Modeling of Spatio-Temporal Seismic Activity and Its Residual Analysis

Yosihiko Ogata, Koichi Katsura and Masaharu Tanemura
The Institute of Statistical Mathematics, Tokyo, 106-8569, Japan
email: ogata@ism.ac.jp

We assume that tectonic seismic activity is given by a superposition of earthquake clusters triggered by relatively large earthquakes. Based on several established empirical laws in the traditional studies of aftershock statistics, we construct a space-time point-process model in terms of the conditional intensity function which needs seven parameters to characterize seismicity in a geophysical region. However, as the data size increases, each characteristic parameter takes significantly different values from place to place. Thus we further need to consider a hierarchical extension of the model such that each parameter is a function of location. Specifically, it is represented by two dimensional piecewise linear function consisting of facets defined on Delaunay tessellated triangles whose vertices are locations of earthquakes in the data. Then the penalized log likelihood is considered for the trade-off between the good fit to the data and the uniformity of each function (i.e., the facets of the piecewise linear function are as flat as possible). A Bayesian method is applied for the optimal tuning of the trade-off to the long-term earthquake occurrence data in and around Japan. Thus we have spatial images of the parameter changes (the maximum a posteriori estimate) which show regional characteristics of seismicity. Our final goal is to detect a space-time volume in which a certain unusual seismicity change is revealed. For this purpose we consider space-time piecewise linear function defined on Delaunay tessellated tetrahedra whose vertices are locations and times of earthquakes in the data. Then this function and the previously estimated space-time conditional intensity function multiply to make a new conditional intensity function, which is applied to the same earthquake data to estimate the piecewise linear function. The estimation is carred out by means of the Bayesian method for the similar trade-off. Thus the estimated function shows a three dimensional image indicating space-time volumes of standard, high or low seismicity relative to the evaluated activity by the previously obtained space-time point-process model, according to that the function takes equal to, larger or smaller than 1, respectively. Our serious interest is particularly placed on the last case called 'relative quiescence' to see whether this could be useful as a precursor to predict the time and location of forthcoming large earthquakes.
KEYWORDS: Bayesian smoothing procedure; Conditional intensity function; Delaunay tessellation (2D and 3D); Hierarchical space-time point-process model; Residual image; Relative seismic quiescence;

1 Introduction

Space-time aspects of earthquake prediction have been developed in seismology on the basis of seismicity data sets. 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. Furthermore, seismic activity before a large earthquake can be quiet not only in the seismic gap but also in its wide neighborhood (Inouye, 1965). Together with the observation of the seismic quiescence, the gap theory gave successful predictions in some cases (Utsu, 1968, and Ohtake et al., 1977, for instance). However, this is not frequently the case and gaps do not always appear very clearly, especially in the 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; and Nishenko and Sykes, 1993).

In seismology it seems very difficult to make any other mechanism-based forecasting of seismicity in time nor space. Namely, seismicity patterns vary substantially from event to event, even though some of the fundamental physical process leading to an earthquake may be common to all events. They postulate that fault zone heterogeneity and complexity are responsible for the observed variations. Therefore, the problem of developing stochastic point-process models for earthquake occurrence may be of great interest to seismologists.

The ultimate objective of our study is to indicate the location of the anomalous area as well as the corresponding temporal anomalies for an intermediate earthquake prediction. In the similar manner to the application of the Epidemic Type Aftershock Sequence (ETAS) model for detection of relatively quiet period (Ogata; 1988, 1989, 1992 and 2001), we believe that the sensitivity in detecting such anomalies can be amplified by contrasting the observed seismic activity with the predicted intensity 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, the hierarchical space-time ETAS (HIST-ETAS) model is derived to represent the various seismicity patterns in and around Japan. Then the modeling and method of the residual analysis of a space-time point process is introduced for the detection of anomalous changes in time relative to the estimated HIST-ETAS model. These are implemented to the earthquake occurrence data in Japan region for illustrating the usefulness of the proposed models and methods. In Section 2 the data for the illustration and the basic frame of space-time point-process modeling are explained. In Section 3 the derivation of the optimal space-time extension of the ETAS model are reviewed to make a further extension to a hierarchical version of the model (HIST-ETAS model) where its parameters are dependent on the location of the earthquakes (Section 4). To estimate the model, we need a Bayesian smoothing method. The practical and numerical aspects of the procedure is also given in Section 4, where the real application is made for the data explained in Section 2. The space-time residual analysis of the model is presented in Section 5. The last section describes the conclusions.

2 Data and point processes

We use the Hypocenter Data File of Japan Meteorological Agency (JMA) as the source of the data in this study. We select the data of earthquakes of magnitude (M) 5.0 or larger with depths shallower than 100 km throughout whole Japan (within the rectangular region bounded by 128°E and 149°E meridians, and 26°N and 47°N parallels) for the period from 1926 through 1995. There are substantial changes in detection capability of earthquakes for the last 75 years as the seismic network of the JMA has been developed. The threshold magnitude M5.0 is taken because this earthquake size or larger are regarded to be almost detected throughout the whole period and the Japan area except the north-end off-shore and southern end of Izu-Ogasawara (Izu-Bornin) Islands in early years. Incidentally, a quake by M5 event is really astonishing, usually causes substantial damages in the region near to the hypocenter. The accuracy of the hypocenter depth of the JMA catalog was not enough for the present analysis until 1983, so that we ignore the depth axis and consider only longitude and latitude for the location of an earthquake restricting ourselves to shallow events down to 100 km depth. Figure 1 shows such data set of 4586 earthquakes in space and time.

*** Figure 1 around here ***

Thus, we are concerned with point-process models for the data of occurrence times and locations of earthquakes whose magnitudes equal to or larger than a certain cut-off magnitude M_c (i.e., M5.0 in the present data). The conditional intensity function $\lambda(t,x,y|H_t)$ of a space-time point process is defined as the occurrence rate at time t and the location (x,y) conditional on the past history of the occurrences such that

$$\begin{aligned} &Prob\{\text{an event occurs in } dt \times dx \times dy \mid H_t\} \\ &\quad = \lambda(t,x,y|H_t)dtdxdy + o(dtdxdy) \end{aligned}$$

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, with the corresponding epicenters $\{(x_i, y_i)\}$ and magnitudes $\{M_i\}$.

Assuming stationarity, Hawkes's self-exciting point-process model (Hawkes, 1971) is extended to the following form

$$\begin{aligned} \lambda(t,x,y|H_t) = \mu(x,y) + &\sum_{\{i:\, t_i<t\}} \kappa(M_i) \times \\ &g(t-t_i;\ M_i) f(x-x_i,\ y-y_i;\ M_i) \quad (1) \end{aligned}$$

at a space-time coordinate $(t, x, y) \in [0, T] \times A$. Here $\kappa(M_i)$ is the expected number of aftershocks with magnitude M_c or larger, triggered by the event i; $g(\tau;\ M_i)$ is probability density in time; and $f(x, y;\ M_i)$ is spatial probability density. There are various parametric forms of these functions as described below. Hereafter we represent the conditional intensity function $\lambda_{\boldsymbol{\theta}}(t, x, y|H_t)$ by the parameter vector $\boldsymbol{\theta}$.

Then, given the data of origin times and space coordinates of earthquakes associated with their magnitudes $\{(t_i, x_i, y_i, M_i);\ M_i \geq M_c,\ i = 1, \cdots, n\}$ during a period $[0, T]$ and in a region A, the log of likelihood of the model conditional on the sequence of the magnitudes is given by

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \lambda_{\boldsymbol{\theta}}(t_i, x_i, y_i|\ H_t) - \int_0^T \int\int_A \lambda_{\boldsymbol{\theta}}(t, x, y|\ H_t) dt dx dy. \quad (2)$$

An important aspect of space-time modeling is parametric form of the functions $\kappa(M_i)$, $g(\tau; M_i)$, and $f(x, y; M_i)$ in (1). Musmeci and Vere-Jones (1992) suggest a diffusion type function and a product-Cauchy form to apply to the Italian historical earthquake data. At about the same time, Kagan (1991) suggested other parametric forms based on the investigations of the second-order statistical features in time and space of hypocenter catalogs in California (e.g., Kagan and Knopoff, 1978, 1980). Ogata (1993) and Rathbun (1993, 1994) considers a function of products of the modified Omori decay in time, the Gaussian kernel in space, and the exponential in magnitude.

3 Space-time ETAS models

3.1 The ETAS model

Ignoring the location coordinates, Ogata (1988) introduced a point-process model for earthquake occurrence data as the superposition of the modified Omori functions (Utsu, 1957, 1961) for aftershocks of each event i such that

$$\lambda(t|H_t) = \mu + \sum_{t_i < t} \frac{K_i}{(t - t_i + c)^p}. \quad (3)$$

A crucial point of the model here is that the restriction among the parameters $\{K_i\}$ is considered as a function of the magnitude M_i of the corresponding event i, besides a reference magnitude M_0 of the data set, according to the exponential function form,

$$K_i = K_0 e^{\alpha(M_i - M_0)}, \quad (4)$$

which is based on the assumption that the expected number of aftershocks are proportional to the aftershock area. Relevantly, the statistical law between aftershock area A (km^2) and the magnitude M of the mainshock was discovered by Utsu and Seki (1955) such that

$$\log_{10} A = M - D, \quad (5)$$

where D is about 4.0 for all earthquakes ($6.0 \leq M_J \leq 8.5$) in and around Japan. Interestingly, the intersect D of the regression line for the earthquakes occurred in land and sea are significantly different to each other, while the slope is not so.

Thus, the maximum likelihood estimates (MLE) of the parameters $\boldsymbol{\theta} = (\mu, c, p, K_0, \alpha)$ is obtained by the maximization of the same log-likelihood as in (2) by substituting the conditional intensity in (3) with (4). The model for ordinary seismicity defined by (3) with (4) in terms of the occurrence rate of shocks is called the *Epidemic Type Aftershock-Sequences* (ETAS) model.

3.2 Extensions of the ETAS model

In order to recover the ETAS model from the space-time models (1) when the intensity is integrated (i.e. superposed) with respect to the space variables (location coordinates), the cluster size function (average number of aftershocks for the event of magnitude M) in (1) is not only set by

$$\kappa(M) = K_0 e^{\alpha(M - M_c)}, \quad (6)$$

but also, for the time probability density function in (1), the modified Omori function (Utsu, 1957, 1961)

$$g(t) = \frac{(p-1)\ c^{p-1}}{(t + c)^p} \quad (7)$$

is adopted, which is dependent on the lapse time t from the occurrence of triggering event but no more dependent of its magnitude.

As for the spatial response function f in (1) for the aftershocks, we see mostly that locations of aftershocks are approximately elliptically distributed (see Utsu, 1969, for instance) depending on the ratio of the length to width of the ruptured fault, its dip angle or the location errors of aftershock hypocenters. Therefore its common form should be

$$f(x,y;M) = \frac{|\Sigma|}{2\pi\sigma(M)} f_0 \left\{ \frac{(x,y)\Sigma(x,y)^t}{\sigma(M)^2} \right\}, \qquad (8)$$

in which we will examine whether or not the scale factor $\sigma(M)$ works corresponding to the Utsu-Seki law (5), particularly, in the case where an inverse power function is assumed for f_0 (like the above Modified Omori function); namely, $\sigma(M)$ may be either dependent of magnitude M or just a constant. Furthermore, the anisotropic feature of the clusters is modeled by a positive-definite 2×2 symmetric matrix Σ such that

$$\begin{aligned} r(x,y) &\equiv (x,y)\Sigma(x,y)^t \\ &= \frac{1}{\sqrt{1-\rho^2}} \left(\frac{\sigma_2}{\sigma_1} x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2} y^2 \right), \qquad (9) \end{aligned}$$

for positive constants σ_1, σ_2 and ρ; and the estimation procedure of these parameters for each earthquake will be discussed in Section 4.1 later. Then the main issues in modeling the spatial response function are:

1. the character of the functional form of $f_0(r)$: either short range decay (e.g., normal etc.) or long range decay (e.g.., inverse power law), namely, either $f_0(r) \propto \exp(-r/d)$ or $\propto (r+d)^{-q}$ for some $d > 0$, and
2. the dependence of the scale factor on magnitude: namely, either $\sigma(M) \propto 1$ or $\propto \exp(\alpha M)$.

Ogata (1998) implemented the comparison of those possible models by fitting the two data sets from plate boundary and intraplate regions, respectively, first assuming a constant back-ground intensity μ and further the identity matrix for Σ. Then the best model was of the form

$$f(x,y;M) = \frac{(q-1)d^{q-1}|\Sigma|}{2\pi e^{\alpha(M-M_c)}} \left\{ \frac{(x,y)\Sigma(x,y)^t}{e^{2\alpha(M-M_c)}} + d \right\}^{-q}. \qquad (10)$$

The result did not change for the different data sets and different threshold magnitudes M_c. Moreover, the differences of the *AIC* values increased with increased data size by decreased threshold magnitude, which implies that the result get clearer for data set with lower threshold magnitude levels.

In summary, the result indicates that: [1] the clusters in space extend beyond the traditional aftershock regions, having a diffuse boundary with power law decay rather than forming a well-defined region with a fairly sharp boundary; [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, and the far field component may relate to the so called the immigrations of earthquake activity or causal relations between the distant regions; and [3] the cluster regions scale with magnitudes firmly according to the Utsu-Seki formula.

For the numerical computations the adopted model is rewritten by the following simplified form of parameterization,

$$\begin{aligned} \lambda(t,x,y|H_t) &= \mu + K_0 \sum_{t_i<t} (t-t_i+c)^{-p} \times \\ &\quad \left\{ \frac{r(x-x_i, y-y_i)}{e^{2\alpha(M_i-M_c)}} + d \right\}^{-q}, \qquad (11) \end{aligned}$$

where $r(x,y)$ is given in (9), and the parameters to be estimated are $\boldsymbol{\theta} = (\mu, K_0, c, \alpha, p, d, q)$.

4 Modeling of spatially heterogeneous seismicity

4.1 Preliminary arrangement of data for anisotropic clusters

It is often the case that the epicenter of a mainshock is located at the margin of its aftershock area because the epicenter listed in the catalog corresponds to the location of earthquake rupture initiation. In such an earthquake (say, i-th event), the epicenter location (x_i, y_i) in the catalog is not quite suitable for the present models in (1) with spatial response function given in (10) with (9), or the intensity (11) with (9). Therefore, in place of the location (x_i, y_i), we consider the *centroid epicenter of aftershocks* which is estimated as follows: First, we have to identify clusters of aftershocks, the algorithm of which is provided below. Then, we take the average of the locations of the cluster members to replace the catalogue's epicenter of the mainshock only when the two are significantly different to each other as determined by the method below.

Also, we see mostly that locations of aftershocks are approximately elliptically distributed (see Utsu, 1969, for instance) as described by the matrix Σ in (9) owing to the several reasons described in Section 3.2. For the location coordinates of events relative to the mainshock's epicenter in each cluster, a bivariate Normal distribution is fitted to obtain the MLEs $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$, only when each of them is respectively significantly different from the null hypothesis (i.e., $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0$). Namely, the model with the minimum AIC value (Akaike, 1974) is adopted among all the nested models that include the null hypothesis (c.f. Ogata, 1998). For the rest of events in the cluster, the null hypothesis is always adopted; namely, the same epicenter as that of catalog and the identity matrix for Σ.

The algorithm for identifying the aftershock clusters starts with selecting the largest shock in the catalog for the *mainshock*. If there are plural largest shocks, the earliest one is adopted for the mainshock. Then, to form a cluster, we set a space-time window where the bounds of distance and time depending on the magnitude of the mainshock which are based on the empirical laws of aftershocks (c.f., Utsu, 1969; Ogata et al, 1995) as explicitly given 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 up 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 $M = 5 \sim 6$ and 1000 days for $M = 8$) after the main shock. The side length of square area for the space window centered at a mainshock epicenter is taken to be $2 \times (0.015 \times 10^{0.5M-2} + \epsilon)$ degrees (i.e., about 70 km for $M = 5$ and 400 km for $M = 8$): here, we took $\epsilon = 0.3$ degrees (33.3 km) taking account of the error of epicenter determination in early years of the JMA catalog.

4.2 Hierarchical space-time ETAS model

We learn by experience that, as the number of data events increases by lowering the magnitude threshold or as the area of the investigation get wider, the difference of parameter values of the model at different subregions get more significant. Therefore, it will be practical to assume that the parameters of the model in (11), except for c and d, are the functions of location (x, y) as follows.

Consider the Delaunay triangulation (e.g., Green and Sibson, 1978) of the whole region A tessellated by the locations of earthquakes and some additional points $\{(x_i, y_i) \in A;\ i = 1, \cdots, N + n\}$, where n is the number of the additional points on the rectangular boundary including the corners. Figure 2b show such tessllation based on the epicenters of the present data set consisting of the N=4586 events (Figure 2a) and the additional points (n=81), which make 9251 Delaunay triangles.

*** Figures 2a and b around here ***

Then, consider piecewise linear function $\varphi(x, y)$ defined on the tessellation where, for each vertex (x_i, y_i) of a triangle, it takes the value $\phi_i = \varphi(x_i, y_i)$. Thus the function value at any location (x, y) is given as follows: let (x_1, y_1), (x_2, y_2) and (x_3, y_3), say, be the coordinates of vertices of the Delaunay triangle which includes (x, y), then consider the linear equation

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= x \\ a_1y_1 + a_2y_2 + a_3y_3 &= y \\ a_1 + a_2 + a_3 &= 1 \end{aligned}$$

to obtain the non-negative solution $\hat{a}_1$, $\hat{a}_2$ and $\hat{a}_3$ so that we have

$$\varphi(x, y) = \hat{a}_1\phi_1 + \hat{a}_2\phi_2 + \hat{a}_3\phi_3$$

Using such piecewise linear functions we define the functions for the parameters μ, K_0, α, p and q of the space-time ETAS model (11) in the form of

$$\begin{aligned} &\mu(x, y) = \hat{\mu}\, e^{\varphi_1(x,y)}; \quad K(x, y) = \hat{K}_0 e^{\varphi_2(x,y)}; \\ &\alpha(x, y) = \hat{\alpha}\, e^{\varphi_3(x,y)}; \quad p(x, y) = \hat{p}\, e^{\varphi_4(x,y)}; \\ &\text{and} \quad q(x, y) = \hat{q}\, e^{\varphi_5(x,y)}, \end{aligned} \tag{12}$$

avoiding negative function values, where $\hat{\mu}$, $\hat{K}_0$, $\hat{\alpha}$, $\hat{p}$ and $\hat{q}$ are suitable reference values. From here on, we call the conditional intensity function (11) with location-dependent parameters (12) the *hierarchical space-time ETAS (HIST-ETAS) model*. This model actually need about five times as many coefficients as the number of events in the data, namely, the unknown parameters are $\boldsymbol{\theta} = \{(\phi_{ki};\ i = 1, \cdots, N + n);\ k = 1, \cdots, 5\}$ with $\phi_{ki} = \varphi_k(x_i, y_i)$ in (12). Therefore we need to formulate the penalized log likelihood (Good and Gaskins, 1971)

$$R(\boldsymbol{\theta}|\ \boldsymbol{w}) = \log L(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}|\ \boldsymbol{w}), \tag{13}$$

where $\boldsymbol{w} = (w_1, \cdots, w_5)$ and the penalty Q is defined by

$$Q(\boldsymbol{\theta}|\ \boldsymbol{w}) = \sum_{k=1}^{5} w_k \int\!\!\int_A \left\{ \left(\frac{\partial \varphi_k}{\partial x}\right)^2 + \left(\frac{\partial \varphi_k}{\partial y}\right)^2 \right\} dxdy. \tag{14}$$

The penalized log likelihood is thus considered for the trade-off between the good fit to the data and the uniformity of each function (i.e., the facets of the piecewise linear function are as flat as possible).

4.3 An objective Bayesian procedure

In order to find optimal weights $\hat{\boldsymbol{w}} = (\hat{w}_1, \cdots, \hat{w}_5)$ we adopt a Bayesian procedure where the exponential to the negative penalty stands for a prior density, denoted by $prior(\boldsymbol{\theta}|\boldsymbol{w})$ hereafter. Since the penalty function in (14) has quadratic form with respect to the parameters $\boldsymbol{\theta}$, the prior is a multivariate Normal distribution in which the variance-covariance matrix is the inverse of the Hessian matrix H_Q that consists of the negative second order partial derivatives of the penalty function Q. Actually, the Hessian matrix in the present case has the diagonal form of five independent sub-matrices corresponding to each φ_k-function such that

$$H_Q = diag\{H_Q^1, H_Q^2, H_Q^3, H_Q^4, H_Q^5\}, \tag{15}$$

since we do not assume any prior restrictions between the different φ_k-functions. Here all sub-matrices of H_Q^k are sparse and have the same configuration of non-zero elements; that is to say, the (i, j)-element is non-zero if and only if the pair of points i and j are vertices of the same Delaunay triangle.

Then, we consider the posterior probability density function

$$p(\boldsymbol{\theta}|\ \boldsymbol{w}) = \frac{L(\boldsymbol{\theta}) \cdot prior(\boldsymbol{\theta}|\ \boldsymbol{w})}{\boldsymbol{\Lambda}(\boldsymbol{w})} \tag{16}$$

with normalizing factor

$$\boldsymbol{\Lambda}(\boldsymbol{w}) = \int_{\Theta} L(\boldsymbol{\theta}) \cdot prior(\boldsymbol{\theta}|\ \boldsymbol{w})\ \boldsymbol{d\theta}. \tag{17}$$

Maximization of (17) or its logarithm with respect to the hyperparameters $\boldsymbol{w}$ is called the method of the type II maximum likelihood, due to Good (1965). Then the minimized solution of the penalized log-likelihood in (13) corresponds to the (optimal) *maximum a posteriori (MAP)* estimate.

However, the integration of the posterior function in (17) cannot be analytically carried out because the likelihood function of the point-process model is not Normal. Nevertheless, by virtue of the Normal prior, the Normal approximation of the posterior function is useful. That is to say, the penalized log-likelihood is well approximated by the quadratic form

$$\begin{aligned} T(\boldsymbol{\theta}|\boldsymbol{w}) &\equiv \log L(\boldsymbol{\theta}) + \log\ \{prior(\boldsymbol{\theta}|\boldsymbol{w})\} \\ &\approx T(\hat{\boldsymbol{\theta}}|\boldsymbol{w}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) H_T(\hat{\boldsymbol{\theta}}|\boldsymbol{w})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^t, \end{aligned} \tag{18}$$

where $\hat{\boldsymbol{\theta}} = arg\{\max_{\boldsymbol{\theta}} T(\boldsymbol{\theta}|\boldsymbol{w})\}$, and $H_T(\boldsymbol{\theta}|\boldsymbol{w})$ is the Hessian matrix of $T(\boldsymbol{\theta}|\boldsymbol{w})$ consisting of its negative second-order partial derivatives with respect to $\boldsymbol{\theta}$. We further assume that the Hessian matrix below Equation (18) is approximated by the diagonal form of the five sub-matrices,

$$H_T = diag\{H_T^1, H_T^2, H_T^3, H_T^4, H_T^5\}, \tag{19}$$

which also assume independency between the coefficients of the different φ_k-functions in the penalized log-likelihood (13).

Thus, the log likelihood of the present Bayesian model is given by

$$\begin{aligned} &\log \Lambda(\boldsymbol{w}) = \log \int L(\boldsymbol{\theta}) \cdot prior(\boldsymbol{\theta}|\boldsymbol{w})\ \boldsymbol{d\theta} \\ &\approx T(\hat{\boldsymbol{\theta}}|\boldsymbol{w}) - \frac{1}{2}\log \|H_T(\hat{\boldsymbol{\theta}}|\boldsymbol{w})\| + \frac{\dim(\boldsymbol{\theta})}{2} \cdot \log 2\pi \\ &= R(\hat{\boldsymbol{\theta}}|\boldsymbol{w}) - \frac{1}{2}\log \|H_R(\hat{\boldsymbol{\theta}}|\ \boldsymbol{w})\| + \frac{1}{2}\log \|H_Q(\hat{\boldsymbol{\theta}}|\ \boldsymbol{w})\|, \end{aligned} \tag{20}$$

where H_R is the similar diagonal Hessian matrix of the functions R in (13), and $\|\cdot\|$ is the determinant of the matrices.

In order to get the optimal hyperparameters, one has to repeat the following calculations:

(A). For a given $\boldsymbol{w}$ being fixed, get the maximizing parameters $\hat{\boldsymbol{\theta}}$ of the penalized log likelihood T in (18) with respect to $\boldsymbol{\theta}$ on the one-dimensional straight line determined by the initial parameter vector $\boldsymbol{\theta}_0$ and gradient vector of the function $T(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ (Linear Search, e.g., Kowalik and Osborne, 1968).

(B). Set this maximizing parameter $\hat{\boldsymbol{\theta}}$ to be the next starting parameter $\boldsymbol{\theta}_0$. Then, using the gradient vector of the function $T(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ and solving Incomplete Cholesky Conjugate Gradient (ICCG) method (e.g., Mori, 1986), find the new vector for the direction of the next Linear Seach to repeat the present step up until the series of maximumized function values of T and $\hat{\boldsymbol{\theta}}$ converge to the max T and the *maximum a posteriori (MAP)* solution, respectively.

(C). Calculate $\log\Lambda(\boldsymbol{w})$ using the approximation in (18) around $\hat{\boldsymbol{\theta}}$. We further need to maximize the log-likelihood with respect to the hyperparameters $\boldsymbol{w}$ by a direct search method such as the simplex method (e.g., Kowalik and Osborne, 1968; Murata, 1992).

(D). The steps (A)-(C) are repeated in turn up until the step (C) converges.

It is noteworthy that the convergence in step (B) have been very fast in spite of the very high dimensionality of $\boldsymbol{\theta}$. This can be expected if quadratic approximations of T are adequate for a wide enough region around the MAP solution. After all, assuming unimodality of the posterior function, one can get the optimal MAP solution $\hat{\boldsymbol{\theta}}$ for the maximum likelihood estimate $\hat{\boldsymbol{w}}$. Also, the Hessian matrix for error assessment is given by $\mathbf{N}(\hat{\boldsymbol{\theta}}, H_T(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{w}})^{-1})$.

4.4 Application of nonhomogeneous Poisson field to the spatial data

In order to demonstrate the present Bayesian procedure, we apply this to the data of spatial point pattern of the earthquake locations $\{(x_i, y_i),\ i = 1, \cdots, 4586\}$ in Figure 2a. The intensity function of non-homogeneous Poisson field of the form

$$\lambda_{\boldsymbol{\theta}}(x,y) = \exp\{\varphi_{\boldsymbol{\theta}}(x,y)\},$$

is considered to avoid taking negative values, where $\varphi_{\boldsymbol{\theta}}(x,y)$ is the same sort of piecewise linear function on the Delaunay tessellation as described in Section 4.2. Then we consider the penalized log-likelihood in (13) where the log-likelihood is given by

$$\begin{aligned}\log L(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log \lambda_{\boldsymbol{\theta}}(x_i, y_i) - \int\int_A \lambda_{\boldsymbol{\theta}}(x,y)dxdy, \\ &= \sum_{i=1}^{n} \varphi_{\boldsymbol{\theta}}(x_i, y_i) - \int\int_A \exp\{\varphi_{\boldsymbol{\theta}}(x,y)\}dxdy,\end{aligned}$$

and the penalty in (14) is restricted by setting $w_2 = w_3 = w_4 = w_5 = 0$. By the same computation procedure as explained in Section 4.3, we have the optimum hyper-parameter value (weight) $\hat{w}_1 = 0.6222$ which attain the maximum likelihood $\log\Lambda(\hat{\boldsymbol{w}}) = 14833.5$ in (20). Under the optimal weight $\hat{w}_1$ we have the MAP estimate for the intensity function $\lambda_{\hat{\boldsymbol{\theta}}}(x,y)$ shown in Figure 3.

*** Figure 3 around here ***

This MAP estimate appears consistent with the real seismic activity in and around Japan. In particular, the changes in the eastern off-shores of the northern Japan appear very large, taking account of the contours drawn in logarithmic scale. Furthermore, the parameterization using Delaunay tessellation does appear very suitable for the observations on highly non-homogeneous or clustered point pattern. That is to say, we can see detailed estimate of changes where the observations are densely populated, while smoother changes are expected in the region of sparsely populated. This is somewhat similar idea to the kernel estimation with variable bandwidths (e.g., Choi and Hall, 1999; Zhuang et al., 2000).

4.5 Application of the HIST-ETAS model

First of all, we apply the space-time ETAS model described in (11) to the data that have been rearranged as described in Section 4.1 to take account of the significant anisotropic features of each event, if any. Then, we obtained the MLEs $\hat{\mu}$, $\hat{K}$, $\hat{c}$, $\hat{\alpha}$, $\hat{p}$, $\hat{d}$ and $\hat{q}$ which are listed in the first row of Table 1. The technical details for the likelihood computation is described in Ogata (1998).

*** Table 1 around here ***

However, it is shown in Ogata (1998) that those MLEs are biased (for example, $\hat{p}$ takes lower value than 1.0) owing to the restriction that the constant background rate μ is assumed throughout the Japan area in the model (11). Therefore, we consider a piecewise-linear function on the Delaunay tessellation (Figure 2b) for the nonhomogeneous background rate $\mu(x,y)$ by which we replace the constant μ in the conditional intensity (11). Fix the other parameters of the space-time ETAS model in (11) to the same MLE values in the first row of Table 1. Then, we get the optimal MAP estimate $\hat{\mu}(x,y)$ by the Bayesian procedure using the log-likelihood function in (2) and the restricted penalty in (14) assuming $w_2 = \cdots = w_5 = 0$. Next, replace μ in (12) again by $\mu(x,y) = \nu\cdot\hat{\mu}(x,y)$ for the fixed MAP solution $\hat{\mu}(x,y)$ in order to get the new MLEs $\hat{\nu}$, $\hat{K}$, $\hat{c}$, $\hat{\alpha}$, $\hat{p}$, $\hat{d}$ and $\hat{q}$, which are listed in the second row of Table 1. Furthermore, we fix $\hat{\nu}$, $\hat{K}$, $\hat{c}$, $\hat{\alpha}$, $\hat{p}$, $\hat{d}$ and $\hat{q}$ to get the MAP estimate $\hat{\mu}(x,y)$ again; then for the fixed $\hat{\mu}(x,y)$ we get the corresponding MLEs

$\hat{\nu}$, $\hat{K}$, $\hat{c}$, $\hat{\alpha}$, $\hat{p}$, $\hat{d}$ and $\hat{q}$, which are listed in the third row of Table 1. We repeat this procedure up until the MLEs converge.

Setting the MLEs, $\hat{\nu}$, $\hat{K}$, $\hat{\alpha}$, $\hat{p}$ and $\hat{q}$ in the last row of Table 1 as the corresponding reference values in (12), we prepare five sets of the piecewise linear functions $\{\varphi_k;\ k = 1, \cdots, 5\}$ for estimating the location dependency of the parameters μ, K_0, α, p and q, while the parameters c and d are always fixed to be the last MLEs $\hat{c}$ and $\hat{d}$ in Table 1. To define those piecewise linear functions on the Delaunay tessellated area, we have about five times as many coefficients $\{\phi_{k,i}\}$ to be estimated as the number of data events, namely, about 23,000 unknown coefficients. Furthermore, we have five hyper-parameters (weights) to be tuned simultaneously by the Bayesian procedure described in Section 4.3.

We start with setting $\phi_{k,i} = \varphi_k(x_i, y_i) = 0$ in (12) for all $k = 1, \cdots, 5$ and all $i = 1, 2, \cdots, N + n$ as the initial coefficients values. By the Bayesian procedure we can objectively obtain the optimal weights simultaneously together with the corresponding MAP estimate of the coefficients of the piecewise linear functions. The attained values of the hyper-parameters $\hat{\boldsymbol{w}} = (\hat{w}_1, \cdots, \hat{w}_5)$ in (20) for the roughness penalty is given in Table 2.

Incidentally, it is confirmed that the all the MLEs in Table 1 except for the MLE in the first row led to the very similar optimal MAP solution of the HIST-ETAS model.

*** Table 2 around here ***

The parameter functions with the optimal maximum posterior (MAP) solutions are shown in Figures 4a-e. Figure 4a shows the estimated spatial background rate $\hat{\mu}(x, y)$. Taking account of the contours being drawn in logarithmic scale, the seismicity changes in the eastern off-shores of Japan appear very large. Here, in comparison with Figure 3a, we can see that clustering effects (aftershocks) have been well removed from the total seismic activity, particularly in the regions where the seismic activity had been low before the large earthquakes took place. Figure 4b shows the regional change of $\hat{K}(x, y)$ which shows not only the average number density of the aftershocks per unit area relative to the magnitude of the mainshock but also significantly distorted spatial distribution of aftershocks relative to the assumed elliptical distribution in the model. In the active swarm areas such as Matsushiro and Izu Islands, this has high values. The low values of α should have expected similarly in the swarm area, but we cannot see it in Figure 4c probably owing to the high threshold magnitude (M5.0) of the present data. The p-value distribution in Figure 4d varies in the reasonable range between 0.98 and 1.33 in which p-values of the most individual aftershock studies in and around Japan (e.g., Utsu, 1969; Guo and Ogata, 1996) are included. Finally, the change of the q-value is very small in the value, but slight systematic decrease in Figure 4e from the west to the east is seen.

*** Figures 4a-e around here ***

5 Space-time residual analysis

5.1 Method

Let $\hat{\lambda}(t, x, y|\ H_t)$ be the conditional intensity function of the HIST-ETAS model with the MAP estimates obtained in Section 4.5. In order to detect the temporal deviations of the seismicity from the one predicted by the conditional intensity, we consider a flexibly parameterized indicator function $\xi(t, x, y; \boldsymbol{\theta})$ that composes a new intensity function such that

$$\eta_{\boldsymbol{\theta}}(t, x, y) = \hat{\lambda}(t, x, y|\ H_t)\ e^{\xi(t,x,y;\boldsymbol{\theta})} \qquad (21)$$

to apply the same the data again. The estimated HIST-ETAS model $\hat{\lambda}(t, x, y|\ H_t)$ is shown to be a good-fit in the space-time volume where $\xi(t, x, y; \boldsymbol{\theta}) \approx 0$ hold. However, we are particularly interested in the significant volume where $\xi(t, x, y; \boldsymbol{\theta})$ takes negative values, which means that the volume is systematically quiet relative to the estimated intensity.

In order to estimate $\xi(t, x, y; \boldsymbol{\theta})$, the whole three-dimensional volume $[0, T] \times A$ is divided into the Delaunay tetrahedra (e.g., Tanemura et al., 1983) whose vertices consist of the hypocenter coordinates of earthquake data set $\{(t_i, x_i, y_i); i = 1, \cdots, N\}$ and some (say, n) additionally placed points on the boundary surface, edges and vertices of the whole space-time volume. Those are all associated with unknown parameters $\{\xi_i; i = 1, \cdots, N+n\}$ to be estimated. Then, three-dimensional piecewise linear function $\xi(t, x, y; \boldsymbol{\theta})$ is defined on the tessellated volume. Specifically, for any location (t, x, y) in the space-time volume, find the tetrahedron which includes (t, x, y). Let these vertices be $\{(t_{i_k}, x_{i_k}, y_{i_k});\ k = 1, 2, 3, 4\}$. Then we have

$$\xi(t, x, y) = \hat{a}_1\xi_{i_1} + \hat{a}_2\xi_{i_2} + \hat{a}_3\xi_{i_3} + \hat{a}_4\xi_{i_4}, \qquad (22)$$

where the values $\{\hat{a}_k,\ k = 1, 2, 3, 4\}$ are non-negative solution of the linear equation

$$\begin{aligned} a_1 t_{i_1} + a_2 t_{i_2} + a_3 t_{i_3} + a_4 t_{i_4} &= t \\ a_1 x_{i_1} + a_2 x_{i_2} + a_3 x_{i_3} + a_4 x_{i_4} &= x \\ a_1 y_{i_1} + a_2 y_{i_2} + a_3 y_{i_3} + a_4 y_{i_4} &= y \\ a_1 + a_2 + a_3 + a_4 &= 1. \end{aligned}$$

If the conditional intensity $\lambda_{\boldsymbol{\theta}}$ in the log likelihood (2) is replaced by the intensity $\eta_{\boldsymbol{\theta}}$ in (21), we can define the log likelihood function of the indicator function $\xi(t, x, y; \boldsymbol{\theta})$. On the other hand, for the smoothness constraint of parameters $\boldsymbol{\theta}$, we consider the penalty

$$Q(\boldsymbol{\theta}) = \int_0^T \int\int_A \left\{ w_1 \xi_t^2 + w_2 \left(\xi_x^2 + \xi_y^2 \right) \right\} dt dx dy, \quad (23)$$

where ξ_t, ξ_x and ξ_y are partial derivative of the function $\xi(t, x, y; \boldsymbol{\theta})$ with respect to the variables t, x and y, respectively. Since the above penalty function is quadratic with respect to the parameters, the prior is multivariate Normal distribution. Therefore, the Normal approximation of the posterior function is useful in order to find the optimal hyper-parameters $\boldsymbol{w} = (\hat{w}_1, \hat{w}_2)$ which maximize the log likelihood in (20). Simultaneously, maximizing the penalized log-likelihood in (13) with the penalty function in (23) for the fixed $\hat{w}_1$ and $\hat{w}_2$, we can also get the MAP estimates $\{\hat{\xi}_i;\ i = 1, \cdots, N + n\}$ attached to the coordinates of vertices $\{(t_i, x_i, y_i);\ i = 1, \cdots, N + n\}$ of the Delaunay tetrahedron. Thus, the image $\{\xi_{ijk}\}$ on the three dimensional lattice (t_i, x_j, y_k) can be calculated by the linear interpolation in (22).

5.2 Application

Based on the estimated intensity function of the HIST-ETAS model in Section 4.5, we want to detect significant temporal changes of the seismicity relative to the one predicted by the estimated model. Thus, we have a space-time conditional intensity function in (21) to fit the same data again to obtain the estimate of the indicator function $\xi_{\boldsymbol{\theta}}$. By the Bayesian optimization procedure, we have the maximizing hyper-parameter (weights) $\hat{w}_1 = 0.0164$ and $\hat{w}_2 = 0.0814$ for the penalty in (23) with $\log \Lambda(\hat{\boldsymbol{w}}) = -2265.37$.

The MAP estimate $\{\hat{\xi}(t, x, y)\}$ provides the residual image $\{\hat{\xi}_{ijk}\}$, which is explored by means of the Application Visualization System (AVS, Stardent Computer Inc.). One of its tools is to show cross-sectional images. An example is shown by Figures 5 and 6. This section cut through one of the most active area, off the east coasts of Japan (141°-145°E, 38°-42°N), where we have dense enough earthquake occurrences for a high precision of the image to see the significant changes. Figure 6 shows such cross-sectional image of latitude versus time associated with the earthquake locations within the volume. The yellow color of the image stands for the $\hat{\xi}$-value around zero, and the warmer and colder colors stand for the $\hat{\xi}$ values larger and smaller than zero, respectively.

*** Figures 5 and 6 around here ***

First of all, general impression of the residual image is gradually tending from warm to colder colors in time, which means that the trend of the seismic activity appears decreasing. In fact, as shown in Figure 7, the trend of earthquake activity ($M \geq 5.0$) in whole Japanese region is decreasing for the later period of 50 years except for the occasional jumps owing to the aftershocks especially that of the 1968 Tokachi-Oki great event of M_J7.9. Incidentally, Ogata and Abe (1989) discuss that the similar trend is seen in the long-term global seismic activity. Now, let us concentrate on the changes in shorter frequencies. Then, we see a number of hollows with colder colors before the large earthquakes of M_J7.5 or larger, about the source (aftershock) region, which show the lowering of the seismicity rates (*relative quiescence*) compared to those predicted by the estimated HIST-ETAS model whose coefficients are shown in Figure 4a-e.

*** Figure 7 around here ***

Although such lowering of the residual image was not always followed by a large event, the present result shown in Figure 6, for example, is encouraging in that we may hope to forecast the time and region of forthcoming great earthquake using abundant space-time data of the smaller earthquakes.

6 Conclusions

In this paper we have reviewed the derivation of the suitable form of the space-time ETAS model in (1) with (10), or in (11), based on the empirical laws in seismology and also on the model selection procedure. Then the hierarchical space-time ETAS (HIST-ETAS) model are considered. Namely, the function for each parameter is expanded by the two dimensional piecewise linear function consisting of facets defined on the Delaunay tessellated triangles whose vertices are locations of the earthquakes in the data and the additional points in the boundary of the whole volume. It is notable that the present function form based on the Delaunay tessellations is very suitable for highly nonhomgeneous or clustered sampling data sets.

Then the penalized log likelihood in (13) is considered for the trade-off between the good fit to the data and the uniformity of each function. The Bayesian method is applied, for solving the trade-off objectiviely, to the long-term earthquake occurrence data in and around Japan. Thus, we have the MAP spatial images of the parameter changes of the HIST-ETAS model which show regional characteristics of seismicity in and around Japan.

For the residual analysis, we fit to the same data again a new conditional intensity function in (21) that is the multiplication of the estimated intensity function of the HIST-ETAS and the indicator function $\xi(t,x,y)$ expanded by the space-time piecewise linear function defined on the Delaunay tessellated tetrahedra whose vertices are locations and times of the earthquakes in the data and the additional points on the boundaries of the whole space-time volume. The estimation is objectively carried out by means of the Bayesian method for the similar trade-off with the roughness penalty in (23).

The estimated three dimensional MAP image $\{\xi_{ijk}\}$ shows not only the decrease trend of seismicity in time on the whole but also the relatively quiet volume before and around the source of the forthcoming great events in the area of active region of plate boundary where enough number of events are available to reveal such a local detail.

The present experiment also demonstrated that parameterization of the indicator function $\xi(t,x,y)$ through Delaunay tessellation has a great advantage for smoothing functions on the three or higher dimensional space.

Finally, I would conclude that the HIST-ETAS model and the proposed residual analysis are useful for measuring characteristic of seismic activity and also for the space-time forecasting of the large earthquakes.

References

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

Choi, E. and Hall, P. (1999), "Nonparametric approach to analysis of space-time data on earthquake occurrences," *J. Comput. Graphic. Statist.* **8**, 733-748.

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.

Good, I.J. and Gaskins, R.A. (1971), "Nonparametric roughness penalties for probability densities," *Biometrika* **58**, 255-277.

Good, I.J. (1965), *The Estimation of Probabilities*, M.I.T. Press, Cambridge, Massachusetts.

Green, P.J. and Sibson, R. (1978), "Computing Dirichlet tessellation in the plane," *Comput. J.* **21**, pp.168-173.

Guo, Z. and Ogata Y. (1996), "Statistical relations between the parameters of aftershocks in time, space and magnitude," *J. Geophys. Res.* **102**, 2857-2873.

Gutenberg, R. and Richter, C.F. (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Amer.*, **34**, pp. 185-188.

Hawkes, A.G. (1971), "Point spectra of some mutually exciting point processes," *J. R. Statist. Soc.*, **B 33**, pp. 438-443.

Inouye, W. (1965), "On the seismicity in the epicentral region and its neighborhood before the Niigata earthquake," *Kenshin-jiho (Quarterly J. Seismol.)* **29**, 139-144 (in Japanese).

Kagan, Y.Y. (1991), "Likelihood analysis of earthquake catalogues," *J. geophys. Res.*, **106**, 135-148.

Kagan, Y.Y. and Jackson, D. (1991), "Seismic gap hypothesis: Ten years after," *J. Geophys. Res.* **96**, 21419-21431.

Kagan, Y.Y. and Knopoff, L. (1978), "Statistical study of the occurrence of shallow earthquake," *Geophys. J. R. Astron. Soc.*, **55**, 67-86.

Kagan, Y.Y. and Knopoff, L. (1980), "Spatial distribution of earthquakes: The two-point correlation function," *Geophys. J. R. Astron. Soc.*, **62**, 303-320.

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

McCann, W.R., Nishenko, L.R., Sykes, L.R., and Krause, J. (1979), "Seismic gaps and plate tectonics: Seismic potential for major boundaries," *Pure Appl. Geophys. (PAGEOPH)* **117**, 1082-1147.

Mogi, K. (1968), "Some features of recent seismic activity in and near Japan (1)," *Bull. Earthq. Res. Inst., Univ. Tokyo* **46**, 1225-1235.

Mori, M. (1986), *FORTRAN 77 Numerical Analysis Programming*, 342pp., Iwanami Publisher, Tokyo, in Japanese.

Murata, Y. (1992), "Estimation of optimum surface density distribution only from gravitational data: an objective Bayesian approach," *J. Geophys. Res.*, **98**, 12097-12109.

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' by Y.Y. Kagan and D.D. Jackson," *J. Geophys. Res.* **98**, 9909-9916.

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *J. Amer. Statist. Assoc.*, **83**, No. 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," *J. Geophys. Res.*, **97**, 19845-19871.

Ogata, Y. (1993), "Space-time modeling of earthquake occurrences," *Bull. Int. Statist. Inst.*, **55**, Book 2, 249-250.

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

Ogata, Y. (2001), "Increased probability of large earthquakes near aftershock regions with relative quiescence," to appear in *J. Geophys. Res.*

Ogata, Y. and Abe, K. (1989), "Some statistical features of the long term variation of the global and regional seismic activity. *Int. Statist. Rev.*," **59**, 139-161.

Ogata, Y., Utsu, T. and Katsura, K. (1995), "Statistical features of foreshocks in comparison with other earthquake clusters," *Geophys. J. Int.* **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 Apple. Geophys.*, **115**, 375-385.

Rathbun S.L. (1993), "Modeling marked spatio-temporal point patterns," *Bull. Int. Statist. Inst.*, **55**, Book 2, pp. 379-396.

Rathbun S.L. (1994), "Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes," in special issue on Spatial Statistic of *J. Statist. Plann. Inf.* **51**, 55-74.

Tanemura, M., Ogawa, T. and Ogita, N. (1983), "A new algorithm for three-dimensional Voronoi tessellation," *J. Comput. Phys.* **51**, pp. 191-207.

Utsu, T. (1961), "Statistical study on the occurrence of aftershocks," *Geophys. Mag.*, **30**, 521-605.

Utsu, T. (1968), "Seismic activity in Hokkaido and its vicinity, *Geophys. Bull. Hokkaido Univ.,* **13**, 99-103 (in Japanese).

Utsu, T. (1969). "Aftershocks and earthquake statistics (I): some parameters which characterize an aftershock sequence and their interaction," *J. Faculty Sci., Hokkaido Univ., Ser. VII (geophysics)* **3**, 129-195.

Utsu, T., and A. Seki (1955), "Relation between the area of aftershock region and the energy of the main shock," *Zisin (J. Seismol. Soc. Japan), 2nd Ser., ii*, *7*, 233-240 (in Japanese with English summary).

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2000), "A model based declustering method for estimating the spatial background seismic activity," *Research Memorandum*, No. 769, The Institute of Statistical Mathematics, Tokyo, 45pp.

Figure captions

Figure 1. Space-time configuration of earthquakes of magnitude 5.0 and larger in and around Japan for the period 1926-1995. The vertical axis stands for time which runs upwards from the bottom.

Figure 2. Earthquakes of magnitude 5.0 and larger in and around Japan for the period 1926-1995. (a) Epicenter locations and (b) Delaunay tessellation connecting the epicenters of the earthquakes of the data.

Figure 3. The MAP estimate of the spatial intensity fitted to the events shown in Figure 2a. The contours are equi-spaced in logarithmic scale, ranging from $1.24\times10^{-6}\sim0.0612/\mathrm{deg}^2/\mathrm{day}$.

Figure 4. The MAP estimates of the hierarchical space-time ETAS model. The contours are equi-spaced in the respective scale below: (a) μ-values, ranging $0.925\times10^{-6}\sim0.0492/\mathrm{deg}^2/\mathrm{day}$ in logarithmic scale; (b) K_0-value, ranging 0.306×10^{-5} $\sim0.134\times10^{-3}/\mathrm{deg}^2/\mathrm{day}$ in linear scale; (c) α-value, ranging $1.46\sim1.76$ in linear scale; (d) p-value, ranging $0.982\sim1.33$ in linear scale; and (e) q-value, ranging $1.86\sim1.91$ in linear scale.

Figure 5. The different perspective of the data from Figure 1 with a plane for the 2-dimension cross-sectional image of the 3-dimensional image $\hat{\xi}(t,x,y)$ at the longitude of 143°E which is shown in Figure 6. The time axis runs from the front to back.

Figure 6. Latitude versus time cross-sectional image of the estimated $\hat{\xi}(t,x,y)$ on the plane shown in Figure 5. Black dots show the earthquakes of M5.0 and larger within a zone of one degree distances from the cross sectional plane.

Figure 7. Earthquakes of magnitude 5.0 and larger in and around Japan for the period 1926-1995. (a) Epicenter location and (b) cumulative number versus time. The dotted line shows the extrapolation of expected seismicity based on the rate of the early one.

See tne next page.

TABLE 1. The MLE of the space-time ETAS model

Step (unit)	AIC	$\hat{\nu}$ ($\hat{\mu}$ for Step 0) (event/deg^2/day)	$\hat{K}$ (event/deg^2/day)	$\hat{c}$ (days)	$\hat{\alpha}$ (1/mag)	$\hat{p}$	$\hat{d}$ (degrees)	$\hat{q}$
0	41722.8	1.92×10^{-4}	7.60×10^{-4}	0.0134	1.42	0.99	0.200	2.84
1	41700.9	2.79×10^{-4}	3.86×10^{-5}	0.0320	1.63	1.11	0.00191	1.74
2	41705.8	2.90×10^{-4}	2.05×10^{-5}	0.0370	1.64	1.14	0.00214	1.88
3	41704.1	2.91×10^{-4}	1.99×10^{-5}	0.0376	1.65	1.14	0.00214	1.88
4	41704.0	2.91×10^{-4}	1.99×10^{-5}	0.0376	1.65	1.14	0.00214	1.88

TABLE 2. The estimated hyperparameters (weights) of the HIST-ETAS model

$\log\Lambda(\hat{\boldsymbol{w}})$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_5$
−22507.75	0.4803	0.4774	137.7	55.79	6112.