

Research Memorandum No. 941

June 07, 2005

**Multi-dimensional second-order residual
analysis of space-time point processes and its
applications in modelling earthquake data**

Jiancang Zhuang

Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data

Jiancang Zhuang

Institute of Statistical Mathematics, Tokyo, Japan.

Summary. This paper proposes the principle of the second-order martingale-based residual analysis for space-time point processes. In order to illustrate the powers of the first- and second-order residuals, I apply them to the space-time epidemic-type aftershock sequence (ETAS) model for testing some important and interesting hypotheses associated with the features of earthquake clusters.

1. Introduction

Temporal and spatio-temporal point-process models have been more and more widely used in modelling the occurrences of earthquake events (see, e.g., Console et al., 2003; Console and Murru, 2001; Kagan, 1991; Musmeci and Vere-Jones, 1992; Ogata, 1998, 2004; Ogata et al., 2003; Ogata and Zhuang, 2005; Rathbun, 1993; Zheng and Vere-Jones, 1991, 1994; Zhuang et al., 2002, 2004). Among the associated statistical interference techniques, such as model specification, parameterizations, model selection and testing goodness-of-fit, the tools of testing goodness-of-fit are quite underdeveloped. This is one motivation for writing this article.

In previous studies, residual analysis has been carried out by transforming the point process into a standard Poisson process (see, e.g., Ogata, 1988). Schoenberg (2004) uses the thinned residuals to analyze the goodness-of-fit of a multidimensional point process model to an earthquake dataset. Baddeley (2004) proposes the principles of first order residual analysis for temporal and spatial point process according to the martingale properties and the Nguyen-Zessin formula (Nguyen and Zessin, 1979), respectively. In this paper, I will generalize his work to second-order residual analysis through an overview of his work on temporal point processes.

Among the point-process models, the space-time epidemic type aftershock (ETAS) model is used to describe the behavior of earthquake clusters (Ogata, 1998; Ogata et al., 2003, 2004; Zhuang et al., 2002, 2004; Console and Murru, 2001; Console et al., 2003). In this model, the seismicity is classified into two components, the background

and the cluster. Background seismicity is modelled as a stationary Poisson process but not homogeneous in space. Once an event occurs, no matter if it is a background event or generated by another event, it produces its own offspring independently according to some rules. Zhuang et al. (2004) develop a stochastic reconstruction method based on the ETAS model and use it to test a series of interesting hypotheses associated with the features of earthquake clusters, which may not be implied by the model. Their method is purely from intuition without any strict theoretical basis. To provide a theoretical basis for their method is another motivation of writing this article.

In this paper, I first give the space-time analogue of the first-order residual developed by Baddeley (2004), and then propose the second-order residual. To illustrate how to use the above residuals, I use the epidemic type aftershock sequence (ETAS) model (Ogata, 1988, 1998) for earthquake occurrences as an example.

2. Point processes and conditional intensities

Consider a space-time point process consisting of event occurring at time $\{t_i\}$ in the interval $[0, T]$ and at corresponding locations $\{x_i\}$ in a region $\subset \mathbb{R}^d$. Assume that the marginal temporal point process is *orderly*; i.e., the probability that more than one event occurs in the time interval $[t, t + \delta)$ is $o(\delta)$ for all $t \geq 0$. Such space-time point processes can be defined through random counting measures N on $[0, \infty) \times X$. Here, $N(C \times B)$ is the number of events falling in a region $B \in \mathcal{X}$ and at times in a set $C \in \mathcal{T}$, where \mathcal{X} and \mathcal{T} are the Borel σ -algebras of subsets of X and $[0, \infty)$, respectively. Let \mathcal{F} be the σ -algebra generated by $\{N((r, t] \times B) : B \in \mathcal{X}, 0 \leq r \leq t\}$, and let Φ be the collection of counting measures N on $X \times [0, \infty)$. Then a *space-time point process* is a measurable mapping of a probability space (Ω, \mathcal{A}, P) onto (Φ, \mathcal{F}) .

For any measurable $B \in \mathcal{X}$, there exists an \mathcal{F} -compensator $A(t, B)$ such that $N([0, t] \times B) - A(t, B)$ is an \mathcal{F} -martingale. Let ℓ and ℓ^d denote Lebesgue measure in \mathbb{R} and \mathbb{R}^d , respectively. Suppose that, for each $x \in X$, there exists an integrable, non-negative, \mathcal{F} -adopted process $\lambda(t, x)$ such that, with probability 1, for all $t \in \mathbb{R}^+$ and $B \in \mathcal{X}$,

$$\int_B \int_0^t \lambda(s, x) \ell(ds) \ell^d(dx) = A(t, B). \quad (1)$$

The conditions and justification for the existence of such a process can be found, for example, in Vere-Jones and

Schoenberg (2004). If it exists, $\lambda(t, x)$ is called an \mathcal{F} -conditional intensity, satisfying

$$\int_B \lambda(t, x) \ell^d(dx) = \lim_{\Delta t \downarrow 0} \frac{\mathbf{E}[N(t + \Delta t, B) - N(t, B) \mid \mathcal{F}_t]}{\Delta t}. \quad (2)$$

3. First order residuals

According to the martingale property of the conditional intensity, for any predictable process (measurable with respects to the σ -algebra generated by events of the form $E \times (s, t] \times B$ where $E \in \mathcal{F}$ and $B \in \mathcal{X}$) $h(t, x) \geq 0$, a.e.,

$$\mathbf{E} \left[\int_D h(t, x) N(dt \times dx) \right] = \mathbf{E} \left[\int_D h(t, x) \lambda(t, x) \mu(dt \times dx) \right], \quad (3)$$

where μ is the Lebesgue measure $\ell \times \ell^d$, $D \in \mathcal{T} \otimes \mathcal{X}$.

Similar to the residuals for spatial point processes proposed by Baddeley (2004), define residuals with respect to a predictable function $h(t, x) \geq 0$, a.e., and a measurable set $D \subset T \times X$ by

$$R(h, D) = \int_D [h(t, x) N(dt \times dx) - h(t, x) \lambda(t, x) \mu(dt \times dx)]. \quad (4)$$

It is clear that $\mathbf{E}[R(h, D)] = 0$ (see also, Brémaud, 1981, Chap. 2; Karr, 1991, Chap. 5).

EXAMPLE 1. *Raw residuals.* Let $h(t, x) = 1$, then the corresponding residual is

$$R(h, D) = N(D) - \int_D \lambda(t, x) \mu(dt \times dx), \quad (5)$$

which is called the *raw residual*. One important property of raw residuals in applications is that, for a realization of the process, $\{(t_i, x_i) : i \in \mathbf{N}\}$

$$\left\{ \tau_i = \int_0^{t_i} \int_{A \in \mathcal{X}} \lambda(t, x) \ell^d(dx) \ell(dt) \right\} \quad (6)$$

is a standard Poisson process (Meyer, 1971; Papangelou, 1972; Ogata 1988; Vere-Jones and Schoenberg 2004).

EXAMPLE 2. *Reciprocal-lambda residuals* (Baddeley 2004; Schoenberg, 2004) If $h(t, x) = 1/\lambda(t, x)$,

$$R(h, D) = \int_D \frac{N(dt \times dx)}{\lambda(t, x)} - \mu(D). \quad (7)$$

This is an analogue of the Stoyan-Grabarnik (1991) weights for Gibbs point processes. The residual analysis on the space-time ETAS model done by Schoenberg (2004) was essentially using the inverse-lambda residual.

EXAMPLE 3. *Pearson residuals* (Baddeley, 2004). If $h(t, x) = 1/\sqrt{\lambda(t, x)}$,

$$R(h, D) = \int_D N(dt \times dx) / \sqrt{\lambda(t, x)} - \int_D \sqrt{\lambda(t, x)} \mu(dt \times dx). \quad (8)$$

EXAMPLE 4. *Score residuals* (Baddeley, 2004). The score residual is defined by

$$R\left(\frac{\partial \log \lambda(t, x)}{\partial \theta}, D\right) = \int_D \frac{\partial \log \lambda(t, x)}{\partial \theta} N(dt \times dx) - \int_D \frac{\partial \lambda(t, x)}{\partial \theta} \mu(dt \times dx), \quad (9)$$

where θ is any of the usual parameters in the model.

$$\mathbf{E}\left[R\left(\frac{\partial \log \lambda(t, x)}{\partial \theta}, D\right)\right] = 0$$

is the condition for maximizing the log-likelihood function in D , i.e.,

$$\log L = \int_D \log \lambda(t, x) N(dt \times dx) - \int_D \lambda(t, x) \mu(dt \times dx). \quad (10)$$

EXAMPLE 5. *Weighted score residuals and localized maximum likelihood estimates* The equality associated with weighted score residuals

$$\mathbf{E}\left[R\left(w(t, x; t_0, x_0) \frac{\partial \log \lambda(t, x)}{\partial \theta}, D\right)\right] = 0, \quad (11)$$

is the condition for maximizing the local log-likelihood

$$WLL = \int_D w(t, x; t_0, x_0) \log \lambda(t, x) N(dt \times dx) - \int_D w(t, x; t_0, x_0) \lambda(t, x) \mu(dt \times dx), \quad (12)$$

where $w(t, x; t_0, x_0)$ is a kernel function centered at (t_0, x_0) .

4. Second-order residual analysis

A second-order predictable function $H(t, x; t', x')$ is a function predictable on $\mathcal{F}_t \otimes \mathcal{F}_{t'}$, i.e., measurable on the σ -algebra generated by the events of the form of $(E \times (s, t] \times B) \times (E' \times (s', t'] \times B')$, where $E \times E' \in \mathcal{F}_t \otimes \mathcal{F}_{t'}$.

It is obvious that each prediction function $H(t, x; t', x')$ can be decomposed uniquely as $H_- + H_+ + H_0$ such that $H_-(t, x; t', x') = 0$ for $t \geq t'$, $H_+(t, x; t', x') = 0$ for $t \leq t'$ and $H_0(t, x; t', x') = 0$ for $t \neq t'$.

If $H = H_-$, then

$$\begin{aligned} & \mathbf{E}\left[\iint_D H(t, x; t', x') N(dt \times dx) N(dt' \times dx')\right] \\ &= \mathbf{E}\left[\iint_D H(t, x; t', x') \lambda(t, x) \lambda(t', x') \mu(dt' \times dx') \mu(dt \times dx)\right] \end{aligned} \quad (13)$$

where D is measurable subset of $(T \times X) \times (T \times X)$. To obtain this equality, notice that now $H(t, x; t', x') = 0$ for $t \geq t'$, which yields

$$\begin{aligned} I &\equiv \mathbf{E} \left\{ \iint H(t, x; t', x') N(dt \times dx) N(dt' \times dx') \right\} \\ &= \mathbf{E} \left\{ \int \mathbf{E} \left[\int H(t, x; t', x') N(dt' \times dx') \mid \mathcal{F}_t \right] N(dt \times dx) \right\} \end{aligned} \quad (14)$$

Because

$$J(s) \equiv \int_{B \cap \{t' \leq s\}} H(t, x; t', x') [N(dt' \times dx') - \lambda(t', x') \mu(dt' \times dx')] \quad (15)$$

is a martingale of s for any measurable subset $B \subset T \times X$, i.e.,

$$\mathbf{E}[J(s) \mid \mathcal{F}_t] = J(t) = 0 \quad (16)$$

or,

$$\mathbf{E} \left[\int H(t, x; t', x') N(dt' \times dx') \mid \mathcal{F}_t \right] = \mathbf{E} \left[\int H(t, x; t', x') \lambda(t', x') \mu(dt' \times dx') \mid \mathcal{F}_t \right], \quad (17)$$

whose both side are predictable functions with respect to \mathcal{F}_t ,

$$\begin{aligned} I &= \mathbf{E} \left\{ \int \mathbf{E} \left[\int H(t, x; t', x') \lambda(t', x') \mu(dt' \times dx') \mid \mathcal{F}_t \right] N(dt \times dx) \right\} \\ &= \mathbf{E} \left\{ \int \mathbf{E} \left[\int H(t, x; t', x') \lambda(t', x') \mu(dt' \times dx') \mid \mathcal{F}_t \right] \lambda(t, x) \mu(dt \times dx) \right\}. \end{aligned} \quad (18)$$

On the other side,

$$\begin{aligned} &\mathbf{E} \left\{ \iint H(t, x; t', x') \lambda(t', x') \lambda(t, x) \mu(dt' \times dx') \mu(dt \times dx) \right\} \\ &= \mathbf{E} \left\{ \int \mathbf{E} \left[\int H(t, x; t', x') \lambda(t', x') \mu(dt' \times dx') \mid \mathcal{F}_t \right] \lambda(t, x) \mu(dt \times dx) \right\}. \end{aligned} \quad (19)$$

If $H = H_+$, we can get a similar formula as (13). If $H = H_0$, it becomes a predictable function of the first order. Consider the decomposability of H , it is obvious that

$$\begin{aligned} &\mathbf{E} \left[\iint_D H(t, x; t', x') N(dt \times dx) N(dt' \times dx') \right] \\ &= \mathbf{E} \left[\iint_{D \setminus \text{diag}(D) \times \text{diag}(D)} H(t, x; t', x') \lambda(t, x) \lambda(t', x') \mu(dt' \times dx') \mu(dt \times dx) \right] \\ &\quad + \mathbf{E} \left[\iint_{\text{diag}(D)} H(t, x; t, x) \lambda(t, x) \mu(dt \times dx) \right] \end{aligned} \quad (20)$$

where D is measurable subset of $(T \times X) \times (T \times X)$ and $\text{diag}(D) = \{(t, x) : (t, x; t, x) \in D\}$

The second-order residual with respect to the predictable function $H(t, x; t', x')$ for D is

$$\begin{aligned} R(H, D) &= \int_{D \setminus \text{diag}(D) \times \text{diag}(D)} H(t, x; t', x') N(dt' \times dx') N(dt \times dx) \\ &\quad - \iint_{D \setminus \text{diag}(D) \times \text{diag}(D)} H(t, x; t', x') \lambda(t, x) \lambda(t', x') \mu(dt' \times dx') \mu(dt \times dx) \end{aligned} \quad (21)$$

EXAMPLE 6. *Poisson processes* For a temporal Poisson process of a stationary rate λ ,

$$\mathbf{E} \left[\int_{[0, T] \times [0, T] \setminus \{t=t'\}} N(dt) N(dt') \right] = \lambda^2 T^2, \quad (22)$$

which yields an estimate of λ as

$$\hat{\lambda} = \frac{\sqrt{n(n-1)}}{T}, \quad (23)$$

where n is the number of events observed in the time interval $[0, T]$.

To illustrate more on the usages of the above first- and second-order residuals, we use the space-time ETAS model as an example.

5. Space-time ETAS models

The space-time ETAS model has a conditional intensity taking the form of

$$\lambda(t, x, s) = \gamma(s) \left[u(x) + \int_D \xi(t - t', x - x' | s') N(dt' \times dx' \times ds') \right] \quad (24)$$

where (t, x, s) represents the occurrence time, location and size (magnitude) of a point in $\mathbf{R} \times X \times \mathbf{R}^+$, $u(x)$ is background (immigration) rate, $D = (-\infty, t) \times X \times \mathbf{R}^+$,

$$\xi(t, x | s) = \kappa(s) g(t) f(x | s), \quad (25)$$

is the response function, $\kappa(s)$ is the mean number of offspring (aftershocks) produced by a parent of size s , g and f are p.d.f.s for occurrence times and locations of offspring, respectively, and γ is p.d.f of sizes for all the events.

In this model, the magnitude component can regarded as a special spatial coordinate in $M = \mathbf{R}^+$.

Given $u(x) \in L^\infty(X)$, i.e., there exists a positive constant C such that $|u(x)| < C$, a.e., the sufficient and necessary condition for $\mathbf{E}[\lambda(t, x, s)] < \infty$, a.e., is

$$\int_M \kappa(s) \gamma(s) \ell(ds) < 1. \quad (26)$$

Proof: Let

$$K(t, x) = \int_D \xi(t - t', x - x' | s') N(dt' \times dx' \times ds'), \quad (27)$$

then

$$\begin{aligned} \mathbf{E}[K(t, x)] &= \mathbf{E} \left[\int_D \xi(t - t', x - x' | s') \lambda(t', x', s') \mu(dt' \times dx' \times ds') \right] \\ &= \mathbf{E} \left\{ \int_D \xi(t - t', x - x' | s') \gamma(s') [u(x') + K(t', x')] \mu(dt' \times dx' \times ds') \right\}, \end{aligned} \quad (28)$$

where

$$K(t', x') = \int_{D'} \xi(t' - t^*, x' - x^* | s^*) N(dt^* \times dx^* \times ds^*), \quad (29)$$

D' being $(-\infty, t') \times X \times \mathbb{R}^+$. Let T be the operator satisfying

$$(Tv)(t, x) = \int_D \xi(t - t', x - x' | s') \gamma(s') v(t', x') \mu(dt' \times dx' \times ds'). \quad (30)$$

Taking (25) into account, it is easy to prove that T is a linear operator from $L^\infty(X)$ to $L^\infty(X)$ and

$$\|T^n\| = \|T\|^n = \left[\int_{\mathbb{R}^+} \kappa(s) \gamma(s) \ell(ds) \right]^n, \quad n = 1, 2, \dots. \quad (31)$$

Equation (28) can be rewritten as

$$\mathbf{E}[K(t, x)] = (Tu)(t, x) + \mathbf{E}[(TK)(t, x)], \quad (32)$$

Substituting (32) into itself,

$$\mathbf{E}[K(t, x)] = (Tu + T^2u + \dots + T^nu + \dots)(t, x). \quad (33)$$

Moreover,

$$\mathbf{E}[\lambda(t, x, s)] = \gamma(s) [u + Tu + T^2u + \dots + T^nu + \dots](t, x) \quad (34)$$

implies that $\mathbf{E}[\lambda(t, x, s)]$ is essentially bounded iff $\|T\| < 1$ and that $\mathbf{E}[\lambda(t, x, s)] = \gamma(s)m(x)$ is a function independent of t , where $m(x) = (u + Tu + T^2u + \dots + T^nu + \dots)(x)$.

6. Data description, model specification and parameter estimation

Assume that the response functions in (25) take the following forms

$$\kappa(s) = A \exp[\alpha s]; \quad (35)$$

$$\gamma(s) = \beta \exp[-\beta s] H(s); \quad (36)$$

$$f(x|s) = \frac{q-1}{\pi D e^{\alpha s}} \left(1 + \frac{\|x\|^2}{D e^{\alpha s}}\right)^{-q}, \quad (37)$$

and

$$g(t) = \frac{p-1}{c} \left(1 + \frac{t}{c}\right)^{-p} H(t), \quad p > 1, \quad (38)$$

where H represents the Heaviside function (Ogata, 1998).

We use the shallow earthquake (with depths less than 100 km) in the Japanese Meteorological Agency (JMA) catalogue for this analysis (see Figure 1). The time span of this catalogue is 1926/01/01 to 1999/12/31. We select the data in the polygon with vertices $(134.0^\circ E, 31.9^\circ N)$, $(137.9^\circ E, 33.0^\circ N)$, $(143.1^\circ E, 33.2^\circ N)$, $(144.9^\circ E, 35.2^\circ N)$, $(147.8^\circ E, 41.3^\circ N)$, $(137.8^\circ E, 44.2^\circ N)$, $(137.4^\circ E, 40.2^\circ N)$, $(135.1^\circ E, 38.0^\circ N)$ and $(130.6^\circ E, 35.4^\circ N)$. The time period after 10000 days as the target region for estimating the parameters by using MLE.

The parameters and the background of the given ETAS model can be iteratively estimated from the following procedures (Zhuang et al., 2002, 2004).

(a) Set up the initial background seismicity rate, for example, let $\hat{u}(x) = 1$.

(b) Set

$$\lambda(t, x, y) = \gamma(s) \left[C \hat{u}(x) + \sum_{i:t_i < t} \kappa(s_i) g(t - t_i) f(x - x_i | s_i) \right], \quad (39)$$

and estimate C and the other model parameters by maximizing the log-likelihood function

$$\log L = \sum_{(t_i, x_i, s_i) \in D} \log \lambda(t_i, x_i, s_i) - \iiint_D \lambda(t, x, s) dt dx ds, \quad (40)$$

where $D \in \mathcal{T} \otimes \mathcal{X}$ is a specified region of interests.

(c) For each event i , set

$$\varphi_i = C \hat{u}(x_i) \gamma(s_i) / \lambda(t_i, x_i, s_i). \quad (41)$$

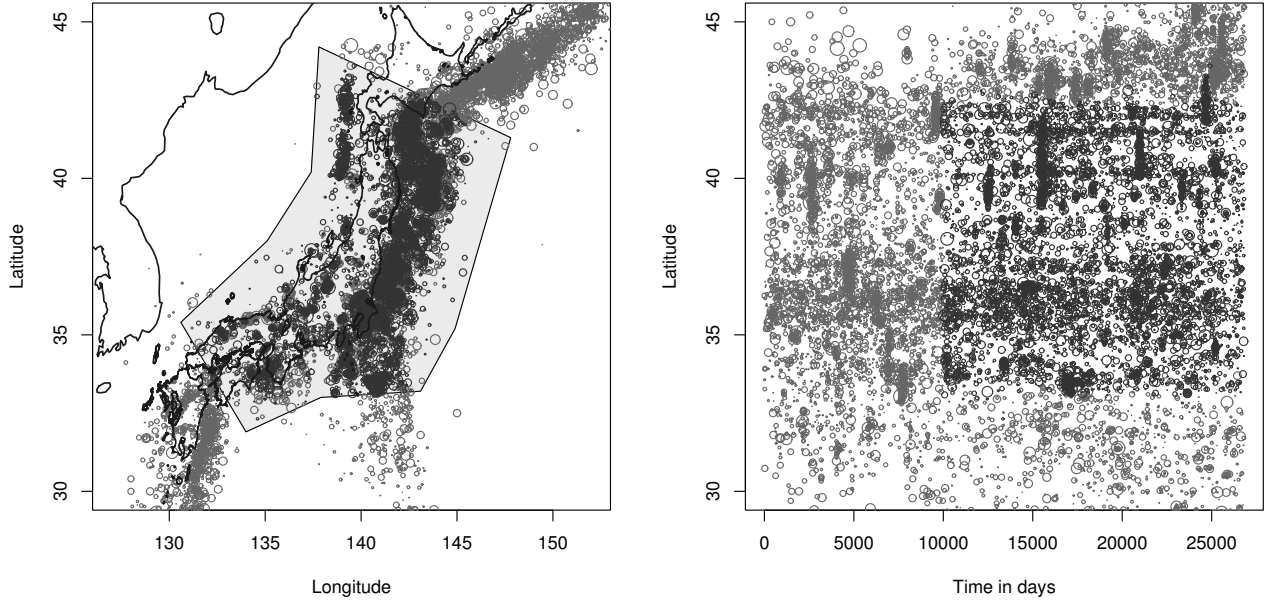


Fig. 1. Occurrence locations and times of shallow earthquakes in the central Japan area. (a) Epicentral locations. The polygon represents the target region. (b) Epicentral latitudes against occurrence times. Black and gray circles are target events and complementary events, respectively. Different sizes of circles show the magnitude of earthquake from 4.2 to 8.1. The polygon represents the target region.

(d) Get a better estimate of the background rate by using the weighted kernel estimates

$$\hat{u}(x) \propto \sum_i \varphi_i Z(x - x_i, h_i) \quad (42)$$

where Z represent the gaussian kernel function, the bandwidth h_i is the distance to the n_p th closest events to i or is a threshold bandwidth if that distance is less than the threshold bandwidth.

(e) Replace the background rate by this better one, and return to Step 2 until the results converge.

7. Residuals for the ETAS model and data analysis

Once the conditional intensity function in (24) is estimated, it provides us a good way to evaluate the probability how an event is likely to be a background event or triggered by others (Kagan and Knoppoff, 1980; Zhuang et al 2002). Consider the contribution of the background seismicity rate relative to total seismicity rate at the occurrence of the i th event,

$$\varphi_i = \frac{u(x_i) \gamma(s_i)}{\lambda(t_i, x_i, s_i)}, \quad (43)$$

If we remove the i th event with probability $1 - \varphi_i$ for all the events in the process, we can realize a process with the occurrence rate of $\mu(x)\gamma(s)$ (see, e.g., Ogata, 1981; Karr, 1991, Chap. 5 for justification). Thus it is natural to

regard φ_i as the probability that the i th event is a background events. Similarly,

$$\rho_{ij} = \frac{\gamma(s_j) \xi(t_j - t_i, x_j - x_i | s_i)}{\lambda(t_j, x_j, s_j)} \quad (44)$$

can be regarded as the probability that the j th event is directly produced by the i th event. Based the above ideas, Zhuang et al. (2004) tested some hypothesis associated with the clustering features of earthquake by building empirical functions with the events weighted by probabilities φ_i or ρ_{ij} . This reconstruction method was purely from intuition. In the coming sections, we show how it works.

For simplification, in the following sections, the notations of the Riemann integral are used, and, without confusion, the integral region is indicated by an index function in the integrand.

7.1. Testing stationarity of background process

Take $h(t, x, s) = u(x)\gamma(s)/\lambda(t, x, s)$, then $\varphi_i = h(t_i, x_i, s_i)$ is the background probability. Define background residuals by

$$R(h, B) = \sum_i \varphi_i I[(t_i, x_i, s_i) \in B] - \iiint_B u(x) \gamma_1(s) dt dx ds, \quad (45)$$

where B is a measurable subset of $T \times X \times M$. To test the stationarity of the background process, choose $B = (0, t) \times B_x \times B_s$, $B_x \subset X$ and $B_s \subset S$. From $\mathbf{E}[R(h, B)] = 0$ (45),

$$\sum_i \varphi_i I[(t_i, x_i, s_i) \in B] \approx t \times \text{constant}. \quad (46)$$

By using the above equation, Zhuang et al. (2005) find that background seismicity become quiescent before the 1999/9/21 Chi-Chi earthquake (MS7.3) in Taiwan. Hainzl and Ogata (2005) show that, in the analysis of the earthquake swarm in 2000 in Vogtland/NW Bohemia, centra Europe, the patterns of changes in the background residuals can be used to identify the pore pressure changes due to fluid intrusion.

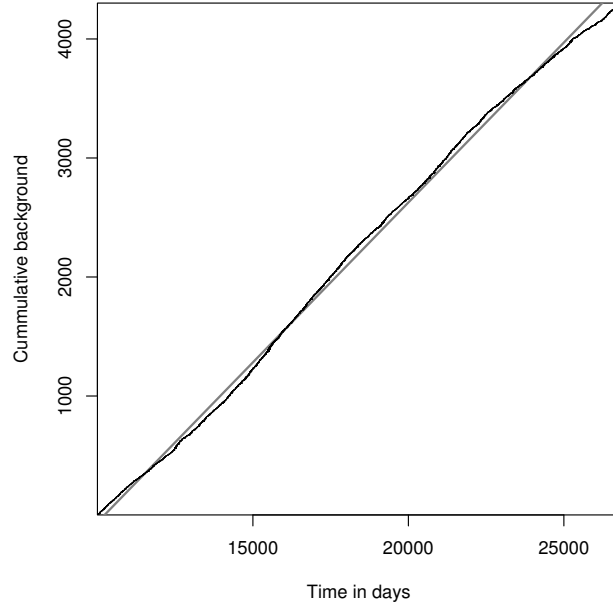


Fig. 2. Cumulative background weights against times in the target region.

7.2. Reconstructing individual functions

Reconstructing the magnitudal distribution Set $h(t, x) = I(s \in S)$ where $S = [s_0 - \delta, s_0 + \delta]$, δ being a small number. Then

$$\begin{aligned}
 \mathbf{E} \left[\sum_i I(s_i \in S) \right] &= \mathbf{E} \iiint I(s \in S) \lambda(t, x, s) dt dx ds \\
 &= \int \gamma(s) I(s \in S) ds \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right] \\
 &\approx 2\gamma(s_0) \delta \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right],
 \end{aligned} \tag{47}$$

where $\lambda(t, x) = u(x) + \sum_{i: t_i < t} \xi(t - t_i, x - x_i | s_i)$. It gives an estimate of $\gamma(s)$,

$$\hat{\gamma}(s) = \frac{\sum_i I(s_i \in [s - \delta, s + \delta])}{2\delta \sum_i 1}. \tag{48}$$

The denominator in the above equation is used for normalization.

Reconstructing the function of triggering ability Let

$$H(t, x, s; t', x', s') = \frac{\kappa(s) g(t' - t) f(x' - x | s) \gamma(s')}{\lambda(t', x', s')}. \tag{49}$$

Then

$$\begin{aligned}
& \mathbf{E} \sum_{i,j} I(s_i \in S) H(t_i, x_i, s_i; t_j, x_j, s_j) \\
&= \mathbf{E} \iiint \iiint I(s \in S) \kappa(s) g(t' - t) f(x' - x | s) \gamma(s') \lambda(t, x, s) dt' dx' ds' dt dx ds \\
&= \int \kappa(s) \gamma(s) I(s \in S) ds \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right] \\
&\approx 2 \delta \kappa(s_0) \gamma(s_0) \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right].
\end{aligned} \tag{50}$$

Consider $\rho_{ij} = H(t_i, x_i, s_i; t_j, x_j, s_j)$, we have the following estimator

$$\hat{\kappa}(s) = \frac{\sum_{i,j} \rho_{ij} I(s_i \in S)}{\sum_i I(s_j \in S)}. \tag{51}$$

Reconstructing distributions for the occurrence times and locations of direct offspring Similarly,

$$\begin{aligned}
& \mathbf{E} \left[\sum_{i,j} I(t_j - t_i \in T) H(t_i, x_i, s_i; t_j, x_j, s_j) \right] \\
&\approx 2 \delta g(u) \int \kappa(s) \gamma(s) ds \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right]
\end{aligned} \tag{52}$$

where $T = [u - \delta, u + \delta]$, corresponds to the estimate

$$\hat{g}(u) = \frac{\sum_{i,j} \rho_{ij} I(|t_j - t_i - u| < \delta)}{2 \delta \sum_{i,j} \rho_{ij}}; \tag{53}$$

and

$$\begin{aligned}
& \mathbf{E} \left[\sum_{i,j} I(x_j - x_i \in X) I(s_i \in S) H(t_i, x_i, s_i; t_j, x_j, s_j) \right] \\
&\approx 2 \mu(X) \delta f(u | s_0) \kappa(s_0) \gamma(s_0) \times \mathbf{E} \left[\iint \lambda(t, x) dt dx \right]
\end{aligned} \tag{54}$$

where X is a small volume around u and $S = [s_0 - \delta, s_0 + \delta]$, corresponds to

$$\hat{f}(u | s) = \frac{\sum_{i,j} \rho_{ij} I(x_j - x_i \in X) I(|s_i - s| < \delta)}{2 \mu(X) \delta \sum_{i,j} \rho_{ij} I(|s_i - s| < \delta)}. \tag{55}$$

EXAMPLE 7. *Reconstructing individual functions for the JMA data* We apply the above reconstruction procedures to the JMA catalogue and the reconstructed $\gamma(s)$, $\kappa(s)$ and $g(t)$ are shown in Figure 7.3. Basically, the functions that we have chosen in the model formulation are good enough to be used as a first approximation for describing the clustering features in seismicity.

7.3. Testing difference between background events and triggered events

The ETAS model assumes that there is no distinction between background events and triggered events. Once an event occurs, its magnitude is from the common unique magnitudal distribution and it triggers its offspring in the same manner as others. In this section, we will test whether background events and triggered events are different in some aspects.

Magnitude difference To test whether background seismicity and triggered seismicity have different magnitudal distribution is to test the model as in (24) against a more complicated model with a conditional intensity function,

$$\lambda_1(t, x, s) = u(x) \gamma_0(s) + \gamma_1(s) \sum_{t_i < t} \kappa(s_i) g(t - t_i) f(x - x_i | s_i). \quad (56)$$

Then, for the new model,

$$\begin{aligned} \mathbf{E} \left[\sum_i \frac{u(x_i) \gamma_0(s_i) I(s_i \in S)}{\lambda_1(t_i, x_i, s_i)} \right] &= \int \gamma_0(s) I(s \in S) ds \iint u(x) dt dx \\ &\approx 2 \delta \gamma_0(s_0) \iint u(x) dt dx \end{aligned} \quad (57)$$

and

$$\begin{aligned} &\mathbf{E} \left[\sum_{i,j} I(s_j \in S) \frac{\gamma_1(s_j) \kappa(s_i) g(t_j - t_i) f(x_j - x_i | s_i)}{\lambda_1(t_j, x_j, s_j)} \right] \\ &= \int \gamma_1(s') I(s' \in S) ds' \times \mathbf{E} \left[\iiint \lambda_i(t, x, s) dt dx ds \right] \\ &\approx 2 \delta \gamma_1(s_0) \times \mathbf{E} \left[\iiint \lambda_1(t, x, s) dt dx ds \right]. \end{aligned} \quad (58)$$

Take

$$\varphi_i \approx \frac{u(x_i) \gamma_0(s_i)}{\lambda_1(t_i, x_i, s_i)} \quad (59)$$

and

$$\rho_{ij} \approx \frac{\gamma_1(s_j) \kappa(s_i) g(t_j - t_i) f(x_j - x_i | s_j)}{\lambda_1(t_j, x_j, s_j)}, \quad (60)$$

$\gamma_0(s)$ and $\gamma_1(s)$ can be reconstructed by

$$\hat{\gamma}_0(s) = \frac{\sum_i \varphi_i I(|s_i - s| < \delta)}{2 \delta \sum_i \varphi_i} \quad (61)$$

and

$$\hat{\gamma}_0(s) = \frac{\sum_{i,j} \rho_{ij} I(|s_j - s| < \delta)}{2\delta \sum_{i,j} \rho_{ij}} = \frac{\sum_i (1 - \varphi_i) I(|s_i - s| < \delta)}{2\delta \sum_i (1 - \varphi_i)}, \quad (62)$$

respectively.

Triggering abilities This is to test the model against a more complicated model with a conditional intensity in the form of

$$\lambda_1(t, x, s) = \lambda_1(t, x, s, \omega) I(\omega = 0) + \lambda_1(t, x, s, \omega) I(\omega = 1), \quad (63)$$

where

$$\lambda_1(t, x, s, \omega) = \begin{cases} u(x) \gamma_0(s), & \text{if } \omega = 0, \\ \gamma_1(s) \sum_{t_i < t} \xi(t, x; t_i, x_i, s_i, \omega_i), & \text{if } \omega = 1. \end{cases} \quad (64)$$

and

$$\xi(t, x; t_i, x_i, s_i, \omega_i) = \begin{cases} \kappa_0(s_i) g_0(t - t_i) f_0(x - x_i | s_i), & \text{if } \omega_i = 0; \\ \kappa_1(s_i) g_1(t - t_i) f_1(x - x_i | s_i), & \text{if } \omega_i = 1. \end{cases} \quad (65)$$

Let

$$\begin{aligned} & H(t, x, s; t', x', s') \\ &= \frac{\gamma_1(s') \xi(t', x'; t, x, s, \omega) \lambda(t, x, s, \omega) I(\omega = 0)}{\lambda_1(t', x', s') \lambda_1(t, x, s)} \\ &= \frac{\gamma_1(s') \kappa_0(s) g_0(t' - t) f_0(x' - x | s) u(x) \gamma_0(s)}{\lambda_1(t', x', s') \lambda_1(t, x, s)}. \end{aligned} \quad (66)$$

Then

$$\begin{aligned} & \mathbf{E} \left[\sum_{i,j} H(t_i, x_i, s_i; t_j, x_j, s_j) I(s_i \in S) \right] \\ &= \int \kappa_0(s) \gamma_0(s) ds \times \iint u(x) dt dx \\ &\approx \delta \kappa_0(s_0) \gamma_0(s_0) \times \iint u(x) dt dx \end{aligned} \quad (67)$$

Use the approximations

$$\varphi_i \approx \frac{u(x_i) \gamma_0(s_i)}{\lambda_1(t_i, x_i, s_i)} \quad (68)$$

and

$$\rho_{ij} \approx \frac{\gamma_1(s_j) \kappa_0(s_i) g_0(t_j - t_i) f_0(x_j - x_i | s_i)}{\lambda(t_j, x_j, s_j)}, \quad (69)$$

we get

$$H(t_i, x_i, s_i; t_j, x_j, s_j) \approx \varphi_i \rho_{ij}. \quad (70)$$

From (67),

$$\hat{\kappa}_0(s) = \frac{\sum_{i,j} \varphi_i \rho_{ij} I(s_i \in S)}{\sum_i \varphi_i I(s_i \in S)}, \quad (71)$$

where S is a neighborhood of s with a reasonable volume.

Similarly, let

$$\begin{aligned} H(t, x, s; t', x', s') &= \frac{\gamma_1(s) \xi(t', x'; t, x, s, \omega) \lambda_1(t, x, s, \omega) I(\omega = 1)}{\lambda_1(t', x', s') \lambda_1(t, x, s)}, \\ &= \frac{\gamma_1(s) \xi(t', x'; t, x, s, \omega)}{\lambda_1(t', x', s')} \left[1 - \frac{\lambda_1(t, x, s, \omega) I(\omega = 0)}{\lambda_1(t, x, s)} \right], \end{aligned} \quad (72)$$

then

$$\begin{aligned} &\mathbf{E} \left[\sum_{i,j} H(t_i, x_i, s_i; t_j, x_j, s_j) I(s_i \in S) \right] \\ &= \int \kappa_1(s) \gamma_1(s) I(s \in S) ds \times \left\{ \mathbf{E} \left[\iint \lambda_1(t, x, s) dt dx ds \right] - \iint u(x) dt dx \right\} \\ &\approx \delta \kappa_1(s_0) \gamma_1(s_0) \left\{ \mathbf{E} \left[\iint \lambda_1(t, x, s) dt dx ds \right] - \iint u(x) dt dx \right\}. \end{aligned} \quad (73)$$

Similar to (71), the triggering abilities of non-background events can be reconstructed through

$$\hat{\kappa}_1(s) = \frac{\sum_{i,j} (1 - \varphi_i) \rho_{ij} I(s_i \in S)}{\sum_i (1 - \varphi_i) I(s_i \in S)}, \quad (74)$$

where S is again a neighborhood of s with a reasonable volume.

The other functions, g_1 , g_2 , f_1 and f_2 , can be reconstructed in similar ways.

7.4. Testing dependence between direct offspring events and parent events

Magnitude dependence To test whether the magnitudes of the offspring is dependent on their direct ancestor's magnitude is equivalent to the problem of testing the original model against a more complicated model with a condition intensity in the form of

$$\lambda_1(t, x, s) = u(x) \gamma_0(s) + \sum_{i: t_i < t} \kappa(s_i) g(t - t_i) f(x - x_i | s_i) \gamma_1(s | s_i). \quad (75)$$

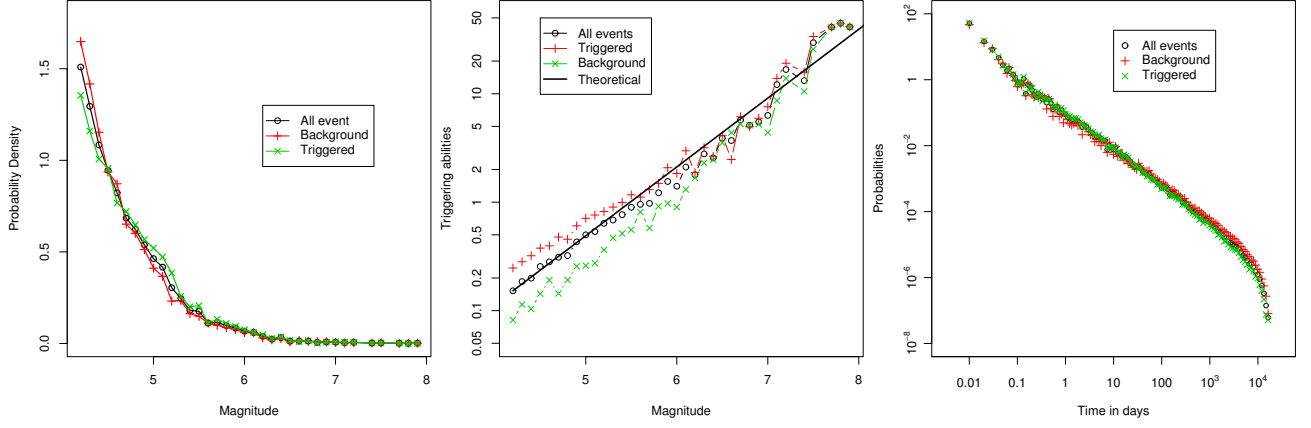


Fig. 3. (a) Reconstructed p.d.f of the magnitude distributions for all the events, background events and triggered events, i.e., as in (48), (61) and (62), respectively; (b) Reconstructed triggering abilities (average number of offspring can be triggered by the events of same magnitude) for all the events, background events and triggered events, as in (51), (71) and (74), respectively; (c) Reconstructed p.d.f of the occurrence times of offspring relatively to the occurrence time of their direct ancestor. In (a) and (b), magnitude 4.2 on the horizontal axes corresponds to $s = 0$.

In the new model, set

$$H(t, x, s; t', x', s') = \frac{\kappa(s) g(t' - t) f(x' - x | s) \gamma_1(s' | s)}{\lambda_1(t', x', s')}, \quad (76)$$

then

$$\begin{aligned} & \mathbf{E} \left[\sum_{i,j} H(t_i, x_i, s_i; t_j, x_j, s_j) I(s_i \in S) I(s_j \in S') \right] \\ &= \mathbf{E} \left[\iiint \iiint \kappa(s) g(t' - t) f(x' - x | s) \gamma_1(s' | s) I(s \in S) I(s' \in S') \lambda_1(t, x, s) dt' dx' ds' dt dx ds \right] \\ &\approx \delta \delta' \gamma_1(s'_0 | s_0) \kappa(s_0) \mathbf{E}[\lambda_1(t, x, s_0)] \end{aligned} \quad (77)$$

where S and S' are neighborhoods of s_0 and s'_0 of volumes δ and δ' , respectively.

Consider $\rho_{ij} \approx H(t_i, x_i, s_i; t_j, x_j, s_j)$ in (77), the function $\gamma_1(s' | s)$ can approximately reconstructed by

$$\hat{\gamma}_1(s' | s) = \frac{\sum_{i,j} \rho_{ij} I(s_i \in S) I(s_j \in S')}{\sum_{i,j} \rho_{ij} I(s_i \in S)}. \quad (78)$$

EXAMPLE 8. *Testing dependence of the magnitudes between the ancestors and the direct offspring in the JMA catalogue* As shown in Figure 7.4, the magnitudal distributions of the offspring are different when the ancestor magnitude changes. This may indicate the existence of such dependence (Zhuang et al., 2004), or, be caused by the low detection ability of monitoring network to the shocks immediately after the mainshocks (Kagan, 2004).

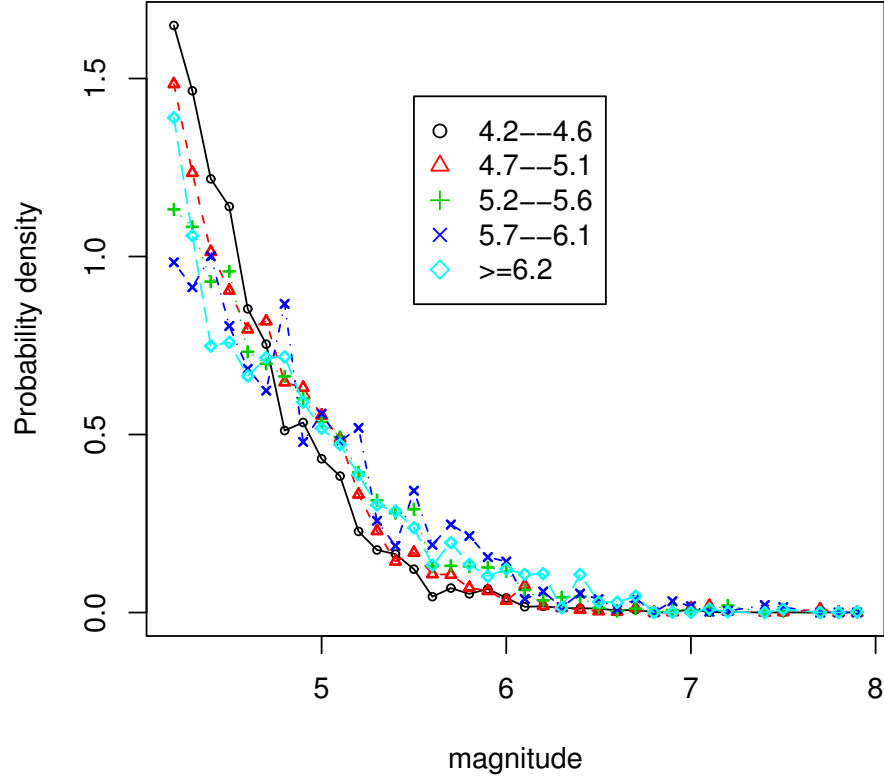


Fig. 4. Magnitude dependence: reconstructed p.d.f of magnitudes of the offspring directly triggered by ancestors of different magnitude classes. Magnitude 4.2 corresponds to $s = 0$.

Spatial scaling factor and triggering abilities Immediate questions to (38) are: (1) Is the scaling factor $De^{\alpha s}$ necessary, i.e., can it be replaced by a constant D_0 ? (2) Should the scaling factor $De^{\alpha s}$ have the same exponent α as in the triggering abilities of (35)? (3) Is the scaling factor an exponential law?

Suppose that the p.d.f for the locations of offspring should be expressed as $f(x | s) = f(x, \sigma(s))$, where $\sigma(s)$ is a continuous function of s . Then for each s , $\sigma(s)$ can be estimate through maximizing the following pseudo-log-likelihood

$$PLL(\sigma(s)) = \sum_{i,j} \rho_{ij} \log f(x_j - x_i, \sigma(s_i)) I(s_i \in [s - \delta, s + \delta]) \quad (79)$$

where δ is a small positive number. Maximizing $PLL(\sigma(s))$ is equivalent to solving

$$\sum_{i,j} \rho_{ij} I(s_i \in S) \left. \frac{\partial}{\partial \sigma(s)} \log f(x_j - x_i; \sigma(s)) \right|_{s=s_i} = 0. \quad (80)$$

To show the validity of (80), let

$$H(t, x, s; t', x', s') = \frac{\gamma(s') \kappa(s) g(t' - t) f(x' - x, \sigma(s)) \frac{\partial}{\partial \sigma(s)} \log f(x' - x, \sigma(s))}{\lambda(t', x', s')}, \quad (81)$$

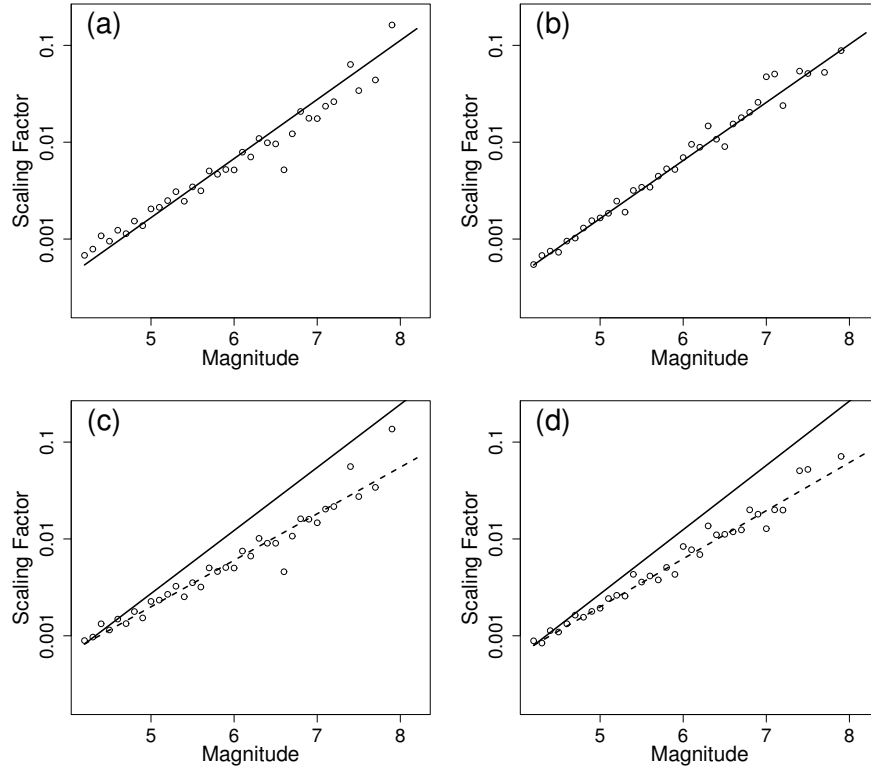


Fig. 5. Reconstructed results of $\sigma(s)$ against the corresponding magnitudes (Magnitude 4.2 corresponds to $s = 0$) of parent earthquakes. Circles indicate the values of $\hat{\sigma}(s)$. (a) shows the reconstructed results from the original model for the JMA catalogue. (b) is the same as (a), but the catalogue is from simulation of the original model. (c) Reconstructed results from the model equipped with (84). (d) is the same as (c), but the catalogue is from simulation of the model equipped with (84). In (c) and (d), the solid lines represent $De^{\alpha s}$ and the dashed lines represent $De^{\alpha' s}$.

and then

$$\begin{aligned}
& \mathbf{E} \left[\sum_{i,j} H(t_i, x_i, s_i; t_j, x_j, s_j) I(s_i \in S) \right] \\
&= \mathbf{E} \left[\iiint \iiint \gamma(s') \kappa(s) g(t' - t) \frac{\partial f(x' - x; \sigma(s))}{\partial \sigma(s)} I(s \in S) \lambda(t, x, s) dt' dx' ds' dt dx ds \right] \\
&= \int \kappa(s) I(s \in S) \gamma(s) ds \times \mathbf{E}[\lambda(t, x)] \times \frac{\partial}{\partial \sigma(s)} \int f(x' - x, \sigma(s)) dx' \\
&= 0
\end{aligned} \tag{82}$$

because $\int f(x' - x, \sigma(s)) dx' = 1$. Equation (80) can be obtained by using the approximation

$$\rho_{ij} \approx \frac{\gamma(s_j) \kappa(s_i) g(t_j - t_i) f(x_j - x_i, \sigma(s_i))}{\lambda(t_j, x_j, s_j)}. \tag{83}$$

EXAMPLE 9. *The scaling law for the locations of clusters in the JMA catalogue* Applying the above procedures

to the JMA catalogue by using the original model and a model equipped with

$$f(x|s) = \frac{q-1}{\pi D e^{\alpha' s}} \exp\left(1 + \frac{\|x\|^2}{D e^{\alpha' s}}\right)^{-q}, \quad (84)$$

the results are shown in Figure 7.4. For comparison the same procedures are also applied to catalogues simulated from the same models and parameters as in the fitting results. It can be seen that the scaling law for the spatial locations of offspring is still an exponential law, but not the same as the one for the triggering abilities. This conclusion results in a new revision of the formulation of the space-time ETAS model in the form of (84) in practice (Ogata and Zhuang, 2005; Zhuang et al., 2005).

8. Conclusions

I propose the principles of the second-order residual analysis for multi-dimension space-time point processes. The uses of these residuals are illustrated by examples, especially the space-time ETAS model with non-homogeneous background rate, which has been used to show the powers of residual analysis in testing a number of hypotheses of general interests associated with seismicity clustering patterns.

Acknowledgement

The author thank Prof. Yosihiko Ogata and Prof. David Vere-Jones for introducing me to the interesting research field of point processes. Prof. Adrian Baddeley has kindly sent me his manuscript on residual analysis for spatial point processes. Discussions with Prof. Daryl Daley are very helpful. This work is carried out when the author is supported by a post-doctoral programme funded by the Japan Society for Promotion of Science.

References

- Baddeley A., Turner R., Møller J. and Hazelton M. (2004) Residual analysis for spatial point processes. School of Mathematics and Statistics, University of Western Australia, Research Report. 2004/08.
- Brémaud P. (1981), *Point Processes and Queues: Martingale Dynamics*, New York: Springer-Verlag.
- Console, R., and M. Murru (2001), A simple and testable model for earthquake clustering, *J. Geophys. Res.*, 106, 8699-8711.

- Console, R., M. Murru, and A. M. Lombardi (2003), Refining earthquake clustering models. *J. Geophys. Res.*, *108*(B10), 2468, doi:10.1029/2002JB002130.
- Daley, D. and D. Vere-Jones (2003), *An Introduction to Theory of Point Processes*. Springer, New York.
- Hainzl, S. and Y. Ogata (2005), Detecting fluid signals in seismicity data through statistical earthquake modeling. *Journal of Geophysical Research*, *110*, B5, B05S07, doi:10.1029/2004JB003247.
- Kagan, Y. (1991), Likelihood analysis of earthquake catalogues. *Geophysical Journal International*, *106*, B7, 135-148.
- Kagan Y. Y. (2004), Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of Seismological Society of America*, *94*, 1207-1228.
- Kagan, Y. Y., and L. Knopoff, (1981). Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research*, *86*, 2853-2862.
- Karr, A.F. (1991). *Point processes and their Statistical Inference*, second edition. New York: Dekker.
- Meyer P. (1971) Demonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités, V*, University of Strasbourg, Lecture Notes in Math, *191*, 191-195.
- Musmeci, F. and D. Vere-Jones (1992). A space-time clustering model for historical earthquakes. *Ann. Inst. Stat. Math.*, *44*, 1-11.
- Nguyen X. X. and Zessin H. (1979). Integral and Differential Characterizations of the Gibbs process. *Mathematische Nachrichten*. *88*: 105-115.
- Ogata Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, *IT-27* (1): 23-31.
- Ogata, Y. (1988), Statistical model for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.*, *83*, 9-27.

- Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, 50, 379-402.
- Ogata, Y. (2004), Space-time model for regional seismicity and detection of crustal stress changes, *J. Geophys. Res.*, 109, B03308, doi:10.1029/2003JB002621.
- Ogata, Y., K. Katsura and M. Tanemura (2003). Modelling heterogeneous space-time occurrences of earthquake and its residual analysis, *Appl. Stat. (J. Roy. Statist. Soc, Ser. C)*, 52, 499-509.
- Ogata, Y. and J. Zhuang (2005), Space-time ETAS models and an improved extension, submitted to *Tectonophysics, special issue on Critical Point Theory and Space-Time Pattern Formation in Precursory Seismicity*.
- Papangelou F. (1972). Integrability of expected increments point processes and a related random change of scale. *Trans. Mer. Math. Soc.*, 165, 483-506.
- Rathbun, S. L. (1993), Modelling marked spatio-temporal point patterns. *Bull. Int. Stat. Inst.*, 55, Book 2, 379-396.
- Schoenberg F. P. (2004). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98, 789-795, doi: 10.1198/016214503000000710.
- Stoyan D. and Grabarnik P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten*, 151, 95-100.
- Vere-Jones D. and Schoenberg F. P. (2004). Rescaling marked point processes. *Australian & New Zealand Journal of Statistics*, 46 (1): 133-143.
- Zheng X. and Vere-Jones D. (1991) Application of stress release models to historical earthquakes from North China. *Pure and Applied Geophysics*, 135, 559-576.
- Zheng X. and Vere-Jones D. (1994) Further application of the stress release model to historical earthquake data. *Tectonophysics*, 229, 101-121.

Zhuang J., Ogata Y. and Vere-Jones D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97: 369-380.

Zhuang J., Ogata Y., Vere-Jones D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109, No. B5, B05301, doi:10.1029/2003JB002879.

Zhuang J., Chang C.-P., Ogata Y. and Chen Y.-I. (2005). A study on the background and clustering seismicity in the Taiwan region by using a point-process model. *Journal of Geophysical Research*, in press.

REGISTRATION
of
Research Memorandum

(month) (day) (year)
Research Memorandum NO. 941; Received on 06, 07, 2005
by

Center for Information on Statistical Sciences
the Institute of Statistical Mathematics

(*Applicant:*
Jianchang Zhuang
phone: 03-5421-8780, email: zhuangjc@ism.ac.jp)

.....

Title: Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data
Author(Affiliation): Zhuang, Jianchang (Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minato-Ku, Tokyo)
Key words: residual analysis: goodness-of-fit: point process: ETAS model: earthquake
Abstract: This paper proposes the principle of the second-order martingale-based residual analysis for space-time point processes. In order to illustrate the powers of the first- and second-order residuals, I apply them to the space-time epidemic-type aftershock sequence (ETAS) model for testing some important and interesting hypotheses associated with the features of earthquake clusters.