Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Stochastic Declustering of Space-Time Earthquake Occurrences

Jiancang Zhuang, Yosihiko Ogata & David Vere-Jones

To cite this article: Jiancang Zhuang, Yosihiko Ogata & David Vere-Jones (2002) Stochastic Declustering of Space-Time Earthquake Occurrences, Journal of the American Statistical Association, 97:458, 369-380, DOI: 10.1198/016214502760046925

To link to this article: https://doi.org/10.1198/016214502760046925

Published online: 31 Dec 2011.

Stochastic Declustering of Space-Time Earthquake Occurrences

Jiancang ZHUANG, Yosihiko OGATA, and David VERE-JONES

This article is concerned with objective estimation of the spatial intensity function of the background earthquake occurrences from an earthquake catalog that includes numerous clustered events in space and time, and also with an algorithm for producing declustered catalogs from the original catalog. A space-time branching process model (the ETAS model) is used for describing how each event generates offspring events. It is shown that the background intensity function can be evaluated if the total spatial seismicity intensity and the branching structure can be estimated. In fact, the whole space-time process is split into two subprocesses, the background events and the clustered events. The proposed algorithm combines a parametric maximum likelihood estimate for the clustering structures using the space-time ETAS model and a nonparametric estimate of the background seismicity that we call a variable weighted kernel estimate. To demonstrate the present methods, we estimate the background seismic activities in the central region of New Zealand and in the central and western regions of Japan, then use these estimates to produce catalogs of background events.

KEY WORDS: Background seismicity; Branching processes; Earthquake declustering; Space-time ETAS model; Thinning method; Variable weighted kernel estimate.

1. INTRODUCTION

The motivation of this article arises from problems in the prediction of large earthquakes with clusters of aftershocks. The earthquake clusters complicate the statistical analysis of the background seismic activity that might be related to changes in the tectonic field. The cluster features differ from place to place and typically lie between two extreme types of spatial clusters of earthquakes: those that eventually decrease in time, such as aftershock sequences and swarms, and those that persist in time at the same location. Ultimately, the persistent background activity prevails over the aftershock activity. To forecast the location of the large earthquakes, it is necessary to analyze the background seismicity, for which removal of temporal cluster members is considered to be of central importance.

A number of methods have been proposed for declustering a catalog to obtain the background seismicity. Most of these methods remove the earthquakes in a space-time window around a large event called the mainshock; the differences among these methods relate to the choice of the window sizes and some other detailed procedures (see, e.g., Utsu 1969; Gardner and Knopoff 1974; Kellis-Borok and Kossobokov 1986). In general, the bigger the magnitude of the mainshock, the larger the window. An alternative to the window-based declustering methods is the link method based on a space-time distance between the events (e.g., Resenberg 1985; Frohlich and Davis 1990; Davis and Frohlich 1991).

All of the aforementioned conventional methods contain arbitrary parameters for defining the aftershock window sizes or the link distances in both space and time. Each choice of parameter gives a different declustered catalog and thus a different estimate of the background seismicity. As a result, the conventional declustering algorithms have difficulty making an optimal choice of parameters. For a dataset from a specific (rather narrow) seismic region and a given threshold magnitude, these choices are usually colored by the user's subjective impression of the declustered output. Also, given a dataset, the concept of what constitutes an aftershock has not been uniquely defined. These are among the reasons why so many different declustering algorithms have been proposed.

To avoid such difficulties, it seems to us inevitable that the declustering method should be based on a stochastic model to objectively quantify the observations, so that any given event has a probability to be either a background event or an offspring (cluster) event generated by others. Therefore, the main task of a declustering method is to estimate this probability for each event. Several authors have proposed space-time models for the way in which earthquakes generate clustered events. To fit their models to Italian historical data, Musmeci and Vere-Jones (1992) assumed that the background seismicity was proportional to a kernel estimate obtained from the spatial locations of all earthquakes that include clusters. Ogata (1998) suggested a space-time version of the epidemic-type aftershock sequence (ETAS) model. He used a conventional declustering algorithm for preliminary estimation of the background rate by fitting bicubic B-spline functions before fitting the space-time ETAS model to all the events. However, as we show in this article, the use of such a preliminary declustering method can be avoided.

In the next section we give a formulation of the space-time point process for earthquake occurrences and then investigates the relationship between the total spatial rate, the background rate, and the branching structure of this model. This relationship suggests an approach to objective methods for esti-

Jiancang Zhuang is a graduate student, Department of Statistical Sciences, Graduate University for Advanced Studies, 4-6-7 Minami-Azabu, Minato-ku, 106-8569 Tokyo, Japan (E-mail: *zhuangjc@ism.ac.jp*). Yosihiko Ogata is Professor, Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-ku, 106-8569 Tokyo, Japan (E-mail: *ogata@ism.ac.jp*). David Vere-Jones is Emeritus Professor, School of Mathematics and Computing Sciences, Victoria University of Wellington, New Zealand. The study was initialized while the first author was a visitor to School of Mathematical and Computing Sciences, Victoria University of Wellington, supported by International Cooperation Fund of the China Seismological Bureau and an Earthquake Forecasting subcontract with Institute of Geological and Nuclear Sciences in New Zealand. It was continued with the support of grant-in-aid 11680334 for the Scientific Research from the Ministry of Education, Science, Sport and Culture, Japan. This study was also supported by a research fund of Yoshiyasu Tamura. The authors thank Ma Li for her continuous encouragement and support. The authors also gratefully acknowledge the assistance of Ray Brownrigg, David Harte, Makoto Taiji, and Koichi Katsura, as well as the helpful comments and encouragement from the referees and the editor.

mating the background intensity. We discuss several technical methods for carrying out such estimation, including the thinning (random deletion) method and a variable bandwidth kernel function method associated with maximum likelihood estimation of the space-time ETAS model. We then obtain the background intensity using an iterative algorithm and generate the declustered catalogs. Finally, we apply this procedure to datasets from the central region of New Zealand and the western and central Honshu area of Japan to demonstrate our methods.

2. SPACE–TIME MODELS FOR EARTHQUAKE OCCURRENCE

Several space-time point-process models have been proposed for describing clustering phenomena in seismic activity (Kagan 1991; Musmeci and Vere-Jones 1992; Rathbun 1993; Ogata 1998). These models have common features that can be outlined as follows:

(a) The background events are regarded as the "immigrants" in the branching process of earthquake occurrence; their occurrence rate is assumed to be a function of spatial location and magnitude, but not of time.

(b) Each "ancestor" event produces offspring independently. The expected number of direct offspring from an individual ancestor is assumed to depend on its magnitude M and is denoted by $\kappa(M)$.

(c) The probability distribution of the time until the appearance of an offspring event is a function of the time lag from its direct ancestor and is independent of magnitude (cf. Utsu 1962); thus its probability density function is assumed to have the form $g(t|\tau) = g(t-\tau)$, where τ is the occurrence time of the ancestor. Moreover, the function is independent of what happens between τ and t.

(d) The probability distributions of the location (x, y) and magnitude M of an offspring event are dependent on the magnitude M^* and the location (ξ, η) of its direct ancestor. These probability density functions are denoted by $f(x-\xi, y-\eta|M^*)$ and $j(M|M^*)$, where ξ, η represents the location and M^* represents the magnitude of the ancestor.

In general, this class of marked branching point processes for earthquake occurrences can be represented completely by the conditional intensity function (e.g., Daley and Vere-Jones 1988, chap. 13) defined by

$$\begin{aligned}\Pr\{&\text{an event in } [t, t+dt)\times[x, x+dx)\\ &\times[y, y+dy)\times[M, M+dM) \mid \mathcal{H}_t\}\\ &= \lambda(t, x, y, M \mid \mathcal{H}_t)\, dt\, dx\, dy\, dM\\ &\quad + o(dt\, dx\, dy\, dM),\end{aligned} \tag{1}$$

provided that it exists, where $\mathcal{H}_t$ denotes the space-time magnitude occurrence history of the earthquakes up to time t. Note that here $\mathcal{H}_t$ is the history of the available earthquake records including not only those during the study period and in the study area, but also those before the study period (e.g., Ogata 1992) and outside of the study area (e.g., this article). In particular, we include information about large earthquakes from outside the study area if they contribute substantially to the seismicity of the study region during the study period.

Based on the assumptions a–d, the conditional intensity function for the space-time model can be written as

$$\begin{aligned}\lambda(t, x, y, M \mid \mathcal{H}_t) = \mu(x, y, M) &+ \sum_{\{k:\ t_k<t\}} \kappa(M_k) g(t-t_k)\\ &\times f(x-x_k, y-y_k|M_k)\, j(M|M_k).\end{aligned} \tag{2}$$

In this equation, $\mu(x, y, M)$ is the background intensity function and is assumed to be independent of time. The functions $g(t)$, $f(x, y|M_k)$, and $j(M|M_k)$ are the normalized response functions (i.e., pdf's) of the occurrence time, location, and magnitude of an offspring from an ancestor of magnitude M_k. From the fact that the kth event excites a nonstationary Poisson process with intensity function $\kappa(M_k)g(t-t_k)f(x-x_k, y-y_k|M_k)j(M|M_k)$, it is easy to see that $\kappa(M_k)$ is the expected number of offspring from an ancestor of size M_k. Note also that the probability function $g(t)$ is independent of the magnitude of the ancestor, as mentioned in assumption c.

To simplify the subsequent discussion, we also make the following stronger assumption:

(e) The magnitude distribution of a background event is independent of its location, so that $\mu(x, y, M) = \mu(x, y) j_\mu(M)$; the magnitude distribution of a direct offspring event is independent of the size of its ancestor, so that $j(M|M^*) = j(M)$; and the magnitude distributions for the background events and their offspring are identical, that is, $j_\mu(M) = j(M)$.

Under these conditions, the conditional intensity function for the model can be decomposed as

$$\lambda(t, x, y, M \mid \mathcal{H}_t) = j(M)\lambda(t, x, y \mid \mathcal{H}_t), \tag{3}$$

where

$$\begin{aligned}\lambda(t, x, y \mid \mathcal{H}_t) = \mu(x, y) &+ \sum_{\{k:\ t_k<t\}} \kappa(M_k) g(t-t_k)\\ &\times f(x-x_k, y-y_k|M_k).\end{aligned} \tag{4}$$

Later, in applications, we make use of the specific function forms

$$\kappa(M) = A\mathrm{e}^{\alpha(M-M_0)}$$

$$g(t) = \begin{cases}(p-1)c^{p-1}(t+c)^{-p} & \text{for } t>0\\ 0 & \text{otherwise,}\end{cases}$$

and

$$f(x, y|M) = \frac{1}{2\pi d\mathrm{e}^{\alpha(M-M_0)}} \exp\left\{-\frac{1}{2}\frac{x^2+y^2}{d\mathrm{e}^{\alpha(M-M_0)}}\right\}$$

as an example. These constitute one form of the space-time ETAS model (Ogata 1998; see also Rathbun 1993 for a similar form).

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

$$\mu(x, y) = \nu u(x, y)$$

for the background rate in (4), where ν is a positive-valued parameter, and maximize the log-likelihood function

$$\log L(\boldsymbol{\theta}) = \sum_{k=1}^{N} \log \lambda_{\boldsymbol{\theta}}(t_k, x_k, y_k \mid \mathcal{H}_{t_k}) - \int_0^T \iint_S \lambda_{\boldsymbol{\theta}}(t, x, y \mid \mathcal{H}_t)\, dx\, dy\, dt \quad (5)$$

to obtain the maximum likelihood estimates (MLEs), $\hat{\boldsymbol{\theta}} = (\hat{\nu}, \hat{A}, \hat{\alpha}, \hat{c}, \hat{p}, \hat{d})$, where the subscript k runs over all of the events occurring in the study region S and the study time interval $[0, T]$. Computational details have been given by Ogata (1998).

Now, assuming stationarity and ergodicity of the space-time point process, the total spatial intensity function (first-order moment density) is equal to

$$m_1(x, y) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \lambda(t, x, y \mid \mathcal{H}_t)\, dt, \quad (6)$$

where T is the length of the observation period. Replacing the limit in (6) by a finite approximation and substituting (4) in (6), we obtain

$$\begin{aligned} m_1(x, y) &\approx \frac{1}{T} \int_0^T \left\{ \mu(x, y) + \sum_{\{k:\, t_k < T\}} \kappa(M_k) g(t - t_k) \times f(x - x_k, y - y_k | M_k) \right\} dt \\ &= \mu(x, y) + \frac{1}{T} \sum_{\{k:\, t_k < T\}} \kappa(M_k) f(x - x_k, y - y_k | M_k) \times \int_{t_k}^T g(t - t_k)\, dt \\ &= \mu(x, y) + \frac{1}{T} \sum_{\{k:\, t_k < T\}} \kappa(M_k) \times f(x - x_k, y - y_k | M_k). \end{aligned} \quad (7)$$

Therefore, an estimator of the background intensity can be derived from

$$\mu(x, y) \approx m_1(x, y) - \frac{1}{T} \sum_{\{k:\, t_k < T\}} \kappa(M_k) f(x - x_k, y - y_k | M_k), \quad (8)$$

where both the total rate $m_1(x, y)$ and the transfer function $\kappa(M) f(x, y|M)$ can be estimated.

3. THINNING PROCEDURE

To obtain a practical version of (8), we consider the following thinning operation (or random deletion). Consider a series of events $\{(t_j, x_j, y_j)\colon j = 1, 2, \ldots, N\}$ associated with a series of probabilities $\{\rho_j\colon j = 1, 2, \ldots, N\}$. Suppose that each event j of this process is deleted with probability ρ_j. Then the remaining events represent a new point process, called the *thinned process*. The simplest form of the thinning operation is to delete each point with a fixed probability (e.g., Daley and Vere-Jones 1972). The thinning method has also been used for the simulation of point processes (see e.g., Lewis and Shedler 1979; Ogata 1981, 1998; Musmeci and Vere-Jones 1992).

Here we interpret ρ_j as the probability of the jth earthquake being an offspring in the process, defined as follows. In the second term in (4), we set

$$\begin{aligned} \rho_{i,j} &= \Pr\{\text{the } j\text{th event is an offspring of } i\text{th event} | \mathcal{H}_{t_j}\} \\ &= \frac{\kappa(M_i) g(t_j - t_i) f(x_j - x_i, y_j - y_i | M_i)}{\lambda(t_j, x_j, y_j \mid \mathcal{H}_{t_j})} \end{aligned} \quad (9)$$

and

$$\begin{aligned} \rho_j &= \Pr\{\text{the } j\text{th event to be an offspring} | \mathcal{H}_{t_j}\} \\ &= \sum_{i=1}^{j-1} \rho_{i,j}, \end{aligned} \quad (10)$$

where $\rho_j \le 1$ because it represents the ratio of the sum term in (4) to the whole of (4). Thus the probability that the jth event belongs to the background is

$$\begin{aligned} \varphi_j &= \Pr\{\text{the } j\text{th event is an immigrant} \mid \mathcal{H}_{t_j}\} \\ &= 1 - \rho_j = \frac{\mu(x_j, y_j | \mathcal{H}_{t_j})}{\lambda(t_j, x_j, y_j \mid \mathcal{H}_{t_j})}. \end{aligned} \quad (11)$$

If we delete the jth event in the process with probability ρ_j for all $j = 1, 2, \ldots, N$, then the thinned process should realize a nonhomogeneous Poisson process of the spatial intensity $\mu(x, y)$ (see Ogata 1981 for the justification). We call this process the background subprocess and call the complementary subprocess the cluster subprocess or the offspring process.

4. VARIABLE KERNEL ESTIMATES OF SEISMICITY

Estimating the mean rate $m_1(x, y)$ of the total seismic activity in (8) can be carried out by several methods, including using kernel estimates (e.g., Vere-Jones 1992) or spline functions (e.g., Ogata and Katsura 1988). Here we adopt the kernel estimate method. The simple kernel estimate with a fixed bandwidth has a serious disadvantage, however. For a spatially clustered point dataset, a small bandwidth gives a noisy estimate for the sparsely populated area, whereas a large bandwidth mixes up the boundaries between the densely populated and sparsely populated areas. Therefore, instead of the kernel estimate

$$\hat{m}_1(x, y) = \frac{1}{T} \sum_{j=1}^{N} k_d(x - x_j, y - y_j), \quad (12)$$

where $k_d(x, y)$ denotes the Gaussian kernel function

$$k_d(x, y) = \frac{1}{2\pi d} \exp\left\{ -\frac{x^2 + y^2}{2d^2} \right\}$$

with a fixed bandwidth d, we adopt

$$\hat{m}_1(x, y) = \frac{1}{T} \sum_{j=1}^{N} k_{d_j}(x - x_j, y - y_j), \quad (13)$$

where d_j represents the varying bandwidth calculated for each event j in the following way. Given a suitable integer n_p between 10 and 100, find the smallest disk centered at the location of the jth event that includes at least n_p other earthquakes and with a radius larger than some small value (e.g., a distance within .02 degree, which is of the order of the location error), and let its radius be d_j (e.g., Silverman 1986, chap. 5). Similar ideas have been given by Choi and Hall (1998) and Musmeci and Vere-Jones (1986).

Using the same kernel function k_{d_j} as in (13), the occurrence rates of the cluster and background subprocesses can then be estimated by

$$\hat{\gamma}(x, y) = \frac{1}{T}\sum_j \rho_j k_{d_j}(x - x_j, y - y_j), \tag{14}$$

where ρ_j is derived from (10) and

$$\hat{\mu}(x,y) = \hat{m}_1(x,y) - \hat{\gamma}(x,y) = \frac{1}{T}\sum_j (1-\rho_j) k_{d_j}(x-x_j, y-y_j). \tag{15}$$

We call the estimates in (14) and (15) the *variable weighted kernel estimates.*

There remains the problem of optimal selection of n_p for the variable bandwidth for the statistics in (13), (14), and (15). Remembering that the clustering intensity term on the right side of (8),

$$I_T(x, y) = \frac{1}{T}\sum_j \kappa(M_j) f(x - x_j, y - y_j | M_j), \tag{16}$$

is also an image of the clustering intensity in space, but not so smoothed, we can in principle select a suitable n_p for $\hat{\gamma}(x, y)$ by minimizing the discrepancy between the low-frequency (smoothed) component of $I_T(x, y)$ in (16) and $\hat{\gamma}(x, y)$ in (14). In practice, the estimates are rather insensitive to the choice of n_p, so that a rough estimate is generally sufficient. At the same time, the mean rate $\hat{m}_1(x, y)$ is also compared to $\int_0^T \hat{\lambda}(t, x, y|\mathcal{H}_t)\, dt/T$ in view of (7), where $[0, T]$ is the observation time interval.

5. ITERATION ALGORITHM

In Sections 2–4, we outlined methodologic solutions to the following issues: (a) how to obtain the MLE of the space-time ETAS model for the branching structure when the background rate is given, (b) how to estimate the total spatial occurrence rate, and (c) how to estimate the background seismicity in the case where the branching structure is provided. But in reality we know only the earthquake data—namely, the records of occurrence times, locations, and magnitudes of the observed events. To estimate the background rate from these data only, we carry out the following iterative algorithm that simultaneously estimates the background rate and the branching structure:

Algorithm 1

1. Given a preliminary parameter n_p, say 20, calculate the bandwidth d_j for each event (t_j, x_j, y_j, M_j), $j = 1, 2, .., N$.
2. Set $l = 0$ and $u^{(0)}(x, y) = 1$.
3. Using the maximum likelihood procedure described by, for example, Ogata (1998), fit the conditional intensity function

$$\lambda(t, x, y|\mathcal{H}_t) = \nu u^{(l)}(x, y) + \sum_{k:t_k<t} \kappa(M_k) g(t - t_k) \times f(x - x_k, y - y_k|M_k) \tag{17}$$

   to the earthquake data, where k, g, and f are defined after (4).
4. Calculate ρ_j from (9) and (10) for each $j = 1, 2, \ldots, N$.
5. Calculate $\hat{\mu}(x, y)$ from (15) and record as $u^{(l+1)}(x, y)$.
6. If $\max_{(x,y)} |u^{(l+1)}(x, y) - u^{(l)}(x, y)| > \varepsilon$, where ε is a small positive number, then set $l = l + 1$ and go to step 3. Otherwise, take $\hat{\nu} u^{(l+1)}(x, y)$ as the background rate and stop.

The n_p is tuned so as to find the value that can minimize the difference between the low-frequency part of $I_T(x, y)$ and $\hat{\gamma}(x, y)$. In this study we find that the adjustment of n_p is not so sensitive for the final estimates of $\hat{m}_1(x, y)$ and $\hat{\gamma}(x, y)$, comparing to $\int_0^T \hat{\lambda}(t, x, y)\, dt$ and $I_T(x, y)$. The estimates only change slightly when n_p changes in the range of 15–100. In this article we adopt $n_p = 20$ as a standard value for use with the New Zealand and Japanese data.

This algorithm converges quickly. Table 1 shows an example of the convergence steps of this algorithm when the space-time ETAS model in (4) was applied to the earthquake data from the central New Zealand region, which is discussed in the next section.

Table 1. Convergence Steps of the Evaluation Procedure in Applying the Space-Time ETAS Model by Algorithm 1 to the Earthquake Data From the Central New Zealand Region

	log L	ν	A	c	α	p	d
1	−7987.49	.43594	.43378	.024737	.73554	1.1573	.39395e − 02
2	−7716.51	1.25850	.31482	.016264	.93397	1.1671	.19980e − 02
3	−7710.75	.95684	.33944	.017698	.88535	1.1638	.23097e − 02
4	−7696.38	1.00945	.33370	.017370	.89776	1.1644	.22326e − 02
5	−7695.60	.99755	.33492	.017441	.89518	1.1643	.22487e − 02
6	−7694.69	1.00011	.33465	.017426	.89574	1.1643	.22452e − 02
7	−7694.88	.99955	.33471	.017429	.89562	1.1643	.22459e − 02
8	−7694.84	.99966	.33470	.017428	.89565	1.1643	.22458e − 02

6. APPLICATIONS

6.1 Central New Zealand Region

The data used in this calculation are from the New Zealand local catalog recorded by the Institute of Geology and Nuclear Sciences. For this study, we select the earthquake data for the period January 1970–August 1999 from the rectangular area 38°–43°S and 171°–179°E (the central New Zealand region), which extends from the boundary of the Bay of Plenty to Arthur's Pass (Fig. 1). We take the magnitude threshold $M_L = 4.0$ and consider shallow events down to the depth of 40 km.

The final MLE values of the space-time ETAS model for this subregion are $\hat{A} = .33470$ event/(deg^2×day), $\hat{\alpha} = .89565$, $\hat{c} = .017428$ day, $\hat{p} = 1.1643$, and $\hat{d} = .0022458$/deg^2. From here on, we use deg^2 to denote the unit of the geographical area, where the unit degree is that of latitude (111.11 km). The converging steps of the calculation are shown in Table 1.

Figure 2(a) shows the temporal change of the function

$$p(t) = \frac{\iint_S \hat{\lambda}(t, x, y|\mathcal{H}_t)\, dx\, dy}{\iint_S \hat{\mu}(x, y)\, dx\, dy}, \tag{18}$$

that is, the integrated conditional intensity $\iint_S \hat{\lambda}(t, x, y|\mathcal{H}_t)\, dx\, dy$ over the whole study region S relative to the overall background rate $\iint_S \hat{\mu}(x, y)\, dx\, dy$. The pattern is similar to that from the simple ETAS model (Ogata 1988). Figure 2(b) shows the total intensity $\hat{m}_1(x, y)$ of the central New Zealand region. Figure 2(c) shows the spatial clustering intensity $\hat{\gamma}(x, y)$, calculated by (14). To enhance the contrast between Figures 2(b) and (c), we introduce the concept of relative clustering effect, which is shown by the contours in Figure 2(d). We do this to measure the cluster effect relative to the overall rate and defined by the ratio of the spatial clustering intensity to the total intensity, that is,

$$C(x, y) = \frac{\hat{\gamma}(x, y)}{\hat{m}(x, y)} = \frac{\sum_j \rho_j k_{d_j}(x - x_j, y - y_j)}{\sum_j k_{d_j}(x - x_j, y - y_j)}. \tag{19}$$

Here $C(x, y)$ can be regarded as a smoothed value of ρ_j.

Figures 2(b)–(d) show that most places in the study area have a total intensity of less than 50 events/deg^2 throughout the whole period of 28.75 years ($T = 10{,}500$ days). The clustering intensity is less than 20 events/deg^2 throughout the period, and the relative clustering effect is less than .3. Some spots get high values for these three quantities, however. They are located mainly in the plate boundary between the Pacific and Australian plates or in the inland volcanic zones.

By subtracting the clustering intensity from the overall intensity as indicated in (15), we can get the estimate of the background intensity function shown in Figure 2(e). It can be seen that the background rate is much smoother than the total rate in Figure 2(b). Most places have values in the range of 20–50 events/deg^2, except for the plate boundary, where the high values form up a strip with values of 50–100 events/deg^2 and some peaks up to 200 events/deg^2. Also, some high values are seen in the volcanic zones and offshore from Wanganui.

6.2 Central and Western Japan

The second example used data from the central and western Honshu area of Japan from the hypocenter catalog compiled by the Japan Meteorological Agency (Fig. 3). We selected data for the period 1926–1995 in the rectangular area 34°–39°N and 131°–140°E with magnitudes $M_J \geq 4.0$ and depths $\leq$100 km. Outside the southern boundary were two great earthquakes, the 1944 Tonankai earthquake of M_J7.9 with epicenter (136.62°E, 33.80°N) and the 1946 Nankai earthquake of M_J8.0 with epicenter (135.62°E, 33.03°N). These events and their large aftershocks apparently affected the seismic activity in the aforementioned selected inland area. However, the detection rate of earthquakes offshore is quite low compared to that inland, and thus it is hard to widen the region of the analysis without raising the threshold magnitude substantially. Instead, we keep the same area and same magnitude threshold as before for the analysis, but allow for boundary effects arising from the occurrence of the earthquakes not more than 1 degree outside the study area, to include the influence of the two great events with magnitude $\geq$7.8. Thus the likelihood calculation is implemented for the enlarged history $\mathcal{H}_t$ containing data from the extended region.

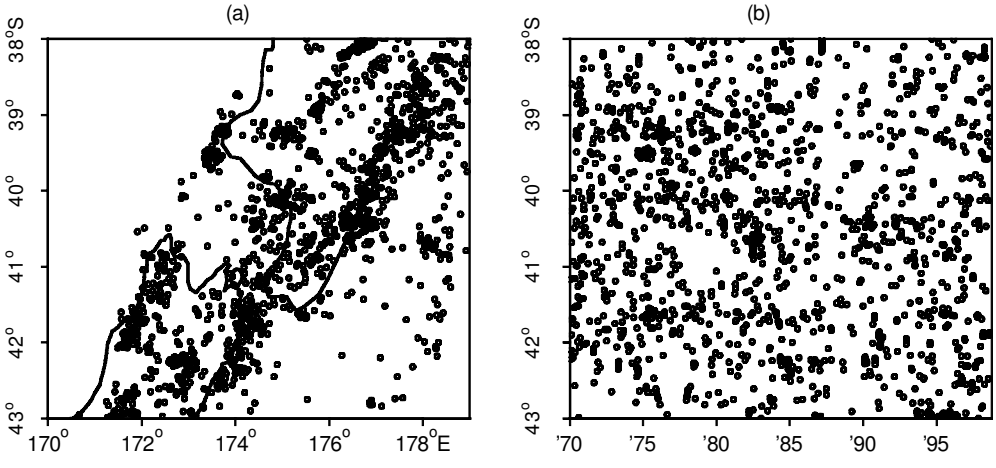

Figure 1. Epicentral Locations in the Central New Zealand Region (a) and Space-Time Plots for the Latitudes Against the Times of All of the Events (b).

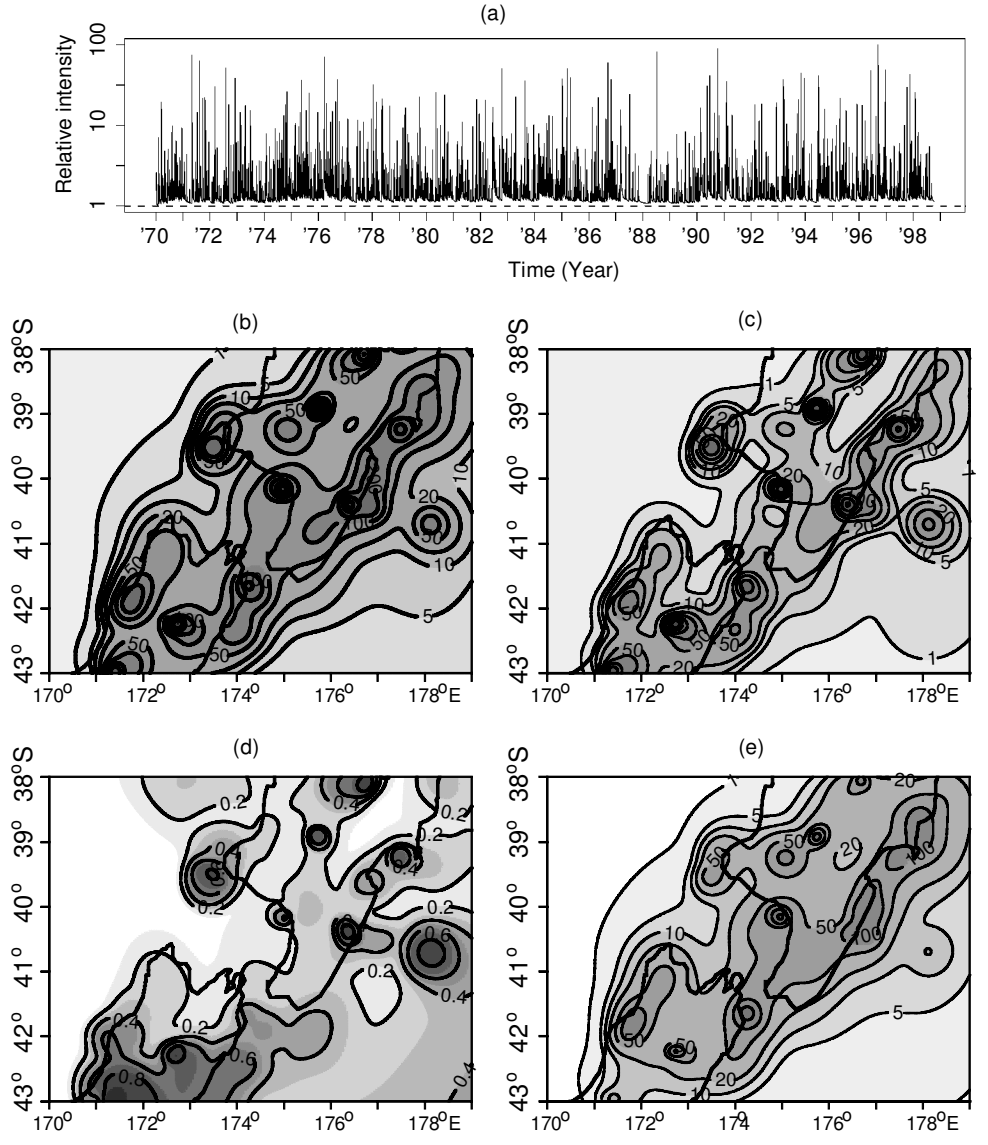

Figure 2. Results From the Central New Zealand Region. (a) Plot of the conditional intensity function of time integrated over the studied region. The horizontal axis represents time in years, and the vertical axis represents the ratio of the conditional intensity to the background intensity over the whole region. (b) Contours of the estimated total rate $\hat{m}_1(x,y)$ [see (13)]. (c) The clustering intensity $\hat{\gamma}(x,y)$ [see (14)]. (d) The estimated relative clustering effect $\widehat{C}(x,y)$ [see (19)]. (e) The estimated background seismicity rate $\hat{\mu}(x,y)$ [see (15)]. In (b), (c), and (e), the values are in events/deg^2 throughout the whole period ($T = 10{,}500$ days), and contours are drawn for approximately equal intervals in a logarithmic scale. The thick solid lines represent the coastlines.

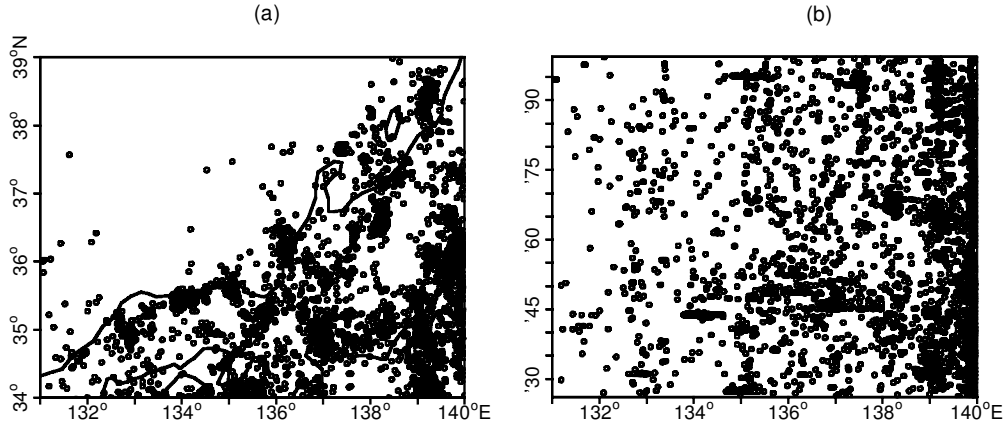

Figure 3. Epicentral Locations and Selected Subregions in the Western and Central Honshu Area of Japan (a) and Space-Time Plots for the Longitudes Against the Times of All of the Events (b).

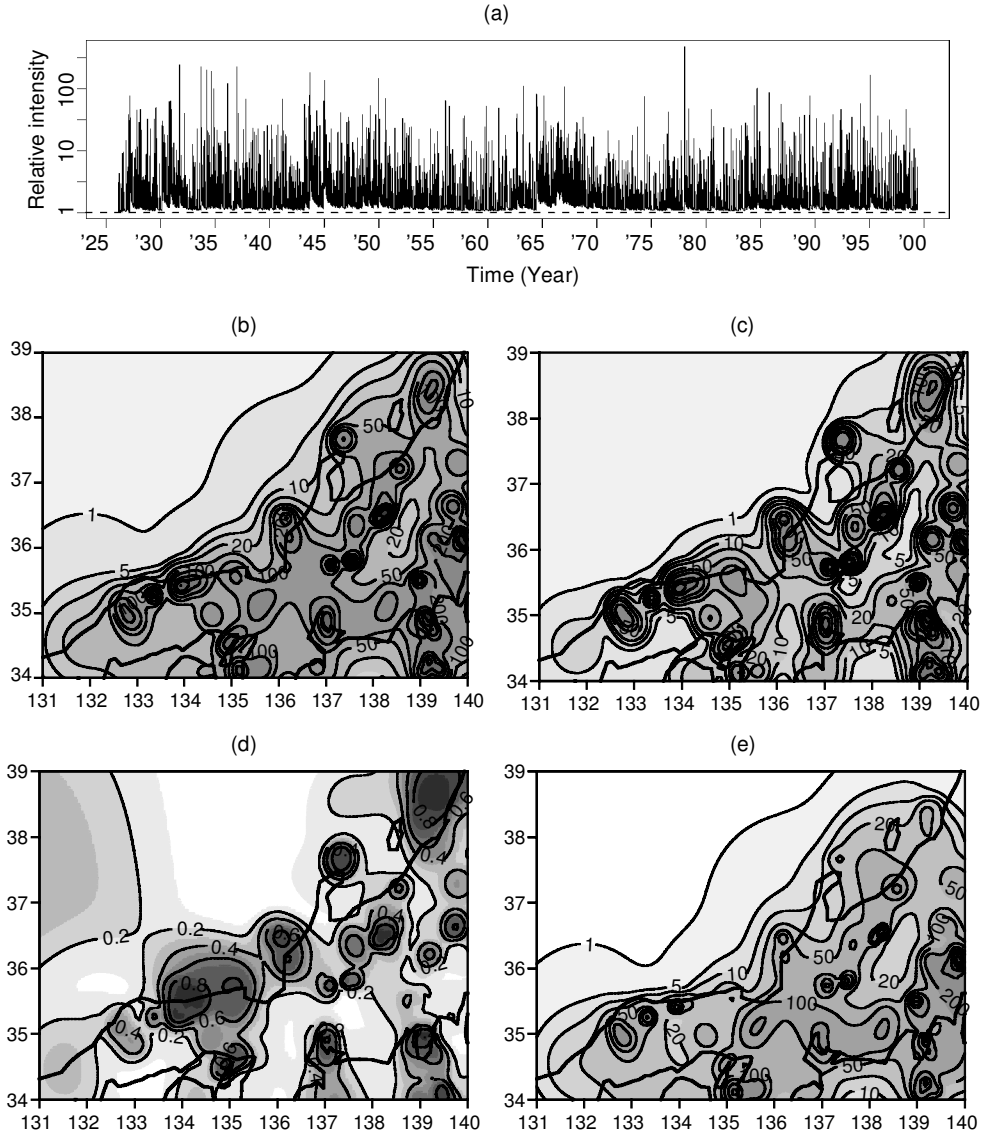

Figure 4. Results From the Central and Western Honshu Area of Japan. (a) Plot of the conditional intensity function of time integrated over the studied region. The horizontal axis represents time in years, and the vertical axis represents the ratio of the conditional intensity to the background intensity over the whole region. (b) Contours of the estimated total rate $\hat{m}_1(x,y)$ [see (13)]. (c) the clustering intensity $\hat{\gamma}(x,y)$ [see (14)]. (d) The estimated relative clustering effect $\widehat{C}(x,y)$ [see (19)]. (e) The estimated background seismicity rate $\hat{\mu}(x,y)$ [see (15)]. In (b), (c) and (e), the values are in events/deg^2 throughout the whole period ($T = 26{,}750$ days), and contours are drawn for approximately equal intervals in a logarithmic scale. The thick solid lines represent the coast lines.

The MLEs for this region are $\widehat{A} = .20376148$ event/(deg$^2 \times$ day), $\hat{\alpha} = 1.1355606$, $\hat{c} = .040436107$ day, $\hat{p} = 1.250478$, and $\hat{d} = .001033081$/deg^2. Figures 4(a)–(e) show estimates of the time variation of the function $p(t)$ in (18), total intensity, relative clustering effect, and background intensity. We can see that the clustering effects arising from typical aftershocks are eliminated in the estimates of the background intensity in Figure 4(d). Those include the 1927 Kita-Tango, 1943 Tottori, 1945 Mikawa, 1948 Fukui, 1964 Niigata, 1993 Noto-Hanto-Oki, and 1995 Kobe earthquakes. In contrast, the background rates are almost the same as the total intensities in the Wakayama and Ibaragi regions, which show almost pure spatial occurrence of nonclustered events.

7. PROBABILITY–BASED DECLUSTERING ALGORITHM

With the estimated background intensity and the branching structure (the space-time ETAS model), we can carry out the following declustering algorithm:

Algorithm 2

1. For each earthquake $j = 1, 2, \ldots, N$, calculate probability ρ_j in (10) from the final solution in algorithm 1.
2. Generate N uniform random numbers $U_1, U_2, \ldots, U_N$ in $[0, 1]$.
3. If $U_j < 1 - \rho_j$, then keep the jth event; otherwise, delete it from the catalog as an offspring. The remaining events can be considered the background events.

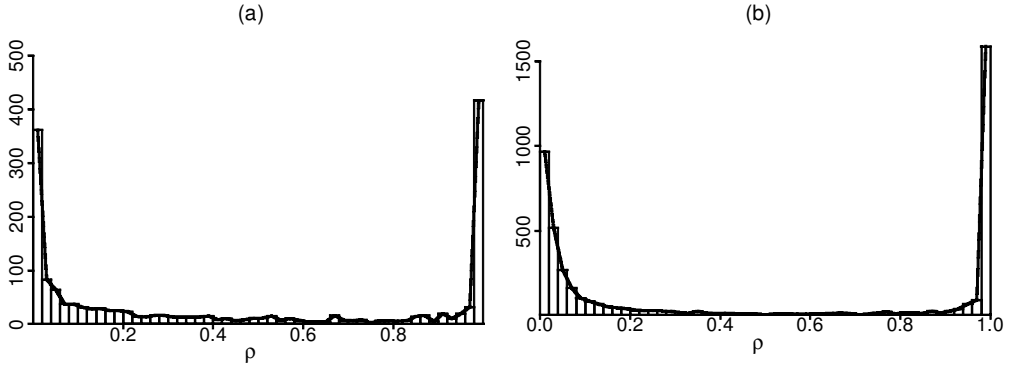

Figure 5. Histograms of Aftershock Probabilities [cf. (10)] for (a) the Central New Zealand Region and (b) the Western and Central Honshu Region.

The essence of the algorithm is the probability ρ_j associated with each event. Figure 5 shows a histogram of these probabilities for the two datasets. It can be seen that with very high probability, most of the events are either independent events or aftershocks, but that quite a few events are somewhere in between; about 31.7% of the New Zealand events and 18.3% of the Japanese events have probability of being an aftershock ranging from .1 to .9.

Using algorithm 2, we obtained examples of catalogs of background events from the two datasets. Typical examples are shown in Figures 6 and 7. Figures 6(a)–(b) and 7(a)–(b) show the epicentral locations of the independent events and the aftershock events; Figures 6(c)–(d) and 7(c)–(d) include space-time plots of the independent events and the aftershock events. Here we used the latitude for the New Zealand data and the longitude for the Japanese data.

Note that in Figure 7(c) for the declustered events, two clusters of linearly aligned events appear in 1945 and 1976 and in the longitude ranges 136°E–138°E and 137°E–140°E (see the arrows in the figure). The first cluster follows the December 1944 Tonankai earthquake of M_J7.9 and precedes the January 1945 Mikawa earthquake of M_J6.8 that occurred at 137.07°E

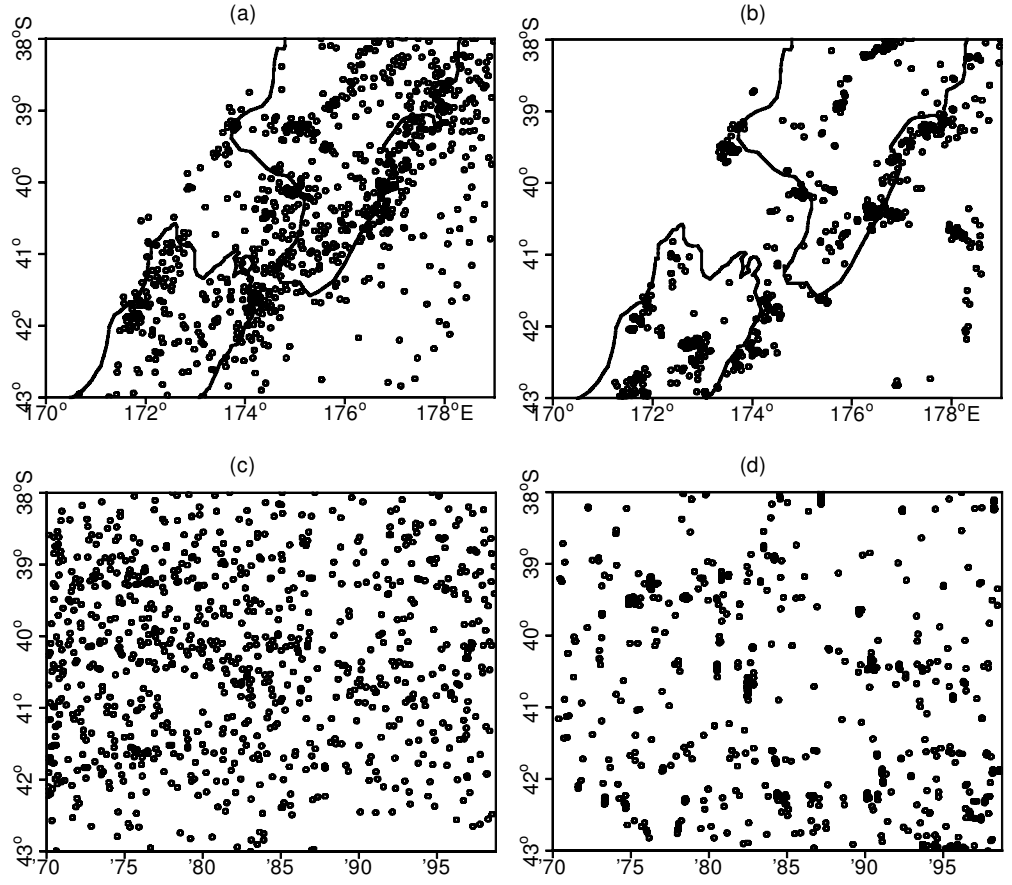

Figure 6. Results From Applying Algorithm 2 to the Central New Zealand Region. (a) Epicenter locations of the background events. (b) Epicenter locations of the offspring events. (c) Space-time plots of latitudes against times of the background events. (d) Space-time plots of latitudes against times of the offspring events.

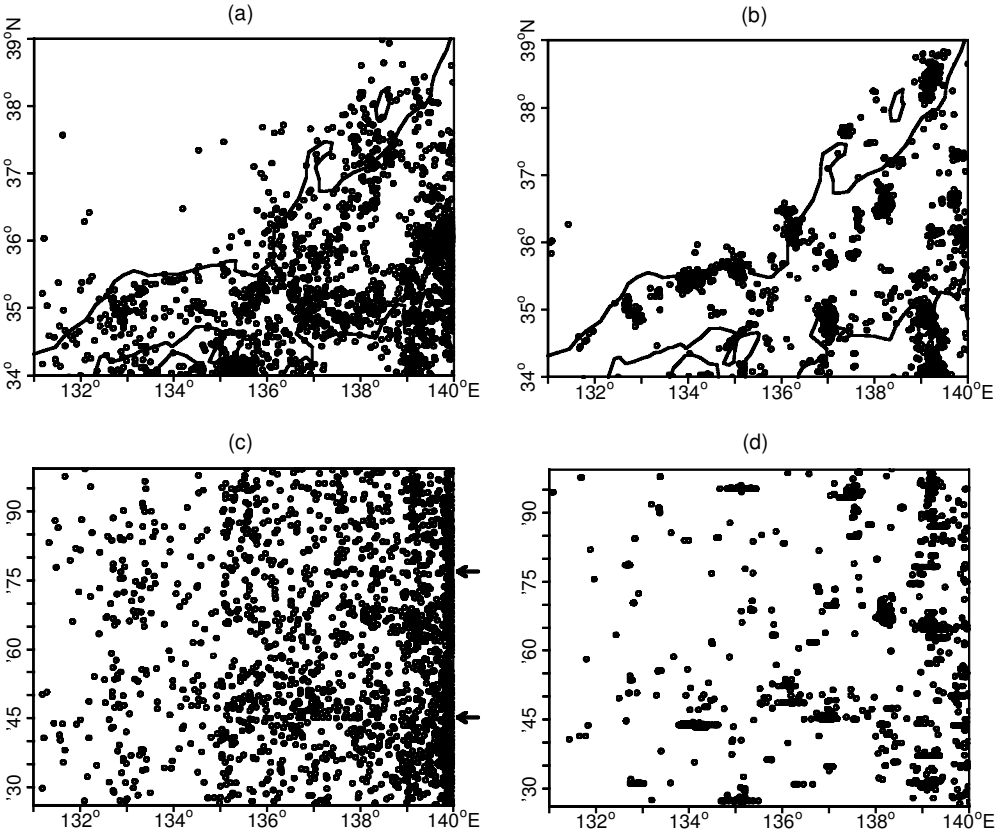

Figure 7. Results From Applying Algorithm 2 to the Western and Central Honshu Area. (a) Epicenter locations of the background events. (b) Epicenter locations of the offspring events. (c) and (d) Space-time plots for the longitudes against the times of the background events and the offspring events. The two arrows mark the occurrence time of the remote triggered events after 1944 and 1975.

and 34.68°N. We have learned that these events include quite a few events remotely triggered by the Tonankai earthquake, whose epicenter (136.18°E, 33.57°N) is outside the studied area. In other words, the remotely triggered clusters in time have not been removed, even though the log-likelihood (5) takes into account the effect of the boundary condition of the Tonankai event. The magnitudes of all of the events in the second cluster are relatively small, and a possible triggering event is the 1975 deep great event of M_J7.7, which occurred at (130.09°E, 38.79°N) at a depth of 549 km. The main reason for the detection of the remote triggering here is probably the short-range range decay of the spatial response function defined after (4) that conforms to the conventional understanding of an aftershock area (Utsu 1969).

Figures 6(d) and 7(d) include quite a few single events from which one may feel that the background patterns have not been well separated from the clusters. However, the single or double events that remain in the cluster catalog are the remains of the numerous double or triple events in the original catalog, most of which represent small clusters from mainshocks with magnitudes close to the lower magnitude threshold.

The following algorithm provides a stochastic realization of family trees based on the estimated space-time ETAS model:

Algorithm 3

1. For each pair of earthquakes $i, j = 1, 2, \ldots, N$ $(i < j)$, calculate probability $\rho_{i,j}$ in (9) and φ_j in (11) from the final solution in algorithm 1.
2. Set $j = 1$.
3. Generate a uniform random number U_j in $[0, 1]$.
4. If $U_j < \varphi_j$, then the jth event is considered to be an immigrant (background event).
5. Otherwise, select the smallest I such that $U_j < \varphi_j + \sum_{i=1}^{I} \rho_{i,j}$. Then the jth event is considered to be a descendant of the Ith event.
6. If $j = N$, then terminate the algorithm; otherwise, set $j = j + 1$ and go to step 3.

This algorithm enables us to identify a set of cluster families each starting from an immigrant. The earthquake dataset is stochastically separated into clusters and sets consisting of single events; the separation is stochastic because it depends on the series of uniform random numbers used in the foregoing algorithm. It should be noted that the events selected for the background here (and, consequently, those generated by algorithm 2) are always the first events occurring in the cluster family, but are not always the largest events in the family. However, if preferred, one can take the set of the largest event in each cluster as an alternative representation of the background seismicity.

8. DISCUSSION

In this section we discuss various additional issues relating to the methods proposed in the previous sections and the preliminary results obtained there.

8.1 Justification of the Underlying Model

It is obvious that the success of the algorithms proposed here depends in the first instance on a successful choice of the model for the underlying clustering process. The ETAS model used in this article has a long history of use for this purpose, initially in the time domain (see, e.g., Ogata 1988, 1999 and Utsu, Ogata, and Matsu'ura 1995 for reviews of earlier work) and more recently in the space-time domain. In particular, different possible forms of the response functions arising within the general framework of the space-time branching cluster model were carried out by Ogata (1998), using the Akaike Information Criterion (Akaike 1974) to discriminate between models.

More generally, when looking to justify a particular model for use as the basis of a declustering algorithm, it is worth bearing in mind that in addition to such likelihood-based tests, several conventional graphic statistics are available for testing the absolute goodness of fit of the model to specific space and time patterns of the data. These include the variance-time curve and the log-survivor function for time sequences (e.g., Ogata 1988; Andersen, Borgan, Gill, and Keiding 1993) and the K statistic and nearest-neighbor statistics for spatial point patterns (e.g., Ripley 1981; Diggle 1983). Graphic comparisons of results from the original data with those from repeated simulations from the model are particularly useful (see Ogata 1998 for details of simulation algorithms for space-time data). More sensitive graphic comparisons can be carried out using similar statistics for the transformed times (e.g., Ogata 1988) and possibly for the transformed higher-dimensional data, including locations (e.g., Schoenberg 1999) to test against the stationary Poisson process.

8.2 Dealing With Boundary Effects

As observed in Section 6.2, boundary effects can be important. Taking these into account, even if in only a rough way, can substantially improve the estimate of the background activity. In the present analysis, only the large events outside the region had a significant effect on estimates within the study region, because of the relatively short range of the spatial clustering terms. Thus it was sufficient to include in the augmented history only the coordinates of the larger events from outside the region. This represented quite a useful compromise, insofar as it reduced both the computational work and the demands on the underlying catalog, without causing significant loss of accuracy.

8.3 Interpretation of Temporal Fluctuations in the Estimated Background Activity

The term $\mu(x, y)$ in (4) for the background seismicity implies that the space-time characteristics of the declustered process are close to those of a Poisson process that is stationary in time even if it is nonhomogeneous in space. In particular, the proposed algorithm assumes a constant mean rate in time for the background seismicity at each location. Thus, if the thinned outcomes for the background events do not appear uniform in time, then this should indicate some departure from the base structure assumed for the clustering effects. For example, this may be caused by temporal activation or quiescence that has some physical cause outside those causes that create the standard ETAS clusters, or it may be due to some hitherto undetected fault in the catalog data. Revealing such phenomena and evaluating their significance, or uncertainty, are the eventual aims of the present declustering method. In effect, the model used for declustering plays the role of a null model, and departures from it represent effects worthy of further study and interpretation.

For example, in our study the previously discussed aligned declustered events in the space-time plots in Figure 7(c) clearly indicate the presence of remotely triggered events. This in turn suggests that the short-range decay function (2) adopted as the spatial response function of the model may require modification, perhaps replacement by a power law decay function, as has been indicated by Ogata (1998).

It is also possible that in some smaller regions, apparent differences from stationarity in time may be caused by artifacts of the dataset, in particular the substantial numbers of smaller aftershocks that may be missed (particularly in the early part of the catalog) in the time immediately after a mainshock. The rate of missed events around the threshold magnitude decreases as time elapses after the mainshock; further, this type of incompleteness of data has been improving as the ability to detect small earthquakes more rapidly has improved. Consequently, the estimates of c and p in the modified Omori function in (2) for the aftershock sequences may be biased by the missing events in early years, giving rise to some underdeclustering of the later aftershocks. Some indications of this effect can be seen in the space-time plot of the declustered events in Figure 7(e) for the earthquakes during the period of 1943–1946 and with the longitude range of 135°–138°E.

8.4 Nonuniqueness of the Declustered Catalog

A stochastic declustered catalog is not unique, because it is dependent on the random numbers used in selecting the events to form the background seismicity. Unlike conventional declustering methods, the stochastic declustering method does not make a fixed judgment on whether or not an event is an aftershock. Instead, it gives the probability of how likely each event is to be an aftershock. Each catalog produced by the stochastic declustering method should be understood as a simulation, or a resampling from the original catalog. This may be considered an advantage of the method rather than a disadvantage, because it allows uncertainty about the declustering to be quantified. Thus, by repeating random declustering, we can easily produce many stochastic copies of the declustered catalog. From these copies, graphic or other statistics can be calculated to evaluate the uncertainty or significance of a particular feature of the declustered catalog.

8.5 Preservation of Information

One may argue that throwing aftershocks out by the thinning (random deletion) method loses potentially useful information for carrying out seismicity analysis. Although this may be true of conventional declustering methods or for one particular realization of the stochastically declustered catalog, the essential output of the stochastic declustering algorithm is not the particular realization of the declustered catalog; rather,

they are probabilities associated with decomposition of the intensity into background and clustering terms. No data is lost in this process; rather, the process adds a new item, the probability ρ_j, to the parameters of the events in the original catalog. It should not be forgotten that the basic purpose of declustering is not the production of declustered catalogs for their own sake, but rather the investigation of features of the data that otherwise may be obscured by the clustering effects. For this purpose, the best representation of the combined information in both data and model is attaching these probabilities to the original events, whether studied directly or studied through the stochastically declustered catalogs to which they give rise.

8.6 Avoiding the Selection of Mainshocks as Aftershocks

As we can see from the declustering algorithm, the magnitude of an earthquake does not play a role in the random deletion. If a mainshock is preceded shortly by one or more foreshocks, then it has some probability of being interpreted as an offspring of one of the foreshocks, and in this case may not be selected as a background event. If this is regarded as a disadvantage, then we can separate the clusters by algorithm 3 and then adopt the largest event in each cluster instead of the initial event as the representative for inclusion in the catalog of background events.

8.7 Prospects for Future Study

Possibilities for further study of the stochastic declustering method can be discussed in two areas: theoretic problems and applications. On the theoretic side, some important aspects of the model remain to be considered. For example, we did not consider the important problem of deciding when the use of a nonhomogeneous background is necessary, either in space or in time, nor of how to estimate such time-varying effects in the background if these indeed are present. On the applications side, the stochastic declustering algorithm gives us a useful tool for the study of seismicity patterns. For example, in this article we have found some interesting differences between the New Zealand catalog and the Japanese catalog, including (a) the Japanese catalog has denser clusters in space than the New Zealand catalog, as can be seen from the parameters estimated by MLE, and (b) the New Zealand catalog has more events that cannot be clearly classified (31.7% events with probability .1–.9 of being aftershocks) than the Japanese catalog (only 18.3% events with probability .1–.9 of being aftershocks). The reasons for such differences require further study. For example, a hierarchical Bayesian space-time ETAS model has been developed recently (Ogata 2001).

Moreover, stochastic declustering can also help answer such questions as whether relative quiescence appears before a large earthquake, whether the occurrence of a large earthquake changes the seismicity patterns in the neighboring region, and what is the probability for an event to produce a larger offspring, that is, the probability of the occurrence of a foreshock. All of these important problems require careful evaluation of uncertainties, and although they all require further study, the stochastic declustering algorithm suggests one way in which such problems might be tackled in a more objective fashion than has been possible in the past.

9. CONCLUSIONS

In this article we have outlined a procedure for objectively estimating the spatial intensity of the declustered or background seismicity from an earthquake catalog. The method is based on an assumed stochastic model (spatial ETAS model) for the clustering patterns. Its core is a two-stage iterative algorithm based on the thinning method presented in Section 3 and estimation of the background seismicity in (13). The procedure stochastically splits the whole process into background events and offspring events grouped into clusters or family trees. Each such splitting can be regarded as on possible realization of how the original catalog could have been produced, taking into account that individual events can be classified as background events or cluster events only with a certain probability. We can estimate the significance and uncertainties of the features of the catalog not explainable in terms of the assumed model for clustering by running the procedure many times.

Using this method, preliminary studies were made of earthquake datasets from the central region of New Zealand and the from the central and western regions of Japan. Although preliminary, the results have revealed several features of interest that could be the subject of future study and interpretation.

[Received November 2000. Revised November 2001.]

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer.

Choi, E., and Hall, P. (1999), "Nonparametric Approach to Analysis of Space-Time Data on Earthquake Occurrences," *Journal of Computational and Graphical Statistics*, 8, 733–748.

Daley, D. J., and Vere-Jones, D. (1972), "A Summary of the Theory of Point Process," in *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, ed. P. A. W. Lewis, New York: Wiley.

Davis, S. D., and Frohlich, C. (1991), "Single-Link Cluster Analysis, Synthetic Earthquake Catalogs, and Aftershock Identification," *Geophysical Journal International*, 104, 289–306.

Diggle, P. (1983), *Statistical Analysis of Spatial Point Patterns*, London: Academic Press.

Frohlich, C., and Davis, S. D. (1990), "Single-Link Cluster Analysis as a Method to Evaluate Spatial and Temporal Properties of Earthquake Catalogs," *Geophysical Journal International*, 100, 19–32.

Gardner, J., and Knopoff, L. (1974), "Is the Sequence of Earthquakes in Southern California, With Aftershock Removed, Poissonian?" *Bulletin of the Seismological Society of America*, 64, 1363–1367.

Kagan, Y. (1991), "Likelihood Analysis of Earthquake Catalogues," *Journal of Geophysical Research*, 106, Ser. B7, 135–148.

Kellis-Borok, V. I., and Kossobokov, V. I. (1986), "Time of Increased Probability for the Great Earthquakes of the World," *Computational Seismology*, 19, 45–58.

Lewis, P. A. W., and Shedler, E. (1979), "Simulation of Non-Homogeneous Poisson Processes by Thinning," *Naval Research Logistics Quarterly*, 26, 403–413.

Musmeci, F., and Vere-Jones, D. (1986), "A Variable-Grid Algorithm for Smoothing Clustered Data," *Biometrics*, 42, 483–494.

——— (1992), "A Space-Time Clustering Model for Historical Earthquakes," *Annals of the Institute of Statistical Mathematics*, 44, 1–11.

Ogata, Y. (1981), "On Lewis' Simulation Method for Point Processes," *IEEE Translations on Information Theory*, IT-27, 23–31.

——— (1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9–27.

——— (1992), "Detection of Precursory Relative Quiescence Before Great Earthquakes Through a Statistical Model," *Journal of Geophysical Research*, 97, 19845–19871.

——— (1998), "Space-Time Point-Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402.

——— (1999), "Seismicity Analyses Through Point-Process Modelling—A Review," in *Seismicity Patterns, Their Statistical Significance and Physical Meaning*, eds. M. Wyss, K. Shimazaki, and A. Ito, Basel: Birkhauser-Verlag, pp. 471–507.

——— (2001), "Modelling of Heterogeneous Space-Time Seismic Activity and Its Residual Analysis," *Research Memorandum* 808, The Institute of Statistical Mathematics, Tokyo, submitted to *Applied Statistics*.

Ogata, Y., and Katsura, K. (1988), "Likelihood Analysis of Spatial Inhomogeneity for Marked Point Patterns," *Annals of the Institute of Statistical Mathematics*, 40, 20–39.

Rathbun, S. L. (1993), "Modeling Marked Spatio-Temporal Point Patterns," *Bulletin of the International Statistical Institute*, 55, Book 2, 379–396.

Resenberg, P. (1985), "Second-Order Moment of Central California Seismicity, 1969–1982," *Journal of Geophysical Research*, 90, Ser. B7, 5479–5495.

Ripley, B. D. (1981), *Spatial Statistics*, New York: Wiley.

Schoenberg, F. (1999), "Transforming Spatial Point Processes Into Poisson Processes," *Stochastic Processes and Their Applications*, 81, 155–164.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Utsu, T. (1962), "On the Nature of Three Alaskan Aftershock Sequences," *Bulletin of the Seismological Society of America*, 52, 279–297.

——— (1969), "Aftershock and Earthquake Statistics (I): Some Parameters Which Characterize an Aftershock Sequence and Their Interrelations," *Journal of the Faculty of Science, Hokkaido University*, 3, Ser. VII (Geophysics), 129–195.

Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995), "The Centenary of the Omori Formula for a Decay Law of Aftershock Activity," *Journal of Physics of the Earth*, 43, 1–33.

Vere-Jones, D. (1992), "Statistical Methods for the Description and Display of Earthquake Catalogues," in *Statistics in the Environmental and Earth Sciences*, eds. A. Walden and P. Guttorp, London: Edward Arnold, pp. 220–236.