Ann Inst Stat Math (2010) 62:127–143
DOI 10.1007/s10463-009-0268-7

Hybrid kernel estimates of space–time earthquake occurrence rates using the epidemic-type aftershock sequence model

Giada Adelfio · Yosihiko Ogata

Received: 23 July 2008 / Revised: 15 January 2009 / Published online: 2 December 2009

Abstract The following steps are suggested for smoothing the occurrence patterns in a clustered space–time process, in particular the data from an earthquake catalogue. First, the original data is fitted by a temporal version of the ETAS model, and the occurrence times are transformed by using the cumulative form of the fitted ETAS model. Then the transformed data (transformed times and original locations) is smoothed by a space–time kernel with bandwidth obtained by optimizing a naive likelihood cross-validation. Finally, the estimated intensity for the original data is obtained by back-transforming the estimated intensity for the transformed data. This technique is used to estimate the intensity for earthquake occurrence data for associated with complex sequences of events off the East Coast of Tohoku district, northern Japan. The intensity so obtained is compared to the conditional intensity estimated from a full space–time ETAS model for the same data.

Keywords Bandwidths · Shape parameters · Cross-validation · ETAS models · Intensity function · Kernel estimates · Space–time point processes · Space–time ETAS model · Transformation of time

1 Objective

One of the important functions of the statistical analysis of space–time point process data is to provide a convenient visualsummary of the original data. In this paper we

G. Adelfio
Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", University of Palermo, Palermo, Italy

Y. Ogata (✉)
The Institute of Statistical Mathematics, Tokyo, Japan
e-mail: ogata@ism.ac.jp

Y. Ogata
Department of Statistical Science, The Graduate University of Advanced Studies, Tokyo, Japan

approach this problem by first assuming, if only for this purpose, that the data can be treated as a non-homogeneous Poisson process, characterized by its space–time intensity function. Direct kernel smoothing can then be used to estimate the intensity of this presumed Poisson process, and has the advantages of simplicity in computation and adaptability to the data. However, in fact, it is difficult for such intensity rate changes estimated by kernels to express very asymmetric, sharp or highly variable features as seen in Choi and Hall, (1999) and Grillenzoni (2005). These features mainly arise from the fact that earthquake occurrences are history dependent, as for example when a large earthquake occurs that is followed by many aftershocks.

In this study, we consider a particular method of space–time intensity estimation where the time axis is transformed by using the temporal ETAS model (Ogata 1988). The transformed data is assumed to form an inhomogeneous space–time Poisson process, and a kernel method is applied to smooth this data. In order to get an optimal estmate of the intensity function, we maximize a cross-validated likelihood score. After back-transforming, the final intensity is compared with that estimated from the parametric space–time ETAS model (Ogata 1998).

The reason for the time transformation is to simplify the time inhomogeneities which can be well described by the epidemic type aftershocks-sequences (ETAS) model (Ogata 1988). Time component of this model, integrating out the spatial variation, uses the well established and robust Omori–Utsu law (Utsu 1961, 1970; Utsu et al. 1995). On the other hand, spatial distributions of aftershocks are too complex (Ogata et al. 2003; Ogata 2004) to be described by a standard function from a parametric model (Ogata 1998; Ogata and Zhuang 2006); hence they are handled by a form of non-parametric estimate.

In the following, a brief introduction of space–time point processes describing earthquake occurrences is provided in Sect. 2, and the dataset of complex seismic activity in North-Eastern Japan is described in Sect. 3. Then, as a particular class of space–time processes for earthquake occurrences, the ETAS model (Ogata 1988) is reviewed in Sect. 4. The hybrid kernel estimation method, together with the computational procedure developed to estimate the intensity function of Poisson point processes is introduced and applied to the dataset in Sect. 5. Then, in Sect. 6, a full space–time ETAS model (Ogata 1998) is applied to the dataset to compare with the hybrid kernel estimates in Sect. 7; specifically, from the data, we take epicenters of three great earthquakes as the fixed locations and the two days after the occurrence of each earthquake as the time instants, and we show and discuss the results of the two approaches from these three episodes. The conclusion is provided in Sect. 8.

2 Point processes for space–time earthquake occurrences

A point process is defined as a collection of randomly located points in some space. In this study, we are concerned with earthquake occurrences, where each point represents the time of occurrence and the location of epicenter of a certain earthquake of magnitude greater than a certain value, in a geophysical region.

To estimate the probability of such a future earthquake occurrence, the conditional intensity function is crucial, and is proportional to the probability that such

an event above a certain threshold magnitude M_{th} will occur at time t at a location (x, y), conditional on the history of the occurrence times t_i, locations (x_i, y_i), and magnitudes M_i of past earthquakes up to time t. Let us denote such a history by $\mathcal{H}_t = \{(t_i, x_i, y_i, M_i);\ t_i < t, M_i > M_{th}\}$ for simplicity. Then, its conditional intensity function is defined by

$$\begin{aligned}&\lambda(t, x, y|\mathcal{H}_t)\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\\&\quad = Pr\{\text{an event in volume } [t, t+\mathrm{d}t) \times [x, x+\mathrm{d}x) \times [y, y+\mathrm{d}y)|\mathcal{H}_t\} \quad (1)\end{aligned}$$

Assuming such a limit exists for each point (t, x, y) in the space–time domain, the conditional intensity process uniquely characterizes the finite-dimensional distributions of the point process (Daley and Vere-Jones 2003). If the conditional intensity function is independent of the past history and dependent only on the current time and spatial location, Eq. (1) determines $\lambda(t, x, y)$ and identifies an inhomogeneous Poisson process. Furthermore, by normalizing the intensity function by the defined space–time rectangular volume $T \times X \times Y$,

$$f(t, x, y) = \lambda(t, x, y) \Big/ \int_T \int_X \int_Y \lambda(t, x, y)\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y, \quad (2)$$

becomes a probability density distribution on $T \times X \times Y$.

3 Data description

We use the hypocenter data compiled by the Japan Meteorological Agency, and consider one of several datasets discussed in Ogata (1998) and Ogata and Zhuang (2006). These data cover the area off the east coast of Tohoku District, one of the most active plate boundary zones in and around Japan; see Fig. 1a. The data consist of earthquakes of magnitude (M) 4.5 and larger that are chosen from the wide region 36° ∼ 42°N and 141° ∼ 145°E for depths down to 100 km and for the time span 1926–1995; see Fig. 1c. In this region, most of large earthquakes occurred on the plate boundary along the subducting Pacific plate beneath northern Japan. We ignore the depth axis due to inaccurate solutions, at offshore in particular, and therefore consider only two-dimensional locations (longitude and latitude) of earthquakes, restricting ourselves to shallow events. In fact, most of the events in this region are distributed within depths down to 60 km.

In calculating the distance between earthquake epicenters, the distance in longitude component is reduced to $\cos(\pi y_0/180°)$ times the distance in latitude (one degree in latitude corresponds to about 111.11 km), where $y_0 = 39°$N is the latitude of the center of the region.

4 Temporal ETAS model and transformation of time

The ETAS model is a temporal self-exciting point process, representing earthquake occurrences in a region during a period (Ogata 1988). This model is a modification

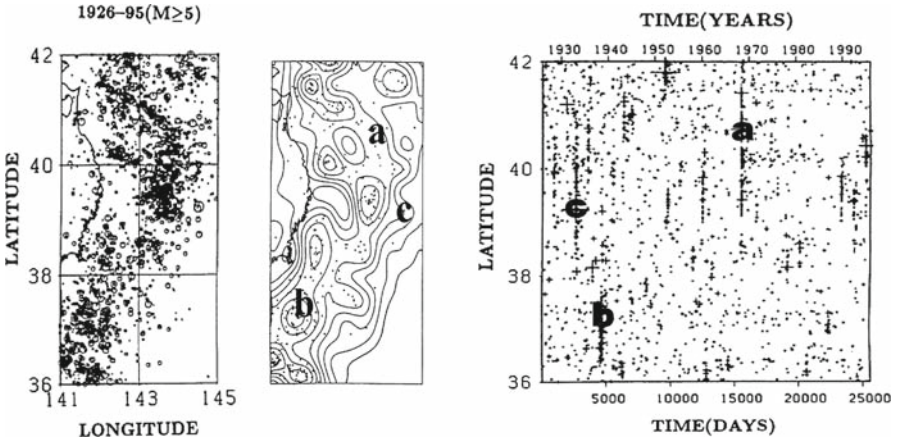

Fig. 1 Earthquakes with $M \geq 4.5$ off the coast of Tohoku District, northern Japan, for the period 1926–1995; *left* epicenters, *middle* the background seismicity rate $\nu(x, y)$ in Eq. (13) with logarithmically equidistant contours, and *right* latitude versus time plot. Sizes of *circles* and *plus signs* correspond to magnitudes of earthquakes, and the bold letters a, b, and c indicate hypocenters of the 1968, 1938, and 1933 great earthquakes, respectively, in this region

of Hawkes' model (Hawkes 1971; Hawkes and Adamopoulos 1973), customized for standard seismic activity based on a series of aftershock studies (Utsu 1961, 1970; Utsu et al. 1995). For a series of earthquakes of magnitude M_{th} and larger, this model is defined by

$$\begin{aligned}\lambda(t|\mathcal{H}_t)\mathrm{d}t &= Pr\{\text{an event in volume } [t, t+\mathrm{d}t)|\mathcal{H}_t\} \\ &= \left\{\mu + \sum_{\{j;\ t_j<t\}} K_0 e^{\alpha(M_j - M_{th})}(t - t_j + c)^{-p}\right\}\mathrm{d}t \end{aligned} \tag{3}$$

where $\mathcal{H}_t = \{(t_i, M_i);\ t_i < t, M_i > M_{th}\}$ is the history of occurrence times associated with magnitudes of past earthquakes.

Given a series of events associated with magnitudes $\{(t_i, M_i),\ 0 \leq t_i \leq T,\ i = 1, 2, \ldots, n\}$, the parameters $\vartheta = (\mu, K_0, c, \alpha, p)$ are estimated by maximizing the log-likelihood function of point processes (Daley and Vere-Jones 2003)

$$\ln L(\vartheta) = \sum_{i=1}^{n} \ln \lambda_\vartheta(t_i|\mathcal{H}_t) - \int_0^T \lambda_\vartheta(t|\mathcal{H}_t)\mathrm{d}t.$$

The maximum likelihood estimates of the ETAS model for the dataset are listed in Table 1.

Here, we can see how well or poorly the estimated model fits an earthquake sequence by comparing the cumulative number of earthquakes with the rate predicted by the

Table 1 Estimated parameters of the temporal ETAS model.

Model	μ	K	c	α	p
(3)	0.48032	0.014816	0.029113	1.5500	1.0362

Here, the both dimension unit of μ and K is events/day, and the units of c and α are day and magnitude^{-1}, respectively, while the coefficient p is dimensionless

estimated model. Suppose that a series of events is generated based on the ETAS model. Then the integral

$$\Lambda(t|\mathcal{H}_t) = \int_0^t \lambda(s|\mathcal{H}_s)\mathrm{d}t \tag{4}$$

defines the theoretical cumulative number of events over the time interval $[0, t]$.

The time transformation $\tau_i = \Lambda(t_i|\mathcal{H}_{t_i})$ from t to τ, then takes the original occurrence time sequence $\{t_1, \ldots, t_n\}$ to the sequence $\{\tau_1, \ldots, \tau_n\}$ in the transformed time interval $[0, \Lambda(T)]$. Assume that the seismicity rate $\lambda(t|\mathcal{H}_t)$ estimated from the events in the interval $[0, T]$ is a good approximation to the real seismicity. Then we can expect that the function curve $\Lambda(t)$ of time t and the empirical cumulative function $N(t)$ of the events $\{t_1, \ldots, t_n\}$ overlap each other until time T. This implies that the transformed events $\{\tau_1, \ldots, \tau_n\}$ are uniformly distributed in the corresponding time interval; i.e., the cumulative numbers of events against transformed times, (τ_i, i), are aligned on a straight line with unit slope (Ogata 1988; Ogata et al. 2003).

Consider then the space–time coordinates $\{(\tau_i, x_i, y_i);\ i = 1, 2, \ldots, n\}$ of a set of hypocenter data with the above transformed times $\{\tau_i\}$, where (x_i, y_i) is the epicenter of an aftershock. As a first approximation, we assume that the space–time events with coordinates $\{(\tau_i, x_i, y_i);\ i = 1, 2, \ldots, n\}$ are distributed according to an inhomogeneous space–time Poisson process. Thus, we propose to make an optimal space–time smoothing of the intensity of this set of data by the kernel method. Then the estimated intensity $\lambda_{\hat{\theta}}(\tau, x, y)$ is inversely transformed back to the corresponding $\hat{\lambda}(t, x, y)$ on the space–time zone of the original regular time $t = \Lambda^{-1}(\tau)$.

5 Spatio-temporal kernel intensity estimates and their applications

As presented in Eq. (2), the normalized intensity function is a probability density function assuming the process is Poissonian. Therefore, a conventional nonparametric estimation technique could be used in this work to estimate the density function on the basis of the sample points. To tackle the optimization problem in terms of the least-square method, one must consider that, as the bandwidth tends to zero, the estimator variance decreases but its bias increases. For example, Adelfio et al. (2006) discusses the seismicity off the north coast of Sicilian Island, Italy, in which the bandwidths of the isotropic kernel are optimized by the minimization principle (Silverman 1986) of the mean integrated square error of the estimator $\hat{f}(\cdot)$ around the true density $f(\cdot)$.

For more sensitive estimation, we adopt the cross-validation approach using the log-likelihood score of Poisson process for the space–time kernel estimation. Consider the

space–time coordinates $\{(\tau_i, \xi_i, \eta_i);\ i = 1, 2, \ldots, n\}$, where the transformed times $\{\tau\}$ were found using relation (4) based on the estimated ETAS model, as described in Sect. 3, and (ξ, η) are the epicenter coordinates in azimuth degrees of longitude and latitude, and 1 degree in latitude covers about 111.11 km. For the spatial isotropy treatment, we consider the transformation of scale

$$x = \xi \cos(\pi y_0/180^\circ) \quad \text{and} \quad y = \eta,$$

where y_0 is the central latitude coordinate in the considered seismicity region. Thus, we consider $\{(\tau_i, x_i, y_i);\ i = 1, 2, \ldots, n\}$, where $0 \le \tau \le \Lambda(T)$ and $(x_i, y_i) \in A$. Then, we assume that it is a spatially isotropic but spatio-temporally inhomogeneous Poisson process according to the intensity function $\lambda(\tau, x, y)$. Thus the space–time kernel intensity estimator is defined by the superposition of the separable kernel densities such that

$$\lambda(\tau, x, y) = \sum_{i=1}^{n} f\,(\tau - \tau_i)\ g(x - x_i,\ y - y_i)\,, \tag{5}$$

where $f(\tau)$ and $g(x, y)$ are the one- and two-dimensional probability density functions, respectively. Here, if we assume isotropic kernels in space, the spatial kernel density $g(x, y)$ is equivalent to the function $2\pi r \cdot g(r)$ of distance r from the origin in polar coordinates. Thus, we consider the following candidates for symmetric kernel functions for the temporal and spatial components:

$$f(\tau) = \left(\sqrt{2\pi\sigma}\right)^{-1} \exp\left\{-\tau^2/(2\sigma^2)\right\} \tag{6}$$

$$f(\tau) = (2\,\sigma)^{-1} \exp\{-|\tau|/\sigma\} \tag{7}$$

$$f(\tau) = (\beta - 1)\ \sigma^{\beta-1}/\left\{2(|\tau| + \sigma)^{\beta}\right\} \tag{8}$$

$$g(r) = 1/(2\pi\rho)\ \exp\left\{-r^2/(2\rho^2)\right\} \tag{9}$$

$$g(r) = (\gamma - 1)\ \rho^{\gamma-1}/\left\{\pi(r^2 + \rho)^{\gamma}\right\}, \tag{10}$$

where τ represents the transformed time span. The application of this technique requires the parameter vector ϑ consisting of σ, β, ρ, and γ, which should be optimized to achieve the best compromise, according to certain criteria, between the surface smoothness and the level of detail in the phenomenal representation. The log-likelihood function of the kernel intensity is described as

$$\begin{aligned} \ln L(\vartheta) &= \sum_{i=1}^{n} \ln \lambda_\vartheta(\tau_i, x_i, y_i) - \int_0^T \iint_A \lambda_\vartheta(\tau, x, y)\mathrm{d}t\mathrm{d}x\mathrm{d}y \\ &= \sum_{i=1}^{n} \ln \lambda_\vartheta(\tau_i, x_i, y_i) - n. \end{aligned}$$

It should be noted again that the maximizing this function leads to the scaling parameters converging to zero, which provides degenerate estimates, putting mass only on

observed points. One way of dealing with this issue so as as to retain good predictive properties for the estimated intensity function is to make use of the entropy maximization principle (Akaike et al. 1998). In the present context, this can be done by considering a penalized log-likelihood function, where the penalty takes the form of a smoothness constraint, or roughness penalty, on the logarithm of the intensity (Ogata and Katsura 1988). The optimal weight for the penalty function is objectively determined by the minimization of the Akaike Bayesian Information Criterion (ABIC) (Akaike 1998; Akaike et al. 1998).

Here, however, we can use an alternative approach which takes advantage of the fact that we are treating the transformed data as if it came from a Poisson process. Because of the independence properties of the Poisson process, we can adopt a naive cross-validation procedure, randomly dividing the data into two equal sets $\{(\tau_i^u, x_i^u, y_i^u);\ i = 1, 2, \ldots, n_u\}$ and $\{(\tau_i^v, x_i^v, y_i^v);\ i = 1, 2, \ldots, n_v\}$, with $n = n_u + n_v$, and assuming that both data sets have the same intensity function. We use the first dataset to evaluate the intensity

$$\lambda_\vartheta^u(\tau, x, y) = \sum_{i=1}^{n_u} f\left(\tau - \tau_i^u\right) g\left(x - x_i^u,\ y - y_i^u\right),$$

assuming the parameters ϑ are known, and then use the second data set to evaluate the predictive performance of this intensity, using the log-likelihood

$$\ln L(\vartheta) = \sum_{i=1}^{n_v} \ln \lambda_\vartheta^u\left(\tau_i^v, x_i^v, y_i^v\right) - n_v, \tag{11}$$

This likelihood is then maximized to get robust optimal estimates of the parameters ϑ.

Using the data presented in Sect. 3 and processed in Sect. 4, we applied the above cross-validation procedure. The length of the entire transformed period is $\Lambda(T) = n = 4333$ (events), and the center of latitude coordinates is $y_0 = 39°\text{N}$; the distance in degree of longitude must be reduced by $\cos(\pi y_0/180°) = 0.7777$. Furthermore, we have compared the goodness-of-fit of the space–time kernel models with various combinations of densities in (6)–(10). The estimated parameters of the kernels are listed in Table 2. According to the cross-validated log-likelihood values, the best fit is attained by the model P-P, which takes inverse power distributions in both time and space components.

Then, each obtained intensity estimate $\lambda_{\hat{\vartheta}}(\tau, x, y)$ is transformed back to the space–time intensity

$$\lambda_{\hat{\vartheta}}(t, x, y) \cong \lambda_{\hat{\vartheta}}\left(\Lambda(t), x, y\right) \cdot \lambda(t|H_t), \tag{12}$$

where the estimated parameter values in Table 1 defines the conditional intensity function $\lambda(t|H_t)$ of the temporal ETAS model given in (3) and its integration $\Lambda(t)$ given in (4), respectively.

Table 2 Estimated parameters of each kernel model and the cross-validated log-likelihood (11), the larger of which is the better fit

Model	$f(\tau)$	$g(r)$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\gamma}$	$\ln \hat{L}_{CV}$
G-G	(6)	(9)	54.64	–	0.19864	–	−8318.19
G-P	(6)	(10)	41.83	–	0.02924	1.952	−8161.44
E-G	(7)	(9)	43.69	–	0.1863	–	−8174.80
E-P	(7)	(10)	36.33	–	0.02443	1.935	−8018.36
P-G	(8)	(9)	8.111	1.642	0.1544	–	−7941.93
P-P	(8)	(10)	6.471	1.631	0.01873	1.550	−7811.41

The model P-P is considered to be the best fit

6 The space–time ETAS model

In the next section we compare the previously obtained intensity estimates by the hybrid kernel methods with conditional intensity estimates of a full space–time ETAS model directly estimated from the dataset in Sect. 3. For this purpose, we here review the space–time ETAS model.

For a series of earthquakes of magnitude M_{th} and larger, the space–time ETAS model is defined by the conditional intensity function

$$\lambda(t, x, y|\mathcal{H}_t) = \mu \cdot \nu(x, y) + \sum_{\{j;\, t_j<t\}} \frac{K_0 e^{\alpha(M_j - M_{th})}}{(t - t_j + c)^p} f(x - x_j, y - y_j|M_j) \tag{13}$$

of time t and location (x, y), where $\mathcal{H}_t = \{(t_i, x_i, y_i, M_i);\ t_i < t,\ M_i > M_{th}\}$ is the history of past occurrences, $\nu(x, y)$ is the spatial background intensity rate estimated either from properly declustered earthquakes or from a set of sufficiently large earthquakes, and $f(x - x_j, y - y_j|M_j)$ is the spatial clustering intensity of triggering an earthquake of magnitude M_j centered at location (x_j, y_j). Here, Ogata (1998) recommends using form

$$f(x - x_j, y - y_j|M_j) = \left\{ \frac{(x - x_j, y - y_j) S_j (x - x_j, y - y_j)'}{e^{\alpha(M_j - M_{th})}} + d \right\}^{-q}, \tag{14}$$

where S_j is a two dimensional positive definite symmetric matrix, and the primed form $(x, y)'$ indicates the transpose of the vector (x, y). The quadratic form within the brackets indicates that the aftershocks are spatially distributed with ellipsoidal contours. In fact, Utsu and Seki (1955) and Utsu (1970) used an ellipsoid instead of a rectangle to measure the area of aftershocks. Such an ellipsoid reflects the simplest shape of the approximated rupture fault with its dip angle, including location errors of aftershock hypocenters. The logarithm of the aftershock area is highly correlated with the main shock's magnitude, leading to the famous Utsu-Seki law, which is represented by the denominator and numerator in model (14) for near-field scaling. The inverse power decay of the aftershock distribution against the distance takes remote (far-field) triggering phenomena into consideration.

Table 3 Estimated parameters of the ETAS and space–time ETAS models. Here, the dimension unit of μ and K is events/day/degree^2, where "degree" is the global distance in latitude (i.e., 111.11 km), and the unit of d is "degree"

Model	μ	K	c	α	p	d	q
(13)	0.1312E-3	0.38156E-4	0.023082	1.6124	1.0426	0.0010128	1.59997

The units of c and α are day and magnitude^{-1}, respectively. The coefficients p and q are dimensionless

Furthermore, it is often the case that the epicenter of a mainshock is located at the margin of its aftershock area because the epicenter corresponds to the location where the earthquake rupture initiates. For such a mainshock, the epicenter location in the catalog is not suitable for the model in (14). Therefore, we replace the mainshock's epicenter coordinate with the centroid of aftershocks (the mean of aftershocks' coordinates) for model (14). Such a centroid of aftershocks could be closely related to the centroid of the ruptured fault determined in the Harvard CMT catalog of Dziewonski et al. (1981). In addition, spatial distributions of aftershock epicenters are not usually isotropic owing to the aforementioned reasons. Ogata (1998) and Ogata and Zhuang (2006) summarize the compiling procedure for the centroid of clusters and also the matrices representing the ellipsoid of anisotropic clusters. Such a recompiled dataset is particularly useful in model (14) for a few large earthquakes; on the other hand, for most of the remaining events, the centroid can be taken as the epicenter of the original catalog and as the identity matrix.

Given such a recompiled dataset of origin times and spatial coordinates of earthquakes together with their magnitudes and matrices $\{(t_i, x_i, y_i, M_i); M_i \geq M_{th}, i = 1, \ldots, n\}$ during a period $[0, T]$ and in a region A, we can calculate the log-likelihood function of the parameter $\theta = (\mu, K_0, \alpha, c, p, d, q)$ characterizing the space–time point-process model, which is given by

$$\ln L(\theta) = \sum_{i=1}^{n} \ln \lambda_\theta(t_i, x_i, y_i|\mathcal{H}_t) - \int_0^T \int \int_A \lambda_\theta(t, x, y|\mathcal{H}_t) \mathrm{d}t \, \mathrm{d}x \, \mathrm{d}y,$$

Readers are referred to Daley and Vere-Jones (2003, Sect. 7) for the derivation of this formula and Ogata (1998) for the numerical calculations. The maximum likelihood estimate (MLE) $\hat{\theta} = (\hat{\mu}, \hat{K}_0, \hat{\alpha}, \hat{c}, \hat{p}, \hat{d}, \hat{q})$ maximizes the function. The recommended space time model (13) with (14) is selected in Ogata (1998) as the best among the possible models by using Akaike's Information Criterion (AIC; Akaike 1974).

The maximum likelihood estimates of the space–time ETAS model for the dataset are listed in Table 3, where parameter μ in this model (13) is a scalar factor multiplying the baseline intensity function $\nu(x, y)$ which is taken from Ogata (1998) where it is represented as a bi-cubic spline surface. It is shown in Fig. 1b.

7 Comparison between kernel and parametric estimates

We have obtained the $40 \times 60 \times 400$ voxel images of the intensity values for two versions of the kernel method, using the Gaussian kernel model G-G and the inverse

power-law model P-P in Table 2, and for the full space–time ETAS model with parameters shown in Table 3. However, since it is difficult to show such voxel images on paper, we separately show one-dimensional images of the time at fixed locations and spatial images (snapshots) of the intensity surface at fixed instants of time. In the following, we take epicenters of three great earthquakes (cf. a, b and c in Fig. 1) as the fixed locations and the 2 days after the occurrence of each earthquake as the time instants. The results from these three episodes are shown and discussed in Sects. 7.1–7.3.

7.1 The 1968 Tokachi-Oki earthquake of M7.9

First we consider data from around the time and place of the Tokachi-Oki great earthquake of M7.9 that occurred on May 16, 1968, which corresponds to a time $t = 15476.41$ days since the beginning of the catalog (1st January 1926). The epicenter is $(x_c, y_c) = (143.59°\text{E}, 40.73°\text{N})$, indicated by the sign "**a**" in Fig. 1b, at which we examine the intensity rate change with time. Figure 2e shows the conditional intensity

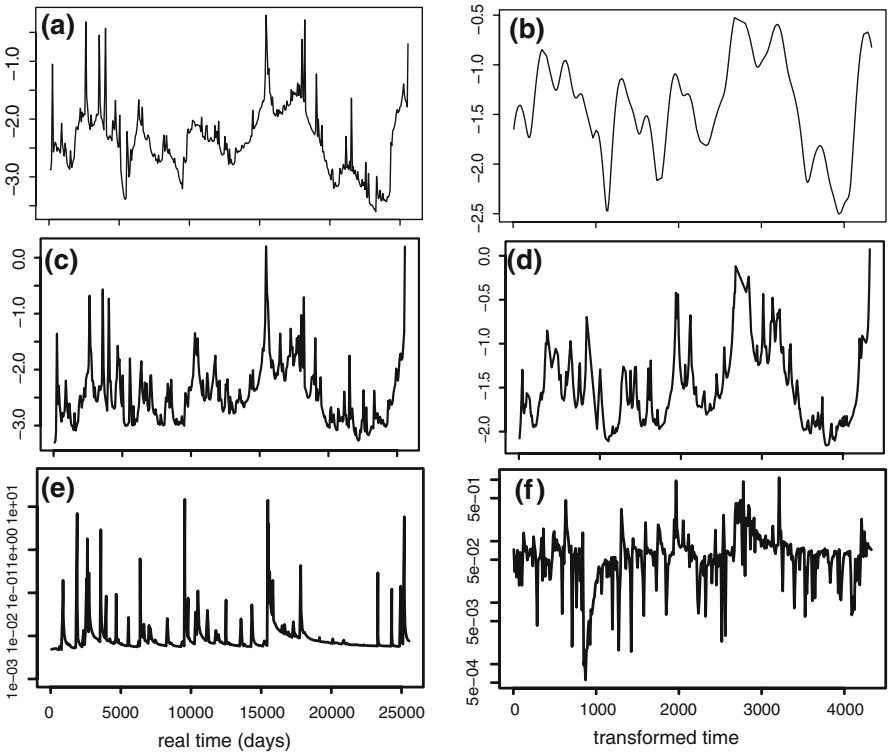

Fig. 2 Log-scaled intensity rate changes in time at the location of the 1968 Tokachi-Oki earthquake epicenter, projected from the space–time Gaussian kernel estimate G-G (panels **a** and **b**), inverse-power kernel estimate P-P (panels **c** and **d**), and space–time ETAS model (panels **e** and **f**), with respect to the transformed time (*right panels*) and the regular time (*left panels*)

of the space–time ETAS model $\hat{\lambda}(t, x_c, y_c|\mathcal{H}_t)$ in (13) with (14). Figure 2f shows the corresponding conditional intensity rates of the transformed time $\tau = \hat{\Lambda}(t)$,

$$\hat{\lambda}(\tau, x_c, y_c|\mathcal{H}_\tau) = \frac{\hat{\lambda}(t, x_c, y_c|\mathcal{H}_t)}{\hat{\lambda}(t|\mathcal{H}_t)}$$

taking the Jacobian coefficient into consideration. Figure 2b and d shows the kernel intensity estimates $\lambda_{\hat{\theta}}(\tau, x_c, y_c)$ of models G-G and P-P, respectively, in (5) at the same fixed location. Finally, Fig. 2a and c is the corresponding kernel intensity estimate $\lambda_{\hat{\theta}}(t, x_c, y_c)$ of the model G-G and P-P, respectively, where the time is transformed back to the ordinary time t by means of relation (12).

Figure 3a, b, and c shows the sliced images of the spatial intensity rates of the kernel intensity $\lambda_{\hat{\theta}}(t_a, x, y)$ of models G-G, P-P, and the space–time ETAS model $\hat{\lambda}(t_a, x, y|\mathcal{H}_\tau)$, respectively, sliced at the fixed time $t_a = 15500.87$ days (June 9, 1968) since January 1, 1926, which is 24.5 days after the mainshock. The plotted epicenters of earthquakes in the figures are those that occurred from January 1, 1968 ($t = 15338.48$ days) through October 15, 1968 ($t = 15627.76$ days), which roughly corresponds to cluster "a" in Fig. 1c. The big difference in the images between the kernel estimates and the space–time ETAS estimate is that the highest peak in the middle of Fig. 3a and b does not appear in Fig. 3c. This is due to the temporally symmetric kernel effect by the secondary aftershocks of the M7.5 aftershock that occurred 10 June 1968 (see also Fig. 2a and c), whereas the space–time ETAS intensity at the time of $t_a = 15{,}500.87$ days (9 June 1968) does not have such effect from the future as is also seen in Fig. 2e.

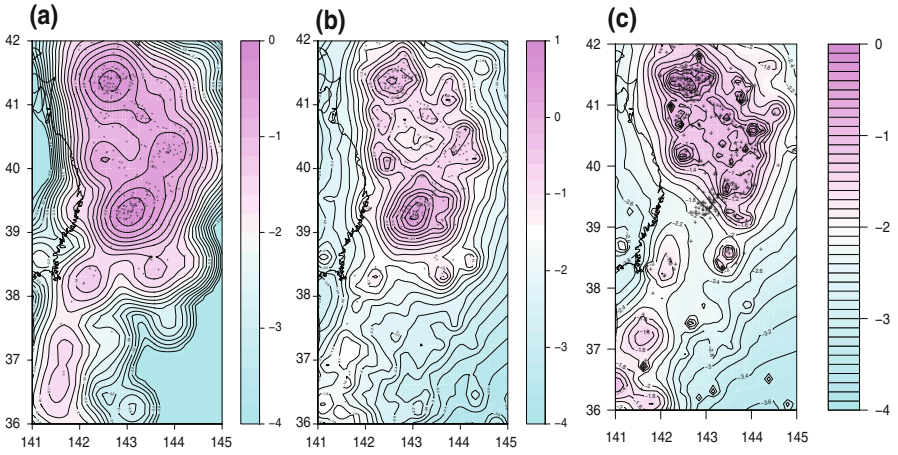

Fig. 3 The spatial intensity rates of the kernel intensities G-G (panel **a**) and P-P (panel **b**), and the space–time ETAS model (panel **c**) sliced at the time of 24.5 days after the 1968 Tokachi-Oki earthquake of M7.9. The contours and color tables are given equidistantly in logarithmic scale in units of events/degree^2/day. The plotted epicenters are earthquakes (mostly aftershocks) occurring from December 30, 1967 through October 14, 1968

7.2 The 1938 Shioya-Oki great swarms

During 1938, the Shioya-Oki great swarm, including several large earthquakes of M7 class, occurred successively off the coast of Fukusima Prefecture. Abe (1977) noted the unusual features of these earthquakes: the swarm of large earthquakes, the large energy release comparable to an earthquake of M8.1, and the inferred fact that there had been no major earthquakes about the focal region for at least the past 800 years, whereas the ordinary recurrence time of earthquakes of $M \geq 7.5$ occurring in other parts of the same plate boundary is usually less than 100 years.

We examine the intensity rate changes in time at the epicenter of the largest earthquake of M7.5 that occurred on November 5, 1938 ($t = 4691.74$ days), which is located at $(x_c, y_c) = (142.18°\text{E}, 37.33°\text{N})$. Figure 4e shows the space–time ETAS intensity $\hat{\lambda}(t, x_c, y_c|\mathcal{H}_t)$ at the location (x_c, y_c) indicated by the sign "**b**" in Fig. 1b. Figure 2f shows the corresponding intensity rates $\hat{\lambda}(\tau, x_c, y_c|\mathcal{H}_\tau)$ in terms of the

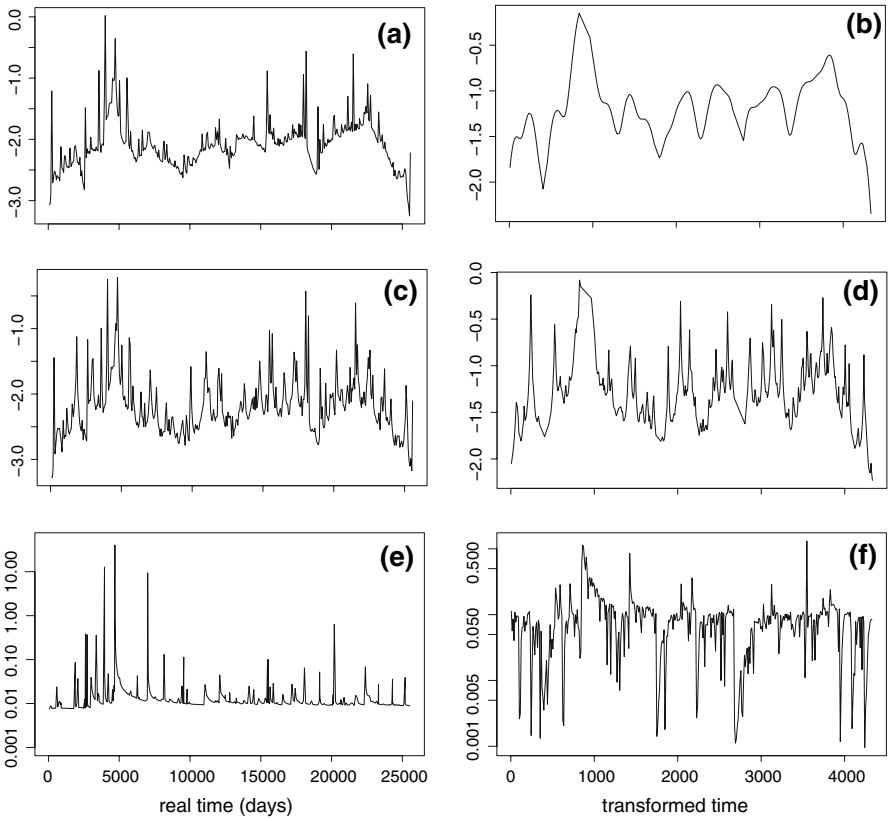

Fig. 4 Log-scaled intensity rate change in time at the location of the 1938 Shioya-Oki earthquake epicenter, projected from the space–time Gaussian kernel estimate G-G (panels **a** and **b**), inverse-power kernel estimate P-P (panels **c** and **d**), and space–time ETAS model (panels **e** and **f**), with respect to the transformed time (*right panels*) and the regular time (*left panels*)

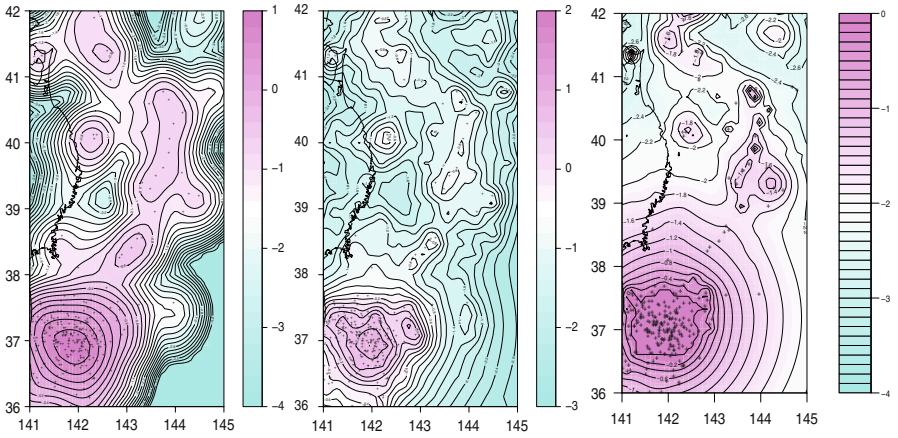

Fig. 5 The spatial intensity rates of the kernel intensities of G-G (*left panel*) and P-P (*middle panel*), and the space–time ETAS model (*right panel*) sliced at 3.1 days after the 1938 Shioya-Oki earthquake of M7.5. The contours and color tables are given equidistantly in logarithmic scale in units of events/degree^2/day. The plotted epicenters are earthquakes (mostly aftershocks) occurring from June 12, 1938 through March 31, 1939

transformed time. Figure 4b and d shows the kernel intensity estimates $\lambda_{\hat{\vartheta}}(\tau, x_c, y_c)$ of models G-G and P-P, respectively, in (5) at the same fixed location. Finally, Fig. 4a and b is the corresponding kernel intensity estimate $\lambda_{\hat{\vartheta}}(t, x_c, y_c)$ of the model G-G and P-P, respectively, in terms of the regular time.

Figure 5 shows the image of the spatial intensity rates of the kernel intensities and the space–time ETAS intensity, sliced at the fixed time $t_a = 4695.85$ days (November 9, 1938), which is 3.1 days after the mainshock. The plotted epicenters of earthquake in the figures are those occurring from June 12, 1938 ($t = 4545.43$ days) through March 31, 1939 ($t = 4835.84$ days), which roughly corresponds to cluster "**b**" in Fig. 1c.

7.3 The 1933 Sanriku-Oki great earthquake

The 1933 Sanriku earthquake of M8.1 is unusual since this intraplate earthquake occurred at the outer rise within the subducting Pacific plate, has the normal fault type mechanism caused by the extensional force (Kanamori 1971), and caused a great tsunami disaster; in contrast most of the major earthquakes in this region occur on the plate boundary and are caused by compressive forces. This earthquake occurred on March 3, 1933 ($t = 2618.10$ days), at the epicenter $(x_c, y_c) = (144.51°\text{E}, 39.23°\text{N})$ which is indicated by "**c**" in Fig. 1b.

We examine the intensity rate change in time at this fixed location. Figure 6e shows the conditional intensity function $\hat{\lambda}(t, x_c, y_c|\mathcal{H}_t)$ of the space–time ETAS model. Figure 6f shows the corresponding conditional intensity rates $\hat{\lambda}(\tau, x_c, y_c|\mathcal{H}_\tau)$ in term of the transformed time. Figure 6b and d shows the kernel intensity estimates $\lambda_{\hat{\vartheta}}(\tau, x_c, y_c)$ of models G-G and P-P, respectively, in (5) at the same fixed location. Finally, Fig. 6a

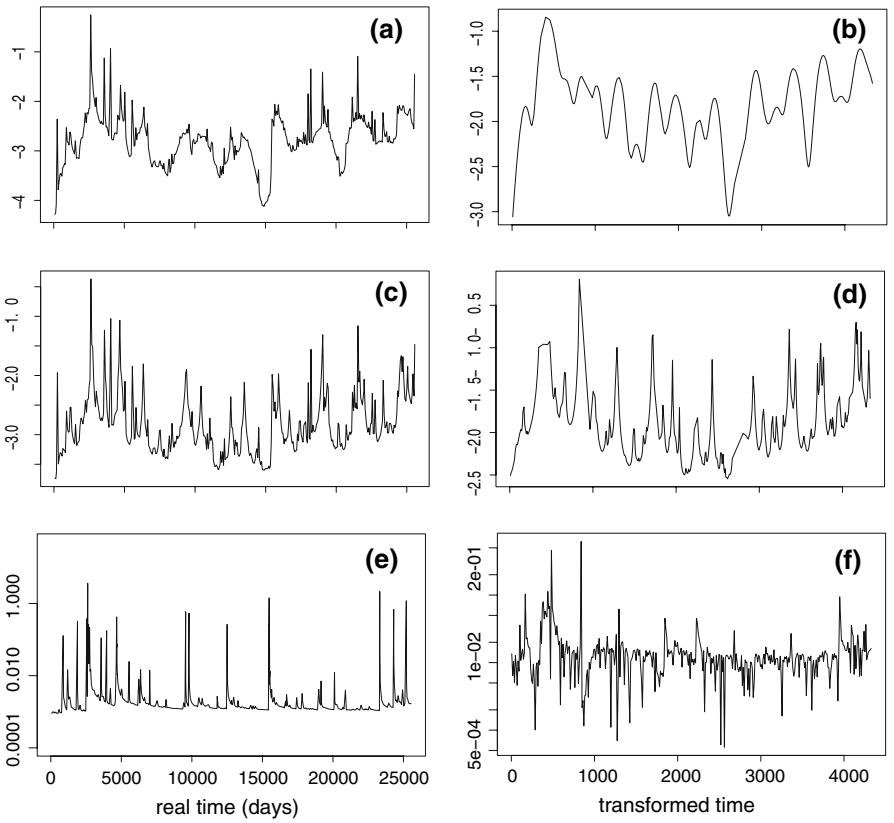

Fig. 6 Log-scaled intensity rate change in time at the location of the 1933 Sanriku-Oki earthquake epicenter, projected from the space–time Gaussian kernel estimate G-G (panels **a** and **b**), inverse-power kernel estimate P-P (panels **c** and **d**), and the space–time ETAS model (panels **e** and **f**), with respect to the transformed time (*right panels*) and the regular time (*left panels*)

and b is the corresponding kernel intensity estimate $\lambda_{\hat{\vartheta}}(t, x_c, y_c)$ of the model G-G and P-P, respectively, where the time is inversely transformed to the ordinary time t by means of relation (4).

Figure 7 shows the spatial intensity rates of the kernel intensities and the space–time ETAS intensity sliced at the fixed time $t_a = 2621.31$ days (March 6, 1933), which is 3.2 days after the mainshock. The plotted epicenters of earthquakes in the figures are those that occurred from November 26, 1932 ($t = 2521.88$ days) through August 29, 1933 ($t = 2797.90$ days), which corresponds roughly to the cluster indicated by "**c**" in Fig. 1c.

7.4 Discussion

With the present naive cross-validated evaluation procedure, it is easy to examine the different distributions for the kernels; therefore, it is worthwhile to compare the

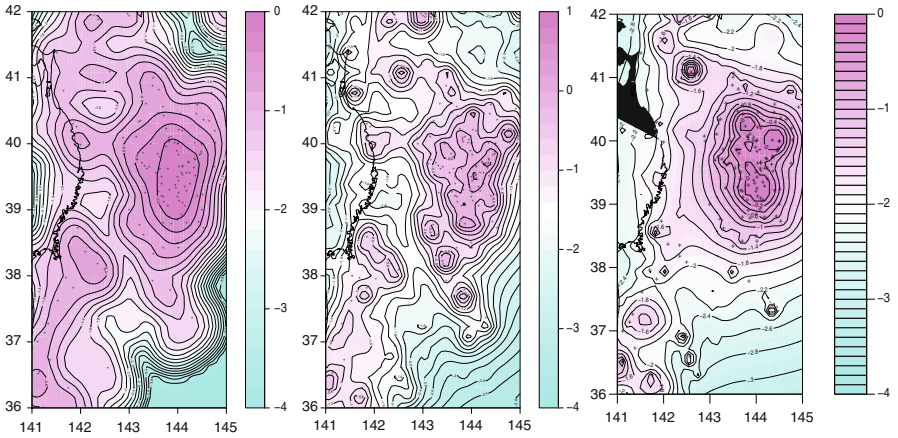

Fig. 7 The spatial intensity rates of the kernel intensities G-G (*left panel*), and P-P (*middle panel*), and the space–time ETAS model (*right panel*) sliced at 3.2 days after the 1933 Sanriku-Oki earthquake of M8.1. The contours and color tables are given equi-distantly in logarithmic scale in units of events/degree^2/day. The plotted epicenters are earthquakes (mostly aftershocks) occurring from November 26, 1932 through August 29, 1933

goodness-of-fit with different kernel distributions. According to the cross-validated likelihood scores in Table 2, the Gaussian kernel is not the best, but the inverse-power kernel is much better fitted to the present data.

From Figs. 3, 5, and 7, we see that the parametric space–time ETAS model represents the spatial intensity variations as flexibly as the kernel estimate does. This indicates that the weighted superposition of ellipsoid contoured functions (14) represents the time evolution of the spatial intensity quite well. From Figs. 2, 4, and 6, we see that the temporal peaks of the two estimates agree with one another. The highest peak of the intensity rate in each panel corresponds to the occurrence of the focal mainshock. On the other hand, the troughs of the temporal intensity variation appear very different in the two approaches. One of the main reasons is that the space–time ETAS model cannot adapt to the temporal change in background seismicity, since the function $\mu(x, y)$ is not time dependent, as seen in Eq. (13), while the kernel approach can adapt to estimate this variation. The background seismicity variations in time include anomalous geophysical information such as seismic quiescence and the type of activation, both of which may result from changing tectonic stresses due to slow seismic slips on nearby faults (Ogata 2007).

Ogata et al. (2003) introduced the hierarchical space–time ETAS (HIST-ETAS) model, in which parameter values vary depending on the location throughout the observed spatial region. This model describes observed seismicity better, since it takes into consideration the individual seismic features that cluster differently from place to place due to the different geological and tectonic regimes. However, the parameters of the HIST-ETAS model still do not vary with time. Instead, the study looks at the residual intensity rates relative to the HIST-ETAS model. In addition, as discussed by Schoenberg (2003), there is a departure from the separability assumption of clustering between the response functions in space and time in equation (13) and (5). Additional

diagnostic analysis of these features was our main motivation for comparing the parametric estimate with the kernel estimates, which should work more realistically and flexibly with seismic activity.

8 Conclusion

We present a practical method of kernel estimation for realistic description of the space–time evolution of seismicity rates with spiky, asymmetric, highly variable changes associated with clusters of earthquakes. This is first implemented by means of a time transformation using the theoretical cumulative function (4) of the temporal ETAS model (3). Namely, by the transformation, we use the fact that the space–time cluster point pattern become closer to an inhomogeneous Poisson process in the space–time volume. In addition, the time transformation using the ETAS model take the effect of earthquake magnitudes into consideration indirectly, whereas conventional kernel methods are based only on the locations of points.

Then the parameters of the space and time kernel bases are objectively determined by means of a naive likelihood cross-validation, for which the events in the dataset are randomly separated into two datasets of equal size: one for prediction and the other for validation. Such kernel intensity estimates are then transformed back to the corresponding intensity functions of the original regular time to compare with the intensity of the parametric space–time ETAS model.

Acknowledgments We thank Jiancang Zhuang for his advice on the use of the graphics software R in drawing the figures in this paper, and to Koichi Katsura for technical assistance. David Vere-Jones generously helped me very much with clarification of the paper. This work is partially supported by Grant-in-Aid 17200021 and 20240027 for Scientific Research (A), Ministry of Education, Science, Sports and Culture, and also by the 2006, 2007 and 2008 projects of the Institute of Statistical Mathematics and the Transdisciplinary Research Integration Center of the Research Organization of Information and Systems.

References

Abe, K. (1977). Tectonic implications of the large Shioya-Oki earthquakes of 1938. *Tectonophysics, 41*, 269–289.

Adelfio, G., Chiodi, M., De Luca, L., Luzio, D., Vitale, M. (2006). Southern-tyrrhenian seismicity in space–time-magnitude domain. *Annals of Geophysics, 49*(6), 1139–1151.

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

Akaike, H. (1998). A new look at the bayes procedure. *Biometrika, 65*, 53–59.

Akaike, H., Parzen, E., Tanabe, K., Kitagawa, G. (1998). *Selected papers of Hirotugu Akaike*. Berlin: Springer.

Choi, E., Hall, P. (1999). Nonparametric approach to analysis of space–time data on earthquake occurrences. *Journal of Computational and Graphical Statistics, 8*(4), 733–748.

Daley, D. J., Vere-Jones, D. (2003). *An introduction to the theory of point processes* (2nd ed.). New York: Springer.

Dziewonski, A., Chou, T., Woodhouse, J. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research, 86*, 2825–2852.

Grillenzoni, C. (2005). Non-parametric smoothing of spatio-temporal point processes. *Journal of Statistical Planning and Inference, 128*(1), 61–78. doi:10.1016/j.jspi.2003.09.030.

Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika, 58*(1), 83–90.

Hawkes, A., Adamopoulos, L. (1973). Cluster models for erthquakes-regional comparison. *Bulletin of the International Statistical Institute, 45*(3), 454–461.

Kanamori, H. (1971). Seismological evidence for a lithospheric normal faulting–the sanriku earthquake of 1933. *Physics of the Earth and Planetary Interiors, 4*, 289–300.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association, 83*(401), 9–27.

Ogata, Y. (1998). Space–time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics, 50*(2), 379–402.

Ogata, Y. (2004). Space–time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research, 109*, B03308. doi:10.1029/2003JB002621.

Ogata, Y. (2007). Seismicity and geodetic anomalies in a wide area preceding the Niigata-ken-Chue-tsu earthquake of 23 October 2004, central japan. *Journal of Geophysical Research, 112*, B10301. doi:10.1029/2006JB004697.

Ogata, Y., Katsura, K. (1988). Likelihood analysis of spatial inhomogenity for marked point patterns. *Annals of the Institute of Statistical Mathematics, 40*(1), 29–39.

Ogata, Y., Zhuang, J. (2006). Space–time ETAS model and an improved extension. *Tectonophysics, 413*, 13–23.

Ogata, Y., Katsura, K., Tanemura, M. (2003). Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. *Applied Statistics, 52*(4), 499–509.

Schoenberg, F. P. (2003). Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association, 98*(464), 789–795.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

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

Utsu, T. (1970). Aftershocks and earthquake statistics (ii): Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences. *Journal of the Faculty of Science, Hokkaido University, Series VII (geophysics), 3*, 198–266.

Utsu, T., Seki, A. (1955). Relation between the area of aftershock region and the energy of the main shock (in japanese). *Zisin (Jouirnal of the Seismological Society of Japan), 2nd Series, ii, 7*, 233–240.

Utsu, T., Ogata, Y., Matsu'ura, R. (1995). The centenary of the omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth, 43*, 1–33.