Diagnostic analysis of space-time branching processes for earthquakes

Jiancang Zhuang[1], *Yosihiko Ogata*[1] *and David Vere-Jones*[2]

[1]*Institute of Statistical Mathematics*

4-6-7 Minami Azabu, Minato-Ku, Tokyo 106-8659, Japan

[2] *Victoria University of Wellington*

P.O.Box 600, Wellington, New Zealand

February 27, 2004

Abstract

This paper discusses several methods to test the goodness-of-fit of a space-time continuous-type branching process with immigration, namely the ETAS (epidemic type aftershock sequence) model, to the occurrences of earthquakes, including: 1) using thinning method to decompose the whole process into a background process and its complementary process, to test the stationarity of the background process, 2) using the residual analysis to test the overall goodness-of-fit, and 3) using the stochastic reconstruction method to compare the characteristics of clustering features between the earthquake data and the model.

1 Introduction

It is natural to use the branching process to describe the occurrence pattern of earthquakes, which is clustered in both space and time. These clustering features are important for seismological studies. For the purpose of long term earthquake prediction, such as zoning, earthquake potential estimation, people try to remove the temporary clustering to estimate the background seismicity; on the other hand, for the short-term or real-time prediction, we need a good understanding of the earthquake clusters. Thus separating the background seismicity from the earthquake clusters is considered to be of central importance.

A number of methods have been proposed for declustering a catalogue or separating the background and the clustering seismicity. Most of these methods remove the earthquakes in a space-time window around a large event called the mainshock; the differences among these methods relate to the choice of the window sizes and some other detailed procedures: for example, see Gardner and Knopoff, 1974; Keilis-Borok and Kossobokov, 1986; Utsu, 1969. In general, the bigger is the magnitude of the mainshock, the larger is the window. An alternative to the window-based declustering methods is the link method based on a space-time distance between

the events (e.g., Reseanberg, 1985; Frohlich and Davis, 1990; Davis and Frohlich, 1991).

All of those conventional methods contain arbitrary parameters, whether for defining the sizes of the aftershock window or the link distances, in both space and time. Each choice of parameters gives a different declustered catalogue, and so a different estimate of the background seismicity. A trouble with the conventional declustering algorithms is thus to make an optimal choice of the parameters. For a data set from a specific (rather narrow) seismic region, and a given threshold magnitude, these choices are usually tuned by the user's subjective impression of the declustered output. Also, given a data set, the concept of what constitutes an aftershock has not been uniquely defined. These are among the reasons why so many declustering algorithms have been proposed.

Alternative to these deterministic declustering methods, ideas of probability treatment to the background component and clustering component first appeared in Kagan and Knoppoff (1978). The assumed a constant background occurrence rate throughout the whole studied region, which is in fact location dependent as we can see in this article. Zhuang et al. (2002, 2004) proposed the stochastic declustering method and brought the probability based declustering method into practice. The core of the stochastic declustering method is an iterative approach to simultaneously estimate the background intensity, assumed to be a function of space but not of time, and the parameters associated with clustering structures. Making use of these estimates and the thinning operation, one can obtain the probabilities for each event being a background event or a triggered event. These probabilities are the key to realizing stochastic versions of the clustering family trees in the catalogue, and of course to separating the background events from the earthquake clusters.

Because these probabilities are estimated through a particular model, the closeness between the formulation of the model and the reality is the essentially important factor influencing the output. The closer is the model to the real data, the more reliable is the output. Of course, some model selection procedures can be used to choose the best model among many models fitted to the same set of data. But these procedures usually give us a number indicating the overall goodness-of-fit for the model, and rarely tell us whether there are some good points in a model if its overall fit is worse than another. It is also difficult to find clues about how to improve the formulation of the clustering models through model selection procedures. In this paper, we are going to show some graphical diagnostic methods useful for improving model formulation.

2 The space-time ETAS model

In the empirical studies, the Omori law (Omori, 1898; and see, Utsu et al., 1995, for a review), describes the decay of aftershock frequencies with time, i.e.,

$$n(t) = \frac{K}{(t+c)^p}, \tag{1}$$

which $n(t)$ is the occurrence rate of events at the time t after the occurrence of the mainshock, K, c and p are constants. Another commonly accepted empirical law is

the Gutenberg-Ritcher law, describing the relationship between the magnitudes and occurrence frequencies of earthquakes and taking the form of

$$\log_{10} N(\geq M) = a - bM, \tag{2}$$

i.e., the number of earthquake events decreases exponentially when we increase the magnitude threshold.

These above empirical laws has been considered as the foundation for statistical modelling in seismicity. Several point process models, in different forms (Ogata, 1988, 1998; Kagan, 1991; Musmeci and Vere-Jones, 1992; Rathbun, 1993; Console and Marru, 2001; Console et al., 2003). Among them, Ogata (1998) suggested a model with a component of Omori type (inverse power law) decay in time and a component for the locations of triggered events, independent from the time component. In general, all of those models classify the seismicity into two components, the background and the cluster, where each earthquake event, no matter it is a background event or generated by another event, produces (triggers) its own offspring (aftershocks) according to some branching rules.

All these models can be formulated in the form of the conditional intensity function, i.e., the process is controlled by an intensity conditional on the observation history (see, e.g., Daley and Vere-Jones, 2002, Chapter 7, for details of mathematical treatment on the conditional intensity), namely,

$$\lambda(t, x, y, M|\mathcal{H}_t)dt\,dx\,dy\,dM = E[\boldsymbol{N}(dt\,dx\,dy\,dM)|\mathcal{H}_t], \tag{3}$$

where $\mathcal{H}_t$ is the observation history up to time t. In this study, we base our analysis on the formulation of the space-time epidemic type aftershock sequence (ETAS) models by Ogata (1998),

$$\lambda(t, x, y, M) = \lambda(t, x, y|\mathcal{H}_t)J(M) \tag{4}$$

$$\lambda(t, x, y|\mathcal{H}_t) = \mu(x, y) + \sum_{i:\, t_i<t} \kappa(M_i)g(t - t_i)f(x - x_i, y - y_i|M_i), \tag{5}$$

where

1. $\mu(x, y)$ is the background intensity, which is a function of space, but not of time;

2. $\kappa(M)$ is the expected number of events triggered from an event of a magnitude M, in the form
$$\kappa(M) = A\exp[\alpha(M - M_C)], \tag{6}$$
where A and α are constant, and M_C is the magnitude threshold (see Utsu, 1969; Yamanaka and Shimazaki, 1990);

3. $g(t)$ is the probability density function of the occurrence times of the triggered events, taking the form
$$g(t) = \frac{p-1}{c}\left(1 + \frac{t}{c}\right)^{-p}, \tag{7}$$
i.e., the p.d.f form of the modified Omori law (Utsu, 1969);

4. $f(x, y|M)$ is the location distribution of the triggered events, which is formulated in either of two ways, a short range Gaussian decay and a long range inverse power decay, or explicitly,

$$f(x, y|M) = \frac{1}{2\pi De^{\alpha(M-M_C)}} \exp\left[-\frac{x^2+y^2}{2De^{\alpha(M-M_C)}}\right] \tag{8}$$

for the short-range decay, and

$$f(x, y|M) = \frac{(q-1)D^{(q-1)}e^{\alpha(q-1)(M-M_C)}}{\pi\left[x^2+y^2+De^{\alpha(M-M_C)}\right]^q}; \quad q > 1, \tag{9}$$

for the long-range decay; and

5. $J(M)$ is the probability density of magnitudes of all the events, independent from other components and taking the form of the Gentenburg-Richter law,

$$J(M) = \beta e^{-\beta(M-M_c)}, \text{ for } M \geq M_c \tag{10}$$

where β is linked with the Gutenberg-Richter's b-value by $\beta = b \log 10$ and M_c is the magnitude threshold considered.

In the text below, we call an ETAS model equipped with (8) as Model I, and Model II if equipped with (9). The expected number of offspring that an event can trigger, namely $\kappa(M)$, is also called its *triggering ability*.

In (6), (8) and (9), the spatial scaling factor for the direct aftershock region is proportional to the triggering ability of the ancestor. This judgment is from Kanamori and Anderson (1975) when they tried to explain the exponential law for the number of aftershocks. In the coming section, we will use the stochastic reconstruction method to verify whether it is a good choice.

Thinning procedure One of the key points of the stochastic declustering method is the thinning operation on a point process determined by (5), which can split the whole process into several sub-processes (Lewis and Shedler 1979; Ogata, 1981). Suppose that events are numbered in time order from 1 to N. The probability of an event j being triggered by the ith event is

$$\rho_{ij} = \begin{cases} \dfrac{\zeta_i(t_j, x_j, y_j)}{\lambda(t_j, x_j, y_j|\mathcal{H}_{t_j})}, & \text{when } j > i, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where

$$\zeta_i(t, x, y) = \kappa(M_i)g(t-t_i)f(x-x_i, y-y_i|M_i), \tag{12}$$

represents the intensity triggered by the ith event. Moreover, the probability of the event j being a triggered event is

$$\rho_j = \sum_{i<j} \rho_{ij}, \tag{13}$$

and the probability that the jth event belongs to the background is

$$\varphi_j = 1 - \rho_j = \frac{\mu(x_j, y_j)}{\lambda(t_j, x_j, y_j|\mathcal{H}_{t_j})}. \tag{14}$$

If we select each event j with probabilities ρ_{ij}, ρ_j or φ_j, we can form up a new process being the triggered process by the ith event, the clustering process or the background process, respectively.

Variable kernel estimates Once a background process is obtained, we can estimate the background intensity by some smoothing techniques. Rather than repeat the thinning procedure and the kernel estimation procedure many times to get an average estimate of the background intensity, we directly estimate the average by weighting, i.e.,

$$\hat{\mu}(x, y) = \frac{1}{T} \sum_i \varphi_i Z_{h_i}(x - x_i, y - y_i), \tag{15}$$

where i runs over all of the events in the whole process, T is the length of the time period of the process, and Z_h is the Gaussian Kernel function with a bandwidth h. The variable bandwidth h_j is determined by

$$h_j = \max\{\epsilon, \inf(r : \boldsymbol{N}[B(x_i, y_i; r)] > n_p)\}, \tag{16}$$

where ϵ is a small real number, $B(x, y; r)$ is the disk centered at (x, y) with a radius of r, and n_p is positive integer, i.e., h_j is the distance to n_pth closest event.

Maximum likelihood estimates Given an estimated intensity function $u(x, y)$, we set

$$\mu(x, y) = \nu u(x, y) \tag{17}$$

for the background rate, where ν is a positive-value parameter, and then obtain the maximum likelihood estimates, $\hat{\theta} = (\hat{\nu}, \hat{A}, \hat{\alpha}, \hat{c}, \hat{p}, \hat{D})$ for Mode I and $\hat{\theta} = (\hat{\nu}, \hat{A}, \hat{\alpha}, \hat{c}, \hat{p}, \hat{D}, \hat{q})$ for Mode II, by maximizing the log likelihood

$$\log L(\theta) = \sum_k \log \lambda_\theta(t_k, x_k, y_k \mid \mathcal{H}_{t_k}) - \int_0^T \iint_S \lambda_\theta(t, x, y|\mathcal{H}_t) dx dy dt \tag{18}$$

where k runs over all the events in the study regions S and the study time interval $[0, T]$ (NOT necessary all the events observed).

Thus, the background seismicity and the parameters in the clustering structures simultaneously can be estimated by an iterative approach (Zhuang et al., 2002). We first assume an initial rough background intensity, calculate the MLE of the branching parameters, calculate φ_j, then use these φ_j to estimate a better evaluation of the background intensity. Replacing the initial background intensity by the newly estimated one, we repeat these procedures until the results convergences.

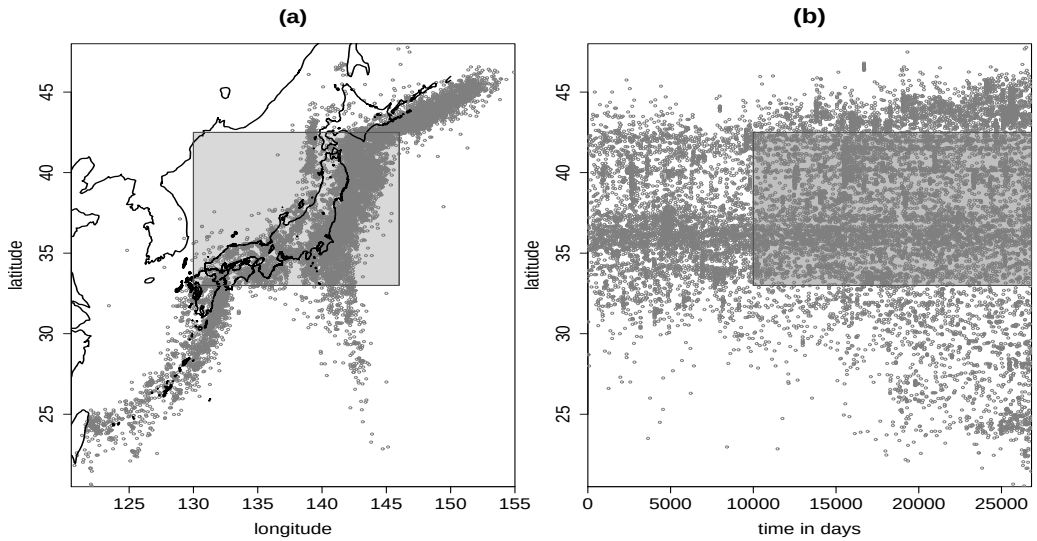

Figure 1: Seismicity in the Japan region and nearby during 1926–1999 ($M_J \geq 4.2$). (a) Epicenter locations; (b) Latitudes of epicenter locations against occurrence times. The shaded region represents the study space-time range.

3 Data and preliminary results

Two datasets are considered in this study. The first dataset is the JMA catalogue. We select the events in the ranges of longitude 121° – 155°E, latitude: 21° – 48°N, depth 0 – 100 km, time 1926/Jan/1 – 1999/Dec/31 and magnitude $\geq M_J 4.2$. There are 19,139 events in this data set.

As other earthquake catalogues covering a long history, completeness and homogeneity are always problems causing troubles for statistical analysis. The incompleteness of the early period and inhomogeneity in this data set can be easily seen from Figures 1. To tackle these problems, we choose a study space–time range, in which the seismicity seems to be relatively complete and homogeneous, with ranges of longitude 130° – 146°E, latitude 33° – 42.5°N, a time period of 10,000 – 26,814 days after 1926/Jan/1, and same depth and magnitude ranges as the whole data set. There are 8283 events occurring in the study space–time range.

The earthquake events not in the target region and time period but in the data set, namely the *boundary catalogue*, are used as the boundary effects in the estimation of the clustering parameters and the background rate in the model, i.e., these events are not counted in computing the likelihood function (18), but are put in the observation history when evaluating the conditional intensity (5).

The second dataset is the Taiwan CWB (Central Weather Bureau) catalogue, we take the events of $M \geq 5.3$ in the period from 1941-1-1 (14975 days from 1900-1-1) to 2001-12-31 and in the region (120∼123°E, 21.5∼25.2°N) for estimating the parameters, non-target events with a magnitude no less than 5.3 in the catalogue are used as boundary events, i.e., they do not directly contribute to the likelihood function, but influence the conditional intensity function. According to the local structures, we then divide each of these three main subregions into smaller ones, as shown in Figure 2.

The background intensity of the study region of JMA catalogue is shown in

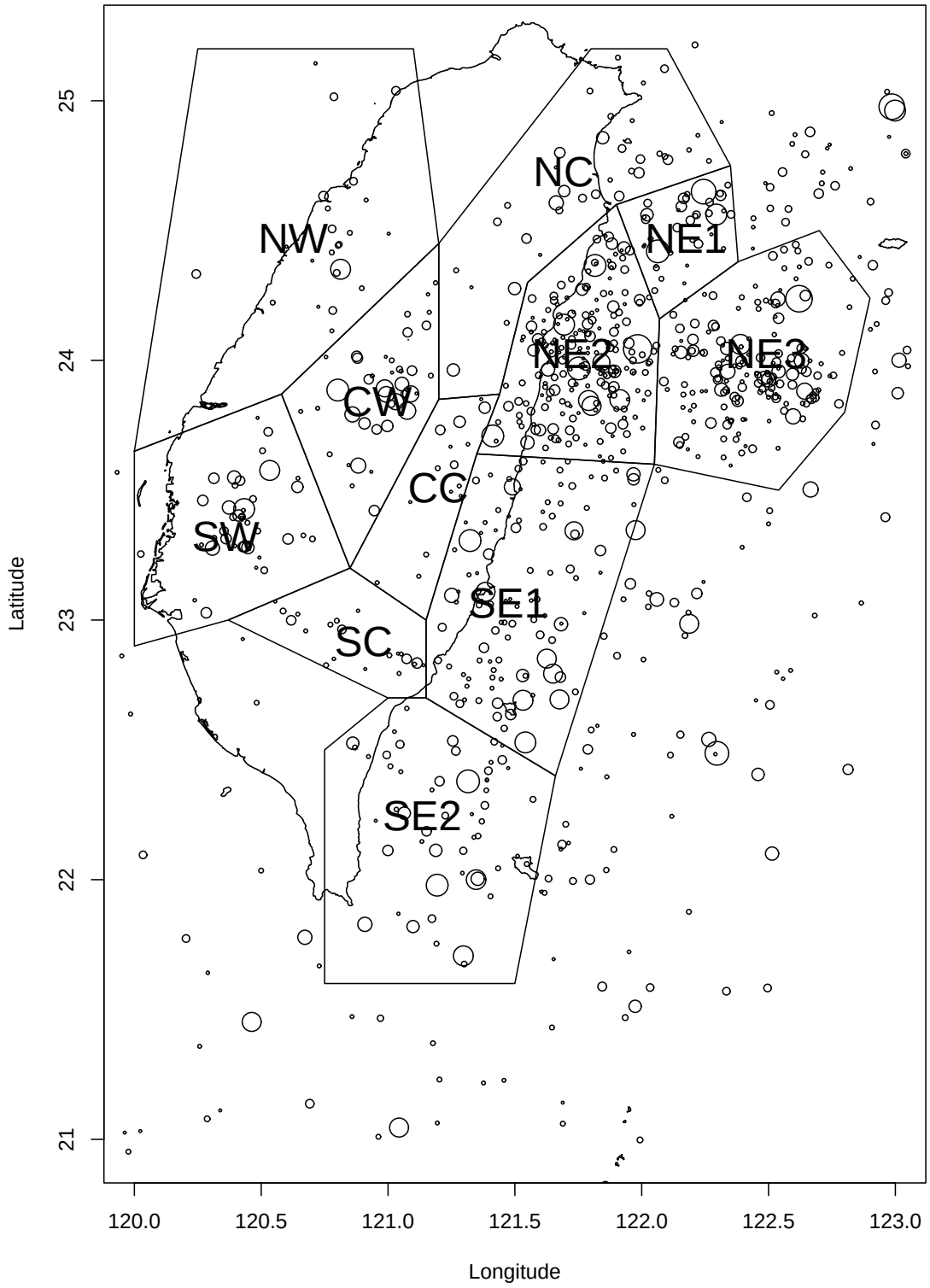

Figure 2: Epicenter locations of earthquakes in the Taiwan region during 1900–2000 ($M \geq 5.3$) and subdivision. Sizes of the circle indicate magnitudes from 5.3 to 8.2.

Figure 3. As a feature complementary to the background, we define

$$\omega(x,y) = 1 - \frac{\mu(x,y)}{\Lambda(x,y)} \tag{19}$$

as the clustering coefficient to measure the clustering effect relative to the total intensity at a location (x,y). This function is as important as $\mu(x,y)$ in earthquake hazard estimation, for damages caused by aftershocks cannot be ignored.

4 Verifying stationarity of the background

In the ETAS model, we assume that the background process is a stationary one. It is not so easy to test the stationarity of this background seismicity by using conventional declustering methods because of the difficulties in choosing declustering parameters. Using the thinning method, we can easily test whether the background seismicity rate changes with the time in a specified region, by constructing the function

$$S(t) = \sum_{t_i<t} \varphi_i, \tag{20}$$

where the subscript i runs over all of the events in this region. The above function $S(t)$ represents the cumulative number of events in the background process in the region considered.

The results of $S(t)$ for some subregions of the Taiwan region are in shown Figure 4, we can see that the earthquake clusters in each subregion are removed, which makes it easier for us to study the background seismicity. A clear feature of the background seismicity in these subregions is the quiescence before the Chichi earthquake occurring subregion CW in 1999. Including Subregion CW itself, all the neighboring subregions show quiescence in seismicity from early 1960 to the occurrence time of the Chichi earthquake. On the other hand, the eastern subregions do not show such a clear quiescence in seismicity.

In summary, the background seismicity seems not to be stationary, and its occurrence rate usually changes before or after great earthquakes.

5 Residual Analysis

The goodness-of-fit of the space-time ETAS model to the earthquake data can be conveniently investigated by using residual analysis. There have been several methods for carrying out the residual analysis. For example, Ogata (2003) re-fitted the data with a time-variant version of the ETAS model and then took the ratio between the conditional intensities of the two models as the residual process. Schoenberg (2003) proposed the thinning residuals, i.e., to keep each event i with probability $\min \lambda(t,x,y|\mathcal{H}_t)/\lambda(t_i,x_i,y_i|\mathcal{H}_{t_i})$ to obtain a homogeneous Poisson process. This method is the essentially the same as Stoyan and Grabarnik (1991). More general discussions on transforming a point process into a Poisson process can be found in Schoenberg (1991), Vere-Jones and Schoenberg (2004).

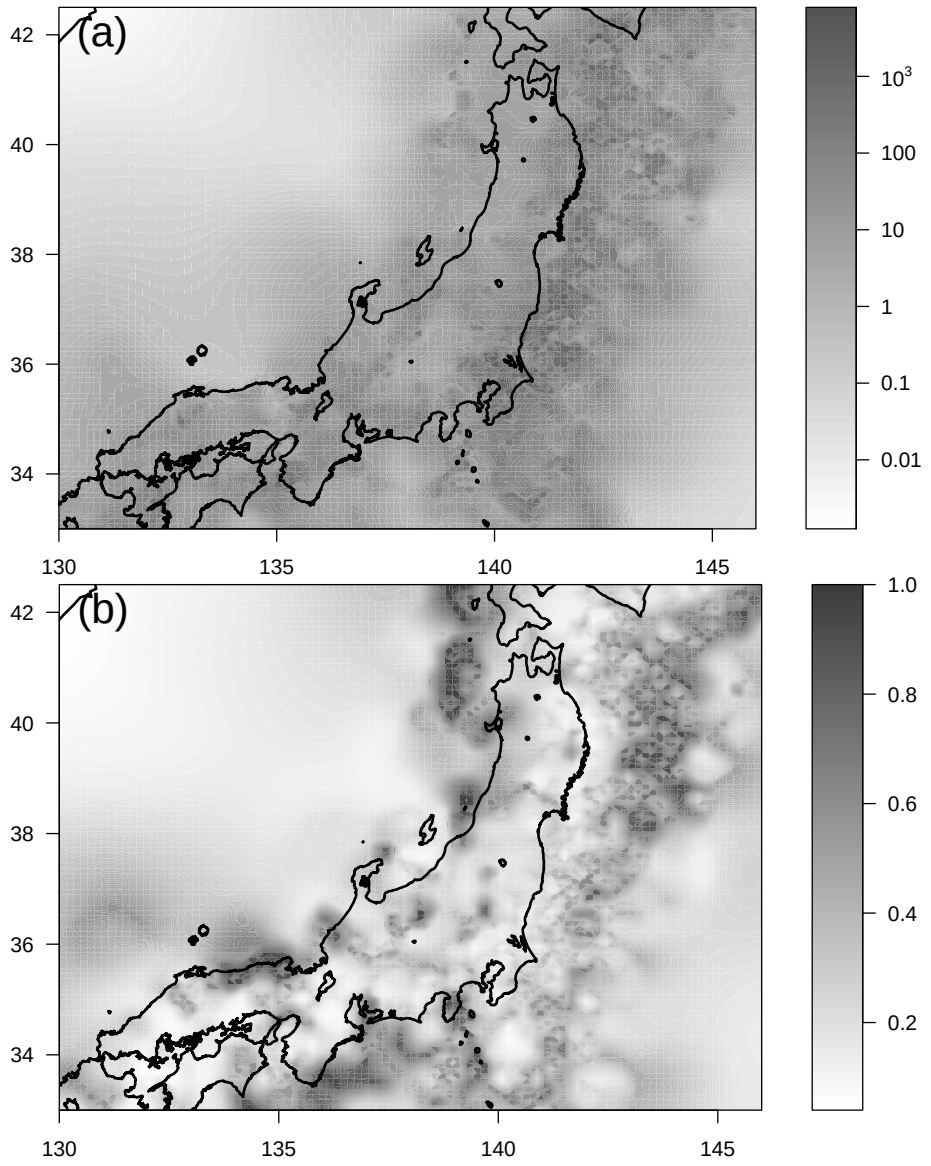

Figure 3: (a) Estimated background rate $\hat{\mu}$ in Equation (15) in the study region (unit: events/(degree^2$\times$74 years)) and (b) the clustering coefficient in Equation (19).

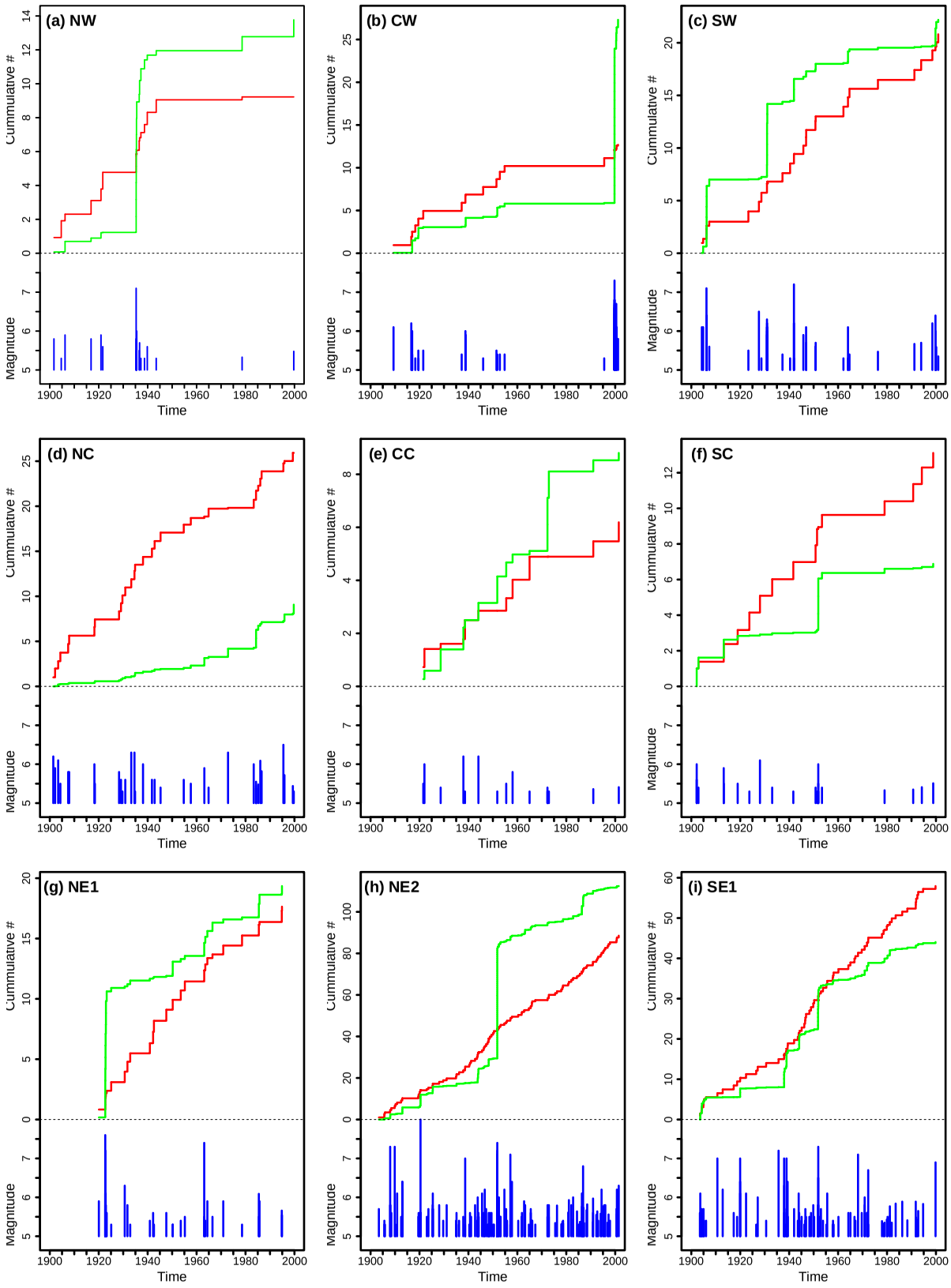

Figure 4: Cumulative background seismicity $S(t)$ in (20) (dark step lines) and clustering seismicity $\#(t) - S(t)$ (gray step lines) in some Taiwan subregions. The magnitudes against the occurrence times of the events are plotted in the lower part of each panel.

Following Stoyan and Grabarnik (1991), to each event, we attach a weight

$$w_i = \frac{1}{\lambda(t_i, x_i, y_i|\mathcal{H}_{t_i})}, \tag{21}$$

then the total weights w_i in a space-time volume B has expectation

$$\boldsymbol{E}[\sum_{(t_i,x_i,y_i)\in B} w_i] = |B|, \tag{22}$$

where $|B|$ denotes the volume of B.

From the residual analysis (Figure 5), we can that there are some deviation in the residual of the data to the ETAS model. In Figure 5(a), the slope of the cumulative wights against the time axis changes at the years around 1940 and 1975, or even around 1994. These changes mainly correspond to the changes of the observation system. On the latitude axis, the cumulative weights change the accumulation rate, showing that this seismic region can be basically divided into three subregion from south to north; while along the longitude axis, the accumulation rate of the weights changes around 120.5, 121.5 and 122.5, showing the difference in seismicity from the west to the east. All these results indicating the difference in clustering seismicity caused by local tectonic environments.

6 Verifying model formulation by stochastic reconstruction

As mentioned in Section 2, if we select each event j with probabilities ρ_{ij}, ρ_j or φ_j, we can get a new process being the triggered process by the ith event, the clustering process or the background processes, respectively. That is to say, we can separate the whole catalogue into different family trees. This method tackles the difficulties in testing hypotheses associated with earthquake clustering features, which are caused by the complicated overlapping of the background seismicity and different earthquake clusters in both space and time. We can repeat such thinning operations for many times to get different stochastic versions of separations of the earthquake clusters. The non-uniqueness of such realizations illustrates the uncertainty in determining earthquake clusters, and thus repetition can help us to evaluate the significance of some properties of seismicity clustering patterns. However, we can also implement these tests by working with the probabilities φ_j and ρ_{ij} directly. In the coming sections, we will show how to use these probabilities to reconstruct the characteristics associated with earthquake clustering features, using the Japanese JMA catalogue as an example.

6.1 Location distributions

Define the standardized distance between a triggered event j and its direct ancestor, assumed i, by

$$r_{ij} = \sqrt{\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{D\exp[\alpha(M_i - M_c)]}}. \tag{23}$$

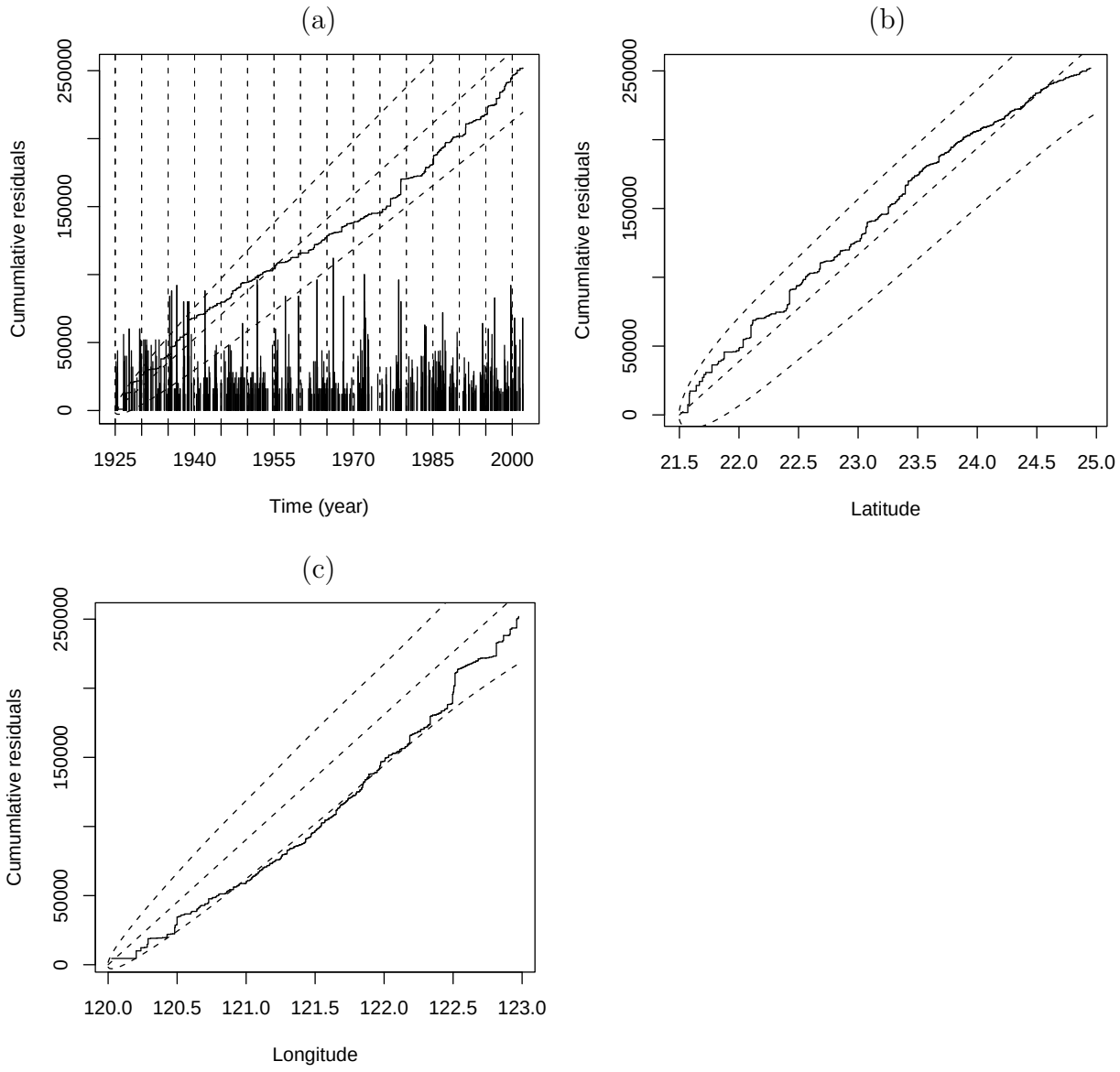

Figure 5: Results from residual analysis for the Taiwan catalogue. The cumulative weights against times, latitudes and longitudes are plotted in (a), (b), and (c), respectively.

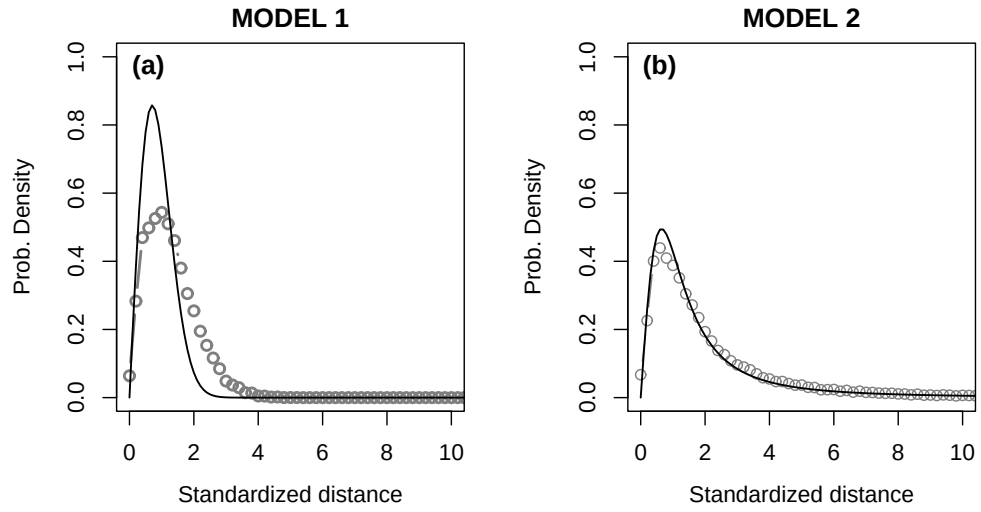

Figure 6: Reconstruction results for the distribution of the standardized triggering distances $\hat{f}_R(r)$ in (26) (gray circles) by using Model I (a) and Model II (b) for the JMA catalogue. The theoretical curves of $f_R(r)$ in (24) and (25) are plotted in solid lines in (a) and (b), respectively.

From (8) and (9), r_{ij} has a density function of

$$f_R(r) = 2re^{-r^2},\ r \geq 0; \tag{24}$$

for Model I and

$$f_R(r) = \frac{2r(q-1)}{(1+r^2)^q},\quad r \geq 0; \tag{25}$$

for Model II, respectively. The distribution with a density of (24) is called a Rayleigh distribution. On the other hand, $f_R(r)$ can be reconstructed through

$$\hat{f}_R(r) = \frac{\sum_{i,j} \rho_{ij} I(|r_{ij} - r| < \Delta r/2)}{\Delta r \sum_{i,j} \rho_{ij}}, \tag{26}$$

where Δr is a small positive number, and I is the index function such that

$$I(x) = \begin{cases} 1, & \text{if the logical statement } x \text{ holds,} \\ 0, & \text{else.} \end{cases} \tag{27}$$

The comparison between $\hat{f}_R$ and f_R for the two models are shown in Figure 6. It can be seen that, if Model I is used, the reconstructed probability density of the transformed distances between the ancestors and the direct offspring is quite different from the theoretical one. When Model II is used, the reconstructed probability density is very close to the theoretical one. These results confirm that the aftershocks decay in a long range in space rather than a short range (Ogata, 1998; Console and Murru, 2001; Console et al. 2003). These results also imply the robustness of the reconstruction method, for we can get a reconstructed probability density function very close to the corresponding function in Model II, even if an improper model like Model I is employed.

Since Model II fits the seismicity much better than Model I, we only consider Model II for reconstruction in the following sections.

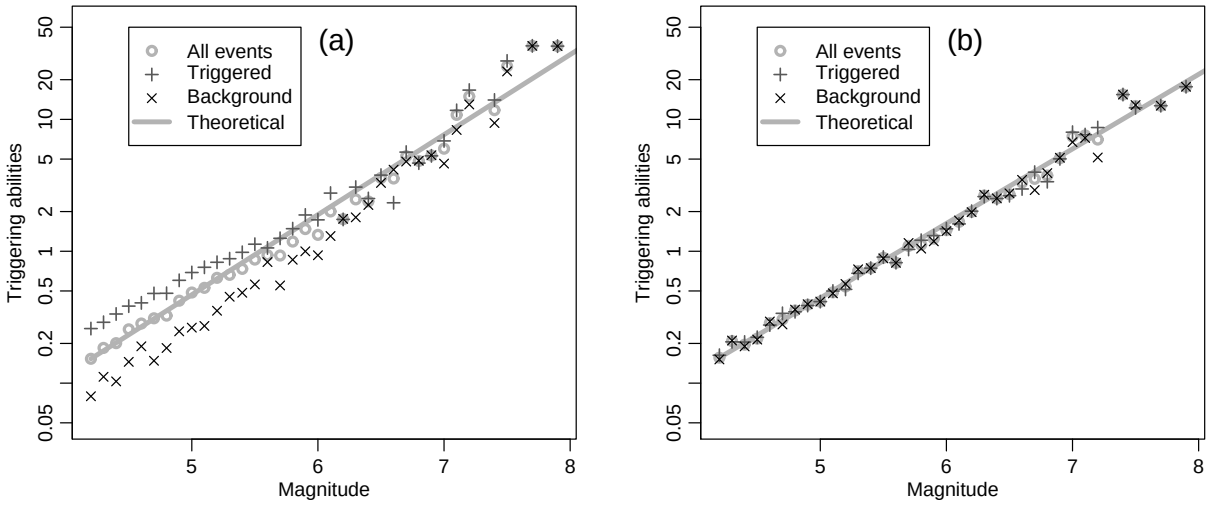

Figure 7: Reconstruction of the triggering abilities, $\hat{\kappa}_b(M)$ in (29) of the background events and $\hat{\kappa}_t(M)$ in (30) of the triggered events for (a) the JMA catalogue and (b) the simulated catalogue. For comparison, the empirical functions of the triggering abilities, $\hat{\kappa}(M)$ in (28) for all the events are plotted in gray circles, and the corresponding theoretical functions, $\kappa(M) = Ae^{\alpha(M-M_C)}$, are represented by the straight lines.

6.2 Difference in triggering ability between the background events and the triggered events

The triggering abilities of an event sized M from all the events, the background events and the triggered events for the JMA catalogue and the simulated catalogue can be reconstructed by using

$$\hat{\kappa}(M) = \frac{\sum_i \sum_j \rho_{ij} I(|M_i - M| < \Delta M/2)}{\sum_i I(|M_i - M| < \Delta M/2)} \tag{28}$$

$$\hat{\kappa}_b(M) = \frac{\sum_i \sum_j \varphi_i \rho_{ij} I(|M_i - M| < \Delta M/2)}{\sum_i \varphi_i I(|M_i - M| < \Delta M/2)} \tag{29}$$

and

$$\hat{\kappa}_t(M) = \frac{\sum_i \sum_j (1-\varphi_i) \rho_{ij} I(|M_i - M| < \Delta M/2)}{\sum_i (1-\varphi_i) I(|M_i - M| < \Delta M/2)}, \tag{30}$$

respectively, as shown in Figure 7. Both the background events and the triggered events generate offspring approximately according to different exponential laws. For the same ancestor magnitude, a triggered event generates more offspring than a background event. The higher is the magnitude, the smaller is the difference. The results for the simulated catalogue show that there is no differences in triggering abilities between these two types of events, indicating that these differences are not caused by numerical procedures.

In the ETAS model, the background events and triggered events generate offspring in the same way. The reconstruction results show that we should have different exponential laws for the triggering abilities of these two types of event, which is

a possible direction on improving the current ETAS models, i.e., to treat the background events and triggered events separately in the model. The higher triggering abilities of the triggered events may be because they occur in an environment where the stress fields is adjusting to the stress changes caused by their ancestors.

6.3 Distributions of offspring locations from different magnitude classes

Because of the historical reasons mentioned in Section 2, we model the distribution of locations of the direct offspring from an earthquake as an inverse power one with a scaling factor associated with the ancestor's magnitude, i.e., $De^{\alpha(M-M_C)}$. Immediate questions for such a choice are:

(a) Is this scaling factor necessary? i.e., Can we use a constant D_0 to replace the scaling factor $De^{\alpha(M-M_C)}$ in the model?

(b) Is it necessary to link the scaling factor to the triggering ability, or should we introduce a new parameter γ instead of α for the scaling factor?

To answer the above questions, for a small interval $\mathcal{M}$ of magnitudes, we select the pairs $\{(i,j)\}$ such that $M_i \in \mathcal{M}$ and then estimate the scaling factor $D_{\mathcal{M}}$ for $\mathcal{M}$, in a way similarly to the estimation of the parameters in mixture models by using the expectation-maximization algorithm (see, e.g., Eggermont and LaRiccia, 2001, Section 2.4). Given ρ_{ij}, consider the following log-likelihood function,

$$\log L = \sum_{M_i \in \mathcal{M}} \sum_j \rho_{ij} \log \left[\frac{(q-1)D^{(q-1)}R_{ij}}{(R_{ij}^2 + D)^q} \right]. \tag{31}$$

To maximize it, let

$$\left. \frac{\partial \log L}{\partial D} \right|_{D=D_{\mathcal{M}}} = 0,$$

i.e.,

$$\frac{q-1}{D_{\mathcal{M}}} \sum_{M_i \in \mathcal{M}} \sum_j \rho_{ij} - q \sum_{M_i \in \mathcal{M}} \sum_j \frac{\rho_{ij}}{R_{ij}^2 + D_{\mathcal{M}}} = 0. \tag{32}$$

Thus, we can construct the following iteration to solve the above equation

$$D_{\mathcal{M}}^{(n+1)} = \frac{(q-1)\sum_{M_i \in \mathcal{M}} \sum_j \rho_{ij}}{q \sum_{M_i \in \mathcal{M}} \sum_j \frac{\rho_{ij}}{R_{ij}^2 + D_{\mathcal{M}}^{(n)}}}. \tag{33}$$

Figure 8 shows the values of $D_{\mathcal{M}}$ against the magnitude classes. We can see that values of $D_{\mathcal{M}}$ have a slope different from $\kappa(M)$, which is 0.5008 by the least square fit. Thus, it is not suitable use $\kappa(M)$ as the scaling factor, a better choice is to introduce another parameter γ as the coefficient in the exponential part .

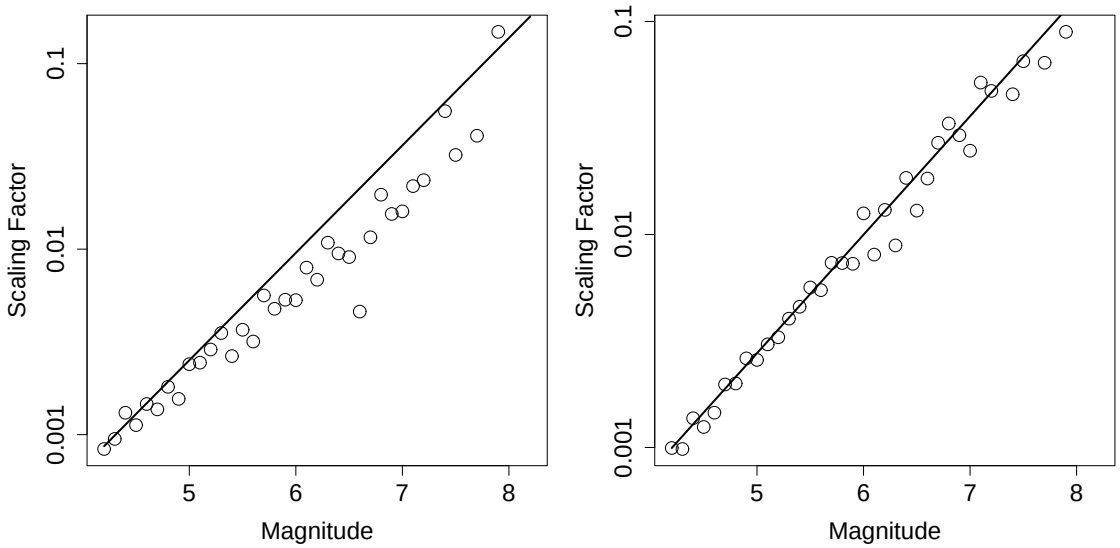

Figure 8: Re-estimated $D_{\mathcal{M}}$ for the JMA catalogue (a) and the simulated catalogue. Theoretical fitting curves, $De^{\alpha(M-M_C)}$, are represented by the straight lines.

7 Conclusion

In this paper, we discussed using the ETAS model to descibe the clustering features of earthquake processes and some methods on how to evaluate the goodness-of-fit. The results show that the ETAS model is a fairly good starting point for modelling the earthquake clusters. Then using different diagnostic methods, we tested the stationary of the background process, the overall goodness-of-fit and each clustering component in the model formulation. These methods are not only useful in justifying the models fitted to the data, but also helpful for us to better understand the data.

References

[1] Console, R., and M. Murru, A simple and testable model for earthquake clustering, *Journal of Geophysical Research*, *106*, 8699-8711, 2001.

[2] Console, R., M. Murru, and A. M. Lombardi, Refining earthquake clustering models. *Journal of Geophysical Research*, *108*(B10), 2468, doi:10.1029/2002JB002130, 2003.

[3] Daley, D., and D. Vere-Jones, *An Introduction to the Theory of Point Process*, *Vol. 1*, Springer, New York, 2002.

[4] Davis, S. D., and C. Frohlich, Single-link cluster analysis, synthetic earthquake catalogs, and aftershock identification, *Geophysical Journal International*, *104*, 289–306, 1991.

[5] Eggermont, P.P.B., and V.N. LaRiccia, *Maximum Penalized Likelihood Estimation*, *Vol I*: Density Estimation, Springer, New York, 2001.

[6] Frohlich, C., and Davis, S. D. Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogs, *Geophysical Journal International*, *100*, 19–32, 100, 1990.

[7] Gardner, J., and L. Knopoff, Is the sequence of earthquakes in southern California, with aftershock removed, Poissonian? *Bulletin of the Seismological society of America*, *64*, 1363 – 1367, 1974.

[8] Kagan, Y. Likelihood analysis of earthquake catalogues," *Journal of Geophysical Research*, *106*(B7), 135–148, 1991.

[9] Kagan, Y. Y., and L. Knopoff Statistical study of the occurrence of shallow earthquakes, *Geophys. J.Roy. astr. Soc.*, *55*, 67–86, 1978.

[10] Kanamori H., and Anderson D., Theoretical basis of some empirical relations in Seismology, *Bulletin of Seismological Society of America*, *65*, 1073–1095, 1975.

[11] Keilis-Borok, V. I., and V. I. Kossobokov, Time of increased probability for the great earthquakes of the world," *Computational Seismology*, *19*, 45–58, 1986.

[12] Lewis, P. A. W., and E. Shedler, Simulation of non-homogeneous Poisson processes by thinning, *Naval Res. Logistics Quart.*, *26*, 403 – 413, 1979.

[13] Musmeci, F., and D. Vere-Jones, A variable-grid algorithm for smoothing clustered data, *Biometrics*, *42*, 483–494, 1986.

[14] Musmeci, F., and D. Vere-Jones, A space-time clustering model for historical earthquakes, *Annals of the Institute of Statistical Mathematics*, *44*, 1–11, 1992.

[15] Ogata, Y., On Lewis' simulation method for point processes, *IEEE translations on information theory*, *IT-27*, 23–31, 1981.

[16] Ogata, Y., Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of American Statistical Association*, *83*, 9–27, 1988.

[17] Ogata, Y., Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, *50*, 379–402, 1998.

[18] Ogata, Y., K. Katsura, and M. Tanemura, Modelling heterogeneous space-time occurrences of earthquake and its residual analysis, *Journal of the Royal Statistical Society: Applied Statistics*, *52*(Part 4), 499–509, 2003.

[19] Omori, F., On after-shocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, *7*, 111–200, 1894.

[20] Rathbun, S. L., Modeling marked spatio-temporal point patterns, *Bulletin of the International Statistical Institute*, *55*, Book 2, 379–396, 1993.

[21] Reasenberg, P., Second-order moment of Central California Seismicity, 1969 – 1982, *Journal of Geophysical Research*, *90*(B7), 5479–5495, 1985.

[22] Schoenberg, F. P., Multi-dimensional residual analysis of point process models for earthquake occurrences, To appear on *Journal of the American Statistical Association*, 2003.

[23] Stoyan D. and Grabarnik P., Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachritchten*, *151*, 95-100, 1991.

[24] Utsu, T., Aftershock and earthquake statistics (I): Some parameters which characterize an aftershock sequence and their interrelations," *Journal of the Faculty of Science, Hokkaido University*, *3*, Ser. VII (Geophysics), 129–195, 1969.

[25] Utsu, T., Y. Ogata and R. S. Matsu'ura, The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, *43*, 1–33, 1995.

[26] Vere-Jones, D. and Schoenberg, F.P., Rescaling marked point processes. *Australian and New Zealand Journal of Statistics*, *to appear*, 2004.

[27] Yamanaka, Y. and K. Shimazaki, Scaling relationship between the number of aftershocks and the size of the main shock, *J. Phys. Earth*, *38*, 305–324, 1990.

[28] Zhuang, J., Y. Ogata and D. Vere-Jones, Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association*, *97*, 369–380, 2002.

[29] Zhuang J., Ogata Y. and Vere-Jones D., Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, *to appear*, 2004.