Ann. Inst. Statist. Math.
Vol. 50, No. 2, 379–402 (1998)

SPACE-TIME POINT-PROCESS MODELS FOR EARTHQUAKE OCCURRENCES

YOSIHIKO OGATA

The Institute of Statistical Mathematics, Minami-Azabu 4-6-7, Minato-ku, Tokyo 106-8569, Japan

(Received January 19, 1996; revised May 9, 1997)

Abstract. Several space-time statistical models are constructed based on both classical empirical studies of clustering and some more speculative hypotheses. Then we discuss the discrimination between models incorporating contrasting assumptions concerning the form of the space-time clusters. We also examine further practical extensions of the model to situations where the background seismicity is spatially non-homogeneous, and the clusters are non-isotropic. The goodness-of-fit of the models, as measured by AIC values, is discussed for two high quality data sets, in different tectonic regions. AIC also allows the details of the clustering structure in space to be clarified. A simulation algorithm for the models is provided, and used to confirm the numerical accuracy of the likelihood calculations. The simulated data sets show the similar spatial distributions to the real ones, but differ from them in some features of space-time clustering. These differences may provide useful indicators of directions for further study.

Key words and phrases: Centroid of aftershock epicenters, ETAS model, inverse power laws, maximum likelihood estimates, magnitude based clustering (MBC) algorithm, modified Omori formula, thinning simulation.

1. Introduction

The Epidemic Type Aftershock-Sequences (ETAS) model (Ogata (1988)) is a point process representing the activity of earthquakes of magnitude M_0 and larger in a region during a period of time. The model includes background activity of constant occurrence rate μ in time (i.e., stationary Poisson process) and also includes aftershocks as described below. Each earthquake, including aftershocks of another earthquake, is followed by its aftershock activity, although only aftershocks of magnitude M_0 and larger are included in the data. The aftershock activity is represented by a nonstationary Poisson process according to the modified Omori formula (Utsu (1961)) in such a way that the occurrence rate of aftershocks at

time t following the i-th earthquake (t_i, M_i) is given by

$$\nu_i(t) = \frac{K_0}{(t - t_i + c)^p} \cdot e^{\alpha(M_i - M_0)}, \tag{1.1}$$

for $t > t_i$, where the parameters K_0, α, c, p are constants common to all i. The rate of occurrence of the whole earthquake series at time t, called the conditional intensity function based on the history of occurrence $H_t = \{(t_i, M_i); t_i < t\}$, then becomes

$$\lambda_\theta(t \mid H_t) = \mu + \sum_{\{i:t_i<t\}} \nu_i(t). \tag{1.2}$$

The ETAS model can also be written in terms of a stochastic integral (e.g. Daley and Vere-Jones (1988)), namely

$$\lambda(t \mid H_t) = \mu + \int_{M_0}^{\infty} \int_0^t \frac{K_0}{(t - s + c)^p} \cdot e^{\alpha(M - M_0)} N(ds, dM), \tag{1.3}$$

where $N(ds, dM) = 1$ if an infinitesimal element (ds, dM) includes an event (t_i, M_i) for some i, otherwise $N(ds, dM) = 0$.

The five parameters (μ, K_0, c, α, p) in (1.3) represent some characteristics of seismic activity of the region. The parameters may correlate with tectonophysical conditions (e.g. structural heterogeneity, stress state, temperature, etc., cf. Guo and Ogata (1997)). Therefore, they vary spatially, and also temporally in some cases. Among the parameters of ETAS model, α and p are particularly useful for characterizing the temporal pattern of seismicity. The p value indicates the decay rate of aftershocks, and the α value measures an efficiency of magnitude of an earthquake in generating its offspring, or aftershocks in a wide sense (e.g. Ogata (1987, 1992)). For example, swarm-type activity has a smaller α value than that of ordinary main shock and aftershock activity, and α is large if there are no conspicuous secondary aftershocks in an aftershock sequence, or the magnitude of the main shock is much larger than the maximum magnitude of its aftershocks.

From the estimated ETAS model, we can predict the expected occurrence rate of earthquakes in normal sequences. Comparing the predicted rate with that of observed occurrence data, periods of relatively decreased or increased seismic activity can be recognized (Ogata (1988, 1989, 1992)): for instance, a significant decrease in seismic activity of a region below the level predicted by the ETAS model, which we call *Relative Quiescence*, is sometimes followed by a large earthquake in the same or neighboring region.

Spatial aspects of earthquake prediction have also been developed in seismology to some extent. From studies of the seismicity of the northwestern Circum-Pacific seismic belt, Fedotov (1965) and Mogi (1968) found that *seismic gaps* in activity have been successively filled, within several tens of years, by a series of great earthquakes without significant overlap of their rupture zones. According to the studies a seismic gap roughly corresponds to the aftershock area of the forthcoming earthquake, that is, the size of the gap leads to an estimation of the

magnitude of the predicted earthquake. The gap theory gave successful predictions in some cases (Utsu (1972) and Ohtake *et al.* (1977), for instance). However, this is not frequently the case and gaps do not always appear very clearly, especially in areas where the background activity is high. Further, the seismicity pattern is usually very complicated, showing various clustering features which make it difficult to evaluate the significance of smaller gaps. Thus, the seismic gap theory seems still under development and even controversial (e.g., McCann *et al.* (1979), Kagan and Jackson (1991, 1995) and Nishenko and Sykes (1993)).

The ultimate objective of our study for earthquake prediction is to indicate the location of the anomalous area as well as the corresponding temporal anomalies. In the similar manner to the application of the ETAS model for detection of relatively quiet period (Ogata (1988), (1989) and (1992)), we believe that the sensitivity in detecting such anomalies can be amplified by contrasting the observed with the predicted seismic activity of the considered space-time volume. In other words, we need a suitable statistical space-time model for the detection of relatively quiet periods and regions from hypocenter data of earthquakes. Such a model has to be good enough to represent the seismicity of the considered wide area throughout the whole period of the available data.

In this paper, several possible extensions of the ETAS model to space-time data are considered, based partly on classical empirical studies of aftershocks, and partly on a number of contrasting speculative hypotheses about the physical nature of the space-time clustering, as specifically described in Subsection 2.3. Thus, our main goal here is the discrimination of the models. For such purpose, goodness-of-fit of the models is compared by the aid of the AIC for two data sets from tectonically distinctive areas in and around Japan (Section 3). In Section 4, further practical extensions of the models are suggested for the realistic but complex features such as non-homogeneous background seismicity and occasionally anisotropic clusters, and further model comparisons are made. A simulation algorithm for these models is described and implemented in Section 5 to compare the space-time patterns with real data sets. The last section describes the conclusions.

2. Space-time models

2.1 *Self-exciting processes*

If we denote by $P_{\Delta t,\Delta x,\Delta y}(t,x,y \mid H_t)$, the history-dependent probability that an earthquake occurs in a small time interval between t and $t+\Delta t$ and in a small region $[x,x+\Delta x)\times[y,y+\Delta y)$, where $H_t=\{(t_i,x_i,y_i,M_i);t_i<t\}$ is the history of occurrence times $\{t_i\}$ up to time t, their corresponding epicenters $\{(x_i,y_i)\}$ and magnitudes $\{M_i\}$, then the conditional intensity function $\lambda(t,x,y \mid H_t)$ of the space-time point process can be defined as

$$\lambda(t,x,y \mid H_t)=\lim_{\Delta t,\Delta x,\Delta y\to 0}\frac{P_{\Delta t,\Delta x,\Delta y}(t,x,y \mid H_t)}{\Delta t\Delta x\Delta y}.$$

As far as stationarity is assumed, Hawkes' type self-exciting point-process model (Hawkes (1971)) is naturally extended to the following form

$$\lambda(t,x,y \mid H_t)=\mu(x,y)+\sum_{\{i:t_i<t\}} g(t-t_i,x-x_i,y-y_i;M_i) \tag{2.1}$$

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

for $(t,x,y) \in [0,T] \times A$, where $N(ds, d\xi, d\eta, dM) = 1$ if an infinitesimal element $(ds, d\xi, d\eta, dM)$ includes an event (t_i, x_i, y_i, M_i) for some i, otherwise $N(ds, d\xi, d\eta, dM) = 0$.

2.2 *Some models from previous studies*

An important aspect for space-time statistical modeling is the parametric form of the response function $g(\cdot,\cdot,\cdot;\cdot)$ of an earthquake described in (2.1). Musmeci and Vere-Jones (1992) suggest a diffusion type function

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

with $\phi = (C, \alpha, \beta, \sigma_x^2, \sigma_y^2)$, and also a product-Cauchy form

$$g_\phi(t,x,y;M) = \frac{Ce^{\alpha M}e^{-\beta t}c_x c_y}{\pi^2(x^2 + t^2c_x^2)(y^2 + t^2c_y^2)}$$

with $\phi = (C, \alpha, \beta, c_x, c_y)$. For the estimation, they first carried out a kernel-type smoothing $\hat{\mu}(x,y)$ for $\mu(x,y)$ in (2.1) using the whole data of locations $\{(x_i,y_i) \in A\}$, and then introduced an additional parameter p such that $0 \le p \le 1$ in order to adjust itself and the remaining parameters ϕ simultaneously in the conditional intensity function

$$\lambda_\theta(t,x,y) = (1-p)\hat{\mu}(x,y) + p \sum_{\{i:t_i<t\}} g_\phi(t-t_i, x-x_i, y-y_i; M_i)$$

with $\theta = (p;\phi)$. This was applied to the Italian historical earthquake data. Due to the decreased activity in the marginal and outer part of the region, the boundary effect in the likelihood calculation is negligible in this particular data.

At about the same time, Kagan (1991) suggested other parametric forms of $g(\cdot,\cdot,\cdot;\cdot)$ based on investigations of the second-order statistical features in time and space of various hypocenter catalogs (e.g., Kagan and Knopoff (1978, 1980)). One of those, for 2-dimensional space, assuming a linear fault model, is given by

$$g_\phi(t,x,y;M) = \frac{\beta T_M^\beta}{t^{1+\beta}} I_{[t_M,\infty)}(t) \cdot K10^{(3/2)\delta M} \cdot \frac{\sqrt{x^2+y^2}}{\epsilon^2 + \sigma 10^{M-M_0}} \exp\left\{\frac{-(x^2+y^2)}{2(\epsilon^2 + \sigma 10^{M-M_0})}\right\},$$

where the parameters $\phi = (\beta, K, \delta, \epsilon, \sigma)$ are to be estimated. $I_{[T_M,\infty)}(t)$ is the indicator function taking value 1 when $t \ge T_M$, otherwise 0; its purpose is to take account missing events in a wave duration time T_M in the seismogram, such that

$T_M \propto 10^{0.5M}$ for each earthquake with magnitude M. The parameter ϵ stands for the standard error of epicenter determination. In this model, Kagan (1991) used the scalar seismic moment $\mathcal{M}$, translated into the magnitude M by the relation (Kanamori and Anderson (1975))

$$\log_{10} \mathcal{M} = 1.5M + \text{const.}$$

Among a variety of spatio-temporal models of marked point processes in ecology and seismology, Rathbun (1993) applied the model

$$g_\phi(t, x, y; M) = \frac{e^{-\beta_0 + \beta(M - M_0)}}{(t + c)^p} \varphi\left(\frac{\sqrt{x^2 + y^2}}{\sigma}\right)$$

to California earthquakes with $M \geq 5.0$ between 1932 and 1992, where the parameters are $\phi = (\beta_0, \beta_1, c, p, \sigma)$ and $\varphi(\cdot)$ is the standard Normal density function $\boldsymbol{N}(0, 1^2)$. Further, Rathbun (1996) provides regularity conditions justifying the standard large sample theory of the maximum likelihood procedure for spatio-temporal point-process models.

2.3 *Extensions of the ETAS model*

Among many reported empirical laws for aftershock statistics, one of the most important with a spatial nature is the Utsu-Seki formula (Utsu and Seki (1955), Utsu (1969)),

$$\log_{10} S = 1.02M + \text{const.},$$

giving the relation between the magnitude M of the main shock and the area of aftershock region S. The aftershock region in their sense is defined as the narrowest convex set that includes almost all the aftershocks within a certain time span from the occurrence time of the main shock. The aftershock region is approximated by an ellipse to calculate the area (Utsu (1969)). Nowadays, the aftershock region is known to correspond to the ruptured fault region of the main shock, and the aftershock area is closely related to the seismic moment cited in the last section. Also, the number of aftershocks N in a certain time interval after the main shock is considered to be proportional to the aftershock area S, so that a similar relation to the Utsu-Seki formula holds (Yamanaka and Shimazaki (1990)). The number density of aftershocks N/S and also the constant terms in both the Utsu-Seki formula (Utsu (1969)) and the moment versus magnitude relation (Kanamori and Anderson (1975)) take significantly different values for earthquakes within a plate and on a plate boundary.

On the other hand, remember that the ETAS model is not concerned with the discrimination between main shocks and aftershocks, but with the appropriate form of the response function for the causal relation with subsequent events (Ogata (1988)). In a similar manner, we are concerned here with the response function for space-time causal relationship, taking the above mentioned quantitative studies into consideration. Specifically, we are concerned with the following questions of physical interest.

1. What is the form of the spatial terms in the model? Are the clusters restricted to well-defined regions, such as aftershock areas, or do they extend beyond the traditional aftershock regions? Can one discriminate between a model with a fairly sharp boundary, and one with a much more diffuse boundary? Are there perhaps two components (near field and far field) with different characteristics?
2. How do the cluster regions scale with magnitude? Do they increase with magnitude, or are they more or less independent of magnitude?

As extensions of the ETAS model we confine ourselves to space-time response functions $g_\phi(t,x,y;M)$ such that the superposed conditional intensity

$$\lambda(t) = \iint_A \lambda(t,x,y)dxdy$$

coincides with the conditional intensity of the ETAS model in (1.3). That is to say, the response functions must satisfy

$$\iint_A g_\phi(t,x,y;M)dxdy \approx \iint_{\mathrm{R}^2} g_\phi(t,x,y;M)dxdy = \frac{K_0 e^{\alpha(M-M_0)}}{(t+c)^p}.$$

Such response functions for the isotropic (i.e. rotation invariant) spatial clustering include:

$$g_\phi(t,x,y;M) = \frac{K_0}{(t+c)^p}\exp\left\{-\frac{1}{2}\frac{x^2+y^2}{de^{\alpha(M-M_0)}}\right\}, \tag{2.2}$$

$$g_\phi(t,x,y;M) = \frac{K_0}{(t+c)^p}\cdot\frac{e^{\alpha(M-M_0)}}{(x^2+y^2+d)^q}, \tag{2.3}$$

and

$$g_\phi(t,x,y;M) = \frac{K_0}{(t+c)^p}\left(\frac{x^2+y^2}{e^{\alpha(M-M_0)}}+d\right)^{-q},$$

where $\phi = (K_0,c,\alpha,p,d)$ for (2.2), and $\phi = (K_0,c,\alpha,p,d,q)$ for (2.3) and (2.4). The models (2.2) and (2.3) are presented in Ogata (1993).

The discriminating features among the above models can be summarized as follows. All the above response functions can be rewritten in the common standard form

$$g_\phi(t,x,y;M) = \kappa(M)\times\frac{(p-1)\,c^{p-1}}{(t+c)^p}\times\left[\frac{1}{\pi\sigma(M)}\cdot f\left\{\frac{x^2+y^2}{\sigma(M)}\right\}\right], \tag{2.5}$$

where $\kappa(M)\propto e^{\alpha M}$ is a cluster size factor (expected number of aftershocks for the event of magnitude M), $(p-1)c^{p-1}/(t+c)^p$ is time probability density distribution, and $\pi^{-1}\sigma(M)^{-1}f\{(x^2+y^2)/\sigma(M)\}$ is space probability density distribution (equivalently $\int_0^\infty f(x)dx = 1$ holds) in which the scale factor $\sigma(M)$ is allowed to depend on magnitude M. Then the main contrasts in modeling the response function taking the points of our questions into consideration are:

1. between the functional forms allowed for $f(\cdot)$; short range decay (i.e. normal etc.) versus long range decay (i.e. inverse power law), and
2. between either $\sigma(M) = \text{const.}$ (independent of M) or $\sigma(M) = e^{\alpha M}$.

It should be noted that the exponential form for $\kappa(M)$ is due to that the superposition of the present space-time intensity models with respect to space is the ETAS model in time, and that the exponential form for $\sigma(M)$ is based on the Utsu-Seki law of the aftershock area relative to the magnitude of the main shock.

2.4 *Estimation and model comparison*

Given occurrence times and space coordinates of earthquakes with their magnitudes $\{(t_i, x_i, y_i, M_i); M_i \geq M_0, i = 1, \ldots, n\}$ during a time interval $[0, T]$ and in a region A, the log-likelihood of the model (2.1) is given by

$$\log L(\mu, \phi) = \sum_{i=1}^{n} \log \lambda(t_i, x_i, y_i \mid H_t) - \int_0^T \iint_A \lambda(t, x, y \mid H_t) dt dx dy. \tag{2.6}$$

The first term can be computed straightforwardly. For the numerical calculation of the second term we have to prepare the following notations; first assume, for simplicity, that $\mu(x, y) = \mu = \text{const.}$ (cf. Subsection 4.2), and that the study region A is a convex set. Let $A - (\xi, \eta) = \{(x - \xi, y - \eta); (x, y) \in A\}$ for any location (ξ, η) in A. Further, for each event i with location (x_i, y_i), suppose the region $A - (x_i, y_i)$ is radially divided into K subregions $\{A_k; k = 1, 2, \ldots, K\}$ by the radial segments connecting between the origin and each of the K knots on the boundary of the convex region $A - (x_i, y_i)$ such that

$$A - (x_i, y_i) = \bigcup_{k=1}^{K} A_k^{(i)}.$$

Let the angles of the dividing line segments of the k-th subregion $A_k^{(i)}$ be θ_k and θ_{k+1} $(0 \leq \theta_1 < \cdots < \theta_k < \theta_{k+1} < \cdots < \theta_K < 2\pi)$ at the origin anti-clockwise from x-axis, and let $\Delta_k = \theta_{k+1} - \theta_k$. Then, changing the order of the integrals, the second term is calculated as follows.

$$\begin{aligned}
&\int_0^T \iint_A \lambda(t, x, y \mid H_t) dt dx dy \\
&\quad = \mu T|A| + \int_0^T \iint_A \int_{M_0}^{\infty} N(ds, d\xi, d\eta, dM) \\
&\qquad \cdot \int_0^{T-s} \iint_{A-(\xi,\eta)} dt dx dy \, g_\phi(t, x, y; M) \\
&\quad = \mu T|A| + \sum_{i=1}^{n} \int_0^{T-t_i} \sum_{k=1}^{K} \iint_{A_k^{(i)}} dt dx dy \, g_\phi(t, x, y; M) \\
&\quad \approx \mu T|A| + \sum_{i=1}^{n} \int_0^{T-t_i} \frac{K_0}{(t+c)^p} dt \times \sum_{k=1}^{K} S_k(x_i, y_i; M_i) \cdot \frac{\Delta_k}{2\pi}.
\end{aligned} \tag{2.7}$$

In the last (approximate) equality, we have transformed and approximated the integral of $g_\phi(\cdot)$ with respect to Cartesian coordinate $dxdy$ over the subregion $A_k^{(i)}$ by the integral with respect to the polar coordinates $rdrd\theta$ over the partial disc centered at the origin with radius r_k which is a distance between (x_i, y_i) and a point on the segment in boundary of the subregion $A_k^{(i)}$. Here, K is taken to be sufficiently large for the accurate numerical approximation of the integral in (2.7). This is important for the accurate log-likelihood calculation avoiding the boundary effect. Specifically, $S_k(\cdot,\cdot;\cdot)$ is the integration of the spatial factor of $g_\phi(t,x,y;M)$ over the disk with the radius r_k such that

$$
\begin{aligned}
(2.8)\quad S_k(x_i, y_i; M_i) &= 2\pi d e^{\alpha(M_i - M_0)} \left[1 - \exp\left\{\frac{-r_k^2}{2de^{\alpha(M_i-M_0)}}\right\}\right] \quad \text{for} \quad (2.2), \\
&= \frac{\pi e^{\alpha(M_i-M_0)}}{1-q}\left\{(r_k^2+d)^{1-q} - d^{1-q}\right\} \quad \text{for} \quad (2.3), \text{ and} \\
&= \frac{\pi e^{\alpha(M_i-M_0)}}{1-q}\left[\left\{\frac{r_k^2}{e^{\alpha(M_i-M_0)}} + d\right\}^{1-q} - d^{1-q}\right] \quad \text{for} \quad (2.4).
\end{aligned}
$$

The maximum log-likelihood and its maximizing parameters $\hat{\theta} = (\hat{\mu}, \hat{\phi})$ in (2.6) are computed by a standard nonlinear optimization procedure such as the Davidon-Fletcher-Powell algorithm which uses gradient vector values of the function in (2.6) with respect to the parameters as the inputs (see Fletcher and Powell (1963) and Kowalik and Osborne (1968)). For our models the shape of the log-likelihood function in (2.6) can be complex, so that the initial values of the parameters to maximize the function should be taken carefully. For instance, μ is taken as a quarter of $N/(T|A|)$; K_0, c, and d are taken small positive values such as 0.01; $p \approx 1.3$, $q \approx 2.0$ and $\alpha \approx 1.0$.

To compare the goodness-of-fit of the models, we use the Akaike's Information Criterion (AIC; Akaike (1974)) which is defined by

$$
(2.9)\quad \text{AIC} = -2 \max_\theta(\text{log-likelihood}) + 2 \times (\text{number of adjusted parameters}).
$$

A model with a smaller AIC value is considered to be a better fit.

3. Comparison of the models

3.1 *Application to the JMA data*

We use the hypocenter data compiled by the Japan Meteorological Agency (JMA), and consider two data sets from tectonically distinctive areas. The first data are chosen from the wide region 36~42°N and 141~145°E (off the east coast of Tohoku District; see Figs. 1a and 1b) for all depth and for the time span 1926–1995. From now on, we refer to this region as Region A. The majority of large earthquakes in Region A took place on the plate boundary between the North American and subducting Pacific plates. The same region was discussed in Ogata and Katsura (1988) for the estimation of the intensity rate of an event with respect

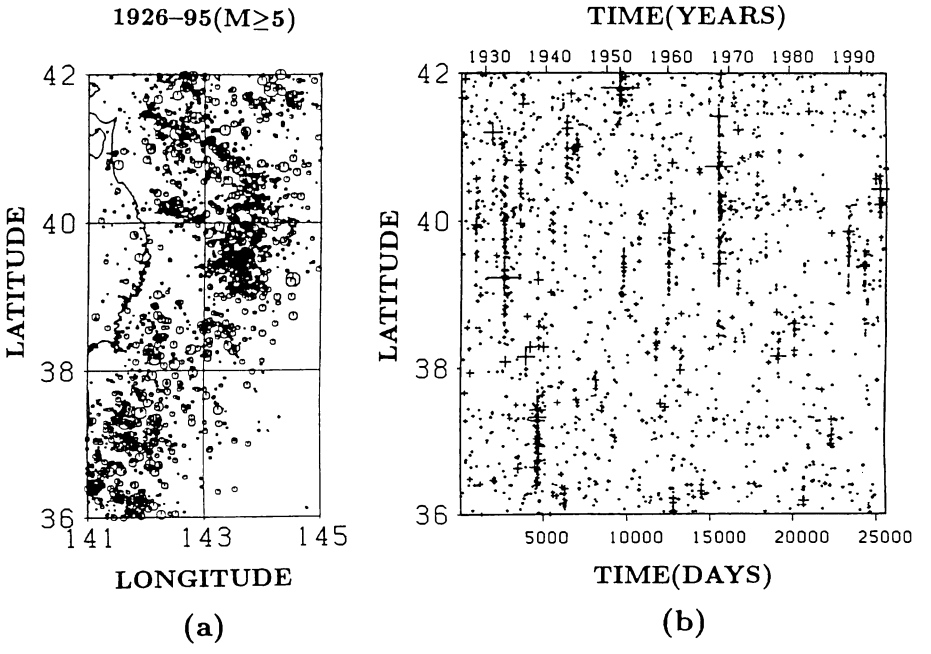

Fig. 1. Earthquakes with $M \geq 5.0$, off the east coast of Tohoku District (Region A), northern Japan, for the time span of 1926–1995; (a) distribution of epicenters and (b) space (latitude) versus time plot. Size of circle and plus sign corresponds magnitude of earthquake.

to location only: a similar method will be used to fit the background seismicity in Subsection 4.1.

Here, we consider several data sets for earthquakes with $M_J \geq M_0$ to examine the stability of the results, where the threshold magnitudes M_0 are taken as 6.0, 5.5, 5.0 and 4.5.

Another area of interest is the region 34~38°N and 131~137°E, the central and western part of Honshu Island, Japan, shown in the Fig. 2a, hereafter referred to as Region B, where the majority of earthquakes are considered to be intraplate events within the Eurasian, North American and Philippine Sea plates. Shallow earthquakes ($h \leq 45$ km) are considered for the time span 1926–1995 (see Fig. 2b for longitude versus time plot of the seismicity). Again, we consider several data sets for earthquakes in this area with $M_J \geq M_0$, where the threshold magnitudes M_0 are taken as 5.5, 5.0, 4.5 and 4.0.

The AIC comparison and parameter estimation of the models are carried out for each data set. We here note that, in calculating epicenter separation (distance between earthquake epicenters), the difference in longitudes is reduced by a factor $\cos(\pi y_0/180°)$ relative to the difference in latitudes (one degree corresponds to about 111 km), where y_0 is taken to be the latitude of the center of the area. Table 1 lists the estimated parameters and the AIC values of the respective models. According to the table, the goodness-of-fit of model (2.4) is significantly better

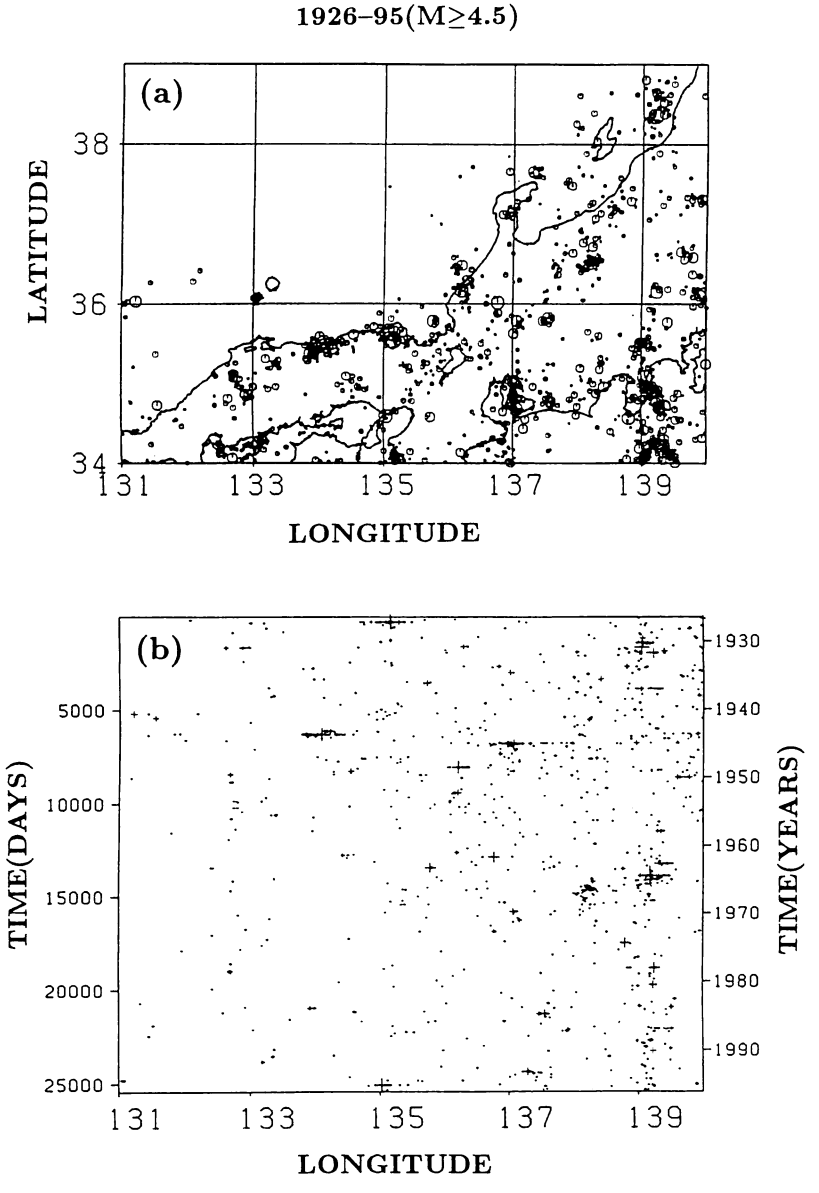

Fig. 2. Shallow earthquakes with $M \geq 4.5$ in central and western Honshu area (Region B), central Japan, for the time span of 1926–1995; (a) epicenter distribution, and (b) time versus space (longitude) plot.

than the others for most of the cutoff magnitude levels from both areas. This gets clearer as the threshold magnitude M_0 increases. Concerning the questions on the clustering features raised in Subsection 2.2, the best model (2.4), in contrast with the others, suggests that the clustering shows long-range decay with distance, and that the size of the spatial cluster scales sharply with the magnitude of the

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

Model	μ (shock/day/degree^2)	K	c (days)	α (M^{-1})	p	d (degree)	q	AIC
Off the east coast of Tohoku District $M \geq 6.0$, 334 shocks								
(2.2)	$.153 \times 10^{-3}$	.174	$.106 \times 10^{-1}$	1.666	0.968	$.211 \times 10^{-1}$		4038.5
(2.3)	$.137 \times 10^{-3}$	$.896 \times 10^{-3}$	$.977 \times 10^{-2}$	1.396	0.971	$.575 \times 10^{-1}$	1.712	4035.6
(2.4)	$.140 \times 10^{-3}$	$.174 \times 10^{-3}$	$.952 \times 10^{-2}$	1.636	0.965	$.164 \times 10^{-1}$	1.860	4008.0
Off the east coast of Tohoku District $M \geq 5.5$, 900 shocks								
(2.2)	$.337 \times 10^{-3}$	.425	$.124 \times 10^{-1}$	1.827	0.964	$.627 \times 10^{-2}$		8874.5
(2.3)	$.251 \times 10^{-3}$	$.700 \times 10^{-3}$	$.122 \times 10^{-1}$	1.247	0.970	$.160 \times 10^{-1}$	1.471	8910.5
(2.4)	$.282 \times 10^{-3}$	$.669 \times 10^{-4}$	$.120 \times 10^{-1}$	1.784	0.965	$.620 \times 10^{-2}$	1.852	8798.1
Off the east coast of Tohoku District $M \geq 5.0$, 2104 shocks								
(2.2)	$.661 \times 10^{-3}$	.566	$.110 \times 10^{-1}$	1.584	0.937	$.454 \times 10^{-2}$		18295.2
(2.3)	$.423 \times 10^{-3}$	$.542 \times 10^{-3}$	$.115 \times 10^{-1}$	1.176	0.949	$.156 \times 10^{-1}$	1.482	18220.3
(2.4)	$.502 \times 10^{-3}$	$.931 \times 10^{-4}$	$.114 \times 10^{-1}$	1.591	0.943	$.265 \times 10^{-2}$	1.612	18085.7
Off the east coast of Tohoku District $M \geq 4.5$, 4333 shocks								
(2.2)	$.111 \times 10^{-2}$	.780	$.777 \times 10^{-2}$	1.360	0.900	$.305 \times 10^{-2}$		33314.2
(2.3)	$.558 \times 10^{-3}$	$.443 \times 10^{-3}$	$.872 \times 10^{-2}$	0.949	0.917	$.640 \times 10^{-2}$	1.438	33124.9
(2.4)	$.707 \times 10^{-3}$	$.967 \times 10^{-4}$	$.836 \times 10^{-2}$	1.281	0.909	$.184 \times 10^{-2}$	1.565	32897.4
Central and Western Honshu Island $M \geq 5.5$, 157 shocks								
(2.2)	$.864 \times 10^{-4}$	.389	$.321 \times 10^{-2}$	0.779	0.973	$.851 \times 10^{-2}$		2306.8
(2.3)	$.669 \times 10^{-4}$	$.342 \times 10^{-3}$	$.319 \times 10^{-2}$	0.955	0.944	$.955 \times 10^{-2}$	1.506	2279.5
(2.4)	$.670 \times 10^{-4}$	$.277 \times 10^{-3}$	$.315 \times 10^{-2}$	0.902	0.944	$.393 \times 10^{-2}$	1.450	2281.5
Central and Western Honshu Island $M \geq 5.0$, 489 shocks								
(2.2)	$.176 \times 10^{-3}$	1.226	$.178 \times 10^{-2}$	0.932	0.962	$.318 \times 10^{-2}$		5030.1
(2.3)	$.121 \times 10^{-3}$	$.254 \times 10^{-3}$	$.172 \times 10^{-2}$	0.843	0.944	$.402 \times 10^{-2}$	1.548	4975.3
(2.4)	$.127 \times 10^{-3}$	$.152 \times 10^{-3}$	$.172 \times 10^{-2}$	0.871	0.946	$.225 \times 10^{-2}$	1.584	4972.7
Central and Western Honshu Island $M \geq 4.5$, 1285 shocks								
(2.2)	$.329 \times 10^{-3}$	1.926	$.369 \times 10^{-2}$	0.853	0.974	$.251 \times 10^{-2}$		9592.3
(2.3)	$.217 \times 10^{-3}$	$.243 \times 10^{-3}$	$.406 \times 10^{-2}$	0.890	0.967	$.192 \times 10^{-2}$	1.490	9395.5
(2.4)	$.226 \times 10^{-3}$	$.152 \times 10^{-3}$	$.408 \times 10^{-2}$	0.950	0.968	$.845 \times 10^{-3}$	1.492	9376.1
Central and Western Honshu Island $M \geq 4.0$, 3007 shocks								
(2.2)	$.633 \times 10^{-3}$	3.562	$.483 \times 10^{-2}$	0.972	0.966	$.121 \times 10^{-2}$		16510.6
(2.3)	$.312 \times 10^{-3}$	$.276 \times 10^{-3}$	$.517 \times 10^{-2}$	0.822	0.962	$.729 \times 10^{-3}$	1.382	15904.4
(2.4)	$.335 \times 10^{-3}$	$.167 \times 10^{-3}$	$.514 \times 10^{-2}$	0.910	0.961	$.341 \times 10^{-3}$	1.405	15836.7

earthquake. We see that α for the plate boundary region is systematically larger than for the intraplate events. Further, p in both regions has a value slightly smaller values than 1; the reason will be suggested in Section 4.

According to the estimate $\hat{\mu}$ in the model (2.4), the background seismicity rate in the Region A (plate boundary) relative to the whole seismicity rate is about twice as high as in Region B (region of intraplate earthquakes) for each of the same threshold magnitudes $M_0 = 5.5$, 5.0 and 4.5. Since the total numbers of earthquakes in Region A is about three to four times as large as the one in Region B for each magnitude threshold, this shows that clustering activity of clusters is higher in Region A than in Region B.

4. More realistic models

4.1 *Cluster identification and centroid of aftershock epicenters*

It is often the case that the epicenter of a main shock is located at the margin of its aftershock area, because the epicenter listed in the catalog corresponds to the location of the rupture initiation. In such cases, the epicenter location in the catalog is not quite suitable for the present models with spatial response functions given in (2.2)–(2.4). Instead, we will consider the *centroid of aftershock epicenters* defined as follows using the JMA hypocenter catalog. In order to estimate the centroid, we have to identify clusters of aftershocks. Then, we use the average of the locations of the aftershocks in a cluster to replace the catalogue's epicenter of the main shock. The following algorithm for identifying aftershock clusters is a modification of the *magnitude-based clustering* (*MBC*) algorithm introduced in Ogata *et al.* (1995) which is based on Utsu-Seki empirical law of the aftershock area in space and the modified Omori law in time.

The MBC algorithm for the present purpose starts with selecting the largest shock in the catalog for the *main shock*. If there are plural largest shocks, the earliest one is adopted for the main shock. Then, to form a cluster, we set a space-time window where the bounds of distance and time depend on the magnitude of the main shock as explained below. All the earthquakes within the window are considered to be the cluster members, and removed from the catalog. Then the largest events in the remainder are selected to continue the same procedure. This procedure lasts until only isolated events remain. The time span of the window is taken to be $\max(100, 10^{0.5M-1})$ days (i.e. 100 days for $M = 4 \sim 6$ and 1000 days for $M = 8$) after the main shock. The side length of square area for the space window centred at a main shock epicenter is taken to be $2 \times (0.015 \times 10^{0.5M-2} + \epsilon)$ degrees (i.e., about 70 km for $M = 4$ and 400 km for $M = 8$): here, we took $\epsilon = 0.3$ degrees for the error of epicenter determination in early years of the JMA catalog.

The difference of the present (modified) MBC with the original MBC in Ogata *et al.* (1995) is that here the clusters include aftershocks only, while the original definition included any 'preshocks' in addition. Examples of the main shocks thus identified at their superposed locations are shown in the maps of Figs. 3a and 3b for both areas.

4.2 *Non-homogeneous background seismicity*

Instead of the constant $\hat{\mu}$ estimated in Subsection 2.4, we will estimate the function $\mu(x, y)$ in (2.1) indirectly as follows. Expanding the function $\log \mu(x, y)$ by the bi-cubic B-splines, its coefficients are estimated by maximizing the penalized log-likelihood where we assume the non-homogeneous Poisson field for the likelihood and a smoothness penalty for $\mu(x, y)$. Ogata and Katsura (1988) describes the detail of the model and of the objective Bayesian estimation procedure to determine the weights for the penalties. In the present case, the data for the estimation is not the whole data but only the *main shocks* identified by the MBC algorithm in Subsection 4.1. The intensity function thus estimated is denoted by $\hat{\mu}_0(x, y)$. This function is estimated for each data set for each of the different magnitude thresholds. An example of the estimated intensity functions of location is

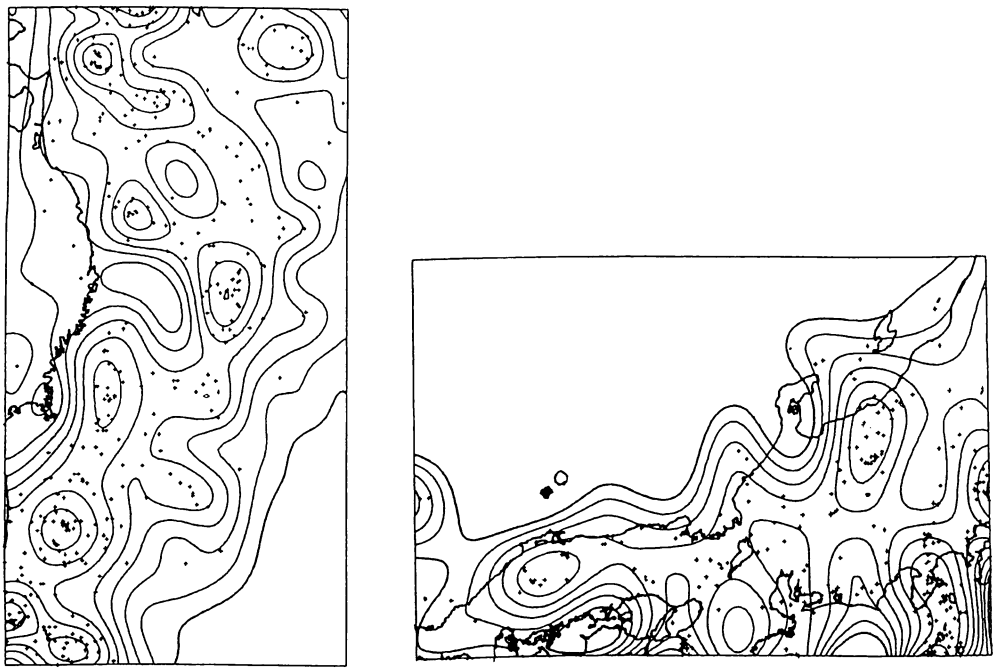

Fig. 3. 'Main shocks' identified by the algorithm of Magnitude based clustering (MBC; see text) in (a) Region A and (b) Region B for the earthquakes with $M \geq 5.0$ and $M \geq 4.5$, respectively. The contours for both (a) and (b) show the height of the spatial intensity functions for the non-homogeneous Poisson fields for the 'main shock' locations (+-signs) in the respective regions; the contour lines increase exponentially such that $1, 2, 4, \ldots$.

shown in Figs. 3a and 3b for Regions A and B with the threshold magnitudes 4.5 and 5.0, respectively.

We then assume that the background intensity rate per unit space-time volume at location (x, y) is given by

$$\mu(x, y) = \nu \cdot \hat{\mu}_0(x, y), \tag{4.1}$$

where ν is a constant to be estimated.

4.3 *Anisotropic clustering*

Usually, estimated epicenter locations of aftershocks are approximately elliptically distributed (see Utsu (1969), for instance), for several reasons such as the dip angle of the slipped fault of an earthquake, proportion of the slipped length of the fault to its width, and the location errors of aftershock hypocenters. Therefore, aiming at a better fit of the models to an earthquake catalog, each response function in (2.2), (2.3) and (2.4) is extended in such a way that the isotropic term

$x^2 + y^2$ in the response functions is replaced by

$$\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\sigma_2}{\sigma_1}x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2}y^2\right) \tag{4.2}$$

so that the corresponding iso-circle and iso-ellipse as a cross-section of the function at the same height have the same area as each other. Then the numerical integral in (2.7) with (2.8) for each of the extended models of (2.2), (2.3) and (2.4) for (4.2) is carried out in the same way, expect that the Euclidian distance r_k from (x_i, y_i) to the boundary's k-th knots (ξ_k, η_k) in (2.7) is replaced by

$$r_k^2 = \frac{1}{\sqrt{1-\rho^2}}\left\{\frac{\sigma_2}{\sigma_1}(\xi_k - x_i)^2 - 2\rho(\xi_k - x_i)(\eta_k - y_i) + \frac{\sigma_1}{\sigma_2}(\eta_k - y_i)^2\right\}. \tag{4.3}$$

In (4.2) and (4.3), the parameters are restricted to satisfy $\sigma_1 > 0$, $\sigma_2 > 0$ and $-1 \leq \rho \leq 1$, but these are allowed to take distinct values depending on the clusters as follows.

Suppose that events in space-time data are decomposed into a number of clusters by the MBC algorithm described in Subsection 4.1. Assume that the locations of the main shock and its aftershocks $\{(x_i, y_i) \in C; i = 1, \ldots, n\}$ in each cluster $C \in \mathcal{C}_{MBC}$ are distributed according to one of the following four models of bivariate Normal distribution;

$$\begin{aligned} &\boldsymbol{N}\left(\begin{pmatrix}x_1\\ y_1\end{pmatrix}, \begin{pmatrix}\tilde{\sigma}^2 & 0\\ 0 & \tilde{\sigma}^2\end{pmatrix}\right), \quad \boldsymbol{N}\left(\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}, \begin{pmatrix}\hat{\sigma}^2 & 0\\ 0 & \hat{\sigma}^2\end{pmatrix}\right), \\ &\boldsymbol{N}\left(\begin{pmatrix}x_1\\ y_1\end{pmatrix}, \begin{pmatrix}\tilde{\sigma}_1^2 & \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2\\ \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2\end{pmatrix}\right) \quad \text{and} \quad \boldsymbol{N}\left(\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}, \begin{pmatrix}\hat{\sigma}_1^2 & \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2\\ \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2\end{pmatrix}\right) \end{aligned} \tag{4.4}$$

where (x_1, y_1) is the epicenter location of the main shock in the cluster, $(\bar{x}, \bar{y})$ is the centroid of aftershock epicenters (i.e. the sample average of the locations over the cluster members), and

$$\tilde{\sigma}^2 = \left\{\sum_j (x_j - x_1)^2 + \sum_j (y_j - y_1)^2\right\}/2n;$$

$$\hat{\sigma}^2 = \left\{\sum_j (x_j - \hat{x})^2 + \sum_j (y_j - \hat{y})^2\right\}/2n;$$

$$\tilde{\sigma}_1^2 = \sum_j (x_j - x_1)^2/n, \quad \tilde{\sigma}_2^2 = \sum_j (y_j - y_1)^2/n,$$

$$\tilde{\rho} = \sum_j (x_j - x_1)(y_j - y_1)/(n\tilde{\sigma}_1\tilde{\sigma}_2);$$

$$\hat{\sigma}_1^2 = \sum_j (x_j - \hat{x})^2/n, \quad \hat{\sigma}_2^2 = \sum_j (y_j - \hat{y})^2/n,$$

$$\hat{\rho} = \sum_j (x_j - \hat{x})(y_j - \hat{y})/(n\hat{\sigma}_1\hat{\sigma}_2),$$

where every sum is taken for $j = 1, \ldots, n$.

Then, for each cluster, the goodness-of-fit of these models are compared by using

$$\text{AIC} = n \log \|S\| + 2K + \text{const.}$$

in the present case; where n is the number of cluster members, S is the estimated variance-covariance matrix of the model, $\|\cdot\|$ is the determinant of the matrix, and K is the number of adjusted parameters. Here, $K = 1, 3, 3$ and 5, respectively, in the order of the above models of Normal distribution in (4.4). The model with the smallest AIC is selected for each cluster. However, it is found that extremely high correlation $\rho \approx \pm 1$ takes place for a number of clusters with only a few members. The reason is that in early years the epicenters were determined on a coarse spatial lattice. Therefore, to avoid this difficulty, the four models in (4.4) are compared only in the case where cluster size is equal to or greater than 6; otherwise, one of the first two models in (4.4) is selected.

If AIC selects any of the models in which the centroid of aftershock epicenters $(\bar{x}, \bar{y})$ is used, then the epicenter coordinate (x_1, y_1) of the main shock in the catalog is replaced by $(\bar{x}, \bar{y})$. Also, if AIC selects any of the last two models for anisotropic clusters in (4.4) then the corresponding estimates of σ_1, σ_2 and ρ are used for the elliptic form in (4.2) in the response function of the corresponding main shock; otherwise they remain the same as for other events which are not identified as main shocks by the MBC algorithm. In this way, the data sets used in Subsection 3.1 for various cut-off magnitudes are modified to compare the goodness-of-fit of the extended versions of those in (2.2)–(2.4).

4.4 *Model comparison among the extended models*

Using the centroid locations and the anisotropic deformation based on the procedure described in the previous sections, we can compare the extended models of (2.2)–(2.4),

$$\lambda(t, x, y \mid H_t) = \nu \cdot \hat{\mu}_0(x, y) + \sum_{\{i: t_i < t\}} g_\phi(t - t_i, x - x_i, y - y_i; M_i), \tag{4.5}$$

where the response functions are of the form

$$g_\phi(t, x, y; M) = \kappa(M) \times \frac{(p-1)c^{p-1}}{(t+c)^p} \times \left[\frac{1}{\pi\sigma(M)} \cdot f\left\{ \frac{r^2(\theta)}{\sigma(M)} \right\} \right] \tag{4.6}$$

where

$$r^2(\theta) = \frac{1}{\sqrt{1-\rho^2}} \left(\frac{\sigma_2}{\sigma_1} x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2} y^2 \right) \tag{4.7}$$

with $\theta = \tan^{-1}(y/x)$. Then the same calculation as in (2.7) with the same $S_k(\cdot, \cdot; \cdot)$ as (2.8) are carried out, except that the Euclidian distance r_k between the origin and the k-th segment on the boundary of the set $A - (x_i, y_i)$ is replaced by the

Table 2. Estimated parameters and the AIC values for the models with the non-homogeneous Poisson field for the background seismicity and the isotropic clustering.

Model	ν (shock/day/degree^2)	K	c (days)	α (M^{-1})	p	d (degree)	q	AIC
Off the east coast of Tohoku District $M \geq 6.0$, 334 shocks								
(2.2)	$.447 \times 10^{-4}$	.249	$.283 \times 10^{-1}$	1.750	1.094	$.142 \times 10^{-1}$		3891.0
(2.3)	$.449 \times 10^{-4}$	$.819 \times 10^{-3}$	$.304 \times 10^{-1}$	1.456	1.122	$.501 \times 10^{-1}$	1.721	3908.3
(2.4)	$.437 \times 10^{-4}$	$.608 \times 10^{-4}$	$.266 \times 10^{-1}$	1.831	1.096	$.230 \times 10^{-1}$	2.327	3868.8
Off the east coast of Tohoku District $M \geq 5.5$, 900 shocks								
(2.2)	$.626 \times 10^{-4}$	.483	$.311 \times 10^{-1}$	1.783	1.101	$.608 \times 10^{-2}$		8597.5
(2.3)	$.627 \times 10^{-4}$	$.619 \times 10^{-3}$	$.354 \times 10^{-1}$	1.419	1.136	$.226 \times 10^{-1}$	1.574	8651.4
(2.4)	$.606 \times 10^{-4}$	$.262 \times 10^{-4}$	$.313 \times 10^{-1}$	1.860	1.104	$.755 \times 10^{-2}$	2.126	8540.1
Off the east coast of Tohoku District $M \geq 5.0$, 2104 shocks								
(2.2)	$.895 \times 10^{-4}$	.771	$.318 \times 10^{-1}$	1.769	1.084	$.281 \times 10^{-2}$		17657.0
(2.3)	$.854 \times 10^{-4}$	$.488 \times 10^{-3}$	$.353 \times 10^{-1}$	1.407	1.121	$.159 \times 10^{-1}$	1.513	17620.2
(2.4)	$.844 \times 10^{-4}$	$.501 \times 10^{-4}$	$.327 \times 10^{-1}$	1.801	1.094	$.204 \times 10^{-2}$	1.668	17477.4
Off the east coast of Tohoku District $M \geq 4.5$, 4333 shocks								
(2.2)	$.139 \times 10^{-3}$	1.027	$.219 \times 10^{-1}$	1.588	1.031	$.175 \times 10^{-2}$		32216.4
(2.3)	$.133 \times 10^{-3}$	$.346 \times 10^{-3}$	$.266 \times 10^{-1}$	1.308	1.072	$.792 \times 10^{-2}$	1.435	32095.1
(2.4)	$.131 \times 10^{-3}$	$.416 \times 10^{-4}$	$.230 \times 10^{-1}$	1.605	1.043	$.103 \times 10^{-2}$	1.587	31859.7
Central and Western Honshu Island $M \geq 5.5$, 157 shocks								
(2.2)	$.500 \times 10^{-4}$	1.272	$.828 \times 10^{-2}$	1.867	1.108	$.881 \times 10^{-3}$		2190.4
(2.3)	$.461 \times 10^{-4}$	$.419 \times 10^{-4}$	$.107 \times 10^{-1}$	1.584	1.109	$.116 \times 10^{-1}$	2.001	2168.2
(2.4)	$.459 \times 10^{-4}$	$.259 \times 10^{-4}$	$.963 \times 10^{-2}$	1.464	1.097	$.155 \times 10^{-2}$	1.731	2172.0
Central and Western Honshu Island $M \geq 5.0$, 489 shocks								
(2.2)	$.505 \times 10^{-4}$	1.401	$.357 \times 10^{-2}$	1.072	1.043	$.242 \times 10^{-2}$		4746.5
(2.3)	$.479 \times 10^{-4}$	$.844 \times 10^{-4}$	$.371 \times 10^{-2}$	1.121	1.041	$.358 \times 10^{-2}$	1.725	4707.9
(2.4)	$.490 \times 10^{-4}$	$.235 \times 10^{-4}$	$.402 \times 10^{-2}$	1.199	1.046	$.135 \times 10^{-2}$	1.818	4688.4
Central and Western Honshu Island $M \geq 4.5$, 1285 shocks								
(2.2)	$.697 \times 10^{-4}$	2.481	$.745 \times 10^{-2}$	0.974	1.056	$.177 \times 10^{-2}$		9114.6
(2.3)	$.634 \times 10^{-4}$	$.919 \times 10^{-4}$	$.793 \times 10^{-2}$	1.120	1.052	$.204 \times 10^{-2}$	1.660	8946.8
(2.4)	$.645 \times 10^{-4}$	$.425 \times 10^{-4}$	$.814 \times 10^{-2}$	1.109	1.054	$.678 \times 10^{-3}$	1.668	8928.0
Central and Western Honshu Island $M \geq 4.0$, 3007 shocks								
(2.2)	$.111 \times 10^{-3}$	4.747	$.883 \times 10^{-2}$	1.052	1.031	$.850 \times 10^{-3}$		15708.5
(2.3)	$.906 \times 10^{-4}$	$.128 \times 10^{-3}$	$.830 \times 10^{-2}$	0.974	1.026	$.804 \times 10^{-3}$	1.510	15269.4
(2.4)	$.947 \times 10^{-4}$	$.511 \times 10^{-4}$	$.855 \times 10^{-2}$	1.040	1.027	$.316 \times 10^{-3}$	1.568	15174.6

metric defined in (4.3). The case where $\sigma_1 = \sigma_2$ and $\rho = 0$ reduces to the isotropic clustering in (2.5).

The extended models are fitted to each modified data set as described in Subsection 4.3. Table 2 summarizes the results for the three models with non-homogeneous background intensity (4.1) and isotropic spatial clustering (2.5), and Table 3 summarizes the results for the three models with non-homogeneous background intensity (4.1) and anisotropic spatial clustering (4.6). The goodness-of-fit of the extended versions of the model (2.4) is again far better than the other two for all the data sets in both Tables 2 and 3. The estimated parameters for each model and data set are rather similar to the corresponding ones in Table 1. It is

Table 3. Estimated parameters and the AIC values for the models with the non-homogeneous Poisson field for the background seismicity and the anisotropic clustering.

Model	ν (shock/day/degree^2)	K	c (days)	α (M^{-1})	p	d (degree)	q	AIC
Off the east coast of Tohoku District $M \geq 6.0$, 334 shocks								
(2.2)	$.441 \times 10^{-4}$	.257	$.276 \times 10^{-1}$	1.802	1.088	$.131 \times 10^{-1}$		3891.2
(2.3)	$.446 \times 10^{-4}$	$.817 \times 10^{-3}$	$.303 \times 10^{-1}$	1.472	1.119	$.474 \times 10^{-1}$	1.693	3910.6
(2.4)	$.432 \times 10^{-4}$	$.576 \times 10^{-4}$	$.266 \times 10^{-1}$	1.860	1.093	$.219 \times 10^{-1}$	2.318	3868.5
Off the east coast of Tohoku District $M \geq 5.5$, 900 shocks								
(2.2)	$.624 \times 10^{-4}$	.493	$.311 \times 10^{-1}$	1.789	1.100	$.592 \times 10^{-2}$		8594.8
(2.3)	$.626 \times 10^{-4}$	$.614 \times 10^{-3}$	$.357 \times 10^{-1}$	1.423	1.136	$.220 \times 10^{-1}$	1.569	8651.8
(2.4)	$.605 \times 10^{-4}$	$.223 \times 10^{-4}$	$.316 \times 10^{-1}$	1.873	1.103	$.762 \times 10^{-2}$	2.165	8536.8
Off the east coast of Tohoku District $M \geq 5.0$, 2104 shocks								
(2.2)	$.894 \times 10^{-4}$	.804	$.322 \times 10^{-1}$	1.782	1.084	$.266 \times 10^{-2}$		17640.1
(2.3)	$.850 \times 10^{-4}$	$.479 \times 10^{-3}$	$.349 \times 10^{-1}$	1.411	1.120	$.156 \times 10^{-1}$	1.513	17608.2
(2.4)	$.845 \times 10^{-4}$	$.433 \times 10^{-4}$	$.327 \times 10^{-1}$	1.809	1.094	$.208 \times 10^{-2}$	1.698	17463.7
Off the east coast of Tohoku District $M \geq 4.5$, 4333 shocks								
(2.2)	$.139 \times 10^{-3}$	1.050	$.220 \times 10^{-1}$	1.595	1.031	$.170 \times 10^{-2}$		32195.0
(2.3)	$.133 \times 10^{-3}$	$.336 \times 10^{-3}$	$.267 \times 10^{-1}$	1.323	1.072	$.799 \times 10^{-2}$	1.441	32064.4
(2.4)	$.131 \times 10^{-3}$	$.382 \times 10^{-4}$	$.231 \times 10^{-1}$	1.612	1.043	$.102 \times 10^{-2}$	1.600	31838.1
Central and Western Honshu Island $M \geq 5.5$, 157 shocks								
(2.2)	$.495 \times 10^{-4}$	1.013	$.882 \times 10^{-2}$	1.463	1.110	$.170 \times 10^{-2}$		2194.7
(2.3)	$.466 \times 10^{-4}$	$.257 \times 10^{-4}$	$.110 \times 10^{-1}$	1.641	1.116	$.175 \times 10^{-1}$	2.267	2172.5
(2.4)	$.462 \times 10^{-4}$	$.200 \times 10^{-4}$	$.977 \times 10^{-2}$	1.409	1.102	$.217 \times 10^{-2}$	1.843	2177.8
Central and Western Honshu Island $M \geq 5.0$, 489 shocks								
(2.2)	$.510 \times 10^{-4}$	1.514	$.364 \times 10^{-2}$	1.065	1.045	$.224 \times 10^{-2}$		4734.9
(2.3)	$.483 \times 10^{-4}$	$.651 \times 10^{-4}$	$.374 \times 10^{-2}$	1.129	1.043	$.448 \times 10^{-2}$	1.825	4705.4
(2.4)	$.495 \times 10^{-4}$	$.134 \times 10^{-4}$	$.408 \times 10^{-2}$	1.196	1.048	$.171 \times 10^{-2}$	1.960	4682.1
Central and Western Honshu Island $M \geq 4.5$, 1285 shocks								
(2.2)	$.697 \times 10^{-4}$	2.519	$.752 \times 10^{-2}$	0.959	1.056	$.177 \times 10^{-2}$		9097.1
(2.3)	$.638 \times 10^{-4}$	$.832 \times 10^{-4}$	$.813 \times 10^{-2}$	1.147	1.054	$.212 \times 10^{-2}$	1.682	8920.7
(2.4)	$.647 \times 10^{-4}$	$.407 \times 10^{-4}$	$.826 \times 10^{-2}$	1.106	1.055	$.675 \times 10^{-3}$	1.676	8909.7
Central and Western Honshu Island $M \geq 4.0$, 3007 shocks								
(2.2)	$.110 \times 10^{-3}$	4.739	$.886 \times 10^{-2}$	1.065	1.031	$.843 \times 10^{-3}$		15695.9
(2.3)	$.907 \times 10^{-4}$	$.124 \times 10^{-3}$	$.846 \times 10^{-2}$	0.997	1.026	$.818 \times 10^{-3}$	1.513	15254.6
(2.4)	$.946 \times 10^{-4}$	$.477 \times 10^{-4}$	$.866 \times 10^{-2}$	1.057	1.027	$.316 \times 10^{-3}$	1.577	15158.1

noteworthy that each AIC in Table 2 is remarkably smaller than the corresponding AIC in Table 1, although the justification is not established for the straightforward comparison by AIC with the models which include the function $\hat{\mu}_0(x, y)$ adjusted beforehand. The p-values which are less than 1 in Table 1 now become larger than 1 for the extended models, suggesting the stationarity of the models. This is consistent with our experience in estimating the ETAS model to the various seismicity data. The non-homogeneous background intensity models also made the estimate of q larger by comparing Tables 1 and 2. We feel that the inclusion of the non-homogeneous background intensity in the space-time modelling provides the significantly better performance for the both two seismic areas.

Comparing between Tables 2 and 3, we see that the parameters of the corresponding models are quite similar and also that the decrease of the AIC is not very large in spite of the implicitly adjusted many parameters for the anisotropic clustering in (4.4). Therefore, we cannot clearly see whether the anisotropic modelling remarkably improved the goodness-of-fit or not. However, throughout Tables 1, 2 and 3, it is clearly confirmed that model (2.4) for the spatial clustering is the best, which provides the answers to the questions raised in Subsection 2.2. Finally, by estimating the quantity $\int \hat{\nu}\mu_0(x,y)dxdy$ for the comparison of seismic features between in Region A (plate boundary) and in Region B (region of intraplate earthquakes), we see that the same statements as in the last paragraph of Subsection 3.1 hold.

5. Comparison between simulated and real catalogs

5.1 *Simulation procedure*

As is briefly described in Musmeci and Vere-Jones (1992), a simulated catalog can be developed following the thinning simulation procedures outlined in Ogata (1981). In particular, the example algorithm for the Hawkes' mutually-exciting (multivariate) point processes provided in the paper can be developed for the present space-time models. Given the intensity (2.1) with (2.2)–(2.4), events in a realization are simulated sequentially in such a way that, for each point, first the time coordinate (steps 1–7 below), and then the space and magnitude values are obtained (steps 8–11), starting from the form of the intensity function at the time of the proceeding event, and recalculating the intensity after the addition of the current point. Incidentally, although there is an alternative and more general thinning method which simultaneously simulates time and space coordinates, this takes much more c.p.u.-time because the rejections are dominant in the procedure.

Consider the superposed intensity with respect to locations

$$\lambda_A(t) = \nu_0 + \sum_{j=1}^{J} \nu_j(t), \tag{5.1}$$

where

$$\nu_0 = \int_A \mu(x,y)dxdy, \qquad J = \max\{j : t_j < t\}$$

and

$$\nu_j(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_\phi(t-t_j,x,y;M_j)dxdy = \frac{K}{(t-t_j+c)^p}\cdot e^{\alpha(M_j-M_0)} \tag{5.2}$$

where K is a constant due to the spatial integration over $\boldsymbol{R}^2$. Notice that $\lambda_A(t)$ is monotonically decreasing in time except increasing jumps from $\lambda_A(t_i)$ to $\lambda_A(t_i+)$ at any occurrence time t_i, where $\lambda(t+)$ denotes the right-limit of the function (the left-continuity is generally assumed), and the jump size at each i is $\nu_i(t_i)$. Then the algorithm to simulate $\{(t_i,x_i,y_i,M_i)\}$ follows:

1. Set the integers $a = b = c = 0$ and $i = 1$.

2. Set $s_a = 0$; generate a uniform random number U_b in $(0, 1]$, and put $\Lambda_c = \nu_0$ and $u_a = -\log(U_b)/\Lambda_c$.
3. If $u_a > T$ then stop; otherwise put $t_i = u_a$, put $J = 0$, and go to step 8.
4. Set b equal to $b + 1$ and generate a uniform random number U_b in $(0, 1]$; set a equal to $a + 1$, and put $u_a = -\log(U_b)/\Lambda_c$.
5. Put $s_a = s_{a-1} + u_a$. If $s_a > T$, stop; otherwise set b equal to $b + 1$, and generate U_b.
6. If $U_b > \lambda_A(s_a)/\Lambda_c$, set c equal to $c+1$ and put $\Lambda_c = \lambda_A(s_a)$; and go to step 4.
7. Put $t_i = s_a$; set b equal to $b + 1$, generate U_b, and select the smallest J such that $\sum_{j=0}^{J} \nu_j(t_i) > U_b \lambda_A(t_i)$.
8. If $J = 0$ then generate $(x_i, y_i) \in A$ according to the non-homogeneous Poisson intensity $\mu(x, y)$ (using the thinning), and go to step 11.
9. Otherwise (i.e., if $J > 0$), set b equal to $b + 1$, generate uniform random numbers U_b and V_b in $(0, 1]$, set $x_i = x_J + r(U_b, J)\cos(2\pi V_b)$ and $y_i = y_J + r(U_b, J)\sin(2\pi V_b)$ where the function $r(\cdot, \cdot)$ is defined below.
10. If (x_i, y_i) is not included in the bounded region A, then go to step 4.
11. Generate a magnitude M_i, set c equal to $c + 1$ and put $\Lambda_c = \lambda_A(t_i+)$; set i equal to $i + 1$, and go to step 4.

In step 9, the distance $r(U_b, J)$ between (x_i, y_i) and (x_J, y_J) is given by

$$\begin{aligned} r(U, j) &= d^{1/2} e^{-\alpha(M_j - M_0)/2} (-2 \log U)^{1/2} \quad \text{for} \quad (2.2), \\ &= d^{1/2} (U^{1/(1-q)} - 1)^{1/2} \quad \text{for} \quad (2.3), \\ &= d^{1/2} e^{-\alpha(M_j - M_0)/2} (U^{1/(1-q)} - 1)^{1/2} \quad \text{for} \quad (2.4), \end{aligned}$$

respectively. For the generation of magnitudes $\{M_i\}$ in step 11, we can either use the same magnitude series as the data or independently sample from the magnitude series. Since I have the strong suspicion that the magnitude sequence is not independent but rather long-range dependent (see Figs. 9 and 10 in Ogata and Abe (1991)), we adopted the former to concentrate on the space-time aspect of the models. Therefore, we terminate the generation of events when we get the same number of the events as the data instead of providing the time span beforehand as described in the above algorithm.

5.2 *Implementation of the simulation*

We performed simulations of space-time events using the selected model (2.4) with the parameters in Table 2 estimated for both Region A with $M \geq 5.0$ and Region B with $M \geq 4.5$. Figures 4a and 5a show the superposed locations of space-time events, and Figs. 4b and 5b show plots of events' latitudes and longitudes against occurrence times, respectively. Comparing Fig. 4a with Fig. 1a we feel a similar pattern in location. The same impression is given by Figs. 2a and 5a. But, from the comparison of the space-time plots in Figs. 1b and 4b, we see discriminating features between these. That is, synchronous and intensive clusters in high contrast with the rest are seen in real seismicity, while the clusters seems to occur independently and somewhat weakly in the simulated data. A somewhat similar impression is gained from the real and simulated seismicity shown in Figs. 2b and 5b. These distinctive features seem important for the ultimate

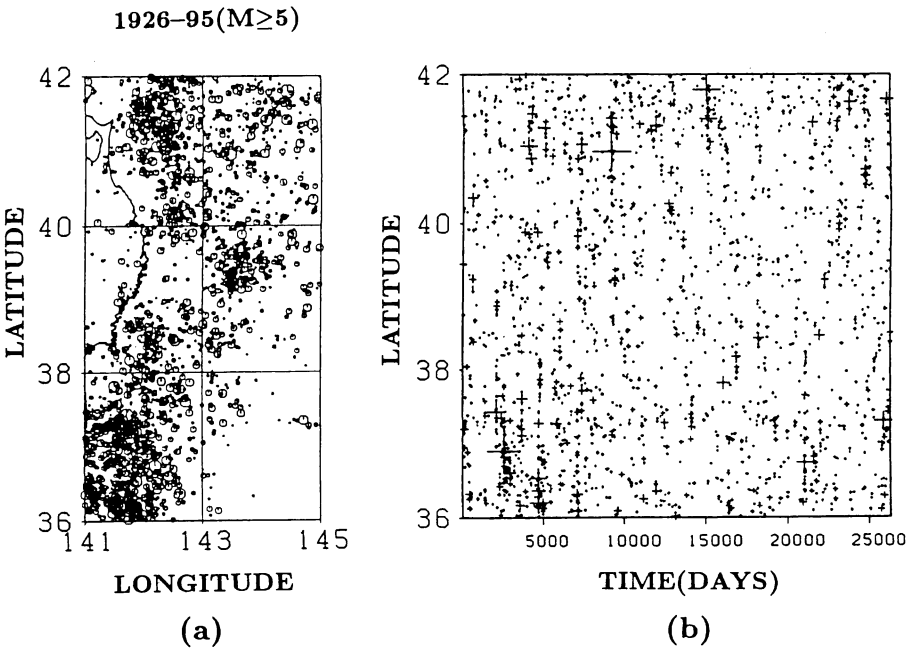

Fig. 4. Simulated events by the selected model (2.4) in Table 2 ($M \geq 5.0$) for Region A (Fig. 1); (a) locations, and (b) space-time plot.

objective of our study, and there are several possible reasons for them. Firstly, the long-range dependence of the occurrences may be stronger than that expected from the model, which assumes a stationary Poisson process for the background seismicity. Secondly, in the model, we have assumed that the magnitude sequence is independent of origin times, spatial locations and the position within a cluster; this is the basic assumption justifying the use of the partial log-likelihood for the estimation of the models. Such cross-correlations may not be negligible. It is also possible that the real seismicity may include forms of the quiescence and activation not present in the seismicity expected from the model. In reality, the seismic activity depends on the dynamical change of the underlying stress-field of an area owing to various geophysical factors, many of which are not considered in our models. Their study is certainly for the next step.

Now, we here fit the model (4.5) with the fixed $\nu_0(x, y)$ to the simulated data in order to examine the likelihood ratio test statistic

$$(-2) \log L(\hat{\theta})/L(\theta_0).$$

Under the null hypothesis $\theta_0 = (\nu_0, K_0, c_0, \alpha_0, p_0, d_0, q_0)$ by which the space-time point process data is simulated, we can expect that the likelihood ratio statistic should be distributed according to the chi-square distribution with 7 degrees of freedom if the simulation algorithm in Subsection 5.1 and the numerical likelihood calculation are both correct: the same approach was used by Ogata (1981) in the

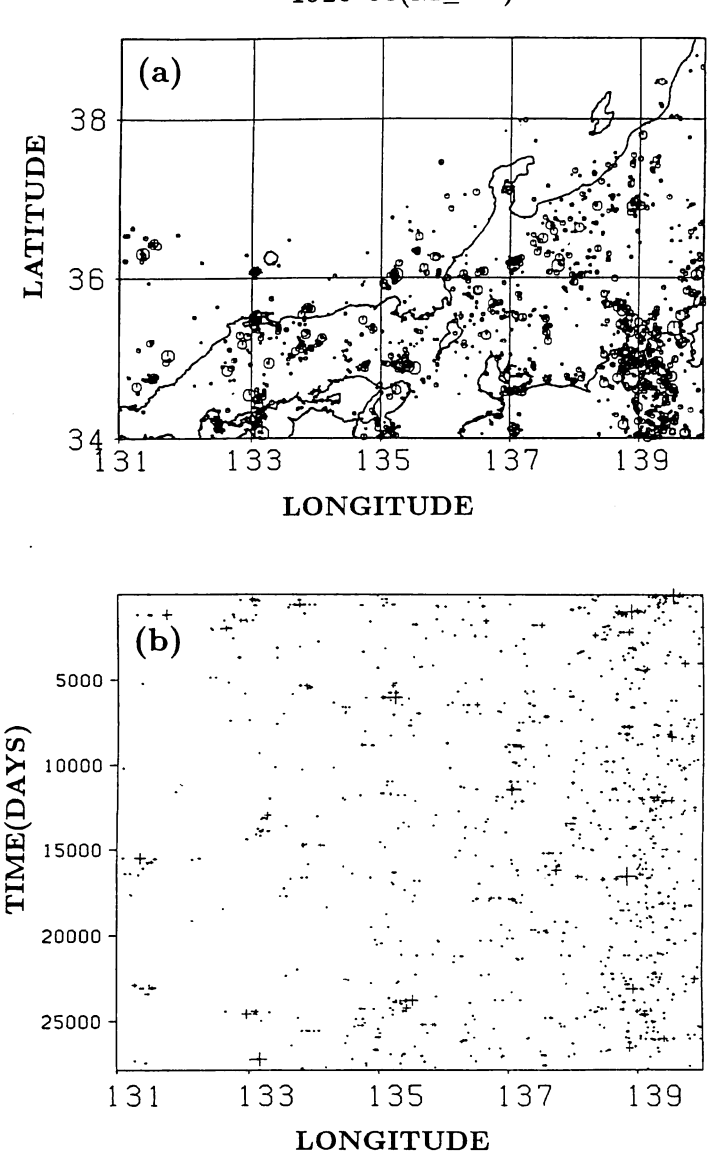

Fig. 5. Simulated events by the selected model (2.4) in Table 2 ($M \geq 4.5$) for Region B (Fig. 2); (a) locations, and (b) space-time plot.

first demonstration of the thinning simulation of intensity based point processes. Actually, in the present case, we found that the log-likelihood ratio values vary around 7.0 for the independent sets of simulated data. This eventually justifies the accuracy of the numerical integrations in (2.7) and also of $\int_A \mu_0(x,y)dxdy$ in the log-likelihood calculations.

6. Concluding remarks

Among the possible space-time extensions of the ETAS model in (2.2)–(2.4), the model with the response function (2.4) and its extensions to location dependent background intensity and anisotropic clustering all give the best fit to the data sets from both the plate boundary (Region A) and intraplate region (Region B). The results are stable for the data sets with different threshold magnitudes of earthquakes. The difference of AIC increases as the size of the data increases, suggesting that the above result gets clearer for the data with the lower threshold magnitude levels. The model (2.4), in contrast with the others, indicates that:

1. the functional form for $f(\cdot)$ in (2.5) and (4.6) extends long range (i.e. inverse power law) rather than short range (i.e. normal etc.), and
2. the scale factor depends on the magnitude in the form $\sigma(M) = e^{\alpha M}$.

Thus, conclusions and consequence of this study are as follows:

1. The clusters in space extend beyond the traditional aftershock regions, showing a much more diffuse boundary with power law decay rather than a more clearly defined region with a fairly sharp boundary converging faster than the exponential decay.
2. There may be perhaps two components (near field and far field) with different characteristics; the near field component corresponds to the traditional aftershock area around the ruptured fault, and the far field component may relate to the so called the 'aftershocks in wide sense' such as immigrations of earthquake activity or causal relations between distant regions, caused by tectonic changes of the stress-field due to the rupture.
3. The cluster regions scale with magnitudes firmly owing to the Utsu-Seki formula.

Further practical extensions of the models have been suggested to include non-homogeneous background seismicity but anisotropic features of earthquake clustering, but the above conclusions are unchanged for the extended models. Among the extensions, the location dependent intensity rate for the background seismicity is found to be effective for ensuring the stationarity of the models in time.

An efficient and practical simulation algorithm by thinning is provided, and the simulations from this revealed some discriminating space-time features between the real and simulated data, which may possibly relate to the ultimate aim of our study for earthquake prediction. The simulated data also justify the accuracy of the approximations used in the likelihood computation.

Acknowledgements

I would like to thank Koichi Katsura for his useful advice in solving the programming difficulties and for his generous help in drawing the figures in the present paper. I am particularly grateful to David Vere-Jones for his interest in the paper, and for constructive comments and suggestions which have led to significant revision and clarification of the present paper. Thanks are extended to Yan Kagan and anonymous referees for their useful comments and suggestions.

The present work was partially supported by Grand-in-Aid for Scientific Research (C) 07680329 from the Ministry of Education, Science, Culture and Sport.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, AC-19, 716–723.

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*, Springer, New York.

Fedotov, S. A. (1965). On regularities in distribution of strong earthquakes in Kamchatka, Kurile Islands and northeastern Japan, *Trudy Inst. Fiz. Zemli*, **36** (203), 66–93.

Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization, *Comput. J.*, **6**, 163–168 (in Russian).

Guo, Z. and Ogata, Y. (1997). Statistical relations between the parameters of aftershocks in time, space and magnitude, *Journal of geophysical Research*, **102**, 2857–2873.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes, *J. Roy. Statist. Soc. Ser. B*, **33**, 438–443.

Kagan, Y. Y. (1991). Likelihood analysis of earthquake catalogues, *Journal of geophysical Research*, **106**, 135–148.

Kagan, Y. Y. and Jackson, D. D. (1991). Seismic gap hypothesis: Ten years after, *Journal of geophysical Research*, **96**, 21419–21431.

Kagan, Y. Y. and Jackson, D. D. (1995). New seismic gap hypothesis: Five years after, *Journal of geophysical Research*, **100**, 3943–3959.

Kagan, Y. Y. and Knopoff, L. (1978). Statistical study of the occurrence of shallow earthquake, *Geophysical Journal of the Royal Astronomical Society*, **55**, 67–86.

Kagan, Y. Y. and Knopoff, L. (1980). Spatial distribution of earthquakes: the two-point correlation function, *Geophysical Journal of the Royal Astronomical Society*, **62**, 303–320.

Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relation in seismology, *Bull. Seismol. Soc. Amer.*, **65**, 1073–1096.

Kowalik, J. and Osborne, M. R. (1968). *Methods for Unconstrained Optimization Problems*, American Elsevier, New York.

McCann, W. R., Nishenko, S. P., Sykes, L. R. and Krause, J. (1979). Seismic gaps and plate tectonics: Seismic potential for major boundaries, *Pure and Applied Geophysics*, **117**, 1082–1147.

Mogi, K. (1968). Sequential occurrences of recent great earthquakes, *Journal of Physics of the Earth*, **16**, 30–36.

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

Nishenko, S. P. and Sykes, L. R. (1993). Comment on "Seismic gap hypothesis: Ten years after" with Reply by Jackson, D. D. and Kagan, Y. Y., *Journal of geophysical Research*, **98**, 9909–9920.

Ogata, Y. (1981). On Lewis' simulation method for point processes, *IEEE Trans. Inform. Theory*, IT-**30**, 23–31.

Ogata, Y. (1987). Long term dependence of earthquake occurrences and statistical models for standard seismic activity (in Japanese), *Suri Zisin Gaku* (*Mathematical Seismology*) *II* (ed. M. Saito), ISM Cooperative Research Report 3, 115–124, *Inst. Statist. Math.*, Tokyo.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.*, **83** (401), 9–27.

Ogata, Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis, *Tectonophysics*, **169**, 159–174.

Ogata, Y. (1992). Detection of precursory seismic quiescence before major earthquakes through a statistical model, *Journal of geophysical Research*, **97**, 19845–19871.

Ogata, Y. (1993). Space-time modeling of earthquake occurrences, *Bulletine of the International Statistical Institute*, **55**, Contributed papers, Book 2, 249–250.

Ogata, Y. and Abe, K. (1991). Some statistical features of the long term variation of the global and regional seismic activity. *International Statistical Review*, **59**, 139–161.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.*, **40**, 29–39.

Ogata, Y., Utsu, T. and Katsura, K. (1995). Statistical features of foreshocks in comparison with other earthquake clusters, *Geophysical Journal International*, **121**, 233–254.

Ohtake, M., Matumoto, T. and Latham, G. V. (1977). Seismicity gap near Oaxaca, southern Mexico as a probable precursor to a large earthquake, *Pure and Applied Geophysics*, **115**, 375–385.

Rathbun S. L. (1993). Modeling marked spatio-temporal point patterns, *Bulletine of the International Statistical Institute*, **55**, Book 2, 379–396.

Rathbun S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes, *J. Statist. Plann. Inference*, **51**, Special issue on Spatial Statistic, Part II, 55–74.

Utsu, T. (1961). Statistical study on the occurrence of aftershocks, *The Geophysical Magazine*, **30**, 521–605.

Utsu, T. (1969). Aftershocks and earthquake statistics (I): some parameters which characterize an aftershock sequence and their interaction, *Journal of the Faculty of Science, Hokkaido University, Ser. VII (geophysics)*, **3**, 129–195.

Utsu, T. (1972). Large earthquakes near Hokkaido and the expectancy of the occurrence of a large earthquake off Nemuro, *Report of the Coordinating Committee for Earthquake Prediction*, **7**, 1–13 (in Japanese).

Utsu, T. and Seki, A. (1955). Relation between the area of aftershock region and the energy of the main shock, *Zisin (Journal of the Seismological Society of Japan) 2nd Ser. ii*, **7**, 233–240 (in Japanese with English summary).

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