesian approach. Finally, I discuss extending it to the general Bayesian approach toward its practical use.

1. Introduction

Mitigating the potential for earthquake-related disasters is a crucial concern within society. An effective countermeasure for this issue is forecasting future significant earthquakes in a probabilistic sense. One such approach is to use the hazard function by regarding the seismic time series at a high cut-off magnitude (i.e., the point process that includes only

*tanaka.hiroki.43s@st.kyoto-u.ac.jp

events greater than or equal to the set cut-off magnitude) as a renewal process.[1,2] While this approach is helpful, it underutilizes abundant data on cut-off earthquakes.[3,4]

A recent study submitted a Bayesian method for incorporating such data into forecasting through the conditional probability $p_{mM}(\tau_m|\tau_M)$ of inter-event time τ_m at the lower cut-off magnitude m given that it is within the inter-event time of length τ_M at the upper cut-off magnitude $M(:= m + \Delta m)$.[3–6] The conditional probability serves as a link between the inter-event time distributions at these two cut-off magnitudes, $p_m(\tau_m)$ and $p_M(\tau_M)$:[6]

$$N_m p_m(\tau_m) = N_M \int_{\tau_m}^{\infty} \frac{\tau_M}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M) p_M(\tau_M) d\tau_M, \tag{1}$$

where $\langle\langle \tau_m \rangle\rangle_{\tau_M} := \int_0^{\infty} \tau_m p_{mM}(\tau_m|\tau_M) d\tau_m$, and N_m (N_M) is the total number of inter-event times at the cut-off magnitude m (M).[6] The integral equation (1) allows us to discuss the characteristics of the inter-event time distribution from the viewpoint of the hierarchical nature of the inter-event times quantified by the conditional probability.[6] Bayes' theorem is expressed with this conditional probability as follows:[3]

$$p_{Mm}(\tau_M|\tau_m) = \frac{z_{mM}(\tau_m|\tau_M) z_M(\tau_M)}{\int_{\tau_m}^{\infty} z_{mM}(\tau_m|\tau_M) z_M(\tau_M)\, d\tau_M}, \tag{2}$$

where $z_m(\tau_m)$ ($z_M(\tau_M)$) and $z_{mM}(\tau_m|\tau_M)$ are generalized probabilities:[3]

$$z_m(\tau_m) := \frac{\tau_m}{\langle \tau_m \rangle} p_m(\tau_m), \tag{3}$$

$$z_{mM}(\tau_m|\tau_M) := \frac{\tau_m}{\langle\langle \tau_m \rangle\rangle_{\tau_M}} p_{mM}(\tau_m|\tau_M), \tag{4}$$

with $\langle \tau_m \rangle := \int_0^{\infty} \tau_m p_m(\tau_m) d\tau_m$, $\langle \tau_M \rangle := \int_0^{\infty} \tau_M p_M(\tau_M) d\tau_M$, and the denominator of Eq. (2) equals $z_m(\tau_m)$.[3] $p_{Mm}(\tau_M|\tau_m)$ yields the inverse probability of the upper inter-event time given the information on a lower inter-event time in it.[3] Equation (2) is the simplest form of the Bayesian updating that incorporates information on multiple inter-event times at the lower cut-off magnitude.[3] The theoretical study of Bayesian updating for the stationary marked Poisson process derived the inverse probability analytically, and the numerical study for the ETAS[7,8] time series tested its effectiveness for forecasting the subsequent large-size event timing; the statistics showed that the forecasting relatively worked when the stationary activity was dominant, while not when the non-stationarity by the Omori – Utsu law[9,10] was dominant.[3] Further, subsequent theoretical work extended this Bayesian framework to a spatiotemporal version.[4]

In this way, previous studies on the Bayesian approach have shown a theoretical framework and tested it using synthetic time series of seismicity. However, the inverse probability

and its approximation function have not been derived for time series with inter-event correlations; the Bayesian approach has yet to achieve probabilistic forecasting in actual seismic time series, and thus, examining the Bayesian framework using seismic catalogs is necessary for its practical forecasting. Therefore, this study aims to examine the properties of the conditional and inverse probabilities in the Bayesian approach using seismic catalogs. As a first step in catalog analysis, I simplify the analysis in two aspects.

First, I examine only Bayes' theorem (Eq. (2)); it is interesting in relation to forecasting whether the inverse probability shows seismicity-dependent unimodality even in this simplest case of Bayesian updating, and it is essential to examine the conditional probability for considering general Bayesian updating. Second, I restrict the subject of analysis to the time series with weak inter-event correlation; such time series can be easy to work on by perturbation from the analytical results for the stationary marked Poisson process. As such, in this study, I analyze two kinds of time series inspired by the preceding studies on the scaling universality in the inter-event time distributions in seismic activity described below.

The scaling universality here refers to the results of catalog analyses[11–13] that the inter-event time distribution at the cut-off magnitude M in a spatial region S (temporarily represented as $p_M^S(\tau_M)$ with superscript S) collapses to around a universal (generalized) gamma distribution ($f(y)$) independent of M and S, except for short time intervals ($y \lesssim 10^{-2}$),[13,14] after rescaled by the occurrence rate (R_M^S):

$$p_M^S(\tau_M) = R_M^S f\left(R_M^S \tau_M\right), \tag{5}$$

$$f(y) \approx C y^{\gamma-1} e^{-y/B}, \tag{6}$$

where $(\gamma, B, C) \approx (0.67, 1.58, 0.5)$.[13] This scaling universality was shown for two kinds of regimes in seismic catalogs.[13] The first is the stationary regime[11–13] in which the cumulative number of events increases nearly linearly over time.[12,13,15,16] In this case, the scaling factor is the average occurrence rate over the stationary time series.[13] The second is the aftershock sequence[13] for which the inter-event times were scaled by the instantaneous Omori – Utsu occurrence rate[13,16] ($R_M^S(t) \propto t^{-p}$, where t is the time period passed since the mainshock occurrence, and p is a parameter referred to as the p-value to characterize the attenuation rate of aftershocks per day[9,10]). As the occurrence rate is almost constant in the stationary time series, the scaling factors in these time ranges can be consolidated into the occurrence rate at a given moment; thus, this scaling universality implies that the instantaneous occurrence rate governs the temporal characteristic of seismic activity.[13]

Subsequent studies of Corral's catalog analysis critically examined this scaling univer-

sality theoretically and numerically using the ETAS model[14,17–21] and elevated it to include multiple time scales other than the occurrence rate.[22–24] I point out that the treatment of the time series differs in these subsequent studies from Corral's original method; these studies regarded the stationarity as the branching ratio[25] is less than one, and the universality was discussed regardless of whether the time series includes major aftershock activities (I refer such time series as nonstationary ones) using the average occurrence rate for entire time series as the scaling factor, leading to approximate- or non-universality. Such an approach overestimates the number of short inter-event times in rescaling inter-event time distributions because the short intervals in major aftershock sequences are not transformed by the Omori – Utsu law. Also, the scaling universality for aftershock sequences by the Omori – Utsu rate[26] has not been considered much since Corral pointed it out.

In this study, on the way of examining Bayes' theorem, I re-examine the scaling universality based on Corral's original view; I consider the scaling of the inter-event time distribution as a temporal unfolding transformation of time series using smooth functions of the event rate[27–29] (Sect. 2 describes details), interpreting the universality as the nature of the temporal fluctuations in the unfolding-transformed time series. This viewpoint can avoid the above-mentioned overestimation and enables us to examine the scaling universality following the original view.

The transformed time series are suitable for analyzing the Bayesian approach in that the inter-event correlations are weak and enable us to consider the scaling universality in light of the hierarchical property of seismic time series while including aftershock sequences. This way, the properties of all three probability distributions associated with Bayes' theorem (Eq. (2)) are analyzed, including their interrelationships.

Thus, this study examines seismic time series after unfolding transformation to explore the characteristics of the probability densities (the conditional and inverse probabilities and the inter-event time distribution) in Bayes' theorem (Eq. (2)); along the way, I discuss the universality of the inter-event time distribution via the integral equation (1), based on the characteristics of the temporal hierarchy of the seismic time series. Section 2 describes the temporal transformation procedure of seismic time series, information on seismic catalogs and how to set the analysis range, the method to extract time series from these catalogs, their temporal transformation, scaling property of inter-event time distributions of after-transformed time series, and methods for analyzing the conditional probability of these time series. Section 3 shows the results of catalog analysis relating to the conditional probability. Section 4, based on the results in Sect. 3, derives the conditional probability, the scaling function of the inter-

event time distribution, and the inverse probability in this order and compares them with the results of catalog analysis. Finally, Sect. 5 summarizes and concludes the study and presents some discussions.

2. Methods

2.1 Transformation of time series

This subsection describes the time series transformation applied in this study. Let t_k and $\tau_k (:= t_{k+1} - t_k)$ be the k-th event's occurrence time and the k-th inter-event time in a time series, respectively. In the scaling of Eq. (5) for an aftershock sequence, time intervals are transformed by instantaneous occurrence rate as follows:[16)]

$$\tau_k \mapsto R(t_{k+1})\tau_k. \tag{7}$$

However, this transformation has ambiguity in terms of the choice of time to take the occurrence rate (any time between t_k and t_{k+1} can be the argument of $R(\cdot)$[16)]). Instead, I apply the following transformation that is a modification of Eq. (7):

$$\tau_k \mapsto R(t_{k+1})t_{k+1} - R(t_k)t_k. \tag{8}$$

This is equivalent to converting the occurrence times:

$$t_k \mapsto R(t_k)t_k. \tag{9}$$

In particular, for stationary time series ($R(t) \approx R$ (const.)) and aftershock sequences ($R(t) \simeq Kt^{-p}$), the transformation of Eq. (9) is:

$$t_k \mapsto \begin{cases} Rt_k & \text{(stationary)}, \\ Kt_k^{1-p} & \text{(aftershock)}. \end{cases} \tag{10}$$

For these time series, the transformation in Eq. (10) is equivalent to transforming the time to the following transformed time (the same transformation for aftershock sequences was already considered in Ref. 27), except for the coefficient:

$$t_k \mapsto z_k := \int_{t_{\min}}^{t_k} R(s)ds, \tag{11}$$

where $t_{\min}$ is its starting time ($t_{\min} = 0$) for stationary time series and the lower bound of the time range in which $R(t) \simeq Kt^{-p}$ holds for aftershock sequences. Actually, Eqs. (11) for the stationary time series and the aftershock sequence ($p \neq 1$) are as follows:

$$t_k \mapsto z_k = \begin{cases} Rt_k & \text{(stationary)}, \\ \frac{Kt_k^{1-p}}{1-p} - \frac{Kt_{\min}^{1-p}}{1-p} & \text{(aftershock)}. \end{cases} \tag{12}$$

The constant ($Kt_{\rm min}^{1-p}/(1-p)$) in Eq. (12) does not affect the inter-event time statistics in $\{z_k\}$.

After these transformations using temporal variation trends ($R(t)$), the time series are aligned to stationary time series with an average occurrence rate of 1. Thus, the temporal fluctuations in different time series become comparable with each other. The statistics of intervals for $\{z_k\}$ suggest the nature of such fluctuations, and Eq. (5) can be regarded as implying the universality of such fluctuations. This type of transformation has been used in the field of quantum chaos and is known as the unfolding transformation.[28,29] In the following, the transformation described above is referred to as the unfolding transformation or unfolding procedure.

Thus, this study performs the following transformation for the stationary and aftershock time series in seismic catalogs:

$$t_k \mapsto w_k = \begin{cases} t_k & \text{(stationary)}, \\ t_k^{1-p} & \text{(aftershock, } p < 1), \\ -t_k^{1-p} & \text{(aftershock, } p > 1). \end{cases} \tag{13}$$

The unfolding procedure is completed by further scaling the transformed time series with the average occurrence rate over the entire time series ($w_k \mapsto z_k := w_k/\langle w \rangle$), which can be done by applying the rescaling of the inter-event time distribution for these transformed time series. These time series ($\{w_k\}$) are the objects of catalog analysis. Note that, hereafter, for the inter-event times of the transformed time series, I use the same symbol of τ as was used for the inter-event times before the transformation.

2.2 Catalog information and procedure of time series transformation

In the approach of this study, I used three seismic catalogs; one is the global Centroid Moment Tensor (CMT) catalog,[30–32] and the two are local catalogs in Southern California (SCEDC)[33,34] and Japan (JMA).[35,36] For each earthquake catalog, I determined the space-time windows to extract stationary and aftershock time series and the magnitude ranges for analysis while considering the catalog completeness in the following way.

2.2.1 CMT catalog

I used the global CMT catalog covering large earthquakes worldwide.[30–32] The moment magnitude is unaffected by saturation at the large scale and thus is appropriate for this study, where the magnitude values are significant to examine the dependence of statistics on cut-off magnitude. The CMT catalog starts from 01/01/1976, though, as shown in Fig. 1(a),

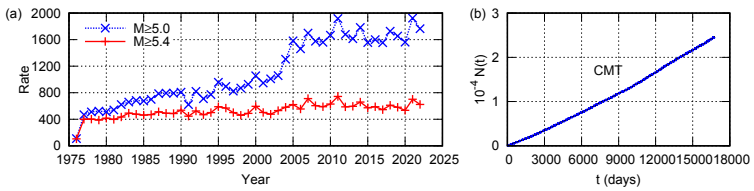

Fig. 1. (Color Online) (a) Annual number of earthquakes recorded in the global CMT catalog (blue × symbol with dotted line for only events with magnitude ≥ 5 and red + symbol with bold line for ≥ 5.4) and (b) the cumulative number of earthquakes ($N(t)$) recorded in the CMT catalog with magnitude ≥ 5.4 against the time (t) from January 1, 1977, at 00:00:00.

the number of recorded events in 1976 is less than in other years, and I chose the term to analyze 01/01/1977 – 12/31/2022. No spatial restriction was imposed. Figure 2(a) shows the probability density of magnitude ($p(M)$, magnitude frequency hereafter) for the CMT catalog in the above period. The completeness magnitude (M_c hereafter) calculated by the MBASS method[37] was 5.4. The figure also shows the b-value ($\hat{b}$ hereafter) calculated for each neighboring two points (at magnitude M and $M + 0.1$, where 0.1 is the increment of magnitude in this study) of the cumulative distribution ($P(M) := \int_M^\infty p(M')dM'$), i.e., $\hat{b} = -(\log_{10} P(M + 0.1)/P(M))/0.1$. Referring to these results (the range larger than or equal to M_c, and where $\hat{b}$ fluctuation is as small as possible, in particular, avoid the large magnitudes where $\hat{b}$ fluctuation becomes very large) and the apparent shape of the graph ($\log_{10} p(M)$ versus M being close to the straight line the Gutenberg – Richter (GR) law ($P(M) \propto 10^{-bM}$)[38] indicates), I determined the magnitude range to set the cut-off magnitude in the following analysis to be $5.4 \leq M \leq 7.5$. The b-value calculated in this magnitude range using the maximum likelihood estimate[39,40] was about 0.990. The cumulative number of earthquakes ($N(t)$, where t is the time since 01/01/1977 at 00:00:00) in the above-determined space-time window with magnitudes ≥ 5.4 is shown in Fig. 1(b); the whole time series can be judged to be almost stationary by the nearly linear form of $N(t)$, $N(t) \approx t$, as in the preceding studies[15,16] (though, precisely speaking, there appears to be gradual increasing trend even for $M \geq 5.4$ as observed in Fig. 1(a), which may due to the improvements in analysis method[32]). The stationarity of time series was judged similarly in other following catalogs.

2.2.2 *Southern California catalog*

The Southern California catalog[33,34] contains the earthquakes in the spatial range 112°W – 123°W and 29°N – 38°N and the time domain 01/01/1981 – 03/31/2022. The process to determine the space-time and magnitude ranges to analyze from this catalog was as follows.

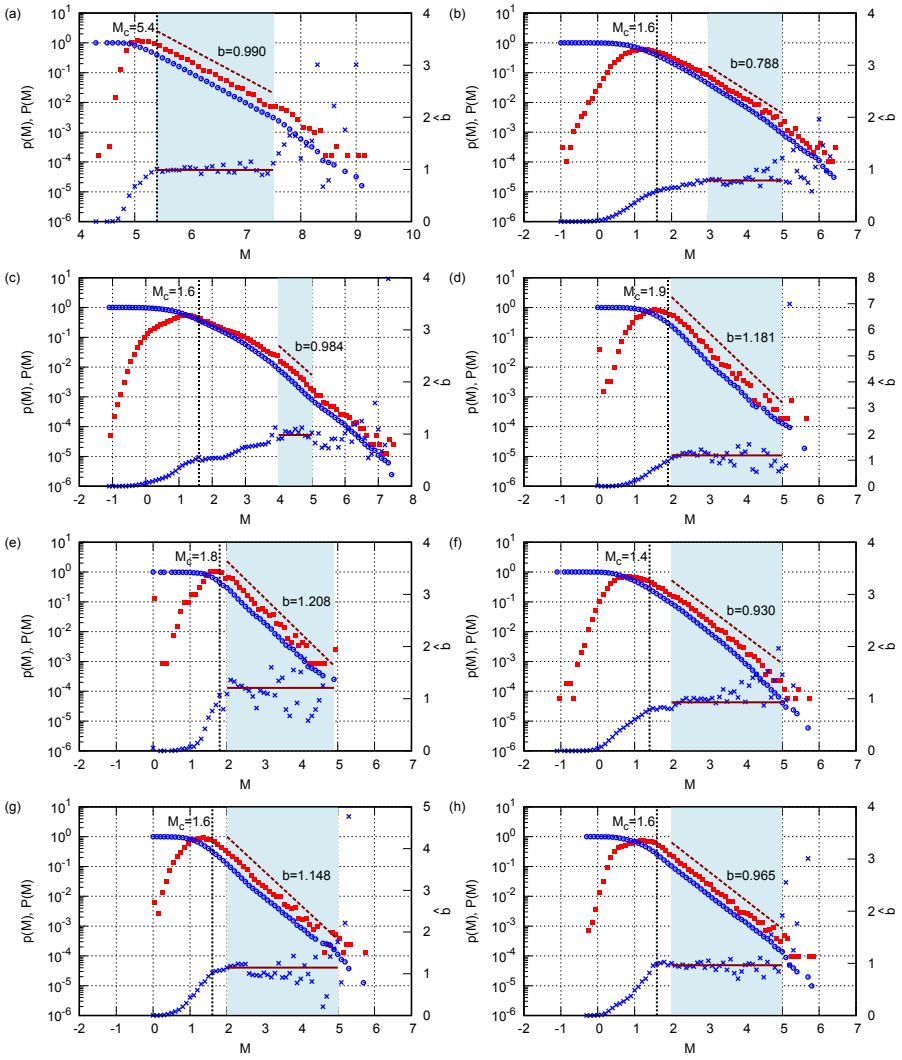

Fig. 2. (Color Online) (Red filled square) $p(M)$, (Blue circle) $P(M)$, and (Blue $\times$) $\hat{b}$ for (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The vertical dotted line shows the completeness magnitude (M_c) calculated by the MBASS method.[37] The magnitude range to set the cut-off magnitude is shaded by light blue, the dotted line shows the GR law with the b-values determined using the maximum likelihood estimate.[39,40] The horizontal line indicates the b-value for comparison with $\hat{b}$.

First, I divided the spatial area into smaller cells of $0.5^\circ \times 0.5^\circ$. Then, I calculated M_c by the MBASS method for each cell with at least 100 events to check its spatial distribution; this process refers to the method used in Ref. 43. Figure 3(a) shows the distribution map of M_c. This result suggests that the M_c at the periphery tends to take higher values (around 3 or higher) than the central part, as preceding studies have already suggested.[41–43] Figure 3(g) shows $N(t)$ for earthquakes with magnitude ≥ 3; based on this figure, I selected two stationary time series (SCS1 and SCS2) and three mainshock-aftershock sequences (SCA1, SCA2, and SCA3; the aftershocks of Landers, Hector Mine, and El Mayor-Cucapah earthquakes[44]) (see

Table I for details on time domains). The distributions of M_c were re-calculated for the earthquakes in each time window; Figures 3(b-f) show the M_c map for each time domain. From Figs. 3(b-f), I determined by sight the quadrilateral area with the vertices at ($120°W, 34.5°N$), ($118.5°W, 36.5°N$), ($117°W, 36.5°N$), and ($114.5°W, 31.5°N$), which is the spatial window excluding the periphery with high M_c as much as possible and including the three mainshock epicenters; the earthquakes included in this area were the target of the analysis below.

After setting the space-time window by the above-mentioned process, I determined the magnitude ranges to set the cut-off magnitude for the two stationary time series. Referring to the M_c values by the MBASS method and $\hat{b}$ values, I visually determined the magnitude range the GR exponential decay holds (see Figs. 2(g, h) and Table I). The maximum likelihood estimate yielded the b-values in these ranges as summarized in Table I. On the other hand, for the three mainshock-aftershock sequences, I further narrowed down the time domain in which $R(t) \propto t^{-p}$ holds. Figure 5 shows the number of aftershocks with magnitude ≥ 2 per day versus time (t) from each mainshock. As shown in the figure, $R(t) \propto t^{-p}$ holds for each aftershock sequence after sufficient time passed from the mainshock. Based on the figure, I chose the time domain where $R(t) \propto t^{-p}$ holds by sight and computed the p-values by fitting within these regimes (see Table II for details). This procedure can exclude the time domain immediately after the mainshock, where M_c tends to become large.[45,46] In this way, the space-time windows for the aftershock regimes were set. For the events within these space-time windows, I determined the magnitude ranges and the b-values in the same manner as for the stationary time series; see Figs. 2(d-f) for the magnitude frequency and Table I for the b-values.

2.2.3 JMA catalog

The JMA catalog[35,36] covers earthquakes throughout Japan, and in this study, I focused on the spatial area around the Tohoku mainshock (M9) in 2011. First, I provisionally set the spatial domain $140°E - 150°E$ and $35°N - 42°N$ and the time domain 01/01/2000 – 03/31/2022. Then, I examined the spatial distribution of M_c in the same way as the Southern California catalog, before (from 01/01/2000 to immediately before the Tohoku mainshock time, Fig. 4(a)) and after (from the Tohoku mainshock time to 03/31/2022, Fig. 4(c)) the Tohoku mainshock. Figures 4(a) and (c) suggest that M_c are partially as high as 3 to 4. Although the threshold magnitude 3 is somewhat low compared to these results, I examined $N(t)$ for earthquakes with magnitude ≥ 3 as shown in Figs. 4(d, e) and selected the stationary time domains (JS) and the aftershock sequence of the Tohoku mainshock (JA) (see Table I

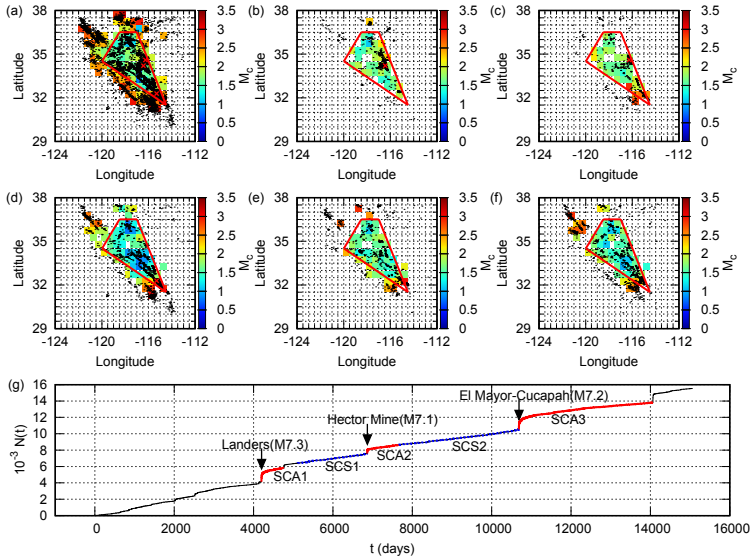

Fig. 3. (Color Online) Spatial distribution of M_c calculated by the MBASS method for the Southern California catalog of the time domain (a) 01/01/1981 – 03/31/2022, (b) SCA1, (c) SCA2, (d) SCA3, (e) SCS1, and (f) SCS2. The black dots show the epicenters of the earthquake with magnitude ≥ 3. The red quadrilateral represents the spatial domain to be analyzed. Red $\times$ symbols in (b-d) indicate the mainshock epicenter in each time domain. (g) $N(t)$ for earthquakes with magnitude ≥ 3, from January 1, 1981, at 00:00:00; the stationary domains (SCS1 and SCS2) are indicated by blue dotted lines and mainshock-aftershocks (SCA1, SCA2, and SCA3) are by bold red curves.

for the respective time domains). Figure 4(b) shows the M_c distribution in JS calculated by the MBASS method. Taking the results in Figs. 4(b, c) into account, I determined the spatial domain to analyze to the inside of the rectangle $140°E$ – $146°E$ and $35°N$ – $42°N$.

Figure 2(b) shows the magnitude frequency in the above-set space-time window of JS. As in the case of Southern California, I determined the magnitude range by visual inspection from the linear part of Fig. 2(b) such that larger than M_c calculated by the MBASS method. On the other hand, for JA, I narrowed down the time domain to the regime obeying $R(t) \propto t^{-p}$; From the occurrence rate graph in Fig. 5, I visually determined such time domain to be $10 < t$. After setting the space-time window for JA, I finally determined the magnitude range as other time series based on Fig. 2(c). The magnitude ranges and the b-values for JS and JA are summarized in Table I.

2.2.4 Time series transformation

Using the above-mentioned process, I obtained four stationary time series (CMT, JS, SCS1, and SCS2) and aftershock sequences (JA, SCA1, SCA2, and SCA3). I applied the transformation in Eq. (13) for these time series. Figure 6 shows the cumulative number of

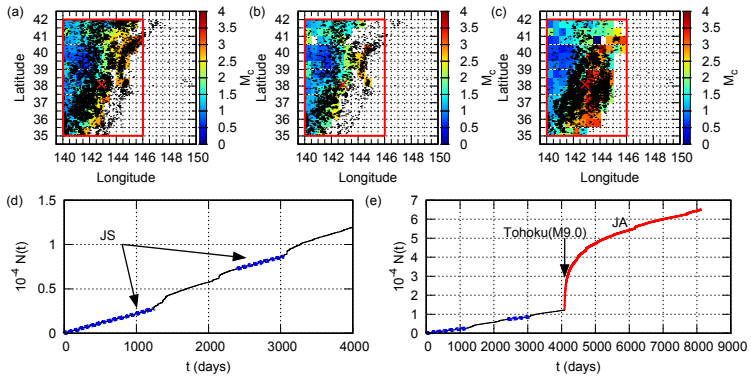

Fig. 4. (Color Online) Spatial distribution of M_c calculated by the MBASS method for the JMA catalog around the epicenter of Tohoku mainshock (shown by red × symbol) in the time domain (a) 01/01/2000 – immediately before the mainshock time, (b) JS, and (c) JA. The black dots show the epicenters of earthquakes with magnitude (a, b) ≥ 3 and (c) ≥ 4 in $140°E - 150°E$ and $35°N - 42°N$ (note that the range shown is slightly wider than this) for each time domain. Inside the red rectangle is the spatial domain to analyze. (d, e) $N(t)$ for earthquakes with magnitude ≥ 3 in the area $140°E - 150°E$, $35°N - 42°N$ from January 1, 2000, at 00:00:00 (JST); the bold red curve indicates the aftershock sequence (JA) and the dotted blue lines stationary regimes (JS).

aftershocks ($\hat{N}(w)$) against the transformed time $w(= \pm t^{1-p})$. Note that the transformation is different for each aftershock sequence depending on the p-value. After the transformation, one can see that $\hat{N}(w)$ increases linearly for the transformed time w in the time ranges shown in Table II. In the following, I analyze these transformed aftershock time series (JA, SCA1, SCA2, SCA3) as well as the stationary time series (CMT, JS, SCS1, SCS2). Hereafter, these unfolding-transformed time series are referred to by the names in Table I. I performed the catalog analysis for these time series by setting the cut-off magnitude in increments of 0.1.

2.3 Scaling of inter-event time distributions

Figure 7 shows the inter-event time distribution after rescaled by the average ($R_m = 1/\langle\tau_m\rangle$) as in Eq. (5), for each unfolding-transformed time series. This figure is drawn using the following method, which is equivalent to changing the bin width according to the inter-event time length.[13] First, the inter-event times at cut-off magnitude m are transformed as $\tau_m \mapsto \sigma_m := \log_{10}\tau_m$. Let $\tilde{N}_m(\sigma_m)$ denote the number of such transformed intervals of length within $[\sigma_m, \sigma_m + \Delta\sigma)$ in the time series, where $\Delta\sigma$ is fixed to 0.1 in this study. Then the probability density function of this transformed intervals ($\tilde{p}_m(\sigma_m)$) is calculated as:

$$\tilde{p}_m(\sigma_m) = \frac{\tilde{N}_m(\sigma_m)}{N_m \Delta\sigma}.$$

Table I. Information on spatial domain, time domain, magnitude range set by the author, and the b-value determined by the maximum likelihood estimate in the magnitude range, the total number of earthquakes in the spatiotemporal domain with magnitude ≥ the lower limit of the magnitude range, of seismic catalogs used in this study. For the mainshock-aftershock sequences (JA, SCA1, SCA2, SCA3), the number of events is bracketed because the events to be analyzed are more limited (see Table II). The first part of JS includes events up to just before the M7.1 earthquake on May 26, 2003, at 18:24:33.42 (JST). Also, the second part of JS includes events up to April 30, 2008; the term is set to avoid activity that could be considered as foreshocks prior to the M7.0 event on May 8, 2008, at 1:45:18.77 (JST). SCA1 includes events up to just before the Northridge earthquake (M6.7) on January 17, 1994, SCS1 just before the Hector Mine earthquake (M7.1), SCS2 just before the El Mayor-Cucapah earthquake (M7.2), and SCA3 just before the foreshock (M6.4) of the 2019 Ridgecrest M7.1 earthquake.[44)]

Catalog	Name	Spatial domain	Time domain	Magnitude range	b-value	Number
global CMT	CMT	Whole world	1977/01/01 – 2022/12/31	$5.4 \le M \le 7.5$	0.990	24648
JMA	JS	$140°E$ – $146°E$	2000/01/01 – 2003/05/26	$3 \le M \le 5$	0.788	2737
		$35°N$ – $42°N$	2006/08/01 – 2008/04/30	$3 \le M \le 5$		1377
	JA		2011/03/11 – 2022/03/31	$4 \le M \le 5$	0.984	(9162)
SCEDC	SCA1	The quadrilateral with vertices:	1992/06/28 – 1994/01/17	$2 \le M \le 5$	1.181	(17308)
	SCS1	$(120°W, 34.5°N)$,	1995/01/01 – 1999/10/16	$2 \le M \le 5$	1.148	9665
	SCA2	$(118.5°W, 36.5°N)$,	1999/10/16 – 2001/12/31	$2 \le M \le 4.9$	1.208	(9890)
	SCS2	$(117°W, 36.5°N)$,	2002/01/01 – 2010/04/04	$2 \le M \le 5$	0.965	10589
	SCA3	and $(114.5°W, 31.5°N)$	2010/04/04 – 2019/07/04	$2 \le M \le 5$	0.930	(18852)

Table II. Time ranges of aftershock sequences to be analyzed. The p-value and the number of events within the time range are also shown. p-values are calculated for events with magnitude ≥ minimum magnitude in Table I.

Name	$R(t) \propto t^{-p}$ range	p-value	Number
JA	$t > 10$	0.895	6706
SCA1	$t > 10$	1.061	12856
SCA2	$10^2 > t > 10^{0.5}$	1.005	3531
SCA3	$t > 10$	0.820	14978

$p_m(\tau_m)$ is obtained by re-transforming $\tilde{p}_m(\sigma_m)$ by $\sigma_m \mapsto \tau_m$. By using the class mark $(\sigma_m + 0.5\Delta\sigma)$ for $d\sigma_m/d\tau_m = 1/(10^{\sigma_m} \ln 10)$,

$$p_m(\tau_m) = \frac{1}{10^{\sigma_m + 0.5\Delta\sigma} \ln 10} \frac{\tilde{N}_m(\sigma_m)}{N_m \Delta\sigma}. \tag{14}$$

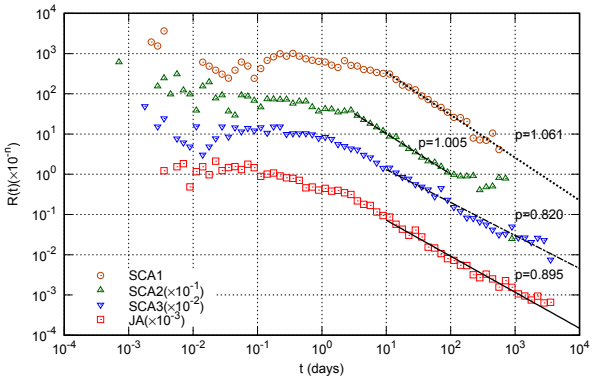

Fig. 5. (Color Online) Number of aftershocks with magnitude ≥ minimum magnitude in Table I per day against time from the mainshock. SCA2, SCA3, and JA results are multiplied by 10^{-1}, 10^{-2}, and 10^{-3}, respectively. Black (dotted) lines are the fitting curves by $R(t) \propto t^{-p}$. Fitting ranges and the p-values are summarized in Table II.

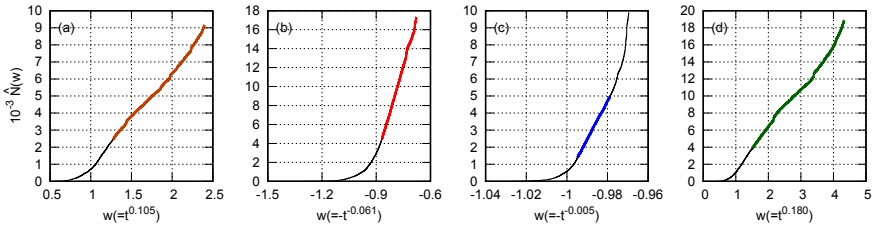

Fig. 6. (Color Online) Cumulative number of earthquakes with magnitude ≥ minimum magnitude in Table I ($\hat{N}(w)$) against transformed time from the mainshock $w = \pm t^{1-p}$ for (a) JA, (b) SCA1, (c) SCA2, and (d) SCA3. Bold lines indicate the stationary regimes in the transformed aftershock sequences.

In this way, the inter-event time distribution is drawn by plotting Eq. (14) against $10^{\sigma_m + 0.5\Delta\sigma}$. Finally, the results shown in Fig. 7 are obtained through further rescaling by $\langle\tau_m\rangle$.

One can see that the rescaled distributions are around the same curve that is expressed in Eq. (6), except for very short intervals ($\tau_m/\langle\tau_m\rangle \lesssim 10^{-2}$). For such short intervals, the rescaled distributions, in many cases, tend to take larger values than those used in Eq. (6); this may be due in part to the limitation in terms of accuracy, and such limitation is often removed by setting a minimum interval (around 10 seconds[23] to 2 minutes[13]). Also, preceding studies provided theoretical explanations for this trend using the ETAS model.[19–21]

These characteristics observed in Fig. 7 are consistent with the results of preceding catalog analyses.[11–13] Thus, the scaling property of the inter-event time distribution was also confirmed for the unfolding-transformed time series.

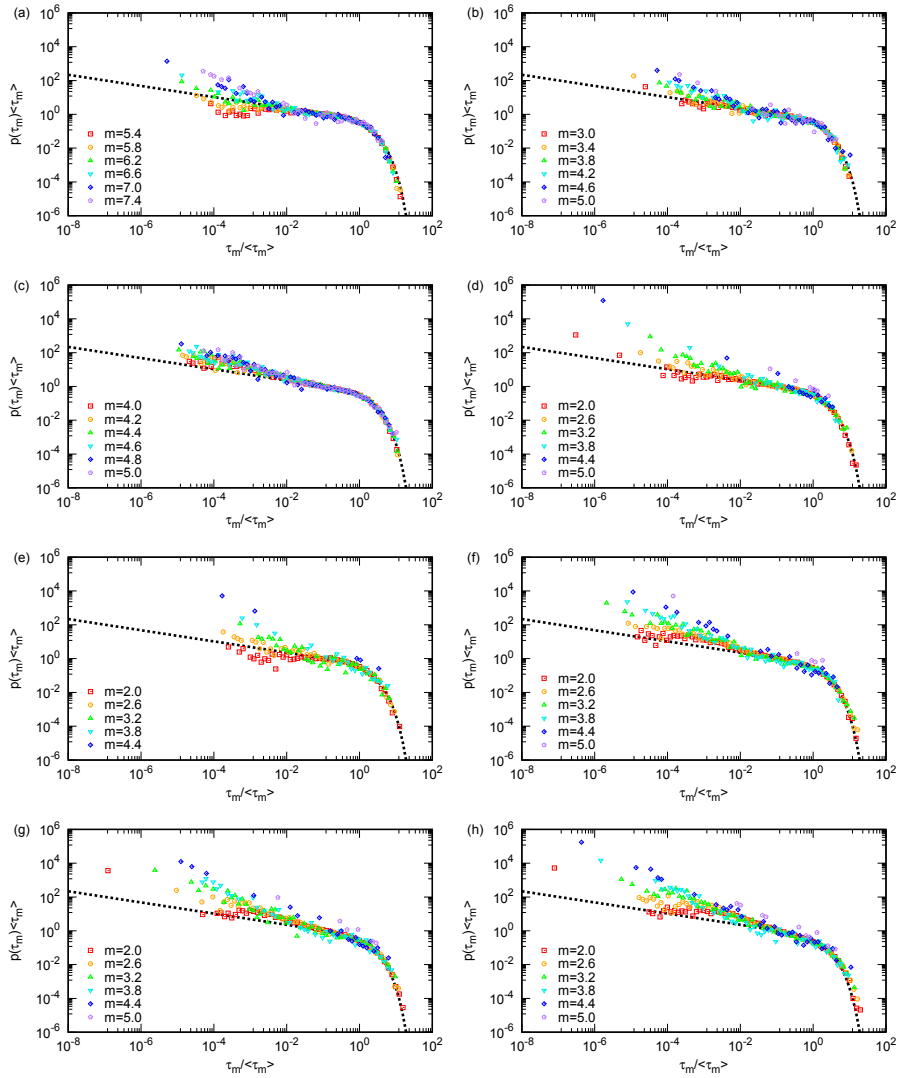

Fig. 7. (Color Online) Rescaled inter-event time distributions by the average interval $\langle\tau_m\rangle$ for the unfolding-transformed time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. For each time series, the results of several cut-off magnitudes are shown. The dotted curve is the gamma distribution in Eq. (6).

2.4 Method to examine the conditional probability

To examine the properties of the conditional probability, I use the rewriting of this function in Eq. (15) introduced in Ref. 3:

$$p_{mM}(\tau_m|\tau_M) = \frac{\sum_{i=1}^{\infty} i\,\rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M)}{\sum_{i=1}^{\infty} i\,\Psi_{mM}(i|\tau_M)}. \tag{15}$$

Here, $\Psi_{mM}(i|\tau_M)$ denotes the probability mass function for the number of inter-event intervals (i) at the lower cut-off magnitude m that are included in the inter-event time interval of length τ_M at the upper cut-off magnitude M.[3)] Also, $\rho_{mM}(\tau_m|i,\tau_M)$ denotes the probability density of

an inter-event time length at the lower cut-off magnitude m, conditioned on it being in the inter-event time interval of length τ_M at the upper cut-off magnitude M and including i pieces of lower intervals; in particular, when $i = 1$, $\rho_{mM}(\tau_m|1, \tau_M)$ is equivalent to the Dirac's delta function $\delta(\tau_M - \tau_m)$.[3)]

For the stationary marked Poisson process where the mark represents the magnitude value that is randomly determined obeying the GR law, these components are as follows:[3,47,48)]

$$\Psi_{mM}(i|\tau_M) = \frac{\left(A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}\right)^{i-1}}{(i-1)!}e^{-A_{\Delta m}\frac{\tau_M}{\langle\tau_M\rangle}}, \tag{16}$$

$$\rho_{mM}(\tau_m|i, \tau_M) = \frac{(i-1)}{\tau_M}\left(1 - \frac{\tau_m}{\tau_M}\right)^{i-2}\theta(\tau_M - \tau_m), \qquad (i \geq 2), \tag{17}$$

where,

$$A_{\Delta m} := \frac{\langle\tau_M\rangle}{\langle\tau_m\rangle} - 1 = 10^{b\Delta m} - 1, \tag{18}$$

and $\theta(x)$ is the unit step function. The conditional probability for the stationary marked Poisson process can be derived by just substituting the above components in Eq. (15).[3)]

Correlations between events in seismic time series change these components from Eqs. (16) and (17). The conditional probability in time series with weak inter-event correlations can be studied by examining such changes. The subsequent section reports the findings from analyzing the unfolding-transformed time series extracted in the preceding subsection using this approach.

3. Results of catalog analysis

3.1 Results for the component $\Psi_{mM}(i|\tau_M)$

3.1.1 Scaling property for the average of the conditional probability

First, I show the scaling property for the average of the conditional probability. By definition, the length of an inter-event time interval at the upper cut-off magnitude (τ_M) divided by this average ($\langle\langle\tau_m\rangle\rangle_{\tau_M}$) is equal to the average number of intervals at the lower cut-off magnitude included in that upper interval, i.e.,

$$\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} = \sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M). \tag{19}$$

This demonstrates that the scaling property associated with Eq. (19) as well as the average of the conditional probability is linked only to the component $\Psi_{mM}(i|\tau_M)$.

Subtracting 1 from Eq. (19) yields the average number of magnitude $\geq m$ events included in the inter-event time of length τ_M at the upper cut-off magnitude M. Figure 8 shows the

results after scaling this average using the factor $A_{\Delta m}$ for magnitude and the average interval $\langle \tau_M \rangle$ for time, across each time series. These results were obtained in the following way.

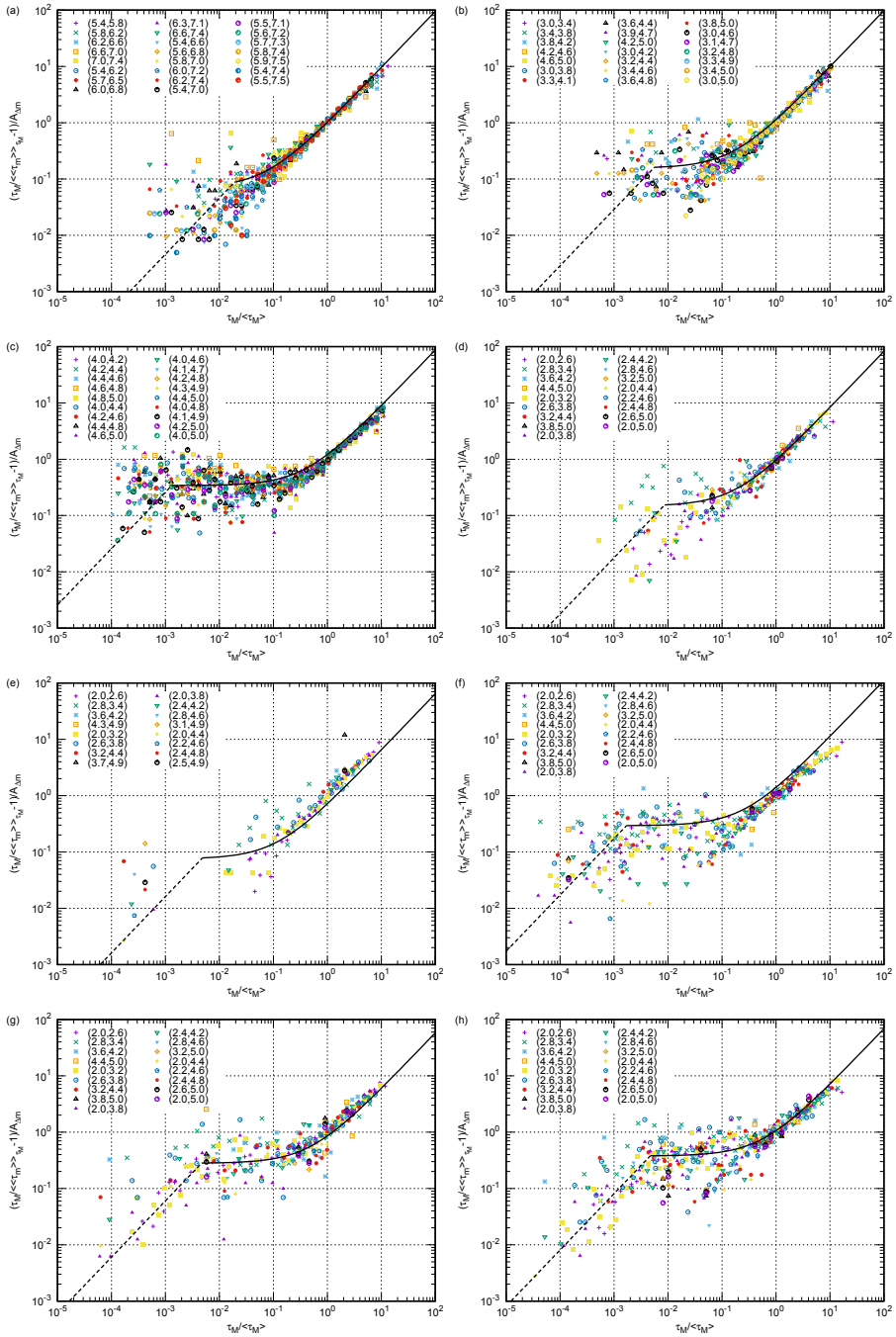

Fig. 8. (Color Online) Results of rescaled $[\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M} - 1]$ for some pairs of specific cut-off magnitudes (m, M) for (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. Equation (23) is also indicated by a black (dotted) line for reference.

Define the transformed upper interval $\sigma_M := \log_{10} \tau_M$. For a set of cut-off magnitudes

(m, M), let $\tilde{N}_M(\sigma_M)$ be the number of inter-event intervals at the upper cut-off magnitude, whose length is within $[\sigma_M, \sigma_M + \Delta\sigma)$. Also, let $\tilde{N}_m(\sigma_M)$ be the total number of inter-event intervals at the lower cut-off magnitude included in these intervals at the upper cut-off magnitude. Then the result shown in Fig. 8 for a specific pair (m, M) is obtained by plotting the following against $10^{\sigma_M+0.5\Delta\sigma}/\langle\tau_M\rangle$:

$$\frac{1}{A_{\Delta m}}\left(\frac{\tilde{N}_m(\sigma_M)}{\tilde{N}_M(\sigma_M)} - 1\right). \tag{20}$$

Figure 8 shows that the average numbers $(\tau_M/\langle\langle\tau_m\rangle\rangle_{\tau_M} - 1)$ computed by the above way collapse on a curve $(\zeta(y))$ independent of the cut-off magnitudes (m, M), after rescaling by $\langle\tau_M\rangle$ and $A_{\Delta m}$, as follows:

$$\frac{\frac{\tau_M}{\langle\langle\tau_m\rangle\rangle_{\tau_M}} - 1}{A_{\Delta m}} = \zeta\left(\frac{\tau_M}{\langle\tau_M\rangle}\right). \tag{21}$$

Hereafter, I use the variable $y := \tau_M/\langle\tau_M\rangle$.

Seismic catalogs have a limitation in the number of events. Therefore, in particular for the combination of (m, M) with large Δm, the number of sample data to calculate Eq. (20) becomes insufficient. This leads to increased fluctuations in the results and makes it difficult to identify a trend. To obtain the scaling function $(\zeta(y))$ utilizing as many sample data as possible, I used all sample data obtained from the combinations of (m, M) with the same Δm together as the population, based on the above result in the following way.

Define the rescaled and transformed upper interval $\sigma := \log_{10} y$. For all (m, M) with the same Δm (i.e., $(m, M) = (m_0, m_0 + \Delta m), (m_1, m_1 + \Delta m), \cdots, (m_n, m_n + \Delta m)$, where $m_j = m_{\min} + 0.1\,j$ ($m_{\min}$ is the minimum magnitude in Table I) and n is the upper limit of j for the given magnitude range and Δm), let $\hat{N}_{m_j+\Delta m}(\sigma)$ be the total number of rescaled and transformed inter-event time intervals of length within $[\sigma, \sigma + \Delta\sigma)$ at the cut-off magnitude $m_j + \Delta m$. Also, let $\hat{N}_{m_j}(\sigma)$ be the total number of inter-event intervals at the cut-off magnitude m_j included in these time intervals at the cut-off magnitude $m_j + \Delta m$. Then the result for Δm is obtained by plotting the following against $10^{\sigma+0.5\Delta\sigma}$:

$$\frac{1}{A_{\Delta m}}\left(\frac{\sum_{j=0}^{n}\hat{N}_{m_j}(\sigma)}{\sum_{j=0}^{n}\hat{N}_{m_j+\Delta m}(\sigma)} - 1\right). \tag{22}$$

Such results for all possible Δm set within the magnitude range in Table I are shown in Fig. 9 by gray + symbols. In particular, the results for several Δm values are indicated by colored symbols. These results are nearly the same as those for the specific pairs of (m, M) in Fig.8, indicating that the scaling property in Eq. (21) holds. The results also suggest that the method using the dependence on the magnitude difference as in Eq. (22) can effectively extract the

scaling function from the limited number of seismic data, as long as the GR law holds.

The following linear functions are used as the fitting function of $\zeta(y)$:

$$\zeta(y) \approx \begin{cases} \alpha y + \beta & (\text{if } y \geq y_0), \\ \gamma y & (\text{if } y < y_0). \end{cases} \tag{23}$$

The parameter values $(\alpha, \beta, \gamma, y_0)$ in Eq. (23) are different for each time series and can be determined as follows. First, β was determined by fitting in the range $y \in [10^{-2}, 10^{-1}]$. Next, as $\zeta(y) \sim \alpha y$ for large y, I determined α by fitting in log-scale $\log_{10}\zeta(y) \sim \log_{10}\alpha + \log_{10} y$, in the range $y \geq 1$. In the same way, γ was determined by fitting in log-scale $\log_{10}\zeta(y) \sim \log_{10}\gamma + \log_{10} y$, in the range $y \leq 10^{-2}$. Lastly, y_0 was determined as the intersection of the above two lines. The resulting parameter values are summarized in Table III, and the fitting curves with these parameter values are shown in Fig. 9 (and also in Fig. 8 for reference).

Table III. Parameter values (to six decimal places) in Eq. (23) for each time series.

Name	α	β	γ	y_0
CMT	0.953828	0.069634	4.603897	0.019077
JS	0.979298	0.156119	28.882991	0.005595
JA	0.869569	0.340167	258.752423	0.001319
SCA1	0.822337	0.147506	17.838524	0.008669
SCA2	0.644104	0.075412	16.157159	0.004861
SCA3	1.106579	0.292309	172.511206	0.001705
SCS1	0.575125	0.279903	60.274019	0.004689
SCS2	0.681038	0.377327	81.000735	0.004698

3.1.2 Functional form of $\Psi_{mM}(i|\tau_M)$

Figure 10 shows the component $\Psi_{mM}(i|\tau_M)$ for $i \leq 5$ and $\Delta m = 0.5$, 1.0, and 1.5 for the case of CMT. The cases of other time series are shown in Figs. S1-(1) – S1-(7).[49] In the figure, the variable τ_M is scaled by the average $\langle\tau_M\rangle$. Only the small Δm results are shown because, when Δm becomes large, the number of samples for calculating the distribution is reduced, and it seemed insufficient to obtain clear outcomes. The results in Fig. 10 were obtained in the following way.

Let $\tilde{N}_M(i|\sigma_M)$ be the number of transformed time intervals at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$), whose length is within $[\sigma_M, \sigma_M + \Delta\sigma)$, and each of them includes i pieces of inter-event intervals at the lower cut-off magnitude. $\Psi_{mM}(i|\tau_M)$ can be drawn by plotting

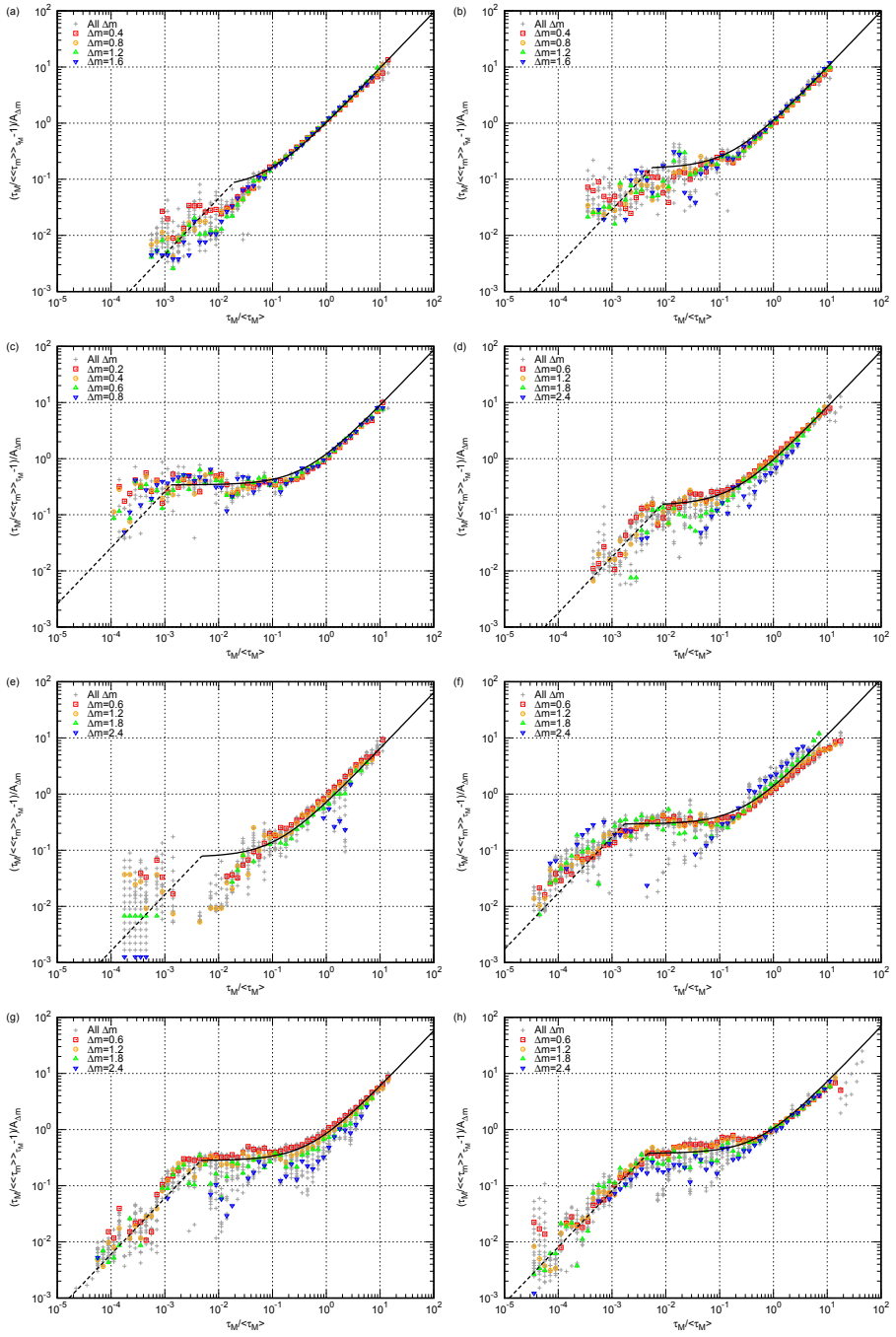

Fig. 9. (Color Online) Results of $[\tau_M/\langle\langle\tau_M\rangle\rangle_{\tau_M}-1]$ calculated by Eq. (22) for the mixed populations of (m, M) with the same Δm. Results of all Δm are shown by gray + symbols. The black solid line shows $\alpha y+\beta$ for $y \geq y_0$ and the black dotted line γy for $y < y_0$ with the parameter values in Table III.

the following against $10^{\sigma_M+0.5\Delta\sigma}$:

$$\frac{\tilde{N}_M(i|\sigma_M)}{\tilde{N}_M(\sigma_M)}. \tag{24}$$

The results shown in Figs. 10 and S1(1) – S1(7)[49] were obtained by further scaling the x-

axis value by the average, that is, plotting Eq. (24) against $10^{\sigma_M+0.5\Delta\sigma}/\langle\tau_M\rangle$. In the figures, the results for several pairs of (m, M) are shown. The figures suggest that the probabilities $\Psi_{mM}(i|\tau_M)$ of the same Δm and i collapse on nearly the same function ($\hat{\Psi}_{\Delta m}(i|y)$) after rescaling by $\langle\tau_M\rangle$:

$$\Psi_{mM}(i|\tau_M) = \hat{\Psi}_{\Delta m}(i|y)\,. \tag{25}$$

Based on this scaling property, I calculated $\hat{\Psi}_{\Delta m}(i|y)$ by taking into consideration all the sample data with the same Δm and y, likewise Sect. 3.1.1. Let $\hat{N}_{m_j+\Delta m}(i|\sigma)$ be the number of rescaled and transformed time intervals ($\sigma = \log_{10} y$) of length within $[\sigma, \sigma+\Delta\sigma)$ at the upper cut-off magnitude ($m_j + \Delta m$), each of which includes i pieces of inter-event intervals at the lower cut-off magnitude m_j. $\hat{\Psi}_{\Delta m}(i|y)$ was also drawn in Figs. 10 and S1 – S7[49)] by plotting the following against $10^{\sigma+0.5\Delta\sigma}$ by + symbols:

$$\frac{\sum_{j=1}^{n}\hat{N}_{m_j+\Delta m}(i|\sigma)}{\sum_{j=1}^{n}\hat{N}_{m_j+\Delta m}(\sigma)}. \tag{26}$$

The results of Eq. (26) resemble those for particular combinations of (m, M) and thus support the scaling property in Eq. (25).

The figures also indicate that $\Psi_{mM}(i|\tau_M)(=\hat{\Psi}_{\Delta m}(i|y))$ can be described empirically by the following negative binomial distribution with the scaling function $\zeta(y)$:

$$\begin{aligned}\Psi_{mM}(i|\tau_M) = {} & \frac{\Gamma(i-1+B_{\Delta m}\zeta(y))}{(i-1)!\Gamma(B_{\Delta m}\zeta(y))} \\ & \times\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1},\end{aligned} \tag{27}$$

where,

$$B_{\Delta m} = \frac{A_{\Delta m}}{b\Delta m\,(b\Delta m+2)}.$$

In the figures, Eq. (27) was drawn with the b-values listed in Table I and the parameter values listed in Table III. One can see that Eq. (27) describes the numerical results of $\Psi_{mM}(i|\tau_M)$, though it tends to deviate at the tail.

Note that Eq. (27) becomes the following Poisson distribution as $\Delta m \to 0$:

$$\Psi_{mM}(i|\tau_M) \xrightarrow[\Delta m\to 0]{} \frac{(A_{\Delta m}\zeta(y))^{(i-1)}}{(i-1)!}e^{-A_{\Delta m}\zeta(y)}. \tag{28}$$

In particular, when $\zeta(y) = y$ is identical to Eq. (16). However, such consistency does not hold for $\Delta m \gneqq 0$, and there is room for improvement in this respect.

Thus, the weak inter-event correlations cause two modifications in $\Psi_{mM}(i|\tau_M)$ compared to the stationary Poisson process. First, the distribution form of $\Psi_{mM}(i|\tau_M)$ changes to a negative

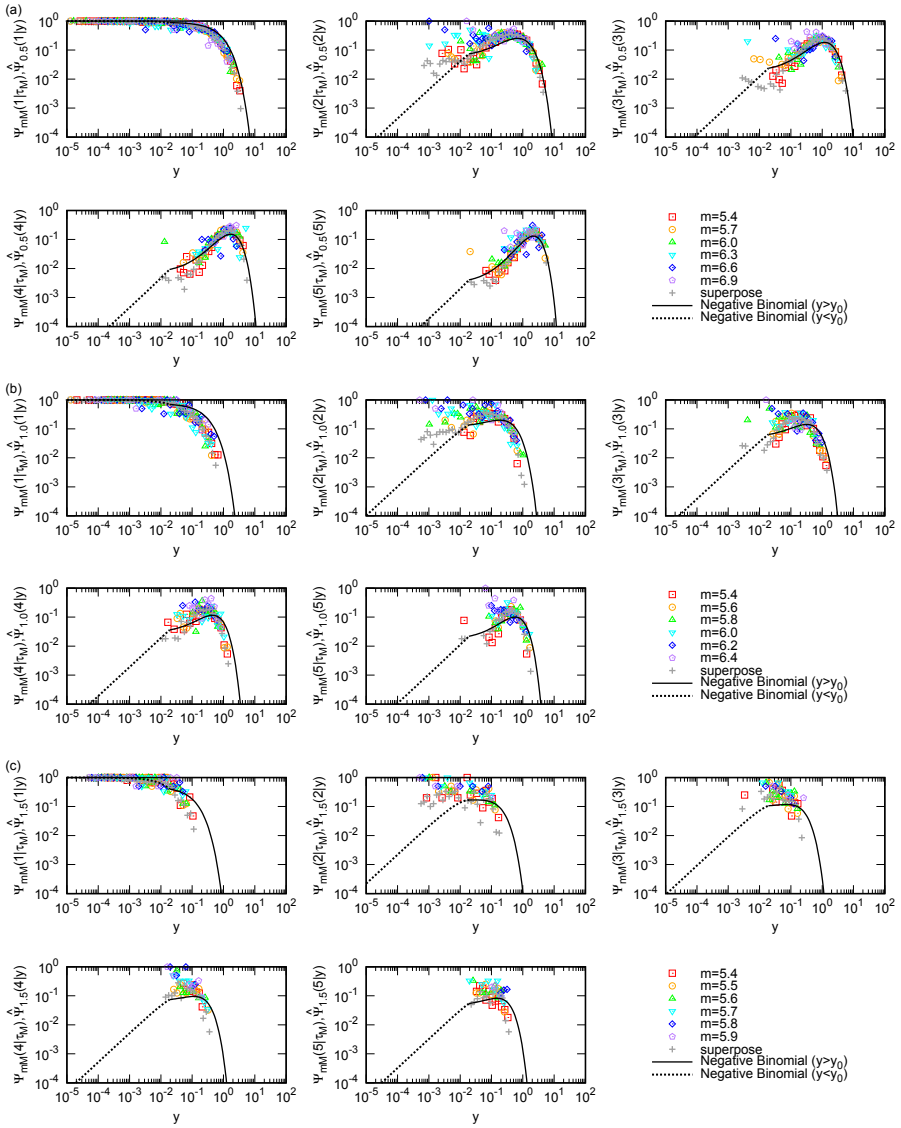

Fig. 10. (Color Online) $\Psi_{mM}(i|\tau_M)$ and $\hat{\Psi}_{\Delta m}(i|y)$ with $i \leq 5$ for CMT, plotted against $y(:= \tau_M/\langle\tau_M\rangle)$. The cases of (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$ are shown. The results of $\Psi_{mM}(i|\tau_M)$ for several specific pairs of cut-off magnitude (m, M) are shown by open-colored symbols. The gray + symbols show $\hat{\Psi}_{\Delta m}(i|y)$ for mixed populations of (m, M) with the same Δm. (Dotted) curves show Eq. (27) with the parameter values listed in Tables I and III.

binomial distribution for $\Delta m > 0$. Second, the scaling function $\zeta(y)$ deviates from $\zeta(y) = y$. In particular, when $\Delta m \to 0$, $\zeta(y)$ represents a perturbation from the stationary Poisson process caused by the weak correlations.

3.2 Results for the component $\rho_{mM}(\tau_m|2,\tau_M)$ and the assumption regarding $\rho_{mM}(\tau_m|i,\tau_M)$ with $i \geq 3$

As for the component $\rho_{mM}(\tau_m|i,\tau_M)$, I quantify its change from Eq. (17) by using $dh^{(i)}_{mM}(\tau_m,\tau_M)$ defined as in Eq. (29).

$$\rho_{mM}(\tau_m|i,\tau_M) = \frac{(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2} \times \left(1+dh^{(i)}_{mM}(\tau_m,\tau_M)\right)\theta(\tau_M-\tau_m). \tag{29}$$

Note that, as pointed out in Sect. 3.1.1, the average of the conditional probability is determined only by the component $\Psi_{mM}(i|\tau_M)$. Therefore, the changes $dh^{(i)}_{mM}(\tau_m|\tau_M)$ affect the higher order moments than 1 in the conditional probability.

In this study, I examined only the case of $i = 2$. By the normalization condition and the symmetry of $\rho_{mM}(\tau_m|2,\tau_M)$, $dh^{(2)}_{mM}(\tau_m,\tau_M)$ must satisfy the followings:

$$\int_0^{\tau_M} dh^{(2)}_{mM}(\tau_m,\tau_M)d\tau_m = 0,$$

$$dh^{(2)}_{mM}(\tau_m,\tau_M) = dh^{(2)}_{mM}(\tau_M-\tau_m,\tau_M).$$

I also examined the distribution function of $\rho_{mM}(\tau_m|2,\tau_M)$:

$$P_{mM}(\tau_m|2,\tau_M) := \int_0^{\tau_m} \rho_{mM}(s|2,\tau_M)ds,$$

which is τ_m/τ_M for the stationary marked Poisson process.

Figure 11 shows the numerical results of $\rho_{mM}(\tau_m|2,\tau_M)$ and $P_{mM}(\tau_m|2,\tau_M)$ for CMT. The latter was scaled by τ_M values in the x-axis direction. The results for other time series are shown in Figs. S2-(1) – S2-(7).[49] The results were obtained in the following way.

Define the transformed lower interval $\sigma_m := \log_{10}\tau_m$. Let $\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)$ be the number of the transformed inter-event time intervals at the lower cut-off magnitude, whose length falls within $[\sigma_m,\sigma_m+\Delta\sigma)$ and each of them is included in the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) such that its length is within $[\sigma_M,\sigma_M+\Delta\sigma)$ and it includes only two ($i = 2$) inter-event intervals at the lower cut-off magnitude. Also, let $\tilde{N}_m(\sigma_M|i=2)$ be the total number of inter-event intervals at the lower cut-off magnitude, included in the inter-event intervals at the upper cut-off magnitude such that its transformed length is within $[\sigma_M,\sigma_M+\Delta\sigma)$ and includes two intervals at the lower cut-off magnitude. Finally, I denote the density $\rho_{mM}(\tau_m|2,\tau_M)$ after applying the above variable

transformations by $\tilde{\rho}_{mM}(\sigma_m|2,\sigma_M)$, which is calculated as follows:

$$\tilde{\rho}_{mM}(\sigma_m|2,\sigma_M) = \frac{\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)}{\tilde{N}_m(\sigma_M|i=2)\Delta s}. \tag{30}$$

$\rho_{mM}(\tau_m|2,\tau_M)$ is obtained by re-transforming this by $\sigma_m \mapsto \tau_m$ and $\sigma_M \mapsto \tau_M$,

$$\rho_{mM}\left(\tau_m|2,\tau_M\right) = \frac{1}{10^{\sigma_m+0.5\Delta s}\ln 10}\frac{\tilde{N}_{mM}(\sigma_m;\sigma_M|i=2)}{\tilde{N}_m(\sigma_M|i=2)\Delta s}. \tag{31}$$

$\rho_{mM}(\tau_m|2,\tau_M)$ in Figs 11 and S2(1) – S2(7)[49] were drawn by plotting Eq. (31) against $10^{\sigma_m+0.5\Delta s}$ and $10^{\sigma_M+0.5\Delta s}$. Also, $P_{mM}(\tau_m|2,\tau_M)$ can be calculated by adding $\tilde{\rho}_{mM}\left(\sigma_m|2,\sigma_M\right)\Delta s$ and was plotted against $10^{\sigma_m+\Delta s}/10^{\sigma_M+0.5\Delta s}$ and $10^{\sigma_M+0.5\Delta s}$.

As shown in these figures, the densities $\rho_{mM}\left(\tau_m|2,\tau_M\right)$ are nearly the same regardless of (m,M). In addition, $\rho_{mM}\left(\tau_m|2,\tau_M\right) \approx 1/\tau_M$, suggesting that $dh^{(2)}_{mM}(\tau_m,\tau_M) \approx 0$. However, as shown in the lower panels of the figures, $dh^{(2)}_{mM}(\tau_m,\tau_M)$ tends to take a positive value for small τ_m. This tendency seems to reflect the temporal clustering of moderate aftershocks in stationary time series and secondary aftershocks in transformed aftershock sequences.

$dh^{(2)}(\tau_m,\tau_M) \approx 0$ implied by the figures reflects the weak inter-event correlations in the time series; the situation seems to be the same for $i \geq 3$, that is, $dh^{(i)}_{mM}(\tau_m,\tau_M) \approx 0$ for $i \geq 3$. Thus, in the following, I assume that inter-event correlations in these time series are sufficiently weak so as not to affect the components $\rho_{mM}(\tau_m|i,\tau_M)$ and there is no change in $\rho_{mM}(\tau_m|i,\tau_M)$ (i.e., $dh^{(i)}_{mM}(\tau_m,\tau_M) \approx 0$ for $i \geq 2$) from the stationary marked Poisson process. Note that this assumption should be re-considered when dealing with time series that include prominent (secondary) aftershocks.

4. Derivation of the probability density functions related to inter-event times

The analysis in the previous section yielded the following results and suggestions for unfolding-transformed seismic time series: the component $\Psi_{mM}(i|\tau_M)$ can be empirically described by the negative binomial distribution (Eq. (27)) with the scaling function $\zeta(y)$ (which can be fit by the linear functions in Eq. (23)) in the newly found scaling property (Eq. (21)), and the other component $\rho_{mM}(\tau_m|i,\tau_M)$ is almost unaffected by weak inter-event correlations and assumed to be the same as in the stationary Poisson process. Based on them, this section derives the functional forms of probability densities in Bayes' theorem (Eq. (2)) and checks whether the derived functions can describe the probability densities.

4.1 Conditional probability

First, I derive the conditional probability. The conditional probability rescaled by the transformations $\tau_m \mapsto x := \tau_m/\langle\tau_M\rangle$ and $\tau_M \mapsto y$ ($\hat{p}_{mM}(x|y)$) is derived as in Eq. (32) (see

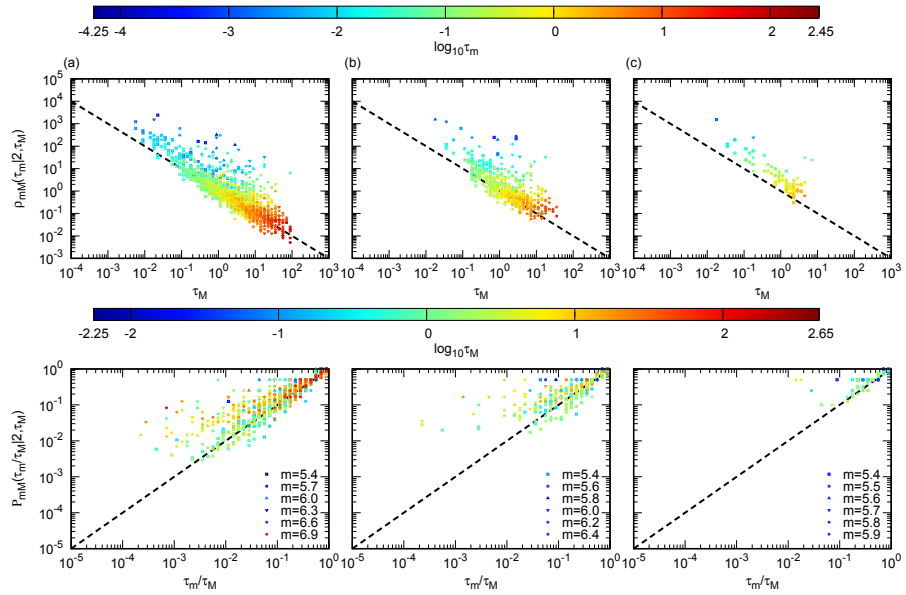

Fig. 11. (Color Online) Numerical results of (upper panels) $\rho_{mM}(\tau_m|2,\tau_M)$ and (lower panels) $P_{mM}(\tau_m|2,\tau_M)$ for several pairs of (m,M) with (a) $\Delta m = 0.5$, (b) $\Delta m = 1.0$, and (c) $\Delta m = 1.5$, for CMT. The upper (lower) panels show $\rho_{mM}(\tau_m|2,\tau_M)$ ($P_{mM}(\tau_m|2,\tau_M)$) in a cross-section parallel to the τ_M (τ_m/τ_M) axis. The colors in the upper (lower) panels indicate τ_m (τ_M) values. The dotted black line in the upper (lower) panels indicates $\rho_{mM}(\tau_m|2,\tau_M) = 1/\tau_M$ ($P_{mM}(\tau_m|2,\tau_M) = \tau_m/\tau_M$), which corresponds to $dh^{(2)}_{mM}(\tau_m,\tau_M) = 0$ in Eq. (29).

Appendix A). Hereafter, variable x is used with this definition.

$$\hat{p}_{mM}(x|y) = (A_{\Delta m}\zeta(y)+1)^{-1}\left\{\delta(y-x)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\right.$$
$$+\theta(y-x)\frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y]}{(A_{\Delta m}x+B_{\Delta m}y)^2}$$
$$\left.\times A_{\Delta m}B_{\Delta m}\zeta(y)\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)}\right\}. \quad (32)$$

As Eq. (32) indicates, the rescaled conditional probability depends on Δm.

To compare with Eq. (32), I numerically obtained the conditional probability similarly to Sect. 3.2, though the condition of $i = 2$ is removed. Let $\tilde{N}_{mM}(\sigma_m;\sigma_M)$ be the number of the transformed inter-event time intervals at the lower cut-off magnitude ($\sigma_m = \log_{10}\tau_m$), whose length is within $[\sigma_m, \sigma_m+\Delta\sigma)$ and each of them is included in the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) of length within $[\sigma_M, \sigma_M+\Delta\sigma)$. This represents the total number of pairs of σ_m and σ_M in the time series. Also, let $\tilde{p}_{mM}(\sigma_m|\sigma_M)$ be the probability density of σ_m under the condition that it is included in the inter-event time interval of transformed length σ_M at the upper cut-off magnitude. Thus, $\tilde{p}_{mM}(\sigma_m|\sigma_M)$ is

calculated as follows:

$$\tilde{p}_{mM}(\sigma_m|\sigma_M) = \frac{\tilde{N}_{mM}(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_M)\Delta\sigma},$$

and $p_{mM}(\tau_m|\tau_M)$ is:

$$p_{mM}(\tau_m|\tau_M) = \frac{1}{10^{\sigma_m+0.5\Delta\sigma}\ln 10}\frac{\tilde{N}_{mM}(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_M)\Delta\sigma}. \tag{33}$$

$\hat{p}_{mM}(x|y)$ can be drawn by plotting the following against (x, y):

$$\hat{p}_{mM}(x|y) = \langle\tau_M\rangle p_{mM}(\tau_m|\tau_M). \tag{34}$$

Figure 12 shows $\hat{P}_{mM}(x|y)$ defined below:

$$\hat{P}_{mM}(x|y) := \int_x^\infty \hat{p}_{mM}(x'|y)dx',$$

numerically calculated for the time series of CMT. In the figure, $\hat{P}_{mM}(x|y)$ is shown in cross-section views for various y ranges. The figure suggests that the conditional probability for the same Δm collapse on the same function, that is,

$$\hat{p}_{mM}(x|y) = \hat{p}_{\Delta m}(x|y). \tag{35}$$

This is consistent with the implication by Eq. (32). The same scaling property was also found for other time series, as shown in Figs. S3-(1) – S3-(7).[49)]

Based on the scaling property in Eq. (35), I calculated the conditional probability using all the pairs of cut-off magnitudes with the same Δm. Define the rescaled and transformed interval $\varsigma := \log_{10} x$, and $\sigma = \log_{10} y$. Let $\hat{N}_{m_j,m_j+\Delta m}(\varsigma;\sigma)$ be the total number of the inter-event time intervals of rescaled and transformed length within $[\varsigma, \varsigma + \Delta\sigma)$ at the cut-off magnitude m_j, that is included in the inter-event time intervals of rescaled and transformed length within $[\sigma, \sigma + \Delta\sigma)$ at the cut-off magnitude $m_j + \Delta m$. The conditional probability for Δm ($\overline{p}_{\Delta m}(x|y)$) is calculated as follows:

$$\overline{p}_{\Delta m}(x|y) = \frac{1}{10^{\varsigma+0.5\Delta\sigma}\ln 10}\frac{\sum_{j=0}^n \hat{N}_{m_j,m_j+\Delta m}(\varsigma;\sigma)}{\sum_{j=0}^n \hat{N}_{m_j}(\sigma)\Delta\sigma}. \tag{36}$$

In Fig. 12, also $\overline{P}_{\Delta m}(x|y)$ defined as follows is shown.

$$\overline{P}_{\Delta m}(x|y) := \int_x^\infty \overline{p}_{\Delta m}(x'|y)dx'.$$

$\overline{P}_{\Delta m}(x|y)$ is almost consistent with $\hat{P}_{mM}(x|y)$, supporting the scaling property in Eq. (35) and the superposition method in Eq. (36).

Figure 13 shows the results of Eq. (36) plotted against $10^{\varsigma+0.5\Delta\sigma}$ and Eq. (32) with the parameter values listed in Table III at a given value of Δm. The results for other Δm values are summarized in Figs. S4-(1) – S4-(8).[49)] The figures show that the results of catalog analyses

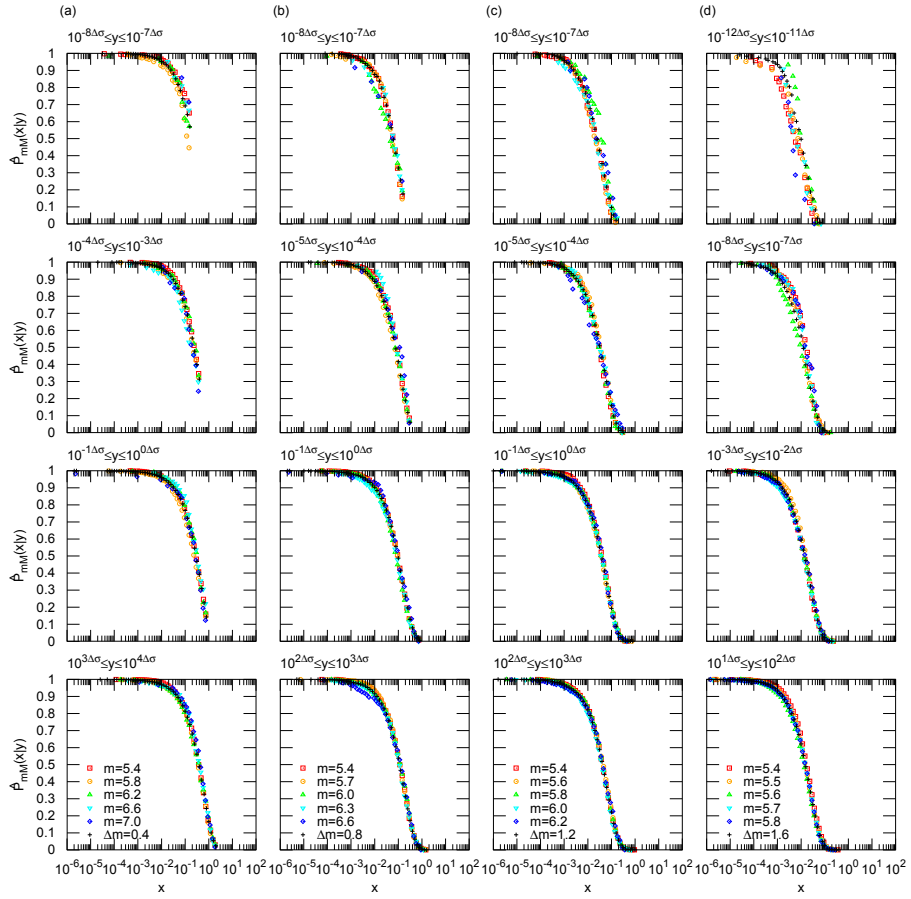

Fig. 12. (Color Online) $\hat{P}_{mM}(x|y)$ for the time series of CMT. Results for certain pairs of cut-off magnitudes (m, M) with (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$, are drawn in cross-sections for the ranges of y shown in the figure. Results for the pairs of (m, M) are represented by colored open symbols. Black + symbols show $\overline{P}_{\Delta m}(x|y)$ where $y = 10^{(j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$. Among the j values for which $\overline{P}_{\Delta m}(x|y)$ is obtained, the values are denoted such that all results for the pairs of (m, M) displayed in the legend are included within $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$ by $j = j_0, j_1, \cdots, j_K$. The ranges of y shown in the figure are $(10^{j\Delta\sigma}, 10^{(j+1)\Delta\sigma})$ for j values nearest to the four points which are dividing $[j_0, j_K]$ into five equal parts.

are described by Eq. (32).

4.2 Inter-event time distribution

Second, I derived the scaling function ($f(y)$ in Eq. (5)) of the inter-event time distribution. Under the assumption that $dh^{(i)}_{mM}(\tau_m, \tau_M) = 0$ for $i \geq 2$, the scaling functions $f(y)$ and $\zeta(y)$ are

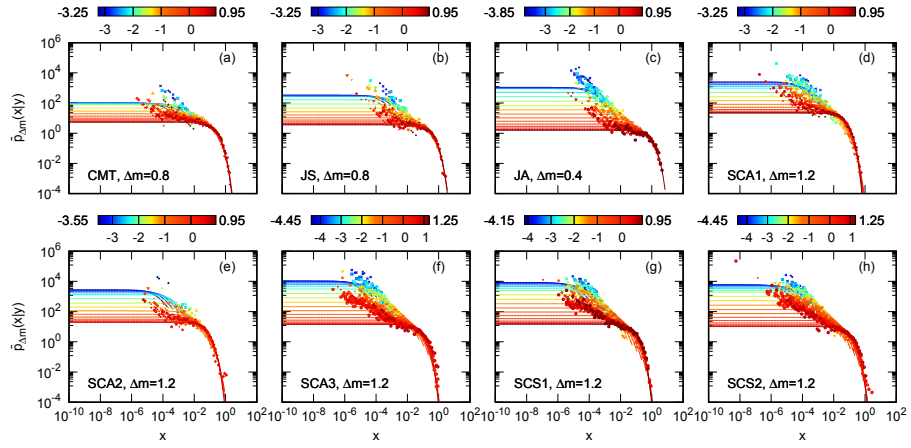

Fig. 13. (Color Online) Equation (32) calculated with the parameter values from Table III (curves) and numerical results of $\overline{p}_{\Delta m}(x|y)$ by catalog analysis (symbols) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The value of Δm is shown in the figure. Results are drawn in cross-sections for $y = 10^{(3j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$. The color corresponds to the $\log_{10} y$ value, and its range is shown in the color bar. Note that when the result of catalog analysis is not obtained for a j, only the corresponding Eq. (32) is shown.

shown to be equivalent, and using Eq. (23) (see Appendix B),

$$f(y) = \begin{cases} \left(\dfrac{y_0^\beta e^{\alpha y_0}}{\frac{e^{\gamma y_0}-1}{\gamma} + \frac{y_0^\beta e^{\alpha y_0}}{\alpha^{1-\beta}} \Gamma(1-\beta, \alpha y_0)} \right) y^{-\beta} e^{-\alpha y} & (\text{for } y \geq y_0), \\ \left(\dfrac{e^{\gamma y_0}}{\frac{e^{\gamma y_0}-1}{\gamma} + \frac{y_0^\beta e^{\alpha y_0}}{\alpha^{1-\beta}} \Gamma(1-\beta, \alpha y_0)} \right) e^{-\gamma y} & (\text{for } y < y_0). \end{cases} \tag{37}$$

Here $\Gamma(\cdot, \cdot)$ is the upper incomplete gamma function.

Figure 14 compares Eq. (37) with the rescaled inter-event time distributions for each time series. Equation (37) was drawn using the parameter values in Table III.

Equation (37) appears to be relatively complicated. However, Fig. 14 also shows that the overall shape, except for very short intervals, of the scaled distribution can be described by the gamma distribution in Eq. (38) that is the extension of $f(y)$ in Eq. (37) for $y \geq y_0$ to $y \in [0, \infty)$ with re-normalization:

$$f(y) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} y^{-\beta} e^{-\alpha y}. \tag{38}$$

4.3 Inverse probability

Third, I derived the inverse probability by substituting the results in Sects. 4.1 and 4.2 into Bayes' theorem (Eq. (2)). The rescaled inverse probability ($\hat{p}_{Mm}(y|x)$) was derived as follows

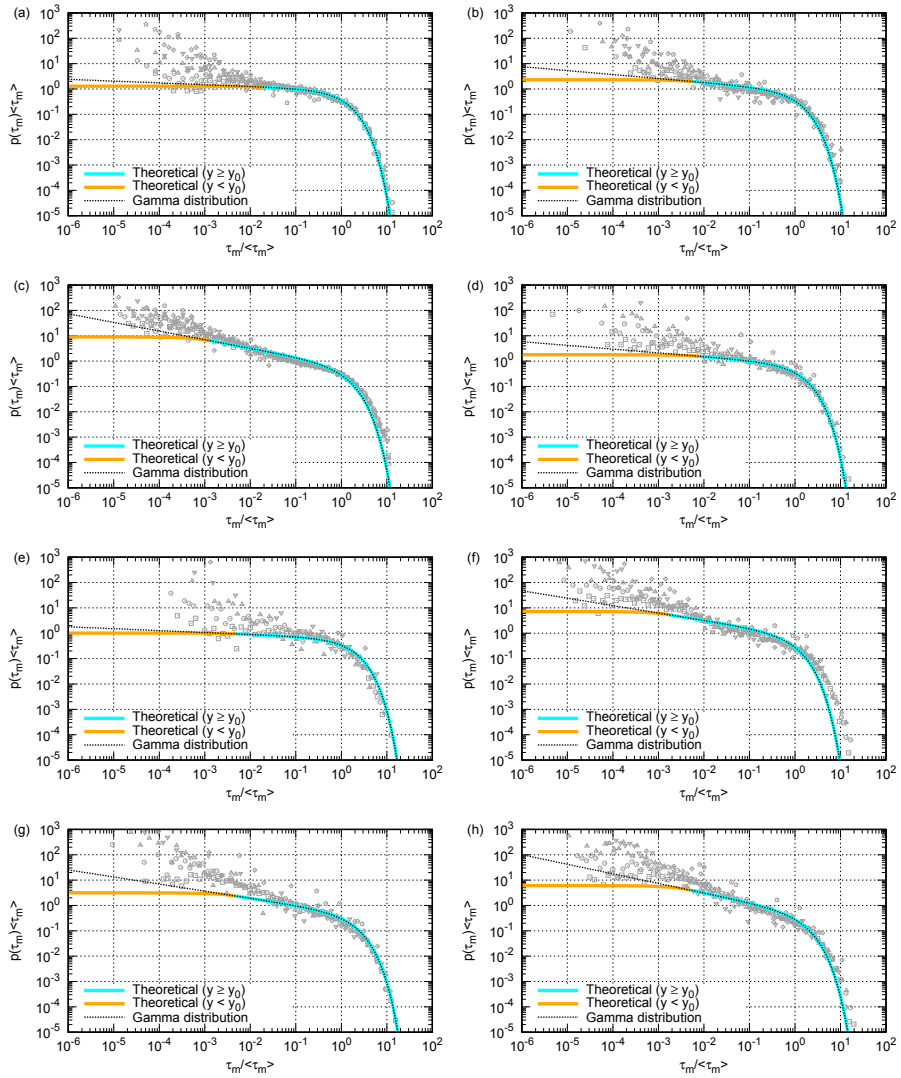

Fig. 14. (Color Online) Equations (37) (orange curve for $y < y_0$ and cyan curve for $y \geq y_0$) and (38) (black dotted curve) as well as numerical results of the rescaled inter-event time distributions by catalog analysis (gray symbols, each symbol represents a result at the same cut-off magnitude as in Fig. 7) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. Equations (37) and (38) are calculated using the parameter values listed in Table III.

(see Appendix C):

$$\hat{p}_{Mm}(y|x) = \frac{\left(\frac{y}{x}\right)^{-\beta} e^{-\alpha(y-(A_{\Delta m}+1)x)}}{(A_{\Delta m}+1)^{2-\beta}} \left\{ \delta(y-x) \left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}} \right)^{B_{\Delta m}\zeta(y)} \right.$$
$$+\theta(y-x) \frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y) + A_{\Delta m}(x+y) + 2B_{\Delta m}y]}{(A_{\Delta m}x + B_{\Delta m}y)^2}$$
$$\left. \times A_{\Delta m}B_{\Delta m}\zeta(y) \left(\frac{B_{\Delta m}y}{A_{\Delta m}x + B_{\Delta m}y} \right)^{B_{\Delta m}\zeta(y)} \right\}. \quad (39)$$

As with the conditional probability, $\hat{p}_{Mm}(y(> x)|x)$ depends on Δm.

As in Sect. 4.1, I numerically computed the inverse probability from catalog data as follows. Let $\tilde{p}_{Mm}(\sigma_M|\sigma_m)$ be the probability density of the transformed inter-event time interval at the upper cut-off magnitude ($\sigma_M = \log_{10}\tau_M$) under the condition that inside of which the transformed time interval of length $\sigma_m(= \log_{10}\tau_m)$ is found. Thus, $\tilde{p}_{Mm}(\sigma_M|\sigma_m)$ is calculated as follows:

$$\tilde{p}_{Mm}(\sigma_M|\sigma_m) = \frac{\tilde{N}_m(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_m)\Delta\sigma},$$

and $p_{Mm}(\tau_M|\tau_m)$ is:

$$p_{Mm}(\tau_M|\tau_m) = \frac{1}{10^{\sigma_M+0.5\Delta\sigma}\ln 10}\frac{\tilde{N}_m(\sigma_m;\sigma_M)}{\tilde{N}_m(\sigma_m)\Delta\sigma}. \tag{40}$$

Thus the scaling of Eq. (40) yields $\hat{p}_{Mm}(y|x)$:

$$\hat{p}_{Mm}(y|x) = \langle\tau_M\rangle p_{Mm}(\tau_M|\tau_m), \tag{41}$$

and,

$$\hat{P}_{Mm}(y|x) := \int_y^\infty \hat{p}_{Mm}(y'|x)dy',$$

is shown for the case of CMT in Fig. 15 as sectional diagrams for several ranges of x. The results for other time series are shown in Figs. S5-(1) – S5-(7).[49] The results suggest the property:

$$\hat{p}_{Mm}(y|x) = \hat{p}_{\Delta m}(y|x), \tag{42}$$

as implied by Eq. (39).

As the Δm dependence was confirmed, I calculated $\hat{p}_{\Delta m}(y|x)$ taking into consideration all the pair of (m, M) with equal Δm as follows:

$$\overline{p}_{\Delta m}(y|x) = \frac{1}{10^{\sigma+0.5\Delta\sigma}\ln 10}\frac{\sum_{j=0}^n \hat{N}_{m_j,m_j+\Delta m}(\varsigma;\sigma)}{\sum_{j=0}^n \hat{N}_{m_j}(\varsigma)\Delta\sigma}. \tag{43}$$

Figure 15 shows $\overline{P}_{\Delta m}(y|x) := \int_y^\infty \overline{p}_{\Delta m}(y'|x)dy'$ and its consistency with $\hat{P}_{Mm}(y|x)$ supports Eqs. (42) and (43).

The results of Eq. (43) are shown in Fig. 16 against $10^{\sigma+0.5\Delta\sigma}$, along with Eq. (39) substituting the parameter values in Table III. Figure 16 shows only the results for a particular Δm; see Figs. S6-(1) – S6-(8)[49] for the results at other Δm values. It can be seen that Eq. (39) illustrates the result of catalog analysis for individual time series.

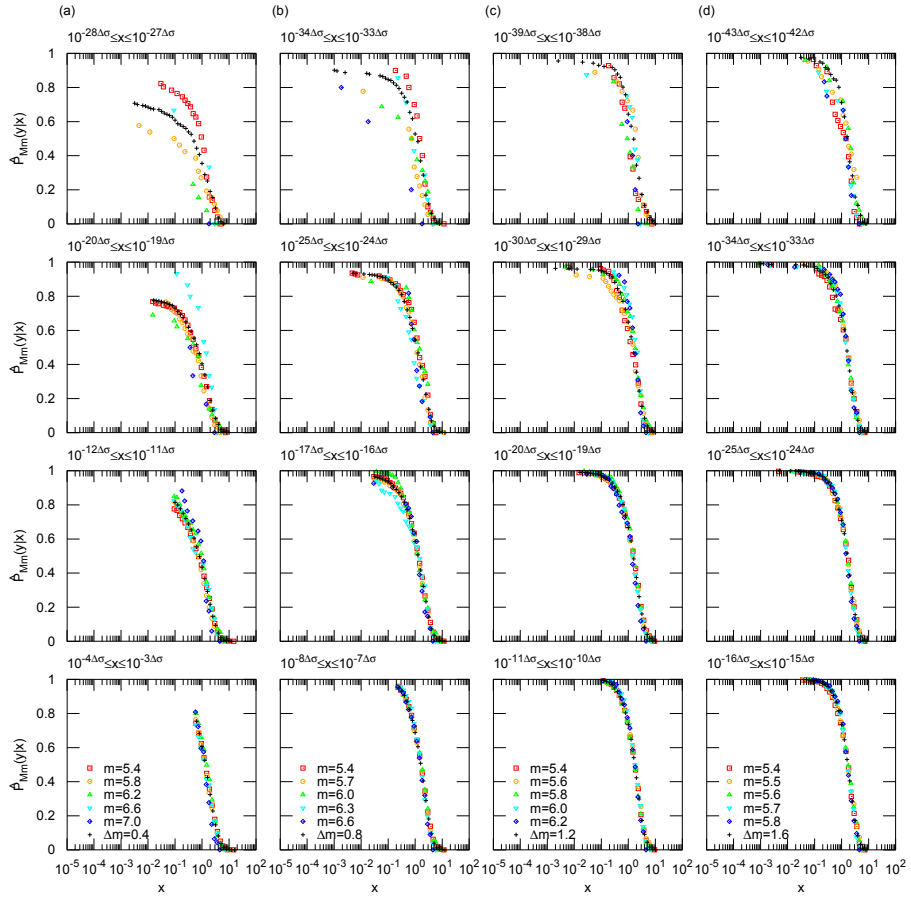

Fig. 15. (Color Online) $\hat{P}_{Mm}(y|x)$ for the time series of CMT. Results for some pairs of (m, M) with (a) $\Delta m = 0.4$, (b) $\Delta m = 0.8$, (c) $\Delta m = 1.2$, and (d) $\Delta m = 1.6$ are shown by colored symbols in cross-sections in the range of x indicated in the figure. Also, $\overline{P}_{\Delta m}(y|x) := \int_y^\infty \overline{p}_{\Delta m}(y'|x)dy'$, where $x = 10^{(j+0.5)\Delta\sigma}$, $j \in \mathbb{Z}$ is shown by black + symbols. The x ranges indicated in the figure are determined by utilizing the same method as outlined in the caption of Fig. 12.

5. Discussion and Conclusions

This study examined the Bayesian forecasting framework of earthquakes' timing using three seismic catalogs toward its practical use. Following Corral's method, I extracted stationary (almost constant occurrence rate) and aftershock (obeying Omori – Utsu occurrence rate) time series from three seismic catalogs and transformed them into weak inter-event correlation ones. For these unfolding-transformed time series, I examined the simplest form of the Bayesian approach, Bayes' theorem, composed of three distribution functions: the conditional probability to quantify temporal hierarchy in time series, the inverse probability directly related to forecasting, and the inter-event time distribution for which scaling universality was re-examined.

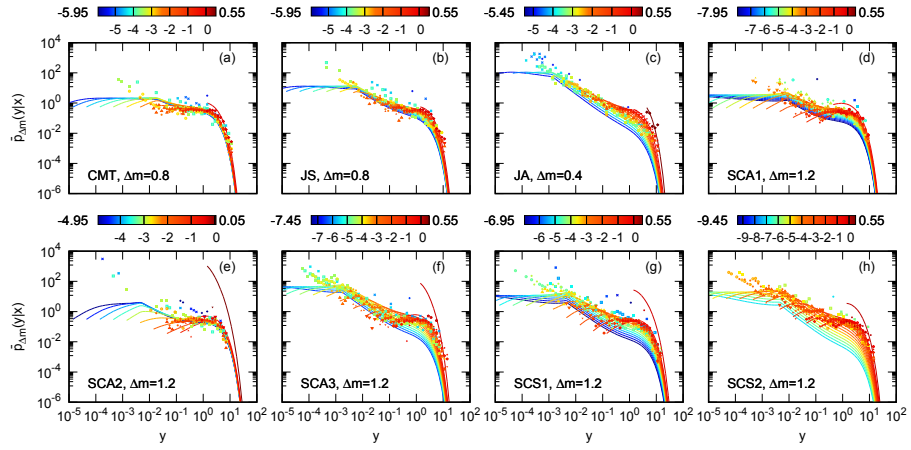

Fig. 16. Equation (39) (curves) with the parameter values in Table III and numerical result of $\overline{p}_{Mm}(y|x)$ by catalog analysis (symbols) for time series of (a) CMT, (b) JS, (c) JA, (d) SCA1, (e) SCA2, (f) SCA3, (g) SCS1, and (h) SCS2. The values of Δm are shown in the figure. The results are drawn in cross-section for $x = 10^{(5j+0.5)\Delta\sigma}$, where $j \in \mathbb{Z}$. The color corresponds to the value of $\log_{10} x$ and its range is shown in the color bar. Note the same possible absence of symbols as in Fig. 13.

A new scaling property on the temporal hierarchy of seismic time series was found by analyzing the average of the conditional probability; the linear functions can fit its scaling function ($\zeta(y)$), though its fitting parameter values differ, reflecting seismic characteristics. The two functional components of the conditional probability were analyzed; one can be empirically described by the negative binomial distribution with the above scaling function ($\zeta(y)$), and the other is almost unchanged from that of the marked stationary Poisson process, which may be attributed to the weakness of inter-event correlation of the unfolding-transformed time series.

Based on them, functional forms of the conditional probability and the scaling function ($f(y)$) of the inter-event time distribution were derived. The theoretical derivation based on the present catalog analysis suggested that the scaling functions for these probability distributions ($\zeta(y)$ and $f(y)$) are essentially equivalent; the difference in the parameter values in the fitting linear functions of $\zeta(y)$ appears as that in the scaling function $f(y)$ of the inter-event time distribution, indicating the negative view of universality. Finally, the inverse probability was derived from Bayes' theorem. These derived distribution functions could describe the characteristics of the results of catalog analysis, such as the functional form and peak emergence in the inverse probability.

5.1 On the scaling universality in the inter-event time distribution in seismicity

The derived scaling function of the inter-event time distribution was described as an exponential distribution for small intervals ($y < y_0$) and a gamma distribution for large intervals ($y \geq y_0$). However, the overall scaling function could be described just by the re-normalized latter gamma distribution. The parameter values to characterize distributions are similar but different for each time series. These results are consistent with the conclusions of the preceding studies in the following respects.

First, in the preceding studies,[11–13] the (generalized) gamma distribution was used to fit the results of catalog analysis, except for short intervals. Theoretical studies under some assumptions on correlations between events derived the distribution near the generalized gamma distribution.[19–21,50] Equation (38) provides another theoretical basis for the fitting instead of such approaches with seismic model time series.

Second, theoretical studies using the ETAS model showed that the universality of the scaling function $f(y)$ is only approximate.[19–21] Our result is also consistent with them; the parameter values in Eq. (37) or (38) were similar but slightly different, reflecting the characteristics of seismic activity aggregated as the difference in $\zeta(y)$.

However, the difference in $f(y)$ for each time series was not as pronounced as in the preceding numerical study.[14] This is because I restricted the analysis only to the stationary time series and transformed aftershock sequences, in accordance with the original Corral's method. The bimodality of the inter-event time distribution pointed out in that study[14] seems to be manifested in the switching of the distribution in Eq. (37), though it could not satisfactorily describe the catalog analysis at short intervals ($y < y_0$). A detailed study of $dh^{(i)}_{mM}(\tau_m, \tau_M)$, set to 0 in this study, may clarify this point.

These results on the scaling universality in the inter-event time distribution from the viewpoint of the unfolding transformation indicate the following: such temporally unfolded time series are indistinguishable regardless of original seismic activity to the extent that their temporal property can be described in the same gamma distribution with slightly different parameter values aggregating seismic character. This conclusion about the temporal nature of the renewal process at a certain cut-off magnitude was derived in the present study from that of the hierarchical relation at two different cut-off magnitudes, i.e., the new scaling property of the conditional probability. Thus, this study extends the above conclusion to the level of the hierarchical structure of seismic time series; unfolding transformed seismic time series are indistinguishable even at the hierarchical level to the extent that the same functional form of

the conditional probability can describe them.

5.2 On the method of unfolding transformation of seismic time series

The following problems still need to be solved with the method I used in this study. Applying the unfolding procedure requires a smooth function representing the temporal variation trend of occurrence rate. In studies of quantum chaos, such functions are given theoretically.[28] However, this study does not have such a theoretical basis, and the temporal variation trends of occurrence rate are given empirically.

This remains the ambiguity in judging whether an aftershock sequence is moderate and thus regarded as fluctuations in a stationary time series or a clear aftershock sequence. Preceding studies using stochastic models considered all aftershock sequences to be the former; thus, there is no ambiguity in this respect. However, in our approach, a modest aftershock sequence can be regarded as a sequence such that its occurrence rate fluctuates around a smooth temporal variation described by the Omori − Utsu rate or just a fluctuation in a stationary time series; a deviation from a constant occurrence rate. For example, this study considered the secondary aftershocks in JA as the deviation in the unfolding-transformed time series, though they could be considered as other aftershock activities. The fitting parameter values of $\zeta(y)$ vary depending on which of the two is adopted, though at present, there is no clear way to determine.

In this way, the unfolding procedure applied in this study leaves some ambiguity in determining the smooth function to describe the trend of the occurrence rate; our conclusion on the universality is tentative in this respect. For a more detailed discussion, it is necessary to establish an objective method to determine the smooth function of the occurrence rate trend adaptable to seismic activity, such as using the moving average.

5.3 Inter-event correlations reflected in the conditional probability

The kind of inter-event correlations reflected in the conditional probability requires further examination. The temporal clustering represented by the Omori − Utsu law seems the primary one. However, preceding studies[43,51] suggested other inter-event correlations; the conditional probability may reflect such correlations. Clarifying this point is important for the theoretical understanding of the Bayesian approach and the properties of the inter-event time distribution.

5.4 Property of the inverse probability and its significance in forecasting earthquakes

This study derived the inverse probability using the GR law and the new scaling property on the hierarchical structure of seismic time series. Although the derivation was based not on analytically rigorous probabilistic calculations but on empirical fitting, assumption, and approximation, this advances the point estimate of the large-earthquake timing in time series with inter-event correlations in the previous study[3] to probabilistic forecasting.

The derived inverse probability has a peak when Δm is large. Such a feature enables us to narrow the timing of future large earthquakes better than the inter-event time distribution, the prior probability without such a peak. The inverse probability is shown to depend on the fitting parameters of $\zeta(y)$ and the b-value reflecting differences in seismic characteristics by time series. Such difference led to the negative conclusion on the scaling universality, though, on the other hand, it suggests the possibility of effective forecasting using the inverse probability. That is, the inverse probability can reflect the characteristics of seismic history in the space-time domain, including minor-scale ones, and varies depending on the length of the lower interval. Therefore, the inverse probability can provide better forecasting than the existing inter-event time distribution by reflecting additional information on the seismic history and the real-time occurrence of small events.

The preceding study showed that the forecasting is relatively ineffective when the non-stationary Omori – Utsu relaxation process that causes inhibiting steady estimate is dominant in activity.[3] The unfolding procedure may improve forecasting in such a relaxation process. This study suggested that the stationary time series are indistinguishable from the unfolding-transformed aftershock sequences at the hierarchy level, and thus, I expect that the Bayesian inference in the transformed aftershock sequence improves to the same extent as other stationary time domains. Therefore, it is conceivable to apply the unfolding transformation to seismic time series before performing Bayesian inference. In such an approach, it is sufficient to consider a less daunting Bayesian framework for time series with weak inter-event correlations. However, in this approach, the already-mentioned problem of determining the smooth function of the occurrence rate remains; in this respect, it seems important to use the moving average in real time.

In this way, the Bayesian approach is expected to contribute to improving the accuracy of probabilistic forecasting of the earthquake's timing. However, toward its practical use, it is significant to examine the Bayesian method in a way that relaxes the two conditions imposed in this study. First, this study considered Bayes' theorem that yields the inverse probability

given only one length of a lower inter-event time as the condition; its extension to Bayesian updating is essential to utilize more information on multiple intervals at the lower cut-off magnitude. Second, this study derived the inverse probability under the condition that the inter-event correlation is weak. I already discussed the Bayesian approach in combination with the unfolding transformation, in which the theory for the time series with weak inter-event correlation is enough. While this approach is advantageous in the theoretical analysis, there remains difficulty in time series treatment. Another more straightforward approach is extending the framework to a general seismic time series that includes notable aftershock activities, which can avoid such difficulty, is theoretically intriguing, and may improve forecasting because such major aftershocks possibly yield correlations between events that can more effectively work in the Bayesian approach. Thus, further examination of the inverse probability in Bayesian updating using general seismic time series is significant for more practical and effective probabilistic forecasting.

5.5 A method to use limited seismic activity data

The catalog analysis found that the conditional and inverse probabilities collapse on Δm-dependent functions after rescaling by the average inter-event times. I utilized this property to use as much data as possible from the limited seismic data. This method seems helpful for further research into analyzing seismic time series with limited data.

However, I point out that this approach only applies in the magnitude range the GR law strictly holds. This can be seen, in particular, in the case of SCA3; as shown in Fig. 14(f), the tail part of the theoretically derived scaling function ($f(y)$) deviated from the result of catalog analysis (a similar trend is also observed in Figs. 14(c) and (h)). The cause of this deviation can be traced back to the poor fitting parameter estimation in Eq. (23) due to the incomplete scaling collapse of $\zeta(y)$ (Fig. 9(f)), and the fundamental cause is considered to be in the minor discrepancy in the magnitude frequency relative to the exact GR law (which can be seen in the slight trend of $\hat{b}$ in Fig. 2(f)). Thus, while the method in this study is effective in compensating for the shortage of seismic data, it should be noted that the results are sensitive to the accurate following of magnitude frequency to the GR law.

Acknowledgments

The author thanks the following organizations for providing seismic data: Japan Meteorological Agency, Ministry of Education, Culture, Sports, Science and Technology Japan; the Caltech/USGS Southern California Seismic Network (SCSN) DOI:10.7914/SN/CI and the

Southern California Earthquake Data Center (SCEDC) DOI: 10.7909/C3WD3xH1 for the Southern California catalog; the Global CMT project for the CMT catalog; and other cooperating organizations for data provision, for each catalog. The author thanks Y. Aizawa for helpful discussions. This work was supported by JST SPRING, Grant Number JPMJSP2110.

Appendix A: Derivation of the conditional probability

I show the derivation of the conditional probability (Eq. (32)) by substituting Eqs. (27) and (29) into Eq. (15). The denominator of Eq. (15) is, from the property of the negative binomial distribution,

$$\sum_{i=1}^{\infty} i\Psi_{mM}(i|\tau_M) = A_{\Delta m}\zeta\left(y\right) + 1. \tag{A·1}$$

On the other hand, the numerator is,

$$\begin{aligned}\sum_{i=1}^{\infty} i\rho_{mM}(\tau_m|i,\tau_M)\Psi_{mM}(i|\tau_M) &= \delta(\tau_M-\tau_m)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)} \\ &+ \theta(\tau_M-\tau_m)\sum_{i=2}^{\infty}\frac{i(i-1)}{\tau_M}\left(1-\frac{x}{y}\right)^{i-2}\left(1+dh^{(i)}_{mM}(\tau_m,\tau_M)\right) \\ &\times\frac{\Gamma\left(i-1+B_{\Delta m}\zeta\left(y\right)\right)}{(i-1)!\Gamma\left(B_{\Delta m}\zeta\left(y\right)\right)} \\ &\times\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1}.\end{aligned} \tag{A·2}$$

The summation without $dh^{(i)}_{mM}(\tau_m,\tau_M)$ of the right hand side (r.h.s.) of Eq. (A·2) is transformed as follows:

$$\begin{aligned}&\sum_{i=2}^{\infty}\frac{i(i-1)}{\tau_M}\left(1-\frac{\tau_m}{\tau_M}\right)^{i-2}\frac{\Gamma\left(i-1+B_{\Delta m}\zeta\left(y\right)\right)}{(i-1)!\Gamma\left(B_{\Delta m}\zeta\left(y\right)\right)} \\ &\quad\times\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{i-1} \\ &= \frac{\left(\frac{B_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)}{\tau_M\Gamma\left(B_{\Delta m}\zeta\left(y\right)\right)} \\ &\quad\times\left(\sum_{i=0}^{\infty}\frac{\Gamma(i+1+D)}{i!}C^{i+1}+2\sum_{i=0}^{\infty}\frac{\Gamma(i+D)}{i!}C^{i}\right),\end{aligned} \tag{A·3}$$

where,

$$C := \frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\left(1-\frac{x}{y}\right).$$

$$D := 1 + B_{\Delta m}\zeta\left(y\right).$$

The first infinite series in the r.h.s. of Eq. (A·3),

$$S_1 = 2\sum_{i=0}^{\infty} a_i, \text{ where } a_i = \frac{\Gamma(i+D)}{i!}C^i, \tag{A·4}$$

satisfies the ratio test as follows:

$$\lim_{i\to\infty}\left|\frac{a_{i+1}}{a_i}\right| = \lim_{i\to\infty}\left|\frac{i+D}{i+1}\right||C| = \left|\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right|\left|1-\frac{x}{y}\right| < 1.$$

Therefore, the first infinite series converges as follows:

$$\begin{aligned} S_1 &= 2\Gamma(D)\left[1+\frac{D}{1!}C+\frac{D(D+1)}{2!}C^2+\cdots\right] \\ &= 2\Gamma(D)(1-C)^{-D}. \end{aligned} \tag{A·5}$$

The second infinite series,

$$S_2 = \sum_{i=0}^{\infty} b_i, \text{ where } b_i = \frac{\Gamma(i+1+D)}{i!}C^{i+1}, \tag{A·6}$$

also satisfies the ratio test as follows.

$$\lim_{i\to\infty}\left|\frac{b_{i+1}}{b_i}\right| = \lim_{i\to\infty}\left|\frac{i+1+D}{i+1}\right||C| = \left|\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right|\left|1-\frac{x}{y}\right| < 1.$$

Equation (A·6) can be transformed in the following way:

$$\begin{aligned} S_2 &= C\sum_{i=0}^{\infty}\frac{(i+D)\Gamma(i+D)}{i!}C^i \\ &= C\left(\sum_{i=0}^{\infty}\frac{\Gamma(i+1+D)}{i!}C^{i+1} + D\sum_{i=0}^{\infty}\frac{\Gamma(i+D)}{i!}C^i\right) \\ &= CS_2 + CD\Gamma(D)(1-C)^{-D} \end{aligned}$$

Therefore,

$$S_2 = CD\Gamma(D)(1-C)^{-D-1}. \tag{A·7}$$

From Eqs. (A·5) and (A·7),

$$\begin{aligned} S_1+S_2 = B_{\Delta m}\zeta(y)\,\Gamma\left(B_{\Delta m}\zeta(y)\right)&\left[\frac{(A_{\Delta m}+B_{\Delta m})y}{A_{\Delta m}x+B_{\Delta m}y}\right]^{B_{\Delta m}\zeta(y)+1} \\ &\times\frac{\left[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y\right]}{(A_{\Delta m}x+B_{\Delta m}y)}. \end{aligned}$$

Equation (A·3) is rewritten as follows:

$$\begin{aligned} &\frac{A_{\Delta m}B_{\Delta m}\zeta(y)}{\langle\tau_M\rangle}\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)} \\ &\times\frac{\{A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y\}}{(A_{\Delta m}x+B_{\Delta m}y)^2}. \end{aligned}$$

Therefore, the conditional probability when $dh^{(i)}_{mM}(\tau_m, \tau_M) \equiv 0$ can be derived as follows:

$$p_{mM}(\tau_m|\tau_M) = (A_{\Delta m}\zeta(y)+1)^{-1}\left\{\delta(\tau_M-\tau_m)\left(\frac{A_{\Delta m}}{A_{\Delta m}+B_{\Delta m}}\right)^{B_{\Delta m}\zeta(y)}\right.$$
$$+\theta(\tau_M-\tau_m)\frac{[A_{\Delta m}B_{\Delta m}(y-x)\zeta(y)+A_{\Delta m}(x+y)+2B_{\Delta m}y]}{(A_{\Delta m}x+B_{\Delta m}y)^2}$$
$$\left.\times\frac{A_{\Delta m}B_{\Delta m}\zeta(y)}{\langle\tau_M\rangle}\left(\frac{B_{\Delta m}y}{A_{\Delta m}x+B_{\Delta m}y}\right)^{B_{\Delta m}\zeta(y)}\right\}. \tag{A·8}$$

Finally, Eq. (32) is obtained by performing the variable transformation $\tau_m \mapsto x$.

Appendix B: Derivation of the scaling function of the inter-event time distribution

Here, I show the derivation of the scaling function of the inter-event time distribution (Eq. (37)). First, I derive the conditional probability for $\Delta m \to 0$. From Eq. (28), the denominator of Eq. (15) is

$$\sum_{i=1}^{\infty} i\frac{(A_{\Delta m}\zeta(y))^{(i-1)}}{(i-1)!}e^{-A_{\Delta m}\zeta(y)} = A_{\Delta m}\zeta(y)+1. \tag{B·1}$$

On the other hand, the numerator of Eq. (15) is, with $dh^{(i)}_{mM}(\tau_m, \tau_M) \equiv 0$,

$$\theta(\tau_M-\tau_m)\sum_{i=2}^{\infty} i\frac{(i-1)}{\tau_M}\left(1-\frac{x}{y}\right)^{i-2}\frac{(A_{\Delta m}\zeta(y))^{(i-1)}}{(i-1)!}e^{-A_{\Delta m}\zeta(y)}$$
$$+\,\delta(\tau_M-\tau_m)e^{-A_{\Delta m}\zeta(y)}. \tag{B·2}$$

The summation in Eq. (B·2) is rewritten as follows:

$$\sum_{i=0}^{\infty}\frac{(i+2)}{\tau_M}\left(1-\frac{x}{y}\right)^{i}\frac{(A_{\Delta m}\zeta(y))^{(i+1)}}{i!}e^{-A_{\Delta m}\zeta(y)}$$
$$=\frac{1}{\tau_M}A_{\Delta m}\zeta(y)\,e^{-A_{\Delta m}\zeta(y)\frac{x}{y}}$$
$$\times\sum_{i=0}^{\infty}(i+2)\frac{\left[A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)\right]^{i}}{i!}e^{-A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)}$$
$$=\frac{1}{\tau_M}A_{\Delta m}\zeta(y)\,e^{-A_{\Delta m}\zeta(y)\frac{x}{y}}\left[A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)+2\right].$$

Therefore, the conditional probability is

$$p_{mM}(\tau_m|\tau_M) = (A_{\Delta m}\zeta(y)+1)^{-1}\left\{e^{-A_{\Delta m}\zeta(y)}\delta(\tau_M-\tau_m)\right.$$
$$\left.+\frac{1}{\tau_M}A_{\Delta m}\zeta(y)\,e^{-A_{\Delta m}\zeta(y)\frac{x}{y}}\left[A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)+2\right]\theta(\tau_M-\tau_m)\right\}. \tag{B·3}$$

Equation (1) is rewritten into the relationship between the scaling functions (Eq. (B·4))

by Eqs. (5), (B·1) and (B·3) with $N_m/N_M = \langle\tau_M\rangle/\langle\tau_m\rangle = 10^{b\Delta m}$.

$$\begin{aligned}
&10^{2b\Delta m} f\left(10^{b\Delta m}x\right) \\
&= e^{-A_{\Delta m}\zeta(x)} f(x) \\
&\quad + A_{\Delta m}\int_x^\infty \frac{\zeta(y)}{y} e^{-A_{\Delta m}\zeta(y)\frac{x}{y}}\left[A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)+2\right] f(y)dy. \qquad \text{(B·4)}
\end{aligned}$$

Each term in Eq. (B·4) is then followed by operating $\lim_{\Delta m\to 0}\partial_{\Delta m}$:

$$\begin{aligned}
&\partial_{\Delta m} 10^{2b\Delta m} f\left(10^{b\Delta m}x\right) \xrightarrow[\Delta m\to 0]{} b\ln 10\left(2f(x)+xf'(x)\right), \\
&\partial_{\Delta m} e^{-A_{\Delta m}\zeta(x)} f(x) \xrightarrow[\Delta m\to 0]{} -b\ln 10\zeta(x)f(x), \\
&\partial_{\Delta m} A_{\Delta m}\int_x^\infty \frac{\zeta(y)}{y} e^{-A_{\Delta m}\zeta(y)\frac{x}{y}}\left[A_{\Delta m}\zeta(y)\left(1-\frac{x}{y}\right)+2\right] f(y)dy \\
&\qquad\qquad \xrightarrow[\Delta m\to 0]{} 2b\ln 10\int_x^\infty \frac{\zeta(y)}{y} f(y)dy.
\end{aligned}$$

Hence,

$$2f(x)+xf'(x) = -\zeta(x)f(x) + 2\int_x^\infty \frac{\zeta(y)}{y} f(y)dy. \qquad \text{(B·5)}$$

Equation (B·5) indicates the essential equivalence between $f(y)$ and $\zeta(y)$.

Equation (B·5) is simplified using $G(x)$ as defined in Eq. (B·6), as follows:

$$2G(x) = -xG'(x),$$

where,

$$G(x) := f(x) - \int_x^\infty \frac{\zeta(y)}{y} f(y)dy, \qquad \text{(B·6)}$$

Therefore, with a constant κ:

$$G(x) = \frac{\kappa}{x^2},$$

or,

$$f'(x) + \frac{\zeta(x)}{x} f(x) = -\frac{2\kappa}{x^3}.$$

Change the variable from x to y, and the solution is with a constant y_0,

$$f(y) = \left(-\int_{y_0}^y \frac{2\kappa}{u^3} e^{\int_{y_0}^u \frac{\zeta(v)}{v}dv} du + f(y_0)\right) e^{-\int_{y_0}^y \frac{\zeta(v)}{v}dv}.$$

In particular, when $\kappa = 0$:

$$f(y) = f(y_0)\exp\left(-\int_{y_0}^y \frac{\zeta(u)}{u} du\right).$$

Using Eq. (23) suggested by the catalog analysis as $\zeta(y)$,

$$f(y) = \begin{cases} f(y_0) y_0^{\beta} e^{\alpha y_0} y^{-\beta} e^{-\alpha y} & (\text{if } y \geq y_0), \\ f(y_0) e^{\gamma y_0} e^{-\gamma y} & (\text{if } y < y_0), \end{cases} \tag{B·7}$$

where $f(y_0)$ is determined by the normalization condition as

$$f(y_0) = \left(\frac{e^{\gamma y_0} - 1}{\gamma} + \frac{y_0^{\beta} e^{\alpha y_0}}{\alpha^{1-\beta}} \Gamma\left(1-\beta, \alpha y_0\right) \right)^{-1}. \tag{B·8}$$

Hence, by Eqs. (B·7) and (B·8), the scaling function of the inter-event time distribution is derived as Eq. (37).

Appendix C: Derivation of the inverse probability

Here, I show the derivation of the inverse probability from Eq. (2) with Eqs. (38) and (A·8).

From Eqs. (3) and (38), $z_m(\tau_m)$ and $z_M(\tau_M)$ are

$$z_m(\tau_m) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} \frac{1}{\langle \tau_m \rangle} \left\{ (A_{\Delta m} + 1) x \right\}^{1-\beta} e^{-\alpha (A_{\Delta m} + 1) x}, \tag{C·1}$$

$$z_M(\tau_M) = \frac{\alpha^{1-\beta}}{\Gamma(1-\beta)} \frac{1}{\langle \tau_M \rangle} y^{1-\beta} e^{-\alpha y}. \tag{C·2}$$

Also, from Eqs. (4) and (A·8), $z_{mM}(\tau_m | \tau_M)$ is

$$\begin{aligned} z_{mM}(\tau_m | \tau_M) = \left(\frac{x}{y} \right) \Bigg\{ & \delta(\tau_M - \tau_m) \left(\frac{A_{\Delta m}}{A_{\Delta m} + B_{\Delta m}} \right)^{B_{\Delta m} \zeta(y)} \\ & + \theta(\tau_M - \tau_m) \frac{\left[A_{\Delta m} B_{\Delta m} (y - x) \zeta(y) + A_{\Delta m} (x + y) + 2 B_{\Delta m} y \right]}{(A_{\Delta m} x + B_{\Delta m} y)^2} \\ & \times \frac{A_{\Delta m} B_{\Delta m} \zeta(y)}{\langle \tau_M \rangle} \left(\frac{B_{\Delta m} y}{A_{\Delta m} x + B_{\Delta m} y} \right)^{B_{\Delta m} \zeta(y)} \Bigg\}. \end{aligned} \tag{C·3}$$

The inverse probability for $dh_{mM}^{(i)}(\tau_m, \tau_M) \equiv 0$ is derived by substituting Eqs. (C·1), (C·2), and (C·3) into Eq. (2):

$$\begin{aligned} p_{Mm}(\tau_M | \tau_m) = & \frac{\left(\frac{y}{x} \right)^{-\beta} e^{-\alpha (y - (A_{\Delta m} + 1) x)}}{(A_{\Delta m} + 1)^{2-\beta}} \\ & \times \Bigg\{ \delta(\tau_M - \tau_m) \left(\frac{A_{\Delta m}}{A_{\Delta m} + B_{\Delta m}} \right)^{B_{\Delta m} \zeta(y)} \\ & + \theta(\tau_M - \tau_m) \frac{\left[A_{\Delta m} B_{\Delta m} (y - x) \zeta(y) + A_{\Delta m} (x + y) + 2 B_{\Delta m} y \right]}{(A_{\Delta m} x + B_{\Delta m} y)^2} \\ & \times \frac{A_{\Delta m} B_{\Delta m} \zeta(y)}{\langle \tau_M \rangle} \left(\frac{B_{\Delta m} y}{A_{\Delta m} x + B_{\Delta m} y} \right)^{B_{\Delta m} \zeta(y)} \Bigg\}. \end{aligned} \tag{C·4}$$

Finally, Eq. (39) is derived by performing the variable transformation $\tau_M \mapsto y$ to Eq. (C·4).

References

1) C. H. Scholz, *The Mechanics of Earthquakes and Faulting* (Cambridge University Press, 2002) 2nd ed., pp.408-pp.412.

2) The Headquarters for Earthquake Research Promotion [https://www.jishin.go.jp/reports/research_report/choukihyoka_01b/] [in Japanese], last access: December 2023.

3) H. Tanaka and K. Umeno: J. Phys. Soc. Jpn. **to be published** (2023).

4) H. Tanaka and K. Umeno, J. Phys. Soc. Jpn. **92**, 113001 (2023).

5) In previous studies (Refs. 3 and 4), when setting the magnitude threshold M, events with magnitude exactly equals to M were not included. However, in this paper, I include such events and refer to such threshold as the cut-off magnitude. This is because the magnitudes recorded in seismic catalogs are discrete with a certain minimum interval.

6) H. Tanaka and Y. Aizawa: Journal of the Physical Society of Japan **86** (2017) 024004.

7) Y. Ogata, J. Am. Stat. Assoc. **83**, 9 (1988).

8) Y. Ogata, Ann. Inst. Statist. Math. **50**, 379 (1998).

9) F. Omori, J. Coll. Sci., Imp. Univ. Tokyo, Japan **7**, 111 (1894).

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

11) Á. Corral, Phys. Rev. E **68**, 035102 (2003).

12) Á. Corral, Physica A **340**, 590 (2004).

13) Á. Corral, Phys. Rev. Lett. **92**, 108501 (2004).

14) S. Touati, M. Naylor, and I. G. Main, Phys. Rev. Lett. **102**, 168501 (2009).

15) Á. Corral: Tectonophysics **424** (2006) 177.

16) Á. Corral, Statistical features of earthquake temporal occurrence, Modelling Critical and Catastrophic Phenomena in Geoscience, pp. 191–221. Springer, 2006.

17) G. Molchan, Pure Appl. Geophys. **162**, 1135 (2005).

18) S. Hainzl, F. Scherbaum, and C. Beauval, Bull. Seismol. Soc. Am. **96**, 313 (2006).

19) A. Saichev and D. Sornette, Phys. Rev. Lett. **97**, 078501 (2006).

20) A. Saichev and D. Sornette, J. Geophys. Res.: Solid Earth **112**, B04313 (2007).

21) D. Sornette, S. Utkin, and A. Saichev, Phys. Rev. E **77**, 066109 (2008).

22) M. Bottiglieri, L. de Arcangelis, C. Godano, and E. Lippiello, Phys. Rev. Lett. **104**, 158501 (2010).

23) E. Lippiello, Á. Corral, M. Bottiglieri, C. Godano, and L. de Arcangelis, Phys. Rev. E **86**, 066119 (2012).

24) L. de Arcangelis, C. Godano, J. R. Grasso, and E. Lippiello, Phys. Rep. **628**, 1, Chap. 3 (2016).

25) A. Helmstetter and D. Sornette, J. Geophys. Res.:Solid Earth **107**, 2237 (2002).

26) E. Marekova: Acta Geophysica **64** (2016) 885.

27) Y. Ogata and K. Shimazaki: Bulletin of the Seismological Society of America **74** 1757 (1984);Y. Ogata, Tectonophysics **169**, 159 (1989).

28) O. Bohigas and M.-J. Giannoni: Mathematical and computational methods in nuclear physics — (1984) 1.

29) H. Hasegawa, Butsurigaku Saizensen (Physics Frontiers) 28, ed. Y. Otsuki (Kyoritsu Shuppan, Tokyo, 1991) p.16-p.17 [in Japanese]; N. Saito, Butsurigaku Saizensen (Physics Frontiers) 30, ed. Y. Otsuki (Kyoritsu Shuppan, Tokyo, 1992) p.74-p.78 [in Japanese].

30) Data downloaded from [https://www.globalcmt.org/], last access: October 2023.

31) A. M. Dziewonski, T.-A. Chou, and J. H. Woodhouse: Journal of Geophysical Research: Solid Earth **86** (1981) 2825.

32) G. Ekström, M. Nettles, and A. Dziewoński: Physics of the Earth and Planetary Interiors **200** (2012) 1.

33) Data downloaded from [http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html], last access: November 2023; Caltech/USGS Southern California Seismic Network (SCSN) DOI: 10.7914/SN/CI; Southern California Earthquake Data Center (SCEDC) DOI: 10.7909/C3WD3xH1.

34) E. Hauksson, W. Yang, and P. M. Shearer: Bulletin of the Seismological Society of America **102** (2012) 2239.

35) Data downloaded from [http://evrrss.eri.u-tokyo.ac.jp/tseis/jma1/index-j.html], last access : November 2023; Japan Meteorological Agency Earthquake catalog.

36) H. Tsuruoka: IPSJ SIG Tech. Rep. **1998** (1998) 65.

37) D. Amorese: Bulletin of the Seismological Society of America **97** (2007) 1742.

38) B. Gutenberg and C. F. Richter, Bull. Seismol. Soc. Am. **34**, 185 (1944).

39) T. Utsu, Geophys bulletin of the Hokkaido University **13** (1965) 99 [in Japanese].

40) K. Aki: Bull. Earthq. Res. Inst., Tokyo Univ. **43** (1965) 237.

41) S. Wiemer and M. Wyss, Bull. Seismol. Soc. Am. **90**, 859 (2000).

42) D. Schorlemmer and J. Woessner, Bull. Seismol. Soc. Am. **98**, 2103 (2008).

43) E. Lippiello, C. Godano, and L. Arcangelis: Geophysical Research Letters **39** (2012).

44) [https://scedc.caltech.edu/earthquake/significant.html].

45) Y. Y. Kagan, Bull. Seismol. Soc. Am. **94**, 1207 (2004).

46) A. Helmstetter, Y. Y. Kagan, and D. D. Jackson, Bull. Seismol. Soc. Am. **96**, 90 (2006).

47) Y. Tohru and O. Atsuyuki, Sogo Toshi Kenkyu (Comprehensive urban studies) **43**, 99 (1991) [in Japanese].

48) D. J. Webb, J. Theor. Biol. **43**, 277 (1974).

49) (Supplemental Material) Refer to the figures provided in the supplemental material for further information not shown in the main text.

50) A. Corral: Physical review letters **95** (2005) 028501.

51) E. Lippiello, L. de Arcangelis, and C. Godano: Physical review letters **100** (2008) 038501.