Theme V – Models and Techniques for Analyzing Seismicity

Basic models of seismicity: Spatiotemporal models

Jiancang Zhuang[1] • Maximilian J. Werner[2] • Sebastian Hainzl[3] • David Harte[4] • Shiyoung Zhou[5]

1. Institute of Statistical Mathematics
2. Department of Geosciences, Princeton University
3. Institute of Geosciences, University of Potsdam
4. Statistics Research Associates
5. Department of Geophysics, Peking University

How to cite this article:

Zhuang, J., M.J. Werner, S. Hainzl, D. Harte, and S. Zhou (2011), Basic models of seismicity: spatiotemporal models, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-07487583. Available at http://www.corssa.org.

Document Information:

Issue date: 26 July 2011 Version: 1.0

Contents

Abstract In this article, we present a review of spatiotemporal point-process models, including the epidemic type aftershock sequence (ETAS) model, the EEPAS (Every Earthquake is Precursor According to Scale) model, the double branching model, and related techniques. Here we emphasize the ETAS model, because it has been well studied and is currently a standard model for testing hypotheses related to seismic activity.

1 Introduction

We assume that the reader has read the articles in this series on purely temporal models and the time independent spatial models (nonhomogeneous Poisson models) and that the reader is familiar with related concepts of point process models, especially the concepts of conditional intensity, likelihood, temporal ETAS model, and methods for estimating spatial rates of nonhomogeneous Poisson models. In this article, we continue with spatiotemporal models.

There have not been very many spatiotemporal models to describe seismicity. One reason is that heavy and complicated numerical computations are always involved in the implementation. Among the few spatiotemporal models that have been proposed, only the spatiotemporal epidemic-type aftershock sequence (ETAS) model has been extensively studied in the context of earthquake short-term clustering. Mathematically, the model is in the class of marked branching processes with immigration. Even though its formulation is based on empirical statistics, recent studies by *Console et al.* (2007) and *Iwata* (2010) suggest that the ETAS model is the best model for describing short-term seismicity, outperforming stochastic implementations of the physics-based model of *Dieterich* (1994). Both the EEPAS ("Every Earthquake is a Precursor According to Scale", *Rhoades and Evison* 2004) and the double branching models (*Marzocchi and Lombardi* 2008) have aspects in common with the ETAS model, i.e., using the ETAS model as the standard clustering component. For these reasons, we focus on issues related to the spatiotemporal ETAS model, such as the estimation of the background rate, stochastic declustering, stochastic reconstruction, etc.

The temporal ETAS model was suggested by *Ogata* (1988) with the conditional intensity function of

$$\lambda(t) = \mu + K_0 \sum_{t_i<t} \frac{\exp[\alpha(m - m_0)]}{(t - t_i + c)^p}, \tag{1}$$

where μ (shocks per unit time) represents the rate of background seismicity, the summation is taken over aftershocks occurring before time t, and m_0 represents the cut-off magnitude of the fitted data. In the above equation, the coefficient α

(magnitude^{-1}) is a measure of the efficiency of a shock in generating aftershock activity relative to its magnitude, K_0 represents the productivity of an event of threshold magnitude m_0, and c (unit of time) and p are the parameters in the Omori-Utsu law for describing the decay of the aftershock sequence.

Before Ogata generalized the temporal ETAS model to a space-time version, Musmeci and Vere-Jones (1992) used space-time diffusion clustering models to analyze seismicity in Italy. The conditional intensity functions for these models have the common form

$$\lambda(t, x, y) = \mu(x, y) + \sum_{i:t_i<t} g_\phi(t - t_i, x - x_i, y - y_i, m_i), \tag{2}$$

where

$$g_\phi(t, x, y, m) = A\frac{e^{\alpha m_i}e^{-ct}}{2\pi\sigma_x\sigma_y t}\exp\left\{-\frac{1}{2t}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right\}, \tag{3}$$

$$g_\phi(t, x, y, m) = \frac{Ae^{\alpha m}e^{-ct}t^2C_xC_y}{\pi^2(x^2 + t^2C_x^2)(y^2 + t^2C_y^2)}, \tag{4}$$

and A, α, σ_x, σ_y, C_x, and C_y are constants. For a fixed point (x, y), when $t \to \infty$ the aftershocks in (3) and (4) decay with time according to $t^{-1}e^{-ct}$ and $t^{-2}e^{-ct}$. *Kagan* (1991) and *Rathbun* (1993) also discussed slightly different forms.

The space-time ETAS model defined by *Ogata* (1998), now the generally accepted definition, had the same general form as in (2), but where $g_\phi(t, x, y, m)$ was defined differently, as

$$g_\phi(t, x, y, m) = \kappa(m)g(t)f(x, y|m). \tag{5}$$

In the above equation,

$$\kappa(m) = Ae^{\alpha(m-m_0)} \tag{6}$$

is the expected number of aftershocks generated from a mainshock of magnitude m,

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

is probability density function of the lagged time distribution of aftershocks, and

$$f(x, y|m) = \frac{1}{\pi\sigma(m)}f\left(\frac{x^2 + y^2}{\sigma(m)}\right) \tag{8}$$

is the density function of the aftershock locations from a mainshock at the origin with magnitude m.

2 Formulation of the spatiotemporal ETAS models

Even though there are some slight differences between the forms adopted by different researchers, the common forms of these models 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 event produces offspring events independently of the others. The expected number of direct offspring from an individual event is assumed to depend on its magnitude m, and will be denoted by $\kappa(m)$, which is called its productivity.
(c) The probability distribution of the time until the appearance of a child event is a function of the time lag from its parent, and is independent of magnitude; 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 a child event are dependent on the magnitude m^* and the location (ξ, η) of its parent. These probability density functions are denoted by $f(x-\xi, y-\eta|m^*)$ and $s(m|m^*)$, respectively, where ξ, η and m^* are the location and magnitude of the ancestor.
(e) The magnitudes of all the events, including background events and their offspring, are independent random variables drawn from the same probability distribution of density $s(m)$.

In general, this class of marked branching point processes for earthquake occurrences can be represented by its conditional intensity function, which is written

$$\begin{aligned} &\Pr\{\text{an event in } [t, t+\mathrm{d}t) \times [x, x+\mathrm{d}x) \times [y, y+\mathrm{d}y) \times [m, m+\mathrm{d}m) |\, \mathcal{H}_t\} \\ &= \lambda(t, x, y, m)\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}m + o(\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}m), \end{aligned} \tag{9}$$

where $\mathcal{H}_t$ denotes the space-time-magnitude occurrence history of the earthquakes up to time t.

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

$$\lambda(t, x, y, m) = s(m)\left[\mu(x, y) + \sum_{\{k:\ t_k<t\}} \kappa(m_k) g(t-t_k) f(x-x_k, y-y_k; m_k)\right]. \tag{10}$$

In the above equation, $\mu(x, y)$ is the background intensity function, which is assumed to be independent of time. The functions $g(t)$, $f(x, y; m_k)$ and $s(m)$ are respectively the normalized response functions (i.e., probability density functions) of the occurrence time, the location, the magnitude of an offspring from an ancestor of magnitude m_k. And from the fact that the k-th event excites a non-stationary Poisson process with intensity function $\kappa(m_k)g(t-t_k)f(x-x_k, y-y_k; m_k)s(m)$, we note that $\kappa(m_k)$ is the expected number of children from a parent of size m_k. This implies that the number of children is a Poisson random variable with a mean of $\kappa(m_k)$. Note also that the probability function $g(t)$ is independent of the magnitude of the parent, as mentioned in assumption (c). Another explanation of (10) is that the risk of earthquake occurrence at time t and location (x, y) consists of the contribution from the background rate μ and the contributions from each previous event, $\xi(t, x, y; t_i, x_i, y_i)$.

According to (10), the conditional intensity function for the model can be decomposed as

$$\lambda(t, x, y, m) = s(m)\lambda(t, x, y), \tag{11}$$

where

$$s(m) = \beta\, \mathrm{e}^{-\beta(m-m_0)} \tag{12}$$

represents the Gutenberg-Richter law for earthquake magnitudes of m_0 or larger in the form of a probability density function, β is linked with the so-called b-value by $\beta = b \ln 10$, and

$$\lambda(t, x, y) = \mu(x, y) + \sum_{\{k:\ t_k<t\}} \kappa(m_k)g(t-t_k)f(x-x_k, y-y_k; m_k). \tag{13}$$

In applications, the following specific functions are often used. The productivity law is given by (6) and the Omori-Utsu law by (7).

For the spatial component, $f(x, y; m)$, the following different functions have been used:

Model 1 [*Rathbun* 1993; *Console et al.* 2003]

$$f(x, y; m) = \frac{1}{2\pi D^2} e^{-\frac{x^2+y^2}{2D^2}}; \tag{14}$$

Model 2 [*Ogata* 1998; *Zhuang et al.* 2002]

$$f(x, y; m) = \frac{1}{2\pi D^2 e^{\alpha(m-m_0)}} e^{-\frac{x^2+y^2}{2D^2 e^{\alpha(m-m_0)}}}, \tag{15}$$

where the parameter α is the same one as in (6);

Model 3 [*Ogata* 1998; *Console et al.* 2003]

$$f(x, y; m) = \frac{q-1}{\pi D^2}\left(1 + \frac{x^2+y^2}{D^2}\right)^{-q}; \tag{16}$$

Model 4 [*Ogata* 1998; *Zhuang et al.* 2002, 2004]

$$f(x, y; m) = \frac{q-1}{\pi D^2 e^{\alpha(m-m_0)}}\left(1 + \frac{x^2+y^2}{D^2 e^{\alpha(m-m_0)}}\right)^{-q}; \tag{17}$$

and
Model 5 [*Zhuang et al.* 2005; *Ogata and Zhuang* 2006]

$$f(x, y; m) = \frac{q-1}{\pi D^2 e^{\gamma(m-m_0)}} \left(1 + \frac{x^2 + y^2}{D^2 e^{\gamma(m-m_0)}}\right)^{-q}. \tag{18}$$

The difference between Models 4 and 5 is that α is identical to the parameter α in the productivity function $\kappa(m)$, whereas γ is completely independent of $\kappa(m)$. As shown in *Zhuang et al.* (2004); *Zhuang* (2006), and *Ogata and Zhuang* (2006), Model 5 (18) usually fits earthquake data the best among those five forms. However, the other four forms are also frequently used by many researchers.

3 Technical points related to the spatiotemporal ETAS model

3.1 Maximum likelihood procedure

Given the background rate $\mu(x, y)$ and the observed earthquake catalog, the maximum likelihood estimates (MLE) of the model, $\hat{\theta} = (\hat{\nu}, \hat{A}, \hat{\alpha}, \hat{c}, \hat{p}, \hat{D})$, are calculated by maximizing the log-likelihood function

$$\log L(\boldsymbol{\theta}) = \sum_{i:t_i \in [0,T], (x_i,y_i) \in S} \log \lambda_{\boldsymbol{\theta}}(t_i, x_i, y_i) - \int_0^T \iint_S \lambda_{\boldsymbol{\theta}}(t, x, y)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t, \tag{19}$$

where the index i runs over all the events occurring in the study region S and the study time interval $[0, T]$. The computational details can be found in *Ogata* (1998). *Veen and Schoenberg* (2008) developed an EM (expectation-maximization) algorithm for a fast estimation by maximizing the expected log-likelihood.

In (19), the space-time window $[0, T] \times S$ is called the target window, and the events inside are called target events. Different from the target events, the index k in (10) runs over all the recorded events in the catalog, whose range is called the auxiliary or complementary window. Usually the complementary window should be taken as large as possible. A more detailed discussion of the influence of the size of the complementary window on the estimated model parameters was provided by *Wang et al.* (2010).

3.2 Thinning procedure

An interesting question in seismicity analysis is to what extent a given earthquake is triggered by a previous earthquake. The so-called thinning procedure (see, e.g., *Lewis and Shedler* 1979; *Ogata* 1981; *Daley and Vere-Jones* 2003), is an important tool for such an analysis and will be briefly described here. Below, we use the concept for assigning a "degree of parenthood" to earthquake pairs and to estimate the spatial background density in the ETAS model. The proportion of the contribution from event i at the occurrence of (t_j, x_j, y_j) could be explained as the probability that event j is triggered by the ith event,

$$\rho_{ij} = \begin{cases} \dfrac{\xi(t_j, x_j, y_j; t_i, x_i, y_i)}{\lambda(t_j, x_j, y_j)}, & \text{when } j > i, \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

Moreover, the probability of the event j being a triggered event is

$$\rho_j = \sum_i \rho_{ij}, \tag{21}$$

and the probability that the jth event belongs to the background is

$$\varphi_j = 1 - \rho_j = \frac{\mu(x_j, y_j)}{\lambda(t_j, x_j, y_j)}. \tag{22}$$

If we select each event j with probabilities ρ_{ij}, ρ_j or φ_j, we can form a new process being the process triggered by the ith event, the clustering process or the background process, respectively. More details and an example are given in Section 4.

3.3 Estimating background rate

Given a dataset of earthquake occurrence times, locations and magnitudes in the observation period T, the spatial distribution of the total seismicity rate is usually estimated by some nonparametric method, such as splines (*Ogata* 1998), kernel functions (*Zhuang et al.* 2002; *Helmstetter et al.* 2007), grid averaging (*Tsukakoshi and Shimazaki* 2006), or tessellation (*Ogata* 2004). In this paper, we consider the variable (adaptive) kernel estimate following *Zhuang et al.* (2002). A similar adaptive power law kernel was used by *Werner et al.* (2011). This adaptive approach is simple and tackles a serious disadvantage of the simple kernel estimate with a fixed bandwidth: for a spatially clustered point dataset, a small bandwidth gives a noisy or variable estimate for the sparsely populated area, while a large bandwidth mixes up the boundaries between the densely populated areas and the sparsely populated areas. Therefore, instead of the kernel estimate

$$\hat{m}_1(x, y) = \frac{1}{T} \sum_{j=1}^{N} k_d(x - x_j, y - y_j) \tag{23}$$

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

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

with a fixed bandwidth d, we adopt

$$\hat{m}_1(x, y) = \frac{1}{T} \sum_{j=1}^{N} k_{d_j}(x - x_j, y - y_j), \tag{24}$$

where d_j represents the varying bandwidth calculated for each event j in the following way: Given a suitable integer n_p, find the smallest disk centered at the location of the j-th event which includes at least n_p other earthquakes and is larger than some minimum value (e.g., a distance within 0.02 degree, which is of the order of the location error, see also CORSSA Article in Theme IV) and let this minimum radius be d_j (e.g., *Silverman* 1986, Chapter 5).

Similar to estimating total seismicity, there have been many approaches for assessing background seismicity rate: (i) proportional to total seismicity rate of all events or only of the big events in the catalog (*Musmeci and Vere-Jones* 1986; *Console et al.* 2003); (ii) using a declustering method to decluster the catalog and use the total rate in the declustered catalog as background rate (*Ogata* 1998); (iii) weighting each event by the probability that it is a background event (*Zhuang et al.* 2002, 2004); and (iv) *Ogata* (2004) introduced a Bayesian smoothness prior on tessellation grids to estimate the spatial variation of the background and the model parameters at the same time. In this study, we use the third method because of its simplicity together with unbiased estimates of sufficient precision.

Using the method described in the last paragraph of Section 3.2, once a background process is obtained, we can estimate the background intensity by applying a smoothing technique to the background catalog. Rather than repeating the thinning procedure and the kernel estimation procedure many times to get an average estimate of the background intensity, we can directly estimate the average by weighting all the events with their corresponding background probabilities, i.e.,

$$\hat{\mu}(x, y) = \frac{1}{T} \sum_{i} \varphi_i k_{d_i}(x - x_i, y - y_i), \tag{25}$$

where i runs over all of the events in the whole process, T is the length of process duration, and k_d is the Gaussian kernel function with a bandwidth d. The variable bandwidth d_i is defined in the same way as in (24).

3.4 Iterative algorithm

Sections 3.1 to 3.3 give us the method for estimating any one of the following three things when the other two are known: the background rate μ, the model parameters θ, and the background probabilities φ_i. In most cases, we have none of them but a catalog of earthquakes. We can use the following algorithm to estimate them simultaneously.

Algorithm A: stochastic declustering

A1. Given a fixed n_p and d_{min}, say 5 and 0.05 degree (roughly equivalent to 5.56 km on Earth's surface, which is close to the location error of earthquakes), calculate the bandwidth d_j for each event (t_j, x_j, M_j), $j = 1, 2, \cdots, N$.

A2. Set $\ell = 0$, and $u^{(0)}(x, y) = 1$.

A3. Using the maximum likelihood procedure (see, *Ogata* 1998), fit the model with conditional intensity function

$$\lambda(t, x, y) = \nu u^{(\ell)}(x, y) + \sum_{i:\, t_i<t} \kappa(M_i) g(t - t_i) f(x - x_i, y - y_i; M_i) \qquad (26)$$

to the earthquake data, where κ, g and f are defined in (10), and ν is the relaxation coefficient, introduced to speed up the convergence of the iterations.

A4. Calculate ρ_{ji}, ρ_i and φ_i for each $j < i$ and $i = 1, 2, \cdots, N$.

A5. Calculate $\mu(x, y)$ from (25) and record it as $u^{(\ell+1)}(x, y)$.

A6. If $\max |u^{(\ell+1)}(x, y) - u^{(\ell)}(x, y)| > \varepsilon$, where ε is a given small positive number (i.e., on the order of 10^{-6}), then set $\ell = \ell + 1$ and go to step A3; otherwise, take $\nu u^{(\ell+1)}(x, y)$ as the background rate and also output ρ_{ij}, ρ_i and φ_i.

Example 1 **Fitting ETAS model to the central Japan region**

We use the earthquake data from the central-western Japan region from the hypocenter catalog compiled by the Japan Meteorological Agency (JMA; see Fig. 1). We selected data for the period 1926-1995 in the rectangular area $34 \sim 39°N$ and $131 \sim 140°E$ with magnitudes $M_J \geq 4.0$ and depths $\leq 100km$.

The space-time ETAS model equipped with the spatial response (18) is fit to the above dataset. The MLEs for this region are $\hat{A} = 0.020\, event/(deg^2 \cdot day)$, $\hat{\alpha} = 1.135$, $\hat{c} = 0.040$ days, $\hat{p} = 1.15$, and $\hat{d} = .0010/deg^2$. Figures 1 (a)-(d) show estimates of the spatial variations of total intensity, clustering intensity, background intensity and relative clustering effect, respectively. We can see that the clustering effects arising from typical aftershocks are eliminated in the estimates of the background intensity in Figure 1(d). In contrast, the background rates are almost the same as the total intensities in the Wakayama and Ibaraki regions, which show almost pure spatial occurrence of nonclustered events.

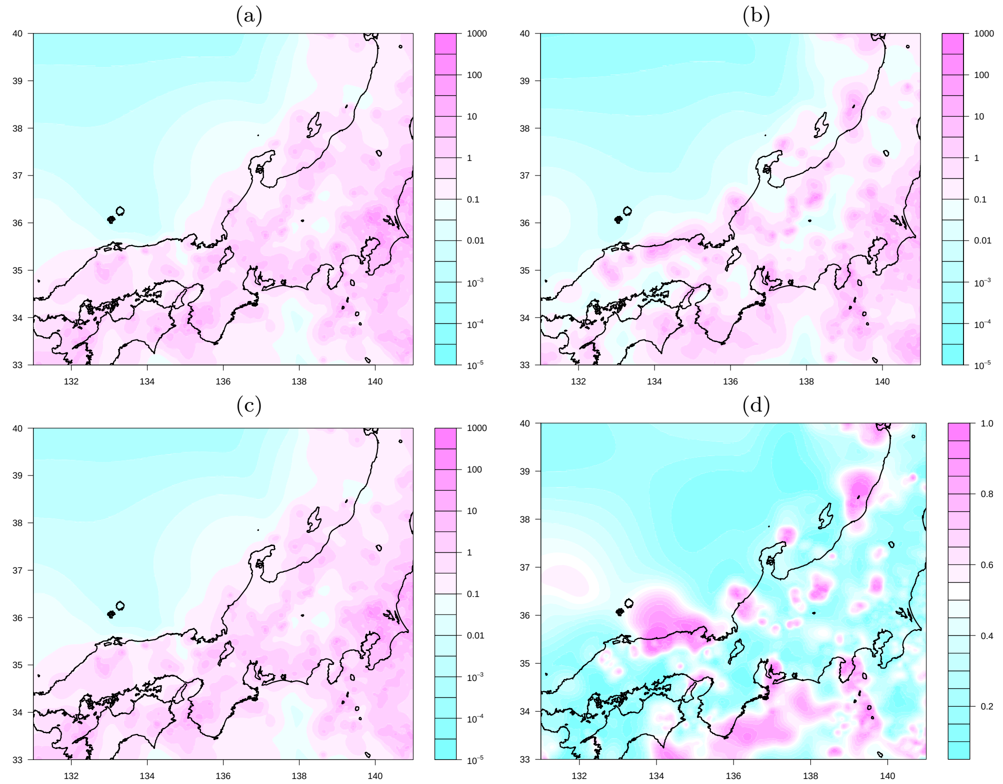

Fig. 1 Estimated seismicity rate of $M_j \geq 4$ earthquakes in central Japan (unit: events/(deg^2· year): (a) total, (b) clustered, and (c) background seismicity. (d) Clustering coefficient (ratio of clustered to total seismicity rate).

4 Stochastic reconstruction

Stochastic reconstruction was introduced to help classify earthquake events in a catalog into different family trees. In testing hypotheses associated with earthquake clustering features, such classification tackles the difficulties caused by the complicated mixture of the background seismicity and different earthquake clusters in both space and time. This algorithm is outlined below.

Algorithm B: Stochastic classification of earthquake clusters [Zhuang et al. (2002), Zhuang et al. (2004)].

B1. Calculate φ_i and ρ_{ji} by using Algorithm A, where $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, i-1$. Here, N is the total number of events.
B2. For each event i, $i = 1, 2, \cdots, N$, generate a random variable U_i uniformly distributed on $[0, 1]$.

B3. For each i, let

$$I_i = \min\{k: \ \varphi_i + \sum_{j=1}^{k} \rho_{ji} \geq U_i \text{ and } 0 \leq k < i\}.$$

If $I_i = 0$, then select i as a background or initial event; else, set the ith event to be a direct offspring of the I_ith event.

We can repeat Algorithm B many times to get different stochastic versions of separations of the earthquake clusters. The non-uniqueness of such realizations illustrates the uncertainty in determining earthquake clusters, and thus repetition can help us to evaluate the significance of some properties of seismicity clustering patterns. However, we can also implement these tests by working with the probabilities φ_j and ρ_{ij} directly. In the coming sections, we will show how to use these probabilities to reconstruct the characteristics associated with earthquake clustering features, using the Japanese JMA catalog as an example. The same reconstruction procedures will be also applied to a simulated catalog for comparison.

Example 2 **Stochastic declustering for the central Japan region** Figure 2 shows a stochastically declustered version of the JMA catalog and illustrates the thinning procedure and the stochastic declustering algorithm, using the same data as in Example 1.

Example 3 **Location distributions of earthquake clusters**
In the early studies of the space-time ETAS model, both scaled Gaussian spatial response function (Model 2, Eq. 15) and scaled power-law spatial response (Model 4, Eq. 17) were adopted by *Ogata* (1998); *Zhuang et al.* (2002, 2004) as the spatial response function. From the viewpoint of hazard mitigation, it is an important question which of the above functions is appropriate for modeling seismicity, i.e., whether the aftershock activity decays quickly over short spatial ranges as in Model 2, or decays in long distances as in Model 4. *Ogata* (1998) found that the second one fit the data better than the first one by using the AIC model selection procedures. Here we obtain the same conclusion from simple histogram techniques.

Define the transformed distance between a triggered event j and its direct ancestor, assumed i, as

$$r_{ij} = \sqrt{\frac{(x_j - x_i)^2 + (y_j - y_i)^2}{D^2 \exp[\alpha(m_i - m_0)]}}. \tag{27}$$

From (15) and (17), r_{ij} has a density function of

$$f_R(r) = 2re^{-r^2}, \ r \geq 0; \tag{28}$$

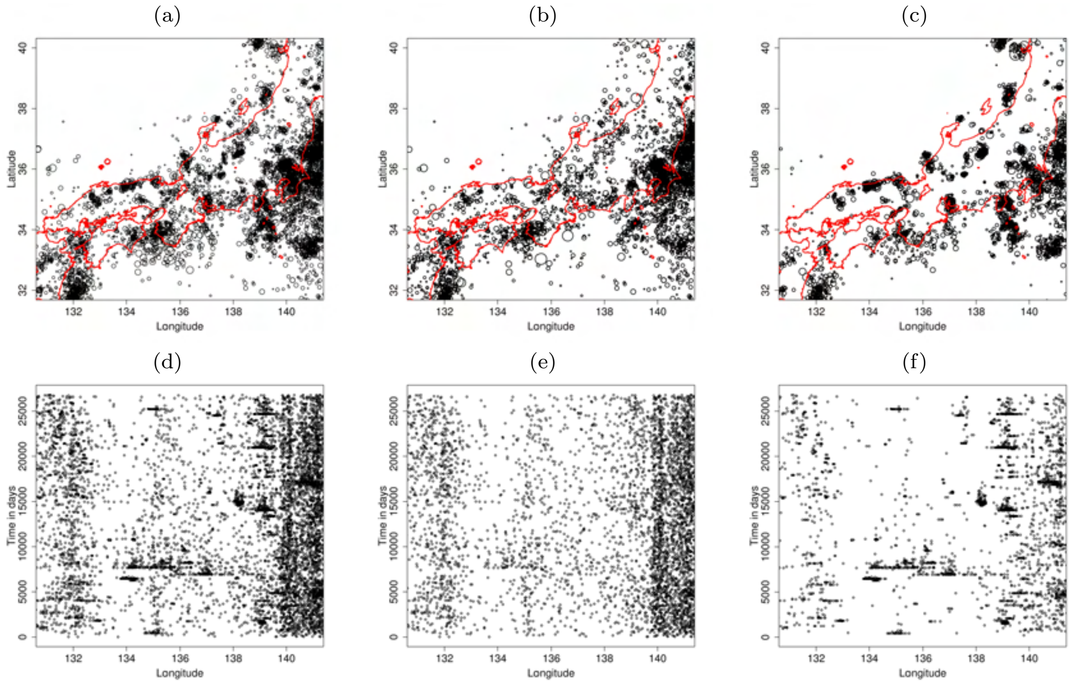

Fig. 2 A realization of stochastic declustering for the central Japan region ($M_J \geq 4$). (a), (b), and (c) are maps of epicenter locations of all the earthquakes, the background events and the clustering events, respectively. (d), (e), and (f) are space-time plots of occurrence times against latitudes for all the earthquakes, the background events and the clustering events, respectively.

and

$$f_R(r) = \frac{2r(q-1)}{(1+r^2)^q}, \quad r \geq 0; \tag{29}$$

for Model 2 and Model 4, respectively. The distribution with a density of (28) is called a Rayleigh distribution. On the other hand, $f_R(r)$ can be reconstructed through

$$\hat{f}_R(r) = \frac{\sum_{i,j} \rho_{ij} H(\Delta r/2 - |r_{ij} - r|)}{\Delta r \sum_{i,j} \rho_{ij}}, \tag{30}$$

where Δr is the bin width and H denotes the Heaviside function. The comparison between $\hat{f}_R$ and f_R for the two models are shown in Figure 3. It can be seen that, when Model 2 is used, the reconstructed probability density of the transformed distances between the ancestors and the direct offspring is quite different from the theoretical one. When Model 4 is used, the reconstructed probability density is very close to the theoretical one. These results confirm that the aftershocks show a long range rather than a short range decay in space (*Ogata* 1998; *Console and*

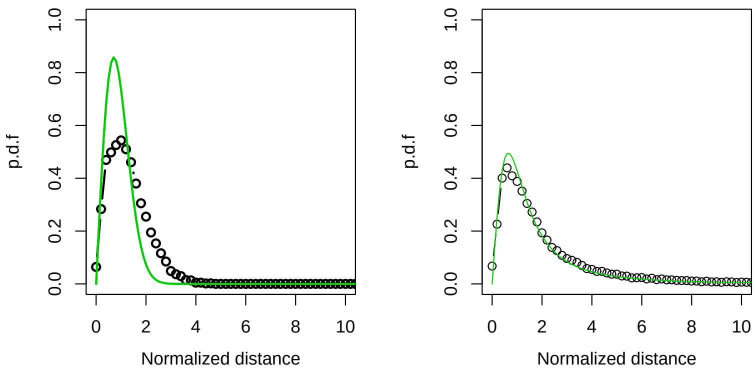

Fig. 3 Reconstruction results for the probability density functions (p.d.f.s) of the standardized triggering distances $\hat{f}_R(r)$ in equation (30) (circles) by using Model 2 (left panel) and Model 4 (right panel). The theoretical curves of $f_R(r)$ in equations (28) and (29) are plotted in solid lines.

Murru 2001; *Console et al.* 2003). These results also imply the robustness of the reconstruction method because one gets a reconstructed probability density function which is very close to the corresponding function in Model 4, even if an improper model like Model 2 is employed.

5 Further reading on the ETAS model

1. *Criticality.* The ETAS model is a branching process, and thus its stability is also related to the concept of criticality. For the ETAS model, the criticality parameter is $\rho = \int \kappa(m)\, s(m)\, \mathrm{d}m$. When $\rho < 1$ the whole process is stable, stationary and ergodic; if $\rho \geq 1$, the population of the process in unit time increases monotonically with time and finally explodes. Detailed discussions can be found in *Helmstetter and Sornette* (2002), *Zhuang* (2003), and *Zhuang and Ogata* (2006).
2. *Relation to foreshocks.* It is an interesting question whether foreshocks can be explained by a clustering model for aftershocks, i.e., whether foreshocks are mainshocks whose aftershocks happen to be bigger. Studies by *Helmstetter and Sornette* (2003a), *Helmstetter et al.* (2003), *Felzer et al.* (2004), *Zhuang and Ogata* (2006), *Zhuang et al.* (2008), *Christophersen and Smith* (2008), *Brodsky* (2011) and *Marzocchi and Zhuang* (2011) suggested that the probability of foreshock phenomena does not exceed the probability of foreshock expected by an ETAS-like generic clustering model.
3. *Relation to Båth's law.* Båth's law (*Båth* 1965) states that the median of the magnitude difference between a mainshock and its largest aftershock is around

1.2. *Utsu* (1957) expressed the median of this difference as a linear function of the mainshock magnitude. *Helmstetter and Sornette* (2003b); *Saichev and Sornette* (2005b); *Zhuang and Ogata* (2006); *Vere-Jones and Zhuang* (2008) showed in different ways that Båth's law can be explained as a double exponential distribution generated by the ETAS model.

4. *HIST-ETAS model.* *Ogata et al.* (2003) and *Ogata* (2004) developed a hierarchical space-time ETAS model to describe the difference of seismicity clustering structures at different locations. He assumed that each parameter in the model is a function of location, which has a smoothness prior. He developed powerful Bayesian tools with penalized likelihoods to estimate the changes of clustering characteristics in the form of spatial variation of model parameters.
5. *Influence of small earthquakes.* In the formulation of the ETAS model, the triggering effect from events smaller than the cut-off magnitude is ignored. The problem was first studied by *Sornette and Werner* (2005a,2005b). Ignoring the triggering effect may influence the conclusion of the study. Here we refer to *Werner* (2007); *Zhuang et al.* (2008); *Wang et al.* (2010) for details.
6. *Impact of anisotropy of earthquake clusters.* Aftershock clusters of larger mainshocks are usually not isotropic around the epicenter of the mainshock but aligned along the mainshock rupture. To account for anisotropy, *Ogata* (1998), *Kagan and Jackson* (1995), *Helmstetter et al.* (2006) and *Werner et al.* (2011) used elliptical aftershock distributions. *Hainzl et al.* (2008) demonstrated by means of ETAS simulations that ignoring this fact can lead to a biased parameter estimation, in particular, to an underestimation of the α-parameter.
7. *Self-similar ETAS models.* *Vere-Jones* (2005) developed a self-similar ETAS model to avoid the problems caused by the cut-off magnitude in the ETAS model and enable fully self-similar features. Because of the missing data problem of immediate aftershocks and small events, however, this model until now remains in theoretical developments (see also, *Saichev and Sornette* 2005a), and has not yet been applied to real seismicity data. Another development to introduce self-similarity is the branching aftershock sequence (BASS) model (e.g., *Turcotte et al.* 2007; *Holliday et al.* 2008).
8. *Earthquake forecasts.* The ETAS model has been used for short-term (one-day) or real-time earthquake probability forecasts. There have been many versions, e.g., *Vere-Jones* (1998); *Helmstetter et al.* (2006); *Werner et al.* (2011); *Zhuang* (2010); *Console et al.* (2006, 2007, 2008, 2010); *Marzocchi and Lombardi* (2009); *Woessner et al.* (2011).

6 The EEPAS model and the double branching model

The EEPAS (Every Earthquake is a Precursor According to Scale) was developed by *Evison and Rhoades* (2004) and *Rhoades and Evison* (2004) in the study of foreshock swarms. Its conditional intensity,

$$\lambda(t,x,y,m) = \mu\lambda_0(t,x,y,m) + \sum_{i:t_i<t} w_i\,\eta(m_i)\,r(m|m_i)\,f(t-t_i)\,g(x-x_i,y-y_i \mid m_i) \tag{31}$$

has a similar form as the ETAS model but is very different in detail. The conditional intensity function is not applied to the whole catalog but only for the events above a second higher threshold, while the summation is taken over all the events above a primary lower threshold. That is to say, this model could be regarded as a self- and mutually exciting process: the target process is of the events above the second higher threshold, and the external process is of the events between the primary and the secondary thresholds. In addition, f here takes the form of the lognormal density, and g and r take the form of normal densities, i.e.,

$$f(u|m_i) = \frac{1}{u\sigma_T\sqrt{2\pi}} \exp\left[-\frac{(\log u - a_T - b_T m_i)^2}{2\sigma_T^2}\right], \tag{32}$$

$$g(x,y|m_i) = \frac{1}{2\pi\sigma_A^2\,10^{m_i\,b_A}} \exp\left[-\frac{x^2+y^2}{2\sigma_A^2\,10^{m_i\,b_A}}\right], \tag{33}$$

$$r(m|m_i) = \frac{1}{\sigma_M\sqrt{2\pi}} \exp\left[-\frac{(m - a_M - b_M m_i)^2}{2\sigma_M^2}\right], \tag{34}$$

where $a_T, b_T, \sigma_T, b_A, \sigma_A, a_M, b_M$, and σ_M are model parameters to be estimated. The weights w_i are usually set to 1, but it has also been suggested to set w_i to the probability that the ith event is a background event, which could be obtained from an initial stochastic declustering using the ETAS model.

This model seems to produce high average probability gains or entropy scores for moderate term forecasts of moderate to large events. Here we refer to *Rhoades and Evison* (2004) for detailed applications.

The double branching model was proposed by *Marzocchi and Lombardi* (2008). In this model, the triggering between earthquakes is divided into 2 steps: one is the short-term clustering due to the elastic response of the upper layers of the Earth, which can be well represented by the ETAS model; and, the other is the interaction among events over longer time-space scales. The conditional intensity of this model takes the form,

$$\lambda(t,x,y) = \mu_2(x,y) + \sum_{i:t_i<t}\left[\frac{K_1\,\mathrm{e}^{\alpha_1(M_i-M_0)}}{(t-t_i+c)^p}\frac{C_1}{(r_i^2+d_1^2)^{q_1}}\right] + \sum_{i:t_i<t}\left[K_2\,\mathrm{e}^{\alpha_2(M_i-M_0)}\mathrm{e}^{-(t-t_i)/\tau}\frac{C_2}{(r_i^2+d_2^2)^{q_2}}\right] \quad (35)$$

where $K_1, K_2, \alpha_1, \alpha_2, c, p, \tau, d_1, d_2, q_1$ and q_2 are model parameters, C_1 and C_2 are normalizing constant such that $C_1\iint(r_i^2+d_1^2)^{-q_1}\,\mathrm{d}x\,\mathrm{d}y = 1$ and $C_2\iint(r_i^2+d_2^2)^{-q_2}\,\mathrm{d}x\,\mathrm{d}y = 1$. *Marzocchi and Lombardi* (2008) used a two-step approach to estimate the model: they firstly carried out Algorithm A to obtain the background probabilities by using the stochastic declustering method, and then fit a self-exciting model with the second summation term in (35) to the catalog of background events. They found that the background seismicity, obtained by removing short-term clustering , is characterized by a second-step long-term clustering with a characteristic time that is compatible with post-seismic relaxation.

7 Summary

This article presents an overview of the spatiotemporal models for seismicity, with an emphasis on some powerful techniques associated with the ETAS model. As can be seen in the article, the most important things are not what the model can describe, but rather the phenomena that the model cannot describe. That is to say, one of the main outcomes of model building is determining aspects of the observed process that the model cannot describe. We introduce the basic techniques of how to decompose the earthquake clusters with uncertainty in the form of probabilities, and how to use these probabilities to find the difference between real data and the model. In summary, even though they are more difficult to implement and to apply to earthquake data, spatiotemporal models are much more powerful in analyzing seismicity than their simple temporal-only counterparts.

Acknowledgements The authors thank two anonymous reviewers for their constructive comments and Jeremy D. Zechar for editing this manuscript.

References

Båth, M. (1965), Lateral inhomogeneities in the upper mantle, *Tectonophysics*, *2*, 483–514. 14

Brodsky, E. E. (2011), The spatial density of foreshocks, *Geophysical Research Letters*, *38*, L10,305, doi:10.1029/2011GL047253. 14

Christophersen, A., and E. G. C. Smith (2008), Foreshock rates from aftershock abundance, *Bulletin of the Seismological Society of Ameerica*, *98*(5), 2133–2148, doi:10.1785/0120060143. 14

Console, R., and M. Murru (2001), A simple and testable model for earthquake clustering, *Journal of Geophysical Research*, *106, B5*, 8699–8711. 13

Console, R., M. Murru, and A. M. Lombardi (2003), Refining earthquake clustering models, *Journal of Geophysical Research*, *108*(B10), 2468, doi:10.1029/2002JB002130. 6, 9, 14

Console, R., D. A. Rhoades, M. Murru, F. F. Evison, E. E. Papadimitriou, and V. Karakostas (2006), Comparative performance of time-invariant, long-range and short-range forecasting models on the earthquake catalogue of Greece, *Journal of Geophysical Research*, *111*, B09,304, doi:10.1029/2005JB0044113. 15

Console, R., M. Murru, F. Catalli, and G. Falcone (2007), Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: comparison with a purely stochastic ETAS model, *Seismological Research Letters*, *78*(1), 49–56. 3, 15

Console, R., M. Murru, and G. Falcone (2008), Real-time earthquake forecasting in Italy, *Tectonophysics*, *470*(3-4), 214–223. 15

Console, R., M. Murru, and G. Falcone (2010), Probability gains of an epidemic-type aftershock sequence model in retrospective forecasting of $m \geq 5$ earthquake in Italy, *Journal of Seismology*, *14*(1), 9–26, doi:10.1007/s10950-009-9161-3. 15

Daley, D. D., and D. Vere-Jones (2003), *An Introduction to Theory of Point Processes – Volume 1: Elementrary Theory and Methods (2nd Edition)*, Springer, New York, NY. 7

Dieterich, J. (1994), A constitutive law for rate of earthquake production and its application to earthquake clustering, *Journal of Geophysical Research*, *99*(B2), 2601–2618. 3

Evison, F. F., and D. A. Rhoades (2004), Demarcation and scaling of long-term seismogenesis, *Pure and Applied Geophysics*, *161*, 21–45, 10.1007/s00024-003-2435-8. 16

Felzer, K. R., R. E. Abercrombie, and G. Ekstrom (2004), A common origin for aftershocks, foreshocks, and multiplets, *Bulletin of the Seismological Society of America*, *94*(1), 88–98, doi:10.1785/0120030069. 14

Hainzl, S., A. Christophersen, and B. Enescu (2008), Impact of earthquake rupture extensions on parameter estimations of point-process models, *Bulletin of the Seismological Society of Ameerica*, *98*(4), 2066–2072, doi:10.1785/0120070256. 15

Helmstetter, A., and D. Sornette (2002), Sub-critical and supercritical regimes in epidemic models of earthquake aftershocks, *Journal of Geophysical Research*, *107*(B10), 2237. 14

Helmstetter, A., and D. Sornette (2003a), Foreshocks explained by cascades of triggered seismicity, *Journal of Geophysical Research*, *108*(B10), 2457, doi:10.1029/2003JB002409. 14

Helmstetter, A., and D. Sornette (2003b), Båth's law derived from the Gutenberg-Richter law and from aftershock properties, *Geophysical Research Letters*, *30*(20), 2069, doi:10.1029/2003GL018186. 15

Helmstetter, A., D. Sornette, and J.-R. Grasso (2003), Mainshocks are aftershocks of conditional foreshocks: How do foreshock statistical properties emerge from aftershock laws?, *Journal of Geophysical Research*, *108*, 2046, doi:10.1029/2002JB001991. 14

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bulletin of the Seismological Society of Ameerica*, *96*(1), 90–106, doi:10.1785/0120050067. 15

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007), High-resolution time-independent forecast for $m \geq 5$ earthquakes in California, *Seismogological Research Letters*, *78*, 59–67. 8

Holliday, J. R., D. L. Turcotte, and J. B. Rundle (2008), Self-similar branching of aftershock sequences, *Physica A: Statistical Mechanics and its Applications*, *387*(4), 933 – 943, doi:DOI:10.1016/j.physa.2007.09.045. 15

Iwata, T. (2010), Effect of secondary aftershocks on aftershock decay in the rate- and state-friction model, *Geophysical Research Abstracts*, *12*(1), EGU2010–7920–2, EGU General Assembly. 3

Kagan, Y. (1991), Likelihood analysis of earthquake catalogues, *Journal of Geophysical Research*, *106, Ser. B7*, 135–148. 4

Kagan, Y. Y., and D. D. Jackson (1995), New seismic gap hypothesis: Five years after, *J. Geophys. Res.*, *100*(B3), 3943–3959. 15

Lewis, P. A. W., and E. Shedler (1979), Simulation of non-nomogeneous Poisson processes by thinning, *Naval Research Logistics Quarterly*, *26*, 403–413. 7

Marzocchi, W., and A. M. Lombardi (2008), A double branching model for earthquake occurrence, *Journal of Geophysical Research*, *113*, B08,317, doi:10.1029/2007JB005472. 3, 16, 17

Marzocchi, W., and A. M. Lombardi (2009), Real-time forecasting following a damaging earthquake, *Geophysical Research Letters*, *36*, L21,302, doi:10.1029/2009GL040233. 15

Marzocchi, W., and J. Zhuang (2011), Statistics between mainshocks and foreshocks in Italy and Southern California, *Geophysical Research Letters*, *38*, L09,310, doi:10.1029/2011GL047165. 14

Musmeci, F., and D. Vere-Jones (1986), A variable-grid algorithm for smoothing clustered data, *Biometrics*, *42*, 483–494. 9

Ogata, Y. (1981), On Lewis' simulation method for point processes, *IEEE Transactions on Information Theory*, *IT-27*(1), 23–31. 7

Ogata, Y. (1988), Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, *83*, 9 – 27. 3

Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, *50*, 379–402. 4, 6, 7, 8, 9, 10, 12, 13, 15

Ogata, Y. (2004), Space-time model for regional seismicity and detection of crustal stress changes, *Journal of Geophysical Research*, *109*(B3), B03,308, doi:10.1029/2003JB002621. 8, 9, 15

Ogata, Y., and J. Zhuang (2006), Space–time ETAS models and an improved extension, *Tectonophysics*, *413*(1-2), 13–23. 7

Ogata, Y., K. Katsura, and M. Tanemura (2003), Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *52*, 499–509(11), doi:doi:10.1111/1467-9876.00420. 15

Rathbun, S. L. (1993), Modeling marked spatio-temporal point patterns, *Bulletin of the International Statistical Institute*, *55, Book 2*, 379–396. 4, 6

Rhoades, D. A., and F. F. Evison (2004), Long-range earthquake forecasting with every earthquake a precursor according to scale, *Pure and Applied Geophysics*, *161*, 47–72, 10.1007/s00024-003-2434-9. 3, 16

Saichev, A., and D. Sornette (2005a), Vere-Jones' self-similar branching model, *Phys. Rev. E*, *72*(5), 056,122, doi: 10.1103/PhysRevE.72.056122. 15

Saichev, A., and D. Sornette (2005b), Distribution of the largest aftershocks in branching models of triggered seismicity: Theory of the universal Båth law, *Physical Review, E*, *71*(5), 056,127, doi:10.1103/PhysRevE.71.056127. 15

Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, Chapman and Hall,, London, New York. 9

Sornette, D., and M. J. Werner (2005a), Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res. Solid Earth*, *110*(B09303), doi:10.1029/2005JB003621. 15

Sornette, D., and M. J. Werner (2005b), Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Båth's law, and observed aftershock sequences, *Journal of Geophysical Research*, *110*, B08,304, doi:10.1029/2004JB003535. 15

Tsukakoshi, Y., and K. Shimazaki (2006), Temporal behavior of the background seismicity rate in central Japan, 1998 to mid-2003, *Tectonophysics*, *417*, 155–168. 8

Turcotte, D., J. Holliday, and J. Rundle (2007), BASS, an alternative to ETAS, *Geophysical Research Letters*, *34*(12), cited By (since 1996) 8. 15

Utsu, T. (1957), Magnitude of earthquakes and occurrence of their aftershocks, *Zisin (J. Seismol. Soc. Jap.)*, *10*, 35–45 (in Japanese). 15

Veen, A., and F. P. Schoenberg (2008), Estimation of space-time branching process models in seismology using an EM-type algorithm, *Journal of the American Statistical Association*, *103*(482), 614–624, doi:10.1198/016214508000000148. 7

Vere-Jones, D. (1998), Probability and information gain for earthquake forecasting, *Computational Seismology*, *30*, 248–263. 15

Vere-Jones, D. (2005), A class of self-similar random measure, *Advances in Applied Probability*, *37*(4), pp. 908–914. 15

Vere-Jones, D., and J. Zhuang (2008), Distribution of the largest event in the critical epidemic-type aftershock-sequence model, *Phys. Rev. E*, *78*(4), 047,102, doi:10.1103/PhysRevE.78.047102. 15

Wang, Q., D. D. Jackson, and J. Zhuang (2010), Missing links in earthquake clustering models, *Geophysical Research Letters*, *37*, L21,307. 7, 15

Werner, M. J. (2007), On the fluctuations of seismicity and uncertainties in earthquake catalogs: Implications and methods for hypothesis testing, Ph.D. thesis, Univ. of Calif., Los Angeles. 15

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011), High-Resolution Long-Term and Short-Term Earthquake Forecasts for California, *Bull. Seism. Soc. Am.*, *101*(4), 1630–1648, doi:10.1785/0120090340. 8, 15

Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, F. Catalli, B. Enescu, M. Cocco, M. C. Gerstenberger, and S. Wiemer (2011), A retrospective comparative forecast test on the 1992 Landers sequence, *Journal of Geophysical Research*, *116*, B05,305, doi:10.1029/2010JB007846. 15

Zhuang, J. (2003), Some applications of point processes in seismicity modelling and prediction, Ph.D. thesis, the Graduate Unversity for Advanced Studies, Kanagawa, Japan. 14

Zhuang, J. (2006), Second-order residual analysis of spatiotemporal point processes and applications in model evaluation, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(4), 635–653, doi: 10.1111/j.1467-9868.2006.00559.x. 7

Zhuang, J. (2010), Next-day earthquake forecasts by using the etas model, *submitted to Earth, Planet, and Space, in press.* 15

Zhuang, J., and Y. Ogata (2006), Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks, *Physical Review, E*, *73*, 046,134, doi:10.1103/PhysRevE.73.046134. 14, 15

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association*, *97*(3), 369–380. 6, 8, 9, 11, 12

Zhuang, J., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using stochastic reconstruction, *Journal of Geophysical Research*, *B5*(3), B05,301, doi:10.1029/2003JB002879. 6, 7, 9, 11, 12

Zhuang, J., C.-P. Chang, Y. Ogata, and Y.-I. Chen (2005), A study on the background and clustering seismicity in the Taiwan region by using a point process model, *Journal of Geophysical Research*, *110*, B05S13, doi: 10.1029/2004JB003157. 7

Zhuang, J., A. Christophosen, M. K. Savage, D. Vere-Jones, Y. Ogata, and D. D. Jackson (2008), Differences between spontaneous and triggered earthquakes: their influences on foreshock probabilities, *Journal of Geophysical Research*, *113*, B11,302, doi:10.1029/2008JB005579. 14, 15