Spatio-temporal epidemic type aftershock sequence model for Tangshan aftershock sequence*

Shaochuan Lü⁺ and Yong Li

School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

Abstract   Shallow earthquakes usually show obvious spatio-temporal clustering patterns. In this study, several spatio-temporal point process models are applied to investigate the clustering characteristics of the well-known Tangshan sequence based on classical empirical laws and a few assumptions. The relative fit of competing models is compared by Akaike Information Criterion. The spatial clustering pattern is well characterized by the model which gives the best fit to the data. A simulated aftershock sequence is generated by thinning algorithm and compared with the real seismicity.

Key words: spatio-temporal model; Tangshan aftershock sequence; Laplace type clustering, thinning simulation; Akaike information criterion

CLC number: P315.01   **Document code**: A

1 Introduction

Nearly all large shallow earthquakes are accompanied by aftershocks which decay in terms of (modified) Omori's law. The comparative abundance of aftershocks with different magnitudes is nearly distributed as exponential distribution. Based on the empirical criterions, the occurrence rate of aftershocks at time t following the ith earthquake (t_i, M_i) is assumed to be proportional to $Ae^{\alpha(m_i-m_n)}/(t-t_i+c)^p$, where $A_i \propto e^{\alpha(m_i-m_n)}$. Assuming also that each aftershock itself can trigger an event, the conditional intensity in epidemic-type aftershock sequence (ETAS) model (Ogata, 1988) is specified by

$$\lambda(t|H_t) = \mu + \sum_{i:t_i<t} \frac{Ae^{\alpha(m_i-m_n)}}{(t-t_i+c)^p}.$$

Five parameters in the conditional intensity function play different roles in characterizing the clustering and are varying spatially. Among them, p and α indicate the decaying rate of aftershock activity and the power

of a shock (main shock or aftershock) generating its offsprings respectively. Through "point process residual analysis" (Ogata, 1988), abnormal earthquake activity relative to the model including "relative activity" and "relative quiescence" can be sensitively detected. It is necessary to develop spatio-temporal models to indicate the abnormal earthquake activity temporally as well as spatially near the period and region of "relative quiescence" for the purpose of earthquake prediction.

We consider several possible spatio-temporal models and apply them to Tangshan aftershock sequence occurred in North China. The goodness-of-fit tests are compared in terms of AIC to highlight some spatial clustering characteristics such as the relationship between earthquake magnitude and clustering scale. A simulated data set is generated from the estimated model and compared to the real seismicity.

2 Spatio-temporal point process model

Hawkes type self-exciting point process models are extensively applied to investigate the clustering characteristics. The conditional intensity function for such a model is given by

$$\lambda(t,x,y|H_t) = \mu(x,y) + \sum_{i:t_i<t} g(t-t_i, x-x_i, y-y_i, M_i) =$$

$$\mu(x,y) + \int_0^t \iint_A \int_{M_0}^\infty g(t-s, x-\xi, y-\eta, M)N(\mathrm{d}s, \mathrm{d}\xi, \mathrm{d}\eta, \mathrm{d}M),$$

for $(t, x, y) \in [0, T) \times A$ where $N(\mathrm{d}s, \mathrm{d}\xi, \mathrm{d}\eta, \mathrm{d}M)=1$, if an infinitesimal element $(\mathrm{d}s, \mathrm{d}\xi, \mathrm{d}\eta, \mathrm{d}M)$ includes an event (t_i, x_i, y_i, M_i) for some i at time t_i with epicenter (x_i, y_i) and magnitude M_i, otherwise $N(\mathrm{d}s, \mathrm{d}\xi, \mathrm{d}\eta, \mathrm{d}M)=0$, H_t is the history of the process. The whole conditional intensity function consists of two parts. The first part is the background rate which is either a constant or not. The summation of the second part means that it is permitted that each earthquake including aftershock can generate its own aftershocks. The trigger intensity in the second part governs the behavior of aftershocks decaying both temporally and spatially.

Musmeci and Vere-Jones (1991) suggested a spatial-temporal clustering model with the parameterization form of the trigger intensity

$$g(t,x,y,M) = \frac{Ce^{\alpha M}e^{-\beta L}}{2\pi\sigma_x\sigma_y t} \exp\left[-\frac{1}{2t}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right],$$

and also a cauchy form

$$g(t,x,y,M) = \frac{Ce^{\alpha M}e^{-\beta L}}{2\pi\sigma_x\sigma_y t} \frac{1}{\pi^2(x^2+t^2c_x^2)(y^2+t^2c_y^2)}.$$

For estimating background rate $u(x, y)$, a kernel type smoothing $\hat{\mu}(x, y)$ for $\mu(x, y)$ is performed and an additional parameter is applied to adjust the background rate and the rest part of the conditional intensity function

$$\lambda(t,x,y) = (1-p)\hat{\mu}(x,y)+$$
$$p\sum_{i:t_i<t} g(t-t_i, x-x_i, y-y_i, M_i),$$

The model was applied to analyze Italian historical earthquake catalogue.

Kagan (1991) and Rathbum (1993) suggested other forms of the trigger intensities. Among a variety of spatio-temporal models, Ogata (1998) scrutinized several models with the form of response function

$$g(t,x,y,M) = \frac{K_0}{(t+c)^p} \exp\left\{-\frac{x^2+y^2}{2de^{\alpha(M-M_0)}}\right\}, \quad (1)$$

$$g(t,x,y,M) = \frac{K_0}{(t+c)^p} \frac{e^{\alpha(M-M_0)}}{(x^2+y^2+d)^q}, \quad (2)$$

$$g(t,x,y,M) = \frac{K_0}{(t+c)^p}\left\{\frac{x^2+y^2}{e^{\alpha(M-M_0)}} + d\right\}^{-q}. \quad (3)$$

Further investigation on identifying main shocks in a cluster was performed by Ogata (1998) based on "MBC" algorithm. According to the algorithm, a space-time window was set to obtain the background rate and the epicenters of aftershock were estimated. However, there exist a number of declustering algorithms. Each one can generate a declustered catalogue and result in a different estimation of background rate. A stochastic declustering algorithm based on the model suggested by Zhuang et al. (2002) was applied to generate declustered catalogue and obtain the background rate, see also Zhuang et al. (2005). The goodness-of-fit test of models shows that the model with heterogeneous background intensity is remarkably better than that with homogeneous Poisson field intensity in terms of AIC. However, the anisotropic clustering models are not significantly better than the isotropic models despite several parameters are introduced to investigate the characteristic of spatial clustering. The conditional intensity rate of the compared anisotropic models is

$$\lambda(t,x,y|H_t) = v\hat{\mu}(x,y)+\sum_{i:t_i<t} g(t-t_i, x-x_i, y-y_i, M_i).$$

The response functions are of the form

$$g(t,x,y,M) = K(M) \times \frac{(p-1)c^{(p-1)}}{(t+c)^p} \times$$
$$\left\{\frac{1}{\pi\sigma(M)}f\left[\frac{r(\Theta)^2}{\sigma(M)}\right]\right\},$$

where $r(\theta)^2 = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{\sigma_2}{\sigma_1}x^2 - 2\rho xy + \frac{\sigma_2}{\sigma_1}y^2\right)$ with $\theta = \arctan\frac{y}{x}$. In addition, among the three classes of models, the heterogeneous background model with the response function in model (3) provided the best performance by AIC for the investigated seismic areas.

2.1 Space-time model for Tangshan sequence

Now, we assume that

(1) The background rate in a studied region is constant or is a function of location and independent of time;

(2) Each event can generate its own aftershocks;

(3) The aftershocks decay with time in terms of modified Omori's law;

(4) The probability distribution of the location (x, y) of an aftershock is independent of the time.

The conditional intensity for such a model is

$$\lambda(t,x,y|H_t) = \mu(x,y) + \sum_{i:t_i<t} K(M_i)g(t-t_i)$$
$$f(x-x_i, y-y_i|M_i).$$

Let conditional intensity functions include

$$\lambda(t,x,y|H_t) = \mu + \sum_{i:t_i<t} A\mathrm{e}^{\alpha(m_i-m_0)}\left(1+\frac{t-t_i}{c}\right)^{-p}\exp\left[-\frac{|x-x_i|}{d_1}-\frac{|y-y_i|}{d_2}\right], \qquad (4)$$

$$\lambda(t,x,y|H_t) = \mu + \sum_{i:t_i<t} A\mathrm{e}^{\alpha(m_i-m_0)}\left(1+\frac{t-t_i}{c}\right)^{-p}\left[\frac{(x-x_i)^2+(y-y_i)^2}{\mathrm{e}^{\alpha(m_i-m_0)}}+d\right]^{-q}, \qquad (5)$$

$$\lambda(t,x,y|H_t) = \mu + \sum_{i:t_i<t} A\mathrm{e}^{\alpha(m_i-m_0)}\left(1+\frac{t-t_i}{c}\right)^{-p}\exp\left[-\frac{|x-x_i|}{d_1\mathrm{e}^{\alpha(m_i-m_0)}}-\frac{|y-y_i|}{d_2\mathrm{e}^{\alpha(m_i-m_0)}}\right], \qquad (6)$$

where $\phi=(A,\alpha,c,p,d_1,d_2)$ for equations (4) and (6), and $\phi=(A,\alpha,c,p,d,q)$ for equation (5).

Given the occurrence times and spatial locations of earthquakes with magnitudes (t_i, x_i, y_i, M_i) during time interval $[T_1, T_2]$ and in a region A, the log-likelihood of the model is

$$\mathrm{Log}L(\mu,\phi) = \sum_{i=1}^{n}\mathrm{Log}\lambda(t_i,x_i,y_i,M_i|H_t)-$$

$$\int_{T_1}^{T_2}\iint_A \lambda(t,x,y|H_t)\mathrm{d}t\mathrm{d}x\mathrm{d}y. \qquad (7)$$

It is evident that the superposed conditional intensity

$$\lambda(t|H_t) = \iint_A \lambda(t,x,y|H_t)\mathrm{d}x\mathrm{d}y$$

is consistent with conditional intensity of ETAS model.

In other words, the response functions (4) and (5) satisfy

$$\iint_A g(t,x,y)\mathrm{d}x\mathrm{d}y \approx \iint_{R^2} g(t,x,y)\mathrm{d}x\mathrm{d}y = \frac{A\mathrm{e}^{\alpha(M-M_0)}}{(t+c)^p}$$

and the response function (6) satisfy

$$\iint_A g(t,x,y)\mathrm{d}x\mathrm{d}y \approx \iint_{R^2} g(t,x,y)\mathrm{d}x\mathrm{d}y = \frac{A\mathrm{e}^{2\alpha(M-M_0)}}{(t+c)^p}.$$

The likelihoods for models (4) and (6) can be explicitly expressed once the studied region is assumed to be a rectangle. However, the likelihood for model (5) cannot be explicitly expressed. We approximate it by numerical method.

3 Application to Tangshan sequence

SSlib includes many useful historical earthquake catalogues. The data are chosen from SSLib ranged within 116.6°E–119.5°E, 38.5°N–40.4°N for the time span 1975-01-01–1988-01-01 which includes the well-known Tangshan sequence. See the epicentral distribution of the aftershock sequence in Figure 1. We consider several data sets in the same region with different cut-off magnitude levels 4 and 4.5. The completeness of the threshold magnitude is about 4.0. Below this level, the catalogue might be incomplete. AIC and the estimated parameters of models (4), (5) and (6) for different cut-off magnitude levels are listed in Table 1. AIC is an information theoretical criterion for model selection which provides a tradeoff to balance the goodness-of-fit measured by the likelihood and the complexity of the model, usually measured by the dimension of the model, to prevent over-fit by otherwise an unnecessary complicated model. It is defined by $-2\mathrm{log}L+2k$, where $\mathrm{log}L$ is the log-likelihood and k is the number of the parameters.

Among the seven parameters, α and p are specially useful for characterizing the clustering of an aftershock sequence. According to Table 1, it is evident that μ and A are increasing with the decreasing cut-off magnitude levels for the same model. However, p and q are reducing
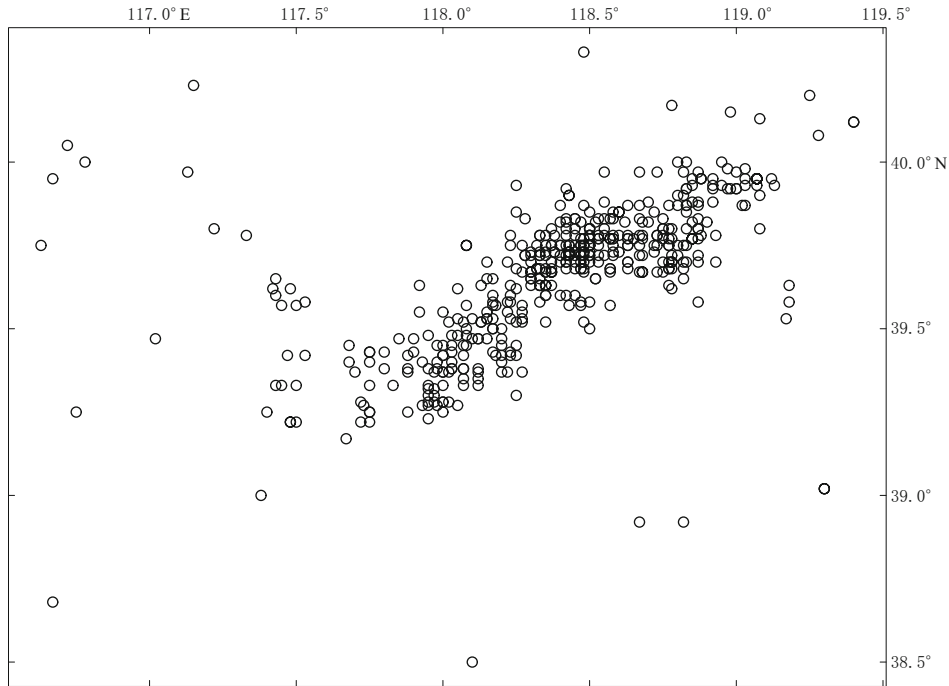
Figure 1   The epicentral distribution of earthquakes with magnitude $M{\geq}4$.

Table 1   Estimated parameters and the AIC values for the models with homogeneous Poisson field for the background seismicity

$M{\geq}4$	μ	A	α	C	p	$d_1(d)$	$d_2(q)$	AIC
Model (4)	1.028e−3	2.186e+2	0.782	1.081e−2	0.928	6.296e−2	3.690e−2	1 472.5
Model (5)	7.332e−4	7.372e−4	0.942	9.750e−3	0.916	1.008e−3	1.950	1 491.9
Model (6)	9.108e−4	7.668e+2	0.534	8.647e−3	0.905	3.038e−2	1.574e−2	1 428.1
$M{\geq}4.5$	μ	A	α	C	p	$d_1(d)$	$d_2(q)$	AIC
Model (4)	4.479e−4	1.030e+2	0.906	3.024e−2	1.007	8.044e−2	4.332e−2	807.1
Model (5)	4.611e−4	6.025e−7	0.945	3.501e−2	1.004	9.486e−3	4.109	823.4
Model (6)	4.195e−4	2.854e+2	0.484	3.107e−2	1.001	4.888e−2	2.289e−2	781.8

with decreasing cut-off magnitude levels. Other parameters such as C, d_1 and d_2 show similar trend as that of p and q for the same model. We also note that p in model (6) is systematically smaller than models (4) and (5) for all the threshold magnitude levels. In addition, Table 1 also shows that model (6) performs better than all other models in terms of AIC. The outcome is stable for all the threshold magnitude levels. From this point of view, the spatial clustering feature of an aftershock sequence correlates with the magnitude of its ancestor,

that is to say, the spatial clustering scale of aftershocks associates with the magnitude of their ancestor.

In most cases, the assumption of homogeneous background rate is impractical. In fact, background rate in a specified region varies from place to place. See Figure 1 for the epicentral distribution of Tangshan sequence. The conditional intensity functions of the alternative models with heterogeneous background rate are as follows:

$$\lambda(t,x,y|H_t) = \mu(x,y) + \sum_{i:t_i<t} A\mathrm{e}^{\alpha(m_i-m_0)}\left(1+\frac{t-t_i}{c}\right)^{-p}\exp\left[-\frac{|x-x_i|}{d_1}-\frac{|y-y_i|}{d_2}\right], \qquad (8)$$

$$\lambda(t, x, y|H_t) = \mu(x, y) + \sum_{i:t_i<t} A\left(1 + \frac{t-t_i}{c}\right)^{-p} \exp\left[-\frac{|x-x_i|}{d_1 e^{\alpha(m_i-m_0)}} - \frac{|y-y_i|}{d_2 e^{\alpha(m_i-m_0)}}\right]. \tag{9}$$

To estimate the background rate, we first perform a kernel type smoothing such that

$$\hat{\mu}(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right),$$

where $K(x, y)$ is taken as a standard multivariate normal kernel function and h is the band width. Here, we use a single smoothing parameter $h=0.96\times n^{(-1/6)}$. Then an additional parameter is applied to adjust background rate and the rest part of the conditional intensity function. Thus the background rate is given by $\mu(x, y)=v\hat{\mu}(x, y)$ with a parameter v to be estimated. The integrals of the log-likelihood in all above models are approximated numerically in terms of heterogeneous background rate in the log-likelihood.

The extended model is fitted to the same data set as aforementioned. Table 2 lists the estimated parameters and AIC values. As far as the models with heterogeneous background rate are concerned, each model performs considerably better than their counterpart with homogeneous background rate. In addition, v and A in Table 2 show the same trend as that in Table 1 with decreasing magnitude thresholds. Furthermore, model (9) gives a better fit than model (8) for all magnitude cutoffs in terms of AIC, suggesting stability of the outcome. In fact, model (9) gives the best fit for all cases. The heterogeneous background intensity models also cause the estimation of α and p larger by comparing Table 1 and Table 2.

Table 2 Estimated parameters and the AIC values for the models with heterogeneous Poisson field for the background seismicity

$M\geq4$	v	A	α	C	p	d_1	d_2	AIC
Model (8)	0.532 5	55.491 6	1.168 3	0.104 0	1.450 4	0.059 8	0.038 7	870.2
Model (9)	0.545 1	396.439 9	0.689 8	0.032 7	1.453 7	1.551e−2	1.116e−2	828.5
$M\geq4.5$	v	A	α	C	p	d_1	d_2	AIC
Model (8)	0.208 5	28.279 9	1.216 0	0.237 4	1.481 7	0.081 7	0.047 9	584.2
Model (9)	0.207 2	121.133 6	0.575 8	0.270 0	1.485 1	4.028e−2	2.167e−2	574.4

4 Simulation implementation

As far as a marked point process is concerned, time and mark coordinates can be simultaneously simulated by thinning method. However, such a straight simulation can be avoided by simulating time and space-magnitude coordinates sequentially in terms of superposed conditional intensity

$$\Lambda(t) = \iint_A \lambda(t, x, y|H_t)\mathrm{d}x\mathrm{d}y$$

as described in Musmeci and Vere-Jones (1990), Ogata (1981, 1998). It is generally assumed that the magnitude is stable and independent of its occurrence time. Here, simulation begins from the most largest main shock and the magnitude of simulated events is taken as the same as that of real catalogue in chronological order. The magnitude-time plot of simulated catalogue from model (9) with cut-off magnitude of 4.0 and the real one are showed in Figure 2. Simulated data set is generated according to following steps:

(1) Simulate a sequence of occurrence times $\{t_i\}_{i=1}^{n}$ in terms of the superposed conditional intensity with respect to locations

$$\Lambda(t) = v_0 + \sum_{j:t_j<t} v_j(t)$$

by thinning method, where $v_0=\int_A \mu(x,y)\mathrm{d}x\mathrm{d}y$ and $v_j(t)=\iint_{R^2} g(t-t_j, x-x_j, y-y_j, M_j)\mathrm{d}x\mathrm{d}y$, generate a magnitude m_i.

(2) generate spatial coordinates as follows:

a) set $i=1$;

b) generate a uniform random number U in (0, 1]; If $U\leq\frac{v_0}{\Lambda(t_i)}$, generate (x_i, y_i) in terms of heterogeneous Poisson field intensity by thinning method, or else put $J=1$;

c) if $U\leq\frac{\Lambda(t_J)}{\Lambda(t_i)}$, set $J=J+1$, repeat current step;

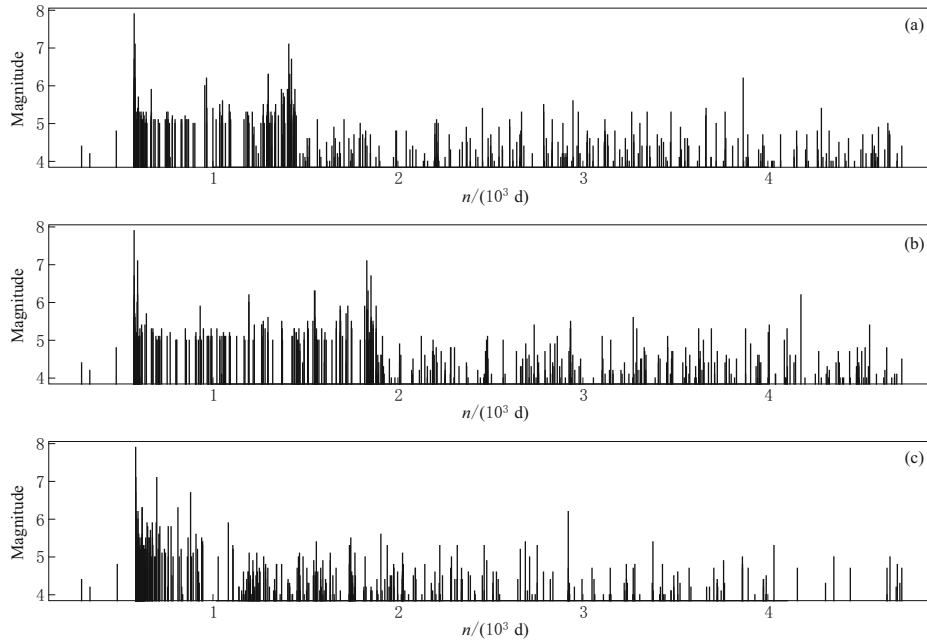d) generate (x_i, y_i) in terms of Laplace distribution, put $x_i=x_j+x_i$ and $y_i=y_j+y_i$;

Figure 2   (a) and (b) are the *M-T* plots for two simulated catalogues based on model (9), the events magnitude (≥4) in two plots are assigned as same as the real catalogue in chronological order. (c) is the *M-T* plot of the real catalogue.

e) put $i=i+1$, go to step b);

f) if $(t_i, x_i, y_i, M_i) \notin A$, delete (t_i, x_i, y_i, M_i), $i=1$, ..., n.

In this study, we are using model (9) with $M_0=4$ to generate simulated data set. The magnitude of simulated events are taken as the same as that of real catalogue in chronological order. Figures 2 and 5 show the magnitude-time and epicentral distribution of the

simulated data set. Figures 3 and 4 show time versus longitude and latitude plots for the simulated sequence. Note that the pattern of simulated data set is similar to that of the real one both spatially and temporally from Figures 2–5. The epicentral distribution of simulated data set is approximate to that of the real one. However, there exist significant differences between the simulated catalogues and the real one. The clustering
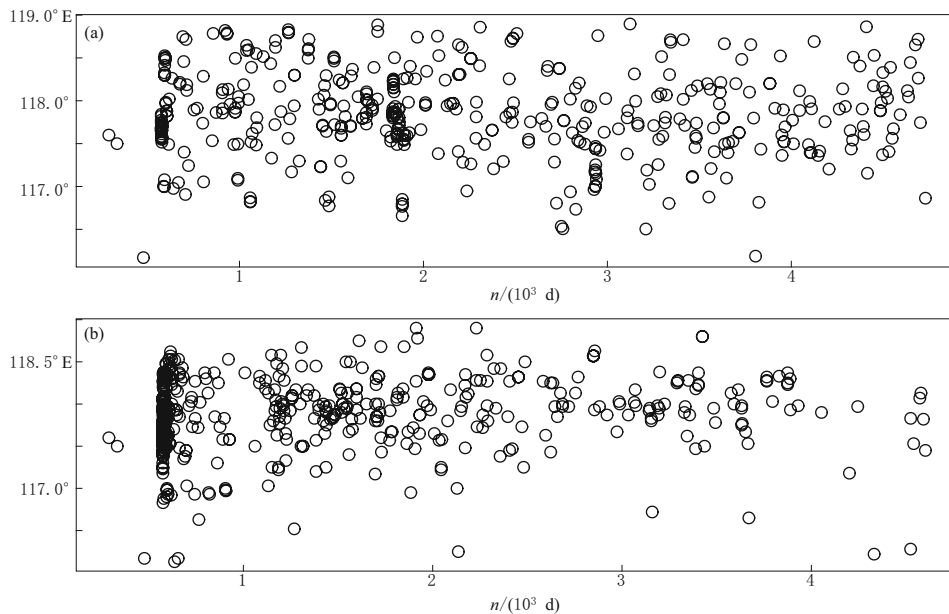

Figure 3   (a) is the time versus longitude plot for the simulated events based on model (9). (b) is the time versus longitude plot for the real catalogue. The threshold magnitude is greater than or equal to 4.
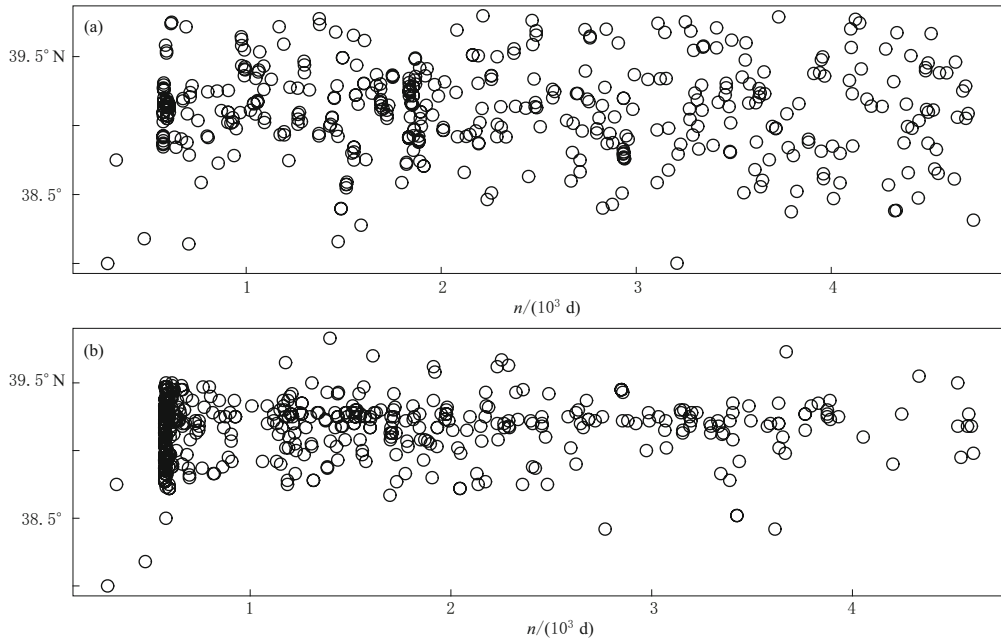
Figure 4   (a) is the time versus latitude plot for the simulated events based on model (9). (b) is the time versus longitude plot for the real catalogue. The threshold magnitude is greater than or equal to 4.
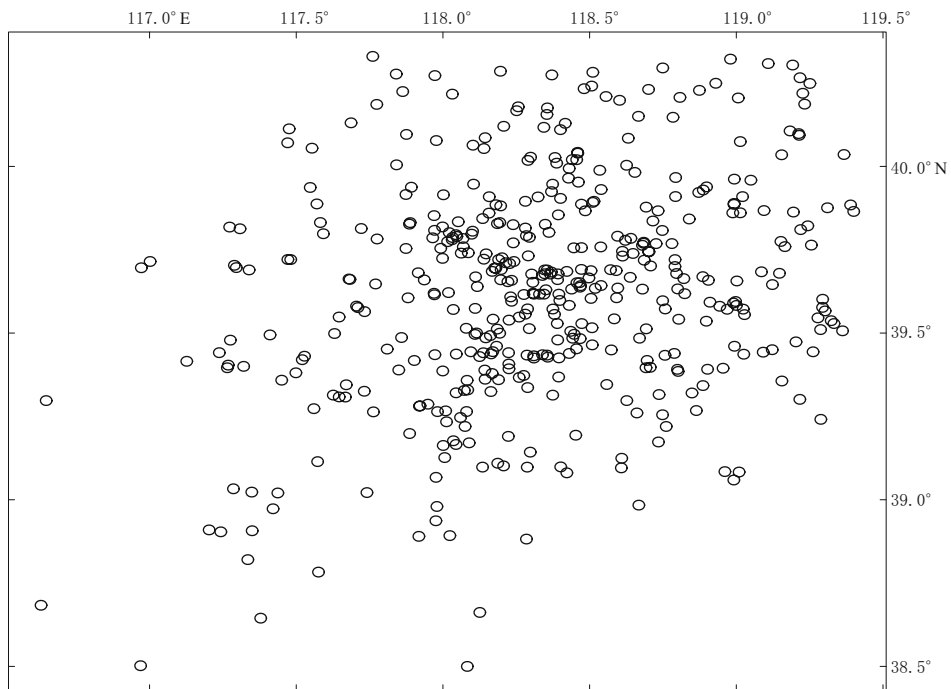

Figure 5   The epicentral distribution of simulated data from model (9) with magnitude $M_0 \geq 4$.

in real data is more intensive both temporally and spatially. The distinctions may be caused by many factors. Firstly, the assumption that background rate is independent of the occurrence time may be impractical. In fact, background rate in a specified region may vary

from time to time with certain trend due to many geophysical factors. Sometimes, the cumulative number of real events is systematically larger or smaller than the expected number predicted from models. Such a systematical deviation of the real pattern from the model can

not be automatically adjusted by current model. At the same time, we assume the magnitude sequence is independent of times, spatial locations. Other assumptions of the model also result in the differences between the simulated catalogue and the real one.

5 Concluding remarks

Among models (4–6) with homogeneous background rate, model (6) gives the best fit to the real seismicity for all the cut-off magnitude levels. Similar results hold for the models with heterogeneous background rate for all the threshold magnitude levels. The difference of AIC increases as the size of the data set increases. The above result gets clearer for the data with the lower threshold magnitude levels. The scale of the clustering is clearly dependents on magnitude with factor $e^{\alpha(m_i-m_0)}$. However, the uncertainty of the ETAS parameter estimates is important (Wang et al., 2010a) and the impact of this uncertainty on the model choices is very interesting and will be investigated by the simulation method described by Wang et al. (2010a) in the future work. In addition, earthquake triggering behavior estimates could be significantly biased by the missing earthquake links due to the choice of minimum magnitude threshold, time and space window, which has been explored by Wang et al. (2010b). The impact of these missing links on our model choices will be investigated in the further research. We generate simulated events in terms of the estimated models by the thinning method. Here, the simulation study is tentative; the genuine seismic activity is fundamentally more complicated. The spatial clustering is so intensive for such a sequence that the bilateral Laplace distribution with scale factor $e^{\alpha(m_i-m_0)}$ gives a better fit to the data. Simulations according to the estimated parameters in model (9) show some similar and also distinctive features which may illuminate further studies.

Remark 1 From Figure 1, we notice that the spatial distribution of the earthquakes is non-isotropic. In fact, the events are distributed in a NE direction along the fault trace. Intuitively, by rotating coordinates to parallel the slip direction of the faults, the preferred model under the new coordinates might be further improved. However, how to compare models under different coordinate systems in terms of some information theoretical criterions such as AIC or BIC is unclear. The limitations may illuminate further improvements for current models.

Remark 2 From Figures 2–4, one notes that the main shock is followed by an intensive aftershock activity relative to the real catalogue, i.e., the expected number of aftershocks in terms of the model is greater than that of the observed. It suggests that the whole seismic activity may be divided into several phases. Thus, the model with different phase changes may provides better performance. Such a phase change may relates to the geophysical intrinsic changes such as "relative activity" and "relative quiescence".

Remark 3 The completeness threshold of the earthquake catalogue for Tangshan sequence is about 4. The analytical results listed in Tables 1 and 2 for events with magnitude greater than 4 and the above is more credible. We admit that one issue associated with current analysis is that some events with missing epicenter records pose influence on the estimation.

References

Kagan Y Y (1991). Likelihood analysis of earthquake catalogues. *J Geophys Res* **106**: 135–148.

Musmeci F and Vere-Jones D (1992). A space-time cluster model for historical earthquakes. *Ann Inst Statist Math* **44**: 1–11.

Ogata Y (1981). *On Lewis' Simulation Method for Point Process, IEEE Transaciton on Information Theory*, IT-27, 27–31.

Ogata Y (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *JASA* **83**: 9–27.

Ogata Y (1998) Space-time point process models for earthquake occurrences. *Ann Inst Statist Math* **50**: 379–402.

Rathbun S L (1993). Modeling marked spatio-temporal point patterns. *Bull Interna Statist Inst* **55**: 379–396.

Wang Q, Jackson D D and Zhuang J (2010b). Missing links in earthquake clustering model. *Geophys Res Lett* **37**: L21307.

Wang Q, Schoenberg F and Jackson D D (2010a). Standard errors of parameter estimates in the ETAS model. *Bull Seismol Soc Am* **100**(5A): 1 989–2 001.

Zhuang J, Chang C-P, Ogata Y and Chen Y-I (2005). A study on the background and clustering seismicity in the Taiwan region by using a point process model. *J Geophys Res* **110**: B05S18.

Zhuang J, Ogata Y and Vere-Jones D (2002). Stochastic declustering of space-time earthquake occurrences. *JASA* **97**: 369–380.